On water and ice classification from Sentinel-2 imagery using machine learning

Rémi Jugier[1], Robin Cremese[1], Hugo Fournier[1], Nuria Duran-Gomez[1], Germain Salgues[1], Chloé Thenoz[1]

[1]MAGELLIUM, Ramonville Saint-Agne, 31520, France

Correspondence to: Rémi Jugier (remi.jugier@magellium.fr)

Abstract. Accurate and dynamic mapping of water and ice surfaces is directly useful to navigation and lake ice cover monitoring to study climate change. Water and ice maps are also useful for various scientific applications such as atmospheric correction of satellite imagery, remote sensing of water quality, and as input data for hydrological, weather and climate models. The existing literature shows that multi-spectral satellite imagery, as provided by Sentinel-2 and Landsat-8, provides a very effective means to discriminate between water, land, and ice. However, most studies focus either on very specific cases (a specific lake for instance), or on general cases but without complex and yet very frequent cases such as turbid waters and salt lakes which can be confused with snow and ice. The Copernicus High-Resolution Snow and Ice Monitoring Service provides an operational Sentinel-2 ice and water classification product at 20m resolution but with a lot of confusion on the aforementioned cases. Using a database of 31 fully hand-labelled Sentinel-2 L2A atmospherically corrected images, and machine learning SVM and RandomForest methods, the current study shows that the classification of land, water, ice, snow, turbid waters, salt lake categories can be achieved with an accuracy over 93%. It is also shown that the atmospheric correction has little to no impact on the results, as training and evaluating from L1C top of the atmosphere images instead of L2A images yields very similar results. This last find is very useful as it means that very accurate surface masks can now be provided to atmospheric processors and may therefore considerably improve the quality of atmospherically corrected images when compared to the current usage of static masks.

1 Introduction

The monitoring of the ice cover over rivers, lakes and seas serves both practical use cases and scientific applications. Indeed, the near real time detection of freezing and thawing events can help regulate mainly naval routes, which are greatly impeded by ice formation. And they can also be useful to regulate land routes on frozen lakes, which are commonplace in northern countries. The monitoring of lake ice is also of great interest to climatologists as it plays a role in the water cycle and surface energy flux, and the global lake ice cover is listed as one of the 50 Essential Climate Variables defined by the Global Climate Observing System (GCOS). Optical and radar satellites are currently the only means for detecting water and ice on a global scale (see review by Duguay et al. 2015) with a reasonably high revisit rate (up to 5 days for Sentinel-1 and Sentinel-2 for instance).

There has been more work in the literature on ice classification using C-band SAR satellites (Radarsat, Sentinel-1, ERS,...) than optical data, because they are not impacted by clouds, shade or polar nights, and are therefore better suited to provide a water/ice mask reliably every few days. The main scientific variables in lake ice monitoring are the freeze-up and break-up days, defined as the first days where the lake is respectively fully frozen and fully water, for two consecutive days. Towards this aim, satellite imagery lacks the temporal accuracy, mainly due to being unable to reliably observe the ground at each cycle, because of the cloud cover. The basis for water and ice discrimination using SAR data is that, at non zero incidence angles, water reflects the radar beam away from the satellite and therefore yields a very low backscatter. On the other hand, ice is usually a much more diffusive material, due to its often fragmented surface, irregularities of the water/ice interface, and the small air bubbles that are contained in the ice (see Nolan et al. 2002). Nghiem and Leshkevich 2007 define, using only HH and VV polarizations from Radarsat and ERS-2, for different ice types, and calm water, backscatter thresholds depending on the satellite's SAR beam angle of incidence. Using those thresholds, Leshkevich and Nghiem 2007 are able to detect and classify different ice types on Lake Superior (USA, Canada) in low wind conditions. Windy conditions are however a major issue for ice detection, as it creates ripples and waves which are much more diffusive than calm water and therefore yield high backscatters that are comparable to ice (see examples in Nolan et al. 2002). Geldsetzer et al. 2010 add thresholds on HV or VH orthogonal polarizations to classify the ice, and show that their values are less affected by wind, but that the signal to noise ratio is often too low for the orthogonal polarizations to be used. They are however able to correctly identify freezing and thawing phases using simple thresholds on HH and HV polarizations. Sobiech et Dierking (2013) obtain similar results with a "k-means" machine learning method. However, both Geldsetzer et al. 2010 and Sobiech et Dierking (2013) note that wind is a major source of water/ice confusion, and that this issue is not resolved by their respective methods. Leigh et al. 2013 and Zakhvatkina et al. 2017 associate texture information to a SVM method and show that they are able to discriminate water and ice even in moderate wind conditions (up to 5m/s). However, they again note that classification errors still occur due to the water roughness in high wind conditions. They also note that grease ice (ice mixed with melted ice or wet snow) is often confused with water. Indeed, C-band microwaves penetrate dry snow up to 20m deep, while only 3% of liquid water reduces this penetration to less than a wavelength i.e. around 5cm (see Rott et Nagler 1994, Nagler et al. 2016), thereby yielding a much lower radar backscatter, which is comparable to open water. Very recently, Tom et al. 2020 have used a deep learning approach (DeepLabV3+ Convolutional Neural Network (CNN)) to classify water and ice over three Swiss lakes, using webcam and Sentinel-2 observations to provide a ground truth for the training phase, and for the evaluation of the network's performance. Taking advantage of the spatial correlation of information, a CNN should be able to interpret texture and contextual information with much more complexity than machine learning methods, and improve classification in windy conditions. However, since the authors were not able to generate sufficiently reliable ground truths during transition phases — i.e. when the lakes were not fully water or fully ice — they only trained and performed the quantitative analysis on non-transition phases. The scope of their results is therefore limited, and the qualitative example they provide on the partially frozen Sils lake tends to confirm this limit as the prediction is not very accurate. Furthermore, the authors do not evaluate the performance of the classifier with respect to wind conditions. Overall,

state of the art water/ice classification using SAR satellite data is currently unreliable in windy conditions and during the transition phases (under-detects thin ice and wet ice). It also requires an a priori static water mask as it cannot discriminate between land and water, and is thus unable to take into account the temporal variations in river and lake surface extents.

