Stability of Marine Communities in Response to Climate Change

A Thesis Presented
by

Maria Salome Perez Rosero

to

The Department of Electrical and Computer Engineering

in partial fulfillment of the requirements
for the degree of

Master of Science

in

Electrical and Computer Engineering

Northeastern University
Boston, Massachusetts

August 2017
ACKNOWLEDGEMENTS

To Those who struggle to be rational in times of independence and achievements.
To Those who support and work for a mutual understanding among humans.
To Those who take a step out of their happiness, their mindset, out of the assurance.

Para Kitty y la Tarde Pega de HCJB,
To BCEC. To River of Life. To Mosaic Boston,
To my 140 The Fenway Friends :)

Thank you for lightening the unseen things by changing the way we think, by inspiring others, by being example of alternative ways in which we can use our inside battles to build our human character. Rather than being far from God, the change of our mindset and way of perceiving the unseen is an invitation that He extends to us, to experience the true meaning behind Him and how He wants to relate to us. When something seems to not make sense; going one step beyond rejection, one may be amazed by the answers of someone who cares not about the things that we can do, but what needs to be connected to our Souls. Chase the clues!
# TABLE OF CONTENTS

**LIST OF FIGURES**

**LIST OF TABLES**

**ABSTRACT**

<table>
<thead>
<tr>
<th>1 Introduction</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Related Work</td>
<td>3</td>
</tr>
<tr>
<td>1.2 Contributions</td>
<td>5</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2 Background</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Multidimensional Scaling</td>
<td>6</td>
</tr>
<tr>
<td>2.2 Principles of Majorization Algorithm</td>
<td>9</td>
</tr>
<tr>
<td>2.3 Minimizing MDS Objective Function by Majorization Algorithm</td>
<td>11</td>
</tr>
<tr>
<td>2.3.1 Majorization Algorithm for nMDS</td>
<td>14</td>
</tr>
<tr>
<td>2.4 Multidimensional Scaling for Ecology</td>
<td>15</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3 Methods for Community Stability in Response to Climate Change</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 Data Acquisition</td>
<td>17</td>
</tr>
<tr>
<td>3.2 Data Processing</td>
<td>18</td>
</tr>
<tr>
<td>3.2.1 Community Matrix</td>
<td>18</td>
</tr>
<tr>
<td>3.2.2 Environmental Matrix</td>
<td>21</td>
</tr>
<tr>
<td>3.3 Data Analysis</td>
<td>22</td>
</tr>
<tr>
<td>3.3.1 Quantifying Climate Change Across Different Latitudes</td>
<td>23</td>
</tr>
<tr>
<td>3.3.2 Quantifying Instability at Poleward Locations</td>
<td>23</td>
</tr>
<tr>
<td>3.3.3 Quantifying Climate Change Patterns Driving Instability</td>
<td>23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4 Results And Discussion</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 Evaluation Metrics</td>
<td>25</td>
</tr>
<tr>
<td>4.1.1 R-Squared Value</td>
<td>26</td>
</tr>
<tr>
<td>4.1.2 The $F$ Test Statistic in the Context of Linear Regression</td>
<td>27</td>
</tr>
<tr>
<td>4.1.3 The $F$ Test Statistic in the Context of ANOVA Linear Models</td>
<td>28</td>
</tr>
<tr>
<td>4.2 Quantifying Climate Change</td>
<td>29</td>
</tr>
<tr>
<td>4.2.1 Statistical Analysis of Climate Change</td>
<td>30</td>
</tr>
<tr>
<td>4.3 Quantifying Instability at Poleward Locations</td>
<td>31</td>
</tr>
</tbody>
</table>
## LIST OF FIGURES

1.1 Hypothesis prediction for community instability in response to climate change...

2.1 Minimizing a function $f(x)$ by iterative majorization approximations $g(x, z)$.

\[ f(x) = 6 + 3x + 10x^2 - 2x^4 \]  
(red curve), and

\[ g(x, z) = 6 + 3x + 10x^2 - 8xz^3 + 6z^4 \]  
(blue, green curves). First iteration point starts at $x_0 = 1.4$ (generates blue curve),

followed by an update at $x = 0.95$ (generates green curve). The process repeats

successively until convergence.

3.1 Block diagram for quantifying climate change and community instability. Our

approach consists of three main blocks (data acquisition, data processing, data analy-

sis); and within the data analysis, we propose a three step study that encompasses

the quantification of climate change, instability at poleward locations, and climate

change patterns driving instability to investigate about the stability of communities

in response to environment/climate change.

3.2 Geographical location of the groundfish trawl survey dataset conducted in the East

Bearing Sea during the summer periods from 1982 to 2016. This is the input data for

the study conducted in this thesis.

3.3 Graphical representation of the quantification of community instability via the com-

putation of pairwise distances between the low-dimensional (latitude, longitude,

year) observation points/site abundance scores.

3.4 Graphical representation of the quantification of community instability via the com-

putation of the mean distance to the centroid point of the low-dimensional (latitude,

longitude, year) observation points/site abundance scores.

3.5 Graphical representation of the quantification of community instability via the com-

putation of the area of a fitted convex hull containing the low-dimensional (latitude,

longitude, year) observation points/site abundance scores.

4.1 Models for Climate Change that relate different Temperature variables as a function

of increasing Latitude. Bottom water temperatures are expected to increase towards

northern latitudes.

4.2 Instability Metrics as a function of increasing Latitude. Communities show an

increasing instability pattern towards poleward locations.

4.3 Scatter plot of Temperature over different years and for different Latitude bins.
### 4.4 Instability Metrics and temporal changes of Temperature variables.

Temporal changes in Bottom Water Temperature show a direct approximation of Latitude; whereas for Surface Water temperature the opposite is observed.

### 4.5 Model fitting for Instability Metrics as a function of Temporal Changes in Temperature.

In all cases, a linear model probed to be enough to describe a significant relationship between these variables.

### I.1 Community Instability Metrics while varying the number of Latitude bins from 5, 10, 20, and 30 (I).

A quasi-uniform distribution of latitudinal variability (right side figures) is desired to interpret spatial patterns of community instability without sampling protocol artifacts. A bin size of 10 was selected for our analysis in the main text.

### I.2 Community Instability Metrics while varying the number of Latitude bins from 5, 10, 20, and 30 (II).

A quasi-uniform distribution of latitudinal variability (right side figures) is desired to interpret spatial patterns of community instability without sampling protocol artifacts. A bin size of 10 was selected for our analysis in the main text.
# LIST OF TABLES

1.1 Experimental ground fish data of related works in [1] ........................................ 4

3.1 Description of the temporal and biological characteristics of the groundfish trawl survey data collected in the East Bearing Sea .................................................. 19

4.1 Statistical results for a linear model fitting of Temperature (and its temporal change variants) as a function of increasing Latitude. Blue labels for p-values highlight a significant relationship; whereas red labels the opposite .......................................................... 31

4.2 Statistical results for a linear regression model to fit Instability Metrics as a function of increasing Latitude. None of the p-values for the different metrics was found non-significant. This suggests the search for a higher polynomial model to best describe the relationship between these variables ............................................. 32

4.3 Second-degree polynomial model for fitting Instability Metrics as a function of increasing Latitude. Blue labels for p-values highlight a significant relationship; whereas red labels the opposite. This second-order model suggest a good significant fit .......................................................... 33

4.4 Third-degree polynomial model for fitting Instability Metrics as a function of increasing Latitude. Blue labels for p-values highlight a significant relationship; whereas red labels the opposite. The cubic term does not justify the additional complexity over a significant relationship or a higher explained variance of the data ............. 33

4.5 Statistical results for a linear regression model fit to relate Instability Metrics as a function of Surface Temperature temporal changes ............................................. 34

4.6 Second-degree polynomial fitting to model Instability Metrics as a function of temporal changes in Surface Temperature Waters. Red labels for p-values indicate a not justifiable complexity introduced by the model .................................................. 35

4.7 Statistical results for a linear regression fit to model Instability Metrics as a function of Bottom Water Temperature Patterns. Blue labels for p-values highlight a significant model for this relationship captured by a linear model ............................................. 35

4.8 Second-degree polynomial fit to model Instability Metrics as a function of temporal changes in Bottom Water Temperatures. Red labels for the p-value in the convex hull case indicate a not justifiable nor general complexity introduced by adding a quadratic term to the model .................................................. 36

4.9 Statistical results for multiple regression for community instability metrics as a function of a linear combination of temporal changes in surface and bottom water temperatures ............................................. 37
ABSTRACT

Stability of Marine Communities in Response to Climate Change

by

Maria Salome Perez Rosero
Master of Science in Electrical and Computer Engineering
Northeastern University, August 2017
Dr. Jennifer Dy, Advisor

This work presents an intuitive method for studying the stability of fish communities in response to climate change via a three-step approach: (i) quantification of climate change across different latitudes, (ii) quantification of community instability, and (iii) quantification of the relationship between climate change and community instability at each location. Abundance data and other biological and environmental variables were recorded for groundfish trawl surveys conducted in the East Bearing Sea during from 1982 to 2011. First, we define climate change as the relationship between (surface, bottom) temperature and time at each location. A significant linear relationship between latitude and bottom water temperature suggests that a change of 1 degree in latitude towards the North Pole leads to an increase in bottom water temperature of 0.035 degrees Celsius. Second, we propose to measure community instability through the computation of three community instability metrics: pairwise distance between observation points/site scores, mean distance to the centroid point, and the area of a fitted convex hull containing all observation points. We compute instability metrics from a low-dimensional representation of the original data, which is built using a non-metric multidimensional scaling technique. Once the instability metrics are calculated, we regress them as a function of latitude to study how the stability of communities changes across locations. We found that all three instability metrics increased with latitude. This trend could be due to the fact that increases in bottom temperature destabilize ecological communities by causing species to migrate to locations with more stable environmental conditions. Finally, with the intuition of how latitude drives the behavior of community stability, we regress the instability metrics as a function of the temporal change in temperature in order to find the environmental factors that are being "proxied" by latitude in the former relationship. We show that towards northern latitudes, where community instability is higher, temporal changes in temperature variables suggest an inverse relationship, with faster temporal changes in the surface water temperatures and slower temporal changes in the bottom water temperatures.
Chapter 1

Introduction

Climate change and ongoing global warming has had large effects in different ecosystems worldwide \[2, 3, 4\]. Many efforts associate climate change mainly with an increase in temperature. Indeed, many studies focus on the effects of increased temperature to predict how climate change will likely affect natural ecosystems \[2, 3\]. This change in temperatures can lower abundances and even lead to extinction. Also, the biological rates of species including metabolism, growth and feeding are dependent on temperature \[4\]. The influence of these changes on population dynamics, however, remains uncertain; thus, the uncertainty about ecological stability in response to climate change remains high.

This work aims to develop an intuitive method to study the stability of marine communities in response to climate change. Climate change is expected to affect different environmental variables; but we will focus on temperature since this environmental variable has been studied extensively and shown to be one of the strongest abiotic manifestations of climate change \[3, 1, 5\]. We predict that since climate change is meant to have a higher impact on latitudes further from the Equator, those communities towards northern latitudes are more likely to be unstable. This prediction can be summarized in Figure 1.1 by which we imply that towards the poles, climate change is producing a higher environmental change; and as a result, communities are experiencing more instability.

