Application of Principal Component Analysis and Unsupervised Classification (ISODATA) to Estimate Arctic Sea Ice Extent Using ISRO's SCATSAT-1 Data

Nanaoba Khoisnam Singh[1,✉], Rajkumar Kamaljit Singh[2], Mamata Maisnam[1], Jayaprasad Palliprad[3], Saroj Maity[3]

[1] National Institute of Technology Manipur, Imphal-795004, Manipur, India;
[2] Manipur Technical University, Imphal-795004, Manipur, India;
[3] Space Application Centre- Indian Space Research Organisation (ISRO), Ahmedabad-380015, Gujarat, India
✉ meiteinanaoba@gmail.com

Keywords: Scatterometer, SCATSAT-1, ISODATA, Principal component analysis, Sea ice extent

1. Introduction

Sea ice plays an active role in ocean circulation, weather and regional climate as well as on the salinity of icy oceans. The insulating property of sea ice reduces evaporation of underlying water and heat loss to the atmosphere. The high albedo of 0.5 to 0.7 of sea ice reduces the global temperature (National Snow and Ice Data Center, 2020).

Expulsion of brine during the formation of sea ice increases salinity and density of surrounding ocean water in the polar regions (Lake and Lewis, 1970). In case of Antarctica, this dense water flows down to the continental shelf to form Antarctic Bottom Water (Dan and Robert, 2006) and in the Arctic, flows down to the ocean floor and moves to enter the Greenland and Norwegian Seas and forms Arctic Bottom Water (Glossary of American Meteorological Society, 2012). These waters lead to the formation of the thermohaline circulation, the largest ocean circulation.

Polar-orbiting satellite data are useful in observing and studying long term and seasonal changes of the Arctic sea ice extent (SIE). SIE is considered as a sensitive indicator of long-term global climate change (Budd, 1975). Estimation of SIE has been carried out using active remote sensing (Remund and Long, 2014) and passive remote sensing (Cavalieri and Parkinson, 2008; Comiso et al., 1996).

In this study, enhanced resolution SCATSAT-1 data are used to estimate the Arctic SIE. Image analysis technique of Principal Component Analysis (PCA) is applied to highlight different features within an image. After masking the land parts, discrimination between sea ice and ocean water is carried out using ISODATA classification. The results obtained are validated by comparing with 1) National Snow and Ice Data Centre (NSIDC) sea ice extent derived from 30 % sea ice concentration and 2) Ocean and Sea Ice Satellite Application Facility (OSI) sea ice edge. Statistical significance is calculated between the comparing data sets.

2. Data

The primary data of the study is SCATSAT-1 data. To validate the results of the SCATSAT-1 data, popular SIE data of NSIDC Sea ice concentration data and OSI sea ice-covered data are utilised.

2.1. SCATSAT-1

Though originally designed for the measurement of ocean winds, scatterometers have vast applications in other Earth observations. Likewise, Scatterometer Satellite-1 (SCATSAT-1), that measures near-surface ocean wind vectors, can be used to study different phenomena on Earth.

This active sensing satellite works at 13.515 GHz (Ku-band microwave). It was launched on 26 September 2016. Despite the gap of two and a half years (from February 2014 to September 2016) between the Oceansat-2 mission and SCATSAT-1, it is considered as the continuity mission of Oceansat-2. These two have similar functions and working principle. SCATSAT-1 data are archived at the Meteorological and Oceanographic Satellite Data Archival Centre, Space Applications Centre-ISRO, India (https://www.mosdac.gov.in).

This satellite gives three important parameters: backscattering coefficient (σ_0), incidence angle normalised backscattering coefficient or gamma nought (γ_0) and noise-derived brightness temperature (T_b). These parameters are measured in horizontal and vertical polarizations. Therefore, six datasets per day are available. Since it is a polar-orbiting satellite there are two passes, the ascending-pass (northward from the equator) and the descending-pass (southward from the equator). The data sets that we have used in this study are the ones obtained by combining these two daily passes, known as "both-pass". The spatial resolution of the data are the enhanced resolution of 2.25 x 2.25 km^2. More on these data sets are given by SCATAT-1 DP Team (2017).