Multi-spectral satellite imagery, as provided by Sentinel-2 and Landsat-8, provides a very effective means to discriminate between water, land, and ice. In fact, a human operator can, in the vast majority of cases, discriminate between water, land and ice, looking only at the RGB bands. This is why satellite imagery is generally used to provide ground truth for training classifiers using SAR data as done by a human operator in Tom et al. 2020. However, because of the aforementioned unreliability of satellite optical data for accurately tracking the lake freeze-up and break-up dates, it has seen much less consideration in literature than SAR. The Copernicus High Resolution Snow and Ice Monitoring Service (Copernicus HR-S&I) provides near real time ice products over the EU-Hydro water mask on continental Europe, using Sentinel-2 data. They use a k-means algorithm trained on a limited number of water and ice polygons from a single Sentinel-2 L2A image (see HR-S&I Sentinel-2 River and Lake Ice Extent Algorthm Theorical Basis Document issue 2.5 at https://land.copernicus.eu/user-corner/technical-library/hrsi-ice-atbd). Those products are generally good over thick ice and clear water, but largely over detect ice over turbid waters and salt lakes, and under-detect thin ice (black ice). Barbieux et al. 2018 use a two-step classification process. They first separate water surfaces, with or without ice, from land, using a threshold method on an index they define for this purpose : Icy Lake Index (Red + SWIR2) / (NIR + SWIR1). They then use a RandomForest approach to separate the water from the ice, using all bands as well as texture information (gradient and standard deviation). Their validation is done on six manually labelled Landsat-8 images which contain deep water, shallow water, opaque ice (thick ice) and clear ice (thin ice). They have however not included cases of turbid waters which are a major cause of ice/water confusion. They have also focused only on large lakes and therefore do not include rivers, seas, small mountain lakes, or salt lakes, all of which could cause issues to their method due to different ice types, turbidity, the presence of snow, salt lakes/ice/snow confusions. Polyapram et al. 2019 have evaluated the performance of deep learning methods on Landsat-8 optical data, using only full ice patches that were manually selected, and full water/land patches labelled automatically using the Pekel water mask (Pekel et al. 2016). For the water/land patched, they excluded all uncertain pixels, i.e. pixels with Pekel water occurrence between specific thresholds. Unfortunately, they do not provide those specific thresholds. The three CNNs used are U-Net, DeepWaterMap (Isikdogan et al. 2019), and a home-made network with no max pooling to preserve pixel-accuracy. The qualitative assessment over four Landsat-8 images shows that the U-Net and DeepWaterMap network fail to adequately classify water and ice over very large areas. However, the quantitative results for those networks are nearly flawless. It seems that, like for Tom et al. 2020, since their training and evaluation database only contains very simple automatically labelled cases, the performance statistics computed are not representative at all of the real performance of the algorithm on generic cases. Their home-made network achieves a much better classification according to their visuals. However, like the other networks, the quantitative results do not capture these improvements well. Furthermore, as we mentionned before, problematic cases for water/land/ice discrimination are turbid waters, salt lakes, and thin ice, and the authors do not mention any of those cases, whether in the quantitative or qualitative evaluation.

The aim of this paper is to provide a thorough analysis of the potential of pixel-based machine learning methods to classify water, ice, snow, and land using Sentinel-2 L2A products (ground-leaving radiance), using a human-operator labelled database, on a variety of cases including cases that are known to cause class confusion such as turbid waters, salt lakes, and thin ice. Section 2 details the construction of the database, and the machine learning methods used. The results are presented in section 3. Accurate water, snow and ice masks are very useful to atmospheric processors. Therefore, the performance of the algorithms using L1C top of the atmosphere radiance products instead of L2A products is evaluated. However, L2A binary cloud masks are still applied to exclude cloudy pixels from the study. Those results are presented in section 4.

2 Method

A database composed of 31 fully classified Sentinel-2 L2A products is used for the training and evaluation. L2A products were processed from Sentinel-2 L1C images using MAJA v4.5.3 atmospheric processor, cloud detector (https://gitlab.orfeo-toolbox.org/maja/maja/-/tree/release_v4.5.3), and the Copernicus global DEM at 30m resolution (COP-DEM_GLO-30-DGED dataset). The database contains standard thick ice and dark water cases but also a large variety of turbid waters, thin ice cases, and salt lakes, i.e. all cases that are cause for confusion between water and ice or land and snow for instance. Figure 1 shows the locations of the tiles used for the database, as well as the sun zenith angle for each corresponding Sentinel-2 image. The sun zenith angle is the best indication of warm/cold seasonal conditions i.e. the higher the angle, the colder the weather. In the case of 33VVF (lake Vanern, Sweden) and 48UVC (lake Baïkal, Russia), two images were used for each, and all four images have a measurement date in February. All other locations are present in the database only once.

