MAPPING SUPRAGLACIAL STREAMS USING MAXIMUM LIKELIHOOD AND MAHALANOBIS DISTANCE CLASSIFICATION ON WORLDVIEW-2 DATA

Shridhar D. Jawak[1], Sandeep T.[2], Alvarinho J. Luis[1]
[1]Polar Remote Sensing Section, Polar Sciences Group, Earth System Science Organization (ESSO), National Centre for Antarctic and Ocean Research (NCAOR), Ministry of Earth Sciences (MoES), Government of India, Headland Sada, Vasco-da-Gama, Goa─403804, India,
Email: shridhar.jawak@gmail.com, alvluis1@gmail.com
2Indian Institute of Information Technology and Management, Kerala─695581, India.
sandeepvenugopal1211@gmail.com

KEYWORDS: Supraglacial streams; Optical satellite imagery; Bathymetry; Supervised classification

ABSTRACT

Supraglacial streams are key geographic features in Antarctica that prevent the disintegration of ice sheets and glaciers. Supraglacial streams also act as a mode of transportation for sediments, solutes and other biotic and abiotic features. This study focuses on geospatial mapping of supraglacial streams from Schirmacher Oasis in Antarctica using high resolution optical satellite imagery. This could help advanced researches on the stream networks such as bathymetry study, hydrological studies etc. Supervised classification methods; Maximum likelihood classification (MLC) and Mahalanobis distance classification (MDC) have been employed and accuracy has been assessed in different levels such as accuracy of few regions that undergone MLC, method-wise accuracy and finally length-wise accuracy assessment. It is seen that MLC is successful overall with average RMSE less than 5 m with a notable limitation of mapping only longer streams of an average length 350 m and above more accurately. Spectral mixing in the available data, presence of sediments, sun angle during data acquisition, streams being dry during most of the year etc., are some of the limitations faced during the study. A higher spectral resolution data or a combined classification system or both together can be employed for better results in the future.

1. INTRODUCTION

Supraglacial streams are the stream network on the surface of the glaciers, ice sheet ice shelves etc. These are formed due to solar radiation (60% – 70% by shortwave radiation) and air temperature. One of the major advantage of this stream networks, is it prevent the disintegration of glaciers, ice shelves, ice sheet etc. which otherwise would have been a reason for sea level rise globally. One example we can sight from the past is the disintegration of Larsen B shelf of Antarctica in 2002. Supraglacial streams are unique from other stream networks around the world by their characteristics like rapid rate of incision, change during the ablation season, due to the presence and influence of shifting moluins and glacial geomorphology defining stream structure and airflow. These streams also carry sediments and other biotic and abiotic entities, which plays a crucial role in defining various characteristics of Antarctica. This paper focuses on mapping supraglacial streams of Schirmacher Oasis in Antarctica, since global warming plays a huge part in melting the ice. Antarctic ice sheet has an area of 14,000 million square kilometre containing 30 million cubic kilometres of ice, which on melting would rise the sea level by 60 m globally. In addition, previous studies on supraglacial streams in Antarctica were only focused in McMurdo Dry valleys. Other regions are also needed to be studied. For most studies on supraglacial streams, only a certain number of supraglacial streams were selected. Availability of very high resolution data for the study is also a crucial factor. Previous studies (Steams et.al, 2008, Fricker et.al, 2010) mostly used data of 15 – 30 meter resolution images. Finding supraglacial streams on these images is challenging since such stream networks have width of 0.15m – 0.17m and in some regions it would go down to 0.005m. It is also difficult to distinguish the supraglacial streams networks in Antarctica, mainly for the reason they are really close to each other. Extracting and mapping these streams network using remote sensing techniques would give opportunities to study hydrology, bathymetry etc. of the stream network. Jawak et al, 2016 also studied various supraglacial features such as debris, streams etc. using WorldView-2 imagery, providing the most accurate geospatial mapping of supraglacial features altogether.

