

2017

Dynamical Typology of Sea Surface Temperature (SST) Fronts Based on Satellite Observations

Yackar L. Mauzole
University of Rhode Island, yackar.mauzole@gmail.com

Follow this and additional works at: https://digitalcommons.uri.edu/oa_diss

Terms of Use

All rights reserved under copyright.

Recommended Citation

Mauzole, Yackar L., "Dynamical Typology of Sea Surface Temperature (SST) Fronts Based on Satellite Observations" (2017). *Open Access Dissertations*. Paper 592.
https://digitalcommons.uri.edu/oa_diss/592

This Dissertation is brought to you by the University of Rhode Island. It has been accepted for inclusion in Open Access Dissertations by an authorized administrator of DigitalCommons@URI. For more information, please contact digitalcommons-group@uri.edu. For permission to reuse copyrighted content, contact the author directly.

DYNAMICAL TYPOLOGY OF SEA SURFACE TEMPERATURE (SST)
FRONTS BASED ON SATELLITE OBSERVATIONS

BY

YACKAR L. MAUZOLE

A DISSERTATION SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
IN
OCEANOGRAPHY

UNIVERSITY OF RHODE ISLAND

2017

DOCTOR OF PHILOSOPHY DISSERTATION
OF
YACKAR L. MAUZOLE

APPROVED:

Dissertation Committee:

Major Professor Peter Cornillon

Melissa Omand

Gavino Puggioni

Nasser H. Zawia

DEAN OF THE GRADUATE SCHOOL

UNIVERSITY OF RHODE ISLAND

2017

ABSTRACT

This dissertation presents a two-part study of persistent sea surface temperature (SST) fronts based on global satellite observations. Its main goal is to achieve a better description of these features at the global scale, while reducing the bias in defining them. To this end, an algorithm was developed to detect and track persistent fronts by relying on frontal probabilities. Frontal probability (or frontal frequency) can be defined as the probability to observe a front relative to the cloud cover. Long-term mean (29 years) and seasonally averaged frequency maps were produced, showing the location of over 2,000 SST fronts for each map. The algorithm, denoted automated detection algorithm or PFDA, is based on a simple concept, which is the determination of local maxima of frontal probability by scanning the fields along lines of constant latitude and of constant longitude. The ADA agrees well with satellite observations, and performs well in detecting fronts with complex shapes or western boundary currents. This method can be applied to other types of datasets, such as chlorophyll fronts, and presents the crucial advantage to remove subjectivity in finding fronts. Frontal properties, derived from the Cayula-Cornillon algorithm were added to the geographic coordinates of the fronts. Furthermore, the bathymetry and bathymetry gradient magnitude are also part of the information provided by the PFDA at each frontal pixel.

Subsequently, we applied multivariate statistical analysis tools to evaluate the presence of patterns among persistent fronts. We performed a principal component analysis, followed by a k-means clustering procedure to partition fronts into different types. Unsupervised machine learning applied to the problem of oceanic fronts resulted in nine clusters, namely four clusters corresponding to shelf and shelf-break fronts, two clusters representing the subpolar frontal system, one matching the Kuroshio and Gulf Stream extensions, and the two last ones group-

ing the boundary currents and upwelling fronts respectively. The partitioning was conducted based on the characteristics of the fronts, in an objective manner. Because this study was performed globally, we gained significant insights on poorly-documented and newly-found SST fronts, but also on fronts that were previously studied. These results also shifts the traditional way of understanding oceanic fronts. In particular, it is interesting to see coastal and equatorial upwelling fronts in the same cluster, just as it is intriguing to observe western and eastern boundary currents being part of one unique cluster. While this study is mostly focused on the statistical description of persistent SST fronts, its findings will certainly bring a broader understanding of ocean's submesoscale dynamics, and is expected to benefit the Oceanography community in the future.

ACKNOWLEDGMENTS

This dissertation is the print left after seven years spent at the Graduate School of the University of Rhode Island. While I did not know where the path would lead to, I embarked on the journey of learning about Oceanography with joy and excitement. I am grateful to have met so many wonderful people here, expanded my world views and learn so much more about myself and what I am capable of.

First, I would like to thank my advisor Peter Cornillon, for taking me with him to work on my PhD three years ago. Without him, this work would not have been possible. I appreciate our bilingual conversations, switching between English and French, and I am grateful for his humor, his insights, his patience and his support over the last few years. I also appreciate how formative my years with him were in terms of personal development: I can look back now and see how much he contributed to my growth as an individual and as a scientist in particular. I am also grateful to Lew Rothstein, for all our discussions around a beer, and his continuous encouragements. Whenever I doubted myself, Lew was never far to push me and tell me I could do it. I would like to acknowledge my committee members, Melissa Omand and Gavino Puggioni for their advice and their support throughout my PhD. More specifically, their contributions were essential to the development of the second chapter of this thesis. Brian Heikes played a key role in my graduate student career here, by "adopting" me for a semester after I completed my Masters and before I started my PhD. I had a great time working with him on Atmospheric Chemistry, and learned a lot thanks to him. I am grateful to have had support from Meredith and David, but also Lisa Conte, all helping me survive the administrative ordeals I had to face during my time at URI, especially as a

foreign student. I know they will be relieved to see me graduate and not tracking them anymore on campus to get yet another form signed. I would also mention the emotional support provided by Cory over the last few years: I have always enjoyed our conversations and making you laugh.

I would like to thank the Grillis, and in particular Annette for being there for me during the hardest times of Graduate School. I have always looked forward to our lunches together, and they will remain a fond memory of my time in Rhode Island. I am not someone who is very social, still I managed to make lifelong friendships here, which definitely helped me through these seven years abroad. While some of my friends are now gone on their own paths, I would like to acknowledge their influence and impact on my life. Lina, Kaoru, Aga and Margaret: you are all great people, and precious friends to me. I clearly remember the taste of the bakhlava and sage tea, as well as our long discussions and I hope we will be able to all meet together again in the near future. On the Bay Campus itself, I would like to thank my friends Christina (for all her Matlab help, and her smile), Clara (for always encouraging me and pushing me) and Victoria (for keeping me alive). I hope we keep our bond beyond Graduate School and well into our lives.

I know, even though I cannot prove it, that Victoria and my mother have made the tacit promise to get me through these last months of PhD and keep me alive at all costs. I truly appreciate that. My mother has been instrumental in me (1) going for a PhD abroad and (2) supporting me through my studies here. Cousins on both sides of the family have encouraged me and believed in me. Delphine, Fatou and Lycka: thank you! I am grateful for my aunts to have never doubted I could make it through: they would always welcome me back home with all my favorite foods on the table and lots of love.

Even though it has been seven years, I realize that I am here because of two of my

former professors: Patrick Huerre from Ecole Polytechnique and Stéphan Fauve from the Ecole Normale Supérieure in Paris. Without them being on my path, I would have never known about Fluid Mechanics and Oceanography. They are both important people, of the kind you meet at crossroads of your life and can change its course forever. I thank you for giving me a chance and for being great teachers.

Lastly, I would like to acknowledge my grandmother, who always believed in the power of education and who always pushed me to never stop learning from the world and from others.

PREFACE

The dissertation is written in manuscript format and consists of two manuscripts presented in separate chapters.

The first one will be submitted to the Journal of Geophysical Research and the second will be submitted to Geophysical Research Letters for publication.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGMENTS	iv
PREFACE	vii
TABLE OF CONTENTS	viii
LIST OF FIGURES	xi
LIST OF TABLES	xiii
MANUSCRIPT	
1 An automated method to track major persistent SST fronts	1
1.1 Introduction	2
1.2 Data and Methods	3
1.2.1 Data	3
1.2.2 CCA	4
1.2.3 Frontal probability	5
1.2.4 Persistent Front Detection Algorithm (Persistent Front Detection Algorithm (PFDA))	6
1.3 Results	10
1.3.1 Persistent fronts - long term mean	11
1.3.2 Seasonal persistent fronts	17
1.3.3 Relation to bathymetry	20
1.4 Discussion	22
1.4.1 New findings and implications	22

	Page
1.4.2 Limitations of the PFDA and possible improvements . . .	23
1.5 Conclusion	27
Bibliography	29
2 Global typology of persistent SST fronts	31
2.1 Introduction	32
2.2 Data and Methods	32
2.2.1 Data	32
2.2.2 k-means clustering	35
2.3 Results	37
2.3.1 Correlations	37
2.3.2 Cluster analysis	38
2.3.3 Summary	51
2.4 Discussion	52
2.4.1 Clustering commentary	53
2.4.2 Future improvements	57
2.5 Conclusion	59
Bibliography	60
 APPENDIX	
A Algorithm details	62
A.1 Validation of CCA	62
A.2 Automated detection algorithm: noise reduction methods	63
B Global maps of persistent SST fronts	66

B.1 Acronyms	71
------------------------	----

LIST OF FIGURES

Figure		Page
1	29-year averaged global frontal probability.	5
2	Flowchart of the automated detection method (ADA).	7
3	Automated detection process.	8
4	Frontal probability and persistent fronts in the Northwest Atlantic	13
5	Frontal probability and persistent fronts in European and Eurasian Seas.	13
6	Persistent sea surface temperature (SST) fronts in the Mediter- ranean Sea.	15
7	Mediterranean Sea fronts, from Belkin et al., 2009.	15
8	Surface geostrophic currents in the Mediterranean Sea, from Poulain et al., 2012	16
9	Seasonal persistent fronts in East Asia.	19
10	Correlation between the bathymetry and the frontal probability	21
11	SST fronts in Patagonia, found by the PFDA (left) and drawn by Belkin (from Belkin et al., 2009, right).	23
12	Coastal upwelling zone delimitation in the California current system.	26
1	Skewness of the SST gradient.	35
2	Davies-Bouldin index.	37
3	Correlation matrix. FP: frontal probability, $\partial_x T$: zonal SST gradient, $\partial_y T$: meridional SST gradient, ∇T : SST gradient magnitude, ΔT : SST cross-frontal step, H: bathymetry, ∇H : bathymetry gradient magnitude.	38
4	Global map of persistent SST fronts clusters.	43

Figure	Page
5	Frontal probability distribution by cluster. 44
6	Zonal SST gradient distribution by cluster. 45
7	Meridional SST gradient distribution by cluster. 46
8	SST gradient magnitude distribution by cluster. 47
9	SST step distribution by cluster. 48
10	Bathymetry distribution by cluster. 49
11	Bathymetry gradient distribution by cluster. 50
12	Tidal energy dissipation, from NASA. 54
13	Antarctic polar frontal system, from Moore et al. (1999) 55
14	SST gradient and SST step off Africa's coast. 56
15	Meridional SST gradient (K/km) for Western Boundary Currents. 57
A.1	Local maxima after the first pass (left) and the second pass (right). 63
B.2	Global map of persistent SST fronts for Winter, based on AVHRR and MODIS 67
B.3	Global map of persistent SST fronts for Spring, based on AVHRR and MODIS 68
B.4	Global map of persistent SST fronts for Summer, based on AVHRR and MODIS 69
B.5	Global map of persistent SST fronts for Fall, based on AVHRR and MODIS 70

LIST OF TABLES

Table		Page
1	Frontal Pixel Properties	10
2	Large Marine Ecosystems, from Belkin et al., 2009	12
3	Frontal clusters by type.	53

MANUSCRIPT 1

An automated method to track major persistent SST fronts

by

Yackar L. Mauzole ¹, Peter C. Cornillon ²

In preparation for submission to Journal of Geophysical Research

¹PhD candidate, Graduate School of Oceanography, The University of Rhode Island, Narragansett RI 02882. Email: yackar_mauzole@my.uri.com

²Prof of Oceanography, Graduate School of Oceanography, The University of Rhode Island, Narragansett RI 02882. Email: pcornillon@me.com

1.1 Introduction

Fronts are ubiquitous features found in the ocean. They can be thought of as narrow boundaries separating water masses with distinct properties. In the case of sea surface temperature (SST) fronts, they separate waters with different sea surface temperature. SST fronts fall into roughly four groups:

1. *Individual fronts, those that form and decay over short periods and are, in general, of relatively small extent and occur as singletons.* Examples of such fronts are those associated with local meteorological forcing. The probability of finding such a front at a given location is low.
2. *Fronts that occur at the same location for long periods of time.* An example of fronts in this group are shelf break fronts. These fronts are generally tied to bathymetry and they cover a broad range of spatial scales depending on the bathymetric feature with which they are associated. The probability of finding such a front at a given location is high.
3. *Fronts delineating currents with large spatial scales.* These tend to exist over long scales, persist for long times but move laterally, hence, the probability of finding such a front at a given location can be relatively small. The currents, which define these fronts, are generally found on a planetary scale such as western boundary currents.
4. *Individual fronts, as in 1 above but, which occur regularly over a given region.* These are characterized, for example, by fronts in upwelling zones. Such fronts tend to be short lived and of relatively small spatial scale but they occur often in an extended region resulting in an overall increased probability of finding a front in the region.

In this study, we are concerned with fronts that can be observed over relatively

long periods of time at approximately the same location; i.e., primarily fronts in group 2, but on occasion those in groups 3 and 4. Hereafter we call these fronts *persistent*, and our efforts are focused on detecting and mapping the major persistent SST fronts found in the global ocean. This study is based on several satellite datasets available from 1982 to 2011, and we derived five global maps of persistent fronts: one long-term mean and four seasonal maps, to account for the seasonality of persistent SST fronts. In the following section, we detail the satellite datasets used for this study and describe the principles of our method. Results are presented in section 3, in which we examine the performance of the algorithm in different scenarios. The major features of the algorithm and its limitations are discussed in section 4, followed by the conclusions in section 5. Appendix A provides additional information regarding the method to automatically locate persistent fronts, while Appendix B shows the global maps of persistent SST fronts.

1.2 Data and Methods

1.2.1 Data

In this study, two types of maps were produced based on satellite observations: a first map was based on Pathfinder Advanced Very High Resolution Radiometer (AVHRR) data alone (Kilpatrick et al., 2001), while the second type, consisting of four maps corresponding to the climatological seasons, rely on Pathfinder as well as MODerate-resolution Imaging Spectroradiometer (MODIS) Terra and Aqua data. The motivation behind using data from both MODIS sensors lies in the fact that Pathfinder fields show limited coverage for the climatological seasons: particularly in Winter, where cloud cover is significant, using data from three satellites instead of one improves the mapping of persistent SST fronts. From our data holdings at the University of Rhode Island (URI), about 30 years of Pathfinder measurements were stored, from 1982 through 2011 with the exclusion of 1995, for

which there was no data. Regarding both MODIS satellites, we chose a time period for which all three datasets were overlapping, i.e. from 2003 to 2011 included. For both Aqua and Terra satellites, we chose the SST4 product, which is derived from measurements in the $4\mu\text{m}$ mid-wave infrared atmospheric window (Kilpatrick et. al, 2015). Within the framework of this study, we used all Level 3 global SST nighttime fields available between 1982 and 2011, about 10,600 Pathfinder fields and slightly more than 6,500 MODIS fields. We excluded daytime fields from our analysis to avoid the bias introduced by diurnal warming. The region of study, defined by a Plate-Carré projection of $0.0417^\circ \times 0.0417^\circ$ pixels (roughly 4 km square on the Equator), extends from 60°N to 60°S

1.2.2 CCA

In contrast to gradient based methods of front detection, the Cayula-Cornillon Algorithm (CCA) is based on a histogram analysis of the SST fields. In the interest of brevity, herein we summarize the principles underlying the CCA. For a detailed description of the algorithm, the interested reader is referred to the original papers (Cayula and Cornillon, 1992; Cayula and Cornillon, 1995; Cayula and Cornillon, 1996), which describe it in detail.

The algorithm relies on three levels of processing: the image level, the window level and the pixel level. Cloud detection and cloud removal are performed at the picture level, while front detection part is primarily done at the window and pixel levels. The CCA examines histograms of the temperature values in 32×32 pixel windows: if the histogram is unimodal, it suggests that there is only one SST population, i.e. the water mass sampled is unique and has a uniform temperature. If however, the distribution is bimodal, this implies that there are two water masses with distinct SST signatures. In this case, the CCA will look for pixels in population A with a pixel to the right or below belonging to population B. These pixels correspond

to the location of a potential front separating the water masses of different SST . At the local level, a contour following algorithm is implemented in order to connect candidate frontal pixels into continuous frontal segments whenever possible. Frontal properties, such as SST gradient and cross-frontal step, are computed and associated with each frontal pixel in addition to their geographic coordinates.

1.2.3 Frontal probability

Major persistent SST fronts, the object of this study, were not detected directly from SST fields. Instead, we relied on maps of frontal probability or frontal frequency. The frontal probability (F) of a space-time region is a measure of the number of pixels in that region flagged as frontal pixels (N) over the number of pixels flagged as clear (C): $F = N/C \cdot 100\%$. In the work presented herein the spatial portion of the space-time region is generally a single pixel, however this need not be the case. Fig. 1 shows the frontal probability for the Pathfinder dataset for the time period 1982-2011.

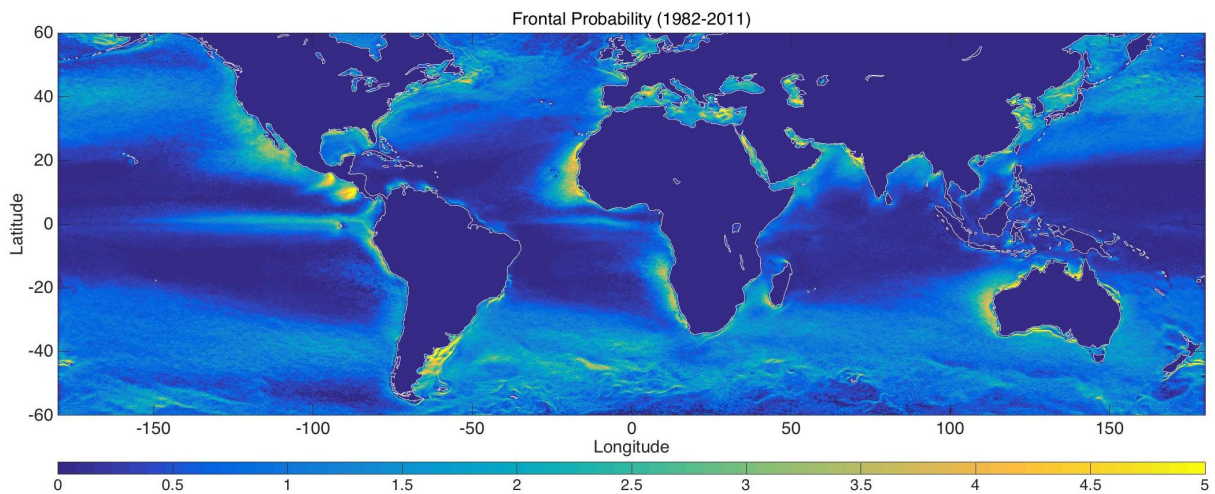


Figure 1: 29-year averaged global frontal probability based on AVHRR Pathfinder 4km SST dataset (1982-2011), processed with the Cayula-Cornillon algorithm.

1.2.4 Persistent Front Detection Algorithm (Persistent Front Detection Algorithm (PFDA))

In this section, we provide a brief overview of the algorithm used to identify persistent frontal locations in fields of frontal probability. The algorithm consists of two major parts. The first part relates to the geometry of persistent frontal regions – the one pixel wide line defining the location of the front and the width of the frontal region. The second part of the algorithm determines properties such as SST gradient, SST step, etc. associated with fronts in the region. The details of the algorithm, the flowchart of which is presented in Fig. 2, can be found in Appendix A. The objective of the algorithm was to reduce subjectivity in establishing maps of SST fronts, and as such involves minimal human intervention. The first step is to locate local maxima of frontal probability by scanning the image both vertically and horizontally. The output of this step consists of binary arrays with 1 for frontal pixels, 0 otherwise. Morphological operations such as dilatation and thinning are then performed in order to connect the peaks (discrete points) into continuous lines whenever possible; this corresponds to the fourth step in the flowchart. The frontal segments indexing is the following step of the ADA: due to Matlab indexing protocol which proceeds in decreasing latitudes, it is necessary to rearrange the pixels of the fronts into monotonic series (step 6). Several noise reductions methods were implemented (shown in the diamond-shaped elements in figure 2) to decrease the noise and remove undesirable points from the fields of local maxima. The first one consists of looking at the properties of the local peaks (relative height and width), in order to discriminate stable fronts from background noise. The second operation aims to decrease the number of spurs, short (several pixel) segments attached to longer segments.

