UC Merced

UC Merced Electronic Theses and Dissertations

Title

DICHOTOMY OR GRADIENT? HOW MARINE LAKE STATES ARE DISTRIBUTED IN PALAU

Permalink

https://escholarship.org/uc/item/0pv1b2k5

ISBN

9798173793881

Author

Marques Klovrza, Anabelle

Publication Date

2026-08-07

Copyright Information

This work is made available under the terms of a Creative Commons Attribution-NoDerivatives License, available at https://creativecommons.org/licenses/by-nd/4.0/

Peer reviewed|Thesis/dissertation

eScholarship.org Powered by the California Digital Library
University of California

UNIVERSITY OF CALIFORNIA, MERCED

DICHOTOMY OR GRADIENT? HOW MARINE LAKE STATES ARE DISTRIBUTED IN PALAU

Thesis

submitted in satisfaction of the requirements

for the degree of

MASTER IN ECOLOGY AND ECOSYSTEM SCIENCES

in Environmental Sciences

by

Anabelle Marques Klovrza

Thesis Committee:

Professor Michael N Dawson, Chair

Professor Marilia Gaiarsa, committee member

Professor Michael Beman, committee member

Professor Sora Kim, committee member

2026

The Thesis of Anabelle Marques Klovrza is approved and is acceptable in quality and form for publication on microfilm and in digital formats:

Marilia Gaiarsa, Ph.D.

Michael Beman, Ph.D.

Sora Kim, Ph.D.

Michael N Dawson, Ph.D., Advisor & Chair

University of California, Merced 2026

© 2026 Anabelle Marques Klovrza

TABLE OF CONTENTS

TABLE OF CONTENTS ii
LIST OF FIGURES iii
LIST OF TABLES v
ACKNOWLEDGEMENTS vi
ABSTRACT viii

CHAPTER 1. DICHOTOMY OR GRADIENT? HOW MARINE LAKE STATES ARE DISTRIBUTED IN PALAU 9
Introduction 9
i. Island biogeography in ecology 9
ii. Environmental factors and physical attributes 2
iii. Marine lakes 3

Methods 6
i. Original files' management 7
ii. Physical attributes data 9
iii. Environmental factors data 10
iv. Statistical analysis 12

Results 14
i. Physical attributes analyses 14
ii. Environmental factors' analyses 15
iii. Statistical analyses 22

Discussion 27
i. Physical attributes 27
ii. Environmental factors 28
iii. Statistical analyses 30
iv. Analysis limitations 33

Additional Statements 35

CHAPTER 1: Summary and conclusions 36
i. Question and hypothesis 36
ii. Statistical analysis that supports a dichotomy pattern in Palau 36
iii. Foundational work for next analyses including biological data 37

REFERENCES 37
SUPPLEMENTARY DATA 42

LIST OF FIGURES

Figure 1. *Scheme of a stratified marine lake transect. (**1**) Ocean level (tide amplitude). (**2**) Tidal amplitude inside the stratified lake. (**3**) Rain that decreases the salinity of shallower layers of water. Some of the rain evaporates with the typical high temperatures. (**4**) Seawater column showing the mixolimnion layer, rarely mixed by the wind (**8**); the chemocline gradient, where the oxygen concentration declines to 0 ppm; and the pycnocline, where a sudden change in the vertical gradient of water density. In the chemocline gradient, there is also a pink layer that represents the presence of anoxic bacteria responsible for this color; the monimolimnion layer, under the chemocline, where there is little or no sunlight and no seawater mixing. The layer thickness can be different across lakes. (**5**) Seawater mixing inside tunnels and channels is represented by the blue cycle icon. (**6**) Colder water (denser water) comes from the ocean to the bottom of the lake. (**7**) Depending on the input of seawater, the chemocline can be mixed, expanding or shrinking the bacterial plate. (**8**) Effects of wind that are lacking in stratified marine lakes due to the structure of the limestone around them (modified from Blanchette et al., 2020).* *5*

Figure 2. *The Palau archipelago map with locations of sampled environmental factors and physical attributes data. (A) Islands of Ngeruktabel and Mecherchar. (B) Island of Koror. Stratified lakes (green), mixed lakes (blue), and ocean sites (red) have their name acronyms in bold black font (**Table 1**). Sites in triangles have only environmental factors data sampled (temperature, dissolved oxygen, pH, and salinity), while those in squares have both environmental and physical attributes data collected.* *11*

Figure 3. *Physical attributes in the marine lakes of Palau. The red color represents the ocean sites, the blue color represents the mixed lakes, and the green color represents the stratified lakes. The black arrows represent the strength of each physical attribute in grey. Vol = Total volume (m^3); VolCh = Volume without the chemocline (m^3); SurAr = Surface Area (m^2); MinDistOcean = Minimum distance to the ocean (m); MeanDistOcean = Mean distance to the ocean (m); MedianDistOcean = Median distance to the ocean (m); TidLag = Tidal lag (min); TidEff = Tidal efficiency (%); Perim = Perimeter (m); and MaxDepth = Maximum depth (m).* *15*

Figure 4. *Time-series plots of Palau marine lakes. (**a.1**) Average temperatures per year and site. (**b.1**) Average salinity per year and site. (**c.1**) Average dissolved oxygen per year and site. (**d.1**) Average pH per year and site. Each line represents one site, and the strongest line, with the gray range, represents a general trend across the years, along with a 95% confidence interval. The red represents ocean sites, the blue represents the mixed lakes, and the green represents the stratified lakes. The density plots on the right (**a.2, b.2, c.2,** and **d.2**) represent the amount of data that was used to make this time series plot.* *17*

Figure 5. *Profile plots for all the averages of environmental factors across all years, grouped by site categories. Ocean sites are in red lines; mixed lakes in blue lines; and stratified lakes in green lines. Each line corresponds to a site and its depth (meters). (**A**) Average temperatures for all years per site and depth. (**B**) Average salinity for all years per site and depth. Both were measured every meter until the bottom of the site.* *18*

Figure 6. *Profile plots for all the averages of environmental factors across all years, grouped by site categories. Ocean sites are in red lines; mixed lakes in blue lines; and stratified lakes in green lines. Each line corresponds to a site and its depth (meters). (**A**) Average dissolved oxygen levels for all years per site and depth. (**B**) Average pH for all years per site and depth. Both were measured every meter until the bottom of the site.* *19*

Figure 7. *Principal Component Analysis (PCA) plot of the average environmental factors per site and year for Palau sites between 2007 and 2019. Each year is represented by a unique shape. The mixed lakes are represented in blue, ocean sites in red, and stratified lakes in green. Each shape represents a year, and the ellipses follow the same color from the sites' categories, with 90% confidence intervals for each site.* *20*

Figure 8. *Correlation plot of environmental factors versus physical attributes for the marine lakes. The coldest colors (from 0 to 1) refer to the positive correlation between variables, while the warmest colors (from 0 to –1) refer to the negative correlation between them. The plot is divided into three subplots: the green rectangle shows correlations among environmental factors; the two purple rectangles show correlations between environmental factors and physical attributes; and the yellow rectangle shows correlations among physical attributes.* *22*

Figure 9. *Distance-based redundancy analysis (dbRDA) showing a strong environmental gradient structured by lake physical attributes, with CAP1 representing the dominant axis of variation driven primarily by distance to the ocean and maximum depth. 27*

Supplementary data: Figures

Figure S1. *Scree plot for physical attributes data. It shows the data variance explained and the cumulative variance in percentage for each principal component (PC). The blue columns represent the individual PC, and the orange line represents the cumulative explanation of the data variance. 40*

Figure S2. *Scree plot of environmental factors' PCA. It shows the data variance explained and the cumulative variance in percentage for each principal component (PC). The blue columns represent the individual PC, and the orange line represents the cumulative explanation of the data variance. 40*

Figure S3. *Within-site environmental variability across years. The black intervals in each row represent the standard deviation of the data across time. The mean pairwise distance was calculated with a distance matrix that grouped the name of each site and their respective categories. 41*

Figure S4. *Variance of each environmental factor data. Dissolved oxygen is represented by the navy blue color, the pH is represented by lime green, the salinity is represented by yellow, and the temperature is represented by the red color. 41*

Figure S5. *Profiles of average environmental factors (temperature, salinity, dissolved oxygen, and pH) for all stratified marine lakes in Palau. The superior axis is fixed for all profiles. Temperature is represented by the dark red lines, salinity is represented by the yellow lines, dissolved oxygen is represented by the navy blue lines, and pH is represented by the green lines. 42*

Figure S6. *Profiles of average environmental factors (temperature, salinity, dissolved oxygen, and pH) for all mixed marine lakes in Palau. The superior axis is fixed for all profiles. Temperature is represented by the dark red lines, salinity is represented by the yellow lines, dissolved oxygen is represented by the navy blue lines, and pH is represented by the green lines. 43*

LIST OF TABLES

Table 1. *Standardized data from each site and data from each year. The numbers six, seven, and eight relate to the modal month chosen to be analyzed; the subtext numbers represent the number of visits each site had by month, followed by the number of profiles that each site had. If there are two profiles in one site, this means that the site has two basins; if there are four profiles, it means that there are two basins and two different meters that were used in that site in that month. 9*

Table 2. *Levene's test for homogeneity of variance analysis for the four environmental factors (temperature, dissolved oxygen, salinity, and pH of the water). Df = degrees of freedom; Fvalue = test to compare the variance between groups with the variance within groups, and Pr(>F) = p-value (probability of observing an F value if the null hypothesis, where the variances are equal, were true). 23*

Table 3. *Outputs for the pairwise PERMANOVA. Df refers to degrees of freedom, SumsOfSqs refers to the sum of squares, F.Model refers to the model used to produce the output, R^2 refers to the strength of the effects between pairs, and p-value and p-adjusted refer to the significance of the model. 23*

Table 4. *(**A**) Permutation test for homogeneity of multivariate dispersions (PERMDISP). Df = degrees of freedom, SumSq = sum of squares, MeanSq = mean of squares, F = F-test, N.Perm = number of permutations, and Pr(>F) = p-value. (**B**) Tukey multiple comparison of means test. 95% family-wise confidence level. Fit: aov(formula = distances ~ group, data = df) 24*

Table 5. *Similarity Percentage (SIMPER) analysis between mixed lakes and ocean sites. The variables below were significant to differentiate mixed lakes (ava) and ocean sites (avb). The ava and avb are related to the mean of each*

group, not to individual lakes. 24

Table 6. *Similarity Percentage (SIMPER) analysis between mixed and stratified lakes. The variables below were significant to differentiate mixed lakes (ava) and ocean sites (avb). The ava and avb are related to the mean of each group, not to individual lakes. 25*

Table 7. *Similarity Percentage (SIMPER) analysis between ocean sites and stratified lakes. The variables below were significant to differentiate ocean sites (ava) and stratified lakes (avb). The ava and avb are related to the mean of each group, not to individual lakes. 25*

Supplementary data: Tables

Table S1. *List of Palau lake name acronyms, full name of the lake, island, stratification, description citations, whether both environmental and physical data are available, and coordinates (latitude and longitude). 44*

Table S2. *Eigenvalues, variance explanation percentage, and cumulative explanation percentage for all principal components analyzed in the physical attributes' PCA. 44*

Table S3. *Physical attribute contribution percentages on each dimension (PC) on the PCA. 45*

Table S4. *SIMPER summary between groups: mixed lakes vs ocean sites; mixed lakes vs stratified lakes; and ocean sites vs stratified lakes. 45*

ACKNOWLEDGEMENTS

Working in research, increasing the knowledge of marine environments, was one of my dreams. I appreciate the opportunity given to me by the University of California, Merced, and Professor Michael Dawson. This work would not have been completed without the help of the collaborators from the Coral Reef Research Foundation, who have been conducting visits to the Palau sites since 1992. All of the data measured per site in Palau were due to their support, sample management, data organization, and good communication. I thank my committee members, Professors Marilia P. Gaiarsa, Michael Beman, and Sora Kim, for all the meetings and support given to me in these last two years of hard work. The writing process was way easier because Professor Marilia lent me resources with tips to write scientifically and manage my time to be able to write the thesis.

This thesis would not have been finished if they had not helped me to be calmer, stronger, and way more organized than before. I also thank my laboratory mates, Carmen Pedraza Marron, Hashel Orquiz, and Kei Fung Chan, for always taking some time to help me with my defense presentation, concepts that were not clear in my mind, and to be able to cheer me up, sometimes being way more positive than me. You helped me to continue in this journey. Additionally, I thank Sam Fellows, Bailey Carlson, Nattanon Wutthituntisil, and Lisa Paggeot for the great welcome when I first got to Merced. I was concerned about whether they would like me or not, but since the first day, I knew that they would not just be colleagues, but the academic family that I needed, always supporting and helping me no matter what.

A huge thank you to my Brazilian friends, whom I met here by surprise. I would never have imagined that there were so many Brazilians in such a small town. A special thanks to Marcelo, his daughters, Stella and Tarsila, Raissa and her husband, John, Iolanda, Mariana, and Mirtes for the support, good laughs, and turning my bad days into good ones with great conversations about any subject. I consider you my third Brazilian family. My second Brazilian family is now spread across different countries. I thank you from the bottom of my heart, Rafael Furlan, for being the light in my darkest days, always making me

smile even when I was crying. Thanks for being my support during the hardest time in my life. Also, I thank you, Juliana W. W. Mendes, for being the best friend that I could even imagine. When you followed your dream of getting your Ph.D. in Vienna, I thought our friendship would be frozen by the distance. However, we have such a strong connection that I feel that we could be several years apart, but when we meet again, our friendship will not have been changed at all. Thanks for always getting me and cheering me up, even when you were not emotionally better than me.

I appreciate the support that I have always had from my blood family. My sestra, Caroline Klovrza, I thank you for letting me be your history professor when you needed me. I would never have imagined that I would have a small sister who is understanding of me when I was vulnerable, opening your arms as a warm sister hug. Thanks to my mother, Simone Klovrza, and to my father, Paulo Eduardo Gomes Klovrza, for being my support and believing in me. Thanks for accepting me as I am and trusting me to chase my dreams in another country. I know how hard it is to see your prole fly away, but I want you to know that I had the best professors to help me learn how to fly. Lastly, I thank you, Edna Caratin Bronzeli, for being the best and craziest grandma that I could ever imagine. I miss you.

Last, but definitely not least, I thank you, Leticia Marques Klovrza, for being the best wife in the world. You believed me when I could not see any positive traits in myself. You showed me that the world can be a beautiful place to live. Today, I can say that because of you, I learned how to be a better person, how to look after other people, and especially how to look at myself with respect. You are and always will be the love of my life, and I thank you from the bottom of my heart for being able to share this life with you and our babies (Maui, Mia, Monet, and Olga, our cats). I love you to the moon and back.

ABSTRACT

Dichotomy or gradient? How marine lake states are distributed in Palau

by

Anabelle Marques Klovrza

Master of Ecology and Ecosystem Sciences in Environmental Systems

University of California, Merced, 2026

Professor Michael N Dawson, Chair

Keywords: marine lakes, environmental factors, physical attributes, mixed lakes, stratified lakes, Palau

Marine lakes are bodies of seawater surrounded by land. Although understudied, they serve as natural laboratories for investigating ecological and evolutionary processes. Formed around 12,000 years ago, they are younger than terrestrial islands and are often described as island-like environments or marine islands. I analyzed environmental factors and physical attributes of 25 Palau sites between 2007 and 2019, classified as stratified and mixed lakes, and ocean sites. The main goal was to determine whether these ecosystems represent a dichotomy or a continuous gradient of lake states. I conducted Principal Component Analyses (PCAs), correlation analyses, and multivariate analyses (pairwise-PERMANOVA, PERMDISP, SIMPER, and dbRDA) to evaluate the influence of environmental and physical variables across lake states. The PCA of physical attributes showed an overlap between TLN (stratified lake) and OLO (mixed lake), indicating similar physical attributes despite belonging to different states. Time-series analyses revealed increasing trends in temperature, salinity, and dissolved oxygen, while pH declined across all sites, consistent with ocean acidification events. The environmental PCA showed partial overlap between mixed lakes and ocean sites. Stratified lakes exhibited greater temperature variability and broader multivariate dispersion within the state. Pairwise-PERMANOVA showed significant differences among all categories ($p = 0.003$), although explanatory strengths varied: mixed and ocean sites differed weakly ($R^2 = 0.124$), whereas stratified lakes were much more distinct ($R^2 = 0.46–0.57$). SIMPER identified lake volume, surface area, depth, isolation level, tidal efficiency, salinity, dissolved oxygen, and pH as primary drivers of state differentiation. Overall, stratified lakes represent the most distinct lake state, while mixed lakes remain more similar to ocean environments. Environmental variation is primarily structured by dissolved oxygen, salinity, pH, depth, and isolation, demonstrating that physical characteristics shape environmental gradients across marine lakes. These findings provide a foundation for understanding island-like ecosystems and future research on small-island effects in marine lakes.

CHAPTER 1. DICHOTOMY OR GRADIENT? HOW MARINE LAKE STATES ARE DISTRIBUTED IN PALAU

Introduction

i. Island biogeography in ecology

Island biogeography has been studied by many scientists (Wallace, 1880; Darwin, 1953; Simberloff and Wilson, 1969; MacArthur and Wilson, 2001; Whittaker et al., 2018). The curiosity about *how species populate islands* has led to models of extinction and immigration rates based on island size and distance from the mainland (i.e., the Equilibrium Theory of Island Biogeography, see MacArthur and Wilson, 1967) and from each other (Whittaker et al., 2018). These models have enhanced our understanding of how changes in distance from the mainland and island size may lead to different island-population dynamics. However, research on islands in the marine realm is still limited (Dawson and Hamner, 2008; Dawson, 2016; Pinheiro et al., 2017; Hirschfeld et al., 2023), as are comparisons between marine and terrestrial ecosystems (Steele, 1985; Dawson and Hamner, 2008; Faria et al., 2021).