The source of the data used for the analysis in this thesis is the East Bearing Sea. Abundance data and other biological and environmental variables were recorded for a groundfish trawl survey from 1982 to 2011. Once the raw input data is arranged into unique combinations of latitude, longitude and year observation points, the related challenges are the high-dimensionality of the data, since the diversity of unique marine species is large; and sparsity, since it is very unlikely to record the abundance of a certain fish specie at multiple locations. Therefore, to be able to extract information
out of this high-dimensional and sparse data, and make these data accessible to visual inspection and exploration, we need to map it into a low-dimensional space \[6\]. In the low dimensional space, the corresponding embedded representation of the groundfish abundance data is the input for quantifying instability via three proposed metrics: mean pairwise distance between site or community sample scores/low dimensional observation points, mean distance between all site scores and the centroid, and the area of a convex hull containing all site scores. Once the instability metrics are computed, we regress them as a function of latitude to study how community stability changes towards northern and southern locations. With the intuition of how latitude drives the behaviour of community instability, we regress the instability metrics as a function of the temporal change in temperature. With this new relationship, we find the environmental factors that approximate the behaviour of latitude in the former relationship, and play a role to explain community stability trends.

Our study suggests that climate change is driving an increase of bottom temperature by 0.035 degrees Celsius towards northern latitudes, and that this rise in bottom temperature is happening at a slow pace when compared with the corresponding bottom temperature changes occurring towards southern latitudes. Furthermore, all three instability metrics computed for the dataset under study show an increasing pattern towards poleward locations. The instability trend can also be explained through the previous bottom temperature behaviour: since bottom temperature at poleward locations is increasing, species might be migrating away from these non-ideal locations and thus increasing community instability.
CHAPTER 1. INTRODUCTION

Finally, we present an inverse relation between the temporal changes of surface and bottom temperature when modeled as environmental representatives to explain the increase in community instability towards northern latitudes.

1.1 Related Work

Efforts to define how marine organisms adapt or move in response to climate change have been explored in [1, 5, 7]. These approaches define marine organism shifts as a function of different climate models and also climate envelope models, which aim to correlate through spatiotemporal factors, how environmental variables are affecting the migration of marine species.

For instance, Pinsky et al. in [1] compiled four decades of scientific surveys of fish invertebrates from continental shelves of North America across the different regions listed in Table 1.1. These surveys captured 128 million organisms from 360 species. They were analyzed to probe that differences in climate velocity, defined as the rate and direction that climate shifts across the landscape, can explain observed species shifts expected as a result of the adaptation and movement of species in response to climate change. Authors in [1] measured range shifts by tracking the location of range centroids, and by defining assemblages as the set of sampled taxa within a geographic region, showed that four assemblages shifted poleward whereas five shifted south. They also, concluded that bottom temperature explained more than half of the variation in assemblage shifts, and that surface temperature trends were not correlated to latitudinal shifts.

Mellin in [5] states that the lack of effective action to reduce global warming has increased the need for effective management to improve ecosystem resilience, understood as the ability to resist and recover from disturbance. Using a unique long-term and broad-scale set of fish of 215 species across 150000 km², differences in community resistance were explored. The authors quantified temporal change in community composition using a multivariate, temporal index of community dissimilarity and compared the timing and magnitude of its responses to disturbance using linear models. The work concludes by discussing the fact that despite of the linear decline in average coral cover on the Australia’s Great Barrier Reef, the study demonstrates that Marine Protected Areas have increased both the resistance and recovery of coral reef community composition in response to a range of disturbances. Protection from fishing was associated with reduced impacts over the five years in which these reef communities showed a rapid recovery.

Moving towards the prediction of impacts of climate change on the distribution of species, Pearson and Dawson in [8] characterize the potential impacts of climate change on the natural
CHAPTER 1. INTRODUCTION

distribution of species by means of a bioclimate envelope. Some of the critiques of this approach have questioned its validity by pointing the many factors (different than climate) that play a role in determining species distributions and the dynamics of distribution changes. The authors propose that, although the complexity of the natural system presents fundamental limits to predictive modelling, the bioclimate envelope approach can provide a useful first approximation as to the potentially dramatic impact of climate change on biodiversity.

<table>
<thead>
<tr>
<th>Region</th>
<th>Season</th>
<th>Years</th>
<th>Freq.</th>
<th>Taxa ≥ 1x/yr</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eastern Bering Sea</td>
<td>Summer</td>
<td>1982-2011</td>
<td>1</td>
<td>74</td>
<td>AFSC</td>
</tr>
<tr>
<td>Gulf of Mexico SEAMAP</td>
<td>Summer</td>
<td>1987-2011</td>
<td>1</td>
<td>106</td>
<td>GSMFC</td>
</tr>
<tr>
<td>Northeast U.S.</td>
<td>Spring</td>
<td>1968-2008</td>
<td>1</td>
<td>41</td>
<td>NEFSC</td>
</tr>
<tr>
<td>Scotian Shelf</td>
<td>Summer</td>
<td>1970-2011</td>
<td>1</td>
<td>28</td>
<td>DFO1</td>
</tr>
<tr>
<td>Southern Gulf of St. Lawrence</td>
<td>Fall</td>
<td>1971-2009</td>
<td>1</td>
<td>14</td>
<td>DFO2</td>
</tr>
<tr>
<td>Newfoundland</td>
<td>Fall</td>
<td>1995-2011</td>
<td>1</td>
<td>83</td>
<td>DFO3</td>
</tr>
</tbody>
</table>

Table 1.1: Experimental ground fish data of related works in [1]

- **AFSC**: Alaska Fisheries Science Center Resource Assessment and Conservation Engineering.
- **GSMFC**: Gulf States Marine Fisheries Commission (GSMFC).
- **NEFSC**: Northeast Fisheries Science Center.
- **DFO1**: Department of Fisheries and Oceans of Scotian Shelf Groundfish.
- **DFO2**: Aquatic Resources Division, Science Branch, Department of Fisheries and Oceans, New Brunswick.
- **DFO3**: Newfoundland and Labrador Autumn multispecies trawl surveys, Science Branch, Department of Fisheries and Oceans, Canada.

In respect to the quantification of community instability, authors in [9] review the definition of stability in Ecology as the knowledge of how species interact with each other and how each is affected by the environment. Some theoretical research in stability has explored how numerous features of ecosystems affect stability, including diversity (number of species), the strength of iterations among species, the topology of food webs, and the sensitivities of species to different types of environmental perturbations. As a matter of fact, there are many concepts related to stability that describe the different properties of the relationships that exist within an ecosystem. In this direction,
authors in [9] recommend: rather than searching for generalities in patterns of diversity-stability relationships, to investigate mechanisms. They state the need of more studies to reveal exactly what these generating mechanisms are; and how they can be reproduced in a statistically robust way, to capture the diversity-stability dynamics of the ecosystem.

Also, authors in [10] provide a description of how resilience, resistance, and stability are measured in Ecology. They define a stable system as the one able to return to a certain initial equilibrium after following perturbation. This idea associates stability with the ability of an ecosystem to return to its initial equilibrium after following perturbations. Perturbations may involve changes in species abundances; and others may involve the removal of some or all of the species in the ecosystem.

1.2 Contributions

This work focuses on the concept of stability that involves an integrated measure of the entire ecosystem (e.g., the average density/abundance of all species); and relates the idea introduced in [10, 11] of associating stability with the ability of an ecosystem to return to its initial condition (e.g., initial abundance score, equilibrium), via three proposed instability metrics.

These instability metrics are designed to capture the concept of how far away are the observation points between the abundance of each fish specie that integrates a marine community via the computation of the mean of pairwise distances in a low dimensional fish abundance space, the mean of these distances to a centroid point, and the area of a fitted convex hull containing these observations. The proposed community instability metrics add practical knowledge to quantify climate-related ecological risk to communities across different locations. This information could be used to devise management plans by identifying regions of high climate risk where relieving local stressers (e.g., harvesting, pollution) could promote stability and ensure the persistence of vulnerable communities.

Also, we present a three-step approach to explore community instability through (i) modelling temperature as a function of latitude, (ii) exploring the relationship of community instability as a function of latitude, and (iii) searching for climate (environmental) patterns that approximate the behaviour of latitude in step (ii). For the data compiled around latitudes of the East Bearing Sea, we conclude that increased community instability was found towards northern latitudes, and that this instability is significant driven by increments in bottom water temperatures at poleward locations.
Chapter 2

Background

This section presents the theory related to the multidimensional scaling technique used for finding a low-dimensional representation of the original data. Section 2.1 introduces the generic formulation for multidimensional scaling techniques (MDS) and its two variants: metric and nonmetric cases. Sections 2.2 and 2.3 provide detail about the techniques used to approach the minimization of the objective function that characterizes multidimensional scaling techniques, known as Stress; and also explain the algorithm that we used to approach the Stress minimization for the nonmetric version of MDS. Finally, Section 2.4 narrows down the practical use of multidimensional scaling techniques for the case of Ecological data.

2.1 Multidimensional Scaling

Multidimensional scaling (MDS) is the method that represents measurements of similarity (or dissimilarity) among pairs of objects as distances between points of a low-dimensional space. MDS performs this representation of the $n$ points in a low dimensional space, so that the inter-point dissimilarities in the new space preserve as well as possible the ranking/ordering of the corresponding dissimilarities in the original space [12].

Let $X$ be the original data comprised of $n$ points in $p$ dimensional space with dissimilarities $d_{ij}$ between points $i$ and $j$ for $i, j \in \{1, \ldots, n\}$.

$$X \in R^{n \times p} \quad (2.1)$$
**CHAPTER 2. BACKGROUND**

\[ d_{ij} = F(x_i, x_j) \quad \forall i, j \in \{1, ..., n\} \]

\[
d_{ij} = \begin{cases} 
    d_{ij} = 0 & \text{if } i = j \\
    d_{ij} = d_{ji} & \text{otherwise}
\end{cases}
\]

(2.2)

Also, let \( \hat{X} \) be a low-dimensional embedding (configuration) of \( X \), such that:

\[
\hat{X} \in \mathbb{R}^{n \times p'} \quad p' << p
\]

(2.3)

Dissimilarities are also computed in the low-dimensional space applying the same function \( F \) as above, but now over the points of \( \hat{X} \):

\[
\hat{d}_{ij} = F(\hat{x}_i, \hat{x}_j) \quad \forall i, j \in \{1, ..., n\}
\]

\[
\hat{d}_{ij} = \begin{cases} 
    \hat{d}_{ij} = 0 & \text{if } i = j \\
    \hat{d}_{ij} = \hat{d}_{ji} & \text{otherwise}
\end{cases}
\]

(2.4)

The goal is to find a low-dimensional embedding \( \hat{X} \) so that the distances between pairs of points \( \hat{d}_{ij} \) and the distances \( d_{ij} \) in the original space match orderwise as well as possible. The quantity that measures the above description is known as Stress \( S \), and it is defined in Equation 2.5:

\[
S(\hat{d}, \tilde{d}) = \sqrt{\frac{\sum_{i<j \leq n} (\hat{d}_{ij} - \tilde{d}_{ij})^2}{\sum_{i<j \leq n} (d_{ij})^2}}
\]

(2.5)

where \( \tilde{d}_{ij} \) is the disparity between points \( x_i \) and \( x_j \) for \( i, j \in \{1, ..., n\} \).