Various classification algorithms are employed in supraglacial feature extraction along with very high resolution images like WorldView–1, 2, 3, IKNOS, QuickBird etc. High accuracy of the process is one of the major reasons for preference of very high resolution data for Antarctic researches. Also most of the geographic features, especially supraglacial streams are distinguishable in such datasets only. This paper focuses on mapping supraglacial streams using supervised classification algorithms such as Mahalanobis distance classification (MDC) and Maximum likelihood classification (MLC) for extracting the streams from the image. Therefore, a comparative analysis of the accuracies of both classification algorithms is performed. In addition, the effects of stream length on classification accuracy is inferred.

2. STUDY AREA AND DATA

11 bit, Ortho-rectified, geometrically corrected and radiometrically corrected WorldView-2 (WV-2) image had been used for the study. DigitalGlobe launched WorldView–2 in 2009. In addition to the already existing bands, Blue, Green, Red, NIR-1 new bandwidth has been used in this generation (Red-Edge, Yellow, Coastal blue, NIR-2). WV–2 images are useful for various analysis in Antarctica (Jawak and Luis, 2014; Jawak and Luis, 2013a; Jawak and Luis, 2013b; Jawak and Luis, 2011a; Jawak and Luis, 2011b). In this study, images acquired over Schirmacher Oasis are analysed using various band combinations for better results. Schirmacher Oasis is the smallest oasis in Antarctica with an area of 25 square kilometre.

Table 1: WV-2 spectral characteristics

No.	Band name	Wavelength (Micrometers)
Pan	Pan	0.45-0.80
1	Coastal	0.40-0.45
2	Blue	0.45-0.51
3	Green	0.51-0.58
4	Yellow	0.585-0.625
5	Red	0.63-0.69
6	Red-Edge	0.705-0.745
7	NIR-1	0.77-0.895
8	NIR-2	0.86-1.04

It is also known as Schirmacher lake plateau. This region contains more than 100 fresh water lakes, one of the reason we believe stream networks are abundant here. Also a very rare research was done here previously on supraglacial streams (Jawak et.al, 2016). Pan-Sharpening was done using Panchromatic image provided by WV-2 satellite. Figure 1 shows the five regions selected for the study based on visual interpretation to identify the supraglacial streams.

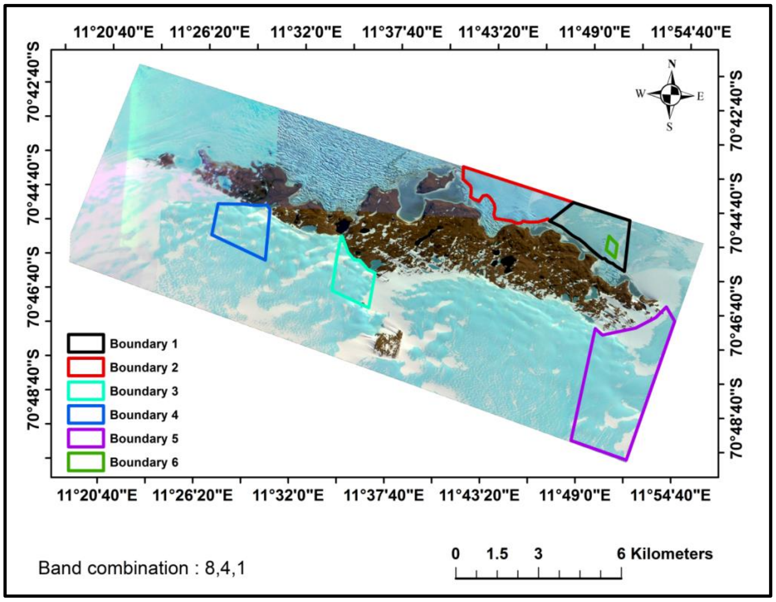

Figure 1. Location map of the study area showing subsets under present consideration.