The Equilibrium Theory of Island Biogeography (MacArthur and Wilson, 1967) suggests that relationships among species diversity, island size, and distance are related to immigration and extinction rates on the island. It comes from an equilibrium number of species, which is controlled by a balance between immigration and extinction rates (MacArthur and Wilson, 1967; Whitehead and Jones, 1969). McArthur & Wilson's (1967) theory can also be applied to other island-like environments, such as mountains, coastal marine environments, and marine lakes, the main subject of this work. To understand the Island Biogeography Theory in the context of other island-like environments, it is essential to have a better understanding and quantitative descriptions of the physical attributes and abiotic factors of these ecosystems that shape environmental gradients.

ii. Environmental factors and physical attributes

Environmental factors and physical attributes of an ecosystem influence processes surrounding community assembly (i.e., natural selection, migration, and ecological or genetic drifts), species diversity in communities, and genetic diversity in different populations. They act independently or together as main filter variables that select and exclude species from the ecosystem, resulting in smaller (or larger) populations and communities. Both can be analyzed as filter sets that act on different spatiotemporal scales to define the structure and composition of different species across communities (Keddy, 1992; Montanyès et al., 2023).

They can also influence the dispersal capabilities of species, which will determine different sets of species reaching different regions (Connor and Simberloff, 1979; Keddy, 1992; Bruno et al., 2003; Montanyès et al., 2023). In this scenario, the species that are dispersing face different environments until they reach another location. In an environment such as the ocean, the environmental factors are seen as gradients that are changing continuously between the place in which the set of species lives and the place where some species (propagules) arrive. After reaching the new ecosystem, this set of species is affected by another set of filters: biotic interactions, which may change the species pool through ecological interactions (i.e., competition, predation, and mutualism), filtering or facilitating a new subset of species in that locale. The abiotic environment, here considered as physical attributes and environmental factors mentioned previously, continues to filter species poorly adapted to these new conditions (Cody and Diamond, 1975; Bruno et al., 2003; Montanyès et al., 2023).

The analyses of environmental factors and physical attributes are essential to expand the knowledge of marine islands and their inhabitants, since both variable sets influence the way that communities assemble across time. Also, to assess different ecosystem processes and patterns, adding them to the Island biogeography area of study.

iii. Marine lakes

Marine lakes are described as pieces of seawater surrounded by land (Dawson and Hamner, 2005; Dawson, 2006). They are understudied island-like systems (Dawson, 2016; Blanchette *et al.*, 2020), which means that impacts of environmental factor variations and physical attributes on marine organisms also remain poorly understood (Dawson and Hamner, 2008; Dawson, 2016; Weigelt *et al.*, 2016; Hirschfeld *et al.*, 2023). Of over 200 global marine lakes, about 35 (Hutchinson, 1975; Walker and Likens, 1975; Hamner and Hamner, 1998; Dawson and Hamner, 2008; Colin, 2009; Dawson, 2016; Blanchette et al., 2020) have been partially described, focusing on their geomorphology, limnology, population genetics, and species diversity (Hutchinson, 1975; Walker and Likens, 1975; Hamner and Hamner, 1998; Cerrano et al., 2006; Dawson and Hamner, 2008; Becking et al., 2011; Saitoh et al., 2011; Strelkov et al., 2014; Dawson, 2016; Meyerhof et al., 2016; Wilson et al., 2017; Patris et al., 2019; Rapacciuolo et al., 2019; Blanchette et al., 2020; Krasnova, 2021; Aji et al., 2023).

The limnology of these ecosystems encompasses different concepts depending on the depth of the marine lakes. The *meromixis* concept was created by Hutchinson (1937), who described it as a water layer that remains partly or fully unmixed compared to the rest of the water column (Hutchinson, 1937 & 1975; Zadereev et al., 2017). In marine lakes, a chemically different part of water in the environment may remain partly or fully unmixed, even during circulation periods. This structure can describe freshwater and marine lakes that are considered meromictic (Hutchinson, 1937; Zadereev et al., 2017). The term *monimolimnion*, which refers to the deep body of water of a meromictic lake, was defined by Findenegg (1937). Hutchinson (1975) introduced the term *mixolimnion*, a term that refers to the body of water above the *monimolimnion*, which can have free circulation periodically (Findenegg, 1937; Hutchinson, 1975; Zadereev et al., 2017). The layer located between the *mixolimnion* and the *monimolimnion* is called the *chemocline* (Hutchinson, 1975; Zadereev et al., 2017).

Hamner and Hamner (1998) discuss the different marine lake state terms that were created by Hutchinson (1937): marine lakes were classified according to how the stratification was created and

developed over time. While *meromictic* lakes have a specific and structured stratified profile, *holomictic* lakes have a mixture throughout the water column (Hall and Northcote, 2012; Zadereev et al., 2017). The differences between the two states raise different questions about why the marine lakes in Palau are so different, whether other lakes from other regions are similar to the ones in Palau, if the Palau marine lakes are unique, and what the differences are between the two lake states (*meromictic* and *holomictic*) on whether there is a dichotomy or a gradient between these states (Hamner and Hamner, 1998).

The marine lakes in Palau have a complex water column to be categorized as simply *meromictic* or *holomictic* lakes (Imboden and Wüest, 1995; Hamner and Hamner, 1998). One marine lake that is an example of this categorization problem is the Hot Water Lake (HLM), located on Mecherchar Island. Its water column is stratified, with a rapid decrease in dissolved oxygen to 0 mg/L. However, there is a tunnel that is located in the middle of the water column of the lake. It changes internal environmental variables through tidal jets (e.g., salinity and level of dissolved oxygen). In deeper layers towards the bottom of the lake, the stratification returns to an anoxic environment. This way, this lake cannot be described as a *meromictic* lake (Hamner and Hamner, 1998), but it is still stratified.

In the small archipelago of Palau, marine lakes vary in age, isolation level from the ocean, depth, surface area, and environmental factors (Hamner and Hamner, 1998; Colin, 2009), with at least 12 stratified lakes documented by Hamner and Hamner (1998). The ocean is seen as analogous to the “mainland”, having the total pool of species that could have migrated to the marine lakes, seen as “marine islands”, through coves and tidal jets in channels or tunnels. In their work, ten out of twelve fit in with the Hutchinson (1975) concept of *meromixis*. However, two of the twelve lakes (Hot Water and Gologuguel lakes) are mixed in midwater by tidal jets, resulting in complex stratified profiles as described previously, but not meromictic (Hamner and Hamner, 1998). The stratified lakes described by Hamner and Hamner (1998) have limited shallow water mixed by the wind (mixolimnion, introduced by Hutchinson (1975)), and their chemocline (Hutchinson, 1975; Zadereev et al., 2017) varies in thickness, dissolved oxygen levels, and free-floating red bacterial plates (Figure 1).

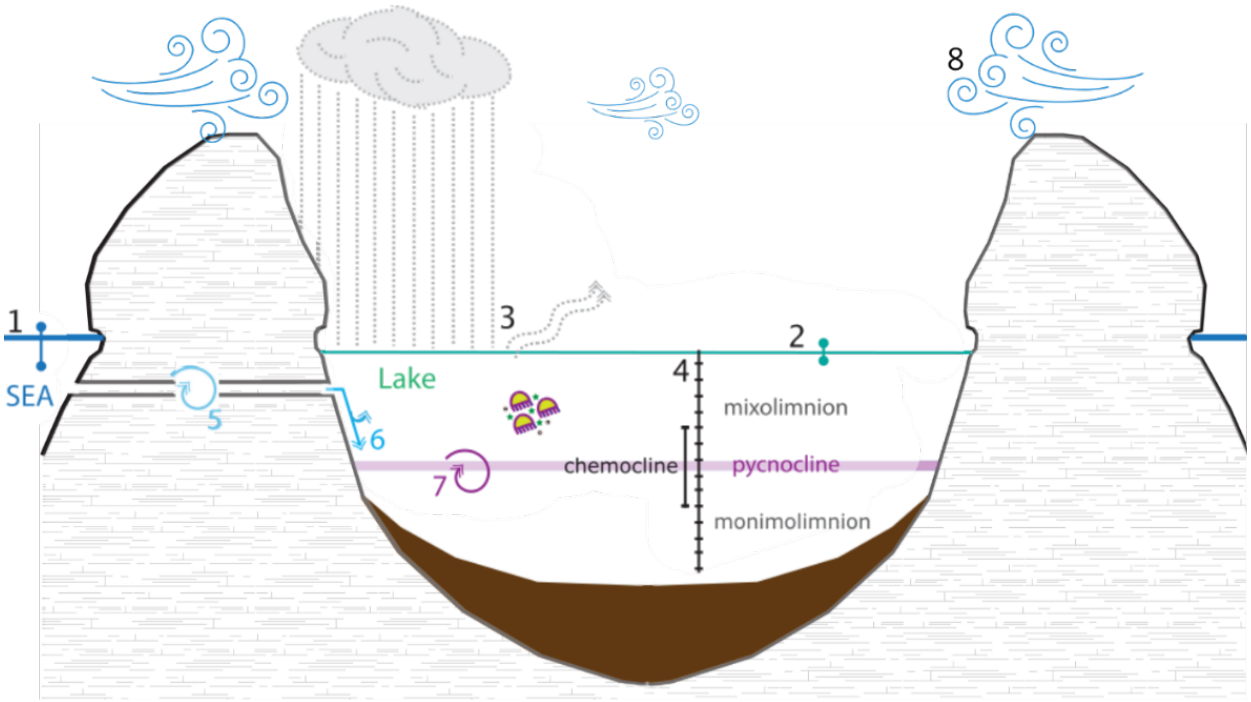


Figure 1. Scheme of a stratified marine lake transect. (**1**) Ocean level (tide amplitude). (**2**) Tidal amplitude inside the stratified lake. (**3**) Rain that decreases the salinity of shallower layers of water. Some of the rain evaporates with the typical high temperatures. (**4**) Seawater column showing the mixolimnion layer, rarely mixed by the wind (**8**); the chemocline gradient, where the oxygen concentration declines to 0 ppm; and the pycnocline, where a sudden change in the vertical gradient of water density. In the chemocline gradient, there is also a pink layer that represents the presence of anoxic bacteria responsible for this color; the monimolimnion layer, under the chemocline, where there is little or no sunlight and no seawater mixing. The layer thickness can be different across lakes. (**5**) Seawater mixing inside tunnels and channels is represented by the blue cycle icon. (**6**) Colder water (denser water) comes from the ocean to the bottom of the lake. (**7**) Depending on the input of seawater, the chemocline can be mixed, expanding or shrinking the bacterial plate. (**8**) Effects of wind that are lacking in stratified marine lakes due to the structure of the limestone around them (modified from Blanchette et al., 2020).

For this work, according to Hamner and Hamner (1998), discussions about the states of marine lakes, I used the terms “mixed” and “stratified” lakes to differentiate the different states of marine lakes. Hamner and Hamner’s descriptions significantly increased the total number of this state of lakes worldwide, increasing the use of marine lakes and their limnology which provide diverse natural laboratory ecosystems for studying how marine communities assemble and which biogeochemical processes influence these assemblages (Burnett et al., 1989; Hamner and Hamner, 1998; Dawson, 2016; Meyerhof et al., 2016; Godsoe et al., 2017; Rapacciuolo et al., 2019; Hirschfeld et al., 2023).

Marine lakes are useful ecosystems to study different ecoevolutionary processes. They are enclosed environments in which we can observe *in situ* population dynamics (Dawson et al., 2015), with low

migration rates when compared to coastal habitats (Blanchette et al., 2020). Since marine lakes are influenced by the ocean and by the runoff from their local watershed, their physical attributes depend on water density due to their salinity stratification (Strelkov et al., 2013). Additionally, marine lakes in Palau allow studying short-term environmental variations, especially in the tropics (Hamner and Hamner, 1998).

The knowledge of marine lakes' limnology (Hamner & Hamner, 1998) and monitoring of environmental factors and physical attributes of the lakes raise the question *"How do environmental factor variations describe the dichotomy (or gradient) between categories of stratified and mixed marine lakes in Palau?"* The main goal of this thesis is to generate new inputs to understand marine lake states in the context of island biogeography, while the specific goal is to analyze whether marine lake states in Palau follow a pattern of a dichotomy or a gradient related to their characteristics. I hypothesize that variations in environmental factors and physical attributes explain the variety of marine lake states (mixed or stratified) across time in Palau. I also hypothesize that environmental and physical variations lead to whether these states form a dichotomy (**H_1**) in which there are no other states between the two, or a continuous gradient (**H_2**) of other marine lakes that are transitioning between stratified and mixed states.

Methods

I used both physical attributes and environmental data gathered by Dr. Dawson, Dr. Hamner, and collaborators from the Coral Reef Research Foundation (CRRF) since December 1998 (Marine Lake Monitoring Project (MLMP); see Martin et al., 2001). Drs. T. Stieglitz & H. Stibor provided the lakes' bathymetry, tidal propagation, and distance from the ocean information. Dr. Dawson and Dr. Carlson calculated the volume, surface area, and perimeter of the sites using Google Earth tools.

i. Original files' management

i.a. Physical attributes

Physical attributes were available for 20 Palau sites. The variables and sites' names are described in the *Methods – iii* section. There were no changes in the physical attributes through time, so the available Excel

sheet was not related to the variation across years as the environmental factors are. Also, the physical attributes data includes information on the respective island names where each site is located.

For some sites, there was incomplete data for some of the variables analyzed. To be sure that the data were complete, I deleted NA-containing rows on the spreadsheet using the same package as for the environmental data set management. I merged the physical and environmental data sets using the *left_join* function from the *dplyr v.1.2.1* package (Wickham et al., 2026) for all comparison analyses between the two. Then, I ran the SIMPER analysis and built a correlation plot to assess the most representative physical attributes across all sites.

i.b. Environmental factors

For each visit to each site, the team sampled two environmental profiles (one per basin) for temperature, salinity, dissolved oxygen, and pH per meter until it reached the bottom of the site. The information about which site, date, time, and weather was saved as individual Excel spreadsheets with visits between 1999 and 2024. I opened each available spreadsheet and copied the environmental information by depth, date (day, month, and year), and lake basin (North, South, East, West, Center, Northwest, and Southeast) into a new data spreadsheet. Then, per site, I added its acronym (three letters representing the name of the site, if it is a lake or cove, and which island the site is on), state (mixed, stratified, or ocean), and environmental meter (Quanta, Manta, or DS5) used to get the profiles. There were visits in which some factors were not sampled, or the meter used was not calibrated.

For missing data, I considered the spreadsheet cells as NAs (not available values). Then, I imported the spreadsheet file (in .csv or .txt format) to RStudio *v.2026.04.0+526* (Posit team, 2026). To clean the NA values, I used the *na.omit* function from the *stats* package *v.4.4.1* (R Core team, 2026). I sorted the data based on the site acronym, then the date (from the oldest to the newest) to be able to organize the data for a clearer visualization using the *dplyr v.1.2.1* package (Wickham et al., 2026).

Because the number of visits per site was different across all sites studied, some sites had far more data than others. To be able to compare the variance within and between sites using the number of visits per

site and date, I first calculated the year range that had the highest number of visits per site for the time series analysis (between 2007 and 2019). Secondly, I calculated the modal month within the time range selected with the highest number of sites available. The modal months that had the highest number of sites within the time range were June, July, and August.

With the modal month chosen for all sites across years, I created the final dataset I used for the analyses. I organized the data for each site, its modal month, the number of available profiles per site (number of basins' profiles), and the number of visits per site (i.e., even if a site has two or more basins – profiles – both basins are considered just one visit, because each profile was collected on the same date) to have a similar number of observations (sample size) across sites (**Table 1**).