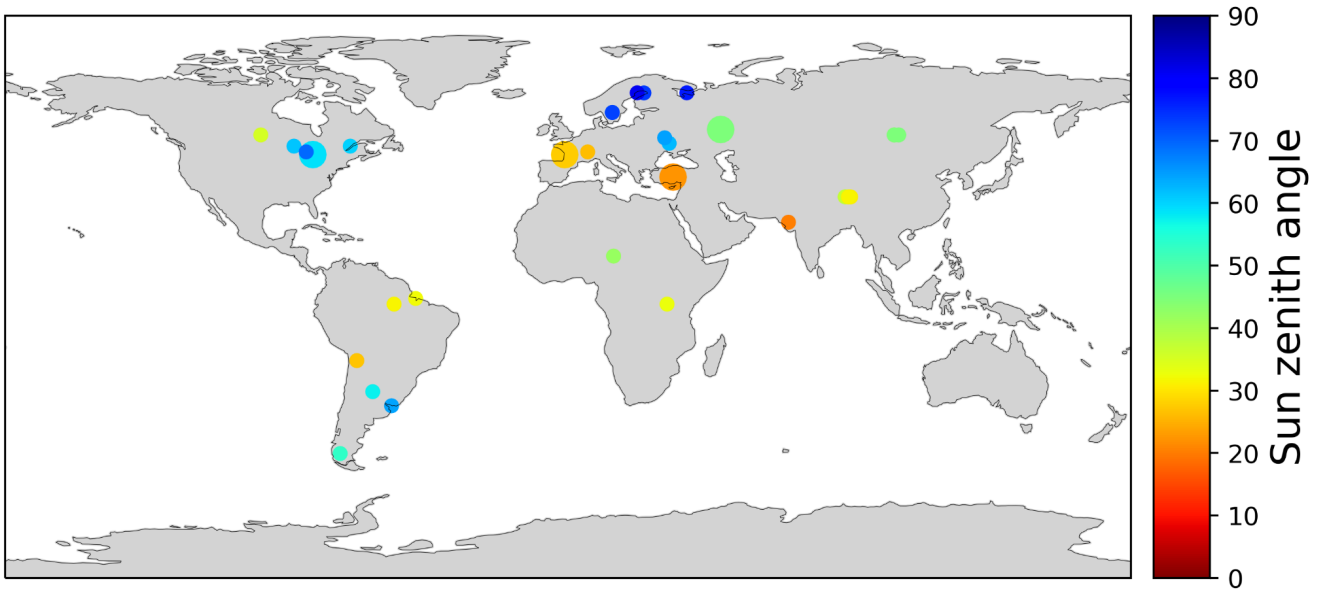

Figure 1. Locations and sun zenith angle of the labelled Sentinel-2 images used for the training and test set. Images selected for the test set are displayed with a larger circle.

A supervised classification tool (ALCD : https://github.com/CNES/ALCD) was used for the initial labelling of the 31 images composing the database. To label an image, the operator first chooses a few dozen of pixels that are selected in QGIS (ALCD uses point shapefile inputs), then runs the ALCD tool that uses a pixel-based RandomForest method to provide an automated labelling of all the pixels in an image at 60m resolution. The operator then iteratively adds new pixels to correct labelling errors from the previous steps. When the operator considers that there is no significant improvement between two successive steps, the resulting classification is reprojected to a 20m resolution, and the operator manually refines and corrects ALCD labelling errors, combining visual information from the true colour RGB image (B2 = Blue, B3 = Green, B4 = Red), the DEM, the Fractional Snow Cover product (https://land.copernicus.eu/user-corner/technical-library/hrsi-snow-pum) for snow labelling, and false colour red/red-edge (B4=Red, B5=Red-edge 1, B6=Red-edge 2) which helps for turbid waters' labelling. The different Sentinel-2 bands, their wavelength, and their resolution are listed in Table 1. Labels included into the database are : ice, snow, water, turbid water, salt lake, land (which includes land, vegetation and human constructions), cloud/cloud shadow, and no-data. Clouds and cloud shadows are a union of the clouds labelled using ALCD and the MAJA cloud mask. All clouds and cloud shadows are excluded from the study.

Band	**Wavelength (nm)**	**Resolution (m)**

B1	443	60
B2	490	10
B3	560	10
B4	665	10
B5	705	20
B6	740	20
B7	783	20
B8	842	10
B8A	865	20
B9	940	60
B10	1375	60
B11	1610	20
B12	2190	20

Table 1. Sentinel-2 bands. L2A products only contain 10m and 20m bands which are highlighted in green.

Two pixel-based machine learning methods are evaluated: Support Vector Machine (SVM) and Random Forest. Both methods were implemented via the scikit-learn Python library, respectively using "sklearn.svmLinearSVC" and "sklearn.ensemble.RandomForestClassifier" classes. The class_weight='balanced' option is used for both, which effectively applies a 1/n coefficient for each class where n is the number of pixels for a specific class. This is necessary to ensure that all classes have the same weight in the learning process. Option max_iter=100K is applied for SVM to allow for the algorithm convergence. Options min_samples_leaf=0.001 and n_estimators=20 are used for RandomForest. The min_samples_leaf option limits the creation of new leafs to those that contain at least 1/1000 of the data. This effectively limits overfitting issues for RandomForest. A higher number of estimators were tried (up to 100), without any noticeable difference to the results, so the n_estimators=20 option was chosen to improve computing time at inference.

All the Sentinel-2 bands from the MAJA L2A products are used (see highlighted bands in Table 1). The bands at 60m resolution are used solely for atmospheric correction. They are therefore not present in L2A products and are not used for the learning process. They will however be used in section 4 for classification using top of the atmosphere L1C products.

We write the normalised difference operator for 2 bands : $ND(B_n - B_m) = \frac{B_n - B_m}{B_n + B_m}$. Several indices, listed in Table 2, are used and included in the learning data. The sun zenith angle σ_s is also included in the learning data, to give additional information to the algorithm about the global scene illumination.

Normalised difference index	Formula
Water index (NDWI)	ND(B3, B8)
Snow index (NDSI)	ND(B3, B11)
Vegetation index (NDVI)	ND(B3, B8)
Vegetation index 2 (NDVI2)	ND(B3, B8A)
Chlorophyll index (NDCI)	ND(B4, B5)
Bare soil index (BSI)	ND(B11+B4, B8+B2)
Vegetation moisture index (NDMI)	ND(B8, B11)
Vegetation moisture index 2 (NDMI2)	ND(B8A, B11)
Blue - Red	ND(B2, B4)
Green - Blue	ND(B3, B2)
Green - Red	ND(B3, B4)

Table 2. Sentinel-2 bands. L2A products only contain 10m and 20m bands which are highlighted in green.

