American Journal of Remote Sensing
2023; 11(2): 32-35
http://www.sciencepublishinggroup.com/j/ajrs
doi: 10.11648/j.ajrs.20231102.11
ISSN: 2328-5788 (Print); ISSN: 2328-580X (Online)

Assessment and Classification of Cloud Coverage Using K-Means Clustering Algorithm for the Sentinel-3 LST Data: A Case Study in the Fujairah Region

Manar Ahmed Mohammed Alblooshi, Sirajul Huda Kalathingal, Shaher Bano Mirza*, Fouad Lamghari Ridouane

Fujairah Research Centre, Fujairah, United Arab Emirates

Email address:

shaher.bano@frc.ae (Shaher Bano Mirza)

*Corresponding author

To cite this article:

Manar Ahmed Mohammed Alblooshi, Sirajul Huda Kalathingal, Shaher Bano Mirza, Fouad Lamghari Ridouane. Assessment and Classification of Cloud Coverage Using K-Means Clustering Algorithm for the Sentinel-3 LST Data: A Case Study in the Fujairah Region. *American Journal of Remote Sensing*. Vol. 11, No. 2, 2023, pp. 32-35. doi: 10.11648/j.ajrs.20231102.11

Received: June 16, 2023; **Accepted**: July 7, 2023; **Published**: July 17, 2023

Abstract: Clouds have a significant impact on the planet's energy balance, climate, and weather. They serve as the primary temperature regulator and function as a blanket to absorb thermal energy or longwave radiation. The present study estimates the percentage of rainfall clouds within a 100-kilometer radius of Fujairah City on the Gulf of Oman using image processing based on machine learning and digital image processing. The data for 9 months starting from January 2022 to October 2022 has been retrieved from the Copernicus satellite data component through the Sentinel 3 LST F2 channel. K-mean cluster analysis has been used to validate the accuracy of an algorithm which is applied to determine cloud cover, with a precision rate of 99.9% for clear weather and 95.5% for overcast conditions. The findings indicate that most of the rainy clouds were observed during the months of January and July. The remaining duration of the year exhibits a reduced occurrence of these clouds. Beginning in February, the region of interest experiences cloud cover accompanied by precipitation subsequent to the month of January. Similarly, the month of July exhibited cloud covers with moisture. Throughout the year, dry clouds are observed with moderate coverage percentages. However, there are no observations of any of these clouds during the months of May and December. In summary, automated systems for observing clouds in the atmosphere are a valuable method for detecting cloud cover and predicting climatic patterns in diverse geographical locations.

Keywords: Cloud Coverage, LST, Land Surface Temperature, K-Mean Clustering, Sentinel-3, Copernicus, UAE, Fujairah

1. Introduction

Clouds have a significant impact on the Earth's energy balance, climate, and weather as being the primary temperature regulator on the planet. Some clouds contribute to cooling because they reflect back into space a portion of the solar energy or shortwave radiation. Other clouds contribute to global warming by acting as a blanket to capture thermal energy or longwave radiation emitted by the Earth's surface and lower atmosphere. Cloud systems also aid in uniformly distributing the sun's energy across the Earth's surface [1]. Cloud-related research requires extensive cloud observations, such as the number and diversity of clouds. Humans have traditionally been the ones to observe these macroscopic factors. However, human perception is subjective and may occasionally produce inaccurate results. When different observers evaluate the same sky condition, there may be a disparity in the number of clouds, for instance. Automated cloud observation systems have been developed due to the limitations of human observation. Using digital images, these systems capture sky conditions and analyse cloud properties. Digital image processing has numerous advantages over analogue image processing as a subcategory or field of digital signal processing. It enables the application of a much broader variety of algorithms to the input data and can prevent problems such as the accumulation of noise and distortion during processing. Considering that images are defined in at least two dimensions, digital image processing can be modelled as multidimensional systems. The generation and development of digital image processing are primarily influenced by three factors: first, the evolution of computers; second, the evolution of mathematics (especially the creation and improvement of discrete mathematics theory); and third, the increased demand for a variety of applications in the environment, agriculture, military, industry, and medicine [2]. There are a variety of image processing techniques, including neural network-based, edge-based, cluster-based, and threshold-based techniques. The clustering approach is one of the most efficient techniques available. Various varieties of clustering include K-means clustering, Fuzzy C-means clustering, the mountain clustering method, and the subtractive clustering method. K-means clustering is one of the most frequently employed clustering algorithms. It is more user-friendly and conducts computations faster than hierarchical clustering [3]. The present study estimates the percentage of rainfall clouds within a 100-kilometer radius of Fujairah City on the Gulf of Oman using image processing based on machine learning and digital image processing.

2. Methodology

2.1. Data Collection

Figure 1. *Area of interest: 100 km off the coastline of Emirates of Fujairah.*

The area of 100km radius, off the coast of emirate of Fujairah has been selected for the estimation of percentage of various types of cloud analysis. The data for 9 months starting from January 2022 to October 2022 has been retrieved from the Copernicus satellite data component through the Sentinel 3 LST F2 channel, utilising Python packages and the Google Earth Engine.