Table 1. *Standardized data from each site and data from each year. The numbers six, seven, and eight relate to the modal month chosen to be analyzed; the subtext numbers represent the number of visits each site had by month, followed by the number of profiles that each site had. If there are two profiles in one site, this means that the site has two basins; if there are four profiles, it means that there are two basins and two different meters that were used in that site in that month.*

Lake	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	Mode /site
BCM					$8_{1,2}$		$8_{1,2}$	$6_{1,2}$			$7_{1,2}$	$7_{1,2}$		N/D*
CLM	$6_{1,2}$	$6_{1,2}$	$7_{1,2}$	$6_{1,2}$	$6_{1,2}$	$6_{1,4}$	$8_{1,2}$	$6_{1,2}$			$6_{1,2}$	$7_{1,2}$		6
FLK	$6_{1,2}$	$7_{1,2}$	$7_{1,4}$	$7_{1,2}$	$7_{1,2}$	$6_{1,2}$	$8_{1,2}$	$6_{1,4}$	$8_{1,2}$		$7_{1,2}$	$6_{1,2}$		7
GLK	$6_{1,2}$	$6_{1,2}$	$6_{1,4}$	$6_{1,4}$	$6_{1,2}$	$6_{1,2}$	$6_{1,2}$	$6_{1,4}$	$6_{1,2}$	$6_{1,2}$	$6_{1,2}$	$6_{1,2}$	$6_{1,2}$	6
HC1	$6_{1,1}$	$6_{1,1}$	$6_{1,2}$	$6_{1,1}$	$6_{1,1}$	$6_{1,1}$	$6_{1,2}$	$6_{1,2}$	$8_{1,1}$		$6_{1,1}$			6
HC2	$8_{1,1}$	$8_{1,1}$	$6_{1,3}$	$8_{1,2}$	$8_{1,1}$	$8_{1,2}$	$8_{1,2}$	$8_{1,1}$	$8_{1,1}$					8
HC3	$7_{1,1}$	$7_{1,1}$	$7_{1,2}$	$7_{1,2}$	$6_{1,1}$	$7_{1,1}$	$7_{1,2}$	$7_{1,1}$	$8_{1,1}$		$6_{1,1}$			7
HLM					$8_{1,2}$	$8_{1,2}$	$8_{1,1}$	$6_{1,2}$			$7_{1,2}$	$6_{1,2}$		8
HLO	$7_{1,2}$	$7_{1,2}$	$7_{1,4}$	$7_{1,4}$	$6_{1,4}$	$7_{1,4}$	$7_{1,4}$	$7_{1,2}$	$8_{1,2}$		$6_{1,2}$	$6_{1,4}$		7
LCM					$8_{1,2}$			$6_{1,2}$			$7_{1,2}$	$8_{1,2}$		8
MLN	$6_{1,2}$	$7_{1,2}$	$7_{1,4}$	$7_{1,2}$	$6_{1,2}$	$6_{1,2}$	$8_{1,4}$	$6_{1,2}$			$7_{1,2}$	$7_{1,2}$		7
NCK	$7_{1,1}$	$7_{1,1}$	$7_{1,2}$	$7_{1,1}$	$6_{1,1}$	$7_{1,1}$	$7_{1,2}$	$6_{1,1}$	$6_{1,1}$	$7_{1,1}$	$7_{1,1}$	$6_{1,1}$	$7_{1,1}$	7
NCN	$7_{1,1}$	$7_{1,1}$	$7_{1,2}$	$7_{1,1}$	$6_{1,1}$	$7_{1,2}$	$7_{1,2}$	$7_{1,1}$	$7_{1,1}$	$7_{1,1}$	$7_{1,1}$	$6_{1,1}$	$7_{1,1}$	7
NLK	$6_{1,2}$	$6_{1,2}$	$6_{1,4}$	$6_{1,2}$	$6_{1,2}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	7
NLN	$7_{1,2}$	$7_{1,2}$	$8_{1,4}$	$7_{1,2}$	$6_{1,2}$	$6_{1,2}$	$8_{1,4}$	$7_{1,2}$			$7_{1,2}$	$7_{1,2}$		7
NLU	$7_{1,2}$	$7_{1,2}$	$8_{1,4}$	$7_{1,4}$	$7_{1,2}$	$7_{1,2}$	$8_{1,4}$				$7_{1,2}$	$6_{1,2}$		7
OLO	$7_{1,2}$	$7_{1,2}$	$8_{1,2}$	$7_{1,2}$	$7_{1,2}$	$6_{1,2}$	$8_{1,4}$	$6_{1,2}$	$8_{1,2}$		$7_{1,2}$	$6_{1,2}$		7
OTM	$7_{1,2}$	$7_{1,2}$	$7_{1,4}$	$7_{1,4}$	$7_{1,2}$	$7_{1,4}$	$7_{1,4}$	$7_{1,2}$	$7_{1,2}$	$7_{2,4}$	$7_{1,2}$	$7_{1,2}$	$7_{1,2}$	7
RCA	$7_{1,1}$	$7_{1,1}$	$7_{1,2}$	$7_{1,1}$	$6_{1,1}$	$7_{1,2}$	$7_{1,2}$	$7_{1,1}$	$7_{1,1}$	$7_{1,1}$	$7_{1,1}$	$6_{1,1}$	$7_{1,1}$	7
SLM	$6_{1,2}$	$7_{1,2}$	$7_{1,4}$	$8_{1,2}$	$6_{1,2}$	$6_{1,2}$		$6_{1,2}$			$6_{1,2}$	$7_{1,2}$		6
SLN	$6_{1,2}$	$7_{1,2}$	$8_{1,4}$	$7_{1,2}$	$7_{1,2}$	$6_{1,4}$	$8_{1,4}$	$6_{1,4}$			$8_{1,2}$	$7_{1,2}$		7
TLN	$7_{1,2}$	$6_{1,2}$	$7_{1,4}$	$6_{1,2}$	$6_{1,2}$	$6_{1,2}$	$8_{1,4}$	$6_{1,2}$	$8_{1,2}$		$7_{1,2}$	$8_{1,2}$		6
ULN	$6_{1,2}$	$7_{1,2}$	$7_{1,4}$	$7_{1,4}$	$7_{1,2}$	$7_{1,2}$	$8_{1,4}$	$8_{1,2}$			$7_{1,2}$	$7_{1,2}$		7
Mode /year	N/D*	7	7	7	6	6	8	6	8	7	7	7	7	**July (7)****

N/D* → There is no modal month because two or more months have the same amount of data.
July (7)** → July was the modal month for all sites and years.

ii. Physical attributes data

I analyzed each site volumes (m^3, total and over the chemocline); surface areas (m^2); minimum, mean, and median distance between the sites and the ocean (m); tidal efficiency (non-dimensional value), tidal lag time (minutes); and the maximum depth (m) of 20 sites located in Palau (eleven stratified lakes, seven mixed lakes, and two ocean sites (NCN and RCA; see Table S1). To assess the relationship and possible correlations between different physical attributes, I used all available data on physical attributes as variables for the Principal Component Analysis (PCA).

The sites that I used to generate the PCA were: Big Crocodile Lake (BCM), Clear Lake (CLM), Flatworm Lake (FLK), Goby Lake (GLK), Hot Water Lake (HLM), Heliofungia Lake (HLO), Little Crocodile Lake (LCM), L-shaped Lake (LSM), Lake Mekaeald (MLN), Ngerchaol Cove (NCN), Big Jellyfish Lake (NLK), Lake Ngerultabel (NLN), Ngel Lake (NLU), Lake Ongael (OLO), Tourist Lake (OTM), Risong Cove (RCA), Spooky Lake (SLM), Shrimp Lake (SLN), T-Lake (TLN), and Lake Ngchas (ULN). Then, to improve the efficiency and accuracy of other statistical methods, I used the Redundancy statistical analysis to filter out unnecessary and repetitive variables.

iii. Environmental factors data

I analyzed available environmental data (e.g., temperature (ºC), salinity (ppt), dissolved oxygen (LDO; mg/L), and pH (units) from the original data. It also had other variables, such as the specific conductance (SpCond; mS/cm), photosynthetically active radiation (PAR; ɲE/s/mɻ), chlorophyll a concentration (CHL; ɲg/L), CHLV (volts), IBVSvr4a (volts), and IBatt (%left). Although the presence of all environmental factors described above, I selected only the first four variables because they were the only ones that were available for the longest number of years within the time range between 2007 and 2019.

With the environmental data organized by site, year, and month, I calculated the averages for each environmental factor for all 25 sites, using the *aggregate* function also from the *stats* package *v.4.4.1* (R Core team, 2026). With the average values for the four interest environmental factors (i.e., temperature,

salinity, oxygen, and pH), I made all sites' profiles with lines for each site at each depth across the years. Then, for the time series analysis, I merged both basins to get the average values per environmental factor, site, and date. After gathering all environmental factors' data, I used RStudio *v.2026.04.0+526* (Posit team, 2026) to clean any NA values, and I made time-series plots to analyze the variation of each factor across time and to understand whether each environmental variable contributes to the differences between lakes' categories.

The samples were collected for each site with 1.0-meter intervals, from 0 to 39 meters, or to the bottom of each site, whichever was reached first. For each marine lake (excluding ocean sites), measurements from at least two different basins were taken using a Hydrolab Quanta, Manta, and/or DS5 meters (e.g., Martin *et al.*, 2001). I analyzed environmental data from 25 sites (seven mixed lakes, twelve stratified lakes, and six ocean sites) between 2007 and 2019 (see Figure 2 and **Table 1**).

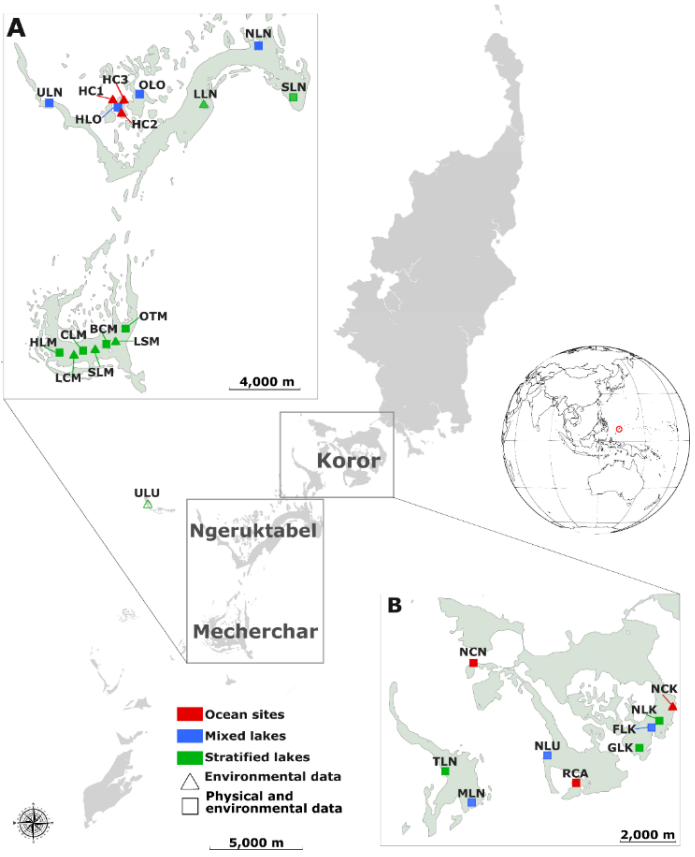


Figure 2. *The Palau archipelago map with locations of sampled environmental factors and physical attributes data. (A) Islands of Ngeruktabel and Mecherchar. (B) Island of Koror. Stratified lakes (green), mixed lakes (blue), and ocean sites (red) have their name acronyms in bold black font. Sites in triangles have only environmental factors data*

sampled (temperature, dissolved oxygen, pH, and salinity), while those in squares have both environmental and physical attributes data collected.

iv. Statistical analysis

To test hypotheses, which talk about the variation of environmental factors being able to significantly explain whether there is a dichotomy or a gradient between stratified and mixed marine lakes in Palau, I analyzed the datasets first in a spatial analysis, then in a temporal range of 12 years (2007-2019). I analyzed the distributions of temperature, salinity, dissolved oxygen, and pH to be able to identify the type of data (if it followed a continuous and normal distribution or not), so I could choose the best way of analysis for the data. All analyses described below were made in RStudio *v.2026.04.0+526* (Posit team, 2026).

I calculated all the possible pairs of variables that might have correlations between physical attributes (independent variables) and environmental factors (dependent variables). The physical attributes were considered independent variables because they do not depend on environmental factors to be formed or changed (i.e., the maximum depth of a marine lake does not depend on salinity or dissolved oxygen). Additionally, the environmental factors depend on the physical attributes to vary (i.e., the salinity and the dissolved oxygen change across different depths). I used all environmental and physical variables to make a correlation plot with information about the level of correlation between physical and environmental variables. I used the *corrplot* v.0.95 package (Wei and Simko, 2024) to make the plot and the *vegan* v.2.7-1 package (Oksanen et al., 2025) to make these calculations.

I created a dissimilarity matrix using the Euclidean method through the *vegan v.2.7-1* package (Oksanen et al., 2025) because of the type of data for both environmental and physical variables (abiotic data). I used the Euclidean matrix to calculate the Pairwise Permutational Multivariate Analysis of Variance (pairwise-PERMANOVA) and the Permutational Multivariate Analysis of Dispersion (PERMDISP) to be able to analyze site variance between and within lake states and to assess whether each state has greater or lower environmental variability. I first calculated the means for each factor by year. The ocean sites normally had just one profile because they do not have basins like the marine lakes. On the other hand, mixed and stratified lakes had at least two profiles. In some cases, there were four profiles because the

sampling was made using two different meters in one visit at each site (**Table 1**). I ran a Tukey test using the *stats* v.4.4.1 package (Odeh et al., 1974; Copenhaver et al., 1988). Additionally, to understand which environmental factors and/or physical attributes contribute significantly to the differences in lake states, I used Similarity Percentages (SIMPER) analysis in the *vegan v.2.7-1* package (Oksanen et al., 2025).

The spatial variation analysis had mixed and stratified lake categories considered discrete responses to physical attributes and environmental factors. To analyze the similarities and differences between and within categories, I made two Principal Component Analyses (PCAs) using the *vegan v.2.7-1* package (Oksanen et al., 2025): one for the physical attributes (total volume, total volume without chemocline, surface area, distance to the ocean, tidal effect, and maximum depth) as predictor variables, and the other for the environmental predictor variables (temperature, salinity, dissolved oxygen, and pH). Additionally, I created individual profile plots (Figure S1 and Figure S6) for each site to assess environmental factors in greater detail and to see whether there is a trend that holds across lakes and within their respective states.

For both physical and environmental PCAs, I analyzed which PC had a higher percentage of variance explained. I also calculated the percentage contributions of each physical attribute and environmental factor at each PC. I associated the physical attributes with each marine lake's acronyms (*Table S1*) to be able to understand the correlation between them and the environmental variables. Also, all plots in this thesis were made using the *ggplot2 v.4.0.2* package (Wickham, 2016). To assess if the environmental factors had any correlations with the physical attributes of available sites, I calculated a correlation plot using the *corrplot v.0.95* package (Wei & Simko, 2024).

For the temporal analysis, I made a time-series plot using all environmental factors (temperature, salinity, dissolved oxygen, and pH) between 2007 and 2019 to assess their increasing or decreasing trends. To analyze the presence or absence of trends and to evaluate the changes of these factors through time, I made a linear regression model (*lm*). All time-series plots were accompanied by density plots, showing the density of the data separated by site categories. I did not make any time-series plots for physical attributes

because the great majority of them did not change significantly throughout the years. Thus, for this thesis, these characteristics are considered constant over time.

To understand which variables influence others, how they are correlated and the strength of the correlations (if *corr*~1, then the correlation between variables would be positive and strong; if *corr*~-1, then the correlation between variables would be negative and also strong; if *corr*~0, then there is no strength between variables), I made a correlation plot following the Pearson Correlation Coefficient, which assessed the values of the correlations and whether the correlation is positive or negative between marine lake states. Then, to analyze how much of the variation in the multivariate response (environmental factors) can be explained by physical attributes that characterize the size of the lakes (i.e., total volume, surface area, maximum depth, and level of isolation), I used a Distance-Based Redundancy Analysis (dbRDA).

Results

i. Physical attributes analyses

The variation of Palau sites is explained by both principal component axes: PC1 explaining 70.74% of the data, and PC2 explaining 17.11% of the data. The attribute that contributed most to Principal Component 1 (PC1) was the mean distance between marine lakes and the ocean, contributing to 12.61% of the explained variance, followed by the median distance to the ocean, with 12.46%. For Principal Component 2 (PC2), the attribute that contributed the most was the mean distance to the ocean with 17.66%, followed by the volume without the chemocline, with 16.78% (for all physical attributes, see Table S3). The other PCs from three to ten showed low explanatory percentages (between 0 and 7.82%; see scree plot in Figure S1).

The PCA plot shows different lake states and their physical attributes, which are different from each other. The minimum, median, and mean distances to the ocean and the perimeter have similar effects on the distribution of sites, with a positive correlation, as well as the surface area, volume of the lakes without the chemocline, and the maximum depth. The tidal efficiency seems to be opposite to the tidal lag, minimum, mean, and median distances from the ocean, which were expected.

The PCA also shows that HLM has the largest volume, followed by OTM. Both lakes are stratified ones, with high depths and a big surface area. Also, the HLM is one of the stratified lakes that could not be defined as a meromictic lake because it has a big channel in the middle of its water column, elevating the levels of dissolved oxygen (see individual profiles in the supplemental material). The stratified lake TLN and the mixed lake OLO show a partial overlap in the multivariate space (Figure 3).

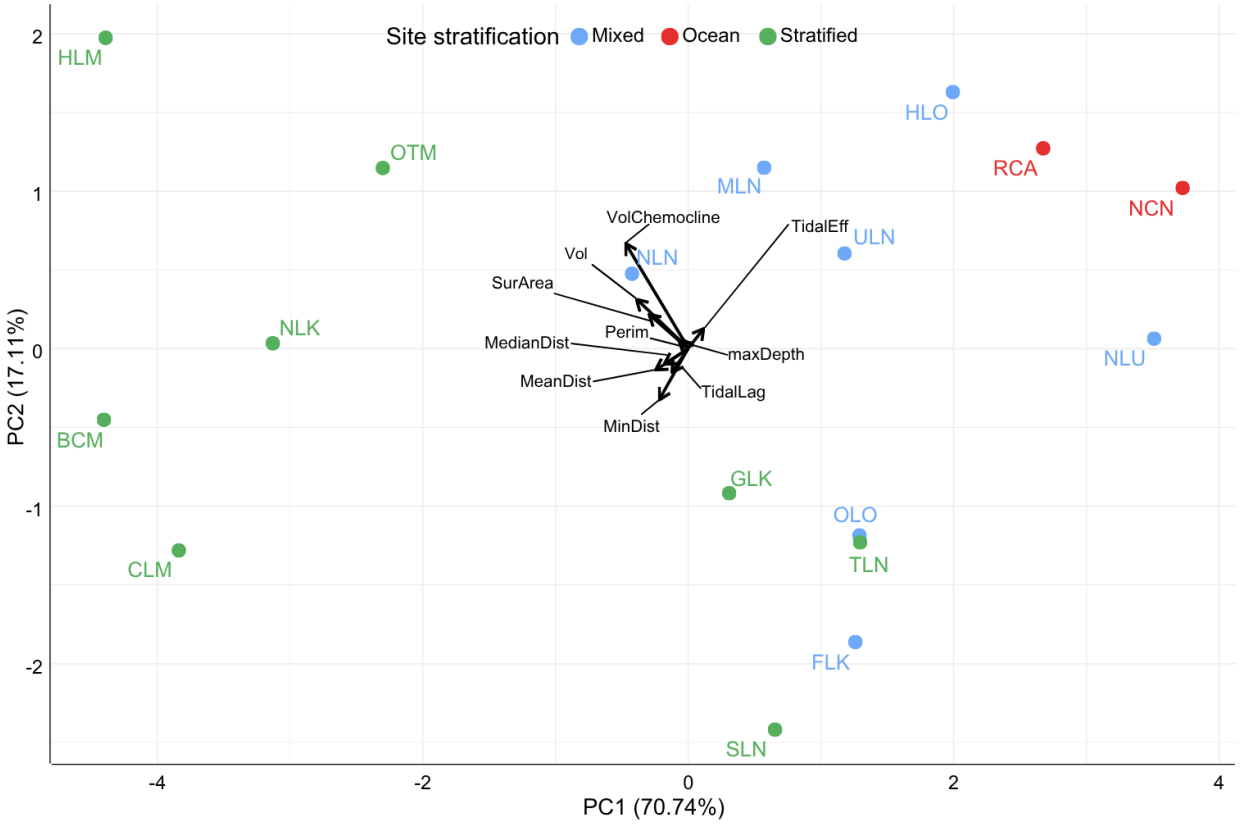


Figure 3. *Physical attributes in the marine lakes of Palau. The red color represents the ocean sites, the blue color represents the mixed lakes, and the green color represents the stratified lakes. The black arrows represent the strength of each physical attribute in grey. Vol = Total volume (m³); VolCh = Volume without the chemocline (m³); SurAr = Surface Area (m²); MinDistOcean = Minimum distance to the ocean (m); MeanDistOcean = Mean distance to the ocean (m); MedianDistOcean = Median distance to the ocean (m); TidLag = Tidal lag (min); TidEff = Tidal efficiency (%); Perim = Perimeter (m); and MaxDepth = Maximum depth (m).*

ii. Environmental factors' analyses

The time series from 2007 to 2019 revealed that stratified lakes had the highest variability across all environmental factors over time, followed by mixed lakes. The density plots (Figure 4.**a.2, b.2, c.2, d.2**) show that ocean sites have a narrower variation for all environmental factors than mixed and stratified lakes. The time series also shows a strong variation in dissolved oxygen and pH between 2010 and 2012 (Figure 4).