The training set is composed of 270K random selected pixels from 27 Sentinel-2 images (10K pixels per image). A first test set (TS1) is composed of all the pixels from the 31 Sentinel-2 images, excluding the 270K pixels used for training the classifiers. A second test set (TS2) is composed of all pixels from 4 distinct Sentinel-2 images from which no pixels were used for training the classifiers. Indeed, as pixels from the same images tend to have similarities, evaluating the classifiers against a fully independent test set is a more robust validation. Figure 1 displays with a larger border the 4 tiles selected for TS2.

3 Results using bottom of the atmosphere L2A products

The confusion matrix, recall, precision and accuracy are presented on Figure 2 for the SVM method for TS2. Figure 3 represents the same results for the RandomForest method. The overall accuracy is 94.4% for SVM and 91.3% for RandomForest. Most confusions are between similar classes : water / turbid water, snow / ice, and land / salt sea. If we merge those classes, the overall accuracy rises to 96.80% for SVM and 95.57% for RandomForest. However, evaluating the performance on the TS1 set – i.e. all 31 Sentinel-2 images with the 270K pixels from the training set removed – is useful because it contains a more diverse number of cases. On TS1, the accuracy is 89.33% for SVM and 90.23% for RandomForest (respectively 92.84% and 93.46% when similar classes are merged).

A more detailed analysis of the statistics displayed on Figure 2 and Figure 3 shows that the ice is very well detected by both methods, with a recall of 87.71% for SVM and 92.56% for RandomForest. However, the slightly higher RandomForest recall comes with a tradeoff as the precision for the ice class is 85.07%, which is lower than SVM's 94.36%. Figure 4 shows the classification results for an image of the TS2 test set, i.e. no pixels from this image were used for training. While the SVM results are very well in line with the ground truth, RandomForest results indeed show commission errors for ice over water. Figure 4 also shows that the thin ice is very well detected by both classifiers.

Figure 2 and Figure 3 show that the confusion between the snow/ice classes, and the salt sea class is extremely low, under 0.01% commission and omission errors. Since salt sea pixels and snow pixels reflect light very similarly (with a slightly higher snow reflectance on average) across the full spectrum of Sentinel-2 bands, it is a surprisingly good result that the classifiers manage to discriminate between those categories so well. Figure 5 indeed shows nearly no snow or ice commission errors from both algorithms on the Taz Golu desert in Turkey.

As mentioned previously, there is a lot of confusion over test sets TS1 and TS2 for similar classes : water / turbid water, land / salt lake, and ice / snow. This behaviour is expected as those classes are very similar, with continuous transitions, and the class annotations in the training / test sets are not very accurate or homogeneous between different images. It is important to note that it is not of any interest to the study to be able to discriminate between salt sea and land, or water and turbid water, so these confusions between similar classes are not problematic. Indeed, the salt sea and turbid water subclasses were added mainly to assess snow and ice commission errors on those areas, because they are identified as a major source of confusion with snow and ice in the Copernicus HR-S&I service snow and ice products.

Figure 2 and Figure 3 show that ice commission errors over turbid waters is very low for both algorithms. Figure 6 indeed shows no ice commission errors over the Gironde estuary in France, which is always very turbid.

To sum up, all classes are detected very well. Most notably, the salt lake and turbid water surfaces that usually present the greatest challenge for snow and ice detection have very low commission errors for snow and ice. The only confusions are between classes that are either not very important to discriminate and that are therefore sometimes loosely defined in the database.

True \ Prediction	Water	Turbid Water	Ice	Snow	Salt sea	Land	**Recall**
Water	13.38	0.61	0.24	0.00	0.00	0.04	93.78%
Turbid Water	0.00	1.16	0.00	0.00	0.06	0.01	94.70%
Ice	0.64	0.05	7.26	0.15	0.00	0.18	87.71%
Snow	0.01	0.00	0.17	4.69	0.00	0.04	95.50%
Salt sea	0.00	0.00	0.00	0.00	1.72	0.01	99.16%
Land	0.05	0.12	0.03	1.85	1.54	69.38	95.08%
Precision	95.05%	59.78%	94.36%	70.12%	51.63%	99.60%	94.40% (accuracy)

Figure 2: Confusion matrix (values in megapixels), recall, precision, and accuracy for SVM method over the TS2 test set, using top of the atmosphere Sentinel-2 L2A products.

True \ Prediction	Water	Turbid Water	Ice	Snow	Salt sea	Land	**Recall**
Water	12.02	1.04	1.18	0.00	0.00	0.03	84.26%
Turbid Water	0.00	1.12	0.00	0.00	0.09	0.01	91.55%
Ice	0.31	0.15	7.66	0.11	0.00	0.04	92.56%
Snow	0.01	0.00	0.15	4.73	0.00	0.02	96.33%
Salt sea	0.00	0.00	0.00	0.00	1.72	0.01	99.46%
Land	0.03	0.09	0.02	2.60	3.11	67.12	91.99%
Precision	97.15%	46.58%	85.07%	63.54%	35.03%	99.84%	91.30% (accuracy)

Figure 3: Confusion matrix (values in megapixels), recall, precision, and accuracy for RandomForest method over the TS2 test set, using top of the atmosphere Sentinel-2 L2A products.