Another way of understanding how the first portion of the PFDA (the portion associated with the geometry of the persistent fronts) functions is presented in

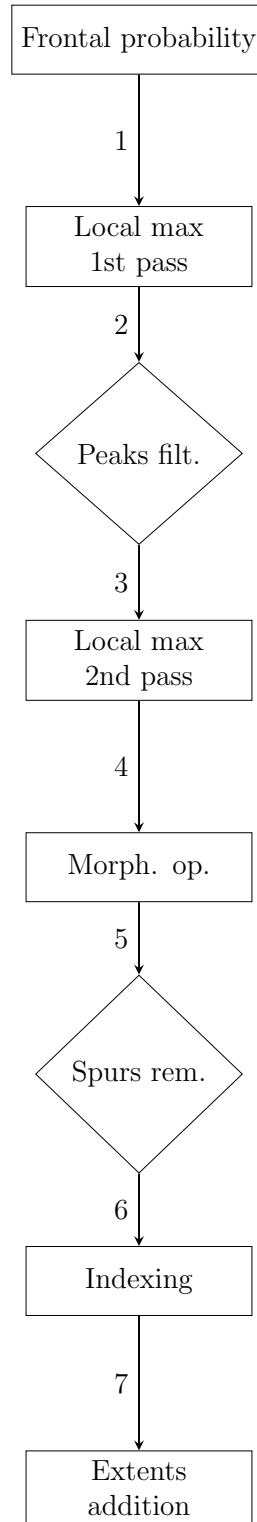


Figure 2: Flowchart of the automated detection method (PFDA). Starting from a frontal probability field, local maxima are detected (step 1) and filtered (step 2) to derive a map of local maxima with less noise (step 3). Morphological operations (dilatation and thinning) are performed to generate continuous lines (step 4). Spurs are removed (step 5) and the pixels of each line are reordered and indexed (step 6). Lastly, extents to the frontal segments are added (step 7).

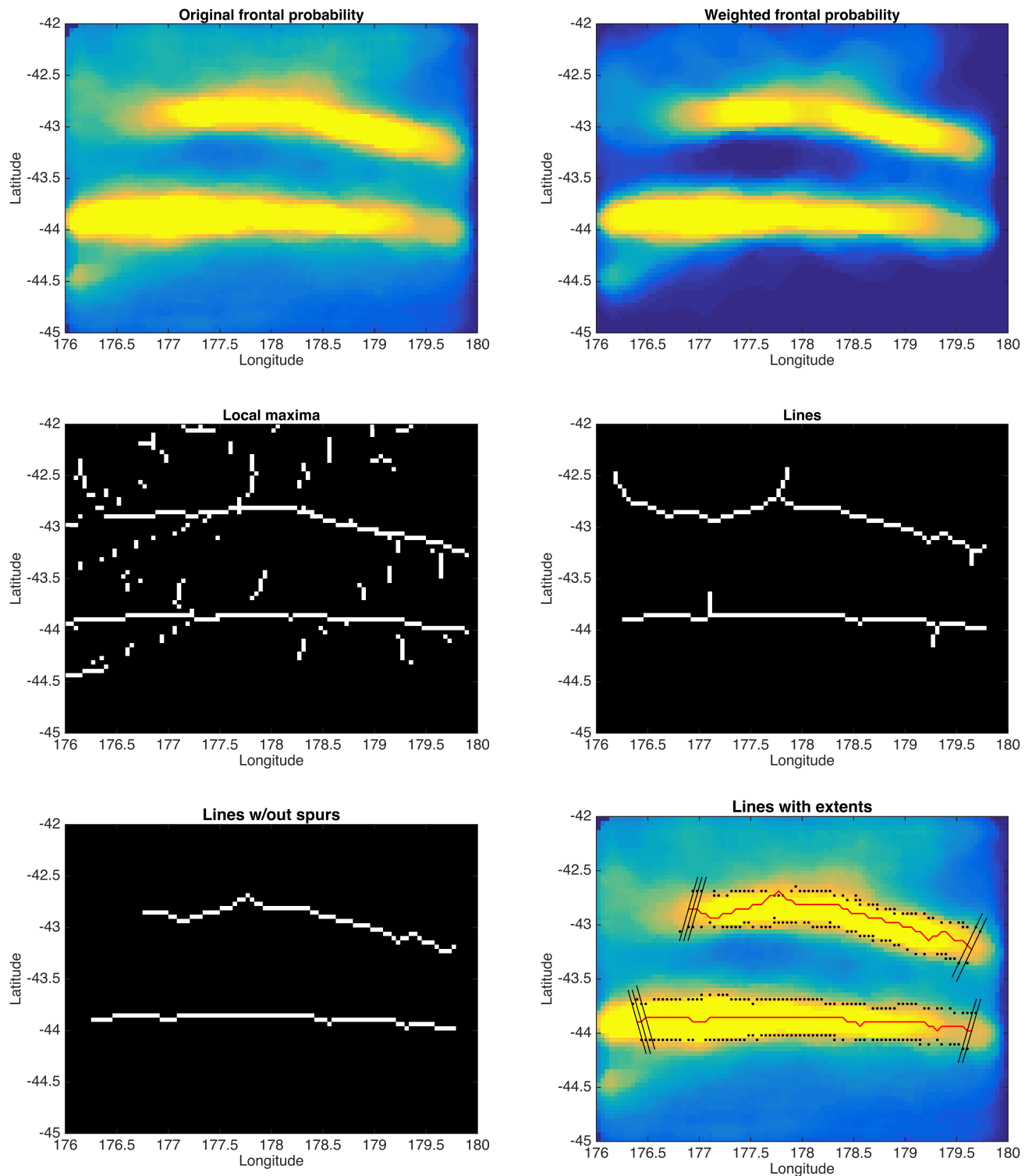


Figure 3: Automated detection process. The frontal probability field (upper left) is weighted to reduce noise (upper right). Local maxima are detected (middle left), and connected through morphological operations (middle right). Spurs are removed (bottom left) and pixels of each segment are indexed (not shown). The extents of the fronts (black dots) are computed by looking at the frontal probability along the local normals (black lines, only plotted for the extremities of the fronts) and saving the locations of the half-prominence of the probability peaks.

Fig. 3. The PFDA begins with the frontal probability field obtained from the 3×3 median filtered SST field (upper left). The weighted frontal probability (upper right) accentuates the difference between high probability features and the surrounding lower probability background and is the basis of the subsequent panels in Fig. 3: the local maxima (after filtering) are presented in the middle left panel. The frequency peaks are then connected into lines, which results in some spurs appearing on the two segments shown in the middle right panel. The spurs are removed in the bottom left panel. This defines the location of pixels along persistent frontal segments. The next step is to determine the locations of the outer edges of the persistent frontal region, the black dots in the lower right panel of Fig. 3. These are based on the probability distribution. The outer edge on a given side is defined by the point, on the line locally normal to the frontal segment, at which the probability is one half of the peak value plus the probability of the first local minimum on that side.

The second major section of the PFDA is the determination of a suite of properties (listed in Table 1) for each persistent frontal pixel, based, in general, on other data sources. The properties are the average values of the parameter over the width of the persistent front; i.e., along the local normal at the persistent frontal pixel of interest from one side of the persistent frontal region to the other. This allows for the consideration of the fronts' properties not just along the persistent front, but over the width of the frontal region at that location as well. Table 1 also lists the source of the data from which the parameters were obtained. *Front probability* is based on the same distributions used to determine the location of persistent fronts. The *bathymetric gradients* and *water depth* were obtained from ETOPO1; the gradient fields by convolving the meridional and zonal Sobel kernel with the bathymetric field. The remaining parameters were obtained from the Pathfinder

and MODIS datasets cited in section 1.2.1. All of the associated SST fields were passed through the URI fronts/gradients workflow used to locate fronts with the Cayula-Cornillon front detection algorithm. This resulted in three fields for each input SST field: a zonal gradient field, a meridional gradient field and a frontal field (with 1s for frontal pixels and missing values otherwise). As with the bathymetric field, the SST vector gradient was obtained with the Sobel gradient operator. Each of the gradient fields was then multiplied by the frontal field resulting in a gradient value where there was a frontal pixel and missing values otherwise. Finally, these fields were averaged, excluding missing values, over the period of interest (seasonal or over the total dataset).

Table 1: Frontal Pixel Properties

Property	Source
Front probability	Probability distribution
Bathymetric gradient magnitude	ETOPO 1
Water depth	ETOPO 1
Meridional SST gradient	http://sstfronts.org/opendap/
Zonal SST gradient	http://sstfronts.org/opendap/
SST gradient magnitude	http://sstfronts.org/opendap/
SST step	http://sstfronts.org/opendap/
Chlorophyll-a concentration	http://sstfronts.org/opendap/

1.3 Results

In this section, we first present the SST fronts located in selected regions from the Pathfinder dataset only; i.e., obtained from the 30 year averaged frontal probability field. We then show the results of the PFDA for seasonal SST fronts, using a combination of Pathfinder and MODIS datasets. Assuming that persistent fronts are highly correlated with bottom topography (Csanady, 1978; Hsueh and Cushman-Roisin, 1983), we propose here three distinct scenarios, corresponding to three different regions:

- a simple bathymetry, and a western boundary current;

- a complex bathymetry, and several marginal and enclosed seas;
- a complex bathymetry, marginal seas and a western boundary current.

The goal is to present how the PFDA fares in various situations and observe the limitations of the algorithm, in particular in the instance of well-known meandering western boundary currents, or when the fronts' profiles are not linear. The first case corresponds to the Northwest Atlantic region, including the Gulf of Mexico. The second case describes the region of Europe and its neighboring seas. In particular, we focus on the Mediterranean Sea by evaluating our results against the map of Belkin et al. (2009) for this sea, as well as with other datasets (surface circulation and a satellite-derived chlorophyll field). Lastly, we investigate the performance of the PFDA in mapping seasonal persistent SST fronts for the East Asian region (case 3). The global maps for the long-term mean and the climatological seasons can be found in Appendix B. For simplicity and consistency with the existing literature, we divide our global domain into the large marine ecosystems (LMEs) delimited in Belkin et al. (2009). The numbers in parentheses indicate the index of the corresponding LME as attributed by Belkin in Table 2. The Caspian Sea was not part of the original LMEs and is given the symbol (*).

1.3.1 Persistent fronts - long term mean

Northwest Atlantic - Gulf of Mexico Fig. 4 shows the frontal probability for the Northeast Continental U.S., the Southeast Continental U.S., and the Gulf of Mexico (LMEs 7, 6, and 5) along with the persistent fronts found by the algorithm (red lines) and several isobaths. The majority of persistent coastal fronts appear to be stable (with higher frontal probability than many of the other fronts to be discussed), and readily found by the PFDA. More surprisingly, a fraction

of the long-term mean paths of the Loop Current and of the Gulf Stream can also be tracked, even though they are intermittent frontal systems (justifying the fragmented paths in both instances). The western and eastern legs of the Gulf Stream, seen south of Cape Hatteras are well detected by the PFDA, however the remaining track of the Gulf Stream is lost east of 65°W .

Europe, European and Eurasian seas Fig. 5 presents the southern part of the Baltic Sea, the North Sea, the Mediterranean Sea, the Black Sea, the Sea of Azov and the Caspian Sea. We chose this region (LMEs 22 to 26, plus the Eurasian seas * and 62) for the complex patterns formed by the fronts. Indeed, in contrast to the Northwest Atlantic, this region contains several enclosed and marginal seas, in addition to a more convoluted bathymetry. In most seas, the vast majority of the fronts mapped by the algorithm agree well with the frontal probability field.

In the Baltic Sea, most of the fronts previously documented (Belkin et al., 2009) can be seen. In the North Sea, fronts along the coasts of Denmark and Norway can be observed on top of the 25 m and 250 m isobaths, . A front is

Table 2: Large Marine Ecosystems, from Belkin et al., 2009

E. Bering Sea (1)	North Sea (22)	N. Australian Shelf (39)
Gulf of Alaska (2)	Baltic Sea (23)	N.E. Australian Shelf (40)
California Current (3)	Celtic Biscay Shelf (24)	E. Central Australian Shelf (41)
Gulf of California (4)	Iberian Coast (25)	S.E. Australian Shelf (42)
Gulf of Mexico (5)	Mediterranean Sea (26)	S.W. Australian Shelf (43)
S.E. U.S. Continental Shelf (6)	Canary Current (27)	W. Central Australian Shelf (44)
N.E. U.S. Continental Shelf (7)	Guinea Current (28)	N.W. Australian Shelf (45)
Scotian Shelf (8)	Benguela Current (29)	New-Zealand Shelf (46)
Newfoundland/Labrador Shelf (9)	Aghulas Current (30)	E. China Sea (47)
Hawaii (10)	Somali Coastal Current (31)	Yellow Sea (48)
Pacific Central America(11)	Arabian Sea (32)	Kuroshio Current (49)
Caribbean Sea (12)	Red Sea (33)	Sea of Japan (50)
Humboldt Current (13)	Bay of Bengal (34)	Oyashio Current (51)
Patagonian Shelf (14)	Gulf of Thailand (35)	Okhotsk Sea (52)
S. Brazil Shelf (15)	S. China Sea (36)	W. Bearing Sea (53)
E. Brazil Shelf (16)	Sulu-Celebes Sea (37)	Caspian Sea (*)
N. Brazil Shelf (17)	Indonesian Sea (38)	Black Sea and Sea of Azov (62)

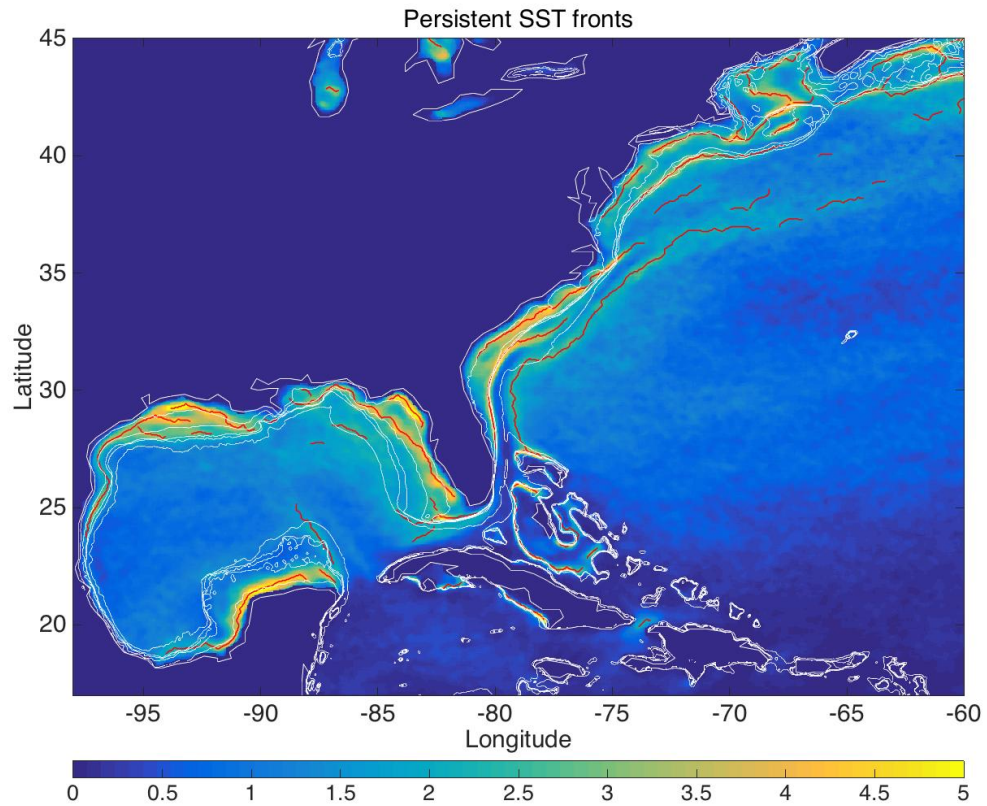


Figure 4: 29-year averaged frontal probability for the Northwest Atlantic and corresponding persistent frontal segments . Red lines: PFDA output; white lines: isobaths at -20m, -50m, -100m, -150m, and -750m.

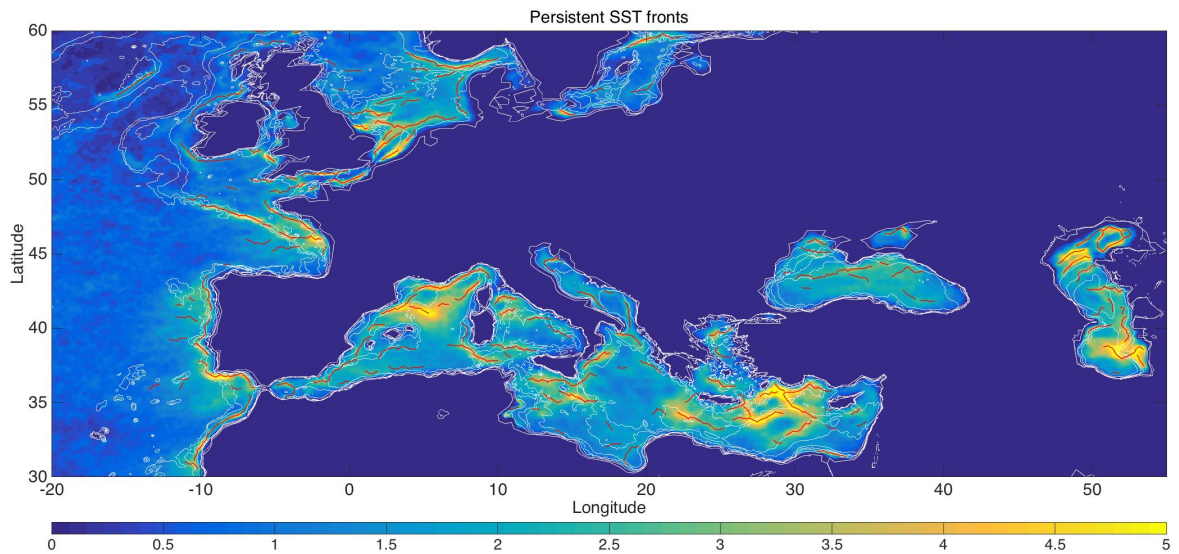


Figure 5: 29-year averaged frontal probability for the European and Eurasian Seas, and corresponding persistent frontal segments . Red lines: PFDA output; white lines: isobaths at -25m, -50m, -75m, -250m, -1500m, and -2500m.

found along Rockall plateau as well, northwest of Ireland. The English Channel fronts were mapped successfully, in contrast to the fronts with lower frontal probability, like the Central front mentioned by Belkin et al. (2009). Neither the Biscay shelf front nor the Ushant front, which are generally considered to be continuous (Belkin et al., 2009), are found to be discontinuous by the PFDA, and the front detected by the algorithm in the vicinity of the Gironde river as well. To the south, fronts are found approximately parallel to and perpendicular to the coastline along the Iberian peninsula. Those perpendicular to the coastline are thought to be upwelling fronts. The northern part of the front associated with the Canary current is detected as one long frontal segment.

Regarding the Mediterranean Sea, we compare our findings, Fig. 6 showing persistent fronts (red lines) superimposed on the bathymetry, with the Fig. IV-7.1 from Belkin et al. (2009), reproduced as Fig. 7 here. Of the 12 fronts reported by Belkin et al., two thirds are evident in Fig. 6, namely the North Balearic front, the Ligurian front, the North Tyrrhenian front, the Sardinia-Sicily front, the North Adriatic front, the Albanian front, the Ionian front, the Crete front and the Cyprus front. With regard to the persistent fronts found by the PFDA but not discussed in Belkin et al., the question naturally arises as to whether they are artifacts or real fronts. We argue the latter in light of their close relationship to surface geostrophic currents described by Poulain et al. (2012). We discuss this relationship in the context of Poulain et al's. Fig. 12 (reproduced here as Fig. 8) from west to east, and from north to south.

Starting at the entrance of the Mediterranean Sea with the Alboran Sea, a stable circular front is observed, associated with the Western Alboran gyre. Fronts are found around the Balearic islands and in the Ligurian Sea, matching the North-

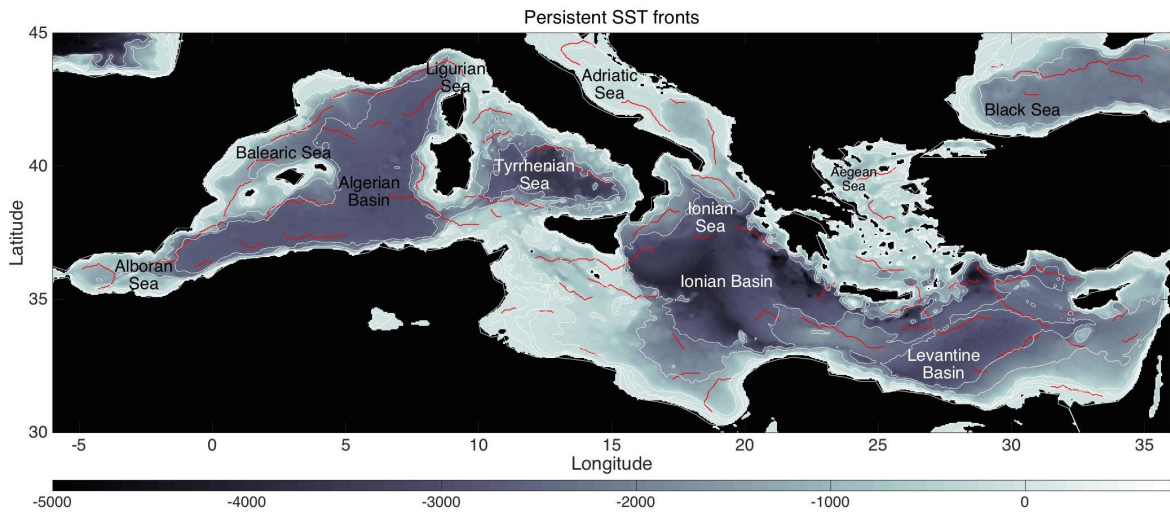


Figure 6: Persistent SST fronts in the Mediterranean Sea, plotted on top of the annotated bathymetry field. The isobaths are the same as in the previous map of Europe.