The linear regressions over the sites' lines revealed that there was an increase in temperature across years for ocean sites and mixed lakes. However, for stratified lakes, the temperature seems to be constant in the same time range. There was a weak increase in dissolved oxygen and salinity for mixed and stratified lakes, and a constant level of the two variables for the ocean sites. However, for all environmental factors and for all categories of sites, there was a decrease in pH, meaning that the water for the three categories (ocean sites, mixed lakes, and stratified lakes) is getting more acidic (Table 4).

The time-series analysis made it feasible to analyze the within-site environmental variability over the years. It showed that stratified lakes had the highest mean pairwise distances, followed by three mixed lakes (FLK, OLO, and NLU) and one ocean site (NCN; see Figure S3). Also, looking into the variance of each environmental factor, salinity was the variable that had the highest variance, followed by dissolved oxygen, temperature, and pH, respectively. The pH had the lowest variance, but it was the variable that had an impact on the three site categories across time (Figure S4).

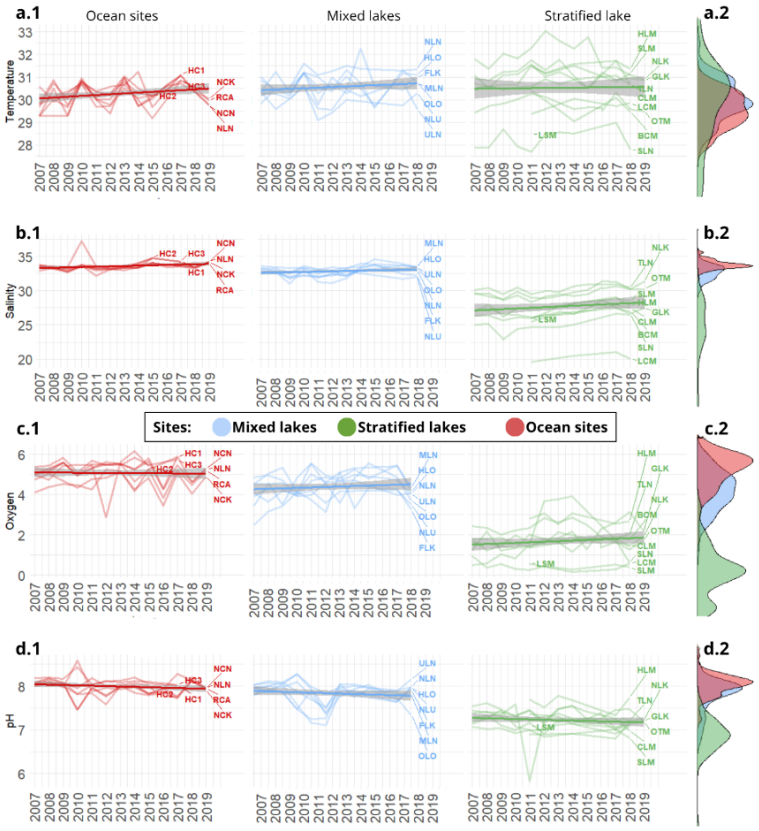


Figure 4. *Time-series plots of Palau marine lakes.* ***(a.1)*** *Average temperatures per year and site.* ***(b.1)*** *Average salinity per year and site.* ***(c.1)*** *Average dissolved oxygen per year and site.* ***(d.1)*** *Average pH per year and site. Each line*

*represents one site, and the strongest line, with the gray range, represents a general trend across the years, along with a 95% confidence interval. The red represents ocean sites, the blue represents the mixed lakes, and the green represents the stratified lakes. The density plots on the right (**a.2, b.2, c.2,** and **d.2**) represent the amount of data that was used to make this time series plot.*

The marine lake profiles show a trend in environmental factors with lake depth. For all marine lake states, the deeper seawater layers have less variation for all environmental factors (stay constant after a certain depth). The stratified lakes show that there is a higher variance for all factors than in mixed lakes, following their depths. Also, the stratified lakes are deeper than the mixed lakes and ocean sites (coves and channels between the ocean and the lake).

The general profiles showed that stratified lakes have a clear pattern of the water column. The temperature variation (ºC) for mixed lakes and ocean sites is narrower than for stratified lakes (**Figure 6. Profile Figure 5**), which reach maximum temperatures up to 36.5ºC, compared with around 31.5ºC (mixed lakes) and 30.7ºC (ocean sites). The salinity (ppt) increased significantly for all lakes as the depth was also increasing, reaching a maximum of 32.8 - 33 ppt at similar depths, around 15 meters (Figure 5).

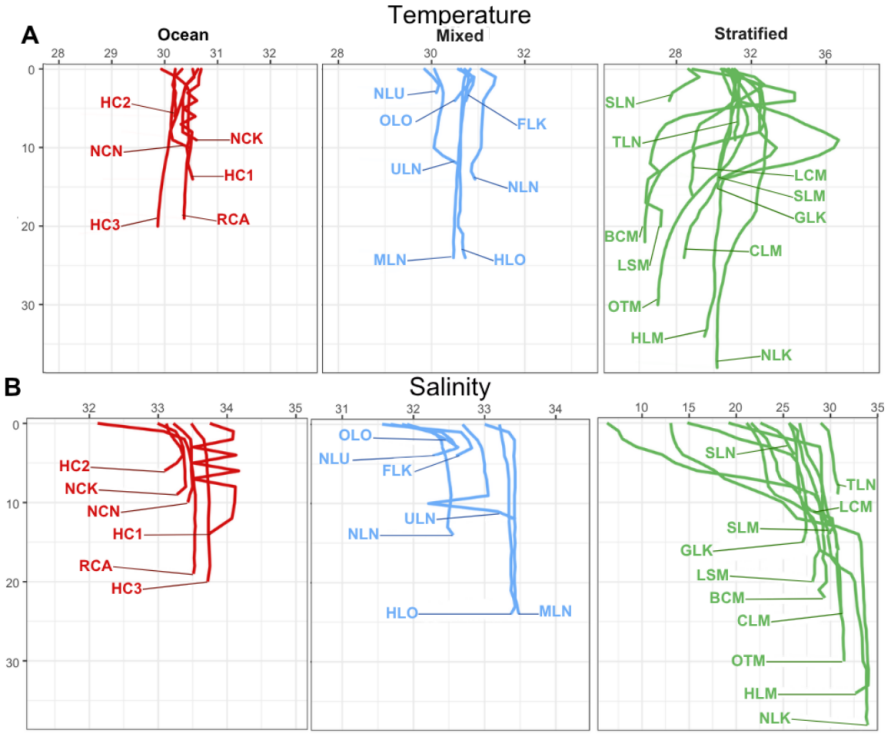


Figure 5. *Profile plots for all the averages of environmental factors across all years, grouped by site categories. Ocean sites are in red lines; mixed lakes in blue lines; and stratified lakes in green lines. Each line corresponds to a*

*site and its depth (meters). (**A**) Average temperatures for all years per site and depth. (**B**) Average salinity for all years per site and depth. Both were measured every meter until the bottom of the site.*

The salinity patterns are similar to the dissolved oxygen (mg/L), which was decreasing until reaching 0 mg/L. The only stratified lake that had a different profile for dissolved oxygen was HLM. For the mixed lakes, the profiles kept relatively constant values for all environmental variables. However, it also showed that the pattern for dissolved oxygen was decreasing as the depth increased. For the ocean sites, the pattern is similar to that of the mixed lakes, with relatively constant levels for all variables. Lastly, the pH variations for ocean sites and mixed lakes also have a very narrow variation across sites in each state. Stratified lakes are more acidic than the other two states, with a range between 6.8 and 7.8, while mixed lakes and ocean sites are more basic environments, between 7.9 and 8 (Figure 6).

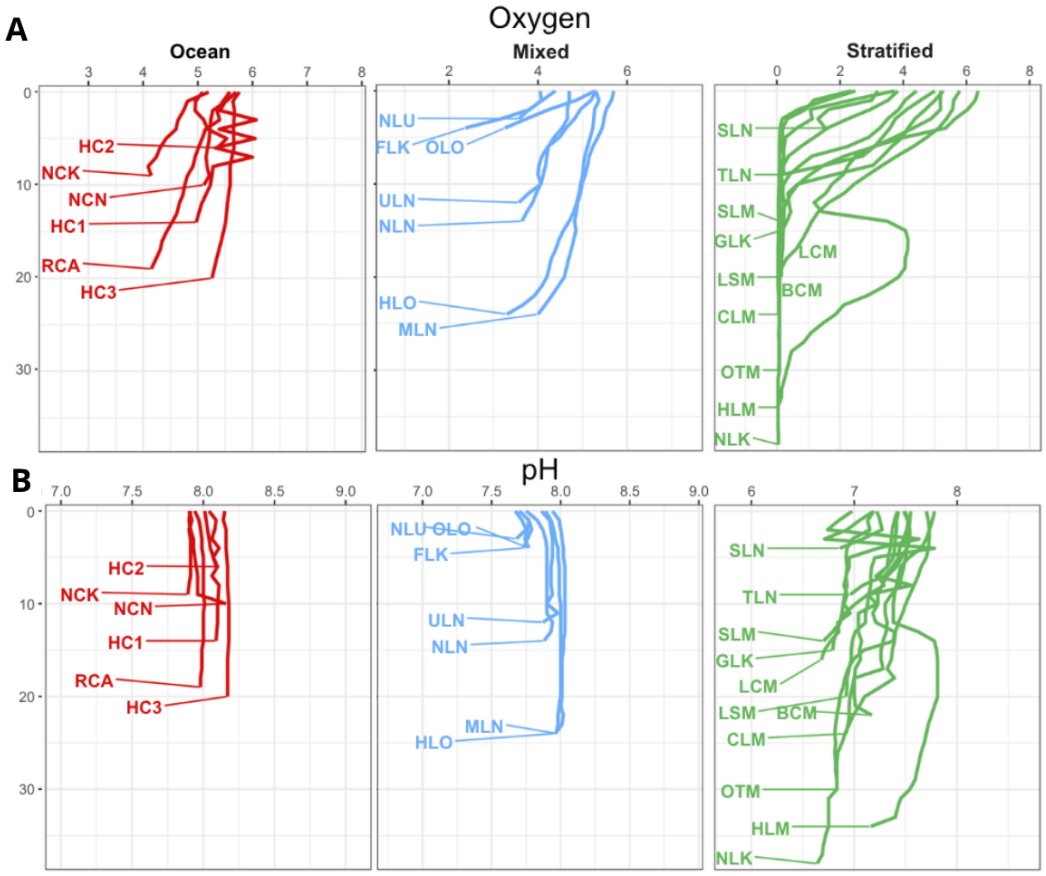


Figure 6. *Profile plots for all the averages of environmental factors across all years, grouped by site categories. Ocean sites are in red lines; mixed lakes in blue lines; and stratified lakes in green lines. Each line corresponds to a site and its depth (meters). (**A**) Average dissolved oxygen levels for all years per site and depth. (**B**) Average pH for all years per site and depth. Both were measured every meter until the bottom of the site.*

The Principal Component Analysis (PCA) for the environmental factors revealed that 90.9% of the variance is explained by PC1 (65.7%) and PC2 (25.2%). The plot shows that mixed lakes largely overlapped with ocean sites, but ocean sites have a narrower distribution than mixed lakes. Also, HLM, NLK, and TLN, all stratified lakes, have minimal overlap with OLO and are very close in the multivariate space to FLK, both mixed lakes. These lakes have similar temperatures (ºC) and are mainly divided by dissolved oxygen levels, which are related to the salinity and pH levels as well (Figure 7).

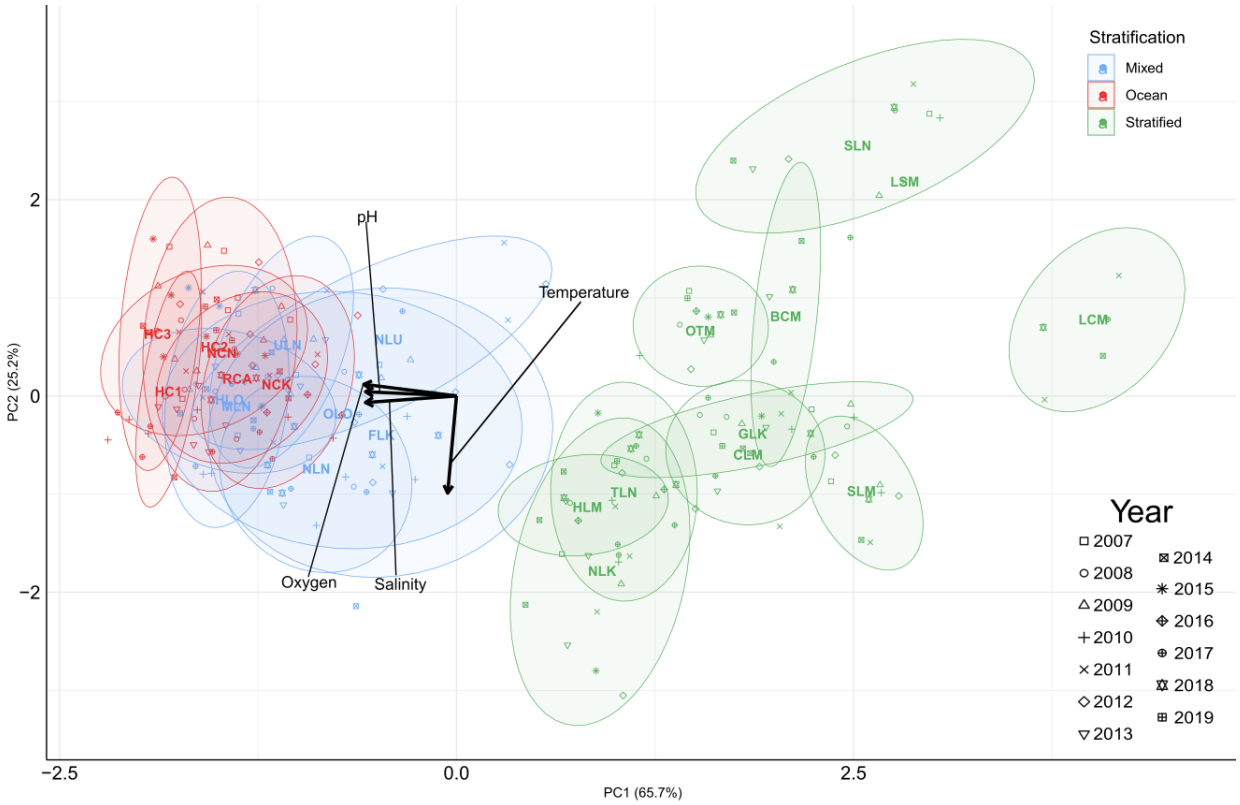


Figure 7. *Principal Component Analysis (PCA) plot of the average environmental factors per site and year for Palau sites between 2007 and 2019. Each year is represented by a unique shape. The mixed lakes are represented in blue, ocean sites in red, and stratified lakes in green. Each shape represents a year, and the ellipses follow the same color from the sites' categories, with 90% confidence intervals for each site.*

The category of stratified lakes has a high variability on the y-axis, which is associated with water temperature. Therefore, the stratified lake SLN is the coolest among this category, and the NLK is the hottest stratified lake according to the plot. On the other hand, mixed lakes are aggregated on the far left of the plot, with an environment that is more saline, basic, and oxygenated, similar to the ocean sites. From all sites in the PCA plot, the Little Crocodile Lake (LCM) is the only one that does not overlap with any other

sites, having an environment that is less saline, more acidic, and more anaerobic compared to the others. The patterns discussed above are more consistent with a dichotomy between mixed and stratified lakes than a continuous environmental gradient, although the variability within the stratified lakes indicates heterogeneity across individual lakes.

With the analysis of physical and environmental data, I also made a plot for the possible correlations between the two. The rectangle in green is the correlation between environmental factors, which shows a high correlation between salinity and oxygen (0.83), salinity and pH (0.77), and oxygen and pH (0.83). The correlations between temperature and the other three factors are pretty weak (0.13, -0.03, and 0.04). Also, there is a negative correlation between temperature and oxygen (-0.03).

