CMAD: Advancing Understanding of Geospatial Clusters of Anomalous Melt Events in Sea Ice Extent

Maloy Kumar Devnath
maloyd1@umbc.edu
Institute for Harnessing Data and Model Revolution in the Polar Regions, University of Maryland, Baltimore County
Maryland, USA

Sudip Chakraborty
sudipc1@umbc.edu
Institute for Harnessing Data and Model Revolution in the Polar Regions, University of Maryland, Baltimore County
Maryland, USA

Vandana P. Janeja
vjaneja@umbc.edu
Institute for Harnessing Data and Model Revolution in the Polar Regions, University of Maryland, Baltimore County
Maryland, USA

ABSTRACT

Traditional statistical analyses do not reveal the spatial locations and the temporal occurrences of clusters of anomalous events that are responsible for a significant loss of sea ice extent. To address this problem, we present a novel method named Convolution Matrix Anomaly Detection (CMAD). The onset and progression of clusters of anomalous melting events over the Antarctic Sea ice are studied as loss in sea ice extent, which are essentially negative values, where the traditional convolutional operation of the Convolutional Neural Network (CNN) approach is ineffective. CMAD is based on an inverse max pooling concept in the convolutional operation of CNN to address this gap. CMAD is developed to offer a solution without using a neural network, and unlike a full CNN, it doesn't require any training or testing processes. Satellite images are utilized to establish the loss in the Antarctic region. Our analysis shows that anomalous melting patterns have significantly affected the Weddell and the Ross Sea regions more than any other regions of the Antarctic, consistent with the largest disappearance in sea ice extent over these two regions. These findings bolster the applicability of the inverse max pooling based CMAD in detecting the spatiotemporal evolution of clusters of anomalous melting events over the Antarctic region. The anomalous melting process was first noticed along the outer boundary of the sea ice extent in early September 2022 and gradually engulfed the entire sea ice region by February 2023 - in tandem with the scientific literature. These findings indicate that there is a necessity to delve deeper into the role of the anomalous melting process on sea ice retreat for a better understanding of the sea ice retreat process. The nature of the problem is to detect clusters of contiguous grids of anomalous melting events rather than detecting discrete grid points. CMAD's ability to perform both data clustering and anomaly detection via the pooling operations allows for a more comprehensive analysis of sea ice melt patterns, facilitating the pinpointing of areas with potentially significant melt events. This method has the potential to apply in other fields of study where anomalous events are detected in clusters. The inverse max pooling concept has successfully detected clusters of anomalous events in sea ice and demonstrated the capability to detect anomalies with 87% accuracy in benchmark data. In contrast to well-established conventional methods such as DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, and Gaussian Mixtures, when applied to dynamic multidimensional data, $\text{CMAD}_{Benchmark}$ (which is a variation of CMAD) exhibits superior capabilities in detecting extreme events. The comparative analysis reveals that $\text{CMAD}_{Benchmark}$ outperforms these traditional approaches, showcasing its heightened sensitivity and efficacy in capturing significant variations within evolving multidimensional datasets over time. This heightens the detection accuracy positions of CMAD as a valuable tool for discerning extreme events in the context of dynamic and changing multidimensional data.

CCS CONCEPTS

• **Computing methodologies** → **Cluster analysis**; **Anomaly detection**.

KEYWORDS

CNN, Antarctic sea ice, Inverse max pooling, Anomaly detection, Clustering

ACM Reference Format:
Maloy Kumar Devnath, Sudip Chakraborty, and Vandana P. Janeja. 2024. CMAD: Advancing Understanding of Geospatial Clusters of Anomalous Melt Events in Sea Ice Extent. In *The 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL '24), October 29-November 1, 2024, Atlanta, GA, USA.* ACM, Atlanta, Georgia, USA, 11 pages. https://doi.org/10.1145/3678717.3691280

1 INTRODUCTION

With about 60% of the global population living within 100 km of a sea or oceanic coast [4], the recent retreat of the Antarctic Sea ice Extent (SIE) and the accelerated melting of the land ice sheets in recent years has raised concern among the scientific community. The Antarctic SIE reached a new record-smashing low of 1.965 million km^2 in February 2023, which is ~ 32% below climatological values [1]. Although Sea ice melt does not directly contribute to the global sea level rise, it acts as a protective blanket to the ice sheets on the land [42, 43, 48]. In this regard, it is important to note that the ice sheet melt has also accelerated in recent times as the Antarctic has lost approximately 1200 Gigatons of ice between February 2017 and February 2022 compared to 147 Gigatons of ice every year since 2002 [39]. The causes behind the retreat of sea ice that was growing till austral winter in 2015 are not well understood. There is a critical need to investigate this understudied

SIGSPATIAL '24, October 29-November 1, 2024, Atlanta, GA, USA

ACM ISBN 979-8-4007-1107-7/24/10
https://doi.org/10.1145/3678717.3691280

phenomenon since the sea ice acts as a buffer from warm currents, wind, and waves to the ice sheets and protects the loss of land ice sheets which leads to sea level rise.

To understand the recent sea ice melt, it is important to detect the nature of the melting pattern in the Antarctic. The current knowledge gap is primarily due the poor representation of the sea ice melting in climate models that pose a large uncertainty in the sea ice disappearance timing [7, 17] and ice thickness [5], a lack of analysis using multi-years of observational datasets [59] and, limited computational resources to analyze the big data. Prior studies have focused on predicting Antarctic Sea ice [57] without considering the melting patterns (anomalous or steady state but gradual), investigating feature importance [11], and drawing causal relationships [35] using deep learning as well as climate models rather than attempting to understand the nature of the melting patterns. Time series-based analysis has been a popular method to study and analyze changes in SIE using a series of observations and other satellite image data sets; however, they are unable to attribute the SIE retreat to spatial locations effectively [6, 30, 49, 58]. Another study [30] primarily examines the spatiotemporal variations and influencing factors of sea ice extent in the Arctic and Antarctic, rather than pinpointing specific anomalous melting areas. [49] revolves around comparing the timing of sea ice retreat and advance in various regions of the Arctic and Antarctic. It delves into the resulting alterations in the duration of the summer ice-free season, emphasizing the significant correlation between anomalies in retreat and subsequent advance timing. However, it does not explicitly detect the specific regions impacted by these anomalies, only maintaining a broader focus on the overall patterns of sea ice behavior in the Arctic and Antarctic. As a result, physical understanding of the melting patterns [1] and prediction are still crude and uncertain [7, 18, 57].

Our study addresses this problem by using satellite images and introducing a novel approach CMAD based on a convolutional operation [8, 16, 32, 37, 44] to investigate whether the massive retreat in sea ice in the Antarctic region is because of periods of anomalous melting events. This is important to understand since anomalous events can cause a sudden and massive loss of sea ice in a very short time [20]. To address these challenges and simultaneously improve our ability to detect clusters of anomalous melting events, we employ the CMAD method to investigate anomalous sea ice melting in the Antarctic region using Sea Ice Concentration (satellite images) between September 2022 and April 2023 (Austral spring and summer). Please note that CMAD is a method that can be used to detect clusters of anomalous events without requiring any training or testing as in full CNN. However, the method can be used in the full CNN model to predict sea ice anomalies.