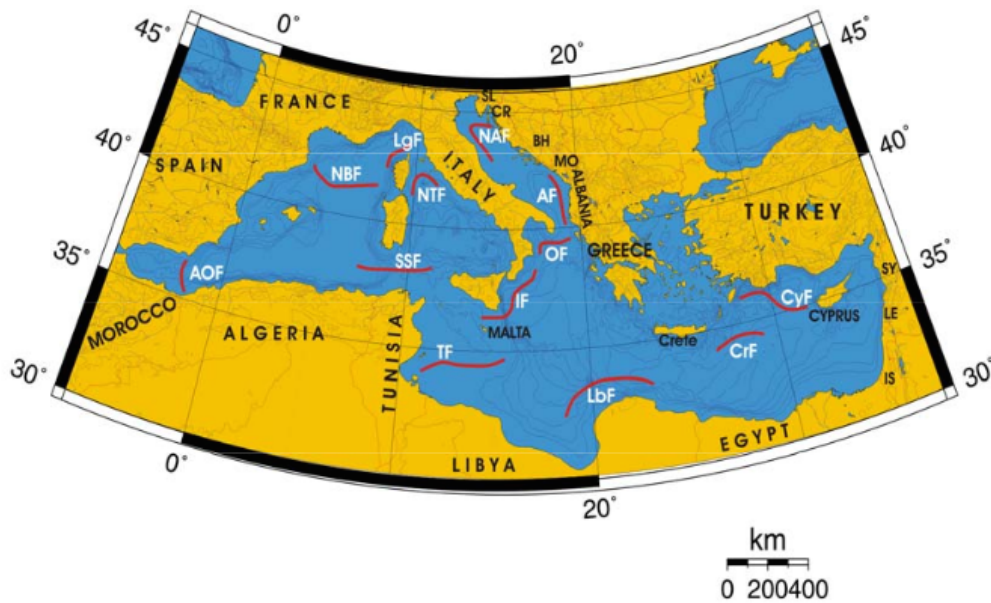


Figure IV-7.1. Fronts of the Mediterranean Sea LME. AF, Albanian Front; AOF, Almeria-Oran Front; CrF, Crete Front; CyF, Cyprus Front; LbF, Libyan Front; LgF, Ligurian Front; NAF, North Adriatic Front; NBF, North Balearic Front; NTF, North Tyrrhenian Front; OF, Otranto Front; SSF, Sardinia-Sicily Front; TF, Tunisian Front. Countries: BH, Bosnia-Herzegovina; CR, Croatia; IS, Israel; LE, Lebanon; MO, Montenegro; SL, Slovenia; SY, Syria. After Belkin et al. 2009).

Figure 7: Mediterranean Sea fronts, from Belkin et al., 2009

ern Current's path. A few frontal segments align with the Algerian Current as well. In the Tyrrhenian Sea, several fronts are present, and they follow the 1500 m and

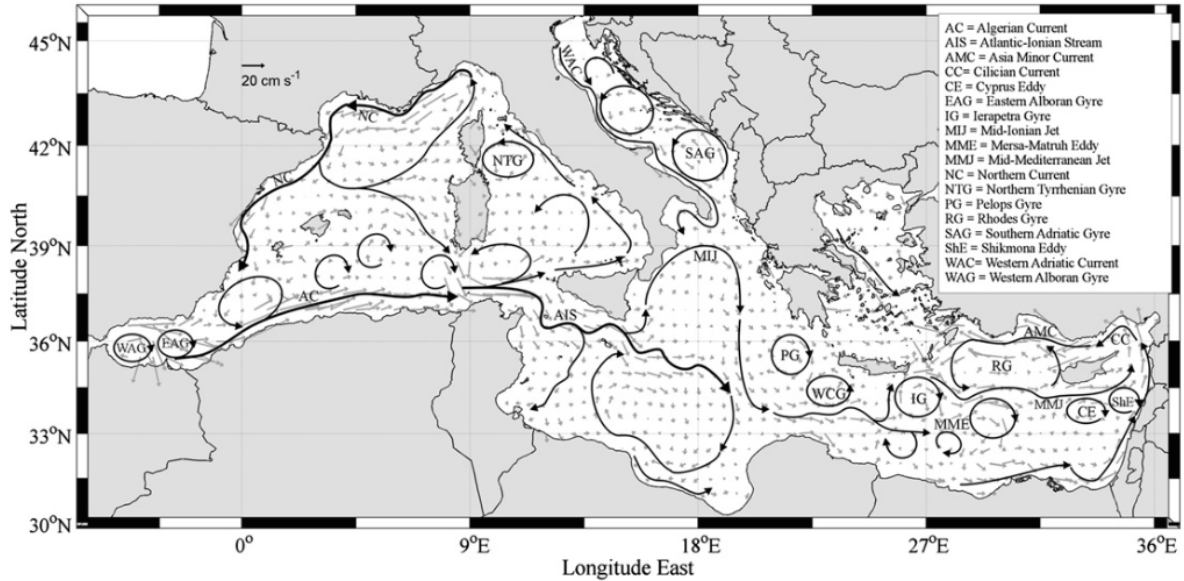


FIG. 12. Schematized representation of the mean surface geostrophic circulation in the Mediterranean Sea in 1992–2010 based on the circulation maps depicted in Figs. 8a,b.

Figure 8: Surface geostrophic currents in the Mediterranean Sea, from Poulain et al., 2012

2500 m isobaths along the Northern Tyrrhenian Gyre in particular. In the Adriatic Sea, several persistent fronts exist: one is associated with the Western Adriatic current, while two others are related to the Northern and Southern Adriatic gyres respectively. The fronts observed south of Sicily and of Italy appear to be related to the Atlantic Ionian stream and the mid-Ionian jet. Several longer fronts are found in the Levantine basin, following closely the 2500 m isobath. Stable frontal activity is observed in the Aegean Sea and fronts found between Crete and Cyprus may be associated with the Rhodes Gyre. Relative to the Ierapetra Gyre, several stable fronts are detected. The front found northwest off Cyprus corresponds to the Asian Minor Current. Lastly, we note that a persistent front is detected off the Nile river along the 25 m isobath.

In their paper, Poulain et al. described the surface geostrophic circulation in the Mediterranean Sea using a combination of drifters and satellite altimetry

data, for the period 1992-2010. Most of the features they observed are collocated with persistent fronts, supporting the idea that stable fronts may be associated to geostrophic currents. Based on the description above, it becomes evident that the PFDA results in a more detailed and more accurate detection of the fronts than previously documented. The surface currents, the bathymetry, and the run-off from the main rivers in the region (Rhone, Po, and Nile) give rise to a complex network of fronts in the region, which are well detected by the PFDA.

Regarding the Eurasian Seas, the Black Sea (Fig. 5) has a moderate frontal probability, which could be explained by a strong seasonal variability (Ginzburg et al., 2004; Karimova, 2014). In fact, the 30-year average does not reflect the stability of the fronts observed in this sea for each climatological season. Because the frontal spatial distribution varies greatly by season, the resulting long-term mean is lower than in the Caspian Sea, and more homogeneous as well. A unique persistent SST front was detected in the Sea of Azov, corroborating the observations of Karimova (2014). The structure of persistent fronts in the Caspian Sea is much better defined than in the Black Sea. In addition, the PFDA frontal patterns in the Caspian Sea align well with the surface circulation patterns described by Ibrayev et al. (2010).

1.3.2 Seasonal persistent fronts

In this section, we focus on the Sea of Japan, the Yellow Sea, the East China Sea as well as the Oyashio and the Kuroshio (LMEs 47 to 51). This corresponds to the last case described earlier; a region with complex bathymetry, marginal seas and a western boundary current. The goal is to evaluate the ability of the algorithm to detect seasonal fronts in this context. Fig. 9 shows the seasonal frontal probability maps for each respective climatological season. As explained

in the previous section, we combined MODIS and Pathfinder datasets to generate seasonal maps, in order to improve the coverage. We define the climatological seasons as follows:

- Winter: January-February-March,
- Spring: April-May-June,
- Summer: July-August-September,
- Fall: October-November-December.

Here, the frontal probability corresponds to the count of frontal pixels over all 29 3-months sets divided by the count of clear pixels found over the same time period for Pathfinder. For both Terra and Aqua, we combined the count of frontal pixels and of clear pixels for the two datasets over the nine year period for which we have MODIS data, resulting in two frontal probability fields for each season, one for Pathfinder and one for MODIS. We applied the PFDA to each map and derived two sets of local maxima, which we then summed. The sum of local maxima is processed through the rest of the algorithm as described in the methods section and Appendix A. In a similar manner to the previous regions, stable fronts obtained with the PFDA are plotted with selected isobaths on the corresponding fields of front probability (Fig. 9).

Planetary currents The long-term mean path of the Oyashio is detected in every season, except Summer where a few short fronts are observed along Japan’s coastline. Regarding the Kuroshio, several parts of the seasonal mean path are readily distinguishable, specifically, the southern branch of the current found between Taiwan and Kyushu, and the northern branch flowing along the southern

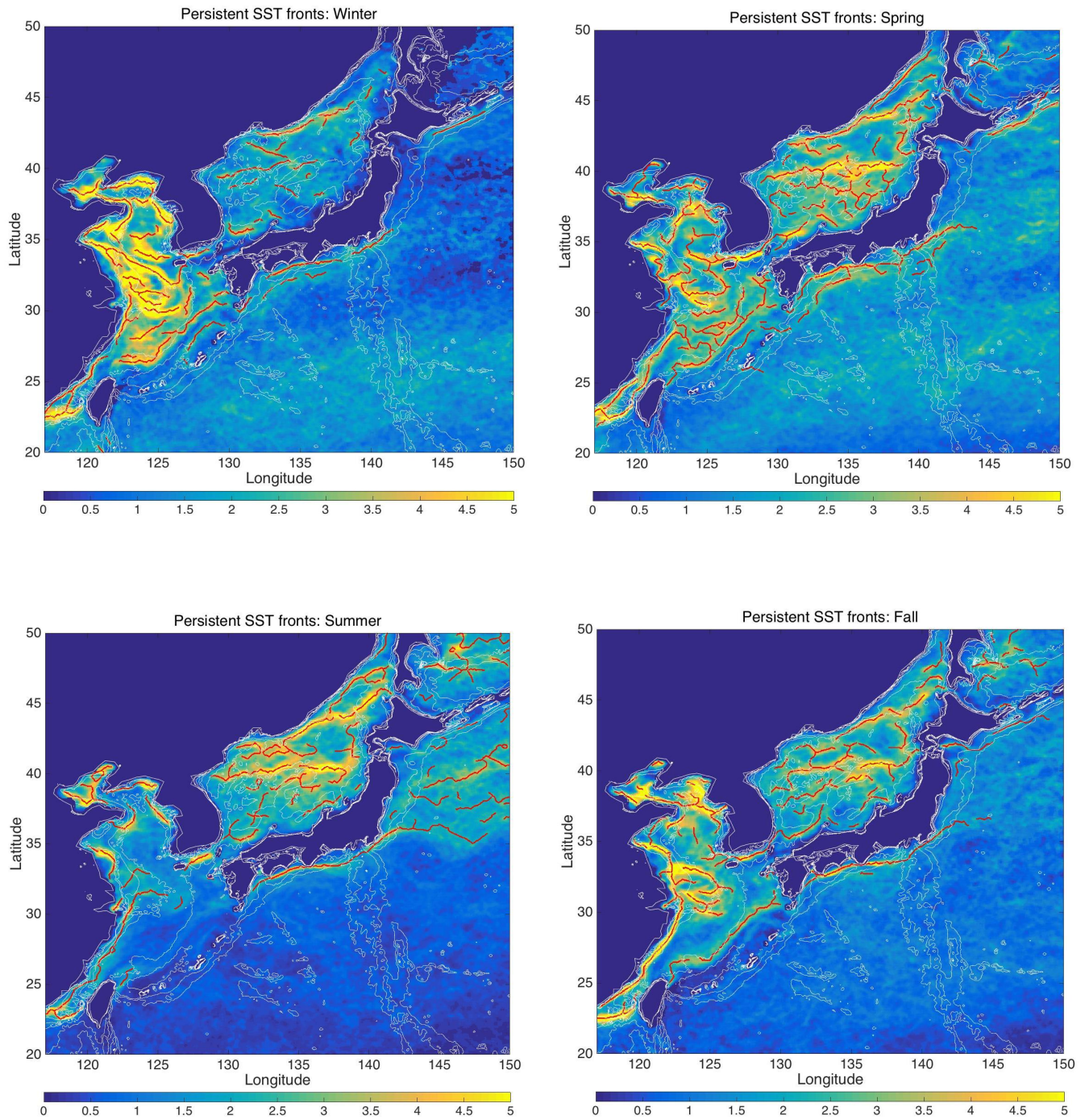


Figure 9: Seasonal persistent fronts in East Asia. Red lines: PFDA output; white lines: isobaths at -25m, -50m, -75m, -250m, -1500m, and -2500m.

coasts of Kyushu, Shikoku and Honshu. The length of the frontal segments corresponding to the western boundary current varies by season. The southern branch is segmented in Winter and Spring, and completely disappears in Summer. It is observed again in Fall as two relatively long segments; the northern branch is at its shortest in Winter (from 133°E to 142°E) and longest in Summer (130°E to 148°E). Interestingly, in contrast to the Oyashio, the Kuroshio does not follow one unique isobath. Besides the main path, frontal segments are also detected by the PFDA in Summer between 35°N and 40°N.

Marginal seas The seasonal variability of the persistent fronts present in the marginal seas (Sea of Japan, Bohai Sea, Yellow Sea and East China Sea) is well captured by the PFDA. Despite a lower frontal probability in Winter, frontal activity was detected in the Sea of Japan. Similarly, in Summer, there are fewer persistent fronts in the Yellow Sea and the East China Sea than during other seasons. In some cases, frontal segments, which we believe should be continuous, are identified by the PFDA as several disconnected segments. We believe that this results from the removal of spurs, which were introduced in the step connecting local maxima into continuous segments. Specifically, the PFDA is able to track, in an objective manner, persistent fronts with a strong seasonal variability, however the strong variability leads in many cases to a larger number of spurs than in the 30-year averaged map. This may in turn alter the results of the algorithm by shortening the original segments. The subpolar front found at 40°N in the summertime Japan Sea is a good example of this.

1.3.3 Relation to bathymetry

We derived the spatial correlation between the bathymetry and the frontal probability, which the determination of the persistent SST fronts relies on. More

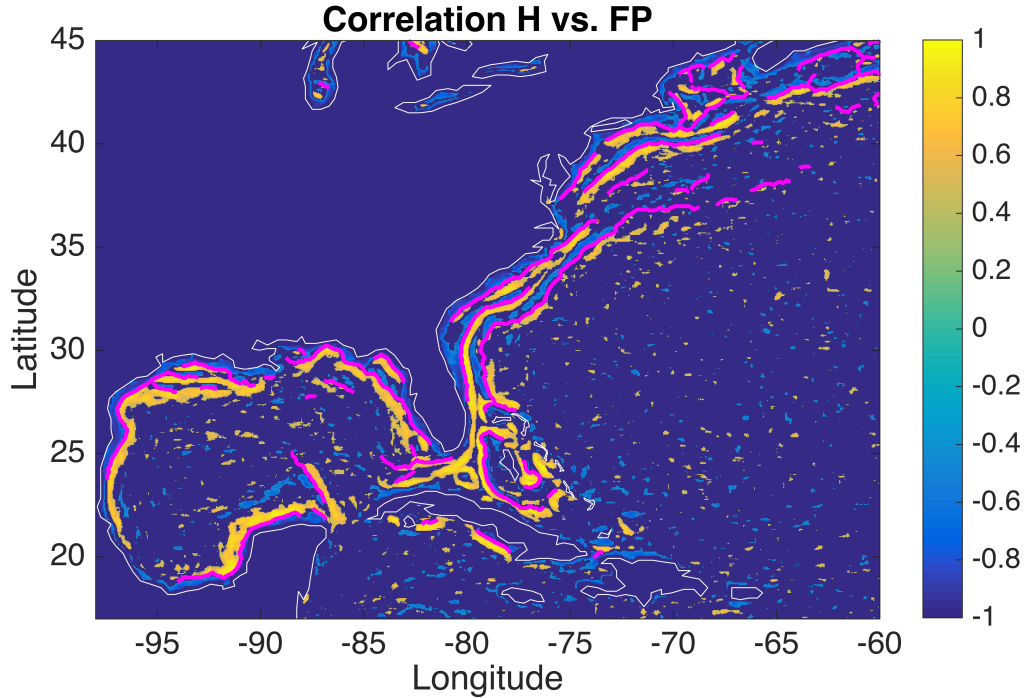


Figure 10: Correlation between the 29-year averaged frontal probability for the Northwest Atlantic and the bathymetry. Values between -0.4 and 0.4 are masked. Magenta lines are the output of the algorithm.

specifically, we computed the Pearson correlation coefficient, defined as follows for the variables X and Y :

$$r = \frac{(X - \bar{X})(Y - \bar{Y})}{\sigma(X) \cdot \sigma(Y)},$$

with σ being the standard deviation, and the overline denoting the mean of X and Y respectively. In our case, the variables are the bathymetry and the long-term mean frontal probability fields. The correlation coefficient was computed over a sliding window of 9 pixels by 9, covering the entire global domain. We present the spatial correlation by masking the values between -0.4 and 0.4, to highlight higher values in figure 10. The magenta lines correspond to the output of the algorithm and are seen often in places where the correlation or anti-correlation is high, indicating that many persistent fronts have a tie to the underlying bathymetry.

1.4 Discussion

1.4.1 New findings and implications

The Persistent Front Detection Algorithm described herein has been applied to the probability distributions of SST fronts associated with three widely used SST products, to produce global fields of persistent SST fronts and their properties. The algorithm finds a significant fraction of the fronts broadly discussed in the literature. In addition, the algorithm finds many persistent fronts not discussed in the literature. This is especially true of regions, typically found in the Southern Hemisphere, although not exclusively. A striking example are the persistent fronts found on the Patagonia shelf (Fig. 11). As for the Mediterranean Sea, a significantly larger number of persistent fronts are found than shown in existing frontal maps of the region. Specifically, small fronts were detected over the continental shelf, between -40°S and -52°S . Many of these fronts are of particular interest in that they tend to be cross-shelf. In some cases, for example, the approximately cross-shelf fronts at 47.5°s , which are not shown in Belkin et al. (2009), follow a cross-shelf topographic features but others do not. One of the reasons that the maps of Belkin et al. may not show these fronts is that their analysis relied on an earlier, 9 km, version of the Pathfinder dataset than the one used in this study.

The newly-found fronts detected by our method provide a more complete picture of the submesoscale dynamics, even on a 30-year averaged field. Smaller stable fronts are observed globally, challenging the notion that shorter fronts have a shorter lifespan than longer ones. These findings could have a profound impact on the understanding of the interactions between biogeochemistry and physics in the ocean, in particular in the context of tracer transport or fluxes estimations at the submesoscale.

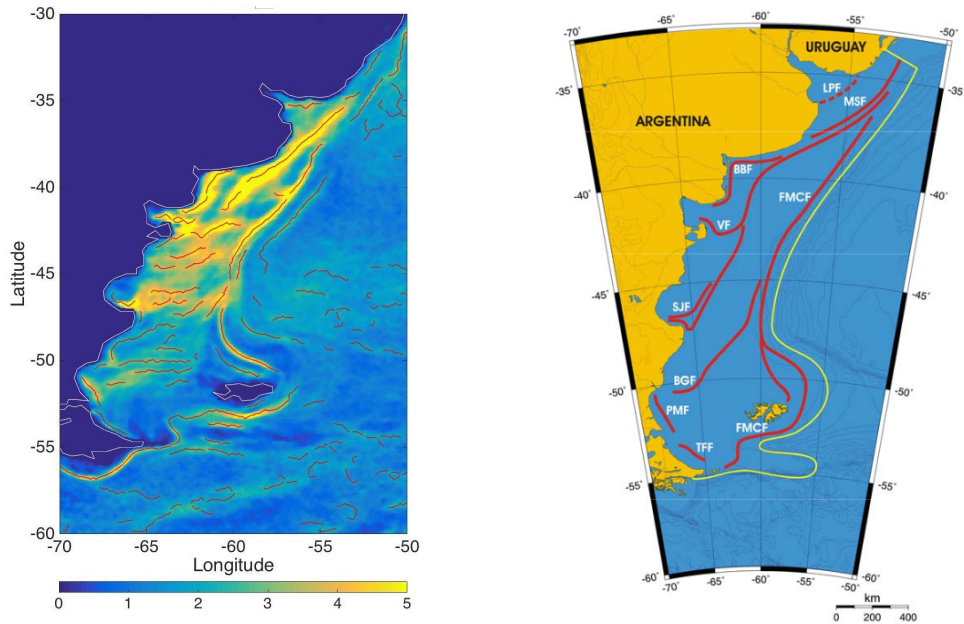


Figure 11: SST fronts in Patagonia, found by the PFDA (left) and drawn by Belkin (from Belkin et al., 2009, right).