Disparities can be defined in two different ways:

- **Metric**: The disparities are assumed to be equal to the original dissimilarities.
  \[
  \tilde{d}_{ij} = d_{ij}
  \]

The dissimilarities are assumed to be a metric distance that satisfies the following conditions:
CHAPTER 2. BACKGROUND

Let $d : O \times O \rightarrow [0, \infty)$, so that $\forall x, y, z \in O, O \in \mathbb{R}^p$:

1. $d(x, y) \geq 0$ \hspace{1cm} \text{Non-negativity}
2. $d(x, y) = 0 \iff x = y$ \hspace{1cm} \text{Identity of indiscernibles}
3. $d(x, y) = d(y, x)$ \hspace{1cm} \text{Symmetry}
4. $d(x, z) \leq d(x, y) + d(y, z)$ \hspace{1cm} \text{Triangle inequality}

The \textit{triangle inequality} states that for any triangle, the sum of the lengths of any two sides must be greater than or equal to the length of the remaining side. In a metric sense, this means that the distance from $x$ to $z$ via $y$ is greater or equal than the distance from $x$ to $z$ directly.

- Non-metric: If we relax the metric requirement, we can define disparities as a monotonic function of the original dissimilarities.

\[
\tilde{d}_{ij} = f(\hat{d}_{ij}, d_{ij})
\]

s.t. \hspace{0.5cm} $\tilde{d}_{ij} \in D_{\tilde{d}}$

\[
D_{\tilde{d}} = \{\tilde{d}, d \in \mathbb{R}^E : \tilde{d}_{ij} \leq \tilde{d}_{i'j'} \iff d_{ij} \leq d_{i'j'}\}
\]

\[
E = \{(i, j) : i \leq j \text{ } \& \text{ } (i, j) \in \{1, \ldots, n\}\}
\]

This monotonic transformation is applied to the low dimensional space distances ($\hat{d}_{ij}$) only considering the ranking (ordering) of the original dissimilarities $d_{ij}$. In other words, the original distances are only considered to define the ordering for the disparities through an isotonic regression process [6].

In the context of MDS, isotonic regression is referred to as a monotone least squares regression of the disparities $\tilde{d}_{ij}$ on $\hat{d}_{ij}$, as presented in Equation 2.7.

\[
\min_{s.t.d_{ij} \in D_{\tilde{d}}} \sum_{i} w_i (\hat{d}_{ij} - \hat{d}_{ij})^2
\]

where $w_i$ are weights $\in \mathbb{R}^n$ (usually set to 1).

Synthesizing the information above, the central concept of Multidimensional Scaling is that dissimilarities in the original space should correspond to the corresponding disparities in the low-dimensional space. In the metric case, we desire to express this relationship by means of a
CHAPTER 2. BACKGROUND

metric formula. However, when the numerical computation of the parameters in a metric formula is not desired/feasible or when it is not possible in practice to achieve the desired fixed relation; it is also possible to search for a relationship which is not described by a metric formula, but by the fact of modeling the low-dimensional disparities as a monotonic increasing function of the original dissimilarities [12].

Now, in order to encompass both metric and non-metric cases, Stress function introduced in Equation 2.5 can be expressed in general as:

\[
S(\hat{X}, \tilde{d}) = \sqrt{\frac{\sum_{i<j\leq n} (F(\hat{x}_i, \hat{x}_j) - \tilde{d}_{ij})^2}{\sum_{i<j\leq n} (F(\hat{x}_i, \hat{x}_j))^2}} \quad (2.8)
\]

Finally, from Equation 2.8 the goal is to find the low dimensional embedding \((\hat{X})\) that produces the lowest stress, such that:

\[
S^* = \min_{\hat{X},\tilde{d} \in D_{\tilde{d}}} S(\hat{X}, \tilde{d}) \quad (2.9)
\]

There are several algorithms one can utilize to minimize Equation 2.9. Some approaches towards this minimization problem utilize iterative steepest descent algorithms [12, 13, 14]; another approach utilizes an iterative majorization algorithm referred to as Scaling by Majorizing a Complicated Function (SMACOF) [15, 6].

Finding the minimum of a function \(f(x)\) is not always viable through the computation of its derivative \(f'(x)\), and set the former equal to zero to solve for \(x\). Sometimes the derivative is not defined everywhere or solving the equation \(f'(x) = 0\) is not possible. For those cases, there are other numerical mathematical techniques such as the method of Iterative Majorization, based on the work of De Leeuw, back in 1997 [15]. The majorization algorithm, introduces a better method in terms of guarantees and rate of convergence for minimizing stress. Here, each step is linked to the minimization of a simple convex function, which bounds Stress from above and touches its surface at a single support tangent point. The following subsections mainly rely on the information provided in [15] and [6].

2.2 Principles of Majorization Algorithm

The Majorization method replaces iteratively an original complicated function \(f(x)\) by an auxiliary function \(g(x, z)\), where \(z\) in \(g(x, z)\) is some fixed value.
CHAPTER 2. BACKGROUND

Figure 2.1: Minimizing a function $f(x)$ by iterative majorization approximations $g(x, z)$.

$f(x) = 6 + 3x + 10x^2 - 2x^4$ (red curve), and $g(x, z) = 6 + 3x + 10x^2 - 8xz^3 + 6z^4$ (blue, green curves). First iteration point starts at $x_0 = 1.4$ (generates blue curve), followed by an update at $x = 0.95$ (generates green curve). The process repeats successively until convergence.

The function $g$ needs to meet the following requirements to be considered a majorization function of $f(x)$:

- The function $g(x, z)$ should be simpler to minimize than $f(x)$. For example, if $g(x, z)$ is a quadratic function in $x$, then the minimum of $g(x, z)$ over $x$ can be computed in one step.

- The original function must always be smaller than or at most equal to the auxiliary function $g(x, z)$; that is, $f(x) \leq g(x, z)$.

- The auxiliary function should touch the surface at the called supporting tangent point $z$; that is, $f(z) = g(z, z)$.

Now, to understand the principle of minimizing a function by majorization, let the minimum of $g(x, z)$ over $x$ be attained at $x^*$. The last two requirements of the majorizing function imply the chain of inequalities:

$$f(x^*) \leq g(x^*, z) \leq g(z, z) = f(z)$$

(2.10)

Note that the minimum of the majorizing function $g(x^*, z)$ is squeezed between $f(x^*)$ and $f(z)$. A graphical representation of these inequalities is included in Figure 2.1. Also, by Equation 2.10 the majorization algorithm yields a non-increasing sequence of function values, which is an attractive
CHAPTER 2. BACKGROUND

aspect of iterative majorization. Furthermore, a necessary condition for a minimum at \( x^* \) is that the derivative of \( f(x) \) at \( x^* \) is 0. Using Equation 2.10 also implies that \( x^* \) minimizes \( g(x, x^*) \) over \( x \) with \( g(x^*, x^*) \) as the minimum. Thus, the necessary condition of a zero derivative at a local minimum may be replaced by the weaker condition that \( g(x^u, z) = f(z) \) and \( x^u = z \).

2.3 Minimizing MDS Objective Function by Majorization Algorithm

The Stress objective function in Equation 2.8 can be written as:

\[
S(\hat{X}) = \sum_{i<j} w_{ij} (d_{ij} - \hat{d}_{ij}(\hat{X}))^2
\]

\[
= \sum_{i<j} w_{ij} d_{ij}^2 + \sum_{i<j} w_{ij} \hat{d}_{ij}(\hat{X}) - 2 \sum_{i<j} w_{ij} d_{ij} \hat{d}_{ij}(\hat{X})
\]

\[
= \eta_d^2 + \eta^2(\hat{X}) - 2\rho(\hat{X}) \tag{2.11}
\]

From Equation 2.11, we obtain an equivalent expression for Stress into three main parts detailed below:

1. \( \eta_d^2 \), is not dependent on \( \hat{X} \) (constant).
2. \( \eta^2(\hat{X}) \) is a weighted sum of the squared distances \( \hat{d}_{ij}(\hat{X}) \)
3. \( 2\rho(\hat{X}) \) is a weighted sum of the distances \( \hat{d}_{ij}(\hat{X}) \).

To proceed, we need to make one additional assumption: the weight matrix \( W \) is irreducible, that is, there exists no partitioning of objects into disjoint subsets, such that \( w_{ij} = 0 \) whenever points \( i \) and \( j \) are in different subsets. We further simplify the second and third terms as follows.

Second Term \( \eta^2(\hat{X}) \):

\[
d_{ij}^2(\hat{X}) = tr(\hat{X}'A_{ij}\hat{X})
\]

\[
w_{ij}d_{ij}^2(\hat{X}) = w_{ij}tr(\hat{X}'A_{ij}\hat{X})
\]

\[
w_{ij}\hat{d}_{ij}(\hat{X}) = tr(\hat{X}'(w_{ij}A_{ij})\hat{X})
\]

Where,

\[
A_{ij} = \begin{cases} 
    a_{ii} = a_{jj} = 1 \\
    a_{ij} = a_{ji} = -1 \\
    \text{All other } a_{ij} = 0
\end{cases}
\]
CHAPTER 2. BACKGROUND

And,

\[ \eta^2(\hat{X}) = \sum_{i<j} w_{ij} \hat{d}^2_{ij}(\hat{X}) = tr \ X' (\sum_{i<j} w_{ij} A_{ij}) X = tr \ (X' V X) \tag{2.14} \]

\[ V = \sum_{i<j} w_{ij} A_{ij} \]

\[ V = \begin{cases} v_{ij} = -w_{ij} & \text{if } i \neq j \\ v_{ii} = \sum_{j=1,j\neq i}^n w_{ij} & \text{if } i = j \end{cases} \tag{2.15} \]

**Third Term \( \rho(\hat{X}) \):**

The \( \rho(\hat{X}) \) is minus a weighted sum of the distances:

\[ -\rho(\hat{X}) = -\sum_{i<j} (w_{ij} d_{ij}) \hat{d}_{ij}(\hat{X}) \tag{2.16} \]

In order to obtain a majorizing inequality for minus the distance \((-d_{ij}(X))\) in Equation 2.16, we apply the Cauchy-Schwarz inequality (Equation 2.17). By substituting \( p_a \) in Equation 2.17 by \((x_{ia} - x_{ja})\) and \( q_a \) by \((z_{ia} - z_{ja})\), we obtain the desired inequality through Equations from 2.18 to 2.23 detailed as follows:

\[ \sum_{a=1}^m p_a q_a \leq (\sum_{a=1}^m p_a^2)^{1/2} (\sum_{a=1}^m q_a^2)^{1/2} \quad \text{if } q_a = c p_a \tag{2.17} \]

\[
\sum_{a=1}^m (\hat{x}_{ia} - \hat{x}_{ja})(z_{ia} - z_{ja}) \leq (\sum_{a=1}^m (\hat{x}_{ia} - \hat{x}_{ja})^2)^{1/2} (\sum_{a=1}^m (z_{ia} - z_{ja})^2)^{1/2} \\
= d_{ij}(\hat{X}) d_{ij}(Z) \quad \text{if } Z = \hat{X} \tag{2.18} \]

Dividing both sides of Equation 2.18 by \( d_{ij}(Z) \), and multiplying the corresponding expression by \(-1\), we obtain Equation 2.19:

\[ -d_{ij}(\hat{X}) \leq -\frac{\sum_{a=1}^m (\hat{x}_{ia} - \hat{x}_{ja})(z_{ia} - z_{ja})}{d_{ij}(Z)} \tag{2.19} \]

Due to the positivity of \( d_{ij}(X) \), it remains true that \(-d_{ij}(X) \leq 0\); and therefore, Equation 2.19 can be expressed as:
CHAPTER 2. BACKGROUND

\[
\sum_{a=1}^{m} (\hat{x}_{ia} - \hat{x}_{ja})(z_{ia} - z_{ja}) = tr(\hat{X}'A_{ij}Z)
\]  
(2.20)

