MATH 524: LINEAR ALGEBRA

ANALYSIS REPORT

Analysis of Melting Sea Ice Caps through Microwave Satellite Data

Zino Meyer

Climate Preservation Team of San Diego, San Diego State University

5500 Campanile Dr, San Diego, CA 92182

Email: zino@cptsd.com

Phone: (619) - 629 5698

Website: cptsd.com

December 7, 2024

Abstract

The decrease of arctic ice caps has a variety of effects on our planet like rising sea levels and change in landscape, affecting coastal communities and marine ecosystems. Through long-term satellite monitoring with microwave imagery, we have access to the detailed history of sea ice levels around the arctic and antarctic since the end-70s.

SVD analysis of this data shows a decreasing trend in all 13 analyzed arctic regions compared to the 1981 - 2000 mean, with the strongest decrease in the East Siberian Sea followed by the Kara Sea and Barents Sea.

With these anomaly calculations and visualizations like heatmaps and arctic maps we hope to (1) help policymakers to develop long-term strategies for protection of coastal communities and accelerate international efforts to reduce greenhouse gas emissions, (2) help scientists anticipate changes in marine ecosystems and assist in global climate modeling, and (3) inform industries operating in the arctic about potential new shipping routes & resource exploration.

Introduction

The primary objective of this paper is the analysis of historical sea ice area data to identify significant trends and patterns of decrease in arctic regions over time. We want to provide insights which help with the prediction of future diminishment, aiding in climate modeling, policy development, marine ecosystem forecasting, and industrial planning. Challenges lie in extracting meaningful patterns from the data amidst natural variability.

This paper uses the Sea Ice Index data of the National Snow and Ice Data Center (NSIDC) from 1978 until today to calculate anomalies in sea ice (NSIDC, n.d. -c). The dataset is divided into 14 arctic regions. We process the raw data into a 14 by 564 space-time-data matrix, with 14 being the arctic regions, and 564 being the consecutive months from 1978 until today.

$$X \in \mathbb{R}^{n imes m}$$

is the data matrix representing sea ice anomalies, where n=14 is the number of arctic regions, m=564 is the number of consecutive months from 1978 to the present, and X_{ij} represents the sea ice area anomaly for region i and month j. We calculate the anomalies by subtracting a baseline mean from the observed data, with a baseline period of 1981 - 2000.

To get an overview, which region decreases more or less compared to others, or in other words, which region contributes how much to the pattern of variability, we can use singular value decomposition (SVD) on the anomaly matrix *X* to extract these patterns and trends. SVD analysis decomposes space-time data into spatial patterns, temporal patterns, and singular values. We are interested in the spatial patterns, especially the first one, which captures the trend that contributes the most to variability.

Singular Value Decomposition can be defined as

$$A = U\Sigma V^T$$

where A is the original space-time data matrix, U are the spatial patterns, Σ are the singular values, also called energy levels, and V^T is the transpose of the temporal patterns V. Our problem can be defined as finding the spatial patterns and evaluating the first column, i.e. the first spatial eigenvector, with the help of visualization techniques like heatmaps and world maps. These visualizations will help illustrate the extent of decline in Arctic ice caps across different regions.

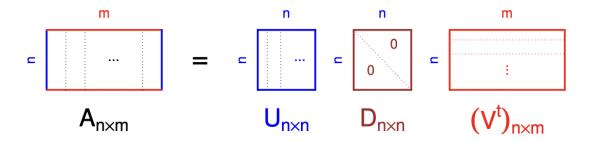


Figure 1: Diagram of SVD analysis (Shen, 2024)

Data and method

The NSIDC is part of CIRES at the University of Colorado Boulder and monitors the cryosphere - the global snow, ice and frozen regions system (NSIDC, n.d. -a). With the help of NASA's Defense Meteorological Satellite Program (DMSP) F-18 Special Sensor Microwave Imager/Sounder (SSMIS) (NSIDC, n.d. -b), the NSIDC created the Sea Ice Index data set (NSIDC, n.d. -c).