Our approach CMAD is based on detecting clusters of anomalous melting events on a spatial scale from each image and then detecting how the anomalous sea ice melting events onset along the outer boundary of the sea ice extent in early spring and propagate towards the coastline in summer over the Antarctic region of the Southern Ocean. We have also implemented the suitability of CMAD on a benchmark Global Aerosol Atmospheric Rivers Data (AAR) [10, 12] and observed superior performance compared to traditional well established methods such as DBSCAN [34, 45, 47, 56, 61], HDBSCAN [9, 28, 36, 38, 45], K-Means [25, 29, 45], Bisecting K-Means [3, 45, 53], BIRCH [33, 45, 46], Agglomerative Clustering [25, 45], OPTICS [13, 26, 45] and Gaussian Mixtures [45, 54, 55]. While these methods are from several years ago, they have held the test of time as robust traditional methods suitable for comparison with our proposed method. Despite their age, these methods remain relevant due to their robustness and continue to be employed in recent anomaly detection research [23, 25, 29, 56, 61]. Our approach excels in handling multidimensional data, effectively detecting extreme values in dense matrices and image data compared to these existing methods. However, CMAD might require further adaptation to achieve optimal performance on sparse data. Thus, our key contributions include:

- This study introduces a novel multi dimensional anomaly detection method called CMAD using convolutional operation from CNN. CMAD is designed to investigate and detect anomalous sea ice melting events in the Antarctic region using satellite images.
- CMAD allows for a spatial understanding of melting patterns over the Antarctic Sea ice. It detects how clusters of anomalous melting events initiate along the outer boundary of the sea ice extent in early September and progress towards April.
- The study addresses challenges related to negative values denoting loss in Sea Ice Extent (SIE). The introduction of inverse max pooling is specifically tailored to handle negative values during melting events, providing a more comprehensive approach to pooling. This approach can be used to detect clusters of anomalous events where negative changes have occurred.
- The study applies the $\text{CMAD}_{Benchmark}$ method to real-world benchmark AAR transportation data, showcasing its superior performance compared to traditional methods. $\text{CMAD}_{Benchmark}$ demonstrates effectiveness in handling multidimensional data and detecting extreme or anomalous values or events in dense multidimensional matrices and image data.

The rest of the paper is organized as follows. Section 2.1 outlines our methodology. In Section 3 we present experimental results, and Section 4 presents the significance and impact of our work. Finally, we present conclusions in and future work in Section 5.

2 METHODOLOGY

2.1 Overall Methodology

Our overall methodology is shown in Figure 1 and includes the following stepsWe first begin by pre-processing the satellite images. We start with a collection of RGB images. Each image has dimensions of height and width. We crop the images to focus on specific regions of interest. We convert the RGB images to grayscale to reduce computational complexity. While RGB images have three channels, grayscale images have only one, simplifying the processing. The resulting grayscale images retain the dimensions H_{original} and W_{original} of the original cropped RGB images.

While applying the convolutional operation of CNN, we encountered an interesting problem while analyzing the images. The loss or gain in SIE is denoted by negative or positive values, respectively. While max pooling was able to extract maximum gain in SIE from the pooling matrix, it was unable to extract the minimum values

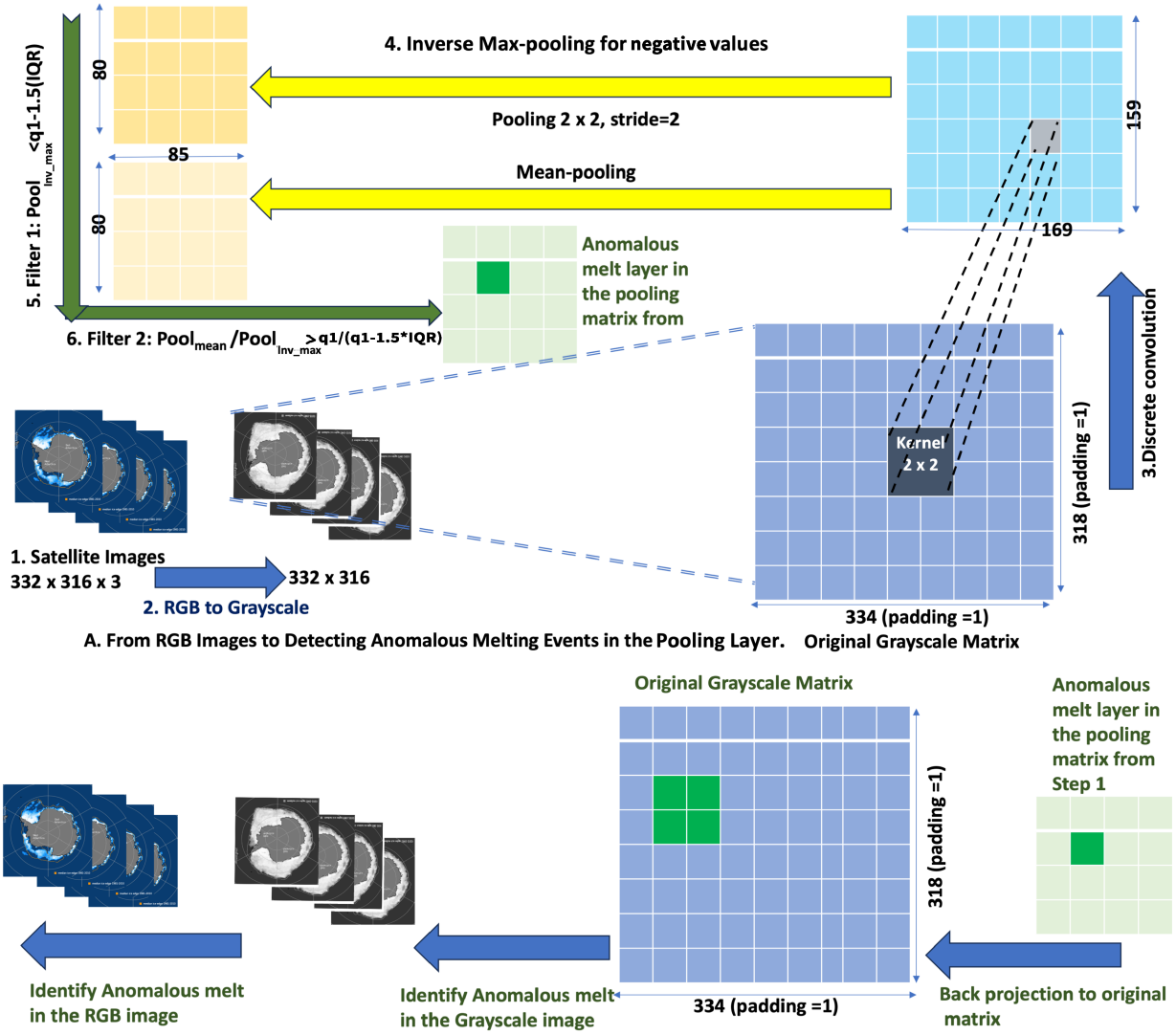

Figure 1: CMAD methodology: A concise overview of the daily satellite image analysis, involving kernel application, inverse max and mean pooling, and subsequent grid-based anomaly detection for spatial and temporal tracking of clusters of anomalous melting events.

during the melting events from images. Max pooling was detecting the lowest amount of SIE loss (δSIE values close to 0). The changes in SIE denote the δSIE on a day from the previous day. To address this limitation, we introduce an inverse max pooling concept shown in Equation 6 - similar to the max pooling concept, but to extract the minima of a set of negative values from the pooling layer.

We proceed by generating difference matrices, computed as the difference between consecutive grayscale images using Equation (1), resulting in a D matrix of dimensions $H_{\text{original}} \times W_{\text{original}}$. Employing an initial 2×2 kernel reduces the resolution of D. Subsequently, threshold calculation and application of inverse max and mean pooling [8] are executed using Equations (3), (4), (6), (7) and (5). Pooling extracted the main features within 2×2 matrix to detect anomalous

melting grids based on Equation (8). The anomaly detection method is then applied to each difference matrix of rr_matrix_List. The anomalies identified through this process are subsequently traced back to the original RGB images using backtracking techniques. The daily sea ice imagery for this study consists of images with a uniform spatial resolution of 25 km. Each image has dimensions of 332 x 316 pixels, providing a detailed representation of the sea ice cover. First, we apply a 2×2 kernel and then we use the inverse max and mean pooling on the convolutional layer. From the pooling layer, we use the mean pooling [8] and inverse max pooling to detect the grids with anomalous melt (Figure 1). In the next step, we project the anomalous grid found in the pooling layers back to the original image to detect locations of anomalous melt within each image. In this way, we use CMAD to detect the anomalous regions of melting on a spatial scale for each day and to detect the propagation of the clusters of anomalous melting events on a temporal scale for the entire period.