The purple rectangles represent the correlation between environmental factors and physical attributes, with a majority of negative values. The strong and positive correlations are between tidal efficiency and three out of four environmental factors: salinity, dissolved oxygen, and pH. The yellow rectangle refers to the correlations between physical attributes. The minimum, mean, and median distances from the site to the ocean have strong correlations, since all three of them are outputs from the same attribute: the distance. The same happens with the total volume and the volume without chemocline. The negative correlations that the plot shows are related to the tidal efficiency, which had negative values for all of the paired correlations with other physical attributes (Figure 8).

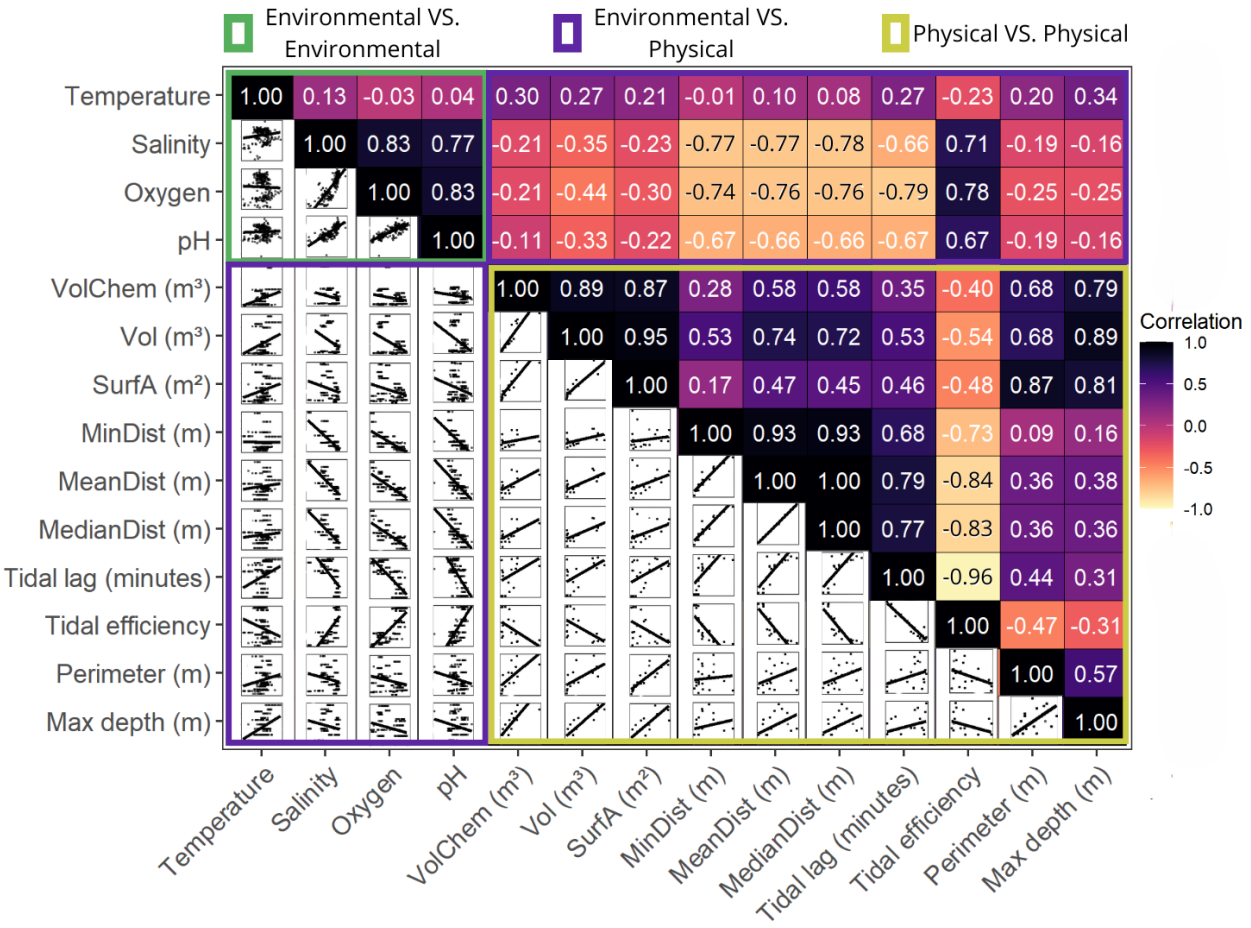


Figure 8. *Correlation plot of environmental factors versus physical attributes for the marine lakes. The coldest colors (from 0 to 1) refer to the positive correlation between variables, while the warmest colors (from 0 to –1) refer to the negative correlation between them. The plot is divided into three subplots: the green rectangle shows correlations among environmental factors; the two purple rectangles show correlations between environmental factors and physical attributes; and the yellow rectangle shows correlations among physical attributes.*

iii. Statistical analyses

The Levene's test below mainly evaluates whether the variance is homogeneous between groups. For all four variables (temperature, dissolved oxygen, salinity, and pH), the test is highly significant (Table 2). Therefore, the assumption of homogeneous variances is violated for all variables, resulting in heterogeneous environments when comparing groups.

Table 2. *Levene's test for homogeneity of variance analysis for the four environmental factors (temperature, dissolved oxygen, salinity, and pH of the water). Df = degrees of freedom; Fvalue = test to compare the variance between groups with the variance within groups, and Pr(>F) = p-value (probability of observing an F value if the null hypothesis, where the variances are equal, were true).*

Environmental factor	Df	Fvalue	Pr(>f)
Temperature	23	200.1	< 2.2e-16 ***
Oxygen	23	81.041	< 2.2e-16 ***
Salinity	23	189.48	< 2.2e-16 ***
Ph	23	63.65	< 2.2e-16 ***
	7651		

*Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

The dispersion of stratified lakes in the PCA and the Levene's test results that were described above suggest heterogeneity across lakes. This variability may be consistent with the hypothesis that processes associated with small, isolated marine systems contribute to different marine lake states (mixed or stratified). The pairwise PERMANOVA revealed significant differences across the three categories (Mixed, Stratified, and Ocean; *p=0.003*). However, the effect strengths varied: mixed and ocean sites differed (R^2=0.124), but not as much as stratified lakes, which showed much stronger separation (R^2=0.46 and R^2=0.57). The stratified lakes, therefore, occupy a distinct multivariate environmental state from the other two, having mixed lakes and ocean sites more similar to each other than stratified lakes, even having a statistically significant difference (Table 3).

Table 3. *Outputs for the pairwise PERMANOVA. Df refers to degrees of freedom, SumsOfSqs refers to the sum of squares, F.Model refers to the model used to produce the output, R^2 refers to the strength of the effects between pairs, and p-value and p-adjusted refer to the significance of the model.*

pairs	Df	SumsOfSqs	F.Model	R2	p.value	p.adjusted
Mixed vs Ocean	1	17.51143	19.81986	0.1240138	0.001	0.003*
Mixed vs Stratified	1	302.90502	139.25709	0.4576955	0.001	0.003*
Ocean vs Stratified	1	449.22415	220.52545	0.5749956	0.001	0.003*

*Significant codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

To verify whether the pairwise PERMANOVA results are not driven primarily by an unbalance within-group dispersion, I ran a Permutational Multivariate Analysis of Dispersion (PERMDISP). The null hypothesis for the PERMDISP is that all states have the same multivariate dispersion (i.e., they have a similar variability within groups). The test showed that at least one of the states has a different dispersion from the others. While the PERMDISP analyzes the dispersions of the data, the Tukey test compares the means for each state pair. The Tukey test showed that the differences between Ocean vs Mixed were not

significant (*p-value* = 0.789), but the other two pairs – Mixed vs Stratified and Stratified vs Ocean – had significant values (*p-values = 0.01* and *0.003*, respectively), showing that the stratified lakes have environmental factors that distinguish them from the mixed lakes and ocean sites (Table 4).

Table 4. *(**A**) Permutation test for homogeneity of multivariate dispersions (PERMDISP). Df = degrees of freedom, SumSq = sum of squares, MeanSq = mean of squares, F = F-test, N.Perm = number of permutations, and Pr(>F) = p-value. (**B**) Tukey multiple comparison of means test. 95% family-wise confidence level. Fit: aov(formula = distances ~ group, data = df)*

A	***Df***	***SumSq***	***MeanSq***	***F***	***N.Perm***	***Pr(>F)***
Groups	2	6.8697	3.4349	9.2359	999	0.002**
Residuals	21	7.8099	0.3719			

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

B	***Comparison***	**Mean difference**	**Adjusted p-value**
	Ocean vs. Mixed	-0.224	0.789
	Stratified vs. Mixed	0.958	0.010**
	Stratified vs. Ocean	1.182	0.0028***

To analyze which environmental factors and physical attributes are the main variables responsible for the site state, I calculated the drivers of differences by running a Similarity Percentage analysis (SIMPER). The analysis focused on each pair of states: first, between mixed lakes and ocean sites, second, between mixed and stratified lakes, and third, between ocean sites and stratified lakes. Between mixed lakes and ocean sites (lagoons, coves, and channels), mixed lakes are farther from the ocean than the coves, channels, and lagoons. Ocean sites tend to be deeper, and there is a significant difference between ocean sites and mixed lakes in tidal efficiency. However, even with the same tidal efficiency averages, the *p-value* was significant. Then, the differences between them appear to be related to the site morphology (i.e., physical attributes), rather than the water chemistry (environmental factors; see Table 5). These results agree with the overlap between mixed lakes and ocean sites seen in the environmental factors' PCA.

Table 5. *Similarity Percentage (SIMPER) analysis between mixed lakes and ocean sites. The variables below were significant to differentiate mixed lakes (ava) and ocean sites (avb). The ava and avb are related to the mean of each group, not to individual lakes.*

Variables	***Mixed (ava)***	***Ocean (avb)***	***p-value***
Mean distance to ocean (m)	128	30	0.002 ***
Max depth (m)	12	15	0.001 ***
Tidal efficiency	1	1	0.016 .

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

The pair containing mixed and stratified lakes revealed that there are strong differences in environmental and physical variables between the two. It shows that the stratified lakes are larger, deeper, less saline, more anaerobic, and more acidic than mixed lakes. Surprisingly, the distance of the lake to the ocean between mixed and stratified lakes was not significant, suggesting that internal lake variables associated with the stratification process are the main and significant differences between the lakes' states, rather than distance-to-the-ocean values (Table 6).

Table 6. *Similarity Percentage (SIMPER) analysis between mixed and stratified lakes. The variables below were significant to differentiate mixed lakes (ava) and ocean sites (avb). The ava and avb are related to the mean of each group, not to individual lakes.*

Variable	***Mixed (ava)***	***Stratified (avb)***	***p-value***
Volume (m^3)	205,000	685,800	<0.001***
Surface area (m^2)	19,140	39,470	0.002***
Maximum depth (m)	12	23	0.016 .
Salinity (ppt)	33	28	<0.001***
Dissolved oxygen (mg/L)	4	2	<0.001***
pH	8	7	0.005***

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

For ocean sites and stratified sites, almost all variables (6 out of 9) had significant differences. Compared to lagoons, coves, and channels (here grouped as ocean sites), the stratified lakes are larger in volume and farther from the ocean. They are characterized by a strongly reduced tidal exchange and exhibit an environment that is less saline, anoxic, and acidic. Therefore, the results are consistent with the PCA, since the stratified lakes are clustered on one side and the ocean sites are clustered on the opposite side of the plot (Table 7).

Table 7. *Similarity Percentage (SIMPER) analysis between ocean sites and stratified lakes. The variables below were significant to differentiate ocean sites (ava) and stratified lakes (avb). The ava and avb are related to the mean of each group, not to individual lakes.*

Variable	***Ocean***	***Stratified***	***p-value***
Volume (m^3)	85,390	685,800	<0.001***
Distance to the ocean (m)	30	253	<0.001***
Salinity (ppt)	33	28	<0.001***
Oxygen (mg/L)	5	2	<0.001***
pH	8	7	0.002***
Tidal efficiency	1	0	<0.001***

*Significance codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1*

The SIMPER results suggest, in a general way, that different variables can differentiate and separate lakes' states. For mixed lakes and ocean sites, the differences were mainly due to physical

attributes (distance to the ocean, depth, and level of isolation), with very similar environmental factor variables. For mixed lakes and ocean sites, both physical (volume, surface area, and depth) and environmental (salinity, dissolved oxygen, and pH) variables are responsible for the differences between them. Lastly, between stratified lakes and ocean sites, the results suggest that there are strong differences in the level of isolation, volume, and environmental factors (except temperature), indicating that the stratified state is the most different state compared to the mixed state and ocean sites.

The dbRDA analysis shows a strong and significant value ($F=7.6511$; $p=0.001$; $R^2=0.718=71.8\%$), which indicates that physical attributes that determine the size of the lakes (i.e., surface area, total volume, maximum depth, and distance to the ocean) significantly explain the multivariate environmental structure (i.e., temperature, salinity, dissolved oxygen, and pH) of the lakes. The analysis also shows that the most important physical attribute driving the environmental factor constraints is the distance of the lakes to the ocean, with the secondary most important being the maximum depth of the site.

The physical attributes that have weak and redundant results are the surface area and the volume of the site. The dbRDA analysis also shows that 71.8% of the environmental factor variations are explained by a strong gradient (represented by the CAP1 axis), which is strongly controlled by the distance of the site to the ocean, followed by the maximum depth of the site. The effect of surface area and volume has a weaker effect on the differentiation of lake states (Figure 9).

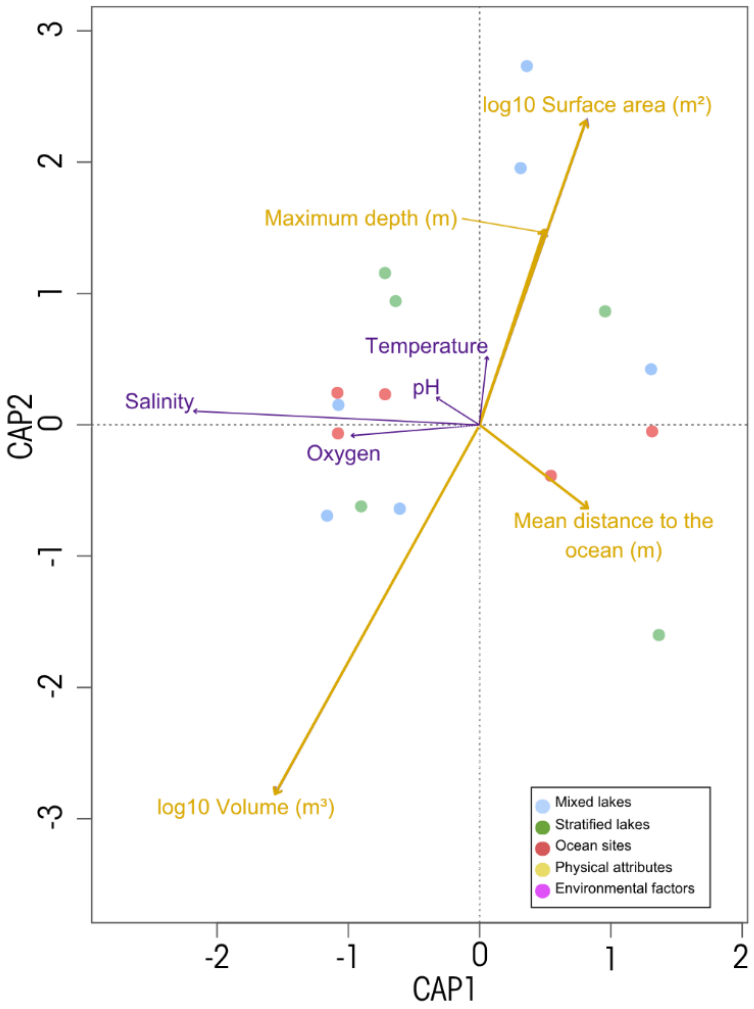


Figure 9. *Distance-based redundancy analysis (dbRDA) showing a strong environmental gradient structured by lake physical attributes, with CAP1 representing the dominant axis of variation driven primarily by distance to the ocean and maximum depth.*

Discussion

i. Physical attributes

Marine lakes exhibit a great variation in physical attributes, including the degree of isolation from the open ocean. In terms of connectivity, most lakes maintained some level of connection to the ocean through subterranean channels and fissures (Hamner & Hamner, 1998; Becking et al., 2024), indicating a continuum from nearly isolated to well-connected ecosystems (Becking et al., 2024). According to the PCA for physical attributes in this work, the variability of the level of isolation between the lakes and the ocean also follows the same pattern, with continuous connectivity, going from extremely connected lakes to lakes with

no or very minimal contact with the ocean (Hamner & Hamner, 1998; Cerrano et al., 2006; Maas, 2021; Becking et al., 2024).

In this work, the total volume had a high contribution to differentiating lake states. This means that besides the levels of isolation, physical attributes such as maximum depths were also contributing to different descriptions of each lake state, with stratified lakes being larger, deeper, and more isolated than mixed lakes and ocean sites. Although Hamner & Hamner (1998) described how the tidal exchange happened (e.g., slowly through fissures in fenestrated karst and rapidly through submarine tunnels), neither tidal efficiency nor tidal lag was able to significantly differentiate Palau's lake states.

Additionally, the PCA in this thesis pointed to the stratified lake (TLN) and the mixed lake (OLO) overlap in the same multivariate space, meaning that both have similar physical attributes but are at the same time in different lake states. This can lead to the hypothesis that OLO may change from a mixed to a stratified state, or the TLN may change from a stratified to a mixed state in the future. As this hypothesis goes against the permanently meromictic lakes in Hamner & Hamner (1998), the evidence of overlap between two different lakes that are in different states highlights the importance of analyzing these marine lakes throughout a longer time series, as this work is showing. The lake's physical attributes as continuous variables between sites contribute to the separation of stratified and mixed lake states, as shown in the following sections of this thesis.

ii. Environmental factors

The main environmental factors analyzed in this work showed that stratified lakes are significantly different from mixed lakes and ocean sites, having little overlap of the stratified lakes TLN, NLK, and HLM with the mixed lakes FLK and OLO. The variations of salinity, pH, and dissolved oxygen influence the gap between the two states. Ocean sites' environmental factors do not vary a lot, resulting in relatively constant environmental factors through depth. The same happens with mixed lakes. However, the lakes' profiles show that the variation per depth within states is still higher than at the ocean sites.