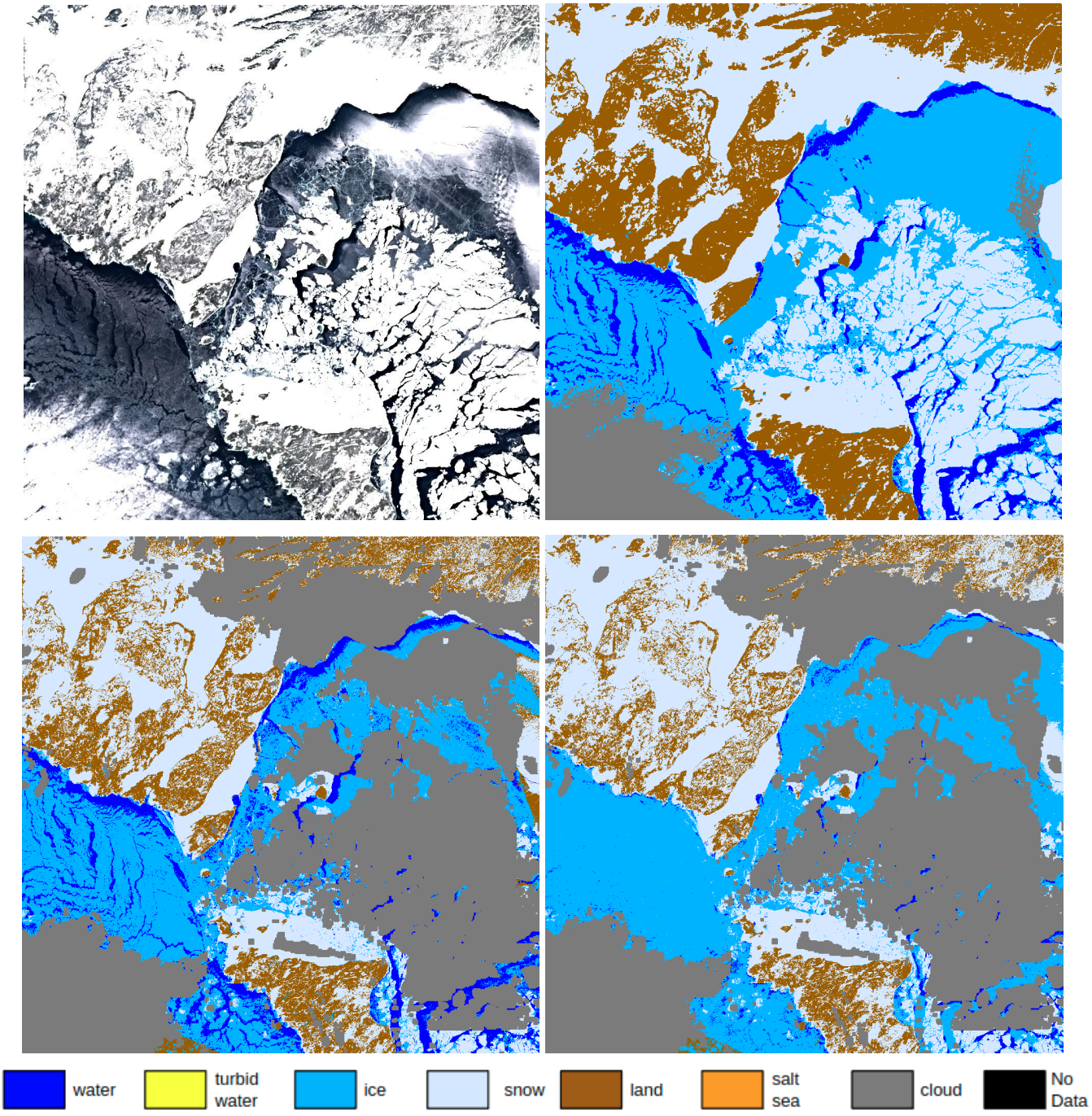

Figure 4: Quicklook in natural colours (upper left), ground truth (upper right), SVM result (bottom left), and RandomForest result (bottom right) for Geargia Bay, Canada. Sentinel-2 tile T17TML, 2019-02-19.