Algorithm 1 CMAD Algorithm

Input: List of RGB images $ImageList$
Output: Anomaly Regions
1: **for** each $image$ in $ImageList$ **do**
2: Read the image in RGB format and store it as $image\_rgb$
3: Convert $image\_rgb$ to $gray\_image$
4: Store $gray\_image$ in $gray\_image\_List$
5: **end for**
6: **for** each pair of consecutive gray images in $gray\_image\_list$ **do**
7: Calculate the difference matrix D as the subtraction of the first image from the second image using Equation (1)
8: Store D in $difference\_gray\_image\_list$
9: **end for**
10: **for** each difference matrix of consecutive gray images **do**
11: Reduce the resolution of the difference matrix D by padding zeros, following Equation (2)
12: Store the reduced resolution matrix in a list named rr_matrix_List
13: **end for**
14: Call the `Reshaping Array` function to reshape the rr_matrix_List
15: Calculate the LB_{IQR} and $Q1$ array using the $reshaped\_array$. $Q1[i, j]$ assigns first quartile of $reshaped\_array[:, i, j, :]$ and $LB_{IQR}[i, j]$ assigns $Q1[i, j] - 1.5\times$ interquartile range of $reshaped\_array[:, i, j, :]$
16: **for** each difference matrix of rr_matrix_List **do**
17: Call the `Optimal Solution` function to apply pooling and detect anomalies
18: Save anomalies in the original RGB images
19: **end for**

2.2 CMAD Algorithm

Our proposed CMAD method processes a list of grayscale images to detect anomaly regions. In the main image processing loop of CMAD, the algorithm 1 iterates through each RGB image in `ImageList`. For each image, it first reads the image in RGB format and stores it as `image_rgb`. Then, it converts `image_rgb` to `gray_image` and stores it to `gray_image_List` which is shown in lines between 1 to 5.

Next, the algorithm calculates the difference matrix D for each pair of consecutive gray images in `gray_image_list`. The difference matrix D is computed using Equation (1), representing the pixel-wise difference between the current and previous images. These difference matrices are stored in `difference_gray_image_list` which is depicted in lines between 6 to 9.

The difference matrix D between two consecutive input matrices at time t and $t + 1$ is computed as follows:

Algorithm 2 Optimal Solution Function

Input: Pooling dimensions ($height = 2$, $width = 2$), Stride height and width (2), Difference matrix, Threshold (LB_{IQR}, $Q1$)
Output: Anomalies
1: Apply inverse max and mean pooling to `matrix` with specified parameters ($height$, $width$, $stride\_height$, $stride\_width$) using $PoolingFunction$.
2: Detect anomalies for each grid of the matrix based on Equation (8).
3: Save anomaly regions in original RGB images.
4: **Return** Anomalies

Algorithm 3 Reshaping Array Function

Input: Reduced resolution matrix list (RRML)
Output: Reshaped array
1: Define pooling parameterspool_size = (2, 2), stride = 2
2: Calculate $output\_array$ dimensions using Equations (3), (4)
3: Initialization$output_array$ = ($RRML.shape[0], output\_height, output\_width, 4$)
4: **for** i in range(output_height) **do**
5: **for** j in range(output_width) **do**
6: Define pooling regionregion = $RRML[:, i \times$ stride$i \times$ stride + pool_size$[0], j \times$ stride$j \times$ stride + pool_size$[1]]$
7: Reshape region to (4,) array and store it in output array output_array$[:, i, j]$ = region.reshape$(-1, 4)$
8: **end for**
9: **end for**
10: **Return** $reshaped\_array$

Algorithm 4 Pooling Function

Input: Input matrix, Pooling parameters (pool_height , pool_width, stride_height, stride_width), Pooling type
Output: Pooled matrix
1: **if** pool_type is 'inverse max' **then**
2: Apply inverse max pooling to input_matrix using specified parameters and Equations (3), (4), (6), and (7).
3: **else if** pool_type is 'mean' **then**
4: Apply mean pooling to input_matrix using specified parameters and Equations (3), (4), and (5).
5: **end if**
6: **Return** Pooled matrix

$$D = Img(t+1) - Img(t) \tag{1}$$

Where:

- D is Difference matrix
- $Img(t)$ is gray image at time t
- $Img(t+1)$ is gray image at time t+1

Then, we apply a reduced resolution operation to the difference matrix D using Equation (2). The Reduced Resolution (RR) at position (i, j) is computed as:

$$\mathrm{RR}[i][j] = \sum_{m=0}^{1} \sum_{n=0}^{1} \mathrm{Input}[2i+m][2j+n] \tag{2}$$

Where

- $\mathrm{RR}[i][j]$ represents the value at position (i, j) in the reduced-resolution matrix.
- The summation terms $\sum_{m=0}^{1} \sum_{n=0}^{1}$ iterate over a 2x2 neighborhood of the input matrix.
- $\mathrm{Input}[2i+m][2j+n]$ represents the values in the input matrix, where each pixel in the reduced-resolution matrix corresponds to a 2x2 block of pixels in the input matrix.

This equation computes the reduced resolution by aggregating the values in the 2x2 neighborhood of the input matrix centered at position (i, j). We have done experimentation with various kernel dimensions, and a consistent choice of a 2×2 kernel in all pooling operations proves effective in capturing crucial features and nuances in our sea ice anomaly detection framework. In the given algorithmic steps, the reduced resolution matrices are generated and stored in a list named *rr_matrix_List*. The whole operation is shown in lines 10 to 13. Following the completion of this storage process, the `Reshaping Array` function is invoked to reshape the list of reduced resolution matrices which is shown in line 14. Subsequently, the LB_{IQR} and $Q1$ arrays are calculated in line 15.

Finally, the `Optimal Solution` function is called to apply to the pool and detect anomalies based on predefined criteria and thresholds shown in Equation (8). The algorithm returns the detected anomalies from the original RGB images. The whole operation is shown between lines 16 to 19.

In summary, this algorithm processes a sequence of grayscale images, computes the difference between consecutive images, and detects anomalies using a combination of mathematical & statistical operations and predefined functions. The time complexity of our proposed methodology is expressed as Θ(number of satellite images $\times$ row and column dimension of the satellite image). This complexity analysis provides insights into the algorithm's performance characteristics based on the following considerations. First, the term "number of satellite images" (p) signifies a linear relationship between the algorithm's time requirement and the number of images to be processed. As the dataset size grows, the time complexity scales accordingly. Second, when examining the row and column dimensions of the satellite image matrix ($m \times n$), the complexity is influenced by the product of the number of rows (m) and the number of columns (n). In cases where m and n are comparable, the complexity may exhibit quadratic behavior ($O(n^2)$). Alternatively, if m and n differ, the expression becomes $\Theta(p \cdot m \cdot n)$, potentially approximated as $O(p \cdot n^2)$ when m is significantly smaller than n. We can approximate it as $O(p \cdot m^2)$ when n is significantly smaller than m. In summary, our algorithm's time complexity is sensitive to both the number of satellite images and the dimensions of each image, showcasing scalable performance with respect to these parameters.

We further explain the anomaly detection process next for satellite image data.

2.3 Anomaly Detection

Detection of anomalous melt onset and progression is practically impossible to ascertain without satellite imagery. We apply inverse max pooling (p=2x2, stride =2 to avoid overlapping) on the convoluted layer to detect grids of anomalous melting events within the convoluted layer. In the next step, we compute mean pooling similarly.

Pooled Matrix Size: The size of the resulting pooled matrix P from an input matrix with dimensions H_{input} (height) and W_{input} (width) is determined as follows:

$$H_{\text{output}} = \left\lfloor \frac{H_{\text{input}} - H_{\text{pool}}}{H_{\text{stride}}} + 1 \right\rfloor \tag{3}$$

$$W_{\text{output}} = \left\lfloor \frac{W_{\text{input}} - W_{\text{pool}}}{W_{\text{stride}}} + 1 \right\rfloor \tag{4}$$

Pooling Operation: For Mean Pooling we use equation 5

$$P[i][j] = \frac{1}{H_{\text{pool}} \cdot W_{\text{pool}}} \sum_{x,y \in R(i,j)} I[x][y] \tag{5}$$

Where $R(i, j)$ represents the set of indices (x, y) within the pooling window corresponding to the position (i, j).