3. METHODOLOGY

As shown in the Figure 2, the methodology starts with preprocessing which is Digital number (DN) values of raw images were converted into top of the atmospheric reflectance values by radiometric calibration, which uses specially developed conversion algorithm for WV-2 images. After mosaicking the images Pan-Sharpening was done using ENVI Gram-Schmidt Pan-Sharpening algorithm, which better serves for a very high resolution images like, WorldView series (Jawak & Luis 2013c). Pre-processing was finished after doing FLAASH atmospheric correction. Reference data was generated by digitizing and able to identify streams in the study area by using different band combinations. As we know Maximum likelihood and Mahalanobis distance classification algorithms are two successful algorithms from the past. Here we focused on which algorithm work perfectly on such an extreme terrain. We selected six subsets where streams were high in numbers and performed classification on those subsets by collecting training sites for the three classes that we found during visual interpretation. The classes identified are streams, fresh snow and blue ice. The number of training sites were fixed based on the total number of training sites, number of bands in the image and size of the image. Training site collection was done using ArcGIS 10.4 and the vector files were converted to ROI in ENVI and used for the classification purpose. Probability thresholds for each feature on the image to be in a particular class were assigned by trial and error method and a best fit was chosen accordingly. Likewise, MLC was applied on each of the six images with a threshold that best fit for that particular image. Different thresholds were assigned for different images, due to topography, surface reflectance, variation in atmospheric gasses etc. The same process was repeated for the Mahalanobis distance classifier. Once classification process is completed in ENVI, the results are exported to shapefile. In order to find accuracy we estimated the Root Mean Square Error (RMSE) by using the formula,

$$\mathrm{RMSE} = \sqrt{\frac{(\mathrm{BIAS})^2}{\mathrm{N}}}, \text{ where } \frac{(\mathrm{BIAS})^2}{\mathrm{N}} \text{ is the Mean Square Error (MSE)} \qquad (1)$$

To find the Bias we needed to find length of the extracted streams and that of reference streams. For that we used ArcGIS 10.4 software, where we got the length of the reference streams in meters (Digitized streams) from attribute table by using calculate geometry. We calculated accuracy at three different levels; Method-wise, Class-wise and comparison between MLC and MDC. Then from each subset, depending on total number of streams representative streams were selected. We found length of extracted and reference streams of these representative streams. Then to calculate bias we followed the equation; Bias = Length of reference streams – Extracted stream. One of our objectives was to find if length of the streams has any impact on classification accuracy. For that in each subset streams were selected in different classes of length (0-100, 100-200 etc.) in a number and accuracy was calculated for each classes of length for every five subset.

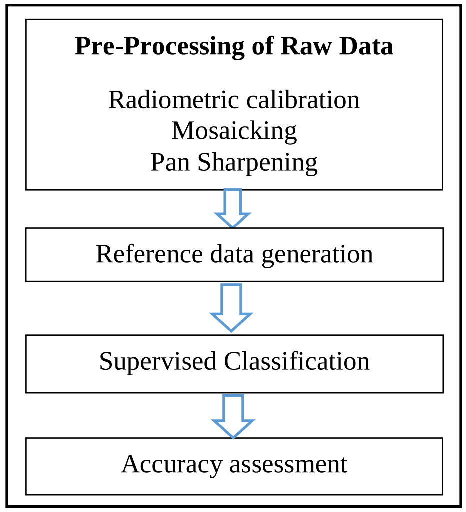

Figure 2: Methodology Protocol

4. RESULTS AND DISCUSSION

As we studied our results, we reached innovative inferences. In the MLC accuracy assessment, all except one subset had better accuracies compared to each other. Subset 1 as explained in table 2 has an accuracy of 10.83 m and almost all the selected streams have lengths between 200-400 m. Table 4. Measures of accuracy for each classification scheme. We believe the reason for lesser accuracy is the presence of other supraglacial features such as debris and the spectral mixing in the WV-2 imagery available for the region. Subset 4 and 5 had sub-meter level accuracy, Subset 4 with 0.24m and subset 5 with 0.53m. Only two streams were overestimated in subset 1. The subset 4 which had the highest accuracy for MLC method and surprisingly 80% of all the streams in subset 4 was comparably shorter streams (<200m).