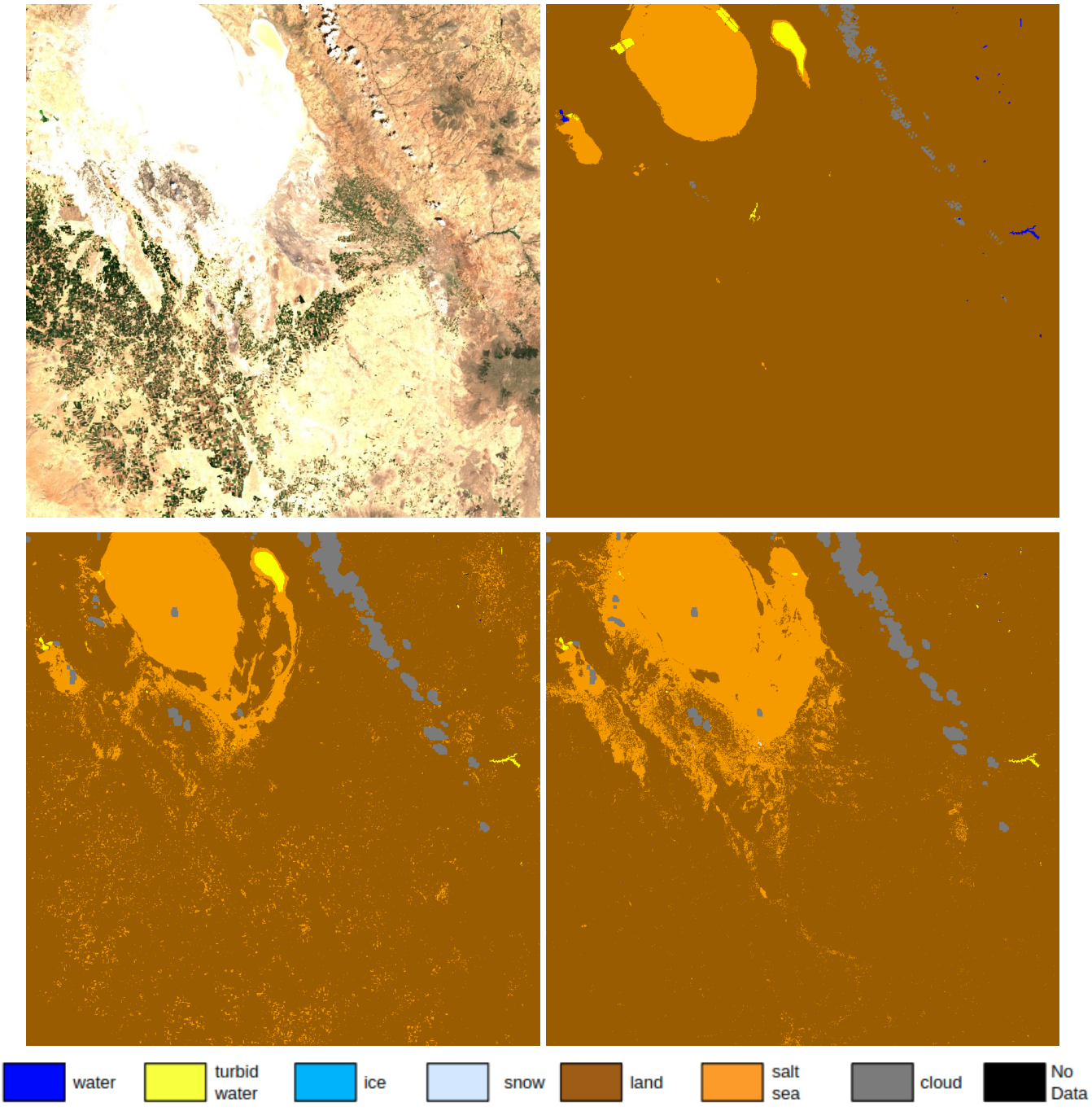

Figure 5: Quicklook in natural colours (upper left), ground truth (upper right), SVM result (bottom left), and RandomForest result (bottom right) for the Taz Golu desert, Turkey. Sentinel-2 tile T36SWH, 2021-07-12.

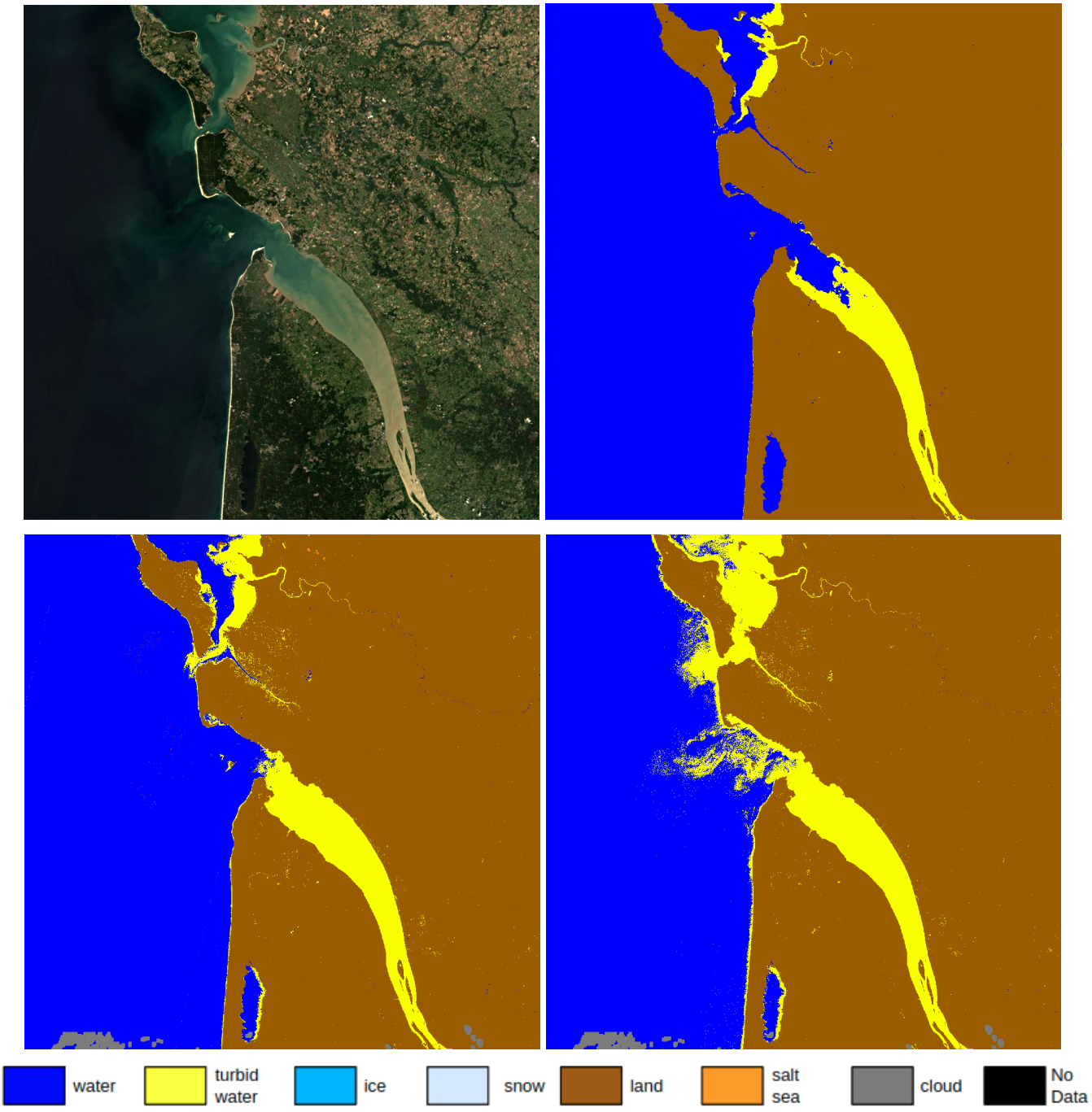

Figure 6: Quicklook in natural colours (upper left), ground truth (upper right), SVM result (bottom left), and RandomForest result (bottom right) for the Gironde estuary, France. Sentinel-2 tile T30TXR, 2021-07-17.