For inverse max pooling, we use the following equation 6:

$$\mathrm{I[i][j]} = \begin{cases} 0 & \text{if input\_matrix[i][j]} > 0 \\ \text{-input\_matrix[i][j]} & \text{otherwise} \end{cases} \tag{6}$$

Here, $input\_matrix[i][j]$ represents the elements of the input matrix. To obtain the inverse max pooling values from the I matrix, we employ the following equation with an inverse sign applied to the result of max pooling:

$$P[i][j] = - \max_{x,y \in R(i,j)} I[x][y] \tag{7}$$

In this Equation 7, $R(i, j)$ denotes the set of indices (x, y) within the pooling window corresponding to the position (i, j).

$$\mathrm{Anomaly}[i][j] = \begin{cases} 1 & \text{if inverse max\_pooled}[i][j] \\ & < \mathrm{LB}_{IQR}[i][j] \\ & \text{and } \frac{\text{mean\_pooled}[i][j]}{\text{inverse max\_pooled}[i][j]} > \\ & \frac{\mathrm{Q1[i][j]}}{\mathrm{LB}_{IQR}[i][j]} \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

Detecting clusters of anomalous melting events in Antarctic Sea ice involves using a key ratio, $\frac{Q1}{LB_{IQR}}$, in the anomaly detection process. The first quartile (Q1) is the median of the lower half of a dataset, representing the value below which 25% of the data falls. LB_{IQR} (Lower Bound using Interquartile Range) is calculated as Q1 minus 1.5 times the interquartile range (IQR) where

IQR is Q3 - Q1 [22]. This ratio, combined with a specific condition, is crucial for detecting unusual melting patterns by considering statistical quartile information and data dispersion. The ratio $\frac{\text{mean\_pooled}[i][j]}{\text{inverse max\_pooled}[i][j]}$ is essential for understanding the sea ice data's mean behavior relative to extreme inverse deviations. This comparison provides valuable insights into the overall characteristics of melting events. The condition $\frac{\text{mean\_pooled}[i][j]}{\text{inverse max\_pooled}[i][j]} > \frac{\text{Q1}[i][j]}{\text{LB}_{IQR}[i][j]}$ sets a criterion for detecting anomalous melting. It ensures that both extreme deviations (captured by inverse max-pooled values) and mean behavior align with the expected range defined by the first quartile and data dispersion. In summary, incorporating the $\frac{\text{mean\_pooled}}{\text{inverse max\_pooled}}$ ratio, alongside statistical quartile information and data dispersion conditions, enhances the anomaly detection process. This comprehensive approach improves the detection of clusters of anomalous melting events in Antarctic Sea ice, offering a more nuanced understanding of sea ice retreat dynamics.

2.4 CMAD$_{Benchmark}$ for Benchmark AAR Data

We have also used benchmark data, with labeled anomalies, from aerosol atmospheric river datasets [10, 12]. This dataset provides information about dust and other aerosol storms in the atmosphere, including their movement and concentration. The data have global coverage with a resolution of 6 hours, and we have focused on the horizontal movement of aerosol mass [10, 12]. The data sets have been derived from MERRA-2 aerosol and meteorological reanalysis datasets [12]. The dataset has been validated against measurements from Aerosol Robotic Network (AERONET) stations, showing good agreement [10, 12]. In adapting the CMAD methodology for AAR data, specific modifications have been introduced, particularly in the context of inverse max pooling. Given the focus on learning features from AAR data without the necessity for differencing between consecutive days (as reflected in Algorithm 1, lines 6 to 8 have been deemed unnecessary for this purpose). In the modified CMAD Algorithm 1, line 15 has undergone adjustments.

To accommodate all non-negative values characteristic of AAR data, *reshaped_array* has been used to calculate the UB_{IQR} and $Q3$ array. $Q3[i, j]$ assigns third quartile of $reshaped\_array[:, i, j, :]$ and $UB_{IQR}[i, j]$ assigns $Q3[i, j]$ + 1.5× interquartile range of $reshaped\_array[:, i, j, :]$ [19].

Given the non-negative nature of the values, the usage of inverse max pooling is no longer required. Instead, the Pooling Function in Algorithm 4 (updating lines 1 and 2) has incorporated a normal max pooling operation [8, 44]. Similarly, in the Optimal Solution Function in Algorithm 2, line 1 has been updated to utilize max pooling and mean pooling functions from the modified Pooling Function in Algorithm 4. This adaptation aligns the CMAD methodology with the specific characteristics and requirements of AAR data, optimizing its functionality for the given context. For detecting the anomaly, we have used the following updated Equation (9). To leverage the benefits of max pooling, we introduce a modified version of CMAD that integrates this technique. We refer to this version as CMAD$_{Benchmark}$ in the paper.

$$\text{Anomaly}[i][j] = \begin{cases} 1 & \text{if max\_pooled}[i][j] \\ & > \text{UB}_{IQR}[i][j] \\ & \text{and } \frac{\text{mean\_pooled}[i][j]}{\text{max\_pooled}[i][j]} > \\ & \frac{\text{Q3[i][j]}}{\text{UB}_{IQR}[i][j]} \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

3 EXPERIMENTS

Table 1 shows the datasets and images used in this study. We next discuss results from our Inverse max pooling based CMAD anomaly detection method. Then we have implemented CMAD$_{Benchmark}$ methods along with DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, and Gaussian Mixtures methods for comparison. We also outline the significance and impact of our work.

Table 1: Data used in this study.

Datasets	Satellite data Sources	Resolution/Pixels/Data Size
Sea Ice Extent Images (satellite images)	SMMR, SSM/I, SSMIS, NIMBUS Passive Microwave .png images	332 × 316 pixels (648 MB)
Sea Ice Concentration	SMMR, SSM/I, SSMIS, Nimbus-7 .nc numerical data	1 day
The movement and concentration of aerosols in the atmosphere (AAR)	Atmospheric rivers dataverse [10, 12]	Resolution (H=576, W=361, T= 6 hours, 24 years (1997-2020, 677 GB)
Integrated Aerosol Transport	MERRA-2 aerosol and meteorological variables [12]	Resolution of 1 hour (H=576, W=361, T= 6 hours, 24 years (1997-2020, 697 GB)

The computation of the CMAD for multi-dimensional data proved to be a computationally intensive effort due to the large number of values that required processing, totaling = 582051776 (332 * 316 * 5548) values. Moreover, the AAR data is 677 GB that has been generated from MERRA-2 aerosol and meteorological reanalysis data (697 GB). This dataset has a resolution of 576 × 361, provided every six hours of temporal resolution. We have processed 24 years of the MERRA-2 reanalysis data sets to estimate the efficiency of different algorithms methods in detecting anomalous AAR events (the benchmark data). Therefore, we have utilized a high-performance computing machine to enhance the efficiency of our program [21].

CMAD required 559.82 seconds for execution, as reported in [21], while CMAD$_{Benchmark}$ demonstrated a longer runtime of 16156.01 seconds, as documented in the same source [21].

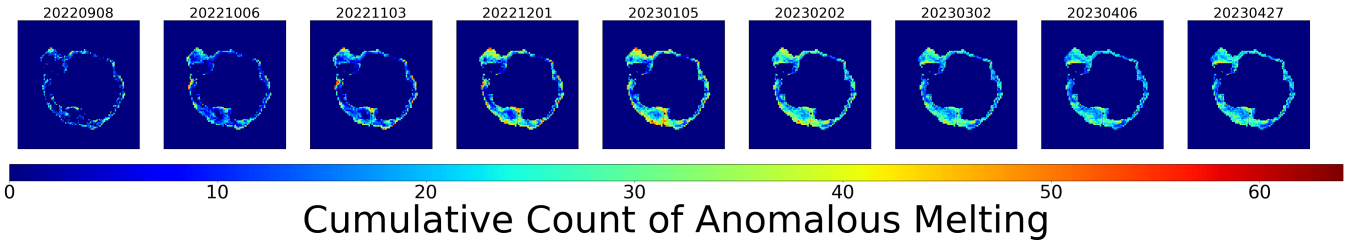

Figure 2: Clusters of Anomalous melting events detected from satellite images using CMAD. Contours of the anomalous melting onset and progress are calculated every 7 days starting from 09082022 to 04272023. Displayed here are sample plots capturing the onset and progression of clusters of anomalous melting events, specifically focusing on the beginning day of each month.