Proceedings of the 5th International YES Congress
DOI: 10.2312/yes19.21

Figure 1: Mask developed from SCATSAT-1.

2.2. NSIDC sea ice concentration

These data are generated from the brightness temperature data obtained from such sensors as the Scanning Multichannel Microwave Radiometer (SMMR) onboard Nimbus-7, Special Sensor Microwave/Imagers (SSM/I) onboard Defense Meteorological Satellite Program (DMSP)-F8, -F11 and -F13, and Special Sensor Microwave Imager/Sounder (SSMIS) on DMSP-F17 (Cavalieri et al., 1995).

The data are stored as a one-byte integer in 25 x 25 km² resolution pixels. The digital value for a pixel can range from 0 to 250. This value when divided by 2.50 will give us the sea ice concentration per pixel in percentage. For this study, a pixel with 30 % sea ice concentration is counted as a pixel of sea ice signature.

The data (data catalogue number NSIDC-0051) are archived at NSIDC and were downloaded from https://nsidc.org/data/nsidc-0051.

2.3. OSI sea ice concentration

The second validation data set is the sea ice concentration generated from both the passive satellites (SSM/Is, Advanced Microwave Scanning Radiometer 2/AMSR2 and Global Change Observation Mission-W1/GCOM-W1) and an active satellite (Advanced Scatterometer/ASCAT). Sea ice edge data of 10 x 10 km² resolution data is developed by combining the data of the active satellite and the passive satellites using the Bayesian method. The pixels of the data are assigned with one-byte integer 1, 2, 3 and -1 which represent No ice (less than 30% ice concentration), Open ice (30 to 70 % ice concentration), Closed ice (more than 70 % ice concentration) and Unclassified or Land, respectively. Open ice and Closed ice are considered as sea ice and OSI sea ice extent is derived by combining these types.

The data sets are produced by the OSI-SAF of EUMETSAT and were downloaded from www.osi-saf.org.

All the data used in this study are in polar stereographic projection but they have different grid sizes. The SCATSAT-1 data set has 4001 columns and 4001 rows (the total area in the projection is 9002.25 x 9002.25 km²); NSIDC has 304 columns and 448 rows (the total area in the projection is 7600 x 11200 km²) and OSI has 790 columns and 1120 rows (the total area in the projection is 7900 x 11200 km²). A primary mask (Fig. 1) is created for this study using SCATSAT-1 data in combination with a sectorial vector file used by Bliss and Anderson (2014) (obtained through personal communication).

3. Methodology

3.1. Principal Component Analysis

Principal component analysis (PCA) is a very useful technique in reducing correlation redundancy in the multi-spectral remotely sensed image data (Lillesand et al., 2008). The mechanism is to extract information from sample data and to transform it into a new set of orthogonal data (Abdi and Williams, 2010). As compared to other well-known transformation methods (e.g. Wavelet transformation), PCA has a drawback of higher distortion of the original spectral information. However, it can express spatial characteristics very well and provide an optimal expression of highly anisotropic edges and contours in images (Shi et al., 2010). PCA has been used in the study of sea ice in the Arctic by Wensnahan et al. (1993) and in Antarctica e. g. by Singh et al. (2018).

In this study, input data for PCA are the SCATSAT-1 data sets of the three parameters mentioned earlier in both polarisations (horizontal and vertical). Therefore, in total there are 6 image data. PCA has been performed in IITVIS ENVI 4.5 using Forward PC rotation transformation with the Correlation matrix. In our analysis, only the first three principal components (PC) are retained to produce the false-colour composite (1st PC, 2nd PC, and 3rd PC as Red, Green, and Blue channel, respectively).