4 Results using top of the atmosphere L1C products

Direct classification from top of the atmosphere L1C products can provide snow/ice, water, land masks for atmospheric processors, and improve efficiency over static masks. If sufficiently accurate, it can also be used to generate snow and ice products without running the atmospheric processor and therefore save up a lot of processing time. In practice, the products would require a cloud mask, ideally processed independently from the atmospheric correction.

To assess the performance of classification using L1C products, we apply the same SVM and RandomForest algorithms that were used in the previous section, but feeding L1C data instead of L2A data, and adding the bands dedicated to atmospheric correction : B1 (coastal aerosols), B9 (water vapour), and B10 (SWIR-CIRRUS). It is worth noting that the 270K random pixels used for training from L1C data are not necessarily the same as the 270K random pixels selected for the training using L2A data, because the selection was run again separately. Figure 7 presents scores for the SVM method over the TS2 test set, and Figure 8 presents the same scores for the RandomForest method. The overall accuracy is 94.5% for SVM and 91.2% for RandomForest which are almost identical to the results for classification using L2A data. The rest of the scores are also very similar to the classification from L2A scores. Those results are therefore indicating that the machine learning algorithms used manage to mitigate the absence of atmospheric correction by learning directly from L1C data. This is a very positive result, as atmospheric correction could therefore be foregone when computing water, snow/ice, and land masks. However, the images included in the database were chosen notably for their low cloud cover. Results could be degraded in the presence of heterogeneous light nebulosity, including aerosols, not marked as cloud by the cloud mask processor.

True \ Prediction	Water	Turbid Water	Ice	Snow	Salt sea	Land	**Recall**
Water	13.57	0.49	0.18	0.00	0.00	0.03	95.10%
Turbid Water	0.00	1.08	0.00	0.00	0.13	0.01	88.26%
Ice	0.91	0.05	7.00	0.13	0.00	0.19	84.56%
Snow	0.00	0.00	0.15	4.72	0.01	0.03	96.17%
Salt sea	0.00	0.00	0.00	0.00	1.72	0.02	99.12%
Land	0.04	0.07	0.08	1.64	1.56	69.59	95.37%
Precision	93.44%	64.17%	94.50%	72.79%	50.32%	99.61%	94.49% (accuracy)

Figure 7: Confusion matrix (values in megapixels), recall, precision, and accuracy for SVM method over the TS2 test set, using top of the atmosphere Sentinel-2 L1C products.

True \ Prediction	Water	Turbid Water	Ice	Snow	Salt sea	Land	**Recall**
Water	13.09	0.79	0.35	0.00	0.00	0.04	91.78%
Turbid Water	0.00	1.07	0.01	0.00	0.13	0.01	87.00%
Ice	0.32	0.15	7.65	0.12	0.00	0.04	92.39%
Snow	0.00	0.00	0.14	4.75	0.00	0.02	96.78%
Salt sea	0.00	0.00	0.00	0.00	1.73	0.01	99.69%
Land	0.01	0.06	0.03	2.32	2.50	68.05	93.26%
Precision	97.58%	51.53%	93.55%	66.02%	39.64%	99.84%	93.19% (accuracy)

Figure 8: Confusion matrix (values in megapixels), recall, precision, and accuracy for RandomForest method over the TS2 test set, using top of the atmosphere Sentinel-2 L1C products.

5 Conclusions

We find that SVM and RandomForest methods achieve very good results in pixel-wise water, snow/ice and land classification, without using texture information. Moreover, the confusion between turbid water and salt lakes with respect to ice and snow is very low, therefore solving the current major issue with the HR-S&I RLIE algorithm. Those methods are easy to implement and run fast on a single CPU: a full Sentinel-2 image can be classified in less than thirty seconds with the SVM method, and less than two minutes for the RamdomForest method. Results on L1C classification also show that the results are very similar to those on L2A classification, signifying that the atmospheric correction could be avoided before estimation of water, land, and snow and ice masks, thus saving the larger part of the processing time.

Code availability

The Python code used for training and evaluation is not publicly available at this stage.

Data availability

The database used for training and evaluation is not publicly available at this stage.

Author contributions

Rémi Jugier, Robin Cremese, Hugo Fournier, and Nuria Duran-Gomez developed the Python codes for the study. Robin Cremese built the database for training and evaluation. Rémi Jugier wrote the paper. Germain Salgues and Chloé Thenoz provided support throughout the study and reviewed the paper.

Acknowledgements

We acknowledge Aurore Dupuis and Santiago Pena Luque from CNES for their support and continuous advice throughout the project. We also acknowledge Simon Gascoin from CESBIO and Alexei Kouarev from LEGOS for lending their expertise on ice and snow remote sensing and providing their advice throughout the project.

Financial support

This study has been proposed by Magellium within the CNES call for research and technology (R&T). It has then been selected and financed at 70% by CNES and 30% by Magellium.

References

Ansari, S., Rennie, C. D., Clark, S. P., & Seidou, O. (2021). IceMaskNet: River ice detection and characterization using deep learning algorithms applied to aerial photography. Cold Regions Science and Technology, 103324.

Baetens, L., Desjardins, C., & Hagolle, O. (2019). Validation of copernicus Sentinel-2 cloud masks obtained from MAJA, Sen2Cor, and FMask processors using reference cloud masks generated with a supervised active learning procedure. Remote Sensing, 11(4), 433.

Barbieux, K., Charitsi, A., & Merminod, B. (2018). Icy lakes extraction and water-ice classification using Landsat 8 OLI multispectral data. International Journal of Remote Sensing, 39(11), 3646-3678.

Belgiu, M., & Drăguţ, L. (2016). Random forest in remote sensing: A review of applications and future directions. ISPRS journal of photogrammetry and remote sensing, 114, 24-31.

Chen, L. C., Zhu, Y., Papandreou, G., Schroff, F., & Adam, H. (2018). Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European conference on computer vision (ECCV) (pp. 801-818).

Duguay, C. R., Bernier, M., Gauthier, Y., & Kouraev, A. (2015). Remote sensing of lake and river ice. Remote sensing of the cryosphere, 273-306.