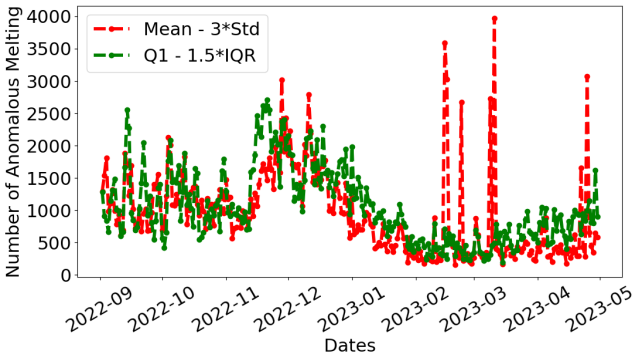

Figure 3: Line plots delineating the count of daily anomalous melting over the Antarctic region using the mean minus 3 times standard deviation based CMAD (depicted in red) and the Q1 minus 1.5 times IQR based CMAD (depicted in green) methods for the period of September 2022 to April 2023.

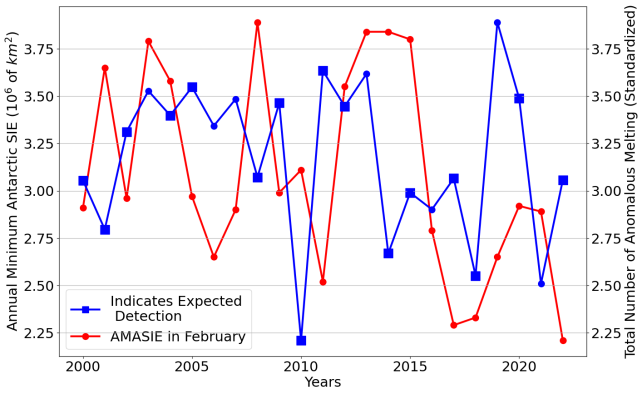

Figure 4: Visualization of the evolution of Annual Minimum Antarctic Sea Ice Extent (AMASIE) in February, spanning the years 2000 to 2022, represented through numerical satellite data [14] shown by the red line. The total number of anomalous melting events per year detected through CMAD analysis on satellite images is shown by the blue line. The blue box indicates the correct detection of CMAD.

3.1 CMAD Based Anomaly Detection in Satellite Images

Cumulative anomalies and regions in the Antarctic

Figure 2 shows the anomalous melting patterns detected by using CMAD. From Figure 2, it appears that the anomalous melting events onset near the outer boundary of the maximum sea ice extent after the end of the winter over the Antarctic region. Gradually with time, those anomalous melting events progress inwards and by the end of December 2022, they cover the entire sea ice region - in tandem with the massive retreat of sea ice in the last austral summer [50, 51].

Figure 2 shows the cumulative number of anomalous melting detected from the images using CMAD. It appears from our analysis that the Weddell and the Ross Sea regions are heavily affected by anomalous melting. Our methods also confirm that the anomalous melting and outliers are low over the other three regions – the Bell Amundsen, Indian Ocean, and Pacific Ocean regions [40]. Figure 2 shows that the Ross Sea and the Weddell regions experienced a heavy loss in SIE extent by February 2023 than any other regions. The outer boundary of SIE is prone to suffer from a higher number of anomalous melting, presumably because of the direct exposure to the warm oceanic water.

Generalizability of anomaly detection statistical tests

The predominant occurrence of clusters of anomalous melting events is observed in December 2022, as illustrated in Figure 3. Notably, considering that each grid corresponds to a 625 km^2 area, instances of daily anomalous melting surpassing 2500 during this period may result in a substantial loss of sea ice area. Specifically, this could translate to an impact of approximately $\sim$ 2500 * 625 = 1562500 km^2, depicting a significant influence on the Earth's environmental dynamics. To validate the effectiveness of CMAD while using any statistical measures, we tested CMAD with standard deviation based outliers in addition to the IQR based anomalies. Figure 3 demonstrates the results of two approachesone based on quantiles (green plot) and the other on 3σ deviation (red plot). While both methods exhibit similar outcomes, it is worth noting that a mean minus three times the standard deviation occasionally detects more anomalies in specific instances. In general, the two plots indicate that the maximum anomalies occur towards the end of December 2022 (austral summer).

Comparison with annual minimum Antarctic sea ice extent

We evaluated the reliability of the CMAD method by comparing its detection with the annual minimum Antarctic sea ice extent (AMASIE) [14]. In Figure 4, the plot in blue, marked with square boxes, signifies the accurate detection achieved through our CMAD method. Meanwhile, the plot in red, distinguished by circles, illustrates the annual minimum Antarctic sea ice extent for each respective year. We have standardized the total number of anomalous melting detected by our CMAD method in the range of minimum Antarctic sea ice extent before incorporating them into the plot. Figure 4 shows that when the number of anomalous melting increases in a particular year, the minimum sea ice extent in that year decreases and vice versa. This is because an increase in anomalous melting may cause a higher disappearance in sea ice. For example, in 2010, when the minimum sea ice extent was higher than in 2011, CMAD correctly detected a lower number of anomalies which is shown in Figure 4. On the contrary, in 2011, when the minimum sea ice extent was lower than in 2010, CMAD detected a higher number of anomalies which is expected. Out of 23 years, CMAD accurately detects this phenomenon for 16 years. This is a noticeable finding since many years after 2010 agree with this sea ice minima and higher annual anomalies relationship. In this regard, it is important to note that this study doesn't consider the steady state melting events, which might also be important and may have affected other years (7/23 years). Thus, it can be inferred that the lower minimum Antarctic sea ice extent is associated with higher anomalous melting, especially in recent years (including 2020-2022) [14].

3.2 CMAD$_{Benchmark}$ in Benchmark AAR Data: A Comparative Evaluation

In evaluating the performance of anomaly detection methods on dynamically changing data like AAR, CMAD$_{Benchmark}$ emerges as a robust and effective solution. The comparison of DBSCAN [34, 45, 47, 56, 61], CMAD$_{Benchmark}$, and K-Means [25, 29, 45] methods reveals notable advantages of CMAD$_{Benchmark}$ in terms of precision, f1-score, and overall accuracy [24, 45]. We know that DBSCAN can provide meaningful results and good performance for clustering spatial data [47]. Specifically, CMAD$_{Benchmark}$ demonstrates superior precision (49%), f1-score (57%), and accuracy (87%) compared to DBSCAN [34, 45, 47, 56, 61] and K-Means [25, 29, 45] indicating its ability to accurately detect extreme events and changes in dynamic changing data which is shown in Table 2. Here, we have considered mean values of precision, recall, f1-score, and overall accuracy to address the data gaps in specific locations. Along with DBSCAN, and K-Means, CMAD$_{Benchmark}$ outperforms HDBSCAN, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, and Gaussian Mixtures across all metrics (precision, f1-score, and accuracy), showcasing its effectiveness in detecting anomalies in dynamically changing data over time. While some methods (HDBSCAN and Gaussian Mixtures) exhibit higher recall by classifying every grid point as an extreme event, this assumption is erroneous and leads to a high number of false positives, a phenomenon corroborated by the analysis of the Area Under the Receiver Operating Characteristic Curve (ROC) curve shown in Figure 5. Here, ROC demonstrates its exceptional ability to distinguish between true anomalies and normal data points compared to CMAD$_{Benchmark}$, DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering (Agg. Clustering), OPTICS, and Gaussian Mixtures.