Finally, combining Equations 2.19 and 2.20, a simple expression for \(-\rho(\hat{X})\) in Equation 2.16 is presented by Equation 2.21.

\[
-\rho(\hat{X}) = -\sum_{i<j} (w_{ij}d_{ij})\hat{d}_{ij}(\hat{X}) \leq -tr(\hat{X}'(\sum_{i<j} b_{ij}A_{ij}) Z)
\]
(2.21)

where \(B(Z)\) is comprised of the elements:

\[
b_{ij} = \begin{cases} 
-\frac{w_{ij}d_{ij}}{d_{ij}(Z)} & \text{for } i \neq j \text{ and } \hat{d}_{ij}(Z) \neq 0 \\
0 & \text{for } i \neq j \text{ and } \hat{d}_{ij}(Z) = 0
\end{cases}
\]
(2.22)

\[
b_{ii} = -\sum_{j=1, j\neq i}^{n} b_{ij}
\]

Since the equality in Equation 2.21 occurs at \(Z = X\), \(-\rho(\hat{X})\) can be majorized by the linear function in terms of \(\hat{X}\) included in Equation 2.23.

\[
-\rho(\hat{X}) = -tr(\hat{X}'B(\hat{X})\hat{X}) \leq -tr(\hat{X}'B(Z)Z)
\]  
(2.23)

Now, by combining all the three former reduced expressions, the Stress function in Equation 2.24 provides a simplified expression for the majorization upper bound auxiliary function:

\[
S(\hat{X}) = \eta^{2}_d + tr(\hat{X}'V\hat{X}) - 2 \cdot tr(\hat{X}'B(\hat{X})\hat{X})
\]

\[
S(\hat{X}) \leq \eta^{2}_d + tr(\hat{X}'V\hat{X}) - 2 \cdot tr(\hat{X}'B(\hat{X}Z)) = \tau(\hat{X}, Z)
\]  
(2.24)

With \(\tau(\hat{X}, Z)\) as the proposed auxiliary function for the Majorization Algorithm, we compute the gradient and equate to zero to find a simplified gradient update in Equation 2.25.

\[
\nabla_{\tau}(\hat{X}, Z) = 2V\hat{X} - 2B(Z)Z = 0
\]
(2.25)

\[
V\hat{X} = B(Z)Z.
\]

13
CHAPTER 2. BACKGROUND

However, the $V$ matrix in the last expression of Equation 2.25 is not full rank, and its inverse can not be computed. In order overcome the rank restriction, we apply a Guttman Transformation (Equation 2.26) to be able to find a closed simplified expression to update $\hat{X}^{k+1}$.

$$\hat{X}^{k+1} = V^+ B(Z)Z,$$

(2.26)

where $V^+ = n^{-1}J$ is the Moore-Penrose inverse with $J$ denoting the centering matrix. Finally, the update for $\hat{X}^{k+1}$ is given by Equation 2.27

$$\hat{X}^{k+1} = n^{-1}B(Z)Z.$$  

(2.27)

2.3.1 Majorization Algorithm for nMDS

Algorithm 1 summarizes the steps required in order to compute the non-metric version of a multidimensional scaling algorithm. The utility of applying a majorization algorithm to minimize the Stress function is linked to the utility of approximating the multivariate gradient by the ratio between the dissimilarities and disparities. This provides a computationally efficient approach to manage Stress significant computational advantage to manage the stress minimization by a closed-form expression for updating the low dimensional points in ($\hat{X}$).

Algorithm 1 Pseudo-code for Multidimensional Scaling: Metric and Non-metric cases.

1: procedure MDS($X, p' = 2, M$)
2: $X =$ original data $\in \mathbb{R}^{n \times p}$ $\rightarrow$ $d_{ij} = F(x_i, x_j)$
3: $\hat{X}^k =$ random initialization $\in \mathbb{R}^{n \times p'}$
4: while ($S < \epsilon$) do
5: $\hat{d}_{ij} = F(\hat{x}^k_i, \hat{x}^k_j)$
6: if $M = 1$ then $\triangleright M = 1$ for Metric MDS
7: $\tilde{d}_{ij} = d_{ij}$
8: else $\triangleright M = 0$ for nonMetric MDS
9: $\tilde{d}_{ij} = \text{isotonicRegression}(\hat{d}_{ij}, \text{ranking} [d_{ij}])$
10: end if
11: $S = \text{Stress}(\hat{d}_{ij}, \tilde{d}_{ij})$
12: $B = -\frac{\text{disparities}(\hat{d}_{ij})}{\text{dissimilarities}(\hat{d}_{ij})}$ $\triangleright$ Gradient approximation (SMACOF)
13: $\hat{X}^{k+1} = n^{-1}(B \cdot \hat{X}^k)$
14: end while
15: return $\hat{X}^k$ $\triangleright$ Low-dimensional representation of $X$ (min. Stress)
16: end procedure
2.4 Multidimensional Scaling for Ecology

In the field of Ecology, Multidimensional Scaling Techniques (the non-metric case primarily), serve as methods to represent dissimilarity data as distances in a low-dimensional space, in order to make these data accessible to visual inspection and exploration via two (2D) or three (3D) dimensional plots \([6]\). Specifically for this work, we map \(X\) to a low dimensional embedding in 2D \((p' = 2)\).

From the literature in \([16, 5, 17, 18, 19]\), the Bray-Curtis dissimilarity is widely used in Ecological data processing. Equation 2.28 presents the corresponding expression for the Stress function based on the Bray-Curtis dissimilarity.

\[
S_{\text{Bray}}(\hat{X}, \tilde{d}) = \sqrt{\frac{\sum_{i<j \leq n} \left( \frac{\sum_{k=1}^{p'} |\hat{x}_{ik} - \hat{x}_{jk}|}{\sum_{k=1}^{p'} (\hat{x}_{ik} + \hat{x}_{jk})} - \tilde{d}_{ij} \right)^2}{\sum_{i<j \leq n} \left( \frac{\sum_{k=1}^{p'} |\hat{x}_{ik} - \hat{x}_{jk}|}{\sum_{k=1}^{p'} (\hat{x}_{ik} + \hat{x}_{jk})} \right)^2}} \tag{2.28}
\]

Some ideal characteristics attributed to Bray-Curtis dissimilarities for Ecological data processing are: its invariant to changes in units; its unaffected by additions or removals of species that are not present in two communities; its unaffected by the addition of a new community; and it can be sensitive to differences in total abundances when relative abundances are the same \([19]\).

Also, authors in \([5, 17, 18]\) have used Bray-Curtis dissimilarities to study resilience and stability in Ecological datasets.
Chapter 3

Methods for Community Stability in Response to Climate Change

This work aims to develop an intuitive method to study the stability of marine communities in response to climate change. Figure 3.1 presents the three main components of our approach: data acquisition, data processing and data analysis. After the corresponding acquisition and processing stages, which are described in detail in Sections 3.1 and 3.2, the main study of community stability in response to climate change is explained in Section 3.3 and is summarized as follows. First, we start defining climate change as the relationship of the change of temperature over time. We focus on temperature since this environmental variable was consistently measured across the dataset used for this study. Second, we propose to measure community instability through the computation of three metrics (e.g., mean pairwise distance between observation points, mean distance to the centroid point, and the area of a fitted convex hull containing the observation points). Once these community instability metrics are computed, we regress them against latitude, as to analyze whether poleward latitudes are more likely to experience instability in their communities. Finally, as for the third step, we regress community instability as a function of environmental variables (e.g., mean value of surface and bottom water temperatures and their corresponding temporal change). This regression allows us to find the environmental factors that approximate the behaviour of latitude in the former relationship of instability modeled as a function of latitude, and play a significant role in explaining the behaviour of these community instability patterns.
CHAPTER 3. METHODS FOR COMMUNITY STABILITY IN RESPONSE TO CLIMATE CHANGE

Figure 3.1: Block diagram for quantifying climate change and community instability. Our approach consists of three main blocks (data acquisition, data processing, data analysis); and within the data analysis, we propose a three step study that encompasses the quantification of climate change, instability at poleward locations, and climate change patterns driving instability to investigate about the stability of communities in response to environment/climate change.

3.1 Data Acquisition

The source of data used for this work is from a groundfish trawl survey performed in the East Bearing Sea during the period from 1982 to 2016. This dataset is publicly available from the Alaska Fisheries Science Center (AFSC) Resource Assessment and Conservation Engineering (RACE)\(^1\). The data contains the abundance of the different species that were found while conducting the experiment. The abundance metric we used from this dataset is specified by the amount of catch fish weight per area that the net swept, also known as WTCPUE and it is measured in kg/Ha (kilogram/hectare). This abundance is reported for each of the fish species that were recorded in the survey. Regarding the biological information in the dataset, the number of fish species reported sum up to 860, and the environmental variables recorded alongside the abundance data are surface and bottom temperatures. Other details are included in the description of Table 3.1.

In Figure 3.1 the input data is labeled as \(X_{[n \times l]}\), where \(n\) is the number of total observations and \(l\) is the number of columns that contain all the biological information in the raw format from the dataset website.

\(^{1}\text{https://www.afsc.noaa.gov/RACE/groundfish/survey_data/default.htm}\)
CHAPTER 3. METHODS FOR COMMUNITY STABILITY IN RESPONSE TO CLIMATE CHANGE

Figure 3.2: Geographical location of the groundfish trawl survey dataset conducted in the East Bearing Sea during the summer periods from 1982 to 2016. This is the input data for the study conducted in this thesis.

3.2 Data Processing

The data processing block consists of all steps that were performed to process the input data into community instability measurements, as well as their corresponding environmental factors/variables. As included in Figure 3.1, two main processing threads can be distinguished as explained below:

3.2.1 Community Matrix

The first thread takes the input matrix $X$ and builds the community matrix $X_c[n \times p]$, where $n$ are the unique (latitude, longitude) observation points, averaged per year, and $p$ represents the number of unique fish species across columns. The type of data in the community matrix is the abundance of each fish specie per (latitude, longitude) point.

Once the community matrix is built, we seek to find its low dimensional embedding (representation), $\hat{X}_c$, since the dimension of the original $X_c$ is complex to visualize and manipulate, we map it to a 2D low-dimensional embedding representation. This low dimensional embedding is found through a non-metric multidimensional technique (nMDS), detailed in Chapter 2. After applying nMDS to the original community matrix, $X_c$ turns into $\hat{X}_c[n \times p']$, where $n$ remains as the unique number of (latitude, longitude observations), and $p'$ is fixed to 2. This new low-dimensional space $R^{n \times p'}$ is known as the ordination space in Ecology.
**CHAPTER 3. METHODS FOR COMMUNITY STABILITY IN RESPONSE TO CLIMATE CHANGE**

<table>
<thead>
<tr>
<th>Category</th>
<th>Groundfish trawl survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>East Bearing Sea</td>
</tr>
<tr>
<td>Period</td>
<td>1982 to 2016 (35 years)</td>
</tr>
<tr>
<td>Latitude (Lat)</td>
<td>54.655°C to 62.019°C</td>
</tr>
<tr>
<td>Longitude (Lon)</td>
<td>−178.232°C to −158.307°C</td>
</tr>
<tr>
<td>No. raw observations</td>
<td>406,927</td>
</tr>
<tr>
<td>No. unique (Lat,Lon,Year)</td>
<td>12,961</td>
</tr>
<tr>
<td>Unique fish species</td>
<td>860</td>
</tr>
<tr>
<td>Environmental variables</td>
<td>Surface water temperature (sT)</td>
</tr>
<tr>
<td></td>
<td>Bottom water temperature (bT)</td>
</tr>
<tr>
<td>Temperature (sT,bT) scale</td>
<td>Tenths of a degree Celsius</td>
</tr>
<tr>
<td>sT range</td>
<td>−1.1 °C to 14.1 °C</td>
</tr>
<tr>
<td>bT range</td>
<td>−2.1 °C to 11.7 °C</td>
</tr>
</tbody>
</table>

Table 3.1: Description of the temporal and biological characteristics of the groundfish trawl survey data collected in the East Bearing Sea.