In this report, we focus on the north sea ice index regional monthly data, specifically the "area" subsets. It consists of monthly averaged area data in square kilometers from 1978 to present, broken up into 14 arctic regional seas shown in Figure 2 (Windnagel, 2017). We focus on the "area" sheets because they show the ocean's total area covered with ice, instead of sea ice "extent" data, where only at least 15% of the ocean needs to be covered with ice (NSIDC, n.d. -b).



Figure 2: Arctic Regions (Windnagel, 2017)

Before analyzing a space-time-matrix with SVD analysis, we need to preprocess the raw Sea Ice Index Spreadsheet in two steps: anomaly calculation and formatting the data in the correct dimensions. The raw spreadsheet retrieved from NSIDC consists of 29 sheets, 14 sea ice area, 14 sea ice extent, and one documentation sheet. Each of the 14 area sheets stores a 47 years by 12 months table of sea ice area values in square kilometers.

Because year-to-year variability is significant for sea ice, we chose to subtract each value by a long-term baseline mean from 1981 - 2000:

$$\overline{S}_{i,m} = rac{1}{Y_b} \sum_{y=y_{start}}^{y_{end}} S_{i,m,y}$$

represents the mean for region i and month m, with $S_{i,m,y}$ being the observed data at region i during month m and year y, $y_{start}=1981-1978+1=4$ being the start year index, $y_{end}=2000-1978+1=23$ being the end year index, $Y_b=y_{end}-y_{start}+1=20$ being the number of years in the baseline period, and $\overline{S}_{i,m}$ being the baseline mean sea ice area for region i during month m. The anomaly $X_{i,m,y}$ for each region, month and year is calculated by $X_{i,m,y}=S_{i,m,y}-\overline{S}_{i,m}$ for all years y=1,2,...,Y, where Y=47, and the final anomaly matrix is $X_i=\in\mathbb{R}^{12\times47}$ for each region i. In R, we can easily achieve that by calculating the column means for each sheet, and subtracting them from the dataset:

```
# Replace values with values minus each colmeans
means <- colMeans(region_area[4:23, ], na.rm = TRUE)
region_area_anomalies <- sweep(region_area, 2, means, "-")</pre>
```

where the array to form the mean from are all area rows from y_{start} until y_{end} . The 2 argument in the sweep() function indicates that the operation is being applied across columns, the "-" indicates a subtraction operation.

After calculating the anomalies, we reshape the data to prepare it for SVD analysis. For each region i, we flatten the matrix X_i into

$$\mathbf{x}_i = \left[\,X_{i,1,1},\; X_{i,2,1},\; \ldots,\; X_{i,12,1},\; X_{i,1,2},\; X_{i,2,2},\; \ldots,\; X_{i,12,Y}\,
ight]_{\,.}$$

a $1 \times T$ matrix, where $T = 12 \times Y = 12 \times 47 = 564$. By stacking all flattened anomaly vectors \mathbf{x}_i for all regions i, we form the final data matrix X:

$$X = egin{bmatrix} \mathbf{x}_1 \ \mathbf{x}_2 \ dots \ \mathbf{x}_{14} \end{bmatrix} \in \mathbb{R}^{14 imes 564}$$

In R, the flattening step is done with the reshape() function. We artificially create a unique id and time column in the cbind() function, which will be used by the reshape() function to create unique columns. We also specify a "-" as a separator, so each column will have names of MONTH-YYYY. In the end, we remove the helper row for the column names to receive a single row, 1 by 564 time matrix:

```
# Flatten the data into one row
region_space_time_df <- reshape(
   cbind(
      region_area_anomalies,
      id = 1,
      time = rownames(region_area_anomalies)
),
   direction = "wide",
   sep = "-",
   new.row.names = 0
)[-1]</pre>
```

An excerpt of the resulting matrix looks like this - anomaly values for every month:

```
August-1979 September-1979 October-1979 November-1979 December-1979 January-1980 138271.1 211882.9 126986.6 -2050.156 3741.244 -2248.052
```