This study covers the period from October 2016 to March 2018. PCA is performed on only three specific days of each month. The selected days are in the beginning, in the middle and in the end of each month. There are no days with missing data and lack of principal components (number of inputs is less than 6) in

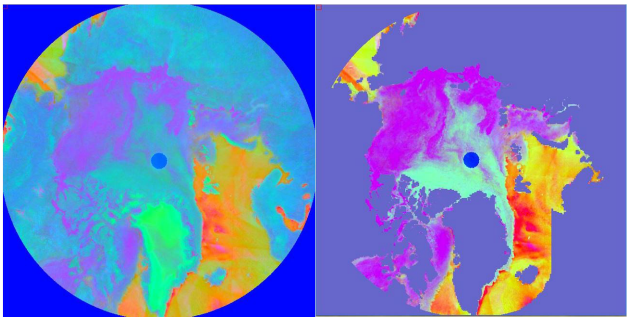

Figure 2: Unmasked image of PCA (left) and masked image of PCA (right).

the images picked for the study.

The result of an unmasked image after PCA and a masked image after PCA is shown side by side in Fig. 2. The unmasked and masked data are from the same day (1 January 2017). However, the colour of the two images is different. This is due to the increase in contrast to the masked image. There is no change in the data except the apparent colour. In the masked image, the violet part and the cyan part represent sea ice: the former representing first-year ice while the latter represents multi-year ice. Yellow, orange and red shades are possibly open oceans.

3.2. Unsupervised classification of image

As mentioned in the previous section, different features are visible in the PCA images. The false-colour composites formed using the first three principal components are then subjected to an unsupervised classification algorithm known as Iterative Self-Organizing Data Analysis Technique (ISODATA) to get a fair distinction between the open ocean and sea ice regions. Unsupervised ISODATA does not require a priori training data and classification proceeds by the formation of clusters using different cluster-means in the data space. Iteration is done with minimum distance technique and it reclassifies clusters and recalculates new means (Ball and Hall, 1965).

ESRI ArcGIS software is used for this classification. Application of PCA creates data sets in a range (standardised by mean = 0 and standard deviation = 1) and it is the requirement for ISODATA classification to yield a more accurate results. The number of output class is assigned to four, considering land (masked area), sea ice, open sea and central “pole hole”.

Iteration is performed until a threshold number of iteration (provided by software) is achieved. Four distinct classes were not obtained for some dates. The dates on which this problem was encountered are excluded from the analysis. Since the central pole hole is a sea ice area for the Arctic, this class is counted together with the surrounding sea ice class.

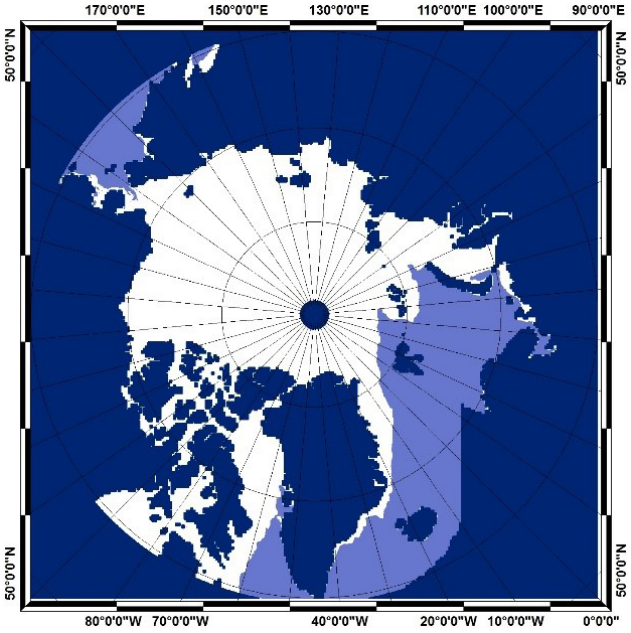

Figure 3: Sea ice extent obtained from the analysis.