After the dimensionality reduction, we quantize the \( n \) (latitude, longitude, year) observation points into \( N \) equally-sized bins/groups (\( \hat{\mathbf{X}}_{ci,i=1:N} \)). The division of the latitude range into \( N \) groups is needed since it is a way to obtain specific measurements for a particular sub-range of latitude and differentiate changes as this sub-ranges goes from northern to southern latitudes.

Now, with the low-dimensional embedding data divided into \( N \) bins, we compute the so-called instability metrics. These metrics are intended to reflect community instability by measuring how far away are the low-dimensional embedding points between each other, for each bin. Specifically, we propose to test three different instability metrics: mean pairwise distance, distance to the centroid, and the area of a convex hull.

1. **Mean Pairwise Distances**

The mean pairwise distance between low-dimensional embedding points, \( \hat{x}_{c,\text{pairwise}} \), is calculated as follows. Given the \( N \) bins of \( k \) (latitude, longitude, year) observation points; with \( i = 1 : N \), for each of the \( i^{th} \) bin we have:

\[
\hat{x}_{c,\text{pairwise}}^i = \frac{1}{k/2(k-1)} \times \sum_{r,s=1 ; r \neq s}^{k/2(k-1)} ||\hat{x}_{c,r}^i - \hat{x}_{c,s}^i||
\]

\[
\hat{\mathbf{x}}_{c,\text{pairwise}} = [\hat{x}_{c,\text{pairwise}}^1, \ldots, \hat{x}_{c,\text{pairwise}}^N]
\]
CHAPTER 3. METHODS FOR COMMUNITY STABILITY IN RESPONSE TO CLIMATE CHANGE

Figure 3.3: Graphical representation of the quantification of community instability via the computation of pairwise distances between the low-dimensional (latitude, longitude, year) observation points/site abundance scores.

2. Distance to the Centroid

The mean distance between all low-dimensional points and its centroid, \( \hat{x}_{c, \text{distCentroid}} \), is calculated as follows. Given the \( N \) bins of \( k \) (latitude, longitude, year) observation points; with \( i = 1 : N \), for each of the \( i^{th} \) bin we have:

\[
\hat{x}^i_{c, \text{centroid}} = \frac{1}{k/2(k-1)} \times \sum_{r,s=1 ; r\neq s ; r,s \in \hat{x}^i_c} \frac{k/2(k-1)}{k/2(k-1)} (\hat{x}^r_c - \hat{x}^s_c)
\]

\[
\hat{x}^i_{c, \text{distCentroid}} = \frac{1}{k/2(k-1)} \times \sum_{t=1 ; t \in \hat{x}^i_c} ||\hat{x}^i_c - \hat{x}^t_{c, \text{centroid}}||
\]

\[
\hat{x}_{c, \text{distCentroid}} = [\hat{x}^1_{c, \text{distCentroid}}, \ldots, \hat{x}^N_{c, \text{distCentroid}}]
\]

3. Area of a Convex Hull

The area of a fitted convex hull, \( \hat{x}_{c, \text{convexHull}} \), is calculated as follows. Given the \( N \) bins of \( k \) (latitude, longitude, year) observation points; with \( i = 1 : N \), for each of the \( i^{th} \) bin, the area of a fitted convex hull containing the \( k \) observation points per bin is given by:
Figure 3.4: Graphical representation of the quantification of community instability via the computation of the mean distance to the centroid point of the low-dimensional (latitude, longitude, year) observation points/site abundance scores.

\[
CH^i = \text{Conv}(\hat{X}^i_c) = \{ \sum_{j=1 : j \in \hat{X}^i_c} \alpha_j \hat{x}^i_j \mid (j : \alpha_j \geq 0), \sum_{j=1}^k \alpha_j = 1 \}
\]

\[
\hat{x}^i_{c,\text{convexHull}} = \text{area}(CH^i)
\]

\[
\hat{x}^i_{c,\text{convexHull}} = [\hat{x}^1_{c,\text{convexHull}}, \ldots, \hat{x}^N_{c,\text{convexHull}}]
\]

3.2.2 Environmental Matrix

The second thread takes the input matrix \( X \) to build the environmental matrix \( X_{e[n \times q]} \), where \( n \) remains as the unique combination of (latitude, longitude) points, and \( q \) represents the environmental variables recorded for the dataset (e.g., surface and bottom water temperatures).

For the case of the environmental matrix, the binning of data was also performed in the same way as for the community matrix – separating the \( n \) observation points into \( N \) equally-sized bins. Now, instead of computing instability metrics as for the community matrix case, we select a bin representative value for each of the environmental variables in two different ways:

- Mean bin temperature: \( X_{em[i \times q]} = \text{mean}(\hat{X}_{ei})_{i=1:N} \)
CHAPTER 3. METHODS FOR COMMUNITY STABILITY IN RESPONSE TO CLIMATE CHANGE

Figure 3.5: Graphical representation of the quantification of community instability via the computation of the area of a fitted convex hull containing the low-dimensional (latitude, longitude, year) observation points/site abundance scores.

This bin representative is computed taking the mean value of each of the temperature variables for all observation points that belong to that bin.

- **Bin temperature slope:** $X_{es[i \times q]} = \text{slope}(\hat{X}_{ei})_{i=1:N}$
  
  This bin representative is computed by a linear regression of each temperature variable as a function of time. The corresponding slope from the resulting linear model is selected as the bin representative.

With the computation of the environment representatives for each of the bins, the data processing block finalizes to provide community instability measurements and their corresponding environmental factors, that will be the input for the next section of data analysis.

### 3.3 Data Analysis

This section details how we perform the study of community stability in response to climate change. This process is divided into three main parts: Subsection 3.3.1 starts with the quantification of climate change as the relation of temperature as a function of latitude; Subsection 3.3.2 continues with the study of how instability metrics are related as a function of latitude; and
3.3.1 Quantifying Climate Change Across Different Latitudes

For the quantification of climate change and following the notation of the block diagram of our approach in Figure 3.1 from the environmental matrix, $X_{e[n \times q]}$, where $n$ is the number of (latitude, longitude, year) unique low-dimensional observations/score points and $q$ the number of environmental variables; for each environmental variable $j$, we model: its mean, $x_{em,j}$, and its temporal change (slope), $x_{es,j}$, as a function of latitude, $f(latitude)$.

Equation 3.4 visualizes these relationships:

\[ x_{em,j} = f(latitude) \quad ; \quad j = 1 : q \]
\[ x_{es,j} = f(latitude) \quad ; \quad j = 1 : q \]  

(3.4)

3.3.2 Quantifying Instability at Poleward Locations

Once climate change is quantified through temperature as a function of latitude, the next step is to model community instability as a function of latitude, $f(latitude)$. This relationship is useful to study the tendency of communities to experience instability towards poleward locations.

Following the notation of the block diagram of our approach in Figure 3.1, the three instability metrics are specified by: pairwise, $\hat{x}_{c,\text{pairwise}}$, distance to the centroid, $\hat{x}_{c,\text{distCentroid}}$, and the area of a fitted convex hull, $\hat{x}_{c,\text{convexHull}}$. These metrics are used for this analysis as to corroborate a generic instability behaviour in the dataset under study. Also, recall that community instability measurements are proposed to capture in different ways how far away observation points are between each other.

Equation 3.5 specifies this relationship:

\[ \hat{x}_{c,\text{pairwise}}, \hat{x}_{c,\text{distCentroid}}, \hat{x}_{c,\text{convexHull}} = f(latitude) \]  

(3.5)

3.3.3 Quantifying Climate Change Patterns Driving Instability

Finally, the question turns into identifying the environmental patterns (related to climate change) that drive community instability at each location. By environmental patterns we refer to
CHAPTER 3. METHODS FOR COMMUNITY STABILITY IN RESPONSE TO CLIMATE CHANGE

those environmental variables that are a suitable approximation of the behaviour of latitude in the former model of instability as a function of latitude. In concrete and following as in previous sections the notation of our block diagram in Figure 3.1, this section models the three community instability metrics ($\hat{x}_{c,\text{pairwise}}$, $\hat{x}_{c,\text{distCentroid}}$, and $\hat{x}_{c,\text{convexHull}}$) as a function of the temporal change (slope) of the $q$ environmental variables (surface and bottom water temperatures), $x_{es}$.

Equation 3.6 presents the relationship described above:

$$\hat{x}_{c,\text{pairwise}}, \hat{x}_{c,\text{distCentroid}}, \hat{x}_{c,\text{convexHull}} = f(x_{es,j}) \ j=1:q$$

(3.6)
Chapter 4

Results And Discussion

This chapter presents the results of the regression models described in Chapter 3. Specifically, these regression models seek to quantify the stability of communities in response to climate change via a three-step approach: (i) quantification of climate change across different latitudes, (ii) quantification of community instability at poleward locations, and (iii) quantification of the relationship between climate change and community instability at each location.

First, we define climate change as the relationship between (surface, bottom) temperature and time at each location or latitudinal bin. Second, we propose to measure community instability via the computation of three metrics based on ordination plots (e.g., mean pairwise distance between site or community sample scores, mean distance between all site scores and the centroid, and the area of a convex hull containing all site scores). Once these community instability metrics are computed, we regress them against latitude to determine which locations are more likely to experience instability in their communities. As a third step, we regress community instability against environmental variables (e.g., mean value of surface and bottom water temperatures and their corresponding temporal change). This regression allows us to find the environmental factors that could potentially drive the latitudinal patterns in instability. These patterns refer to the ones observed in the previous model of community instability as a function of latitude.

4.1 Evaluation Metrics

Each of the three steps in the approach towards quantifying community stability in response to climate change relies on looking for a significant relationship between the response $y$ and explanatory $x$ variables defined for each case. Once these variables are defined, we propose to relate...
CHAPTER 4. RESULTS AND DISCUSSION

them by the model presented in Equation 4.1. Here, the goal is to find a significant relationship \( f(x) \) to relate \( y \) as a function of \( x \).

\[
y = f(x) + \epsilon
\]  

(4.1)

where \( \epsilon \) is the residual noise.

In order to find a significant relationship for \( f(x) \), we tried different regression models (e.g., linear, second-degree polynomial, third-degree polynomial) under the following assumptions:

1. **Normality**
   The residuals are normally distributed with zero mean and variance \( \sigma^2 \): \( \epsilon_i = \mathcal{N}(0, \sigma^2) \).

2. **Independence**
   By the normality assumption, the residuals are independent and uncorrelated: \( \text{cov} [\epsilon_i, \epsilon_j] = 0 \).

3. **Homoscedasticity**
   All the residuals are drawn from the same distribution and share the same variance: \( \sigma_i^2 = \sigma_j^2 \).

Out of the different regression models that were fitted to the data, a model selection procedure was performed to identify the model that fits the data best and does so with the fewest parameters. In particular, we used the forward selection method, which starts with the simplest model (linear) and then adds higher order versions of the explanatory variables (second and third degree polynomials) to determine whether doing so yields greater explanatory power given the added model complexity. We repeat this process until none of the remaining explanatory variables explains a sufficiently large portion of the unexplained variation in the data. Here, selection was conducted via a series of \( F \)-tests.