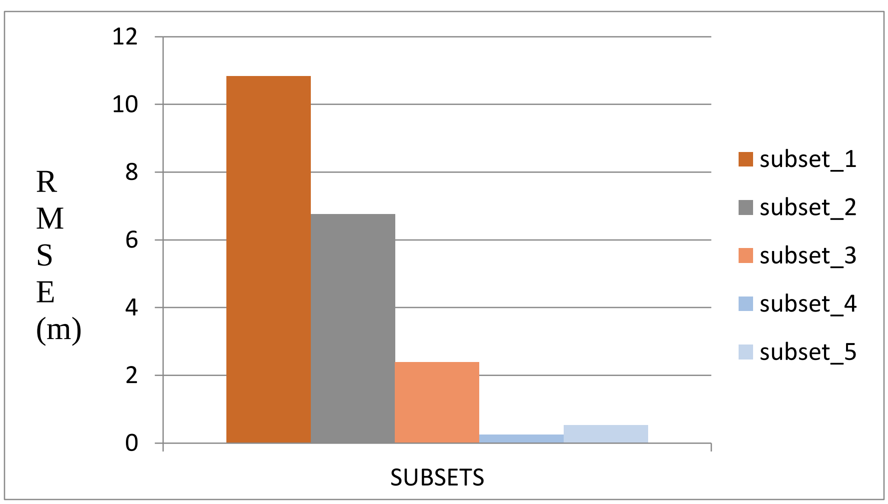

Figure 3: Graphical representation of Accuracy of MLC method

Table 2: RMSE values for MLC method

Subset	Sr. No.	Reference Streams Length (meter)	Extracted Streams Length (meter)	Average RMSE
Subset 1	1	141.10	142.87	10.83
	2	216.82	205.56	
	3	321.82	319.63	
	4	414.12	410.69	
	5	483.21	474.54	
	6	284.41	283.12	
	7	156.68	151.73	
	8	244.50	243.29	
	9	192.89	190.84	
	10	268.25	267.26	
Subset 2	1	247.56	247.50	6.76
	2	242.75	239.62	
	3	340.04	337.35	
	4	455.43	448.30	
	5	177.27	173.75	
	6	160.46	160.09	
	7	1168.62	1150.33	
	8	274.00	268.48	
	9	139.18	135.67	
	10	330.45	330.04	
Subset 3	1	223.87	220.94	2.40
	2	497.14	497.95	
	3	124.86	124.80	
	4	223.87	220.15	
Subset 4	1	52.96	52.45	0.24
	2	35.02	34.95	
	3	61.65	61.55	
	4	188.77	188.60	
	5	102.54	102.51	
Subset 5	1	223.04	222.95	0.53
	2	224.73	224.69	
	3	361.25	360.14	
	4	305.95	306.41	
	5	208.56	208.40	
	6	401.63	400.76	
	7	146.74	146.01	
	8	232.55	232.33	
	9	245.51	245.43	
	10	380.06	380.01	

Overall error percentage of the MLC method is 4% and for the subset on which MDC was carried out, error percentage is 2.5%. For the MDC method extraction was also better which effected on the result (Table 4). MDC given a par accuracy only on one subset. On all other subsets, it was below par. Accuracy assessment was carried out on that subset and it showed far better accuracy than MLC on that set. Class-wise accuracy (Based on length) showed that length of the streams is an important factor in extracting the supraglacial streams more accurately. As the length of the streams increases, more accuracy is achieved. Table 4 shows the results of accuracy assessment carried out on subset 1 in the above-described way. As we have expected, shorter streams have lesser accuracy than the longer one. Moreover, we were able to see that shorter streams was negatively biased. Which in turn shows that over estimation was high for the shorter streams and extraction was poorly done for such streams. Figure 4 shows all the resultant extracted images.