1.4.2 Limitations of the PFDA and possible improvements

While the PFDA appears to perform well in general, there are a few areas of concern. In this section, we discuss four conditions for which the algorithm appears to underperform and we propose improvements in three of the cases. The first two cases relate to fronts in category 2 defined in the Introduction; i.e., fronts occurring in the same location for long periods of time. The third case relates to upwelling fronts, those in Category 4 and the final case relates to fronts associated with western boundary currents, Category 3. Although we have made a distinction here between problems associated with upwelling fronts and those with western boundary currents, the distinction is weak in that western boundary currents are often associated with other fronts in the region – these regions tend to be dynamically active hence fronts related to other features also occur there, tending to broaden the probability distribution in addition to the broadening associated with the meandering of the front. Furthermore, western boundary currents are only on

example of fronts in Group 3, other fronts, such as eastern boundary currents, also fall in Group 3 and may cause similar problems for the algorithm.

Short frontal segments In some cases the presence of numerous spurs occurring after joining local maxima may result in a well defined region of high frontal probability not being identified as a persistent front. The high probability ridge along the coasts of Belgium and the Netherlands (Fig. 5) is a case in point. The PFDA identifies the two ends of this ridge as persistent fronts, but misses the middle portion. In order to remove some of the undesirable segments, we included a threshold on the minimal length of the persistent fronts; i.e., segments with less than 5 pixels were removed from the subsequent analysis. In the case of the English Channel front, the local maxima formed two long series of points (which resulted in the two persistent frontal segments (red lines) shown in Fig. 5), separated by a few very short series of points in the middle of the front (which are missing from the final output of the ADA). In order to remediate this problem, a similar “reconstructing” algorithm to the one suggested above could be used to join frontal segments which are likely to be continuous. Again, caution must be taken to not reintroduce noise eliminated in the current version of the algorithm.

Upwelling fronts As discussed in the Introduction, SST fronts may be divided into four general groups. One of the traits, which characterize these groups, relates to the associated probability distributions. From this perspective, SST fronts in Group 1 simply contribute to the very weak background in frontal probability, while fronts in Groups 2 tend to be found in narrow bands of high frontal probability such as those observed off the east coast of the U.S. (Fig. 4). By contrast, fronts associated with Group 4, upwelling fronts, are found in “frontal clouds” extending from the coast, across the continental shelf well into deep water, such

as the distribution seen off the coast of California (Fig. 1). These distributions are comprised of many small-scale fronts, or upwelling filaments. Because these frontal segments tend to be quite small (a few kilometers long at most) and ephemeral, the diffuse clouds with which they are associated tend to be ‘lumpy’ with probabilities ranging from 1.5% to 4.5% at most. As a result, the PFDA, which keys on local maxima, does not perform well in such regions; the majority of the local maxima detected in the first pass of the algorithm are flagged as noisy pixels and discarded. However, in some cases, statistical anomalies associated with ‘lumps’ in the probability distribution result in segments, which pass the tests used to eliminate non-persistent fronts. For example, in the (relatively weak) upwelling region off the Iberian Peninsula, the along-coast fronts are detected accurately, but the algorithm also detects persistent fronts normal to the coast extending west of 10°W. A potential remedy to this problem would be to add an additional level of analysis for coastal upwelling regions, one designed to delineate regions of diffuse probability. Such an algorithm, relying on discriminating frontal objects from their background, has been implemented and tested for the California Current system (Fig. 12). The results of this method are complementary to the output of the original PFDA.

Planetary-scale fronts The probability distribution of fronts found in Group 3, those associated with large scale currents, tend to have aspects of fronts in both Groups 2 and 4. Specifically, where the currents are topographically constrained, the probability distribution consists of well defined ridges, which the algorithm readily identifies as persistent fronts – the inner edge of the Gulf Stream off the southeastern coast of the U.S. prior to separation from the continental margin (Fig. 1) – and, where the currents are not topographically constrained, the probability distribution tends to be broad and ‘lumpy’, as in the case of the Gulf

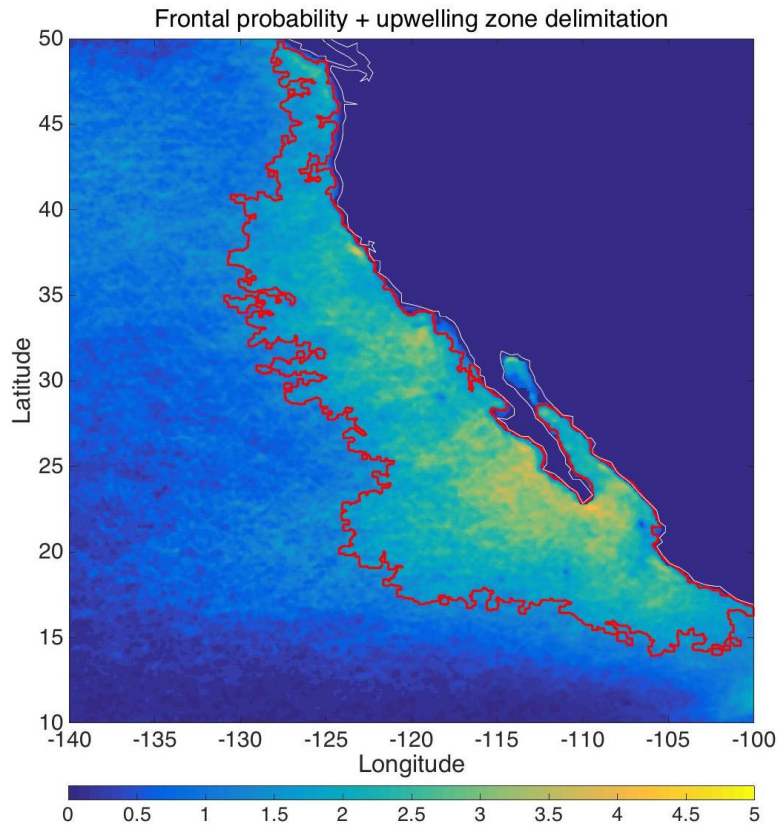


Figure 12: Frontal probability and coastal upwelling zone delimitation (red contour) in the California current system.

Stream following separation from the continental margin (also Fig. 1). These large / planetary scale fronts are associated with:

- western-boundary currents,
- eastern-boundary currents,
- Equatorial currents, primarily those associated with the Intertropical Convergence Zone (ITCZ), and
- polar currents, primarily associated with the Antarctic Circumpolar Current (ACC).

The long-term mean path of the Gulf Stream was identified as a persistent front

west of 65°W and the seasonal path of the Kuroshio was accurately located upstream of the point of separation from the continental margin off of the coast of Japan. However, the algorithm struggles to find a persistent front over significant distances once these currents have moved into the open ocean. This is mainly because the currents tend to meander significantly, resulting in the ‘lumpy’ probability distribution alluded to above, presenting the same algorithmic problems found with upwelling fronts. A similar issue arises regarding fronts belonging to the ITCZ, as their frontal probability is very low (below 1% typically) and diffuse. Again, the majority of the local maxima detected in the first pass of the algorithm are flagged as noisy pixels and discarded. In fact, no ITCZ fronts are evident in the global maps presented in Appendix B. Lastly, with the exception of topography-steered fronts, such as those found along the Kerguelen plateau, the southern subpolar and polar fronts are not found by the algorithm. This is thought to be the case because these fronts are composed of several branches, which tend to meander both factors militating against identification of the fronts as persistent

1.5 Conclusion

In this study, a method to detect persistent SST fronts based on frontal probability fields derived from the Cayula-Cornillon algorithm was presented. Five global maps of persistent SST fronts were produced, one based on the 4 km, 29-year AVHRR Pathfinder dataset and one for each of the climatological seasons based on the 4 km MODIS Aqua and Terra datasets, together with the Pathfinder dataset covering the same ≈ 10 -year period as the MODIS datasets.

In addition to locating persistent fronts, the algorithm also identifies a number of properties associated with each frontal pixel and the approximate edge of the persistent frontal region at each frontal pixel location. The latter are of particular interest in that they allow for the characterization of user defined properties

associated with a persistent frontal segment or as a function of distance along the front. For example, it would be straightforward to examine the change in frontal step over time, or to examine fish catch in the persistent frontal region versus that in the same general area but outside of the frontal zone.

Bibliography

Belkin I.M., and Cornillon P.C., 2007: Fronts in the world ocean's Large Marine Ecosystems, in ICES CM 2007/D:21, 33pp

Belkin I.M., Cornillon P.C., and Sherman K., 2009: Fronts in Large Marine Ecosystems, in *Progress in Oceanography*, 81, pp 223-236

Buongiorno Nardelli B., Colella S., Santoleri R., Guarracino M., and Kholod A., 2010: A re-analysis of Black Sea surface temperature, in *Journal of Marine Systems*, 79, pp 50-64

Cayula J.F., and Cornillon P.C. 1992. Edge detection algorithm for SST images, in *Journal of Atmospheric and Oceanic Technology*, 9(1), 67-80.

Cayula J.F., and Cornillon P.C. 1995. Multi-image edge detection for SST images, in *Journal of Atmospheric and Oceanic Technology*, 12(4), 821-829.

Csanady G.T., 1978: Wind effects on surface to bottom fronts, in *Journal of Geophysical Research*, 83, pp 4633-4640

Fedorov K.N. 1986. The physical nature and structure of oceanic fronts, Springer-Verlag, New-York.

Ginzburg A.I, Andrey G.K., and Sheremet N.A., 2004: Seasonal and inter-annual variability of the Black Sea surface temperature as revealed from satellite data (1982-2000), in *Journal of Marine Systems*, 52, pp 33-50

Hsueh Y., and Cushman-Roisin B., 1983: On the formation of surface to bottom fronts over steep topography, in *Journal of Geophysical Research*, 88, pp 743-750

Ibrayev R.A., zsoy E., Schrum C. and Sur H.I, 2010: Seasonal variability of the Caspian Sea three-dimensional circulation, sea level and air-sea interaction, in

Ocean Science, 6, pp 311-329

Karimova S., 2014: Hydrological fronts seen in visible and infrared MODIS imagery of the Black Sea, in *International Journal of Remote Sensing*, 35, pp 6113-6134

Kilpatrick K.A., Podesta G., and Evans R., 2001: Overview of the NOAA/NASA advanced very high resolution radiometer Pathfinder algorithm for sea surface temperature and associated matchup database, in *Journal of Geophysical Research*, 106, pp 9179-9197

Kilpatrick K.A., Podesta G., Walsh S., Williams E., Halliwell V., Szczodrak M., Brown O.B., Minett P.J, and Evans R., 2015: A decade of sea surface temperature from MODIS, in *Remote Sensing of Environment*, 165, pp 27-41

Poulain P.-M., Menna M., and Mauri E., 2012: Surface geostrophic circulation of the Mediterranean Sea derived from drifter and satellite altimeter data, in *Journal of Physical Oceanography*, 42, pp 973-990

Otsu N., 1979: A threshold selection method from gray-level histograms, in *IEEE Transactions on Systems, Man, and Cybernetics*, 9, pp 62-66

MANUSCRIPT 2

Global typology of persistent SST fronts

by

Yackar L. Mauzole ¹, Peter C. Cornillon ²

In preparation for submission to Geophysical Research Letters

¹PhD candidate, Graduate School of Oceanography, The University of Rhode Island, Narragansett RI 02882. Email: yackar_mauzole@my.uri.com

²Prof of Oceanography, Graduate School of Oceanography, The University of Rhode Island, Narragansett RI 02882. Email: pcornillon@me.com

2.1 Introduction

Fronts are features found in the ocean, which are characterized by strong gradients, separating distinct water masses. The first paper presented in this thesis focused on detecting persistent sea surface temperature fronts, which are fronts being observed at the same location over long periods of time (seasons, years, or decades), acting as boundaries between waters with different surface temperatures. Here, we investigate more in depth the properties of the persistent fronts detected from the global long-term mean frontal probability field. In particular, we proceed to a multivariate analysis to analyze how the frontal properties may or may not be related to one another by relying on unsupervised machine learning techniques (clustering). In the next section, we will introduce the statistical methods used for this study and describe how they apply to the specific problem of oceanic fronts. Results will be presented in section 3, where we will present the various clusters found and their respective characteristics. Then, we will discuss whether the results make sense in regards to the partitioning of the persistent sea surface temperature (SST) fronts into different types or not in section 4, followed by the conclusion in section 5.

2.2 Data and Methods

2.2.1 Data

As stated in the previous section, this study is a continuation of the work on the automated detection of persistent SST fronts in the global ocean. Within the framework of this thesis' first chapter, we applied the Cayula-Cornillon algorithm (Cayula and Cornillon, 1992; 1995) to several time series of SST fields, measured through Advanced Very High Resolution Radiometer (AVHRR) and MODerate-resolution Imaging Spectroradiometer (MODIS) over 29 and 9 years respectively. As a result, frontal probability fields were derived for the long-term

mean (AVHRR only) and for each climatological season (using a combination of AVHRR and MODIS). We developed a method to detect and track SST fronts based on the presence of local maxima in the frontal frequency maps, and obtained five global maps showing the location of persistent fronts. These maps are presented in Appendix B of this thesis.

This second study is focused on the statistical analysis of the SST fronts found from the long-term mean probability field only. Besides the map showing the location of persistent fronts, the Persistent Front Detection Algorithm (PFDA) also determines several frontal properties, for each pixel. The characteristics of the fronts can be divided into two categories, namely the intrinsic properties (SST gradient, SST cross-frontal step, frontal probability), and complementary properties, derived from other fields than SST. More specifically, we used the information provided by ETOPO1, which is a 1 arc-minute global model of the Earth's surface developed by the National Centers for Environmental Information (NCEI) - NOAA, to retrieve the ocean bathymetry and compute the bathymetry gradient magnitude. The gradient was computed using the Sobel operator, which consists in a 3x3 kernel, to obtain the zonal and meridional bathymetry gradient components. The operator is given here for the zonal gradient. By using the transposed of the kernel, we obtain the meridional gradient.

$$\text{Sobel x-derivative kernel: } \begin{bmatrix} +1 & 0 & -1 \\ +2 & 0 & -2 \\ +1 & 0 & -1 \end{bmatrix}$$

We recall that the magnitude of a gradient can be expressed as:

$$|\nabla H| = \sqrt{(\partial_x H)^2 + (\partial_y H)^2},$$

with H being the water depth.

Persistent fronts were found to have an average length of 25 pixels. This led to over 100,000 pixels belonging to almost 4,000 persistent fronts for the long-term mean field. While an exhaustive statistical analysis of the persistent fronts may have been desirable, we chose to reduce our dataset by considering the mean properties of each frontal line. Indeed, we assumed that the characteristics of each front did not change much along its pixels, so that the mean could serve as a good representative for the frontal pixels without too much loss of information. To assert this assumption, we computed the skewness of the SST gradient magnitude along each front over the global ocean. The figure 1 shows the distribution of the skewness for all the fronts. We see that it is bounded between 2 and -2 for the most part, indicating that the distribution of the SST gradient magnitude can be approximated to a normal univariate distribution for each frontal line (George and Mallery, 2010), and that the mean (and not the median) can be used in our case. The skewness was computed for other properties as well and looked very similar to figure 1 (not shown).

By reducing each front to its mean point, we can proceed to a multivariate analysis of the persistent fronts by using only 3932 points, instead of over 100,000. Thus for each front, we compute the mean value of each variable, namely:

- frontal probability,
- SST zonal gradient,
- SST meridional gradient,
- SST gradient magnitude,
- SST cross-frontal step,

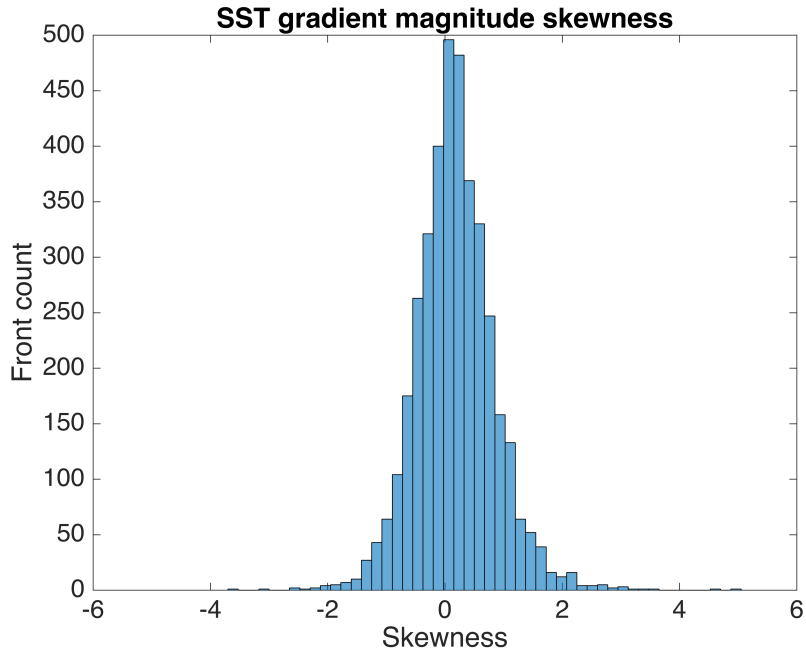


Figure 1: Skewness of the SST gradient magnitude for each persistent SST front detected by the PFDA from the long-term mean observations. Black horizontal lines indicate the ± 2 limit.

- bathymetry,
- bathymetry gradient.

All the values are then gathered into a large matrix containing 3932 rows and 7 columns. This defines the dataset of our study.

2.2.2 k-means clustering

There are many available statistical tools to find patterns in a dataset. In particular, cluster analysis allows to group observations that are similar by clusters, while separating observations that are dissimilar into distinct clusters (Rencher, 2002). Within the framework of multivariate analysis, classification can be understood as a supervised machine learning technique. While it is an attractive machine learning technique to determine the various types of persistent SST fronts, we chose to use cluster analysis instead since it does not require to know the fronts types

beforehand. Thus, we do not introduce bias from previous knowledge or experience, which could have influenced the results.

A popular clustering method is called *k-means*, which consists in partitioning observations into k clusters. It is an iterative method which requires to specify the number of clusters ahead. First, k points are placed randomly in the variable space; these are the seeds that will become the respective centroids of the clusters. The method computes the distance between each point in the dataset and the closest seed, and defines the first version of the clusters. The k points are readjusted as centroids of the clusters, and the distance between the points and the newly-located centroids is calculated again. Thus, the position of the centroids are updated at each iteration, until the method converges to an optimal division of the k clusters. The k-means procedure can be seen as a powerful tool to cluster oceanic fronts based on their properties. Nevertheless, the difficulty of applying it within the framework of this study lies in the determination of the optimal number of clusters, which is unknown a priori. We evaluate how many clusters we should ideally work with by computing the Davies-Bouldin Index (DBI) (Davies and Bouldin, 1979; Matlab documentation). The DBI relies on the ratio of the intra-cluster distance and of the inter-cluster distance:

$$D_{ij} = \frac{\bar{d}_i + \bar{d}_j}{d_{i,j}},$$

with \bar{d}_i (\bar{d}_j) being the average distance between each point and the centroid of the cluster i (j) to which they belong. The distance between the centroids of the clusters i and j is $d_{i,j}$. Figure 2 shows the DBI for various values of k . We restrict the maximum number of potential clusters to 10, and observe that the DBI reaches a minimum for $k=9$, which corresponds to the optimal clustering solution. We decided to proceed to the k-means clustering of our data with 9 clusters. The results will be explained in the following section.

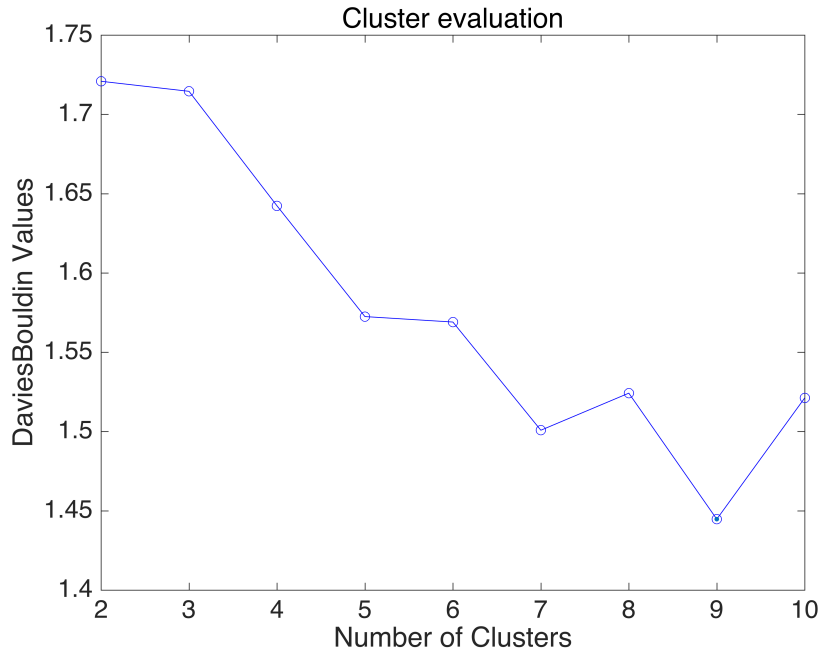


Figure 2: Evolution of the DBI showing the optimal number of clusters (9) for the k-means method.