Then, we append this single row matrix of each region to a list, and bind them together to form a 14 by 564 space-time-data matrix:

```
for(region in regions) {
      # All flattening stepts
      # Add region row to list
      all_regions[[region_name]] <- region_space_time_df</pre>
 }
 all_regions_dataframe <- do.call(rbind, all_regions)</pre>
> print(dim(area_space_time_df))
[1] 14 564
> print(area_space_time_df[, c("January-1980", "August-2023")])
                    January-1980 August-2023
Baffin
                     -170135.316 -22065.144
Barents
                      100882.761 -29245.142
Beaufort
                       -5230.364 -388040.521
Bering
                       62756.835
                                    1534.153
CanadianArchipelago
                       -8095.271 -157822.103
Central Arctic
                     -548663.485 167422.468
Chukchi
                       -3963.411 -230159.130
East Siberian
                      -17528.031 -551485.567
Greenland
                      -38140.543 -37812.968
Hudson
                      -29963.638 -20391.661
Kara
                        2912.393 -142113.517
Laptev
                       -2248.052
                                  -88850.224
0khotsk
                      238576.435
                                    5700.700
St Lawrence
                      -33828.452
                                       0.000
```

Regional sea ice area anomalies in September over time [sqkm]

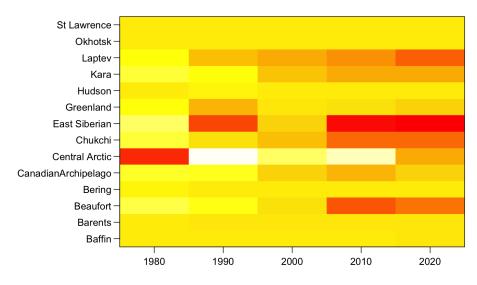


Figure 3: Sea ice area anomalies in September over time (darker = less area)

Based on these preprocessing steps, we can easily decompose the anomaly matrix with SVD intro three matrices:

$$X = U \Sigma V^T$$
 , where $U = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{14}] \in \mathbb{R}^{14 imes 14}$ $\Sigma = \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_{14}) \in \mathbb{R}^{14 imes 14}$ $V^ op \in \mathbb{R}^{14 imes 564}$, with rows $\mathbf{v}_k^ op \in \mathbb{R}^{1 imes 564}$

The left singular vectors U are equivalent to Empirical Orthogonal Functions (EOFs) in climate science. Each EOF represents a spatial pattern of sea ice variability. We found that the first four EOFs represent > 90% of the total variance:

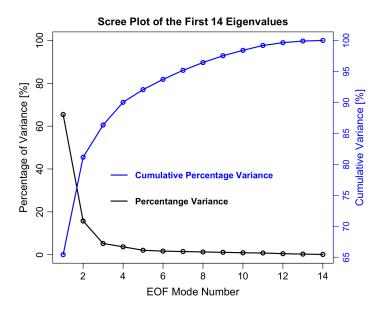


Figure 4: Variance of EOFs

Results

We plot the first EOF on an arctic map to get an understanding about which regions contribute the most to the overall trend. The result was not what we would have expected: the Central Arctic was dancing out of the line. All results were small negative values, while the central arctic had an almost 0.9 positive contribution:

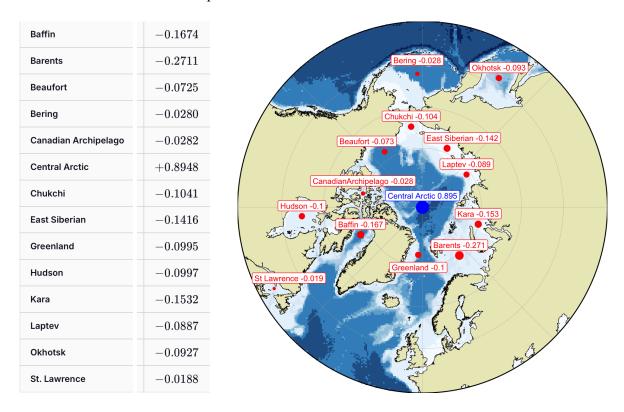


Figure 5: Unexpected EOF results by region on arctic map

This can be explained with the Arctic pole hole, which is part of the Central Arctic region. The satellite responsible for the Sea Ice Index did not collect the full arctic region, but left out a circular area above the pole due to orbit inclination. Over the years, this circular absence area decreased in size thanks to better satellite technology (Windnagel, 2015). However, this means that there is a discontinuity in the Central Arctic area and the net sea ice area "increases" in the Sea Ice Index data (smaller absence area of data = higher area values). Since the goal of this paper is to evaluate the melting ice caps for coastal regions and explore industrial operation