Mixed lakes and ocean sites in Palau had an increasing trend of temperature throughout the years, while stratified lakes had a relatively constant trend. The stratified lakes had a higher variability of temperatures than the other two groups, having stratified lakes with 29ºC and others with up to 33ºC. These high temperatures for stratified lakes were also analyzed in Raja Ampat lakes, which ranged from 30ºC to 36.8ºC (Maas, 2021; Aji et al., 2023). Ha Long Bay experienced even more extreme thermal conditions, with water temperatures up to 36ºC (Cerrano et al., 2006). In the Berau Reef System, temperatures in the lakes were higher than their surrounding environments (Cleary et al., 2018).

The extreme thermal results observed in other places were different from the scenario analyzed of marine lakes in Palau, where there are stratified lakes with high temperatures throughout the years, but also stratified lakes with colder temperatures for the same time period. For temperate marine lakes, the seasonal variation is stronger than that of tropical marine lakes, with months of the year that have a frozen water body, such as Lake Mogil'noye, analyzed by Krasnova (2021). For this Russian lake, the temperature stabilizes around four meters, while in marine lakes in Palau, there is no strong evidence of constant temperatures throughout the depths.

The salinity levels in the marine lakes of Palau exhibited the highest variability between and within lake states, consistent with other papers' analyses (Hamner & Hamner, 1998; Cerrano et al., 2018; Cleary et al., 2018; Becking et al., 2024). The time series plots pointed out relatively constant trends for ocean sites and mixed lakes. Besides these trends, the stratified lakes had a big increase throughout the years, moving from ~27.5 to ~30 ppt. Stratified lakes also have a higher variability in salinity than mixed lakes and ocean sites, reflecting that some lakes have a more saline environment, while other lakes have a very saline environment. Raja Ampat marine lakes had a salinity range from 26.4 to 29.1 ppt, resulting in brackish conditions (Aji et al., 2023; Becking et al., 2024). The salinity for these lakes was higher in lakes with a bigger connection to the ocean, while more isolated lakes were less saline. Ha Long Bay lakes had a very low salinity (Cerrano et al., 2018).

In Hamner & Hamner's (1998) work, they showed that the stratified marine lakes in Palau displayed a permanent stratification, also with brackish waters that overlapped the anoxic saline hypolimnion, a region of the water column that is located under the chemocline, usually at the bottom of the lake. Their descriptions are very similar to this work because the group of stratified lakes had a higher salinity as the depths increased and achieved the constancy at similar depths that also had a near-zero concentration of dissolved oxygen (end of the *chemocline* and beginning of the *hypolimnion*; see Hamner & Hamner, 1998).

The dissolved oxygen levels clearly divided mixed lakes and ocean sites as a group, and stratified lakes as another. Stratified lakes had dissolved oxygen levels in the same pattern as salinity, where there is an increasing trend throughout the years, varying from ~1.9 mg/L to ~2.0 mg/L. Mixed lakes had similar environments to ocean sites, but they still had a significant difference between their environmental factors, showing that even having similar patterns, the values within each pattern were significantly different.

There is a decreasing trend in the pH for all sites throughout the years analyzed. While mixed lakes and ocean sites kept a more neutral pH of ~8, stratified lakes had a more acidic pH, of around 7.25. Raja Ampat (Aji et al., 2023; Becking et al., 2024) had similar results to this thesis, having a more acidic environment in the lakes compared to the ones in the open sea, also seen in Berau's Reef system, Indonesia (Cleary et al., 2018). The fact of having a lower pH in all sites, not just one lake state, can be linked to the recent acidification of the ocean.

iii. Statistical analyses

The statistical analyses support distinct environmental states, with stratified lakes being different from mixed lakes and ocean sites. Stratified lakes also exhibit substantially greater internal variability than the other two groups, indicating environmental heterogeneity not just between site states, but also within state rather than a single uniform condition. Lakes' patterns and the SIMPER analysis suggested that the volume, surface area, maximum depth, distance to the ocean, tidal efficiency, salinity, dissolved oxygen, and pH are the main drivers responsible for separating stratified and mixed states. This means that the physical

attributes of a lake, combined with the connectivity from the lake to the ocean (distance to ocean), and the environmental factors (dissolved oxygen, salinity, and pH), together support the existence of different ecosystem states. Additionally, marine lakes do not form a simple continuous gradient from mixed to stratified states. The PCA from environmental data reveals a clear separation between these states.

The pairwise PERMANOVA comparing marine lake states confirms statistically that differences across the three groups are significant (stratified lakes are strongly different from both other ecosystems). The marine lake states differ not only in their average environmental conditions but also in their within-group environmental variations, meaning that stratified lakes are more heterogeneous environments than mixed lakes and ocean sites, which was also seen in the PERMDISP analysis and the Tukey test (discussed below). This can lead to different ecological niches that species that inhabit the lakes may occupy.

The SIMPER analysis demonstrates that both physical attributes and environmental factors contribute to lake state differences, and that the physical attributes also influence the environmental factor variations across lakes. The presence of distinct ecological states among marine lakes in Palau, rather than a single and continuous gradient that separates both states, may be described as continuous isolation levels and depths that influence the environmental factors that can act together to differentiate distinct and dichotomous marine lake states. All multivariate analyses support that stratified lakes are consistently different from mixed lakes and ocean sites by both physical and environmental variables. However, the significant PERMDISP results indicate that the stratified lakes are also heterogeneous within the state group, as assessed in the time series and profiles for the stratified state, strengthening the argument that while they represent a distinct state, they also have significant variability relative to differences in physical attributes, connectivity to the ocean, and local environmental processes.

The combined multivariate analyses in this work suggest that marine lakes in Palau are structured by both continuous environmental and physical gradients, which give rise to distinct ecological states, in this case, the mixed and stratified lake states. The dbRDA analysis demonstrated that the physical attributes explain a large proportion of the variation in environmental conditions, specifically the level of isolation

(i.e., distance to the ocean) and maximum depth being the most influential predictors of the model. The results suggest that seawater connectivity and the lake limnology generate a strong gradient across the marine lakes. However, the results also demonstrated that the surface area and lake volumes, together with the limited evidence for a robust size threshold, provide little support for a classical Small Island Effect (based primarily on 'island' area). Thus, there is a strong effect from the isolation and maximum depth of the lakes, which may be an alternative small-island effect for these environments.

Despite the continuous isolation levels and maximum depths of the dbRDA model, the PCAs, pairwise PERMANOVA, PERMDISP, and SIMPER analyses consistently support the existence of distinct mixed and stratified lake states. The PCA for the environmental data revealed a clear separation between these states in the multivariate space of the plot, while the pairwise PERMANOVA confirmed that the centroids of each state differed significantly, indicating that the environmental conditions of the two marine lake states are statistically distinct. The SIMPER analysis identified which environmental factor (or factors) were responsible for marine lake states.

These findings suggest that marine lakes are organized across a continuous level of isolation, but exhibit threshold-like environmental responses that produce the distinct marine lake states in Palau. In other words, gradual changes in isolation and maximum depths appear to drive environmental conditions continuously, while physicochemical processes respond nonlinearly, resulting in two contrasting lake states. This framework reconciles the apparent discrepancy between the dbRDA and the clustering analyses and supports that marine lake states represent discrete expressions in an environmental continuum rather than mutually exclusive categories, in which different marine lakes described as mixed and stratified seemed to be very close in the same multivariate space, that is, having very similar lake structures and environmental factors.

iv. Analysis limitations

This study provides a comprehensive assessment of which physical attributes and environmental factors structure marine lake states in Palau. However, limitations should be considered when interpreting its

results. The analyses focused exclusively on abiotic variables, including water chemistry, lake morphology, and connectivity to the ocean (also mentioned as the level of isolation of the site from the ocean). Although these variables explained a substantial proportion of the observed environmental variation among lake states, they do not fully capture the ecological complexity of marine lake ecosystems.

Additionally, some sites exhibited a *zigzag* pattern and other unexpected values of environmental factor profiles across different depths. The North Cove (HC1) of the mixed Heliofungia lake (HLO) showed the *zigzag* pattern for all environmental factors. Some stratified lakes – i.e., the stratified lakes Little Crocodile Lake (LCM), Spooky Lake (SLM), L-Shaped Lake (LSM), and Clear Lake (CLM) – showed unexpected patterns in the middle of the water column for pH levels, with a rapid increase followed by a decrease from 0 to around 9 meters deep. The mixed lake Uet Era Ngchas (ULN) had a decrease followed by an increase in salinity levels, which had a different pattern than the other mixed lakes analyzed.

The original files from all of the Palau sites show that these unexpected patterns happened not just because of the different meters that were used, but also because the original values were used to calculate the averages of different years per depth for the analyses. For the HC1, the Quanta meter measurements were sampled every two meters, while the DS5 meter measurements were sampled every meter. When I calculated the means per depth, the measurements for odd depths had higher values than for even depths, generating this '*zigzag*' pattern. For even depths, where both meters were used, the measurements from the Quanta meter were lower than the DS5 meter, lowering the mean across years for those depths. For depths where only the DS5 meter was used, the values were higher because the means did not include the lower values from the Quanta meter.

The unexpected patterns for the mixed lake ULN can be due to the same differences in meters used to get the data, the meter touching the bottom of the site, or the differentiation in the values from different calibrated meters. When calculating the averages per depth for all years, if there was a big discrepancy between them, the value of that specific depth would not be representative of the depth for all years. Lastly, for the stratified lakes' pH levels, another reason for the unexpected decreases and increases may be the

different years in which events such as El Niño and La Niña occurred. The environmental factors were different than the years without these events, as seen in the time series plot (Figure 4). This means that in El Niño/La Niña years, such as in 2011-2012 and 2015-2016, the levels of pH had some variations compared to the other years for the same sites.

Lastly, other mixed – Ngel Lake (NLU), Ngeruktabel Lake (NLN), Ongael Lake (OLO), Flatworm Lake (FLK), Heliofungia Lake (HLO), and Mekeald Lake (MLN) – and stratified marine lakes – Shrimp Lake (SLN), Big Crocodile Lake (BCM), and Hot Water Lake (HLM) – exhibited unexpected increases or decreases in environmental factor values near the deepest depths. These differences may be due to the meters touching the bottom of the lakes, changing their results. This may happen depending on the tide levels inside the lake, connected to the time of each sampling. If the team visited the lake at different times of the day (for the same months), the tide level would be different, resulting in different maximum depths of each lake. For this work, a troubleshooting method could include the identification of outliers before calculating the averages per year and depth, combined with interpolation to estimate the values between two different values in the dataset before calculating the averages. Also, choosing the minimum depth for each lake across the analyzed years would solve the problem of decreases or increases at the bottom of the lakes.

Marine lakes, also considered marine islands (Carlson et al., submitted), are shaped by interactions between abiotic (environmental factors and physical attributes) and biotic (populations and communities inhabiting the lakes) characteristics. Consequently, the inclusion of biological data (i.e., metabarcoding sequence data from phyto- and zooplankton, microbial, plant, and vertebrate communities) would provide a more complete understanding of the mechanisms underlying not just the differentiation among lake states, but also how this ecosystem heterogeneity influences lakes' community assemblages. Integrating community composition and structure, functional traits from different species, or species interactions with the abiotic dataset produced by this work would allow direct evaluation of how environmental gradients influence ecological structure and ecosystem functioning. Additionally, for

physical attribute data, multiple variables from the same attribute were considered independent from each other (e.g., the minimum, median, and mean distance from the lake to the ocean) in the correlation plot. This may have resulted in a bigger weight of these variables on the physical attributes PCA, inducing a non-representative differentiation of states that can change significant results.

The data set used for this work was analyzed in different ways to have a complete overview of the marine lakes in Palau. The variance of each site was analyzed not just on the PCA, but also using the pairwise PERMANOVA, PERMDISP, and Tukey test as a whole to assess the differences between the marine lake states. Also, the fact that the analyses for the environmental factors and physical attributes were made separately and then as a big group of variables decreased the uncertainty of the individual effects of these variables on the lake states, as well as the effects of the physical attributes on the environmental factors, as shown by the correlation plot.

Given these limitations, the present study offers a useful environmental baseline for understanding the structure of marine lake states in Palau. The analytical framework developed here may serve as a starting point for future integrative studies combining abiotic, biotic, and temporal data, contributing to a broader understanding of the ecological processes that could shape marine lake diversity, stability, and responses to environmental change.

Additional Statements

Conflict of interest statement: There was no conflict of interest for this project by any of its authors.

AI statement: The AI was used to troubleshoot R code chunk problems and correct orthography. I did not use it to write the following thesis.

CHAPTER 1: SUMMARY AND CONCLUSIONS

i. Question and hypothesis

The main goal of this thesis was to understand how environmental factors and physical attributes influence marine lake states in island biogeography, while the specific goal was to analyze whether marine lake states in Palau follow a dichotomous or a gradient pattern related to their abiotic characteristics. The main

hypotheses of this work were related to whether the variation in environmental factors through time, combined with the physical attributes of lakes, would significantly explain a dichotomy or a continuous gradient between stratified and mixed states.

The statistical analyses comparing lake states and ocean sites were sufficient to accept the hypothesis that the environmental factors and physical attributes of the marine lakes significantly influence a gap between them, indicating a dichotomous pattern for these ecosystems.

ii. Statistical analysis that supports a dichotomy pattern in Palau

Different statistical analyses show the same patterns across marine lake states: both physical attributes and environmental factors together influence the differentiation of marine lake states. The differences between the marine lakes and the ocean sites are significant for all lake states, showing that even having similar ecosystems (physical attributes and environmental factors working together), mixed and stratified lakes and ocean sites continue to be different environments.

The analyses also showed that the stratified state had a greater environmental factor variation across stratified lakes, suggesting that this state has a bigger environmental heterogeneity when compared to the mixed lake state and the ocean sites. Also, for the gap between mixed and stratified lakes, the main environmental factors that may increase or decrease the dichotomy between the two states are salinity, dissolved oxygen, and pH, whereas temperature is the main factor that may differentiate the environmental heterogeneity within marine lake states.

iii. Foundational work for next analyses including biological data

The species-area relationship (Schoener, 1976; Lomolino and Weiser, 2001) studies summarize patterns of biodiversity for a high number of taxa across different spatiotemporal scales (Arrhenius, 1921; Whitehead and Jones, 1969; Brooks et al., 1997; Brown & Lomolino, 1998; Lomolino & Weiser, 2001). While large islands show an area-dependent species richness, small islands' species diversity may increase or decrease independently of the island area (Mcarthur and Wilson, 1967; Whitehead and Jones, 1969; Lomolino, 2000; Lomolino & Weiser, 2001), seen as the Small-Island Effect (SIE; see Whitehead and Wilson, 1969;

Lomolino, 2000). Marine lakes' limnology shows a great ecosystem heterogeneity across marine lake states. Considered as small bodies of seawater, they may be identified similarly to small islands. They are environments with different levels of isolation, maximum depths, total volumes, and surface areas, which may increase the differentiation of marine lake states.

The results from this study can help future work to analyze the species diversity variations based on descriptions of the lakes' states assessed by this thesis. If the analysis of variance of the species richness between and within states can be highlighted, then the answer to the questions: *"Can marine islands be under the effect of the small-island effects? How do environmental and physical variables influence the SIE in marine islands?"* might elucidate the similar patterns of small islands and marine lakes, generating new evidence to define marine lakes as true marine islands.

REFERENCES

Agrawal, A.A. 2003. ***Community genetics: New insights into community ecology by integrating population genetics***. Ecology, 84: 543–544.

Aji, L. P. et al. **The diversity of molluscan faunas in marine lakes of Raja Ampat, West Papua, Indonesia.** Contributions to Zoology 92, 391–430 (2023).

Antonovics, J. 1976. **The input from population genetics: "The new ecological genetics**". Systematic Botany, 1: 233–245.

Argueta-Guzmán, M., McFrederick, Q.S., Spasojevic, M.J. 2025. **Multitrophic assembly influences β-diversity across a tripartite system of flowering plants, bees, and bee-gut microbiomes.** Ecography, 2025: e07490.

Armbrecht, L. 2020. **The potential of sedimentary ancient DNA to reconstruct past ocean ecosystems.** Oceanog, 33.

Arrhenius, O., 1921. **Species and area**. Journal of Ecology, 9, 95±99.

Becking, L.E., Renema, W., Santodomingo, N.K., Hoeksema, B.W., Tuti, Y. and Voogd, N.J. de. 2011. **Recently discovered landlocked basins in Indonesia reveal high habitat diversity in anchialine systems**. Hydrobiologia, 677: 89–105.

Blanchette, F., Montroy, S., Patris, S., and Dawson, M.N. 2020. **Marine lakes as biogeographical islands: A physical model for ecological dynamics in an insular marine lake, Palau.** Frontiers of Biogeography, 12.

Brown, J.H. & Lomolino, M.V. (1998). **Biogeography**. Sinauer Associates, Inc., Sunderland, MA.

Bruno, J.F., Stachowicz, J.J., and Bertness, M.D. 2003. **Inclusion of facilitation into ecological theory.** Trends in Ecology & Evolution, 18: 119–125.

Burnett, W.C., Landing, W.M., Lyons, W.B., and Orem, W. 1989. **Jellyfish lake, Palau.** EoS Transactions, 70: 777–783.

Butt, N., Halpern, B.S., O'Hara, C.C., Allcock, A.L., Polidoro, B., Sherman, S., et al. 2022. **A trait-based framework for assessing the vulnerability of marine species to human impacts**. Ecosphere, 13: e3919.