This was particularly evident in testing scenarios involving dust AAR data in June 2020, where CMAD$_{Benchmark}$ excelled in pinpointing locations experiencing dynamic changes with extreme values (for example, over the Sahara Desert and the regions surrounding it). The method has exhibited a nuanced understanding of spatial variations, successfully detecting aerosol atmospheric rivers and other changing features. The Receiver Operating Characteristic (ROC) [31] curve further substantiates the efficacy of CMAD$_{Benchmark}$, highlighting its competence in detecting extreme values when compared to other existing methods depicted in Figure 5. We have computed the average of True Positive Rate and False Positive Rate to address data gaps in specific locations. These findings underscore the applicability of CMAD$_{Benchmark}$ and position it as a valuable tool for anomaly detection in multidimensional datasets.

Our proposed CMAD$_{Benchmark}$ method aims to address some of the key limitations of existing clustering approaches. Unlike methods like K-Means [45], Affinity Propagation [45, 60], and Gaussian Mixtures [45] that struggle with large datasets, CMAD$_{Benchmark}$ is designed to be highly scalable, capable of handling very large $n_{samples}$ efficiently. It does not make strong assumptions about the underlying geometry of the data, allowing it to capture non-flat [45], complex cluster shapes akin to methods like DBSCAN [45] and Mean-Shift [45]. Furthermore, CMAD$_{Benchmark}$ does not impose constraints on the number of clusters or their sizes, automatically detecting an appropriate number of clusters of varying densities and sizes. This flexibility makes it suitable for a wide range of clustering problems without prior knowledge of the cluster structure. Importantly, CMAD$_{Benchmark}$ takes an inductive approach, enabling it to generalize to new data points and assign them to existing or new clusters as needed, unlike transductive methods like Spectral Clustering [15, 41, 45]. The Spectral Clustering algorithm is not well-suited for detecting aerosol atmospheric river events, as it relies on constructing a connected similarity graph representation of the data. This approach is ineffective for capturing the spatial and geometric characteristics inherent to aerosol atmospheric river events. Additionally, the Affinity Propagation [52] and Mean Shift algorithms [27] have failed to converge within the specified number of iterations, suggesting that the obtained cluster centers and labels may be degenerate or suboptimal for the given dataset [45]. By combining scalability, flexibility in geometry and cluster shapes, automatic determination of the number of clusters, and an inductive nature, CMAD$_{Benchmark}$ aims to provide a powerful and versatile clustering solution applicable across diverse domains and datasets.

4 SIGNIFICANCE AND IMPACT OF OUR WORK

This study advances our knowledge about the presence of clusters of anomalous melting events over the Antarctic Sea ice region. While the Arctic sea ice has been retreating for decades, the Antarctic sea ice was able to maintain its extent for a long time till 2015. Recently the sea ice retreat has tremendously accelerated in the

Table 2: Comparative evaluation of anomaly detection methods (CMAD$_{Benchmark}$, DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, and Gaussian Mixtures) on dynamically changing data using precision, recall, f1-score, and accuracy. CMAD$_{Benchmark}$ exhibits superior performance in accurately detecting extreme changes over time, particularly demonstrated in its application to dynamically changing AAR data.

Methods	Precision	Recall	F1-Score	Accuracy
DBSCAN	0.41	0.45	0.36	0.79
Agg. Clustering	0.14	0.77	0.22	0.29
HDBSCAN	0.15	0.87	0.24	0.23
K-Means	0.28	0.55	0.34	0.71
Bisecting K-Means	0.15	0.71	0.23	0.34
BIRCH	0.14	0.74	0.22	0.28
OPTICS	0.15	0.77	0.23	0.29
Gaussian Mixtures	0.20	0.85	0.30	0.53
CMAD$_{Benchmark}$	**0.49**	**0.80**	**0.56**	**0.87**

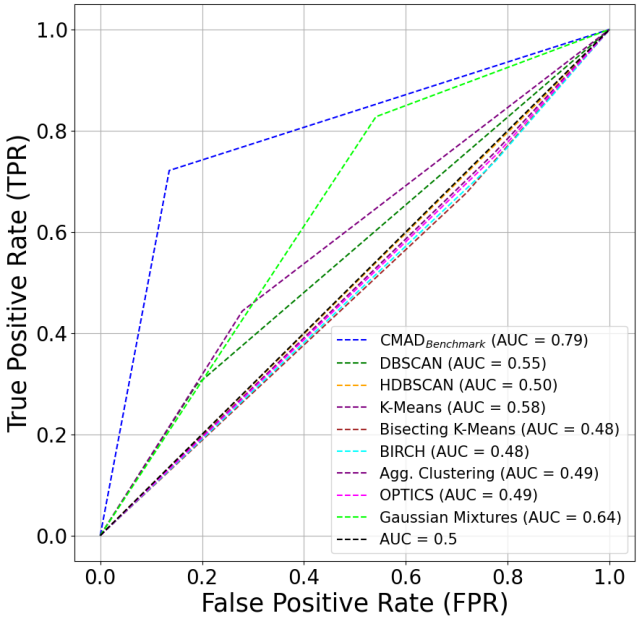

Figure 5: ROC curve illustrating the performance of anomaly detection methods (CMAD$_{Benchmark}$, DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering (Agg. Clustering), OPTICS, and Gaussian Mixtures) in dynamically changing data. CMAD$_{Benchmark}$ demonstrates robust performance across varying conditions, effectively capturing changes in extreme values over time in comparison to other methods

Antarctic region. Since the mass of the ice sheet over the Antarctic is equivalent to a sea level rise of 200 feet [2], the rapid retreat of its protective blanket (or, sea ice) is of significant concern. Our study discovers anomalous melting patterns on sea ice retreat which warrant further investigation and help in understanding the importance of the role of sea ice retreat.

5 CONCLUSION AND FUTURE WORK

In this paper, we introduced the concept of inverse max pooling in CNN since the study primarily focuses on negative values of SIE or retreats in SIE from its previous day. Furthermore, this research addressed the challenge of applying conventional methods to dynamic datasets. It addressed this issue by incorporating a grid-wise matrix profile method to detect extreme values in time series, utilizing the gridded version of the data. Additionally, the study explored the application of DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, and Gaussian Mixtures to capture patterns in dynamically changing data. Our results show that the use of traditional methods like DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, Affinity Propagation, Mean Shift, Spectral Clustering and Gaussian Mixtures on time series data may not yield meaningful results, especially on a dynamic dataset that deals with a temporal variation of spatial objects (AAR). It appears from our analysis that inverse max pooling-based CMAD can be used to detect the spatial distributions and sub-regional variations in anomalous melting over the Antarctic region. CMAD$_{Benchmark}$ is more suitable for finding the temporal evolution of clusters of anomalous events than DBSCAN, HDBSCAN, K-Means, Bisecting K-Means, BIRCH, Agglomerative Clustering, OPTICS, and Gaussian Mixtures for AAR data. Nevertheless, when sparse matrices are employed as input, CMAD tends to underestimate the occurrence of anomalous melting. In the future, we plan to improve the CMAD method to better estimate anomalous melting and steady state melting (melting rate close to the mean melting rate instead of detecting melting events beyond 3σ and lower bound). Building on the successful detection of clusters of anomalous events in sea ice using the inverse max pooling concept, which has demonstrated an 87% accuracy in detecting anomalies within benchmark data, our future work will extend this approach. We aim to integrate the inverse max pooling technique into a comprehensive CNN model. This model will be designed to predict anomalous sea ice extent events using satellite imagery. We will include an extension of the inverse max pooling study to encompass full CNN. This idea can be extended by applying multiple hidden layers in a full CNN model leveraging the inverse max pooling concept to predict clusters of anomalous events. In our study, we have utilized satellite images spanning from 2000 to 2023 to calculate the threshold for anomalous melting events. It's important to note that considering data from past years introduces the possibility of changes in threshold over time. To address this uncertainty and understand how variations may occur, we aim to investigate the bounds of this uncertainty. Examining the range of potential changes will contribute to a more comprehensive understanding of the dynamics and uncertainties associated with the threshold calculation, enhancing the robustness of our findings. This deeper understanding is crucial for developing accurate machine-learning models that can predict the melting of land ice sheets and the rise of sea levels in the future, especially if the sea ice in the Antarctic continues to retreat.