4. Results

The sea ice extent map generated using PCA and unsupervised ISODATA classification from 1st of January 2017 is shown in Fig. 3. The white part of the map represents sea ice including the central hole, light blue represents the open sea and dark blue represents the masked part. A vast expansion of sea ice is also visible in the image. However further studies are needed to validate the method. Therefore, a comparison of sea ice extent was performed.

4.1. Qualitative comparison

Fifty-three days of SIE are available for comparison between the SCATSAT-1 and OSI data for the duration from October 2016 to March 2018. However, only 43 days of SIE are available for comparison between SCATSAT-1 and NSIDC during the duration from October 2016 to December 2017 due to unavailability of later NSIDC data at the time of this study.

SIEs of two selected days (1 January 2017 and 11 September 2017) derived from the above three sources are illustrated in Fig. 4. The month of January represents peak arctic winter while September represents peak summer melting season. Therefore, the growth and decay of sea ice should be visible in the figures.

The three derived SIE have shown a similar pattern in both summer and winter. However, in winter, NSIDC SIE is apparently more than SCATSAT-1 and OSI at the Bering Sea near Providence, Russia. However, at the Barents Sea along the coastline of Russia, the SIE is seen to be less in NSIDC < SACATSAT-1 < OSI.

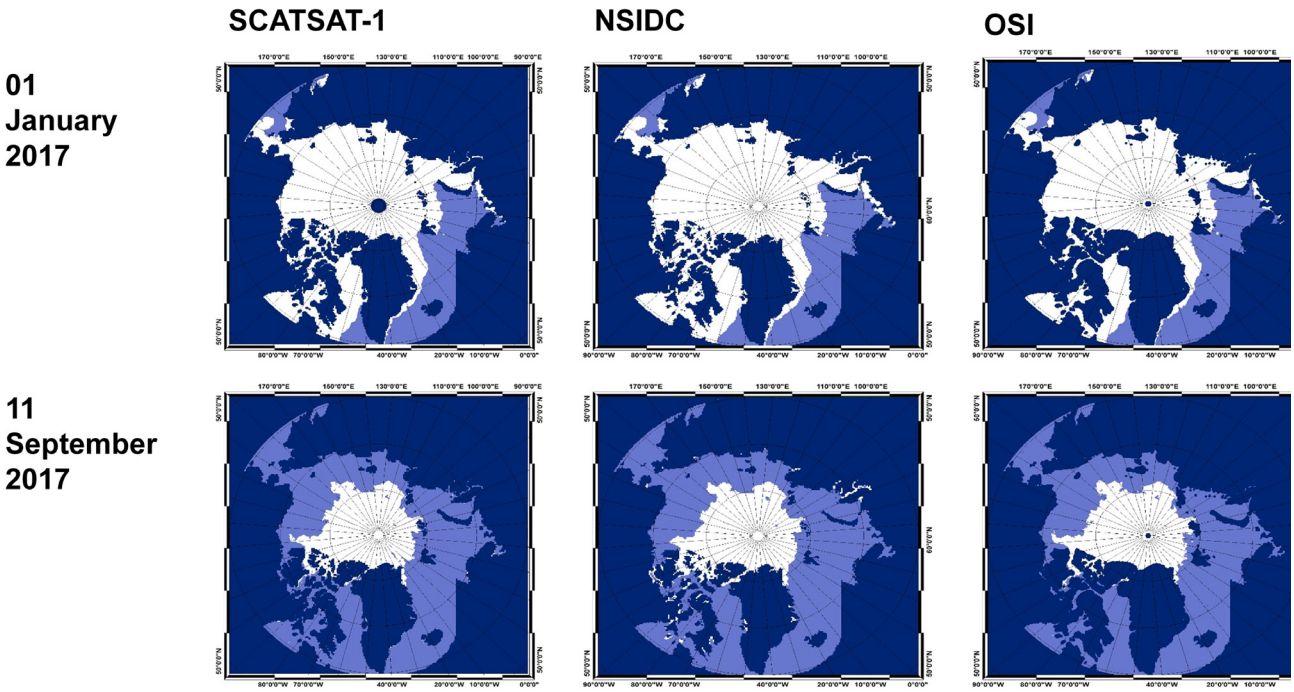

Figure 4: Qualitative comparison of SIEs.