The Sentinel-3 mission is to measures sea surface topography, temperature, and colour to enable ocean forecasting, environmental monitoring, and climate monitoring [4]. Land-surface temperature (LST) is an essential climate variable that describes processes such as energy and water exchange between the land surface, atmosphere, and plant growth [5]. This temperature map can be used to estimate the cloud. The data produced by the satellite has mainly five cloud combinations Null area with no data available, non-Cloud area and cloud with moistures, rainy clouds, and dry clouds (Figure 2). The K mean clustering has been used to classify the area with cloud and cloud combinations [6].

2.2. Data Pre-processing and Algorithm of K-Means Cluster

To mitigate the risk of misclassification by the k-means clustering algorithm, a standardised colour bar was incorporated alongside each image to ensure accurate categorization (Figure 2). The label colours are selected by carefully examining the existing colour scheme in the data. The area surrounding the image consistently exhibits a grey hue, while the interior of the all-in-one (AIO) displays a range of colours including blue, white, pale blue, orange, pale yellow, among others. The process of image processing with the use of a colour bar attachment is carried out through the implementation of Python programming [7].

Figure 2. *Standard colour bar for cloud classification.*

Following the data collection, a cluster k value of 6 was selected. The null area is characterised by a dark brown colour, indicating the absence of available data. The region depicted in orange on the map is characterised as a non-cloud area. A cloud exhibiting a blue hue due to the presence of moisture and the dry clouds are represented by white colour (Figure 3). A total of k data points was randomly selected as the centre, and the Euclidean distance was calculated by assigning each data point to its closest centre. The centroid and mean of the cluster were determined by utilising all available data points. The procedure was iterated until the fulfilment of the completion criteria which can only be initiated upon reaching the maximum number of iterations. Centroid and data points remained in the same clusters [8].

Classification keeps the order of colours in the picture. Due to an equal number of colour labels in both photos, the output image's colour labels have been re-labelled programmatically to match the input images in the same order [9].

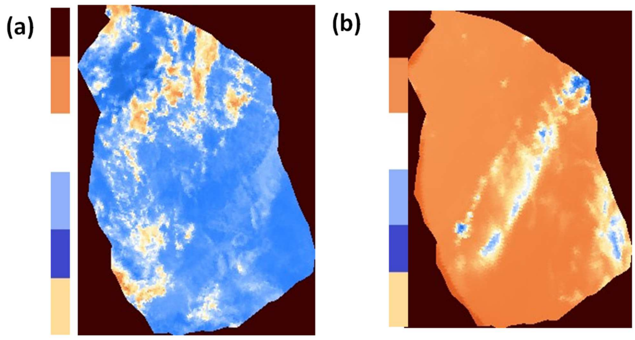

Figure 3. The cloud cover over the area of Interest. (a) Rainy cloud coverage in the month of January (b) few instances of no cloud coverage in the same area for the month of January. Light Blue: Cloud with moisture; Blue: rainy cloud; White/yellow: Dry clouds; Dark brown: Null area with no available data; Orange: Non cloud area.

2.3. Pixel Calculation

Following the colour relabeling from the output to the input using the same order, the number of pixels for each colour was calculated to determine the percentage of each colour by dividing the number for each of the six colours by the sum of all the picture pixels, which was then multiplied by

$$100.pixel\ percentage\% = \frac{number\ of\ pixels}{total\ of\ image\ pixels} * 100$$

3. Results and Discussion

A cluster analysis was conducted on image set of 272, depicting various cloud cover scenarios at different times of the day, to verify the validity of applied methodology. The findings indicate that the approach was able to accurately determine the state of cloud coverage, achieving a precision rate exceeding 99% for clear weather conditions and 95% for overcast conditions. The accuracies were determined by evaluating the algorithm's ability to correctly identify pixels as determined by human observers. The classifications were determined by calculating the percentage of pixels within the full image that corresponded to each category. Results from an automatic categorization process are shown in the tables below, with percentages for rainy clouds, clouds with moisture and dry clouds.

Table 1. *The percentage values of rainy cloud, cloud with moisture and dry cloud have been mentioned.*

Month	Rainy Clouds		Clouds with Moisture		Dry Clouds	
	Maximum	Minimum	Maximum	Minimum	Maximum	Minimum
January	92.7	18.48	24.07	5.83	42.7	1.29
February	0.446	0	74.6	6.76	19.1	1.79
March	24.15	2	23.49	8.33	28.2	1.80
April	9.23	0	8.70	0	22.2	3.79
May	0	0	0	0	0	0
June	0	0	14	7	14.43	1.38
July	47.20	4.4	97.9	3.8	33.27	1.77
August	2.96	0	6.8	2.2	25.11	1.25
September	0	0	0	0	0	0