ACKNOWLEDGMENTS

This work is supported by NSF Award #2118285, "iHARP: NSF HDR Institute for Harnessing Data and Model Revolution in the Polar Region."

REFERENCES

[1] Giant iceberg breaks away from Antarctic ice shelf [n. d.]. *Why Ice Sheets Matter.* Giant iceberg breaks away from Antarctic ice shelf. https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Giant_iceberg_breaks_away_from_Antarctic_ice_shelf Accessed Sep. 22, 2023.

[2] National Snow and Ice Data Center [n. d.]. *Why Ice Sheets Matter.* National Snow and Ice Data Center. https://nsidc.org/learn/parts-cryosphere/ice-sheets/why-ice-sheets-matter Accessed Aug. 03, 2023.

[3] Timo Balz, Ruonan Gao, Walter Musakwa, James Magidi, and Zhenfeng Shao. 2024. South-to-South Cooperation in Multi-Source Satellite Data for Improving Food Security. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 48 (2024), 19–24.

[4] J. M. Barragán and M. de Andrés. 2015. Analysis and trends of the world's coastal cities and agglomerations. *Ocean & Coastal Management* 114 (Sep 2015), 11–20. https://doi.org/10.1016/j.ocecoaman.2015.06.004

[5] Cecilia M Bitz. 2008. Some aspects of uncertainty in predicting sea ice thinning. *Arctic Sea Ice Decline: Observations, Projections, Mechanisms, and Implications, Geophys. Monogr* 180 (2008), 63–76.

[6] Einar Bjørgo, Ola M Johannessen, and Martin W Miles. 1997. Analysis of merged SMMR-SSMI time series of Arctic and Antarctic sea ice parameters 1978–1995. *Geophysical Research Letters* 24, 4 (1997), 413–416.

[7] Julien Boé, Alex Hall, and Xin Qu. 2009. September sea-ice cover in the Arctic Ocean projected to vanish by 2100. *Nature Geoscience* 2, 5 (2009), 341–343.

[8] Jake Bouvrie. 2006. Notes on convolutional neural networks. (2006).

[9] Ricardo JGB Campello, Davoud Moulavi, and Jörg Sander. 2013. Density-based clustering based on hierarchical density estimates. In *Pacific-Asia conference on knowledge discovery and data mining*. Springer, 160–172.

[10] SUDIP CHAKRABORTY. 2022. [Data] Global Aerosol Atmospheric Rivers Database, Version 1. https://doi.org/10.25346/S6/CXO9PD

[11] S. Chakraborty et al. 2023. Understanding the Role of 2019 Amazon Wildfires on Antarctic Sea Ice Extent Using Data Science Approaches. In *Proceedings of KDD'23 Fragile Earth Workshop*. ACM, Long Beach, CA, USA, 8 pages. Accepted.

[12] Sudip Chakraborty, Bin Guan, Duane Waliser, and Arlindo da Silva. 2022. Aerosol atmospheric rivers: climatology, event characteristics, and detection algorithm sensitivities. *Atmospheric Chemistry and Physics Discussions* 2022 (2022), 1–41.

[13] Liqiu Chen, Chongshi Gu, Sen Zheng, and Yanbo Wang. 2024. A method for identifying gross errors in dam monitoring data. *Water* 16, 7 (2024), 978.

[14] Climate.gov. n.d.. *Understanding Climate: Antarctic Sea Ice Extent.* https://www.climate.gov/news-features/understanding-climate/understanding-climate-antarctic-sea-ice-extent

[15] Anil Damle, Victor Minden, and Lexing Ying. 2019. Simple, direct and efficient multi-way spectral clustering. *Information and Inference: A Journal of the IMA* 8, 1 (2019), 181–203.

[16] Maloy Kumar Devnath, Avijoy Chakma, Mohammad Saeid Anwar, Emon Dey, Zahid Hasan, Marc Conn, Biplab Pal, and Nirmalya Roy. 2023. A Systematic Study on Object Recognition Using Millimeter-wave Radar. In *2023 IEEE International Conference on Smart Computing (SMARTCOMP)*. 57–64. https://doi.org/10.1109/SMARTCOMP58114.2023.00025

[17] Francis X Diebold, Maximilian Göbel, Philippe Goulet Coulombe, Glenn D Rudebusch, and Boyuan Zhang. 2021. Optimal combination of Arctic sea ice extent measures: A dynamic factor modeling approach. *International Journal of Forecasting* 37, 4 (2021), 1509–1519.

[18] F. X. Diebold, G. D. Rudebusch, M. Göbel, P. Goulet Coulombe, and B. Zhang. 2023. When will Arctic sea ice disappear? Projections of area, extent, thickness, and volume. *Journal of Econometrics* 236, 2 (Oct 2023), 105479. https://doi.org/10.1016/j.jeconom.2023.105479

[19] Paweł D Domański. 2020. Study on statistical outlier detection and labelling. *International Journal of Automation and Computing* 17, 6 (2020), 788–811.

[20] ESA. [n. d.]. *Giant iceberg breaks away from Antarctic ice shelf.* https://www.esa.int/Applications/Observing_the_Earth/Copernicus/Giant_iceberg_breaks_away_from_Antarctic_ice_shelf Accessed: 08 February, 2024.

[21] UMBC High Performance Computing Facility. 2023. *UMBC High Performance Computing Facility.* https://hpcf.umbc.edu/ Accessed: 08 February, 2024.

[22] Alejandro C Frery. 2023. Interquartile Range. In *Encyclopedia of Mathematical Geosciences*. Springer, 664–666.

[23] Mina Bagherzade Ghazvini, Miquel Sànchez-Marrè, Davood Naderi, and Cecilio Angulo. 2024. Anomaly Detection in Gas Turbines Using Outlet Energy Analysis with Cluster-Based Matrix Profile. *Energies* 17, 3 (2024), 653.

[24] Cyril Goutte and Eric Gaussier. 2005. A probabilistic interpretation of precision, recall and F-score, with implication for evaluation. In *European conference on information retrieval*. Springer, 345–359.

[25] Hongyuan Guo, Xingpo Liu, and Qichen Zhang. 2024. Identifying daily water consumption patterns based on K-means Clustering, Agglomerative Hierarchical Clustering, and Spectral Clustering algorithms. *AQUA—Water Infrastructure, Ecosystems and Society* (2024), jws2024294.

[26] Mahsa Hajihosseinlou, Abbas Maghsoudi, and Reza Ghezelbash. 2024. A comprehensive evaluation of OPTICS, GMM and K-means clustering methodologies for geochemical anomaly detection connected with sample catchment basins. *Geochemistry* (2024), 126094.

[27] Mahsa Hajihosseinlou, Abbas Maghsoudi, and Reza Ghezelbash. 2024. Intelligent mapping of geochemical anomalies: Adaptation of DBSCAN and mean-shift clustering approaches. *Journal of Geochemical Exploration* 258 (2024), 107393.

[28] Feifei Han, Xueyu Zhang, Jingshan Yu, Shugao Xu, Guihuan Zhou, and Shuang Li. 2024. Study on spatiotemporal dynamic characteristics of precipitation and causes of waterlogging based on a data-driven framework. *Science of The Total Environment* 913 (2024), 169796.

[29] Mohammad Hasib, Bagas Anwar Arif Nur, Huffaz Muhammad Abdurrofi Baith, Abdullah Mu'adz Muflih, Cahli Suhendi, Bana Fitro Ghifari, Hafidz Bagus Prasetyo Adi, Estu Kriswati, Titi Anggono, Febty Febriani, et al. 2024. Event classification of volcanic earthquakes based on K-Means clustering: Application on Anak Krakatau Volcano, Sunda Strait. In *IOP Conference Series: Earth and Environmental Science*, Vol. 1314. IOP Publishing, 012045.