Table 3: Method-Wise RMSE comparison

Method	Sr. No.	Reference Streams	Extracted Streams	RMSE
MDC	1	177.31	178.13	2.55
	2	114.73	114.73	
	3	254.19	247.78	
	4	214.98	214.9	
	5	114.73	116.45	
	6	325.44	326.41	
	7	134.03	133.91	
MLC	1	325.44	321.81	14.16
	2	103.69	107.42	
	3	51.21	47.36	
	4	34.70	36.14	
	5	114.73	119.23	
	6	45.07	47.40	
	7	85.84	74.40	

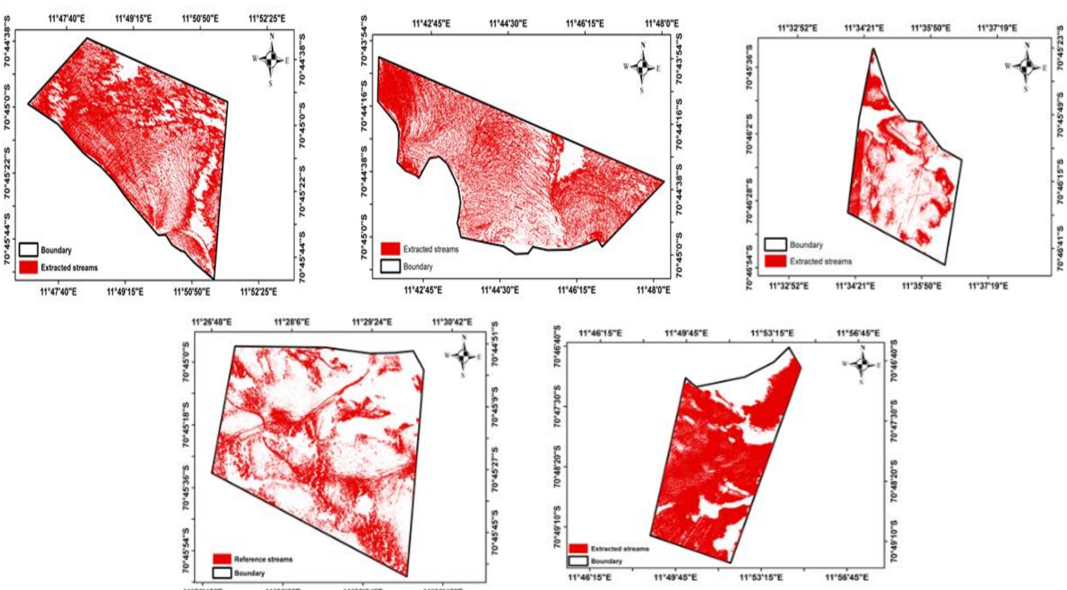

Figure 4: Extracted streams (from the top left corner Subset 1 to subset 5)

5. CONCLUSION

Monitoring the supraglacial streams for studying about their hydrological parameters, geomorphological parameters etc. need a geospatial map of the same. We were able to contribute to that with Maximum Likelihood Classification giving only 4% total error and Mahalanobis distance classification with only 2.5% RMSE error percentage on the accuracy assessment. Even though Mahalanobis distance classification had higher accuracy on the subset, since it prevailed only on that subset we suggest to use MLC method for better results. Some challenges we faced during our research is, disability of the data to distinguish between blue ice and streams in some regions due to spectral mixing, disability of the data to scan dry streams etc. What we propose for future researches is to use a combined classification system that could encompass the morphological characteristics of the supraglacial streams and a multi-synergic approach would give better results even on the shorter streams.

ACKNOWLEDGEMENTS

The authors would like to thank DigitalGlobe for providing the imagery used in this study. We also acknowledge Dr. M. Ravichandran, Director, ESSO-NCAOR, for their encouragement and motivation for this research.

REFERENCES

Alger, et al., (1997). Alger et al.pdf. Institute of Arctic and Alpine Research.