Geldsetzer, T., Sanden, J. V. D., & Brisco, B. (2010). Monitoring lake ice during spring melt using RADARSAT-2 SAR. Canadian Journal of Remote Sensing, 36(sup2), S391-S400.

Hoeser, T., & Kuenzer, C. (2020). Object detection and image segmentation with deep learning on earth observation data: A review-part i: Evolution and recent trends. Remote Sensing, 12(10), 1667.

Isikdogan, F., Bovik, A. C., & Passalacqua, P. (2017). Surface water mapping by deep learning. IEEE journal of selected topics in applied earth observations and remote sensing, 10(11), 4909-4918.

Kouraev, A. V., Semovski, S. V., Shimaraev, M. N., Mognard, N. M., Légresy, B., & Remy, F. (2007). Observations of Lake Baikal ice from satellite altimetry and radiometry. Remote Sensing of Environment, 108(3), 240-253.

Leigh, S., Wang, Z., & Clausi, D. A. (2013). Automated ice–water classification using dual polarization SAR satellite imagery. IEEE Transactions on Geoscience and Remote Sensing, 52(9), 5529-5539.

Leshkevich, G. A., & Nghiem, S. V. (2007). Satellite SAR remote sensing of Great Lakes ice cover, part 2. Ice classification and mapping. Journal of Great Lakes Research, 33(4), 736-750.

Lin, G., Milan, A., Shen, C., & Reid, I. (2017). Refinenet: Multi-path refinement networks for high-resolution semantic segmentation. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1925-1934).

Nagler, T., Rott, H., Ripper, E., Bippus, G., & Hetzenecker, M. (2016). Advancements for snowmelt monitoring by means of Sentinel-1 SAR. Remote Sensing, 8(4), 348.

Nghiem, S. V., & Leshkevich, G. A. (2007). Satellite SAR remote sensing of Great Lakes ice cover, Part 1. Ice backscatter signatures at C band. Journal of Great Lakes Research, 33(4), 722-735.

Nolan, M., Liston, G., Prokein, P., Brigham‐Grette, J., Sharpton, V. L., & Huntzinger, R. (2002). Analysis of lake ice dynamics and morphology on Lake El'gygytgyn, NE Siberia, using synthetic aperture radar (SAR) and Landsat. Journal of Geophysical Research: Atmospheres, 107(D2), ALT-3.

Pekel, J. F., Cottam, A., Gorelick, N., & Belward, A. S. (2016). High-resolution mapping of global surface water and its long-term changes. Nature, 540(7633), 418-422.

Poliyapram, V., Imamoglu, N., & Nakamura, R. (2019, July). Deep learning model for water/ice/land classification using large-scale medium resolution satellite images. In IGARSS 2019-2019 IEEE International Geoscience and Remote Sensing Symposium (pp. 3884-3887). IEEE.

Prabha, R., Tom, M., Rothermel, M., Baltsavias, E., Leal-Taixe, L., & Schindler, K. (2020). Lake ice monitoring with webcams and crowd-sourced images. arXiv preprint arXiv:2002.07875.

Ronneberger, O., Fischer, P., & Brox, T. (2015, October). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical image computing and computer-assisted intervention (pp. 234-241). Springer, Cham.

Rott, H., & Nagler, T. (1994). Capabilities of ERS-1 SAR for snow and glacier monitoring in alpine areas. European Space Agency-Publications-ESA SP, 361, 965-965.

Sobiech, J., & Dierking, W. (2013). Observing lake-and river-ice decay with SAR: advantages and limitations of the unsupervised k-means classification approach. Annals of Glaciology, 54(62), 65-72.

Tan, M., & Le, Q. (2019, May). Efficientnet: Rethinking model scaling for convolutional neural networks. In International Conference on Machine Learning (pp. 6105-6114). PMLR.

Tom, M., Wu, T., Baltsavias, E., & Schindler, K. (2021). Recent Ice Trends in Swiss Mountain Lakes: 20-year Analysis of MODIS Imagery. arXiv preprint arXiv:2103.12434.

Tom, M., Aguilar, R., Imhof, P., Leinss, S., Baltsavias, E., & Schindler, K. (2020). Lake ice detection from sentinel-1 sar with deep learning. arXiv preprint arXiv:2002.07040.

Tom, M., Kälin, U., Sütterlin, M., Baltsavias, E., & Schindler, K. (2018). Lake ice detection in low-resolution optical satellite images. International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, 4, 279-286.

Wangchuk, S., & Bolch, T. (2020). Mapping of glacial lakes using Sentinel-1 and Sentinel-2 data and a random forest classifier: Strengths and challenges. Science of Remote Sensing, 2, 100008.

Yang, T. Y., Kessler, J., Mason, L., Chu, P. Y., & Wang, J. (2020). A consistent Great Lakes ice cover digital data set for winters 1973–2019. Scientific Data, 7(1), 1-12.

Yuan, X., Shi, J., & Gu, L. (2020). A review of deep learning methods for semantic segmentation of remote sensing imagery. Expert Systems with Applications, 114417.

Zakhvatkina, N., Korosov, A., Muckenhuber, S., Sandven, S., & Babiker, M. (2017). Operational algorithm for ice–water classification on dual-polarized RADARSAT-2 images. The Cryosphere, 11(1), 33-46.

Zhang, G., Yao, T., Xie, H., Wang, W., & Yang, W. (2015). An inventory of glacial lakes in the Third Pole region and their changes in response to global warming. Global and Planetary Change, 131, 148-157.

Zhang, W., Liljedahl, A. K., Kanevskiy, M., Epstein, H. E., Jones, B. M., Jorgenson, M. T., & Kent, K. (2020). Transferability of the deep learning mask R-CNN model for automated mapping of ice-wedge polygons in high-resolution satellite and UAV images. Remote Sensing, 12(7), 1085.