### 4.1.1 R-Squared Value

The coefficient of determination (R-squared) represents the proportion of variance in the response variable that is predicted by the explanatory variables. In other words, how well is the total variation in the observations explained by the model. Equation 4.2 summarizes how the R-squared \( R^2 \) value is computed: \( \hat{y} \) values represent the predicted or modeled values, \( SS_{tot} \) the sum of squares of the residuals \( \epsilon_i \), and \( SS_{tot} \) the total sum of squares.

\[
R^2 = 1 - \frac{\sum_{i=1}^{N} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{N} (y_i - \bar{y})^2}
\]  

(4.2)


CHAPTER 4. RESULTS AND DISCUSSION

4.1.2 The $F$ Test Statistic in the Context of Linear Regression

To evaluate the significance of a linear regression model of the form: $Y_i = \beta_0 + \beta_1 \cdot X_i + \epsilon_i$, the first is to find the parameter estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ that minimize the squared sum of residuals in Equation 4.3. To do so, we expand this sum, take partial derivatives with respect to each parameter, and set them to zero to find the corresponding estimates for $\hat{\beta}_0$ and $\hat{\beta}_1$ in Equations 4.4 and 4.5 respectively.

$$SSR = \sum_{i=1}^{N} \epsilon_i^2 = \sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2$$  \hspace{1cm} (4.3)

where $\hat{Y}_i = \beta_0 + \beta_1 \cdot X_i$

The estimate of $\hat{\beta}_0$ in Equation 4.4 shows that the regression line goes through the mean of both variables $\bar{X}$ and $\bar{Y}$.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \cdot \bar{X}$$  \hspace{1cm} (4.4)

And, the estimate of $\hat{\beta}_1$ in Equation 4.5 shows the relationship between the slope $\hat{\beta}_1$ and the variance and covariance of $X$ and $Y$.

$$\hat{\beta}_1 = \frac{\text{cov}[X,Y]}{\text{var}[X]} = \frac{\sum_{i=1}^{N} (Y_i - \bar{Y})(X_i - \bar{X})}{\sum_{i=1}^{N} (X_i - \bar{X})^2}$$  \hspace{1cm} (4.5)

Once the expressions for the parameter estimates are obtained, we can perform a significance test for the slope and intercept of the linear model under analysis, as detailed below:

For the slope significance test, we set the Null Hypothesis to $H_0 : \beta_1 = b$, and the Alternate Hypothesis to $H_a : \beta_1 \neq b$. The slope $\hat{\beta}_1$ is $t$-distributed with $N - 2$ degrees of freedom since the numerator is normally distributed and the denominator is Chi-squared ($\chi^2$) distributed, as included in Equation 4.6.

$$t_{\hat{\beta}_1} = \frac{\hat{\beta}_1 - b}{SE_{\hat{\beta}_1}}$$  \hspace{1cm} (4.6)

where the standard error of the slope $SE_{\hat{\beta}_1}$ is given by:

$$SE_{\hat{\beta}_1} = \sqrt{\frac{\sum_{i=1}^{N} (Y_i - \hat{Y}_i)^2}{(N - 2) \sum_{i=1}^{N} (X_i - \bar{X})^2}}$$  \hspace{1cm} (4.7)

The statistical significance of this model is determined then by conducting a one-sample $t$-test, and comparing the observed $t_{\hat{\beta}_1}$ to the critical value drawn from a $t$-distribution with $N - 2$
CHAPTER 4. RESULTS AND DISCUSSION

degrees of freedom and the specified \( \alpha \) level. Similarly, for the corresponding *intercept significance test* \[20\], the Null Hypothesis is set to \( H_0 : \beta_0 = a \), and the Alternative Hypothesis to \( H_a : \beta_0 \neq a \). Here, the intercept \( \hat{\beta}_0 \) is also \( t \)-distributed with \( N - 2 \) degrees of freedom as presented in Equation 4.8

\[
t_{\hat{\beta}_0} = \frac{\hat{\beta}_0 - a}{SE_{\hat{\beta}_0}}
\] (4.8)

where the standard error of the slope \( SE_{\hat{\beta}_0} \) equals to:

\[
SE_{\hat{\beta}_0} = SE_{\hat{\beta}_1} \sqrt{\frac{1}{N} \sum_{i=1}^{N} X_i^2}
\] (4.9)

Here, the statistical significance is determined by conducting a one-sample \( t \)-test, and comparing the observed \( t_{\hat{\beta}_0} \) to the critical value drawn from a \( t \)-distribution with \( N - 2 \) degrees of freedom at the specified \( \alpha \) level.

### 4.1.3 The \( F \) Test Statistic in the Context of ANOVA Linear Models

Generally the null hypothesis (\( H_0 \)) is formulated to test if all group means are equal to each other, and the alternative hypothesis implies that at least two groups differ in their means. By groups here we refer to the different models we want to evaluate its significance \[21\].

If the null hypothesis is not true, the total variance among group means (\( \sigma_Y^2 \)), Equation 4.10 will be equal to the variance due to sampling error (\( \sigma_Y^2 / n \)) plus the variance due to the group means coming from distinct statistical populations with different means \( \mu_i \) represented by (\( Var[\mu_i] \)).

\[
\begin{align*}
\sigma_Y^2 & = \sigma_Y^2 / n + Var[\mu_i] \\
n \cdot \sigma_Y^2 & = \sigma_Y^2 / + n \cdot Var[\mu_i]
\end{align*}
\] (4.10)

The objective of the statistical tests is to determine whether the variance among groups (\( n \cdot \sigma_Y^2 \)) is greater than expected under \( H_0 \) (i.e., \( n \cdot \sigma_Y^2 > \sigma_Y^2 \)). Also, the parameter \( n \cdot \sigma_Y^2 \) is a measure of among group variance and is estimated by the Mean Squares Group (\( MS_{\text{group}} \)). The parameter \( \sigma_Y^2 \) is the variance within groups and is estimated by the Mean Squares Error (\( MS_{\text{error}} \)).
CHAPTER 4. RESULTS AND DISCUSSION

Equations 4.11 and 4.12 detail the former expressions.

\[ MS_{group} = \frac{SS_{group}}{df_{group}} = \frac{\sum_i n_i (\bar{y}_i - \bar{y})^2}{K - 1} \]  \hspace{1cm} (4.11)

where \( \bar{y}_i \) is the mean of group \( i \), \( n_i \) is the number of observations in group \( i \), and \( \bar{y} \) is the grand mean of all observations across all groups. Also, \( K \) is the number of groups that are considered to compute the degrees of freedom \( df_{group} \).

\[ MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{\sum_i (n_i - 1)s_i^2}{N - K} \]  \hspace{1cm} (4.12)

Where, \( s_i^2 \) is the variance in group \( i \), \( n_i \) is the number of observations in group \( i \), and \( \bar{y} \) is the grand mean of all observations across all groups. \( N \) is the total number of observations that are considered to compute the degrees of freedom \( df_{error} \).

Now, the \( F \)-test statistic is related to the case in which the null hypothesis is true, then \( n \cdot \sigma_y^2 = 0 \). If this holds, the ratio of total group variance to sampling variance (\( F \)) is 1. In order to estimate the \( F \)-score, we use the corresponding sample estimates of group variance and sample variance to compute Equation 4.13.

\[ F = \frac{MS_{group}}{MS_{error}} \]  \hspace{1cm} (4.13)

The \( F \)-test is one-tailed and its null distribution is based on the degrees of freedom in the numerator (i.e., number of groups \( K \) minus 1), and the degrees of freedom in the denominator (i.e., number of observations \( N - K \)). Therefore, we can reject \( H_0 \) at \( \alpha \) level if Equation 4.14 holds.

\[ F > F_{\alpha, df_{num}=K-1, df_{den}=N-K} \]  \hspace{1cm} (4.14)

Finally, since a model is statistically significant if it rejects the null hypothesis, the smaller the p-value, the larger the significance. For our analysis, a model is considered as significant if the corresponding p-value from an \( F \)-test statistic is less than or equal to a threshold value set to 0.05.

4.2 Quantifying Climate Change

By means of climate change we propose to model temperature and some of its variants (e.g., temporal change of temperature) as a function of latitude.
CHAPTER 4. RESULTS AND DISCUSSION

Figure 4.1: Models for Climate Change that relate different Temperature variables as a function of increasing Latitude. Bottom water temperatures are expected to increase towards northern latitudes.

We tested linear regression models to describe this relationship between the response variables: surface water temperature, bottom water temperature, temporal change of surface and bottom water temperatures; and the explanatory variable: increasing latitude. Figure 4.1 presents the graphical output for the regression models performed.

4.2.1 Statistical Analysis of Climate Change

Table 4.1 summarizes the statistical results for the linear regression models performed to find the relationship of temperature (and its temporal change variants) as a function of latitude. Here, \( p \)-values in blue indicate a significant model (\(< 0.05\)), and \( p \)-values in red the opposite.

The significant relationships between latitude and both surface (\( p \)-value = 0.026) and
CHAPTER 4. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>Temperature variants (y)</th>
<th>slope (m)</th>
<th>intercept (b)</th>
<th>p-value</th>
<th>( R^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surface Temperature</td>
<td>0.005</td>
<td>6.150</td>
<td>0.722</td>
<td>0.016</td>
</tr>
<tr>
<td>Bottom Temperature</td>
<td>0.035</td>
<td>-0.197</td>
<td>0.026</td>
<td>0.479</td>
</tr>
<tr>
<td>Surface Temperature Slope</td>
<td>0.002</td>
<td>-0.112</td>
<td>0.082</td>
<td>0.331</td>
</tr>
<tr>
<td>Bottom Temperature Slope</td>
<td>-0.004</td>
<td>0.255</td>
<td>0.008</td>
<td>0.596</td>
</tr>
</tbody>
</table>

Table 4.1: Statistical results for a linear model fitting of Temperature (and its temporal change variants) as a function of increasing Latitude. Blue labels for p-values highlight a significant relationship; whereas red labels the opposite.

Bottom (p-value = 0.008) temperature demonstrate that there is a significant latitudinal gradient in temperature over the range of our data. Although both relationships were significant, the model for bottom temperature slope had a lower p-value and also explained a higher \( R^2 \) (59.6% for bottom temperature vs. 33.1% for surface temperature).

For the case bottom temperature, the linear model suggests that a change of 1 degree in latitude towards the north pole leads to an increase in bottom water temperature of 0.035 degrees Celsius. On the other hand, the linear model for the temporal change of bottom temperature suggests that for a change of 1 degree in latitude towards the north pole, the temporal change in bottom water temperature (i.e., degrees Celsius/year) decreases by 0.004.

Moreover, while bottom temperature increases with latitude, its temporal change, which was obtained by regressing temperature against time, decreases with latitude. Hence, the latitudinal differences in bottom temperature are expected to become attenuated over time due to their predicted reductions at higher latitudes and their predicted rise at lower latitudes.

4.3 Quantifying Instability at Poleward Locations

This section presents how community instability is quantified via the computation of three metrics. Once these metrics are computed, we regress them against latitude to determine which locations are more likely to experience instability in their communities. For this step, regression models were tested to investigate the relationship between the response variables: pairwise, centroid, and convex hull instability metrics; and the explanatory variable: increasing latitude.