2.3 Results

In this section, we present the characteristics of the nine clusters found by the k-means method. We will first show the correlations between the variables of our study (the frontal properties defined in the previous section). We performed a Principal Component Analysis (PCA) on our dataset, to determine how the variance of the data may be explained by the different variables. Unfortunately, no conclusive findings were obtained. Following the presentation of the correlations, we will discuss the differences between the various clusters.

2.3.1 Correlations

The matrix below corresponds to the correlation matrix, and is a 7x7 square (and symmetric) matrix. Each column represents a variable:

We note that the correlations between variables are never high: the largest

$$\begin{bmatrix} \text{FP} & \partial_x T & \partial_y T & \nabla T & \Delta T & H & \nabla H \\ 1 & -0.0435 & -0.1216 & 0.0001 & 0.1082 & 0.4664 & -0.3023 \\ -0.0435 & 1 & 0.0477 & 0.1191 & -0.1811 & -0.0698 & 0.0254 \\ -0.1216 & 0.0477 & 1 & 0.0214 & -0.1676 & -0.1770 & 0.2193 \\ 0.0001 & 0.1191 & 0.0214 & 1 & -0.2841 & -0.0444 & 0.0834 \\ 0.1082 & -0.1811 & -0.1676 & -0.2841 & 1 & 0.2835 & -0.1345 \\ 0.4664 & -0.0698 & -0.1770 & -0.0444 & 0.2835 & 1 & -0.3984 \\ -0.3023 & 0.0254 & 0.2193 & 0.0834 & -0.1345 & -0.3984 & 1 \end{bmatrix}$$

Correlation matrix. FP: frontal probability, $\partial_x T$: zonal SST gradient, $\partial_y T$: meridional SST gradient, ∇T : SST gradient magnitude, ΔT : SST cross-frontal step, H: bathymetry, ∇H : bathymetry gradient magnitude.

correlation corresponds to the one between frontal probability and bathymetry, at 0.46. This goes along with the fact that no variable is highly correlated to one another globally.

2.3.2 Cluster analysis

Clustering was performed on persistent SST fronts, for the long-term mean case. Fig. 7 shows the results of the k-means clustering, the different colors representing the nine clusters. For each cluster, the distributions of the variable are plotted as histogram in the following pages. We decide to describe the various clusters by discriminating them based on the average water depth corresponding to the fronts belonging to each clusters. We consider three different categories of fronts, namely:

- shallow-water fronts ($H > -1000\text{m}$),
- intermediate-water fronts ($-3000\text{m} < H < -1000\text{m}$),
- deep-water fronts ($H < -3000\text{m}$).

Each category counts three clusters each. Within the framework of this study, we first apply the clustering procedure to the dataset as is. The results presented here

take into account the sign of the SST gradient components (zonal and meridional), and of the SST step. In particular, we used the sign of the meridional gradient to separate further the clusters within each category. However, when performing the clustering analysis to the dataset without the effect of the sign on the SST gradients: it lead to an optimal Davies-Bouldin index equal to 10, with a slightly different partition of the fronts into the 10 clusters. We suppose that the sign of the SST gradients may not help explain the physics behind the partitioning of the oceanic fronts.

Shallow-water fronts Clusters 2, 4 and 6 correspond to persistent fronts found in shallow regions. The average water depth below the observed fronts is respectively -360m for cluster 2 (cyan) , -95m for cluster 4 (light blue), and -260m for cluster 6 (dark blue), as seen in Fig. 7.

Unsurprisingly, the fronts belonging to these three clusters are found near the coasts and in most marginal seas, such as the Bering Sea or the East China Sea. It should be noted that in contrast to clusters 2 and 4, cluster 6 shows a dominantly southward SST gradient . On the other hand, cluster 2 presents a larger frontal probability than both clusters 4 and 6. These observations can be interpreted as follows: fronts part of cluster 2 are the most persistent SST fronts present globally, with an average frontal probability of 4.9%. They are mostly found over the first few hundred meters, as are the fronts belonging to cluster 4 and 6. We can see from figure 10 that the tail of the distribution for cluster 4 is much shorter, resulting in an average water depth much smaller (fronts are found in areas where the water depth is shallower than 100m).

Fronts from cluster 6, such as the Ushant front, the fronts in the Yellow Sea and the fronts in the Gulf of Saint-Lawrence do not show a preferred direction for the zonal SST gradient but have clearly a southward SST gradient, seen to be always

in the direction away from the coastline. Similarly to fronts of cluster 2, the SST step is strictly negative, meaning that the waters near the coast are colder than on the other side of the fronts. We computed the correlation coefficients between variables within each cluster. Fronts in cluster 6 are the only shallow fronts for which the correlation between the meridional SST gradient and the cross-frontal SST step is significant, at 0.8.

Intermediate-water fronts The clusters 3, 8 and 9 correspond to persistent fronts found in regions where the bathymetry ranges from -1000m and -3000m. The average water depth below the observed fronts is respectively -1665m for cluster 3 (purple), -1740m for cluster 8 (red), and -2900m for cluster 9 (orange).

Persistent fronts found in cluster 3 are found near the coast, most of the time following closely the local isobaths below -500m. These fronts are usually seen on the western coast of U.S., Brazil, Somalia, India, China, and Tasmania, right over the edge of the continental shelf. The average value of the SST gradient magnitude is the second largest of all clusters with a value of 0.0864K/km, implying that the fronts of this cluster are very energetic. A good example is the Kuroshio front, and we note that some of the fronts in the Sea of Japan like the subpolar front are also part of cluster 3.

Fronts from cluster 8 are mainly found on the eastern boundary of the various ocean basins (Pacific: along Peru's and Chile's coastlines, Atlantic: the Canary and the Benguela currents, and short frontal segments along western Australia's coast for the Indian Ocean). Interestingly, the seaward segment of the Gulf Stream and the Malvinas current are also part of cluster 8. While it is common knowledge that western boundary currents and eastern boundary currents exhibit distinct characteristics, the k-means method groups some of them (not all: the Kuroshio being an exception) in the same cluster. It is not clear why boundary currents

belong to the same group. Remarkably, the two branches of the Gulf Stream each belong to a different cluster. This may be explained when considering the sign of their zonal and meridional SST gradients, showing that there is convergence between the branches. In fact, from figure 6, we see that the zonal SST gradient of cluster 8 is predominantly westward (away from the coast for the eastern boundary currents, and inward for the Gulf Stream and the Malvinas current), which contrasts with clusters 3 and 9, with respectively an eastward SST gradient and a gradient partly eastward, centered around 0. The main similarity between clusters 3, 8 and 9 is their large bathymetry gradient magnitude (above 2.6 in average). The ninth and last cluster groups fronts in the equatorial upwelling and coastal upwelling systems, in particular in the Pacific, and off Northwestern Africa. Corresponding to the nature of upwelling SST fronts, the average SST gradient magnitude is the lowest of all the groups, at 0.0566K/km. While we would have expected equatorial and coastal upwelling to be part of distinct clusters, fronts belonging to these systems have more in common with one another than with fronts from any other cluster.

Deep-water fronts The clusters 1, 5 and 7 correspond to persistent fronts found in deep-water regions. The average water depth below the observed fronts is respectively -3950m for cluster 1 (green), -4840m for cluster 5 (yellow), and -4040m for cluster 7 (light green).

From looking at figure 4, it is evident that clusters 1 and 7 represent the persistent fronts of the Subpolar Front (SPF). Both clusters show numerous similarities (in figure 5, with the frontal probability distributions, as well as in figure 11 with the bathymetry gradient distributions). Furthermore, the seventh cluster seems to group the fraction of the SPF found above bathymetric features, such as fault lines and underwater relief, in contrast to the first cluster. The main characteristics of

these two clusters are strong anti-correlations between the meridional SST gradient and the cross-frontal SST step (-0.85 for cluster 1, and -0.91 for cluster 7) which is unique among all nine clusters. The average meridional SST gradient is the largest of all clusters for cluster 7 at 0.0540K/km, slightly higher than the value for cluster 1, at 0.0392K/km. Cluster 7 also has the largest SST gradient magnitude at 0.0899 K/km, and the largest temperature difference across the fronts in absolute, at -0.8171K . Persistent SST fronts detected as part of the SPF are very energetic, just as the fronts belonging to cluster 5. The major difference between cluster 5 and the other two "deep" clusters is the direction of the meridional SST gradient: it is southwards for cluster 5 and northwards for clusters 1 and 7. Cluster 5 counts the fronts from the Kuroshio extension, the Gulf Stream extension, and part of the California coastal upwelling system. A few fronts can also be observed in the Sea of Japan. When computed exclusively for cluster 5, the correlation matrix presents some interesting values, namely the correlation between the meridional SST gradient and the SST step (0.92), the correlation between the meridional SST gradient and the SST gradient magnitude (-0.61) and lastly the correlation between the SST step and the SST gradient magnitude (-0.62). Cluster 5 is unique in that there exists a relationship between the SST step, the meridional SST gradient and the SST gradient magnitude. These fronts are mostly found in the open ocean, in contrast to fronts from cluster 3, which appear in their vicinity.

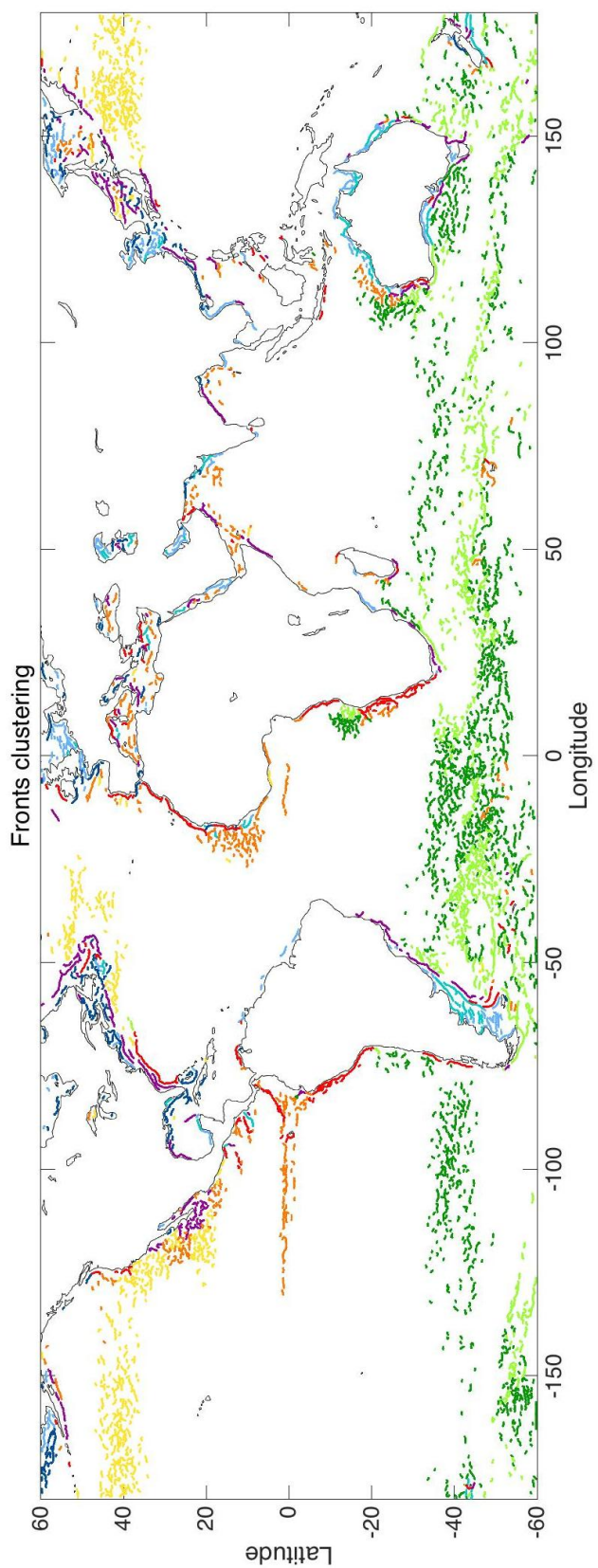


Figure 4: Global map of persistent SST fronts clusters. Legend: green: cluster 1; cyan: cluster 2; purple: cluster 3; light blue: cluster 4; yellow: cluster 5; blue: cluster 6; light green: cluster 7; red: cluster 8; orange: cluster 9.

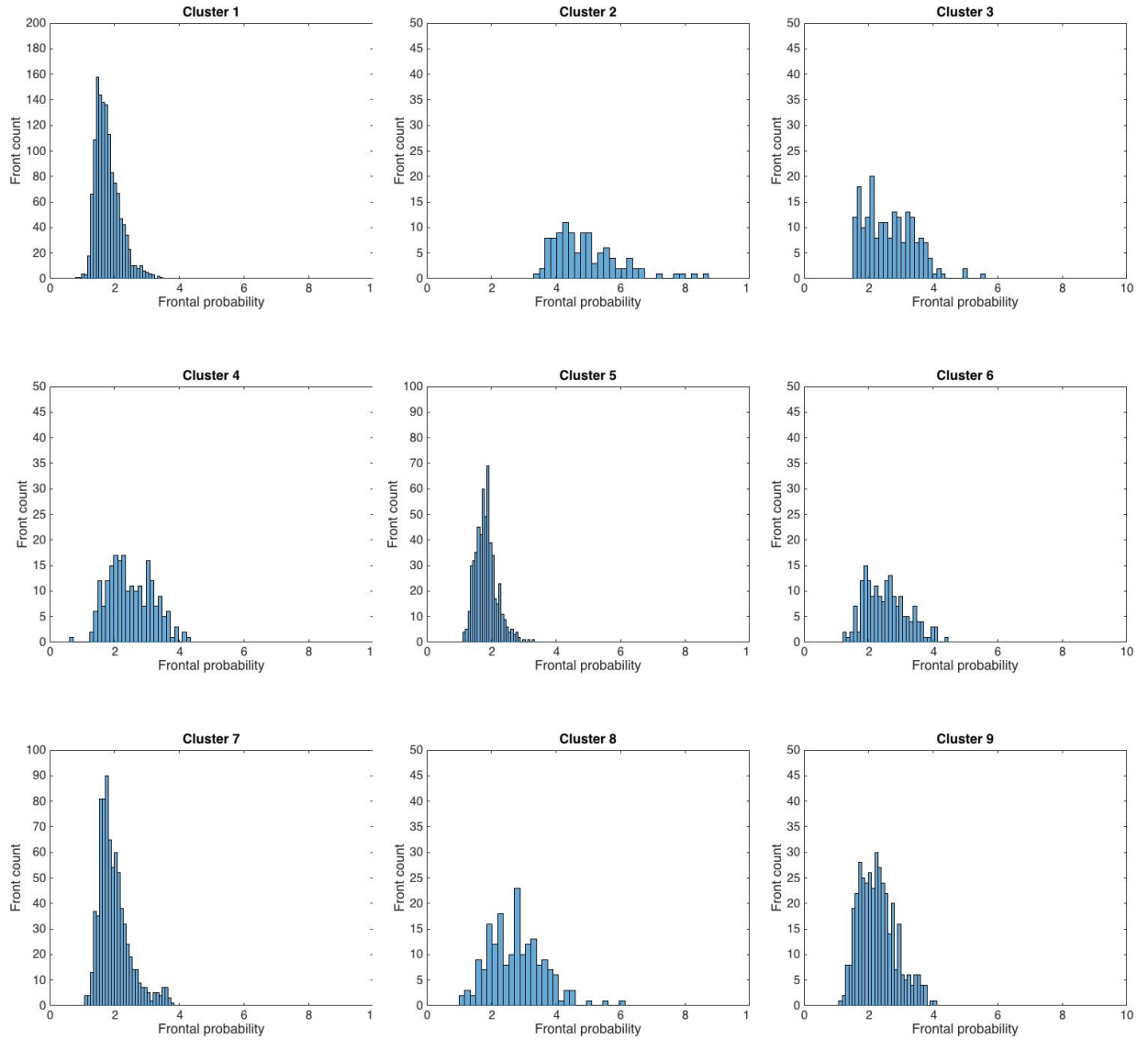


Figure 5: Frontal probability distribution by cluster.

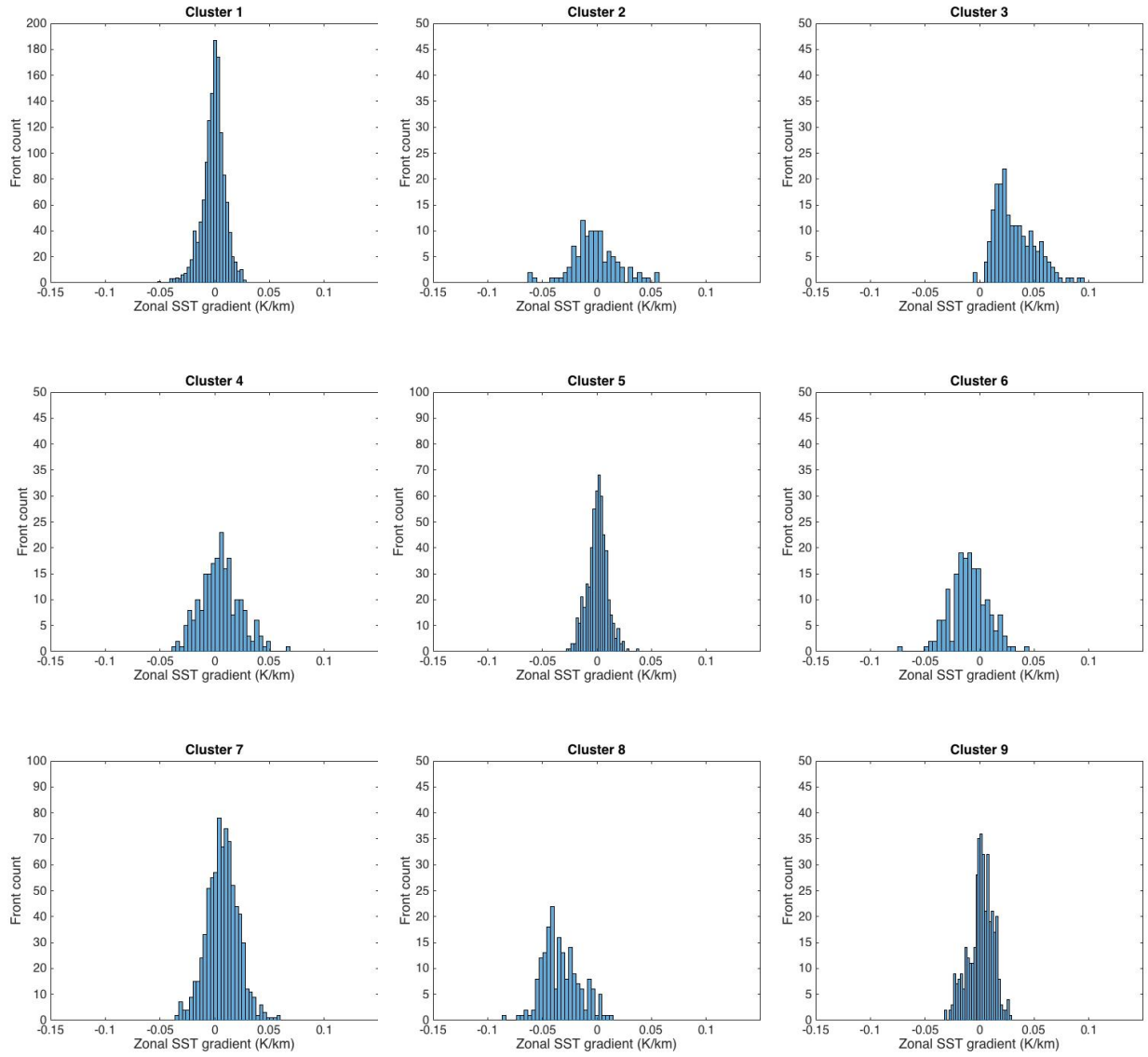


Figure 6: Zonal SST gradient distribution by cluster.