opportunities, and the Central Arctic does not touch coastal regions, we can remove the central arctic area from the SVD calculation to account for the arctic pole hole.

Without the Central Arctic region, we see promising results. After taking the magnitude of the first EOF values and rounding to three digits, we see that the East Siberian Sea contributes the most to the overall trend, followed by Kara Sea and Barents Sea:

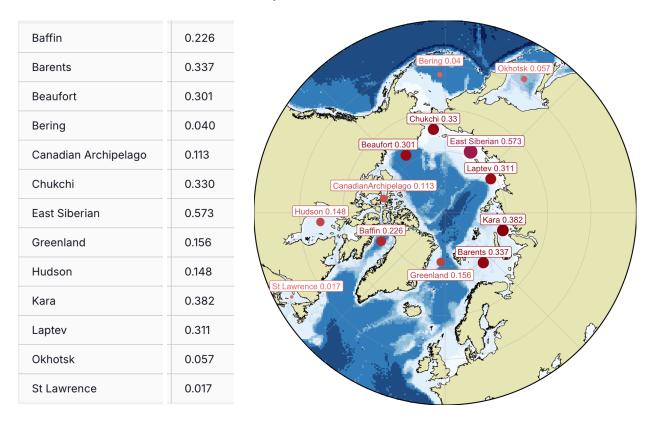


Figure 6: EOF1 by region on arctic map (darker = more trend impact)

The arctic regions of Chukchi, East Siberia, Laptev, Kara, Barents and Beaufort have the largest impact on the overall arctic sea ice system. Figure 6 shows which regions are impacted most (darker red) by the overall trend in sea ice decline. These results can be used as a basis for policymakers and scientists to prepare for more drastic changes to the ecosystem of those regions. At the same time, these most impacted regions are also the most interesting for new shipping routes from countries close to the North Pacific Ocean like China, Taiwan and Tokyo to

countries close to the North Atlantic Ocean like Canada, the Eastern US and Northern Europe. These trends can be used to further forecast the continuing decline in arctic sea ice in critical regions.

Conclusion

In this report, we analyzed sea ice area data from the Sea Ice Index published by the National Sea and Ice Data Center (NSIDC) to identify significant trends of decreasing arctic ice by region. The primary objective was to provide insights that help with climate modeling, policy development, marine ecosystem forecasting, and industrial planning.

We used Singular Vector Decomposition to decompose the space-time data matrix of sea ice area in square kilometers by region over time, and extracted the dominant spatial patterns of 13 arctic regions. After excluding the central arctic region because of data discontinuity due to the arctic pole hole, our analysis revealed a decreasing trend in all analyzed regions of the Arctic compared to the 1981-2000 mean, with the strongest decrease observed in the East Siberian Sea, followed by the Kara Sea, Barents Sea, and Chukchi Sea.

Alternative methods for this insight include time series analysis with principal components or exponential smoothing, and machine learning techniques like random forests. Advantages of the chosen method of SVD are the consideration of all regions simultaneously, capturing trends that are common across all regions, and extracting the most significant patterns amidst natural variability, reducing potential noise. It is also flexible in the sense that it can capture linear and non-linear trends in data. On the other hand, a big disadvantage lies in outlier sensitivity, which we experience first-hand with the discontinuity of the central arctic region data and its result in the first, not useful SVD which had to be adjusted manually.