Recall that instability metrics refer in detail to:

1. Pairwise: Mean pairwise distance between the site/community scores (low-dimensional points).
2. Centroid: Mean distance between all site/community scores and its centroid.
CHAPTER 4. RESULTS AND DISCUSSION

3. Convex Hull: Area of a convex hull containing all these site/community scores.

Figure 4.2 shows the graphical output for the regression models performed to fit the behaviour of community instability as a function of latitude. The first row in this figure presents the performance of the three proposed community instability metrics when plotted against latitude. On the right side, the standard deviation of the latitude values that belong to each of the latitude bins is presented. This plot helps to identify if there are any latitudinal trends in the variability of sample locations within each bin. Such trends could mean that variability in the location of sampling locations within a bin could potentially explain any observed latitudinal trend in community instability because samples taken from locations that differ in their coordinates are more likely to be different than those that are taken from more similar coordinates. Hence, a quasi-uniform distribution of latitudinal variability as a function of latitude is important for interpreting spatial patterns of community instability because they indicate that these trends are not artifacts of the sampling protocol. Finally, the second and third rows in Figure 4.2 present linear and polynomial fits for the community instability metrics as a function of latitude.

4.3.1 Statistical Analysis of Instability at Poleward Locations

Tables 4.3.1, 4.3.1, and 4.3.1 present the statistical results obtained for the linear and polynomial models for instability as a function of latitude. Here, $p$-values in blue indicate a significant model ($< 0.05$), and $p$-values in red the opposite.

<table>
<thead>
<tr>
<th>Instability metric</th>
<th>slope ($m$)</th>
<th>intercept ($b$)</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise</td>
<td>0.02921</td>
<td>-1.56015</td>
<td>0.5927</td>
<td>0.0092</td>
</tr>
<tr>
<td>Centroid</td>
<td>0.01994</td>
<td>-1.06383</td>
<td>0.6079</td>
<td>0.0078</td>
</tr>
<tr>
<td>Convex Hull</td>
<td>0.254</td>
<td>-13.680</td>
<td>0.5699</td>
<td>0.0116</td>
</tr>
</tbody>
</table>

Table 4.2: Statistical results for a linear regression model to fit Instability Metrics as a function of increasing Latitude. None of the p-values for the different metrics was found non-significant. This suggests the search for a higher polynomial model to best describe the relationship between these variables.

A polynomial of degree 2 suffices to describe this relationship since all p-values of all terms in the polynomial are significant. All metrics agree that instability increases with latitude, and that there is not a highly biased sample-size latitude bin that might explain this trend. Also, all metrics for the polynomial of order 2 was able to explain on average 91% of the variance of the data as reflected by the $R^2$ value. For the case of the polynomial of degree 3, the 2 first two instability
CHAPTER 4. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>Instability Metric ($y$)</th>
<th>$a_2$</th>
<th>p-value ($a_2$)</th>
<th>$a_1$</th>
<th>p-value ($a_1$)</th>
<th>intercept ($b$)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise</td>
<td>0.1489</td>
<td>4.49e-4</td>
<td>0.1953</td>
<td>8.28e-5</td>
<td>0.1436</td>
<td>0.9371</td>
</tr>
<tr>
<td>Centroid</td>
<td>0.10</td>
<td>2.3e-4</td>
<td>0.1334</td>
<td>3.66e-5</td>
<td>0.0994</td>
<td>0.9499</td>
</tr>
<tr>
<td>Convex Hull</td>
<td>1.2289</td>
<td>5.32e-3</td>
<td>1.6985</td>
<td>9.06e-4</td>
<td>1.1346</td>
<td>0.8682</td>
</tr>
</tbody>
</table>

Table 4.3: Second-degree polynomial model for fitting Instability Metrics as a function of increasing Latitude. Blue labels for p-values highlight a significant relationship; whereas red labels the opposite. This second-order model suggest a good significant fit.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Pairwise</th>
<th>Instability Metric</th>
</tr>
</thead>
<tbody>
<tr>
<td>$a_3$</td>
<td>p-value</td>
<td>0.047 0.036</td>
</tr>
<tr>
<td>$a_2$</td>
<td>p-value</td>
<td>0.149 1.45e-4</td>
</tr>
<tr>
<td>$a_1$</td>
<td>p-value</td>
<td>0.195 3.11e-5</td>
</tr>
<tr>
<td>intercept</td>
<td></td>
<td>0.144 0.099</td>
</tr>
<tr>
<td>$R^2$</td>
<td></td>
<td>0.971 0.976</td>
</tr>
</tbody>
</table>

Table 4.4: Third-degree polynomial model for fitting Instability Metrics as a function of increasing Latitude. Blue labels for p-values highlight a significant relationship; whereas red labels the opposite. The cubic term does not justify the additional complexity over a significant relationship or a higher explained variance of the data.

metrics continue to report a significant p-value; while for the last instability metric (convex hull), the p-value fails to be significant (red text in Table 4.3.1). Therefore, the polynomial of degree 2 is kept as it provides significant results for describing the relationship of instability as a function of latitude. Also, comparing the R-squared value (e.g., the % of variance of data that the model explains) for the second (Table 4.3.1) and third degree polynomial (Table 4.3.1) coefficients in the polynomial of order 3 fit, the addition of the cubic term does not have a big impact on the model predictions – a significant increment on the variance of the data that is already explained by the quadratic coefficient.

Finally, it is important to note that we observe a trend of increased instability towards northern latitudes. This can also be understood through the bottom temperature behaviour – as the increase of bottom temperature is higher towards northern latitudes, communities might want to escape from those non-ideal conditions, and this generates instability in their community structure.
4.4 Quantifying Climate Change Patterns Driving Instability

This section shifts the study to find the environmental patterns that might explain the latitudinal patterns in community instability described in Section 3.3.2. We considered the temporal change of the surface and bottom water temperatures as the environmental patterns to describe the behaviour of the instability metrics. Also, as for this third step in our approach towards the study of community instability in the data analysis block of Figure 3.1, the regression models where tested to investigate the relationship between the response variables defined by the three instability metrics; and the explanatory variables denoted by the temporal changes (slope) in surface and bottom water temperatures.

Figure 4.3 presents scatter plots of how the surface and bottom temperature variables changed over the different years in which the survey was conducted, for some of the latitude bins in which the observations were grouped. To quantify the temporal change for each of the temperature trends per bin, we fit a linear regression model of temperature as a function of time, and suggest the corresponding slope of that model as the representative of the change of environmental variables over time.

A scatter plot of instability metrics and the slope (temporal change) of temperature variables is presented in Figure 4.4. Also, in Figure 4.5, instability metrics are modeled as a function of these temporal changes in surface and bottom temperatures.

4.4.1 Statistical Analysis of Climate Change Patterns Driving Instability

Tables 4.1 and 4.1 present the statistical results obtained for the linear and polynomial models fitted for instability as a function of the temporal change in surface and bottom water temperatures respectively. As in previous sections, p-values in blue indicate a significant model (< 0.05), and p-values in red the opposite.

<table>
<thead>
<tr>
<th>Instability Metric (y)</th>
<th>slope (m)</th>
<th>intercept (b)</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise</td>
<td>5.056</td>
<td>0.125</td>
<td>0.865</td>
<td>2.74e-4</td>
</tr>
<tr>
<td>Centroid</td>
<td>3.570</td>
<td>0.087</td>
<td>0.853</td>
<td>3.78e-4</td>
</tr>
<tr>
<td>Convex Hull</td>
<td>38.629</td>
<td>0.955</td>
<td>0.897</td>
<td>1.07e-4</td>
</tr>
</tbody>
</table>

Table 4.5: Statistical results for a linear regression model fit to relate Instability Metrics as a function of Surface Temperature temporal changes.
CHAPTER 4. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>Instability Metric</th>
<th>$a_2$</th>
<th>p-value ($a_2$)</th>
<th>$a_1$</th>
<th>p-value ($a_1$)</th>
<th>intercept ($b$)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise</td>
<td>0.017</td>
<td>0.341</td>
<td>0.111</td>
<td>5.17e-4</td>
<td>0.120</td>
<td>0.886</td>
</tr>
<tr>
<td>Centroid</td>
<td>0.012</td>
<td>0.362</td>
<td>0.078</td>
<td>7.10e-4</td>
<td>0.084</td>
<td>0.873</td>
</tr>
<tr>
<td>Convex Hull</td>
<td>-0.034</td>
<td>0.779</td>
<td>0.849</td>
<td>3.43e-4</td>
<td>0.917</td>
<td>0.898</td>
</tr>
</tbody>
</table>

Table 4.6: Second-degree polynomial fitting to model Instability Metrics as a function of temporal changes in Surface Temperature Waters. Red labels for p-values indicate a not justifiable complexity introduced by the model.

<table>
<thead>
<tr>
<th>Instability Metric</th>
<th>slope ($m$)</th>
<th>intercept ($b$)</th>
<th>$R^2$</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise</td>
<td>-5.200</td>
<td>0.141</td>
<td>0.874</td>
<td>2.19e-4</td>
</tr>
<tr>
<td>Centroid</td>
<td>-3.716</td>
<td>0.098</td>
<td>0.882</td>
<td>1.71e-4</td>
</tr>
<tr>
<td>Convex Hull</td>
<td>-36.82</td>
<td>1.067</td>
<td>0.778</td>
<td>1.65e-3</td>
</tr>
</tbody>
</table>

Table 4.7: Statistical results for a linear regression fit to model Instability Metrics as a function of Bottom Water Temperature Patterns. Blue labels for p-values highlight a significant model for this relationship captured by a linear model.

All community instability metrics are significantly related to both surface and bottom temperature slopes, while also explain an average of 80% of the variance of the data. Furthermore, since the p-value for the quadratic term in the following second-order polynomial model representation (for all instability metrics) is not significant any more ($p > 0.05$), a simple linear regression is sufficient to explain the relationship between community instability and temperature change.

In the previous section, the goal was to find the relationship between instability and latitude. The quadratic relationship found suggested that instability tends to increase with latitude more quickly at northern latitudes. Now, for this section, the question shifted to find the environmental variables that can explain the role of latitude in the previous model. The slope of surface and bottom temperatures are good candidates for understanding the latitudinal patterns in community instability:

- The linear relationship between surface temperature slope and instability suggests that an increase in surface temperature leads to a constant increase in community instability. This stands in contrast with the curvilinear relationship between latitude and community instability, which suggests that an increase in latitude leads to faster increase in community instability at higher latitudes. Specifically, for the different instability metrics:

1. Mean pairwise distance between sample scores: an increase of 1 unit in the temporal change of surface water temperature causes this instability metric to increase by 5.06 units.

35
CHAPTER 4. RESULTS AND DISCUSSION

<table>
<thead>
<tr>
<th>Instability Metric</th>
<th>$a^2$</th>
<th>p-value ($a^2$)</th>
<th>$a_1$</th>
<th>p-value ($a_1$)</th>
<th>intercept ($b$)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pairwise</td>
<td>0.0323</td>
<td>0.028</td>
<td>-0.111</td>
<td>6.03e-5</td>
<td>0.120</td>
<td>0.947</td>
</tr>
<tr>
<td>Centroid</td>
<td>0.022</td>
<td>0.027</td>
<td>-0.079</td>
<td>4.66e-5</td>
<td>0.084</td>
<td>0.951</td>
</tr>
<tr>
<td>Convex Hull</td>
<td>0.211</td>
<td>0.208</td>
<td>-0.792</td>
<td>18.5e-4</td>
<td>0.917</td>
<td>0.8332</td>
</tr>
</tbody>
</table>

Table 4.8: Second-degree polynomial fit to model Instability Metrics as a function of temporal changes in Bottom Water Temperatures. Red labels for the p-value in the convex hull case indicate a not justifiable nor general complexity introduced by adding a quadratic term to the model.