Calbet, A. 2024. **Plankton size classification. In: Plankton in a changing world: The impact of global change on marine ecosystems** (A. Calbet, ed), pp. 11–14. Cham: Springer Nature Switzerland.

Cardinale, B.J., Duffy, J.E., Gonzalez, A., Hooper, D.U., Perrings, C., Venail, P., et al. 2012. **Biodiversity loss and its impact on humanity.** Nature, 486: 59–67.

Cerrano, C., Azzini, F., Bavestrello, G., Calcinai, B., Pansini, M., Sarti, M., et al. 2006**. Marine lakes of karst islands in Ha Long Bay (Vietnam).** Chemistry and Ecology, 22: 489–500.

Cerrano C, Bavestrello G, Bertolino M, et al. 2020. **The Ha Long Bay Marine Ecosystem. An Unprecedented Opportunity for Evolutionary Studies on Marine Taxa.** https://doi.org/10.1007/978-3-030-51260-6_5

Cleary D., Polónia A., Voogd N. 2018. **Bacterial Communities Inhabiting the Sponge *Biemna fortis*, Sediment and Water in Marine Lakes and the Open Sea.** Microbial Ecology. https://doi.org/10.1007/s00248-018-1156-6

Cody, M.L. and Diamond, J.M. 1975. **Ecology and evolution of communities**.

Colin, P.L. 2009. **Marine environments of Palau. Koror, Palau** : [San Diego, Calif.]: Coral Reef Research Foundation: Indo-Pacific Press.

Connor, E.F. and Simberloff, D. 1979. **The assembly of species communities: Chance or competition?** Ecology, 60: 1132–1140.

Copenhaver, M. D., and Holland, B.S. (1988). **Computation of the distribution of the maximum studentized range statistic with application to multiple significance testing of simple effects.** Journal of Statistical Computation and Simulation, 30, 1–15. doi:10.1080/00949658808811082.

Csardi, G. and Nepusz, T. 2006**. The igraph software package for complex network research**.

Darwin, C. 1953**. On the origin of species by means of natural selection**. American Midland Naturalist, 49: 937.

Dawson, M.N. 2016. **Island and island-like marine environments**. Global Ecology and Biogeography, 25: 831–846.

Dawson, M.N. and Hamner, W.M. 2008. **A biophysical perspective on dispersal and the geography of evolution in marine and terrestrial systems**. J. R. Soc. Interface., 5: 135–150.

Djeghri, N. et al. **High trophic plasticity in the mixotrophic Mastigias papua-Symbiodiniaceae holobiont: implications for the ecology of zooxanthellate jellyfishes**. Marine Ecology Progress Series 666, 73–88 (2021).

Elton, C.S. and Elton, C.S. 1927. **Animal ecology**. New York: Macmillan Co.

Faria, R., Johannesson, K. & Stankowski, S. **Speciation in marine environments: Diving under the surface**. j. evol. Biol. 34, 4–15 (2021).

Gause, G.F. 1935. **Experimental demonstration of Volterra's periodic oscillations in the number of animals**. Journal of Experimental Biology, 12: 44–48.

Godsoe, W., Jankowski, J., Holt, R.D., and Gravel, D. 2017. **Integrating biogeography with contemporary niche theory**. Trends in Ecology & Evolution, 32: 488–499.

Govaert, L. et al. **Integrating fundamental processes to understand eco-evolutionary community dynamics and patterns**. Functional Ecology 35, 2138–2155 (2021).

Grinnell, J. 1917. **The niche-relationships of the California thrasher.** The Auk, 34: 427–433.

Hamner, W.M. and Hamner, P.P. 1998. **STRATIFIED MARINE LAKES OF PALAU (WESTERN CAROLINE ISLANDS)**. Physical Geography, 19: 175–220.

Hirschfeld, M., Barnett, A., Sheaves, M., and Dudgeon, C. 2023. **What Darwin could not see: Island formation and historical sea levels shape genetic divergence and island biogeography in a coastal marine species**. Heredity, 131: 189–200.

Hubbell, S.P. 2001. **The unified neutral theory of biodiversity and biogeography (MPB-32)**. Princeton University Press.

Hubbell, S.P. 1979. **Tree dispersion, abundance, and diversity in a tropical dry forest: That tropical trees are clumped, not spaced, alters conceptions of the organization and dynamics**. Science, 203: 1299–1309.

Hutchinson, G.E. (1975). **A treatise on limnology / vol. 1: Geography, physics, and chemistry. Part 2: chemistry of lakes.** New York; London: Wiley-Interscience.

Keddy, P.A. 1992. **Assembly and response rules: Two goals for predictive community ecology**. Journal of Vegetation Science, 3: 157–164.

Krasnova, E.D. 2021**. Ecology of meromictic lakes of Russia**. 1. Coastal marine waterbodies. Water Resour, 48: 427–438.

Lavorel, S. and Garnier, E. 2002. **Predicting changes in community composition and ecosystem functioning from plant traits: Revisiting the holy grail.** Functional Ecology, 16: 545–556.

Lawton, J. H. **Are There General Laws in Ecology**? Oikos 84, 177–192 (1999).

Leibold, M.A., Holyoak, M., Mouquet, N., Amarasekare, P., Chase, J.M., Hoopes, M.F., et al. 2004. **The metacommunity concept: A framework for multi-scale community ecology**. Ecology Letters, 7: 601–613.

Leray, M., Yang, J.Y., Meyer, C.P., Mills, S.C., Agudelo, N., Ranwez, V., et al. 2013. **A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: Application for characterizing coral reef fish gut contents**. Front Zool, 10: 34.

Levin, S.A. 1992**. The problem of pattern and scale in ecology: The Robert H. MacArthur award lecture**. Ecology, 73: 1943–1967.

Lomolino, M.V. (2000) **Ecology's most general, yet protean pattern: the species area relationship.** Journal of Biogeography, 27, 17±26.

Maas D. 2021. **Islands of Sea: the evolution, ecology and conservation of marine lake invertebrates.** https://doi.org/10.18174/543115

MacArthur, R.H. and Wilson, E.O. 1967**. The theory of island biogeography**. Princeton University Press.

Martin, L.E., Bell, L.J., Dawson, M.N, and Box, P.O. 2001. **Marine lake monitoring project update, 2001: Biological and physical properties of goby lake, big jellyfish lake, and ongeim'l tketau**.

May, R.M. 1974. **Stability and complexity in model ecosystems**. Princeton University Press.

McCann, K., Hastings, A., and Huxel, G.R. 1998. **Weak trophic interactions and the balance of nature**. Nature, 395: 794–798.

McLaren, I. A. (1967) **Physical and Chemical Characteristics of Ogac Lake, a Landlocked Fiord on Baffin Island.** Journal of the Fisheries Research Board of Canada. 24(5): 981-1015. https://doi.org/10.1139/f67-088

Meyerhof, M.S., Wilson, J.M., Dawson, M.N., and Michael Beman, J. 2016. **Microbial community diversity, structure, and assembly across oxygen gradients in meromictic marine lakes, Palau**. Environmental Microbiology, 18: 4907–4919.

Montanyès, M., Weigel, B., and Lindegren, M. 2023. **Community assembly processes and drivers shaping marine fish community structure in the North Sea**. Ecography, 2023: e06642.

Murphy, E., Fredette-Roman, C., Rochman, C., Gerber, L.R., and Polidoro, B. 2023. **A trait-based framework for assessing the relative vulnerability of marine biodiversity to macroplastics**. SSRN.

Newsome, S.D., Martinez del Rio, C., Bearhop, S. and Phillips, D.L. 2007. **A niche for isotopic ecology**. Frontiers in Ecology and the Environment, 5: 429–436.

Odeh, R. E., and Evans, J. O. (1974). **Algorithm AS 70: Percentage Points of the Normal Distribution.** Applied Statistics, 23, 96–97. doi:10.2307/2347061.

Oksanen, J., Simpson, G.L., Blanchet, F.G., Kindt, R., Legendre, P., Minchin, P.R., et al. 2025. **Vegan: Community ecology package.**

Patris, S., Martin, L. E., Bell, L. J., & Dawson, M. N. **Expansion of an introduced sea anemone population, and its associations with native species in a tropical marine lake (Jellyfish Lake, Palau)**. Frontiers of Biogeography 11, (2019).

Pimm, S.L. and Lawton, J.H. 1978. **On feeding on more than one trophic level**. Nature, 275: 542–544.

Pinheiro, H.T., Bernardi, G., Simon, T., Joyeux, J.-C., Macieira, R.M., Gasparini, J.L., et al. 2017. **Island biogeography of marine organisms.** Nature, 549: 82–85.

Posit team, (2025). **RStudio: Integrated Development Environment for R**. Posit Software, PBC, Boston, MA. URL http://www.posit.co/.

Rapacciuolo, G., Beman, J.M., Schiebelhut, L.M., and Dawson, M.N. 2019. **Microbes and macro-invertebrates show parallel β-diversity but contrasting α-diversity patterns in a marine natural experiment**. Proc. R. Soc. B., 286: 20190999.

Rapacciuolo, G. and Blois, J.L. 2019. **Understanding ecological change across large spatial, temporal, and taxonomic scales: Integrating data and methods in light of theory.** Ecography, 42: 1247–1266.

Ricklefs, R.E. 1987. **Community diversity: Relative roles of local and regional processes**. Science, 235: 167–171.

Ricotta, C. 2017. **Of beta diversity, variance, evenness, and dissimilarity**. Ecology and Evolution, 7: 4835–4843.

Schoener, T.W. (1976) **The species area relationship within archipelagoes: models and evidence from island birds**. Proceedings of the XVI International Ornithological Congress, 6, 629±642.

Simberloff, D.S. and Wilson, E.O. 1969. **Experimental zoogeography of islands: The colonization of empty islands**. Ecology, 50: 278–296.

Skelly, D.K., Post, D.M., Smith, M.D., Lovejoy, T.E., and Hutchinson, G.E. 2011. **The art of ecology: The writings of G. Evelyn Hutchinson.** New Haven, UNITED STATES: Yale University Press.

Steele, J. H. **A comparison of terrestrial and marine ecological systems**. Nature 313, 355–358 (1985).

Strelkov, P., Shunatova, N., Fokin, M., Usov, N., Fedyuk, M., Malavenda, S., *et al.* 2014. **Marine Lake Mogilnoe (Kildin Island, the Barents Sea): one hundred years of solitude**. *Polar Biol*, **37**: 297–310.

Vellend, M. 2010. **Conceptual synthesis in community ecology**. The Quarterly Review of Biology, 85: 183–206.

Vellend, M. 2016. **The theory of ecological communities**. Princeton (N.J.) Oxford (GB): Princeton University Press.

Vellend, M. and Geber, M.A. 2005. **Connections between species diversity and genetic diversity**. Ecology Letters, 8: 767–781.

Walker, K.F. and Likens, G.E. 1975. **Meromixis and a reconsidered typology of lake circulation patterns: With 3 tables in the text.** SIL Proceedings, 1922-2010, 19: 442–458.

Wallace, A.R. 1880. **Island life, or, the phenomena and causes of insular faunas and floras: Including a revision and attempted solution of the problem of geological climates**. London: Macmillan.

Weigelt, P., Steinbauer, M.J., Cabral, J.S., and Kreft, H. 2016. **Late Quaternary climate change shapes island biodiversit**y. Nature, 532: 99–102.

Whittaker, R.H. 1960. **Vegetation of the Siskiyou Mountains, Oregon and California**. 30: 279–332.

Whittaker, R.J., Fernández-Palacios, J.M., Matthews, T.J., Rigal, F. and Triantis, K.A. 2018. **Archipelagos and meta-archipelagos**. Frontiers of Biogeography, 10.

Wickham, H. 2016. **Ggplot2: Elegant graphics for data analysis**, 2nd ed. Cham, Switzerland: Springer International Publishing.

Wickham H, François R, Henry L, Müller K, Vaughan D (2026). **dplyr: A Grammar of Data Manipulation. R package version 1.2.1**, https://dplyr.tidyverse.org.

Wiens, J.A. 1989. **Spatial scaling in ecology**. Functional Ecology, 3: 385.

Wilson, J., Abboud, S., & Beman, J. M. **Primary Production, Community Respiration, and Net Community Production along Oxygen and Nutrient Gradients: Environmental Controls and Biogeochemical Feedbacks within and across "Marine Lakes".** Front. Mar. Sci. 4, (2017).

Zacharias, M.A. and Gregr, E.J. 2005. **Sensitivity and vulnerability in marine environments: An approach to identifying vulnerable marine areas**. Conservation Biology, 19: 86–97.

Zadereev, E.S., Boehrer, B., and Gulati, R.D. 2017. **Introduction: Meromictic lakes, their terminology, and geographic distribution**. In: Ecology of meromictic lakes (R. D. Gulati, E. S. Zadereev, and A. G. Degermendzhi, eds), pp. 1–11. Cham: Springer International Publishing.

SUPPLEMENTARY DATA

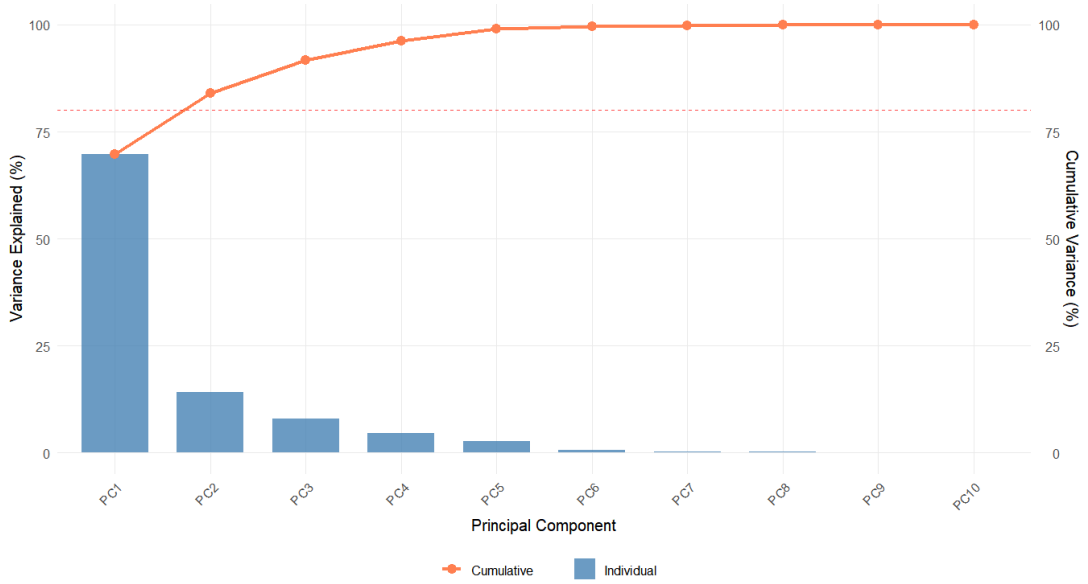


Figure S1. Scree plot for physical attributes data. It shows the data variance explained and the cumulative variance in percentage for each principal component (PC). The blue columns represent the individual PC, and the orange line represents the cumulative explanation of the data variance.

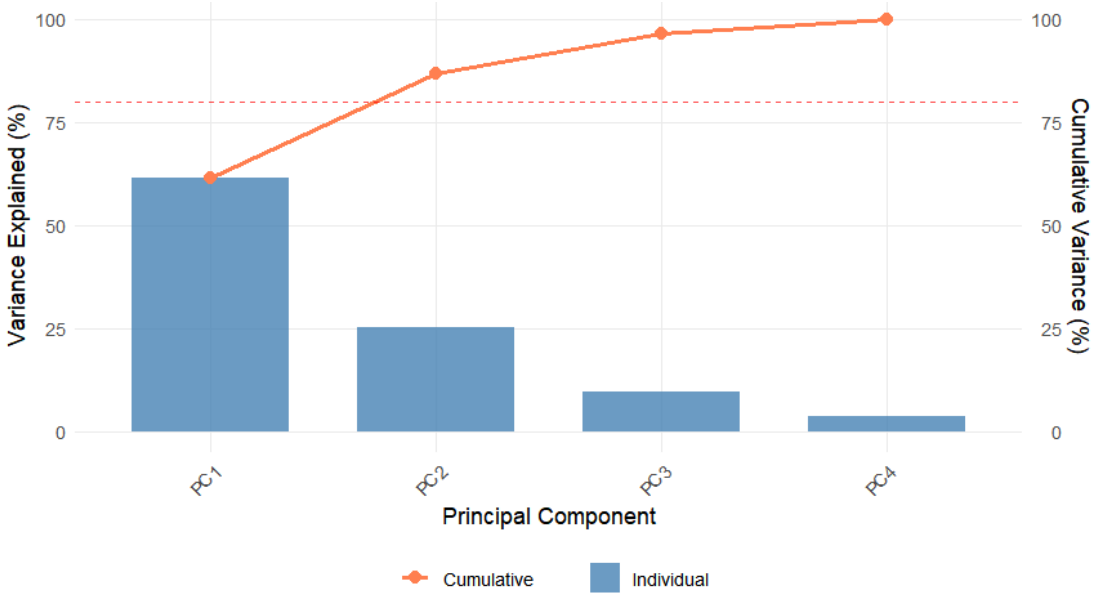


Figure S2. Scree plot of environmental factors' PCA. It shows the data variance explained and the cumulative variance in percentage for each principal component (PC). The blue columns represent the individual PC, and the orange line represents the cumulative explanation of the data variance.