References

- NSIDC. (n.d. -b). *About the data*. NSIDC. Retrieved December 05, 2024, from https://nsidc.org/sea-ice-today/about-data
- NSIDC. (n.d. -c). *Index of /NOAA/G02135/seaice_analysis/*. Retrieved December 05, 2024, from https://noaadata.apps.nsidc.org/NOAA/G02135/seaice_analysis/
- NSIDC. (n.d. -a). *Overview*. Our Research. Retrieved December 05, 2024, from https://nsidc.org/our-research
- Shen, S. S.P. (2024, August). *Interactive Linear Algebra with R and Python*.
- Windnagel, A. (2015, March). *Arctic Pole Hole*. Sea Ice Index, Version 3.0. Retrieved December 07, 2024, from
 - https://nsidc.org/sites/default/files/g02135-v003-userguide 1 1.pdf#page=42
- Windnagel, A. (2017, June). *Sea Ice Index Ice Analysis Spreadsheets*. Retrieved December 05, 2024, from
 - https://nsidc.org/sites/nsidc.org/files/files/data/noaa/g02135/Sea-Ice-Analysis-Spreadshee ts-Overview.pdf

Appendix: R code

```
# Load necessary libraries
install.packages("readxl")
install.packages("dplyr")
install.packages("tidyr")
install.packages("ggOceanMaps")
install.packages("ggspatial")
library(readxl)
library(dplyr)
library(tidyr)
library(ggOceanMaps)
library(ggspatial)
library(grDevices)
# Define the path to the Excel file
file_path <- "data/N_Sea_Ice_Index_Regional_Monthly_Data_G02135_v3.0.xlsx"</pre>
# Get all sheet names
sheet_names <- excel_sheets(file_path)</pre>
# Separate sheets into area and extent
area_sheets <- sheet_names[grepl("area", sheet_names, ignore.case = TRUE)]
months <- c("January", "February", "March", "April", "May", "June", "July", "August",
    "September", "October", "November", "December")
process_sheets <- function(sheets, avg_rows) {</pre>
    all_regions <- list()
    for (sheet in sheets) {
        xls_data <- read_excel(file_path, sheet = sheet)</pre>
        region_data_df <- data.frame(xls_data)</pre>
        # Get years
        years <- region_data_df[3:49, 1]
        # Remove unimportant / empty rows and columns
        region\_area <- data.frame(region\_data\_df[3:49, seq(2, ncol(region\_data\_df), by = 2)])
        # Cell formatting (numeric, no NA, 2 decimal points)
        region_area <- as.data.frame(lapply(region_area, as.numeric))</pre>
        region_area <- round(region_area, 2)</pre>
        # Set row and column names
        rownames(region_area) <- years</pre>
        colnames(region_area) <- months</pre>
        # Replace values with values minus each colmeans
        means <- colMeans(region_area[avg_rows, ], na.rm = TRUE)</pre>
        region_area_anomalies <- sweep(region_area, 2, means, "-")</pre>
        # Remove NA values
        region_area_anomalies[is.na(region_area_anomalies)] <- 0</pre>
        # Flatten the data into one row
        region_space_time_df <- reshape(</pre>
            cbind(
                 region_area_anomalies,
```

```
id = 1,
                time = rownames(region_area_anomalies)
            ),
            direction = "wide",
            sep = "-",
            new.row.names = 0
        )[-1]
        # Format region name
        region_name <- sub("-(Area|Extent).*", "", sheet)</pre>
        region_name <- gsub("-", " ", region_name)</pre>
        # Exclude "Central Arctic" region
        if (region_name != "Central Arctic") {
            # Add region row to list
            all_regions[[region_name]] <- region_space_time_df
        }
    }
    all_regions_dataframe <- do.call(rbind, all_regions)</pre>
    return(all_regions_dataframe)
}
# Set median time frame in years.
# Df starts at 1978 (1 = 1978, e.g. 4:23 = 1981-2000)
rows_for_mean_calc <- 4:23</pre>
# Process and merge all area sheets
area_space_time_df <- process_sheets(area_sheets, rows_for_mean_calc)</pre>
# Display an excerpt of the final data frames
print(dim(area_space_time_df))
print(area_space_time_df[, c("January-1980", "August-2023")])
# Export the data frame to CSV
# write.csv(area_space_time_df, "data/sea-ice-regional-area-anomalies.csv")
# Visualize data -----
# Heat map
generate_heat_map <- function(data, title, month = "August") {</pre>
    month_data <- data[, grepl(month, colnames(data))]</pre>
    month_data <- month_data[, seq(3, ncol(month_data), by = 10)]</pre>
    region_names <- rownames(month_data)</pre>
    time_stamps <- gsub(".*-(\d{4})", "\d{1}", colnames(month_data))
    par(mar = c(4, 10, 4, 4) + 0.1)
    image(t(month_data),
        col = heat.colors(64),
        axes = FALSE