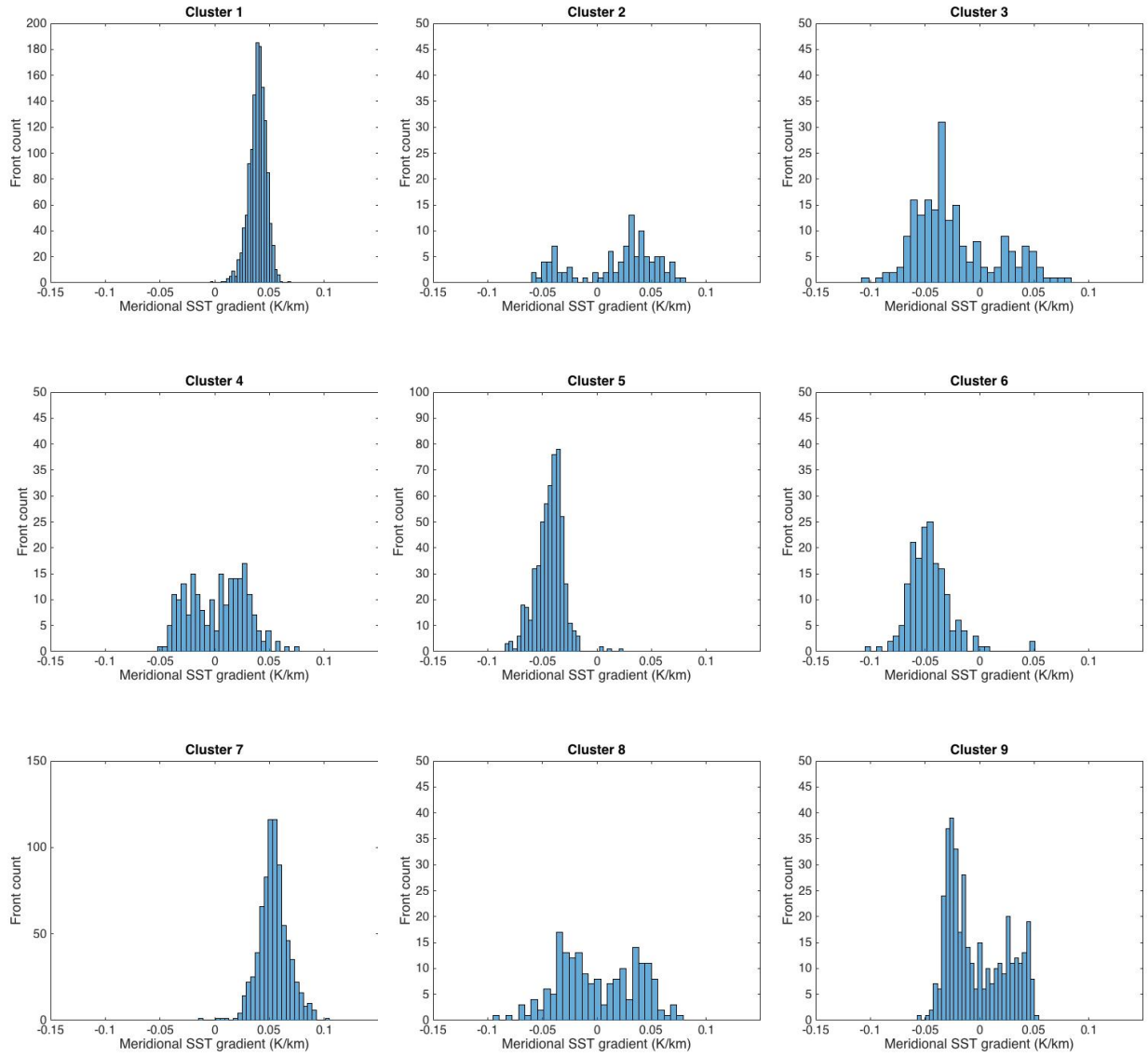


Figure 7: Meridional SST gradient distribution by cluster.

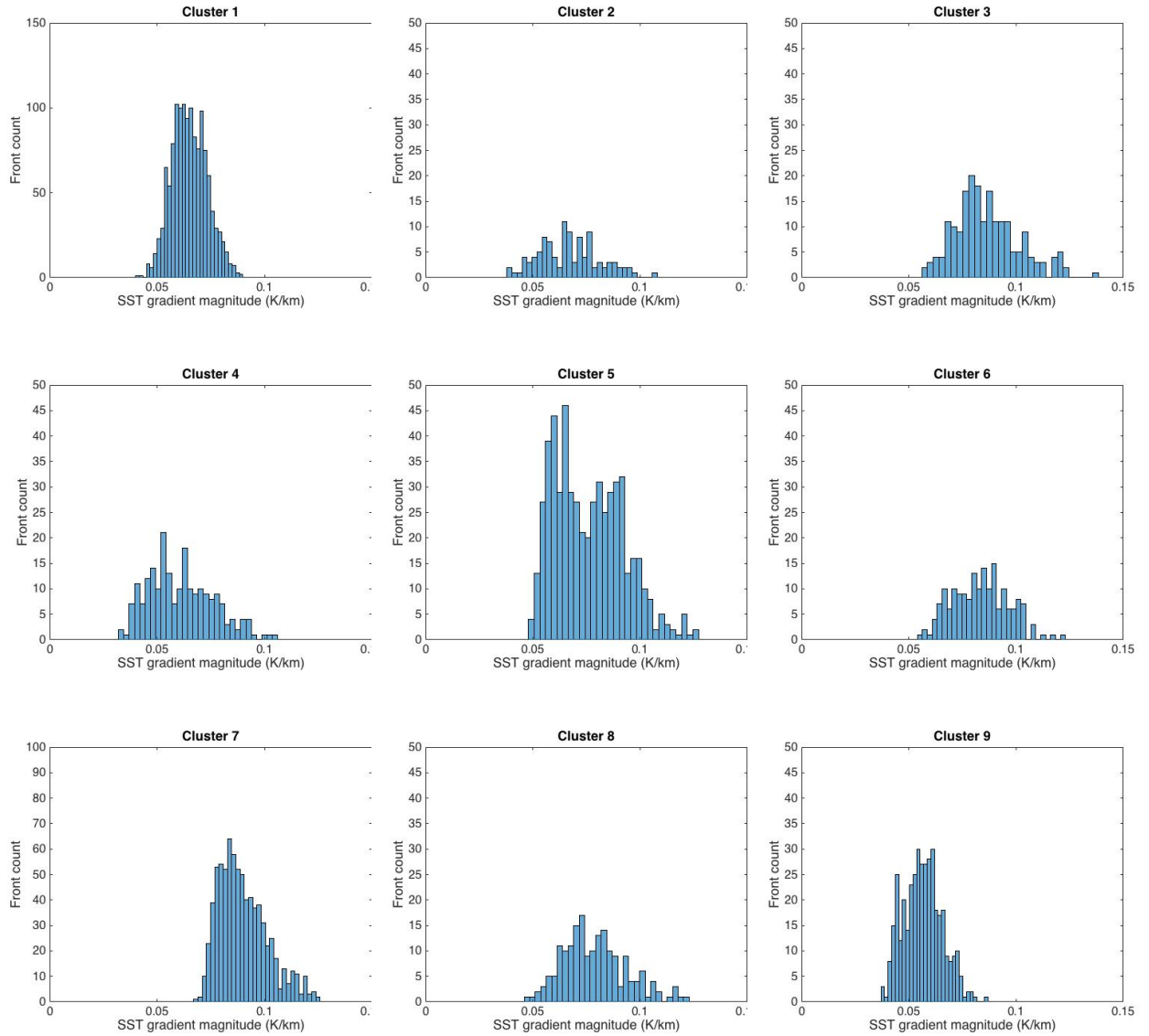


Figure 8: SST gradient magnitude distribution by cluster.

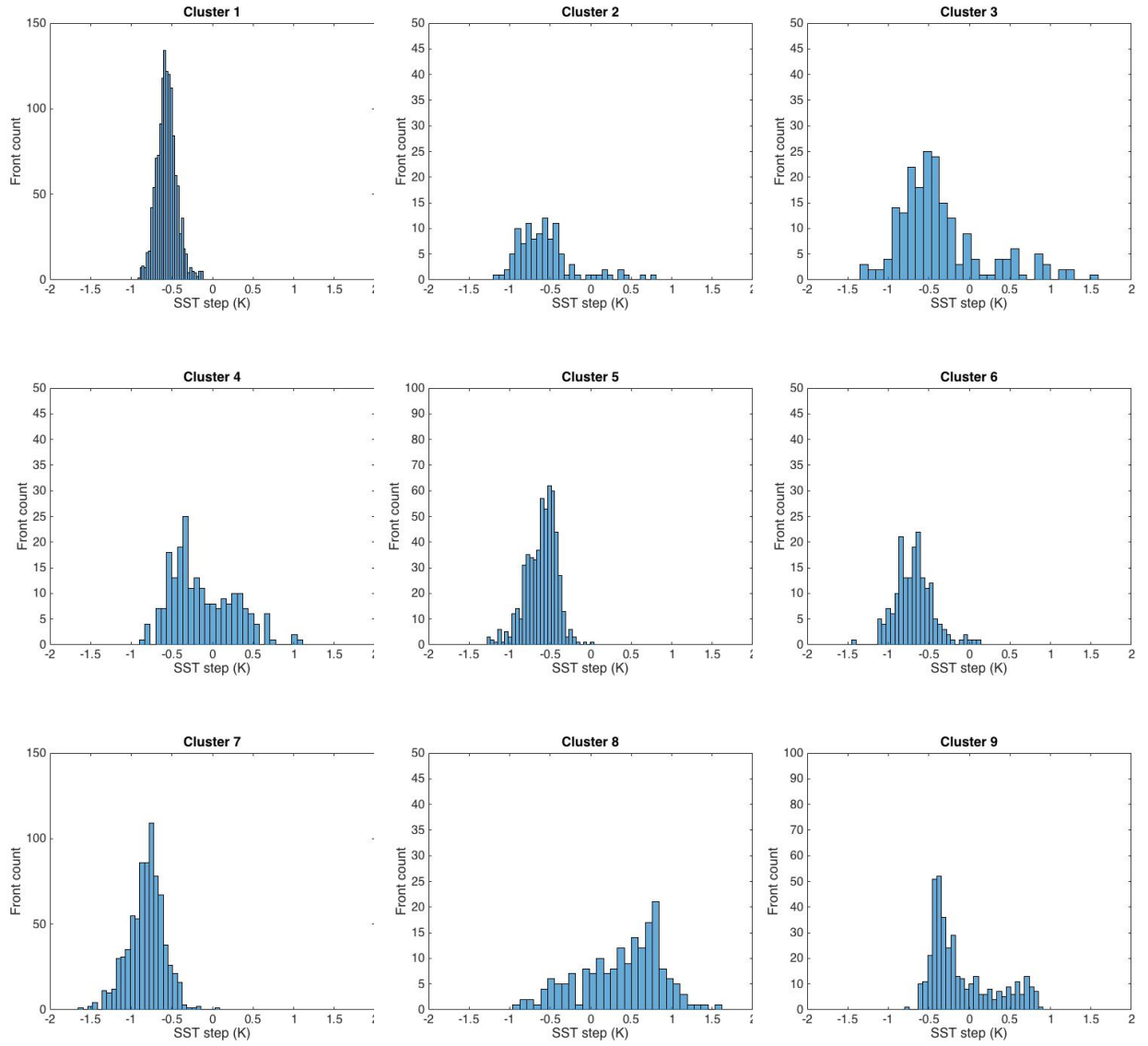


Figure 9: Cross-frontal SST step distribution by cluster.

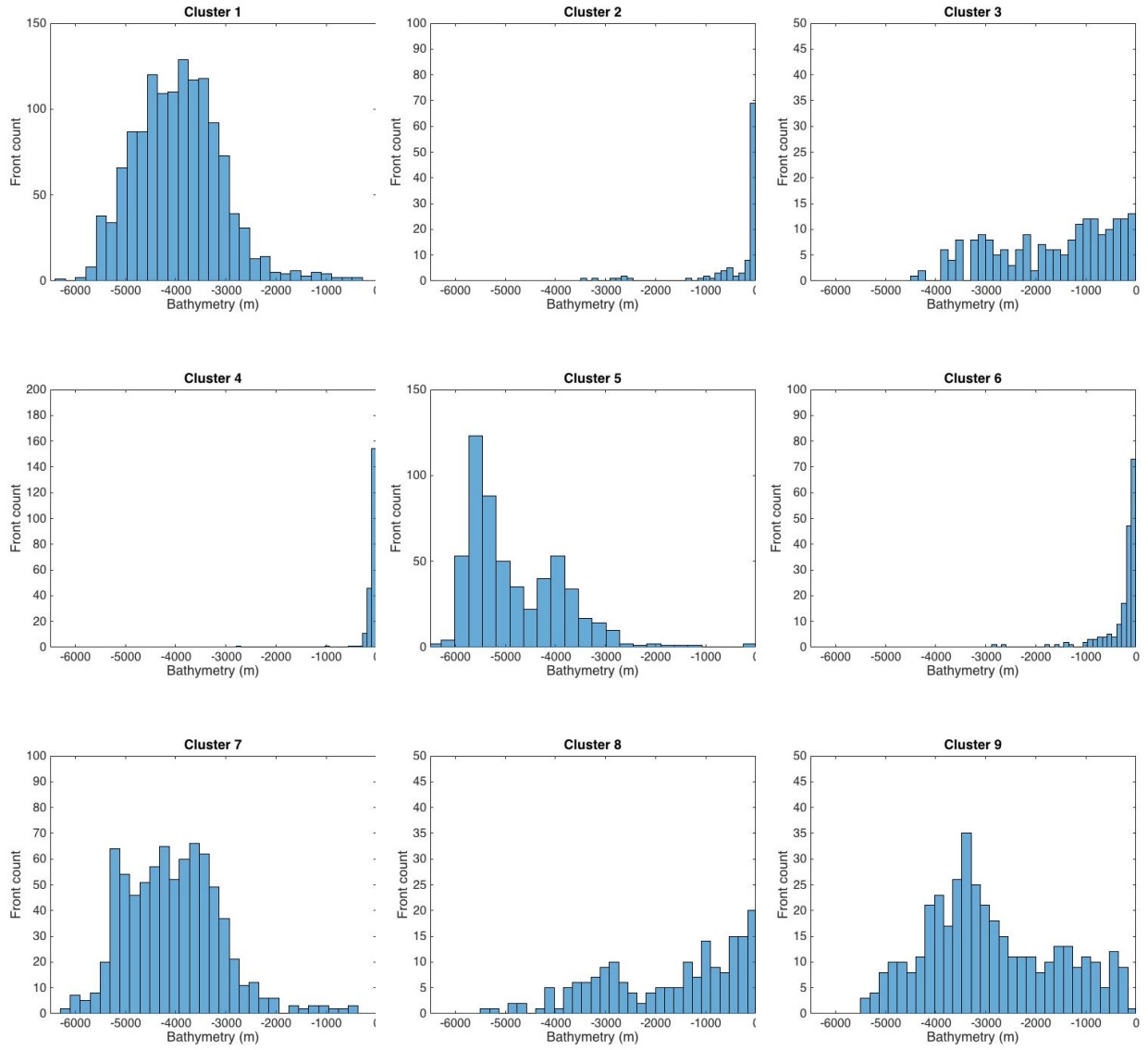


Figure 10: Bathymetry distribution by cluster.

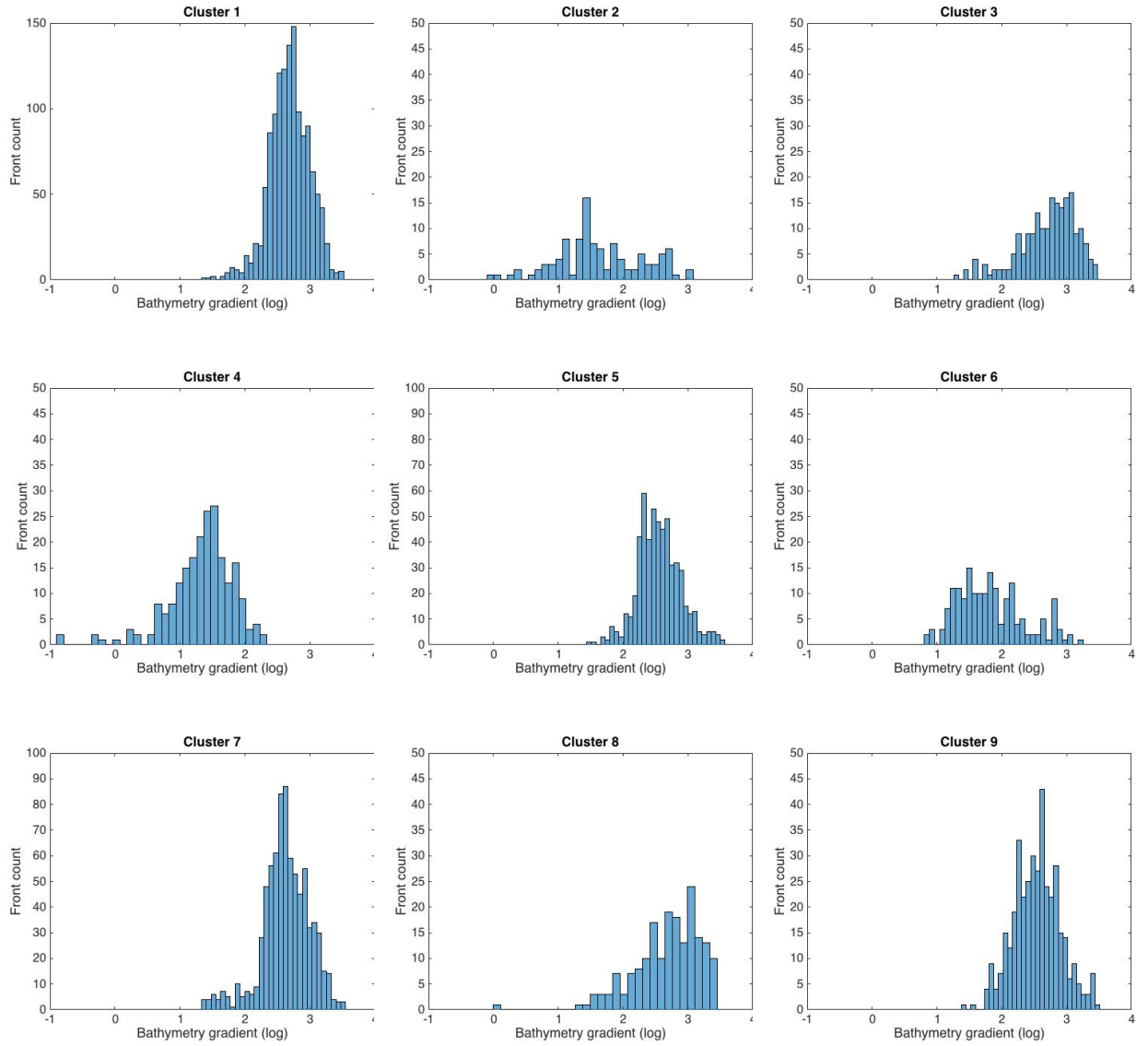


Figure 11: Bathymetry gradient distribution by cluster.

2.3.3 Summary

In this section, we recall succinctly our results from the Principle Components Analysis (PCA) and the k-means clustering. First, from applying the PCA to our dataset, we observe that the first three principal components explain only 60% of the total variance of the observations. The bathymetry is the variable which contributes the most to the first principal component, while the SST gradient magnitude and the meridional SST gradient are the main contributors to the second and third components respectively. We chose to distinguish between the nine clusters through their bathymetry, which lead to three groups of three clusters each.

Cluster 1, along with cluster 7, group fronts belonging to the Subpolar Front. Both clusters have a strong northward SST gradient, anti-correlated to the SST step. The fronts part of these clusters are very energetic, and while they are very similar, the main difference between them lies in the fact that the fronts of cluster 7 seem steered by topographic features, in contrast to the fronts of cluster 1.

Clusters 2 and 4 corresponds to fronts found in very shallow areas (less than 500m depth). Typically, these fronts are found over the continental shelf like in Patagonia, or in Australia for instance. Fronts from cluster 6 are also found in shallow waters, but they have a clear southward SST gradient in contrast to fronts from clusters 2 and 4. Typical fronts grouped in cluster 6 are the fronts found in the Bering Sea, in the Gulf of Maine, and in the South China Sea.

Fronts from cluster 3 are also part of the coastal fronts, but they differ from the fronts described above by being located exactly on the edge of the continental shelf, as seen in the Labrador region, in Patagonia and near Japan. They also have a large SST gradient magnitude, and their meridional SST gradient is correlated to the SST step, which is unique among shelf fronts. These fronts are observed in

regions where the water depth is intermediate, just like fronts from clusters 8 and 9. Cluster 8 represents fronts part of boundary currents systems (both western -Gulf Stream, Malvinas current; and eastern -Humboldt, Canary and Benguela currents-). The associated bathymetry gradient is large, as in the case of cluster 9. Fronts from cluster 9 correspond to fronts in equatorial and coastal upwelling for the most part. While intuitively, one would distinguish between eastern and western boundary currents on one hand, and between coastal upwelling (like in Northwestern Africa) and equatorial upwelling in the Pacific on the other hand, the k-means method find that persistent SST fronts from each category are matched up in only two clusters.

Finally, cluster 5 is a bit unique in the sense that it contains fronts that are part of the Kuroshio and the Gulf Stream extensions across the Pacific and the Atlantic. These fronts are found over the deepest regions, with a water depth past -4800m in average. They also have a very strong positive correlation between their SST step and their meridional SST gradient, which is strictly negative (southward).

The table 3 summarizes our findings. For convenience, we introduce types to fronts with similar patterns but belonging to distinct clusters. Type I corresponds to a cluster with similar patterns as another cluster, then called type II. In the case where there are three clusters with similarities, there will be three types.

2.4 Discussion

In this section, we will argue whether the k-means clustering makes sense from a dynamical point of view, i.e. do the clusters found by this method correspond to existing types of SST fronts or not. We will also discuss some possible developments to expand the scope of the present study.

2.4.1 Clustering commentary

Shelf fronts We ended the previous section with three different types of shelf fronts, represented by three distinct clusters (2, 4 and 6). By comparing our findings to existing literature on SST fronts, we can attest that shelf fronts of type II correspond to tidal fronts. Indeed, the mechanisms behind the occurrence of tidal fronts in shallow seas, such as the North Sea and other European Seas were investigated several decades ago (Simpson and Pingree, 1978; Simpson and Bowers, 1981; Bowers and Simpson, 1987). As mentioned by the authors, the European continental shelf is an area where the tidal dissipation accounts for about 10% of the total global dissipation of the M2 tides. Through the development of satellite altimetry, an estimate of the global dissipation of tidal energy (Egbert and Ray, 2000) was derived. Here, we show the map of tidal energy dissipation in figure 12. Regions where the tidal energy dissipation is very high in absolute value often correspond to areas where light blue fronts from cluster 4 are observed (Bering Sea, Australian Shelf, East China Sea, Patagonia, Gulf of Maine and English Channel). Based on the existing literature on these fronts, We propose that these fronts are associated to tides.