During summer, in the NSIDC data some pixels along the coastlines of Greenland, the Canadian Archipelago and the Kara Sea are shown to be sea ice. But those are the open ocean in SCATSAT-1 and OSI.

4.2. Quantitative comparison

For quantitative comparison, SIEs inside the mask developed for this study are calculated for SCATSAT-1, NSIDC and OSI. For SCATSAT-1, the number of pixels classified as sea ice is counted and multiplied by 2.25 x 2.25 km^2 (resolution of the data) to calculate Arctic SIE. Similarly, the number of sea ice pixels are multiplied by 25 x 25 km^2 and 10 x 10 km^2 for NSIDC and OSI, respectively.

Temporal variations of three derived SIE are plotted in Fig. 5. A similar trend is observed despite the differences in these three extents. Our SCATSAT-1 SIE lies in between the NSIDC and OSI SIE curves. Thus, the observed SIE is in the order from large to small as NSIDC SIE, SCATSAT-1 SIE and OSI SIE.

The differences among the SIE are greatest in the period from January to April (winter to spring) and are the least during the period from August to October (summer to autumn). Between SCATSAT-1 and OSI SIE, the largest difference observed is ~$1.09 \cdot 10^6$ km^2 during the winter period while during the summer period, the largest difference observed is $0.57 \cdot 10^6$ km^2. These values for the SCATSAT-1 and NSIDC comparison are $0.91 \cdot 10^6$ km^2 and $0.66 \cdot 10^6$ km^2, respectively. Thus, the three estimations show convergence in summer rather than in winter. This should be investigated in our future work to understand why the difference smaller during summer than during winter.

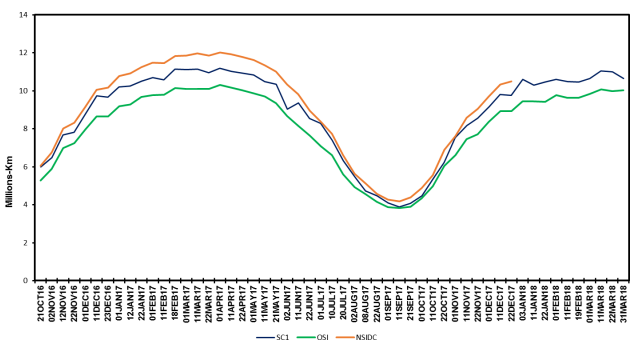

Figure 5: Temporal plot of SCATSAT-1 SIE, OSI SIE and NSIDC SIE.

A R^2 high value of 0.99 in the regression plot of Fig. 6 signifies the closeness between the data. It shows the similar variation of SCATSAT-1 SIE with NSIDC SIE and OSI SIE.

4.3. Student's t-test

We performed a statistical hypothesis test of comparing means of data samples, the null hypothesis (H_0) of which the samples have the same mean and variance (Kim, 2015).

Since data are unrelated, the independent type of a t-test is performed at a confidence level = 0.95 for both the SCATSAT-1/OSI comparison and SCATSAT-1/NSIDC.

The t-test parameters are shown in Table 1. We interpreted the result of this test using these parameters. If the obtained p-value is greater than present

Comparing datasets	Calculated t (t-stat)	Critical t (t-crit)	p-value
SCATSAT-1/OSI	1.8	1.983	0.07
SCATSAT-1/NSIDC	-0.97	1.988	0.33

Figure 1: Table 1: Result of t-test.