    axis(1, at = seq(0, 1, length.out = 5), labels = time_stamps)
    axis(2, at = seq(0, 1, length.out = 13), labels = region_names, las = 2)
    title(main = title)
    box()
}
generate_heat_map(
    area_space_time_df,
    "Regional sea ice area anomalies in September over time [sqkm]",
    "September"
```

```
)
# SVD analysis -----
svd_area <- svd(area_space_time_df)</pre>
U_area <- svd_area$u
D_area <- svd_area$d
V_area <- svd_area$v
# Display the first EOF for each matrix
print(U_area[, 1])
# Scree plot of variance with cumulative percentage ---
# Set needed values
SVDd <- svd area$d
lam <- SVDd^2
K <- 13
lamK <- lam[1:K]</pre>
percentD <- 100 * (lam) / sum(lam)</pre>
cumpercentD <- cumsum(percentD)</pre>
modeK <- 1:length(SVDd)</pre>
# Plot variance
par(mar = c(4, 4, 2, 4), mgp = c(2.2, 0.7, 0))
plot(modeK[1:K], percentD[1:K],
    type = "o",
   ylim = c(0, 100),
   ylab = "Percentage of Variance [%]",
   xlab = "EOF Mode Number",
    cex.lab = 1.2, cex.axis = 1.1, lwd = 2,
   main = paste("Scree Plot of the First", K, "Eigenvalues")
legend(3, 30,
    col = c("black"), lty = 1, lwd = 2.0,
    legend = c("Percentange Variance"), bty = "n",
    text.font = 2, cex = 1.0, text.col = "black"
)
# Plot cumulative variance
par(new = TRUE)
plot(modeK[1:K], cumpercentD,
    ylim = c(min(cumpercentD), 100), type = "o",
    col = "blue", lwd = 2, axes = FALSE,
   xlab = "", ylab = ""
legend(3, 80,
    col = c("blue"), lty = 1, lwd = 2.0,
    legend = c("Cumulative Percentage Variance"), bty = "n",
    text.font = 2, cex = 1.0, text.col = "blue"
axis(4, col = "blue", col.axis = "blue", mgp = c(3, 0.7, 0))
mtext("Cumulative Variance [%]",
    col = "blue".
    cex = 1.2, side = 4, line = 2
)
# ----- Plot EOF mode 1 on Map -----
# Run those lines at top of file:
# install.packages("ggOceanMaps")
# install.packages("ggspatial")
```

```
# library(gg0ceanMaps)
# library(ggspatial)
# Coords (by Google) & Order for all regions
# baffin = 72.809094, -66.175709
\# barents = 74.576217, 37.519449
\# beaufort = 72.961356, -145.563116
\# bering = 56.855056, -177.762477
\# canadian = 74.449727, -102.979623
# central arctic = 90, 0
# chukchi = 69.463991, -171.928849
# east siberian = 73.817176, 157.271205
\# greenland = 77.850046, -5.159874
# hudson = 59.710451, -85.768872
# kara = 75.151568, 73.153283
# laptec = 76.032029, 126.462445
# okhotsk = 52.978659, 149.296564
# st lawrence = 48.552296, -61.379056
# Get first EOF
EOF1_area <- U_area[, 1]
# Combine data for easy access
locations <- data.frame(</pre>
    name = rownames(area_space_time_df),
    lat = c(
        72.809094, 74.576217, 72.961356,
        56.855056, 74.449727,
        69.463991, 73.817176, 77.850046,
        59.710451, 75.151568, 76.032029,
        52.978659, 48.552296
    ),
    lon = c(
        -66.175709, 37.519449, -145.563116,
        -177.762477, -102.979623,
        -171.928849, 157.271205, -5.159874,
        -85.768872, 73.153283, 126.462445,
        149.296564, -61.379056
    EOF_value = EOF1_area,
    hjust = c(
        0.5, \ 0.5, \ 0.8, \ 0.5, \ 0.5, \ 0.5, \ 0.5, \ 0.5, \ 0.5, \ 0.5, \ 0.5, \ 0.5
    viust = c(
        -0.5, -0.5, -0.5, -0.8, -0.7, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5, -0.5
    )
)
# Define a color palette
color_palette <- colorRampPalette(c("maroon", "darkred", "firebrick", "indianred",</pre>
    "lightcoral"))
\# Normalize the EOF values to be between 0 and 1
normalized_values <- (locations$E0F_value - min(locations$E0F_value)) /
    (max(locations$EOF_value) - min(locations$EOF_value))
# Assign colors based on the normalized values
locations$cols <- color_palette(100)[as.numeric(cut(normalized_values, breaks = 100))]</pre>
# Plot Mode 1
```

```
basemap(
   limits = 45,
   shapefiles = "Arctic",
   bathymetry = TRUE, # optional
   bathy.style = "raster_binned_blues", # optional
   land.col = "#eeeac4", # optional
   legends = FALSE
) +
   geom_point(
       data = transform_coord(locations),
        aes(x = lon, y = lat, size = abs(locations$EOF_value)),
       color = locations$cols,
       show.legend = FALSE
   ) +
   geom_label(
        data = transform_coord(locations),
        aes(x = lon, y = lat),
       label = paste(locations$name, round(abs(locations$EOF_value), 3)),
       hjust = locations$hjust,
       vjust = locations$vjust,
       size = 3,
       color = locations$col
   )
dev.off()
```