Blaschke, et al., (2014). Geographic Object-Based Image Analysis - Towards a new paradigm. ISPRS Journal of Photogrammetry and Remote Sensing, 87, 180–191.

Congalton, R. G. (1991). A review of assessing the accuracy of classifications of remotely sensed data. Remote Sensing of Environment, 37(1), 35–46. Derenyi, et al., (n.d.). Researchers at the Department of Surveying Engineering , 976–978.

Du, et al.,. (2012). Multiple classifier system for remote sensing image classification: A review. Sensors, 12(4), 4764–4792.

Fricker, et al., (2010). Synthesizing multiple remote-sensing techniques for subglacial hydrologic mapping: Application to a lake system beneath MacAyeal Ice Stream, West Antarctica. Journal of Glaciology, 56(196), 187–199.

Hussin, , & Atmopawiro,. (2004). Sub-pixel and maximum likelihood classification of Landsat ETM+ images for detecting illegal logging and mapping tropical rain forest cover types in Berau, East Kalimantan, Indonesia. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences - ISPRS Archives, 35.

Jawak, S.D. and Luis, A.J., 2011a. Applications of WorldView-2 satellite data for extraction of polar spatial information and DEM of Larsemann Hills, East Antarctica. In: International Conference on Fuzzy Systems and Neural Computing, Vol. 3, pp. 978-1.

Jawak, S.D. and Luis, A.J., 2011b. High Resolution 8-Band WorldView-2 Satellite Remote Sensing Data for Polar Geospatial Information Mining and Thematic Elevation Mapping of Larsemann Hills, East Antarctica. East Antarctica. 11th ISAES-2011, pp.10-16.

Jawak, S.D. and Luis, A.J., 2013a. A Spectral Index Ratio-Based Antarctic Land-Cover Mapping Using Hyperspatial 8-Band WorldView-2 Imagery. Polar Science, 7, 18-38.

Jawak, S.D. and Luis, A.J., 2013b. Improved land cover mapping using high resolution multiangle 8-band WorldView-2 satellite remote sensing data. Journal of Applied Remote Sensing, 7(1), pp.073573-073573.

Jawak, S.D. and Luis, A.J., 2013c. Very high-resolution satellite data for improved land cover extraction of Larsemann Hills, Eastern Antarctica. Journal of Applied Remote Sensing, 7(1), pp.073460-073460.

Jawak, S.D. and Luis, A.J., 2014. A semiautomatic extraction of Antarctic lake features using WorldView-2 imagery. Photogrammetric Engineering & Remote Sensing, 80(10), pp.939-952.

Jawak., et al., (2016). Semi-automatic extraction of supra-glacial features using fuzzy logic approach for object-oriented classification on WorldView-2 imagery, (April), 98801S.

McKnight, et al.,. (1999). Dry Valley Streams in Antarctica: Ecosystems Waiting for Water. BioScience, 49(12), 985.

Qian, et al., (2005). Spatial Contextual Noise Removal for Post Classification Smoothing of Remotely Sensed Images. Proceedings of the 2005 ACM Symposium on Applied Computing, 524–528.

Sanclements et al., (2016). Biogeophysical properties of an expansive Antarctic supraglacial stream, 12(October), 1–12.

Stearns et al., (2008). Increased flow speed on a large East Antarctic outlet glacier caused by subglacial floods. Nature Geoscience, 1(12), 827–831.

Thomas M. Lillesand, Ralph W. Kiefer, J. W. C. (1989). Remote Sensing and Image Interpretation. Journal of Chemical Information and Modeling (Vol.53).

Tso, B ., & Mather, P. M. (2009). Classification methods for remotely sensed data. Methods.

Welch, et al., (2010). Spatial variations in the geochemistry of glacial meltwater streams in the Taylor Valley, Antarctica. Antarctic Science, 22(6), 662–672.

Yi, X et al., (2007). A Knowledge Based Classification for Urban Mapping Using High Resolution Remote Sensing Data. Digital Image Computing Techniques and Applications, 586–591.