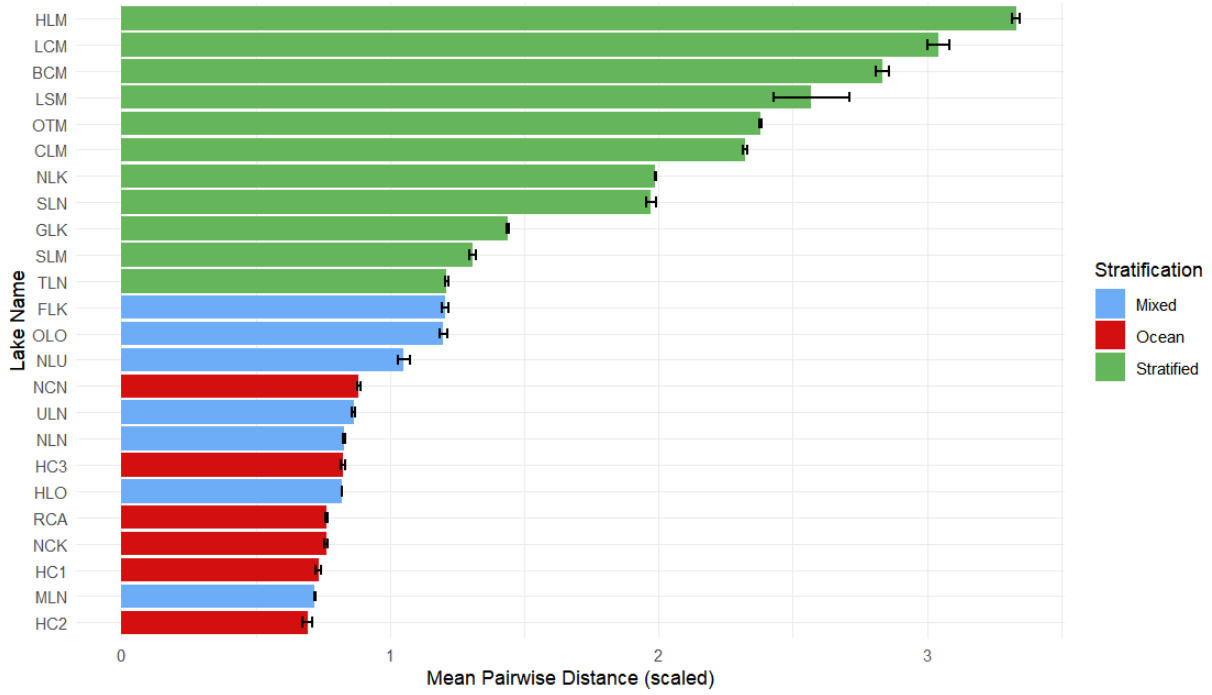


Figure S3. Within-site environmental variability across years. The black intervals in each row represent the standard deviation of the data across time. The mean pairwise distance was calculated with a distance matrix that grouped the name of each site and their respective categories.

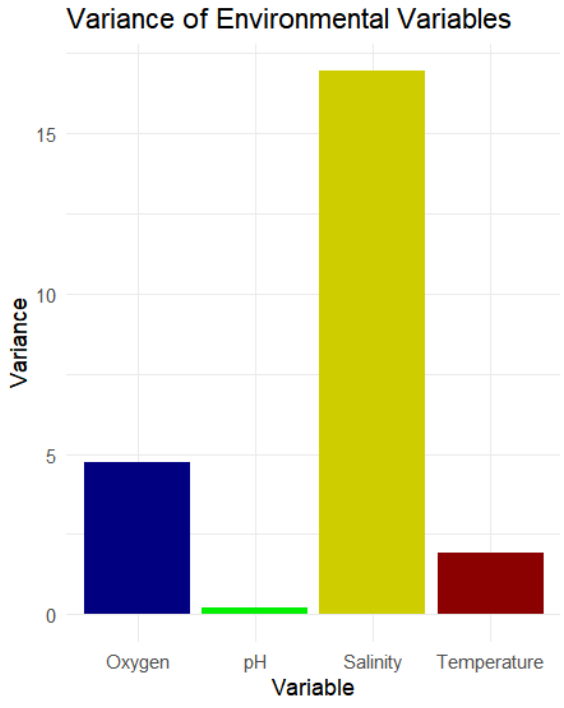


Figure S4. Variance of each environmental factor data. Dissolved oxygen is represented by the navy blue color, the pH is represented by lime green, the salinity is represented by yellow, and the temperature is represented by the red color.

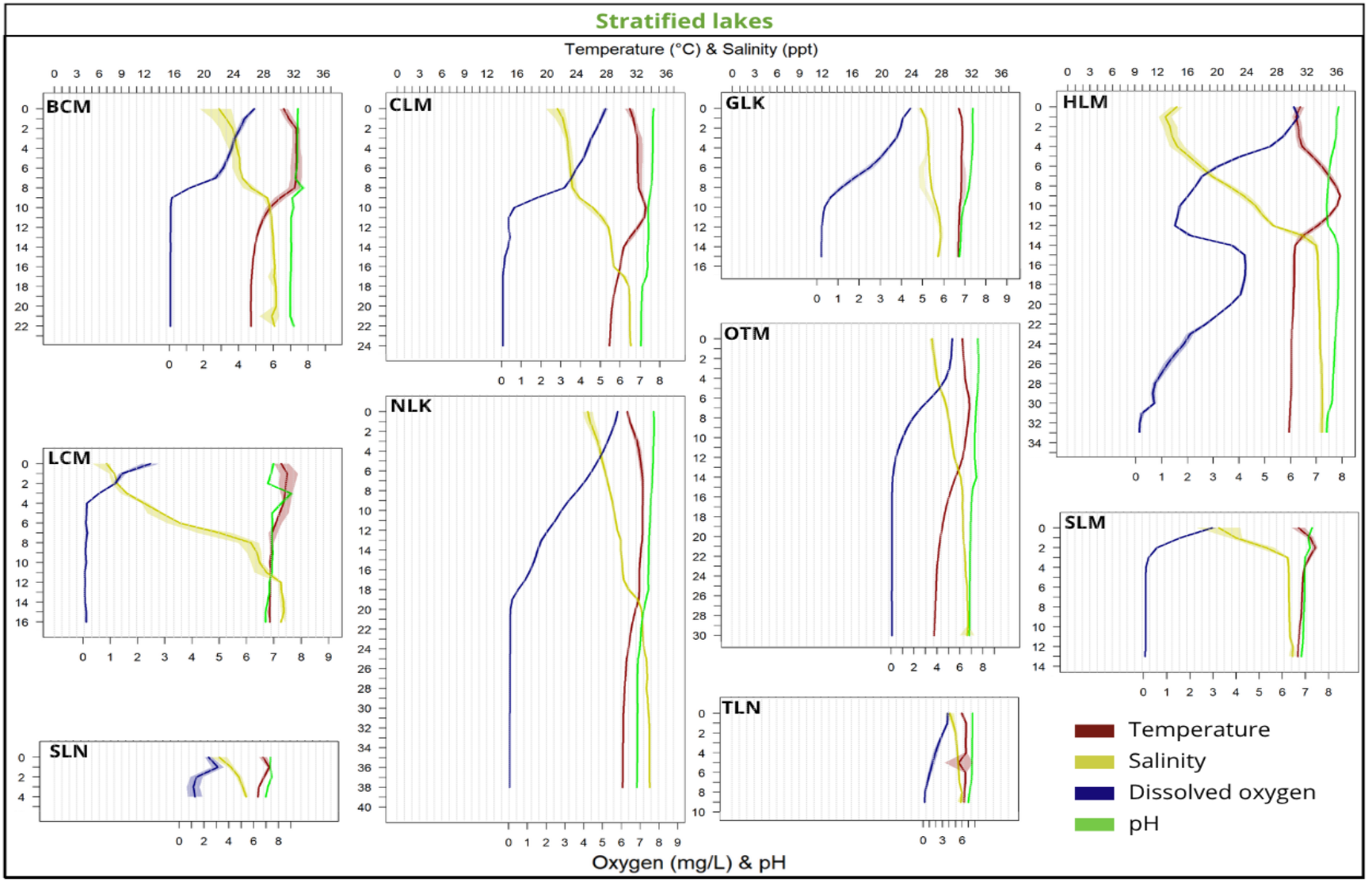


Figure S5. Profiles of average environmental factors (temperature, salinity, dissolved oxygen, and pH) for all stratified marine lakes in Palau. The superior axis is fixed for all profiles. Temperature is represented by the dark red lines, salinity is represented by the yellow lines, dissolved oxygen is represented by the navy blue lines, and pH is represented by the green lines.

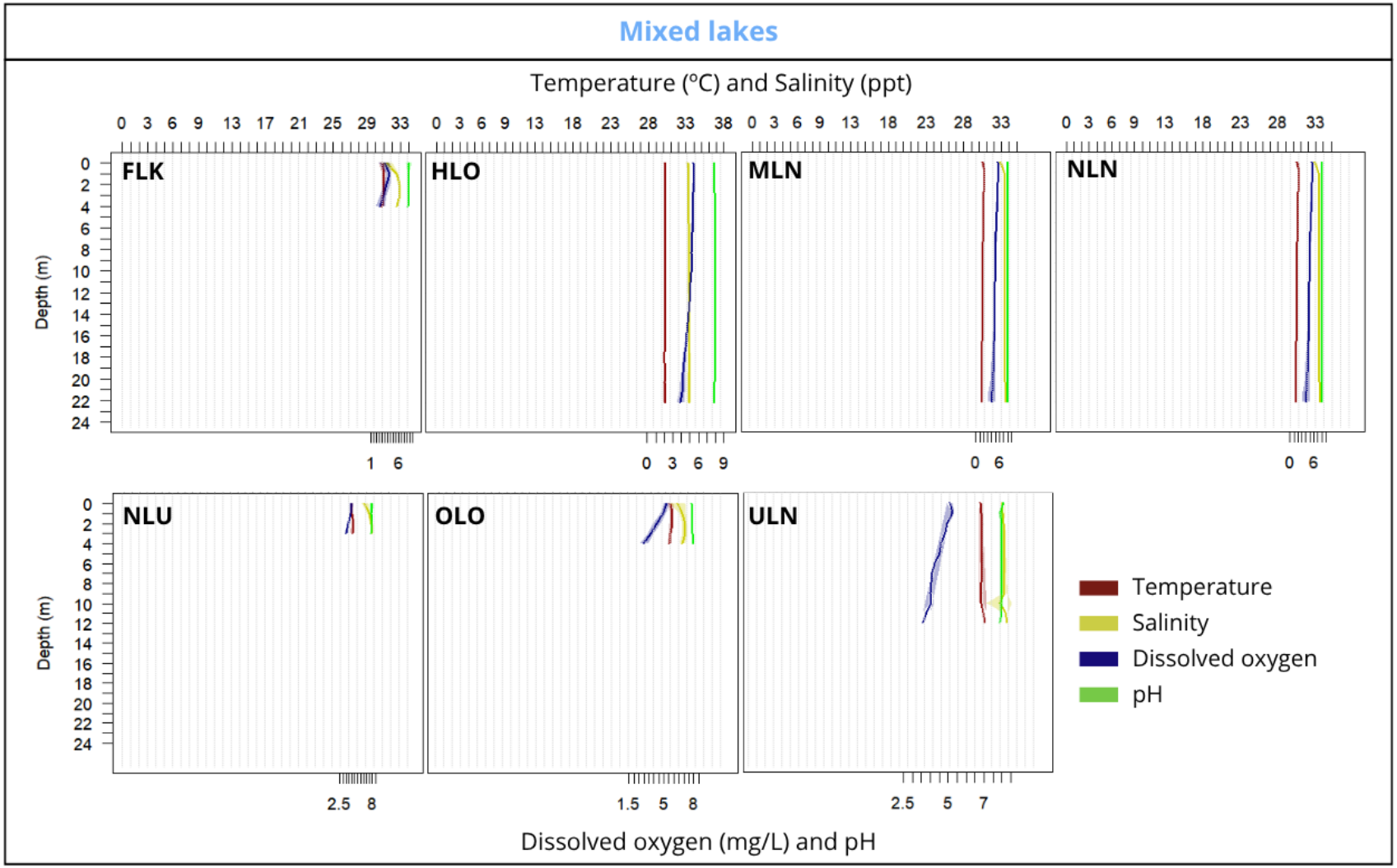


Figure S6. Profiles of average environmental factors (temperature, salinity, dissolved oxygen, and pH) for all mixed marine lakes in Palau. The superior axis is fixed for all profiles. Temperature is represented by the dark red lines, salinity is represented by the yellow lines, dissolved oxygen is represented by the navy blue lines, and pH is represented by the green lines.

Table S1. *List of Palau lake name acronyms, full name of the lake, island, stratification, description citations, whether both environmental and physical data are available, and coordinates (latitude and longitude).*

Lake acronym	Lake name	Island name	Stratification	PHYS attributes	ENV factors	Latitude	Longitude
BCM	Big Crocodile Lake	Mecherchar	Stratified	☑	☑	7°9'18"N	134°22' 9"E
CLM	Clear Lake			☑	☑	7°09'10.06"N	134°21'34.18"E
HLM	Hot Water Lake			☑	☑	7°9.084'N	134°20.947'E
LCM	Little Crocodile Lake			☑	☑	7°09'02.21"N	134°21'20.53"E
LSM	L-shaped Lake			☑	☑	7°09.317' N	134°22.367' E
OTM	Tourist Lake			☑	☑	7°09'39.62"N	134°22'34.02"E
SLM	Spooky Lake			☑	☑	7°09'09.41"N	134°21'48.86"E
NLK	Big Jellyfish Lake	Ngermeuangel		☑	☑	7°19'24.64" N	134°30' 29.55" E
GLK	Goby Lake			☑	☑	7°18'54.15"N	134°30'07.65"E
SLN	Shrimp Lake	Ngeruktabel		☑	☑	7°15'21.58"N	134°26'45.63"E
TLN	T-Lake			☑	☑	7°18'27.51"N	134°26'31.55"E
FLK	Flatworm Lake	Ngermeuangel	Mixed	☑	☑	7°19'17.86"N	134°30'21.84"E
LLN	Long Lake	Ngeruktabel		☑	☑	7°15.442'N	134°24.579' E
MLN	Mekaeald Lake			☑	☑	7°17'52.37"N	134°26'55.63"E
ULN	Ngchas Lake			☑	☑	7°15'11.32"N	134°20'42.70"E
NLN	Ngeruktabel Lake			☑	☑	7°16'38.55"N	134°25'51.39"E
HLO	Heliofungia Lake	Ongael		☑	☑	7°15.117'N	134°22.417' E
OLO	Ongael Lake			☑	☑	7°15'26.75"N	134°22'57.36"E
NLU	Ngel Lake	Ulebsechel		☑	☑	7°18'40.85"N	134°28'20.30"E
ULU	Ulong Lake	Ulong		☑	☑	7°17.142'N	134°16.902' E
NCK	NLK Cove	Ngermeuangel	Ocean	☐	☑	7°19'45.33"N	134°30'48.66"E
HC1	North HLO Cove	Ongael		☐	☑	7°15'12.90"N	134°22'28.44"E
HC2	South HLO Cove			☐	☑	7°15'03.70"N	134°22'29.14"E
HC3	HLO Channel			☐	☑	7°15'01.02"N	134°22'23.32"E

Table S2. *Eigenvalues, variance explanation percentage, and cumulative explanation percentage for all principal components analyzed in the physical attributes' PCA.*

PC	Eigenvalue	Variance explanation percentage (%)	Cumulative explanation percentage (%)
PC1	7.0735	70.74	70.74
PC2	1.7109	17.11	87.84
PC3	0.5208	5.21	93.05
PC4	0.3499	3.5	96.55
PC5	0.2246	2.25	98.8
PC6	0.0673	0.67	99.47
PC7	0.0297	0.3	99.77
PC8	0.0197	0.2	99.96
PC9	0.0032	0.03	100
PC10	4e-4	0	100

Table S3. *Physical attribute contribution percentages on each dimension (PC) on the PCA.*

Physical attributes	Dim.1	Dim.2	Dim.3	Dim.4	Dim.5	Dim.6	Dim.7	Dim.8	Dim.9
Volume without chemocline	8.33	16.78	0.065	2.05	49.34	0.95	0.12	22.16	0.19
Volume	11.59	7.49	4.11	1.21	0.01	33.81	3.64	8.51	25.34
Surface area	12.02	6.54	1.23	6.07	0.39	9.15	0.99	6.83	49.57
Minimum distance to ocean	7.93	17.66	15.60	10.44	3.29	5.09	3.97	35.77	0.23
Mean distance to ocean	12.61	3.87	3.74	2.92	1.96	6.25	6.05	7.12	10.99
Median distance to ocean	12.46	3.79	3.25	5.62	2.87	9.31	6.85	10.36	2.44
Tidal lag	8.71	12.92	10.15	26.52	0.56	10.45	28.34	1.50	0.85
Tidal efficiency	9.90	11.48	7.58	12.02	0.79	7.14	49.47	1.58	0.05

Table S4. *SIMPER summary between groups: mixed lakes vs ocean sites; mixed lakes vs stratified lakes; and ocean sites vs stratified lakes.*

Variable	Mixed vs. Ocean	Mixed vs. Stratified	Ocean vs. Stratified
Volume (m³)	**** (p = 0.998*)*	**** (p = 0.001)*	**** (p = 0.001)*
Surface area (m²)	*NS (p = 0.115)*	*** (p = 0.002)*	*NS (p = 0.652)*
Distance to ocean	*** (p = 0.002)*	*NS (p = 0.926)*	**** (p = 0.001)*
Maximum depth	**** (p = 0.001)*	** (p = 0.016)*	*NS (p = 0.835)*
Salinity	*NS (p = 1.000)*	**** (p = 0.001)*	**** (p = 0.001)*
Dissolved oxygen	*NS (p = 0.772)*	**** (p = 0.001)*	**** (p = 0.001)*
Temperature	*NS (p = 0.804)*	*Marginal (p = 0.083)*	*NS (p = 0.396)*
pH	*NS (p = 0.746)*	*** (p = 0.005)*	*** (p = 0.002)*
Tidal efficiency	** (p = 0.016)*	*NS (p = 0.512)*	**** (p = 0.001)*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.

NS = non-significant, and Marginal = very close to 0.05, but still bigger.