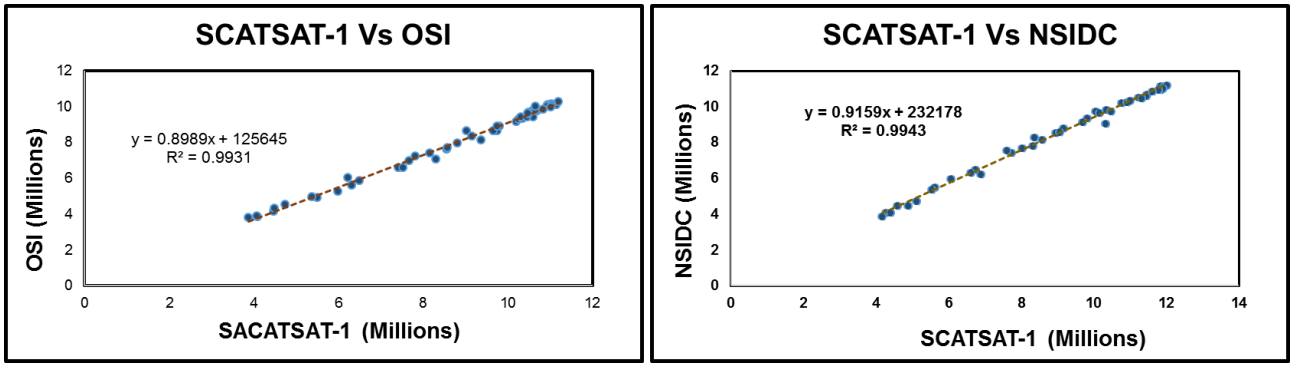

Figure 6: Regression plot between SCATSAT-1 SIE and OSI SIE (left) and SCATSAT-1 SIE and NSIDC SIE.

confidence level (95%), we have to accept the H_0 that the means that the data do not differ statistically. Another way is by considering the absolute value of the calculated t-value (t-stat). If it is less than t-crit (which is a standard value of the t value available in the standard table for given degrees of freedom and present confidence levels), we have to accept the null hypothesis.

In our comparison, we see that both the comparing data sets show statistically significant results.

5. Conclusion

An algorithm to estimate sea ice extent is developed for the Arctic region. High-resolution L4-data (2.25 km) of SCATSAT-1 data is used to produce the SIE. 54 days are selected to check the temporal variation of SIE and to compare with well-known SIE products.

The algorithm is developed by using an aesthetic combination of Principal Component Analysis (image redundancy reduction technique) and Unsupervised ISO classification (a popular image classification technique). SIE estimates from this method are found to be significantly comparable with the other popular data sets of SIE. Small differences may be attributed to the differences in techniques of estimations and the sensors employed. However, we will look further into this in our subsequent publications.

Comparisons between SIE derived from SCATSAT (1) versus derived SIE from NASA team sea ice concentration threshold at 30 % concentration provided by National Snow and Ice Data Centre (NSIDC), and (2) versus European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) SIE from Ocean and Sea Ice Satellite Application Facility (OSI-SAF) for the period October 2016 to March 2018 are performed. Comparison results show high correlation ($R^2 \sim 0.99$) at 95 % confidence level. Statistically, the comparing data sets show significant results.

Overall observations and comparisons confirmed the satisfactory of the estimation method. The SIE derived from the SCATSAT-1 will be useful to study the changes of arctic SIE throughout the life span of the SCATSAT-1 satellite. The method can also be applied to calibrated data of other scatterometer satellites with SCATSAT-1.

6. References

Abdi, H., and Williams, L. J., 2010. Principal component analysis. WIREs Computational Statistics 2, pp. 433–459. https://doi.org/10.1002/wics.101.

Ball, G. H., and Hall, D. J., 1965. ISODATA, A novel method of data analysis and pattern classification, Stanford Research Institute, Menlo Park.

Bliss, A. C., and Anderson, M. R., 2014, Daily Area of Snow Melt Onset on Arctic Sea Ice from Passive Microwave Satellite Observations 1979–2012. Remote Sensing. 6, pp. 11283-11314. https://doi.org/10.3390/rs61111283.

Budd, W. F., 1975. Antarctic sea-ice variations from satellite sensing in relation to climate: Journal of Glaciology 15, pp. 417–427. https://doi.org/10.3189/S0022143000034523.