[30] Kenza Himmich, Martin Vancoppenolle, Gurvan Madec, Jean-Baptiste Sallée, Paul R Holland, and Marion Lebrun. 2023. Drivers of Antarctic sea ice advance. *Nature Communications* 14, 1 (2023), 6219.

[31] Zhe Hui Hoo, Jane Candlish, and Dawn Teare. 2017. What is an ROC curve?

[32] R. Huang, C. Wang, J. Li, and Y. Sui. 2023. DF-UHRNet: A Modified CNN-Based Deep Learning Method for Automatic Sea Ice Classification from Sentinel-1A/B SAR Images. *Remote Sensing* 15, 9 (Jan 2023), Art. no. 9. https://doi.org/10.3390/rs15092448

[33] Divya Jayabalan and Sivasankar Elango. 2024. ICE-VDOP: an integrated clustering and ensemble machine learning methods for an enhanced vector-borne disease outbreak prediction using climatic variables. *International Journal of Information Technology* 16, 4 (2024), 2077–2088.

[34] Kamran Khan, Saif Ur Rehman, Kamran Aziz, Simon Fong, and Sababady Sarasvady. 2014. DBSCAN: Past, present and future. In *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*. IEEE, 232–238.

[35] J. Kromer and L. Trusel. 2021. Using Causal Discovery Applied to Observational Datasets to Evaluate Ocean, Atmosphere, and Ice Sheet Surface Interactions in Antarctica. 2021 (Dec 2021), C43A–06.

[36] Sascha Krysmon, Stefan Pischinger, Johannes Claßen, Georgi Trendafilov, Marc Düzgün, Frank Dorscheidt, Martin Nijs, and Michael Görgen. 2024. Applying Density-Based Clustering for the Analysis of Emission Events in Real Driving Emissions Calibration. *Future Transportation* 4, 1 (2024), 46–66.

[37] Quanhong Liu, Ren Zhang, Yangjun Wang, Hengqian Yan, and Mei Hong. 2021. Daily prediction of the arctic sea ice concentration using reanalysis data based on a convolutional lstm network. *Journal of Marine Science and Engineering* 9, 3 (2021), 330.

[38] Leland McInnes and John Healy. 2017. Accelerated hierarchical density based clustering. In *2017 IEEE international conference on data mining workshops (ICDMW)*. IEEE, 33–42.

[39] NASA. [n. d.]. *Vital Signs: Ice Sheets.* https://climate.nasa.gov/vital-signs/ice-sheets/ Accessed: 08 February, 2024.

[40] NASA. 2014. *NASA IceBridge Blog.* https://blogs.nasa.gov/icebridge/2014/11/19/east-and-west-the-geography-of-antarctica/ Accessed on: February 08,2024.

[41] Shantikumar S Ningombam, EJL Larson, G Indira, BL Madhavan, and Pradeep Khatri. 2024. Aerosol classification by application of machine learning spectral clustering algorithm. *Atmospheric Pollution Research* 15, 3 (2024), 102026.

[42] NASA Earth Observatory. [n. d.]. NASA Earth Observatory. https://earthobservatory.nasa.gov/features/SeaIce Accessed Jun. 7, 2024.

[43] Intergovernmental Panel on Climate Change (IPCC). 2013. Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.

[44] Subrata Kumer Paul, Md Abul Ala Walid, Rakhi Rani Paul, Md Jamal Uddin, Md Sohel Rana, Maloy Kumar Devnath, Ishaat Rahman Dipu, and Md Momenul Haque. 2024. An Adam based CNN and LSTM approach for sign language recognition in real time for deaf people. *Bulletin of Electrical Engineering and Informatics* 13, 1 (2024), 499–509.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. 2024. *Clustering.* https://scikit-learn.org/stable/modules/clustering.html Accessed: 2024-05-17.

[46] D Sathiaraj, Xiaoqi Huang, and J Chen. 2019. Predicting climate types for the Continental United States using unsupervised clustering techniques. *Environmetrics* 30, 4 (2019), e2524.

[47] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN.

ACM Transactions on Database Systems (TODS) 42, 3 (2017), 1–21.

[48] National Snow and Ice Data Center (NSIDC). [n. d.]. Albedo. https://nsidc.org/learn/cryosphere-glossary/albedo Accessed Jun. 7, 2024.

[49] Sharon Stammerjohn, Robert Massom, David Rind, and Douglas Martinson. 2012. Regions of rapid sea ice change: An inter-hemispheric seasonal comparison. *Geophysical Research Letters* 39, 6 (2012).

[50] J. Turner et al. 2020. Recent Decrease of Summer Sea Ice in the Weddell Sea, Antarctica. *Geophysical Research Letters* 47, 11 (2020), e2020GL087127. https://doi.org/10.1029/2020GL087127

[51] John Turner, Caroline Holmes, Thomas Caton Harrison, Tony Phillips, Babula Jena, Tylei Reeves-Francois, Ryan Fogt, Elizabeth R Thomas, and CC Bajish. 2022. Record low Antarctic sea ice cover in February 2022. *Geophysical Research Letters* 49, 12 (2022), e2022GL098904.

[52] Abhishek Varma, Leela Krishna, and Muthukumaran Malarvel. 2024. Locality Recommendation System Using Exploratory Data Analysis and Affinity Propagation. In *Next-Gen Technologies in Computational Intelligence: Proceeding of the International Conference on Next-Gen Technologies in Computational Intelligence (NGTCA 2023)*. CRC Press, 285.

[53] A Vijay and S Shilpa. 2024. A Study of Clustering Algorithms on Big Data using Spark. In *2024 4th International Conference on Data Engineering and Communication Systems (ICDECS)*. IEEE, 1–6.

[54] Maofa Wang, Guangda Gao, Hongliang Huang, Ali Asghar Heidari, Qian Zhang, Huiling Chen, and Weiyu Tang. 2021. A principal component analysis-boosted dynamic Gaussian mixture clustering model for ignition factors of Brazil's rainforests. *IEEE Access* 9 (2021), 145748–145762.

[55] Mengxin Wang and Zhonghua Gou. 2024. Gaussian Mixture Model based classification for analyzing longitudinal outdoor thermal environment data to evaluate comfort conditions in urban open spaces. *Urban Climate* 53 (2024), 101792.

[56] Ya Wang, Gang Huang, Baoxiang Pan, Pengfei Lin, Niklas Boers, Weichen Tao, Yutong Chen, Bo Liu, and Haijie Li. 2024. Correcting Climate Model Sea Surface Temperature Simulations with Generative Adversarial Networks: Climatology, Interannual Variability, and Extremes. *Advances in Atmospheric Sciences* (2024), 1–14.

[57] Yunhe Wang, Xiaojun Yuan, Yibin Ren, Mitchell Bushuk, Qi Shu, Cuihua Li, and Xiaofeng Li. 2023. Subseasonal prediction of regional Antarctic sea ice by a deep learning model. *Geophysical Research Letters* 50, 17 (2023), e2023GL104347.

[58] Zongliang Wang, Zhen Li, Jiangyuan Zeng, Shuang Liang, Ping Zhang, Fuquan Tang, Siyuan Chen, and Xiongwei Ma. 2020. Spatial and temporal variations of Arctic sea ice from 2002 to 2017. *Earth and Space Science* 7, 9 (2020), e2020EA001278.

[59] Katherine Wilson, Andrew Arreak, Trevor Bell, Gita Ljubicic, Sikumiut Committee, et al. 2021. The Mittimatalik Siku Asijjipallianinga (Sea Ice Climate Atlas): how Inuit knowledge, earth observations, and sea ice charts can fill IPCC climate knowledge gaps. *Frontiers in Climate* 3 (2021), 715105.

[60] Ding-yin Xia, Fei Wu, Xu-qing Zhang, and Yue-ting Zhuang. 2008. Local and global approaches of affinity propagation clustering for large scale data. *Journal of Zhejiang University-Science A* 9, 10 (2008), 1373–1381.

[61] Jing Zhang, Min Zhang, Yang Yu, and Ruide Yu. 2024. An innovative method integrating run theory and DBSCAN for complete three-dimensional drought structures. *Science of the Total Environment* 926 (2024), 171901.