As discussed in the previous section, the main difference between the clusters 2 and 6 is the distribution of the meridional SST gradient, which is bimodal for

1	Subpolar	type I
2	Shelf	type I
3	Shelf-break	-
4	Shelf	type II
5	G.S. and Kuroshio extensions	-
6	Shelf	type III
7	Subpolar	type II
8	Boundary currents	Eastern/Western
9	Upwelling	Coastal/Equatorial

Table 3: Frontal clusters. The first column corresponds to the cluster number.

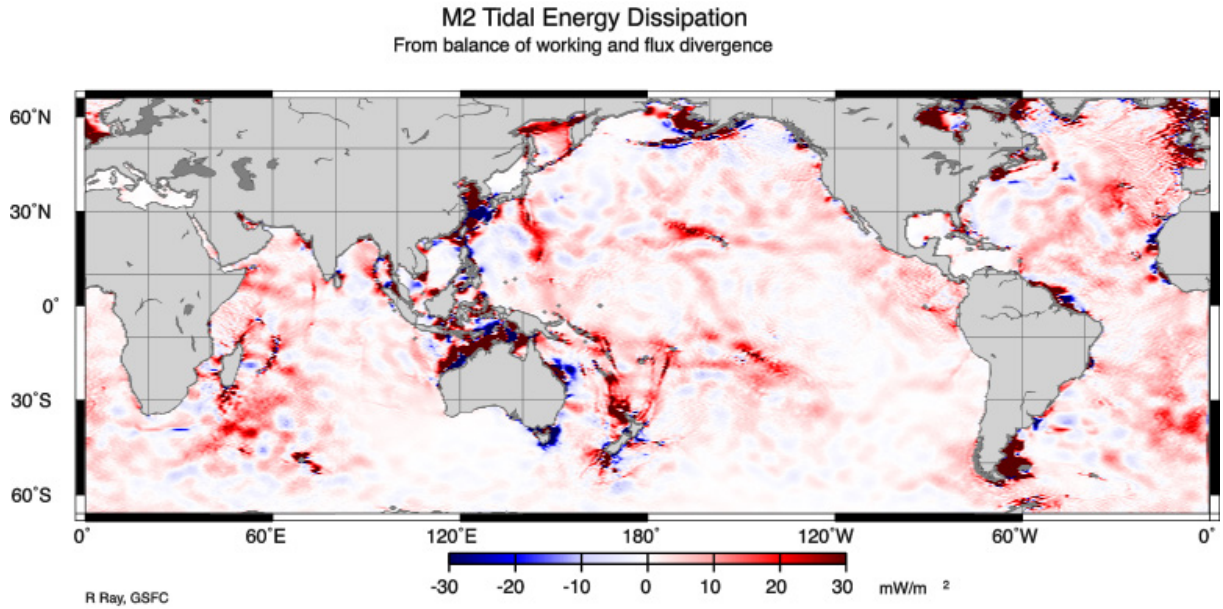


Figure 12: Tidal energy dissipation, from NASA, credits: Ray R.

cluster 2 and negative for cluster 6. In addition, fronts of cluster 2 present the highest frontal probability in average, indicating that they are more stable than fronts from cluster 4. Based on the seminal works of Chapman and Lentz (1994) and Hsueh and Cushman-Roisin (1983), we can safely assume that fronts of cluster 2 are anchored by bathymetry and can be defined as topography-controlled fronts. In regions like the Bering Sea, the Gulf of Maine or the Okhotsk Sea, shelf fronts of type III (cluster 6) appear along other types of shelf fronts. From the analysis of their properties, it seems that fronts of type III are similar to fronts of type II, except for their meridional SST gradient and their bathymetry gradient. It is not clear that these fronts are exclusively due to tides, and represent most likely a combination of type I and type II.

Subpolar frontal system Looking at the clusters 1 and 7, we see that they both correspond to the Subpolar front. As illustrated in figure 13, where the mean path of the SPF is drawn over the bathymetry field (right plot), we can see that

cluster 7 matches the segments following closely the bathymetric features, such as the Kerguelen Plateau and the various ridges. The interesting point is the inclusion of persistent fronts along the coasts of South Africa in these clusters. In particular, around -20°S , we see what could be counted as coastal upwelling (matching the description of cluster 9), however when looking at the cross-frontal SST step and the meridional SST gradient at that location, it is clear that the SST step is negative, and that the SST gradient at this location is also positive, both in the same order of magnitude than for the Subpolar fronts, as shown in figure 14.

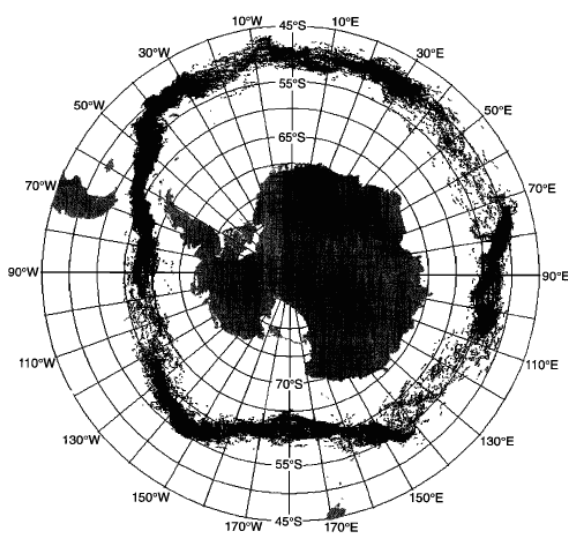


Figure 1. Displayed are all paths for the Antarctic Polar Front digitized for the years 1987–1993.

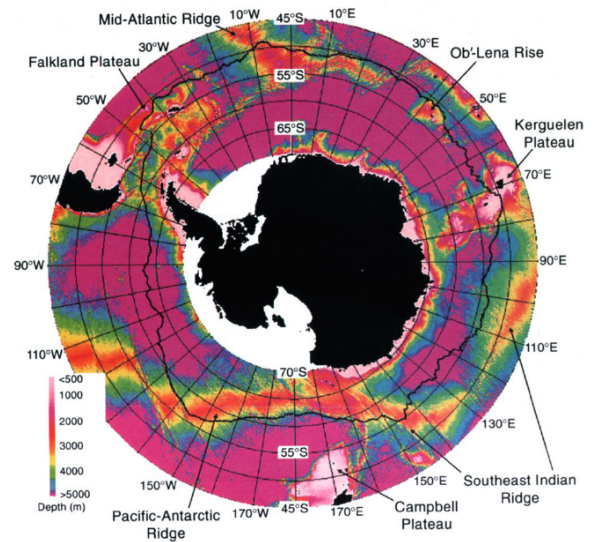


Plate 1. The calculated mean path for the Antarctic Polar Front is shown over the topography of the Southern Ocean [Smith and Sandwell, 1994].

Figure 13: Antarctic polar frontal system: all paths digitized from 1987 to 1993: left, and mean path over bottom topography: right. From Moore et al. (1999)

Boundary currents We noticed in the previous section that the boundary currents were found in a unique cluster (cluster 8), which seemed counter-intuitive given the vast contrast between western and eastern boundary currents. And while the Gulf Stream was part of that cluster, its Pacific counterpart, the Kuroshio, belongs to cluster 3 (which corresponds to shelf-break fronts), just like one branch

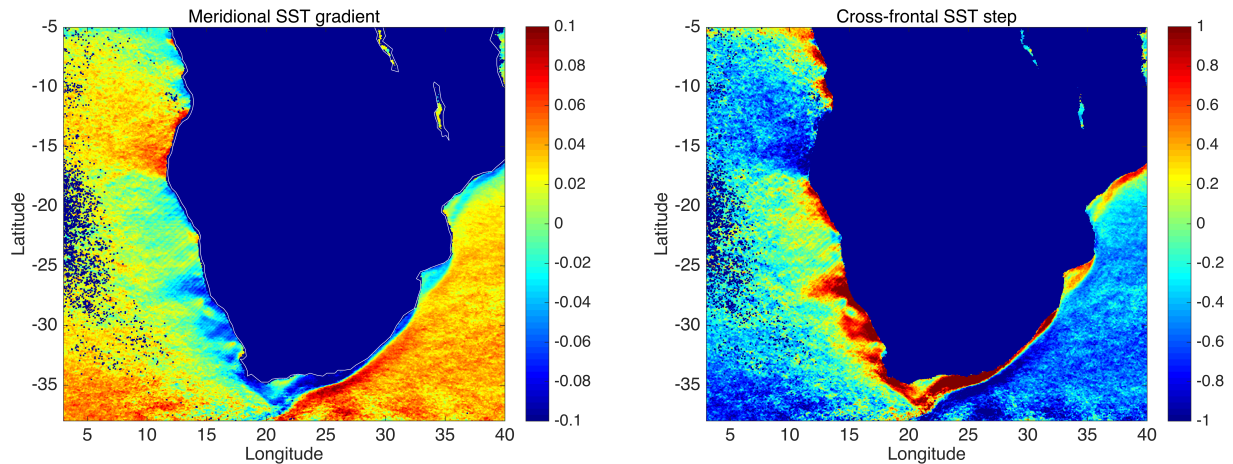


Figure 14: SST meridional gradient (K/km) and cross-frontal SST step (K) off Africa’s coast.

of the Agulhas current. Looking more carefully at the western boundary currents, we remark that the two branches of the Gulf Stream are red and purple (8 and 3); the Agulhas branches are purple and light green (3 and 7); and lastly the Kuroshio front is purple, with a very short red segment right under (8 and 3). When looking at the zonal and meridional gradients, we see a "convergence", i.e. each of these boundary current can be thought as dipoles with one inner branch and one outer branch between which the thermal flux converges. In the Southern Hemisphere, the light green front seen at the tip of South Africa is the outer branch of the Agulhas front. The reason that western boundary currents are associated with two, roughly parallel, fronts is that they are advecting warm water very quickly poleward. This makes them warmer than the water on both sides. On the shelf side the gradient will be larger and facing seaward because the shelf waters are cooler than those of the subtropical interior. The gradients on the outer edge are weaker and pointing shoreward. This feature uniquely distinguishes western boundary currents from other planetary frontal systems, as shown in figure 15.

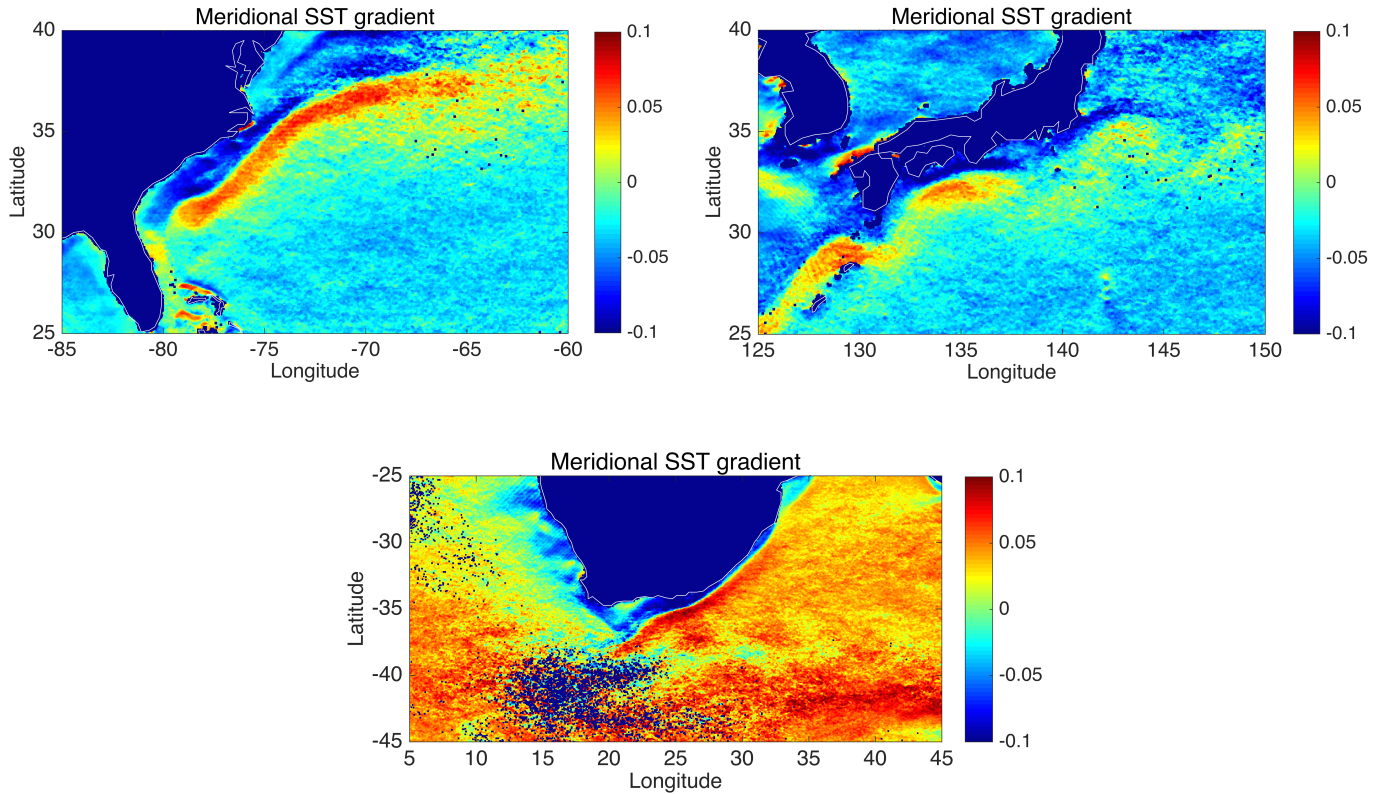


Figure 15: Meridional SST gradient (K/km) for Western Boundary Currents, Gulf Stream: left, Kuroshio: right, and Agulhas current: bottom.

2.4.2 Future improvements

From PCA to FPCA In the Data and Methods section, we introduced the assumption that fronts could be represented by their mean value, thus reducing their dimension from curves to single points. This approximation allowed us to apply the Principal Component Analysis to our dataset, and to use the results in the subsequent k-means clustering procedure with success. However, there exists in statistics an extension of the PCA which takes into account the fact that the frontal properties can have different values along each frontal line detected by the automated method described in chapter 1. The functional Functional - PCA (FPCA) (Tran, 2008) is considered an alternative to PCA to deal with curves such as oceanic fronts, without compromising their complex nature. In this context, the FPCA

could be applied instead of the PCA to define persistent fronts types.

Unsupervised clustering and pixel level Along with the assumption of using the mean value of each front, we also deliberately chose a method of unsupervised machine learning instead of a supervised classification procedure such as support vector machine. The main reason lies in the fact that we did not want to introduce bias in the clustering, but instead let the fronts and their properties "speak for themselves" without us assigning a front to one type or another. The caveat of tagging fronts beforehand with a given type is that it is partly subjective: even though the type of some well-documented fronts is already known, there are many more persistent fronts for which the type is not known. Also, we do not exclude the possibility of fronts being mixed-typed, which may hamper the results of a classification. A limitation we did observe through this study was that by selecting the number of clusters through the Davies-Bouldin index, our methodology became partially supervised.

Another direction which was considered during this project was to apply the PCA to the frontal pixels detected by the PFDA directly, as individual points instead of an ensemble of curves. Therefore the dataset would consist of over 100,000 points instead of 4000 as mentioned in the Data and Methods section. The motivation behind using all the points instead of just the mean points was to see if indeed some fronts could be of mixed types, in which case parts of their segments would belong to one cluster while other parts would be found in a different cluster. Preliminary results show that many of the shelf fronts described in our study are in fact mixed, of type I and II (clusters 2 and 4), as defined in the previous section.

Seasonality While the automated detection method tracked persistent SST fronts globally for the long-term mean (1982-2011) and each climatological sea-

son, this chapter only dealt with the long-term mean dataset. Seasonal variability is a crucial aspect of SST fronts, in particular when one wants to distinguish their dynamics and types. We intend to apply the cluster analysis to the seasonal dataset in the near future, and compare our results to the clustering of the long-term mean dataset.

2.5 Conclusion

Within the framework of this study, we applied multivariate analysis methods to the problem of sorting oceanic fronts, more specifically, persistent SST fronts into different categories. We used the approximation that each front could be represented by its mean point to proceed to a principal component analysis and a k-mean clustering. The clustering resulted in the partition of our dataset into nine clusters. Two of them represent the Subpolar front, three correspond to the shelf fronts, with the distinction between topography-controlled fronts and tidal-mixing fronts, plus one other comprising shelf-break fronts. In addition, one cluster groups coastal and equatorial upwelling fronts, while the two last clusters represent the boundary currents and the Gulf Stream and Kuroshio extensions respectively. From our findings, we can conclude that clustering produces great results regarding SST fronts and their various patterns. This chapter is a follow-up of the first chapter of this dissertation where fronts were objectively detected via an automated method. Here, we distanced ourselves from any subjectivity by relying on an unsupervised machine learning technique to sort out SST fronts. We believe that these results will provide insight in understanding under-documented or previously unknown fronts, as well as shed some light into some similarities between fronts thought before to be very different.

Bibliography

Bowers, D.G. and Simpson, J.H. 1987. Mean position of tidal fronts in European-shelf seas, in *Continental Shelf Research*, 7(1), pp. 35-44.

Chapman, D.C. and Lentz, S.J. 1994. Trapping of a coastal density front by the bottom boundary layer, in *Journal of Physical Oceanography*, 24(7), pp 1464-1479.

Davies D. L., and Bouldin D. W. 1979. A Cluster Separation Measure, in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1, pp. 224227.

Egbert G.D., and Ray R.D. 2000. Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data, in *Nature*, 405 pp. 775778.

Garrett, C.J.R., Keeley, J.R. and Greenberg, D.A. 1978. Tidal mixing versus thermal stratification in the Bay of Fundy and Gulf of Maine, in *Atmosphere-Ocean*, 16(4), pp 403-423.

George D., and Mallery P. 2010. *SPSS for Windows Step by Step: A Simple Guide and Reference 17.0 Update*. 10th Edition, Pearson, Boston.

Hsueh, Y. and Cushman-Roisin, B. 1983. On the formation of surface to bottom fronts over steep topography, in *Journal of Geophysical Research: Oceans*, 88(C1), pp. 743-750.

Moore J.K., Abbott M.R., and Richman J.G. 1999. Location and dynamics of the Antarctic Polar Front from satellite sea surface temperature data, in *Journal of Geophysical Research*, 104, C2, pp. 30593073.

Rencher, A. C. 2002. *Methods of Multivariate Analysis*, Second Edition, John Wiley & Sons, Inc., New York, NY, USA.

Pisoni, J.P., Rivas, A.L. and Piola, A.R. 2015. On the variability of tidal fronts on a macrotidal continental shelf, Northern Patagonia, Argentina, in *Deep Sea Research Part II: Topical Studies in Oceanography*, 119, pp. 61-68.

Rivas, A.L. and Pisoni, J.P. 2010. Identification, characteristics and seasonal evolution of surface thermal fronts in the Argentinean Continental Shelf, in *Journal of Marine Systems*, 79(1), pp. 134-143.

Simpson J.H., and Pingree R.D. 1978. Shallow sea fronts produced by tidal stirring, in *Oceanic fronts in coastal processes*. Springer Berlin Heidelberg, pp. 29-42.

Simpson, J.H., and Bowers, D. 1981. Models of stratification and frontal movement in shelf seas, in *Deep Sea Research Part A. Oceanographic Research Papers*, 28(7), pp. 727-738.

Tran N.M. 2008. An introduction to theoretical properties of functional principal component analysis. Doctoral dissertation, Department of Mathematics and Statistics, The University of Melbourne.