Table 1 displays the occurrence of rainy clouds from the months of January to September. The month of January exhibited the highest proportion of rainy clouds, registering a percentage of 92.7%. This month was among the months with the highest number of observations for the year 2022. July exhibited the highest proportion of rainy clouds, accounting for 47.20% of the total, followed by March with 24.15%. The months of April and August exhibited low rates, with April registering a rate of 9.23% and August recording a rate of 2.96%. No potential rainy clouds were observed for the months of February, May, June, and September. The months of February and July exhibited the greatest concentration of clouds with moisture. Specifically, July demonstrated the highest proportion with a percentage of 97.9%, while February followed with a percentage of 74%. The data indicates that the middle to low percentages for January, March, April, June, and August were 24.07%, 23.49%, 8.7%, 14%, and 6.8%, respectively. No wet clouds observed to detect clouds with moisture during the months of May and September. The months of January and July exhibit the highest percentage of dry cloud cover. During the months of January, February, March, April, June, July, and August, an observation was made indicating that dry clouds were present in moderate to reduced amounts, accounting for more than 20% of the total cloud cover during each of these months. The percentage data for the months of May and September was not detected. Image data features and machine learning methods are being utilized excessively to address the task of cloud coverage assessment. It holds significant importance in weather forecasting and monitoring which offers distinct advantages over traditional methods as it automates the process and reduces reliance on human-eye observations. One of the studies has compared cloud cover results obtained through machine learning with human-eye observations. The pattern similarity of the obtained results from both approached explain the effectivity of the generated model [10-14].

4. Conclusion

In conclusion, it is necessary to conduct thorough observations of clouds, including but not limited to their quantity and variety. This research highlights the importance and potential of machine learning in advancing cloud coverage assessment, thereby facilitating more accurate and efficient forecasting systems. Machine learning techniques enable the extraction of valuable information from the image data, including temporal and meteorological parameters, as well as statistical characteristics. Hence, Automated cloud observation systems are a valuable technique for detecting cloud cover and forecasting climate patterns in various locations.

Conflicts of Interest

The authors declare that they have no conflict of interests.

References

[1] A. Ward, "NASA Facts the Importance of Understanding Clouds." [Online]. Available: www.nasa.gov

[2] "Digital image processing - Wikipedia." https://en.wikipedia.org/wiki/Digital_image_processing (accessed Oct. 25, 2022).

[3] C. W. Chen, J. Luo, and K. J. Parker, "Image segmentation via adaptive K-mean clustering and knowledge-based morphological operations with biomedical applications," *IEEE transactions on image processing*, vol. 7, no. 12, pp. 1673–1683, 1998.

[4] "Sentinel-3 - Wikipedia." https://en.wikipedia.org/wiki/Sentinel-3 (accessed Oct. 27, 2022).

[5] "Land Surface Temperature - Applications - Sentinel-3 SLSTR - Sentinel Online - Sentinel Online." https://sentinels.copernicus.eu/web/sentinel/user-guides/sentinel-3-slstr/overview/geophysical-measurements/land-surface-temperature (accessed Dec. 08, 2022).

[6] J. Yang, J. Zhou, F.-M. Göttsche, Z. Long, J. Ma, and R. Luo, "Investigation and validation of algorithms for estimating land surface temperature from Sentinel-3 SLSTR data," *International Journal of Applied Earth Observation and Geoinformation*, vol. 91, p. 102136, 2020.

[7] M. Blazek and P. Pata, "Colour transformations and K-means segmentation for automatic cloud detection," *Meteorologische Zeitschrift*, vol. 24, no. 5, pp. 503–509, 2015.

[8] N. Dhanachandra, K. Manglem, and Y. J. Chanu, "Image segmentation using K-means clustering algorithm and subtractive clustering algorithm," *Procedia Comput Sci*, vol. 54, pp. 764–771, 2015.

[9] E. Grilli, F. Menna, and F. Remondino, "A review of point clouds segmentation and classification algorithms," *The International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 42, p. 339, 2017.

[10] B.-Y. Kim, J. W. Cha, and K.-H. Chang, "Twenty-four-hour cloud cover calculation using a ground-based imager with machine learning," Atmos Meas Tech, vol. 14, no. 10, pp. 6695–6710, 2021.

[11] F. Rashed, M. O. Alhefeiti, and S. B. Mirza, "Investigation of the spatiotemporal changes in the vegetation of the emirates of Fujairah using normalized difference vegetation index (NDVI) and geographical information system (GIS)," 2022, [Online]. Available: www.biosciencejournals.com

[12] A. Saeed, S. Aldhanhani, M. Sirajul, H. Kalathingal, S. B. Mirza, and F. L. Ridouane, "Efficient comprehension of weather pattern from the meteorological dataset acquired from airport sectors of Fujairah, UAE, using machine learning and data mining approaches," 2023, [Online]. Available: https://www.windfinder.com/forecast/fujairah

[13] M. A. M. Alblooshi, F. R. M. O. Alhefeiti, M. S. Huda, S. B. M. Kalathingal, and F. L. Ridouane, "Assessment of the potential of various types of long short-term memory (LSTM) artificial neural networks and its application in weather forecasting," *International Journal of Advanced Engineering and Technology*, 2023.

[14] A. S. J. Alabdouli, M. S. H. Kalathingal, S. B. Mirza, and F. L. Ridouane, "Effective Oil Spill Monitoring Approach over the Gulf of Oman by Using Advanced Machine Learning and Data Mining Tools," *Int J Swarm Evol Comput*, vol. 12, p. 276, 2023.