2. Mean distance to the centroid: an increase of 1 unit in the temporal change of surface water temperature causes this instability metric to increase by 3.57 units.

3. Area of the fitted convex hull: an increase of 1 unit in the temporal change of surface water temperature causes this instability metric to increase by 38.63 units.

- Bottom temperature slope has the opposite effect on community instability. This suggests that increasing bottom temperature slopes (which occur at lower latitudes) decrease community instability linearly:

1. Mean pairwise distance between sample scores: an increase of 1 unit in the temporal change of bottom water temperature causes this instability metric to decrease in 5.20 units.

2. Mean distance to the centroid: an increase of 1 unit in the temporal change of bottom water temperature causes this instability metric to decrease in 3.72 units.

3. Area of the fitted convex hull: an increase of 1 unit in the temporal change of bottom water temperature causes this instability metric to decrease in 36.82 units.

4.4.2 Further Analysis on Climate Change and Community Instability

The previous observations can be summarized as follows: community instability is expected to increase towards northern latitudes. Towards northern latitudes, temporal changes in temperature (environmental) variables suggest an inverse relation – faster temporal changes in the surface water temperatures and a slower temporal changes in the bottom water temperatures towards northern locations. However, to state the former, we also need to indicate that the latitudinal changes in surface vs. bottom temperatures behave in the opposite direction: bottom temperature slope decreases with latitude whereas surface temperature slope increases with latitude. Hence, not only the rate of change differs but also does the direction. This pattern can be captured by a multiple regression model.
CHAPTER 4. RESULTS AND DISCUSSION

(Equation 4.15) of community instability metrics as a function of both temporal changes in surface and bottom water temperatures. It is important to note that the results of the multiple regression were not tainted by multicollinearity because the correlation between surface and bottom temperature were not too correlated \( \rho < 0.7 \) \[22\].

\[
y = a_1 x_1 + a_2 x_2 + b \quad \text{(4.15)}
\]

where \( y \) remains as the representation for the 3 instability metrics, \( x_1 \) represents the temporal change in surface water temperatures and \( x_2 \) the corresponding temporal change for bottom water temperatures.

\[
\rho(x_1, x_2) = \frac{1}{N-1} \sum_{i=1}^{N} \left( \frac{x_{1i} - \mu_{x1}}{\sigma_{x1}} \right) \times \left( \frac{x_{2i} - \mu_{x2}}{\sigma_{x2}} \right) \quad \text{(4.16)}
\]

Following Pearson correlation coefficient computation (Equation 4.16), the correlation between the temporal change of surface and bottom water temperatures was found to be -0.8105, which follows the requirement of \( \rho < 0.7 \) as to be able to determine net effects of both explanatory variables involved. Also, Table 4.9 presents the statistical results obtained for the multiple regression fit. Recall that \( p \)-values in blue indicate a significant model \(( < 0.05 \)) , and \( p \)-values in red the opposite.

<table>
<thead>
<tr>
<th>Parameter ( a_1 )</th>
<th>p-value ( a_1 )</th>
<th>\begin{tabular}[c]{c} Instability Metric \ \hline \end{tabular}</th>
<th>\begin{tabular}[c]{c} Pairwise \ \hline \end{tabular}</th>
<th>\begin{tabular}[c]{c} Centroid \ \hline \end{tabular}</th>
<th>\begin{tabular}[c]{c} Convex Hull \ \hline \end{tabular}</th>
</tr>
</thead>
<tbody>
<tr>
<td>( a_1 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( a_2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>( b )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Multiple ( R^2 )</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 4.9: Statistical results for multiple regression for community instability metrics as a function of a linear combination of temporal changes in surface and bottom water temperatures.

From Table 4.9 coefficients \( a_1 \) and \( a_2 \) show that for all instability metrics, the temporal changes in surface water temperatures are inversely related to their corresponding temporal changes in bottom water temperatures. For \( \text{pairwise} \) and \( \text{centroid} \) metrics, the magnitude of each of the explanatory variable is equivalent; whereas for the \( \text{convex hull} \) metric, the magnitude of the slope of surface water temperature is approximately twice as big as its corresponding magnitude value for the slope of bottom temperature.
CHAPTER 4. RESULTS AND DISCUSSION

Figure 4.2: Instability Metrics as a function of increasing Latitude. Communities show an increasing instability pattern towards poleward locations.
CHAPTER 4. RESULTS AND DISCUSSION

Figure 4.3: Scatter plot of Temperature over different years and for different Latitude bins.

Figure 4.4: Instability Metrics and temporal changes of Temperature variables. Temporal changes in Bottom Water Temperature show a direct approximation of Latitude; whereas for Surface Water temperature the opposite is observed.
CHAPTER 4. RESULTS AND DISCUSSION

Figure 4.5: Model fitting for Instability Metrics as a function of Temporal Changes in Temperature. In all cases, a linear model probed to be enough to describe a significant relationship between these variables.
Chapter 5

Conclusions

This chapter summarizes the contributions found for each of the three steps we proposed to study community instability: (i) quantification of climate change across different latitudes in Section 5.1, (ii) quantification of community instability across different latitudes in Section 5.2, and (iii) quantification of the relationship between climate change and community instability at each location in Section 5.3.

5.1 Quantifying Climate Change

We propose to quantify the effects of climate change by modeling how temperature and some of its variants (e.g., temporal change of temperature) change with latitude. We found that there is a significant relationship between latitude and bottom (p-value = 0.008) water temperature over the range of our data. This model explains a high proportion of the variance in the data (59.6%). In detail, this linear relationship of bottom temperature and latitude suggests that a change of 1 degree in latitude towards the north pole leads to an increase in bottom water temperature of 0.035 degrees Celsius.

5.2 Community Instability as a Function of Latitude

A second degree polynomial was used to document the significant relationship between each of the three instability metrics and latitude. Here, all the three proposed instability measurements agree that instability increases with latitude. Importantly, the variability (standard deviation) of the sampling locations within each latitudinal bin was quasi-uniform indicating that the significant
CHAPTER 5. CONCLUSIONS

latitudinal trends in instability were not artifacts of the sampling protocol. Also, the second degree polynomial models explain on average 91% of the variance in the instability metrics. This trend of increased instability towards northern latitudes can be also understood through the bottom temperature behaviour – as the increase of bottom temperature is higher towards northern latitudes, species might want to escape from those non-ideal conditions, and this generates instability in the structure of ecological communities.

5.3 Climate Change Patterns Driving Community Instability

All community instability metrics are significantly related to both surface and bottom temperature slopes and these explanatory variables explain an average of 80% of the variance of the data. A simple linear regression is sufficient to explain the relationship between community instability and environmental factors defined through the temporal change of temperature.

In the previous subsection, a quadratic relationship was suggested to relate community instability to latitude – increased instability towards northern latitudes. For this subsection, the objective is to find the environmental variables that approximate the role of latitude in the previous model.

The slope of surface and bottom temperatures are good candidates for understanding the latitudinal patterns in community instability. A linear relationship between surface temperature slope and instability suggests that an increase in surface temperature leads to a constant increase in community instability. This contrasts with the curvilinear relationship between latitude and community instability, which suggests that an increase in latitude leads to faster increase in community instability at higher latitudes. On the other hand, bottom temperature slope has the opposite effect on community instability, and is directly related to latitude as in the case of increased instability towards poleward locations.

5.3.1 Further analysis on climate change and community instability

The previous sections can be summarized as follows: community instability is expected to increase towards northern latitudes. Towards northern latitudes, temporal changes in temperature (environmental) variables suggest an inverse relationship – faster temporal changes in the surface water temperatures and slower temporal changes in bottom water temperatures towards northern locations.
CHAPTER 5. CONCLUSIONS

To confirm the former relation, latitudinal changes in surface vs. bottom temperatures were shown to behave in the opposite direction by fitting a multiple regression model: bottom temperature slope decreases with latitude whereas the opposite holds for surface temperature slope and latitude.

Also, the correlation factor of these two temperature variables $\rho$ was determined to be equal to $-0.8105$, hence multicollinearity was not a problem in this case.

5.4 Contributions

Through this work, we propose three metrics to capture the concept of community instability related to the notion of the ability of an ecosystem to return to its initial condition (e.g., initial abundance score, equilibrium). Specifically, we introduce instability metrics via the computation of the mean of pairwise distances in low dimensional space, the mean of distances to the centroid, and the area of a fitted convex hull that contains the low-dimensional observation points.

These metrics proposed to capture community instability add practical knowledge because they allow us to quantify climate-related ecological risk to communities across different geographical locations. For instance, this information could be used to devise effective spatial management plans by identifying regions of high climate risk where relieving local stressers such as harvesting and pollution could promote stability and ensure the persistence of vulnerable ecological communities.

Also, a three-step approach is proposed to explore community instability: We propose a model based on temperature as a function of latitude, followed by the exploration of the relationship of instability metrics as a function of latitude, and finally the search for climate (environmental) patterns that approximate the behaviour of latitude in the previous model.

Finally, we conclude that over the data compiled for this analysis around the latitudes of the East Bearing Sea, increased instability was found towards poleward locations. We observed that this instability trend is mainly driven by a linear relationship with the increments of bottom water temperatures at northern latitudes.
REFERENCES


REFERENCES


REFERENCES


APPENDIX A

Sensitivity Analysis for Different Latitude Bins
APPENDIX A. SENSITIVITY ANALYSIS FOR DIFFERENT LATITUDE BINS

Our approach towards quantifying community instability integrates a binning process after the dimensionality reduction of the data. This binning process splits the (latitude, longitude) observation points into \( N \) equally-sized bins/groups as to obtain specific measurements for a particular sub-range of latitude and differentiate changes as this sub-ranges goes from northern to southern latitudes. In the study presented in the main text, we included the results generated for \( N = 10 \), and in this appendix we provide a more detailed exploration about the changes in community instability trends as the bin size changes.

Therefore, Figure 2 shows the trends obtained for the 3 community instability metrics while varying the number of latitude bins from 5, 10, 20, and 30. While increasing the number of bins (which implies a reduction in the number of observation points per bin), the latitudinal standard deviation computed for each group starts to meaningful differ from one another. Here, we selected a bin size of 10 for our main analysis since it presents the desired quasi-uniform distribution of latitudinal variability. This allows to interpret spatial patterns of community instability without implying artifacts of the sampling protocol. Recall that the study of the standard deviation of the latitude values per bin helps to avoid considering those latitudinal trends in the variability of sample locations that potentially explain the observed latitudinal trend in community instability from samples taken from similar coordinates, which opposites the desired quasi-uniform distribution of the sampling protocol over the entire latitudinal range.
APPENDIX A. SENSITIVITY ANALYSIS FOR DIFFERENT LATITUDE BINS

Figure 1: Community Instability Metrics while varying the number of Latitude bins from 5, 10, 20, and 30. A quasi-uniform distribution of latitudinal variability (right side figures) is desired to interpret spatial patterns of community instability without sampling protocol artifacts. A bin size of 10 was selected for our analysis in the main text.
APPENDIX A. SENSITIVITY ANALYSIS FOR DIFFERENT LATITUDE BINS

Figure .2: Community Instability Metrics while varying the number of Latitude bins from 5, 10, 20, and 30 (II). A quasi-uniform distribution of latitudinal variability (right side figures) is desired to interpret spatial patterns of community instability without sampling protocol artifacts. A bin size of 10 was selected for our analysis in the main text.