APPENDIX A

Algorithm details

A.1 Validation of CCA

As a preliminary step, we first carried out a validation of the Cayula-Cornillon Algorithm (CCA) to ensure the quality of the frontal pixels detected by the algorithm. We designed two tests to evaluate the performance of the CCA relying on two assumptions on fronts dynamics. We recall here that fronts are typically thought as regions of strong gradients, and we will postulate that the in-front sea surface temperature (SST) gradient at a given frontal pixel should be larger than the SST gradient surrounding it. We set the threshold for the ratio to 1.5, i.e. the in-front gradient should be at least 1.5 times larger than the out-of-front gradient. Regarding the second hypothesis, for simplicity of the argument we assume that fronts are linear objects. We compare the direction of the in-front SST gradient to the local normal of a SST front and detect whenever the angle difference becomes too large. Ideally, the SST gradient vector should coincide with the local normal vector of the fronts detected by the CCA but we allow a range of $+/- 30^\circ$. It appears that the angle difference between the fronts' normal and their SST gradient is a more restrictive condition than the SST gradient ratio: while 80% of the pixels were passing the test of the SST gradient ratio, only 40% passed the second test. When combined, about 60% of the frontal pixels passed both tests. A large number of pixels detected by the CCA showed a discrepancy between the local normal direction and the SST gradient direction. We found out that their spatial distribution was uniform over the global domain with no particular spatial preference. In the event where frontal pixels belonging to a same frontal segment did not all passed both tests, a system of recuperation was set up so that if one "bad" pixel was found in the midst of a series of "good" pixels, it would be re-

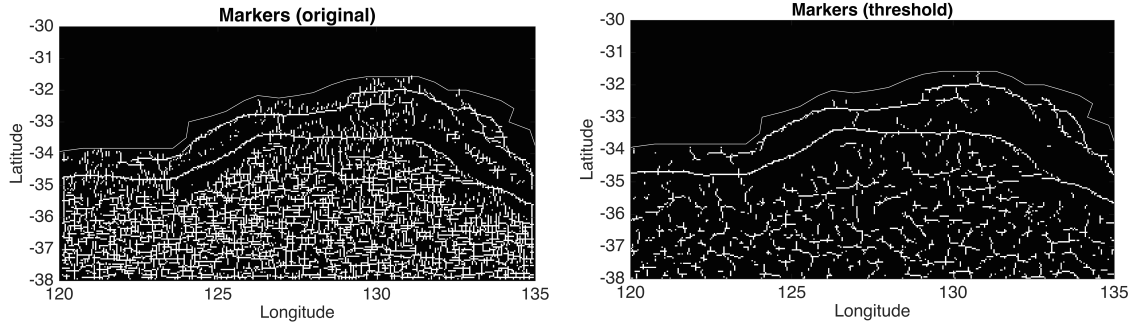


Figure A.1: Local maxima after the first pass (left) and the second pass (right).

stored to preserve the integrity of the frontal segment as often as possible. This last operation raised the percentage of kept pixels to 70%. The CCA was tested for all three satellite datasets, and the valid pixels were saved to serve as base input data, instead of the frontal pixels detected from the original SST fields.

A.2 Automated detection algorithm: noise reduction methods

From the flowchart presented in figure 2, we realized that local maxima were first retrieved even in regions of very low frontal frequency, which was not desirable. Going further, we looked at the properties of the frontal probability peaks, such as their relative height (prominence) and their width at mid-prominence value. Peaks corresponding to stable fronts tend to have a very clear signal, when compared to peaks attached to meandering fronts for instance. By implementing threshold values to the prominence and the width of the local maxima peaks, we were able to reduce the noise considerably as seen in figure A.1.

In many instances, points that should not have been connected were now forming "branches", and a large number of spurs were observed, especially for the seasonal maps where the ratio signal to noise was smaller than in the Pathfinder dataset alone. We thus designed a side-algorithm to detect branching points and disconnect some of the faulty parts in the frontal lines found by the ADA.

The two criteria were the local direction of the branches, as well as their frontal probability. In the case where three branches were connected, we would look at the direction of each branch at the branching point and try to cut the branch that would maximize the direction change of a segment. In the same manner, we looked at the values of frontal probability surrounding the branching point to cut the branch with the largest discrepancy. Following this operation, we assigned a number to each frontal line mapped out, and made sure that their pixels were indexed properly. Indeed, Matlab tends to order pixel by decreasing latitude which lead to problematic indexing for non-monotonic segments.

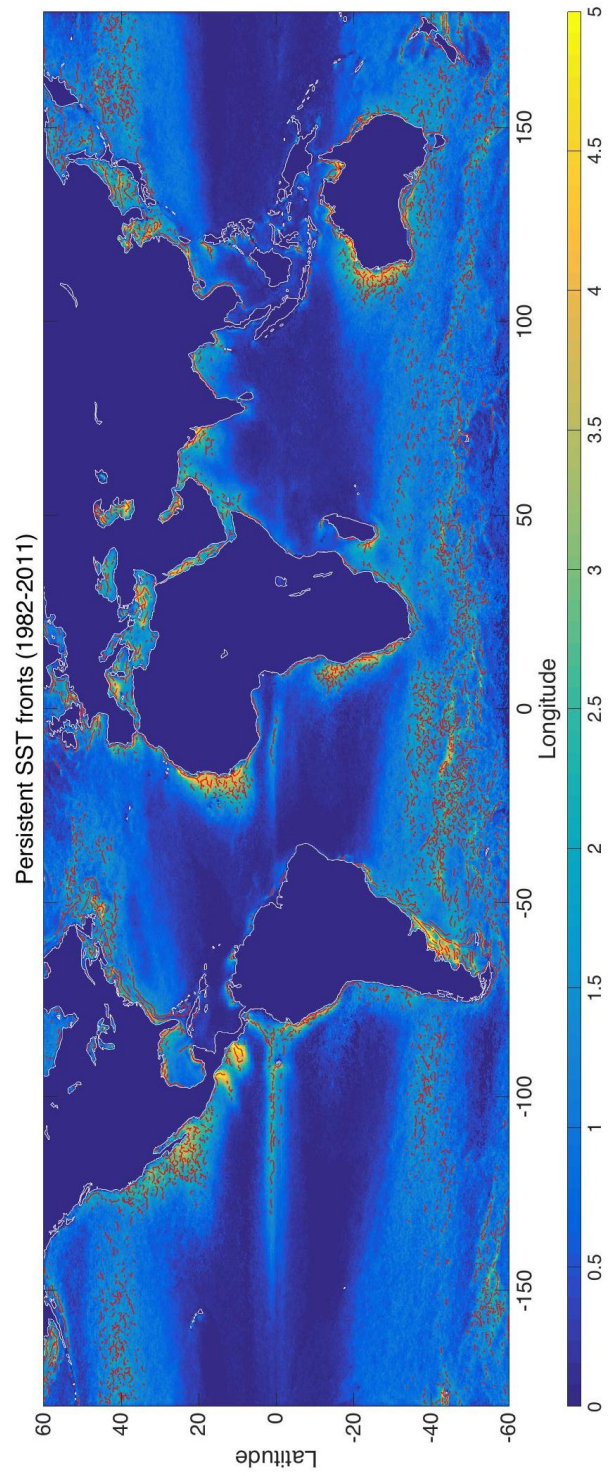
Beyond selecting peaks based on their properties, we rely on the Otsu's thresholding method (Otsu, 1979) to further discriminate between local maxima of interest and undesirable ones. The Otsu method is a well-known image processing technique that allows one to separate objects from their background by performing a histogram analysis of a given image. By assuming that there are two distinct classes of pixels in the image (those belonging to the objects and those part of the background), the method looks for an optimal threshold for which the inter-class variance is maximized. Detailed derivations can be found in the original paper and in Appendix A. We combined the Otsu algorithm with a sliding window, passed over the whole frontal probability fields to determine a local value of the Otsu threshold at any given pixel. This indicates for which value of frontal probability we would be able to set the fronts apart from their background.

After a first pass through the morphological operations, the branch cutting and the indexing methods, we manually tagged valid fronts over a large region

(Northern hemisphere) and compared the distribution of the local threshold for this sample to the distribution for the general population of fronts. Having a preexisting knowledge of what should be considered a valid front enabled us to define a measure of the likelihood for a local maximum to be part of a valid front based on their Otsu value. We introduce a weight $w(k)$, which can be expressed as the ratio of the area under the curve up to the value k to the area under the whole curve. Accordingly, points with a threshold of 3% are 99% likely to be valid stable fronts, which makes sense since higher Otsu thresholds are found in regions of higher frontal frequencies, and larger frequencies indicate more stable fronts. On the other hand, the probability for points with a threshold value of 1% to be valid is about 8%. By multiplying the original frontal probability field by the Otsu weight computed at each pixel, we derive a "weighted" frontal probability field for which high frequency values are accentuated while very low values are weighted down and become negligible parts of the map. Figure 3 shows the resulting field, where regions of low probability have become much less noisy than before (upper left and right plots).

APPENDIX B

Global maps of persistent SST fronts



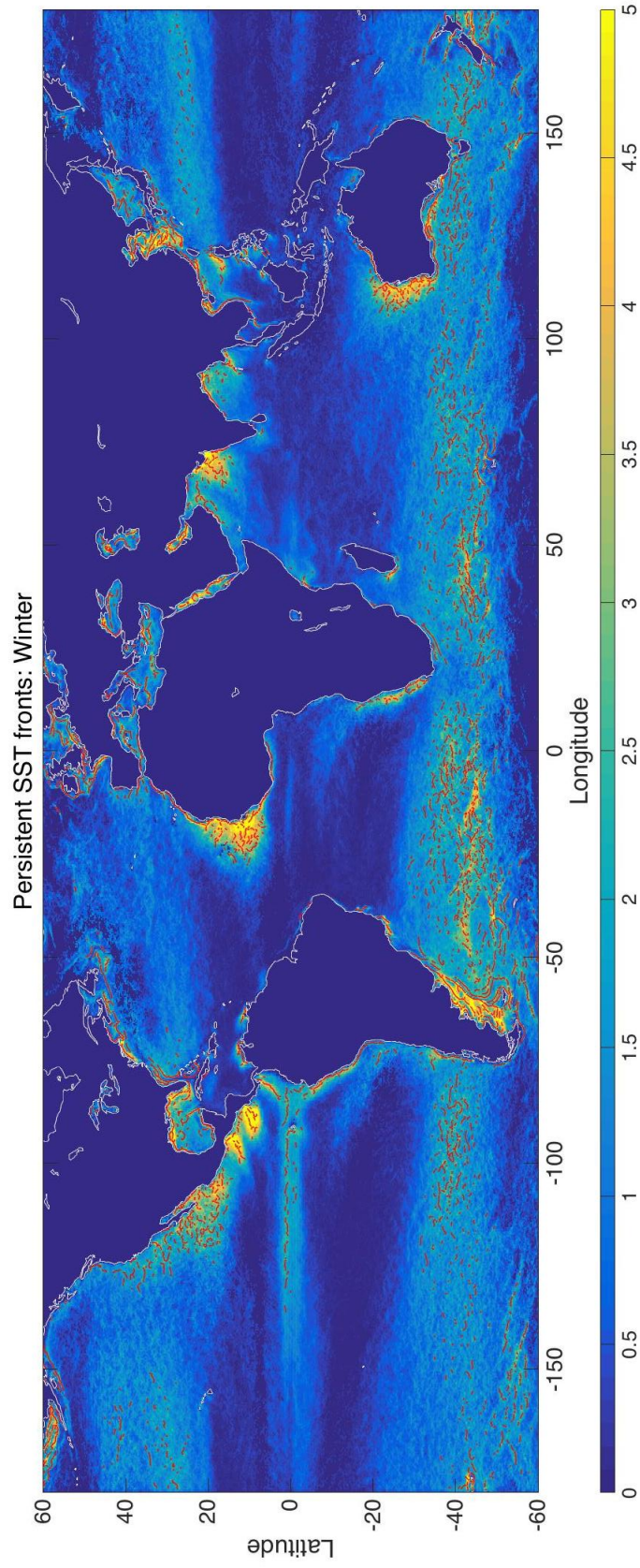


Figure B.2: Global map of persistent SST fronts for Winter, based on AVHRR and MODIS

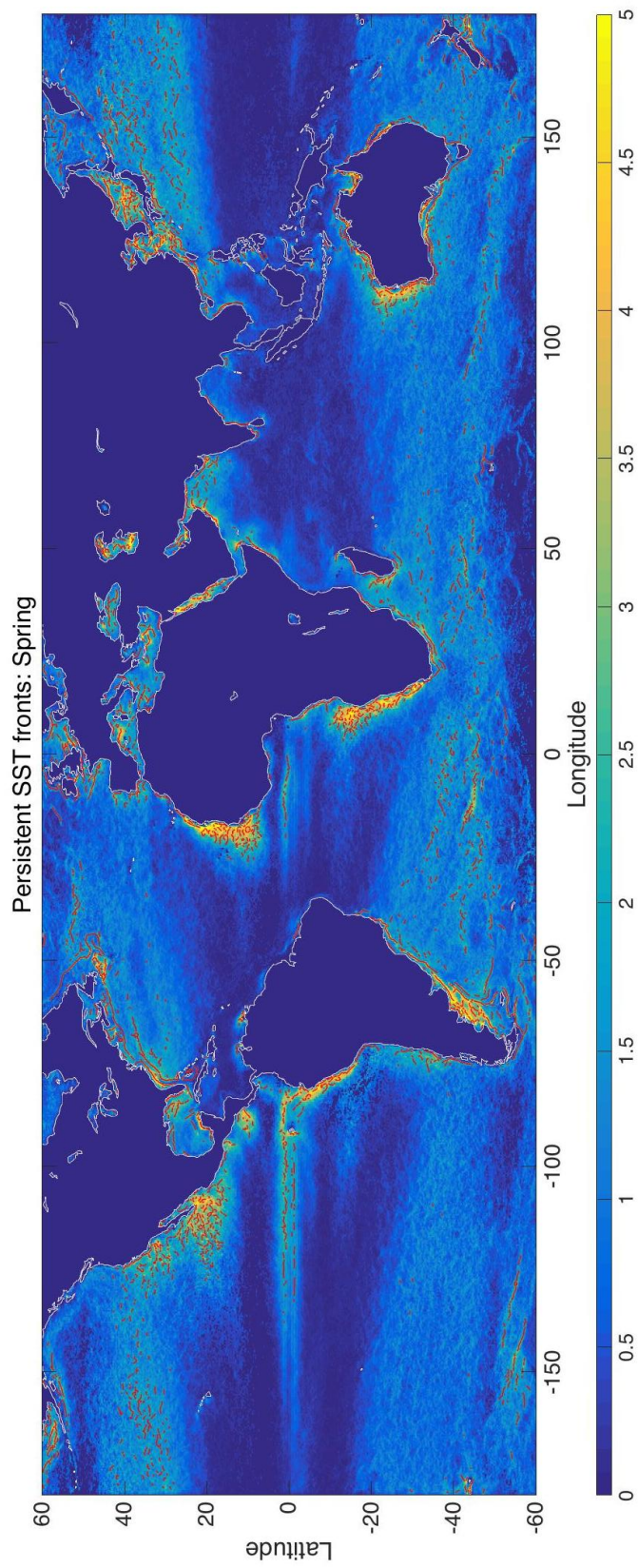


Figure B.3: Global map of persistent SST fronts for Spring, based on AVHRR and MODIS

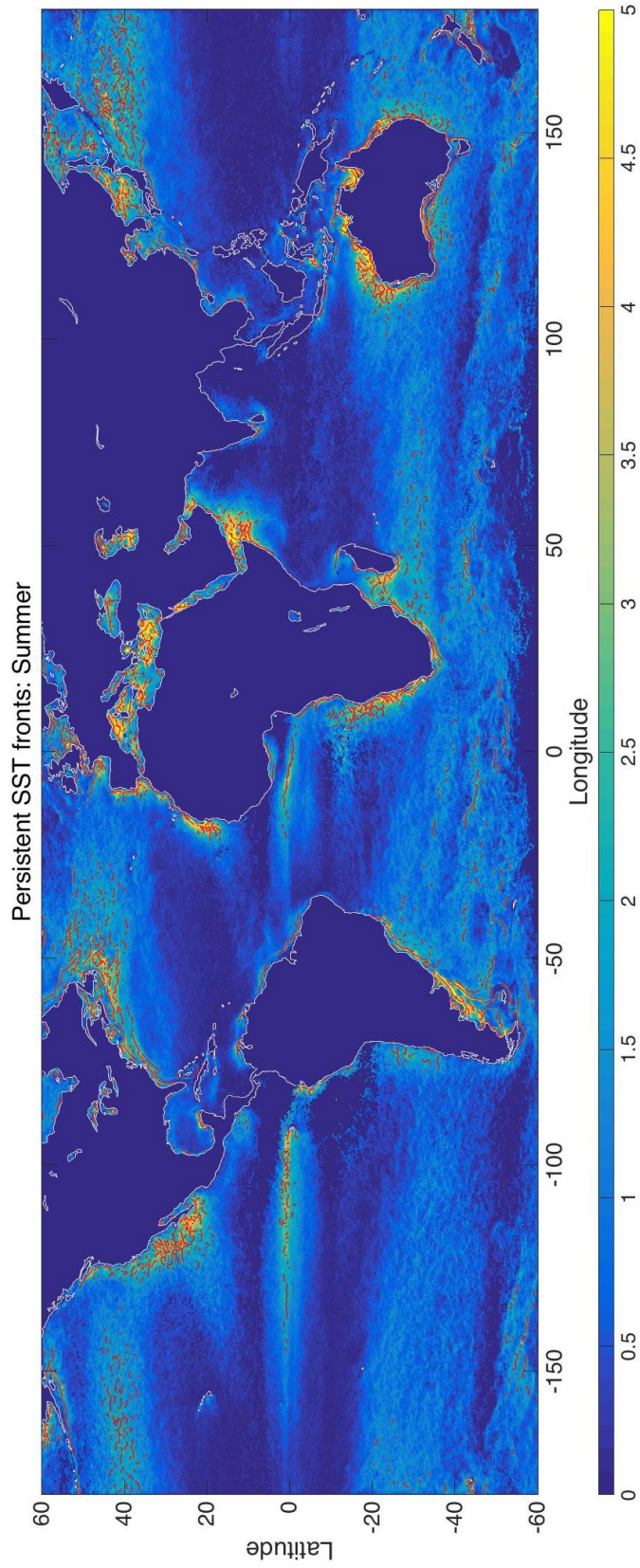


Figure B.4: Global map of persistent SST fronts for Summer, based on AVHRR and MODIS

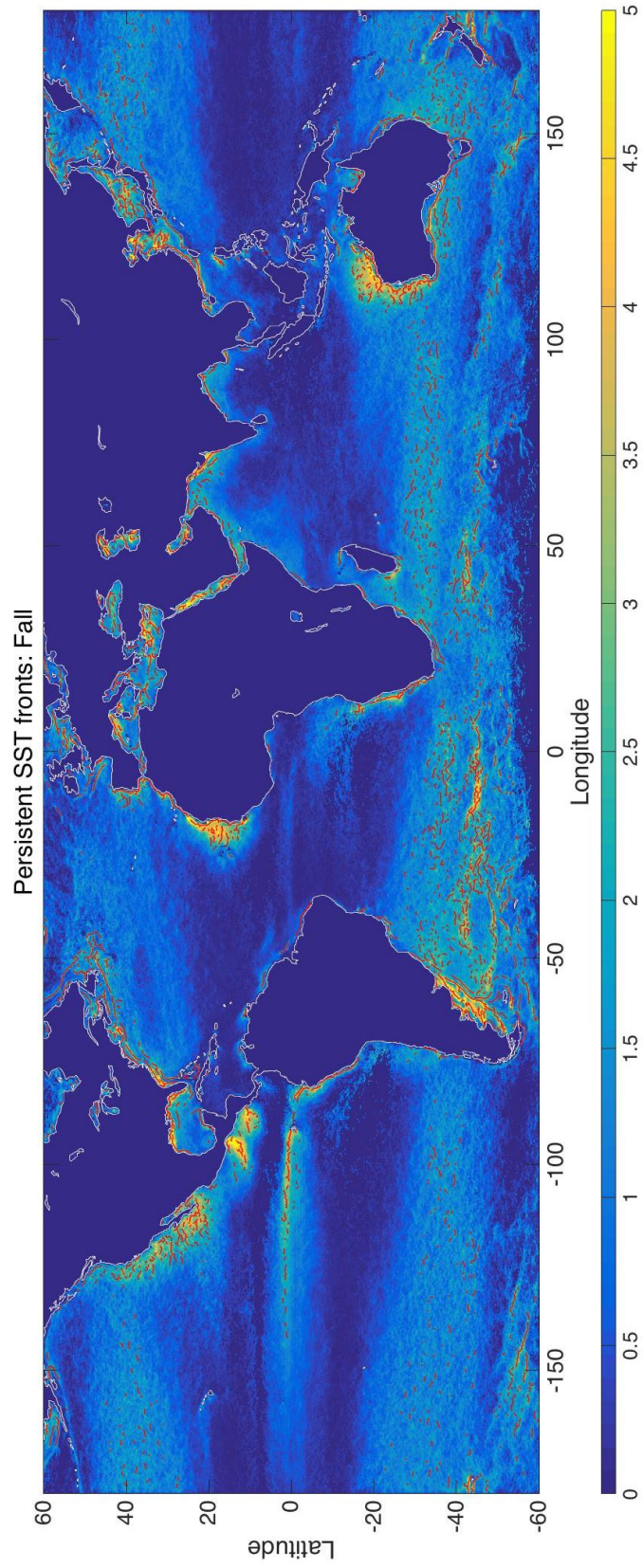


Figure B.5: Global map of persistent SST fronts for Fall, based on AVHRR and MODIS

B.1 Acronyms

ACC	Antarctic Circumpolar Current
AVHRR	Advanced Very High Resolution Radiometer
CCA	Cayula-Cornillon Algorithm
DBI	Davies-Bouldin Index
FPCA	Functional - PCA
ITCZ	Intertropical Convergence Zone
LME	large marine ecosystem
MODIS	MODerate-resolution Imaging Spectroradiometer
NCEI	National Centers for Environmental Information
NOAA	National Oceanic and Atmospheric Administration
PCA	Principle Components Analysis
PFDA	Persistent Front Detection Algorithm
SPF	Subpolar Front
SST	sea surface temperature
URI	University of Rhode Island