Cavalieri, D. J., and C. L. Parkinson 2008. Antarctic sea ice variability and trends, 1979–2006. Journal of Geophysical Research Oceans 113, C07004. https://doi.org/10.1029/2007JC004564.

Cavalieri, D. J., Parkinson, C. L., Gloersen P., and Zwally, H. J., 1995, updated yearly. Sea ice concentration from Nimbus-7 SMMR and DMSP SSM/I-SSMIS passive microwave data Version-1, NASA National Snow and Ice Data Center Distributed Active Archive Center, Boulder, Colorado, USA . https://doi.org/10.5067/8GQ8LZQVL0VL.

Comiso, J. C., Cavalieri, D. J., Parkinson, C. L., and Gloersen P., 1996. Passive microwave algorithms for sea ice concentration: A comparison of two techniques. Remote Sensing of Environment 60, pp. 357–384. https://doi.org/10.1016/S0034-4257(96)00220-9.

Dan, L., and Robert, M., 2006. Polar remote sensing volume I: Atmosphere and Oceans. Springer, Berlin, 756 pp. https://doi.org/10.1007/3-540-30785-0.

Glossary of American Meteorological Society, Arctic bottom water, 2012. Available at http://glossary.ametsoc.org/wiki/Arctic_bottom_water [Accessed 20 January 2021]

Kim, T. K., 2015. T-test as parametric statistic. Korean Journal of Anesthesiology 68, pp. 540–546. https://doi.org/10.4097/kjae.2015.68.6.540.

Lake, R. A., and Lewis E. L., 1970. Salt Rejection by Sea Ice during Growth. Journal of Geophysical Research 75, pp. 583–597. https://doi.org/10.1029/JC075i003p00583.

Lillesand, T. M., Kiefer, R. W., and Chipman, J. W., 2008. Remote Sensing and Image Interpretation. 6th ed. Wiley, 756 pp.

National Snow and Ice Data Center, 2020. All About Sea Ice, Thermodynamics: Albedo. Available at https://nsidc.org/cryosphere/seaice/processes/albedo.html [Accessed 20 January 2021]

Remund, Q., and Long, D., 2014. A Decade of QuikSCAT Scatterometer Sea Ice Extent Data: IEEE Transactions on Geoscience and Remote Sensing 52, pp. 4281–4290. https://doi.org/10.1109/TGRS.2013.2281056.

Scatsat1 DP Team, 2017: SCATSAT-1 Level 4 Data Products Format Document, Microwave Data Processing Division/Sig-

nal and Image Processing Group, Space Applications Centre, Ahmedabad, ISRO Scientific Report- SC1/DP/L4FORMAT-DOC/V1.1/JUL2017. Available at ftp://ftp.mosdac.gov.in/docs/scatsat1_L4_data_products_formats_v1.1_jul_2017.pdf [Accessed 20.01.2021]

Shi, H., Tian, B., and Wang, Y., 2010. Fusion of Multispectral and Panchromatic Satellite Images using Principal Component Analysis and Nonsubsampled Contourlet Transform. In: 2010 Seventh International Conference on Fuzzy Systems and Knowledge Discovery. pp. 2312–2315. https://doi.org/10.1109/FSKD.2010.5569820.

Singh, R. K., Singh, N. K., Mamata, M., and Jayaprasad, P., 2018. Antarctic sea ice extent from ISRO's SCATSAT-1 using PCA and an unsupervised classification. Proceedings 2 (2nd International Electronic Conference on Remote Sensing), 340. https://doi.org/10.3390/ecrs-2-05153.

Wensnahan, M., Maykut, G. A., Grenfell, T. C., and Winebrenner, D. P., 1993. Passive Microwave Remote Sensing of Thin Sea Ice Using Principal Component Analysis. Journal of Geophysical Research 98, pp. 12453–12468. https://doi.org/10.1029/93JC00939.