Review Report

Name: Amanda Hamblin-Trué

Date: December 7, 2024

This review evaluates the report titled "Analysis of Melting Ice Caps through Microwave Satellite Data" by Zino Meyer. The report analyzes historical sea ice area data to identify trends and patterns of decrease in Arctic regions, providing insights for climate modeling, policy development, and industrial planning. The title page is well-structured, featuring the title, author's name, affiliation, contact information, and date.

The abstract concisely captures the essence of the report, highlighting the decrease in Arctic ice caps and its implications for policymakers, scientists, and industry stakeholders. The introduction effectively outlines the problem of decreasing Arctic ice caps and its global significance, using mathematical symbols and equations to provide a strong foundation for the research. The use of the Sea Ice Index data from the NSIDC is well-cited.

The data and method section provides a thorough explanation of the SVD model formulation and solution, with detailed descriptions of the datasets, diagrams, and formulas. The results section presents the SVD solution and interprets the findings effectively, using R code, tables, and figures to enhance clarity. The conclusion summarizes the work and discusses alternative methods like time series analysis and machine learning techniques, addressing the SVD method's strengths and limitations, particularly its sensitivity to outliers. The references section is adequate and professionally formatted.

Areas for improvement could include expanding the discussion on alternative methods and their applicability, ensuring the central Arctic region's data discontinuity is addressed more explicitly in the results section, and including case studies or real-world examples that could provide practical context and enhance the report's relevance. Additionally, a more detailed discussion on the limitations of the data and analysis, including potential biases or uncertainties, would be beneficial.

Overall, the report is well-organized, clear, and effective in communicating its findings. Its strengths include comprehensive data analysis, effective use of visualizations, and clear interpretation of results. The report provides valuable insights into the trends of Arctic ice cap melting, with significant implications for policymakers, scientists, and industries. The analysis is robust, and the findings are presented in a compelling manner, contributing to the understanding of climate change impacts.