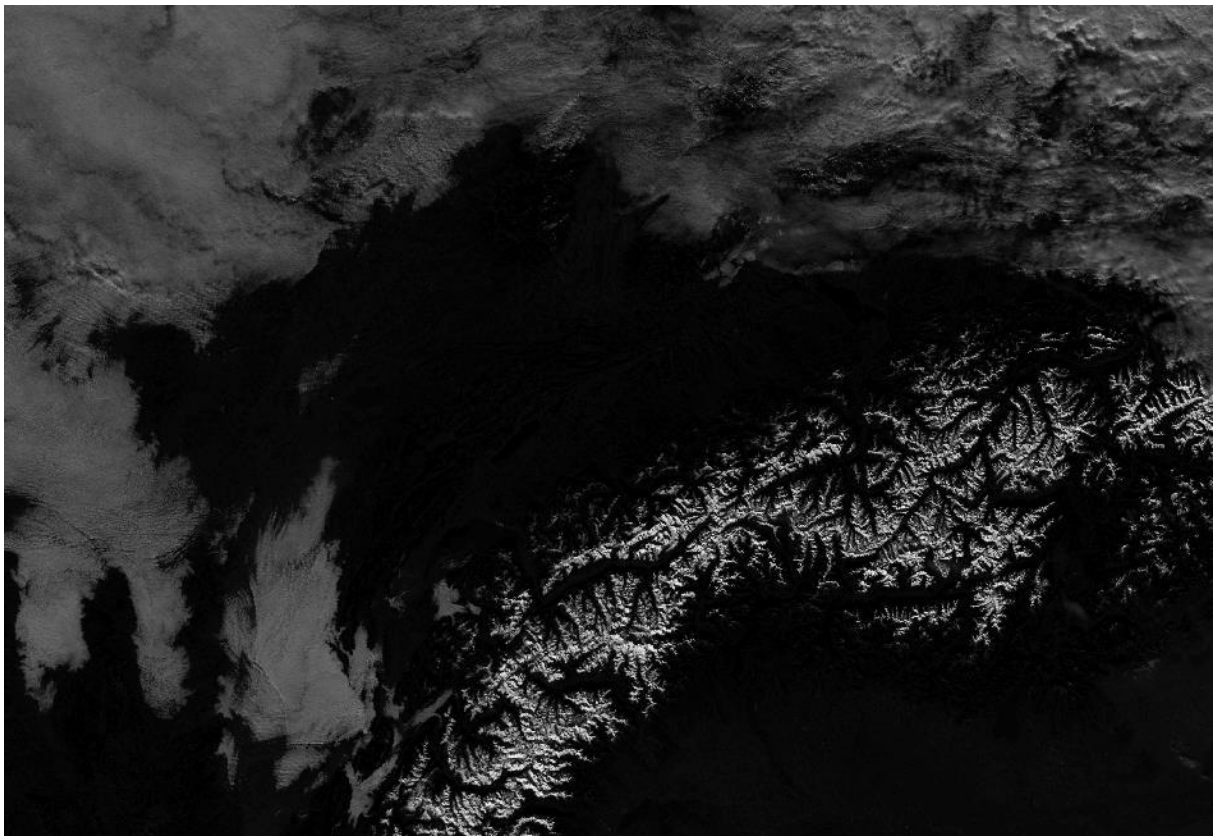
Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

IGP
Institute of Geodesy and Photogrammetry

Lake Ice Detection Using Low Spatial Resolution Satellite Images



Interdisciplinary Project Work 2017, 12.01.2018
Institute of Geodesy and Photogrammetry, ETH Zürich
Photogrammetry and Remote Sensing
Prof. Dr. Konrad Schindler

Author: Ursula Kälin, kaelinu@student.ethz.ch
Supervisors: Manu Tom
Dr. Emmanuel Baltsavias
Dr. Fabio Fontana (MeteoSwiss)
Michelle Stalder (MeteoSwiss)

TABLE OF CONTENTS

ABSTRACT

Lake ice is an observable that is part of the Essential Climate Variables (ECV) in the Global Observing System for Climate. This system is meant to monitor the changes of the climate and the effects of global warming and provide observations for this monitoring.
Low spatial resolution satellite imagery from the VIIRS sensor with a ground sampling distance of 375 m is used to determine lake ice in four target lakes in Switzerland: Sihlsee, Silsersee, Silvaplanersee and St. Moritzersee. The study period includes the three months October 2016, December 2016 and Feburary 2017. For the supervised lake ice classification, only clean pixels, i.e. pixels that lie completely within the lake, that are also cloud-free are considered. Because there is no clean pixel in St. Moritzersee, it has to be excluded from this study. A manual labeling of frozen and non-frozen dates based on webcam images forms the ground truth. Linear Support Vector Machine (SVM) classification and regression are trained with different combinations of VIIRS channels as feature vector and 5-fold cross-validated. The classification results show for each lake high overall accuracies that reach values of >99%. But the results must also be treated with caution because of the low amount of the processed data, the uncertainty of the ground truth (difficult manual interpretation of the webcam images) and the probable cloud mask errors.

1. INTRODUCTION

1.1 Motivation

The climate is changing and for a better understanding of those changes the Essential Climate Variables (ECVs) have been specified by the Global Climate Observing System (GCOS). They are now seen as the key observations to provide us with empirical evidence for the understanding and prediction of the evolution of climate. One of the forty-five ECVs is the *lake ice cover* that contains the observation of the lake ice extent and duration (World Meteorological Organization, 2017).

Up to now there is no harmonized, operational database that includes lake ice observations of Swiss lakes over longer periods. The Swiss GCOS Office at MeteoSwiss started in Summer 2016 together with the project partners ETH Zurich, University of Bern and EAWAG the project *Integrated Monitoring of Ice in selected Swiss Lakes* that evaluates, compares and integrates various input data and processing methods for lake ice estimation. (MeteoSwiss, 2017). Therefore, different input data are analyzed for their feasibility: Webcam images, in-situ measurements and satellite images (ETH Zürich PRS, 2017). In this work (which is integrated in the above project), the focus is on VIIRS satellite images. These images are free and after establishing a processing chain, their processing can be fast. They have a high temporal resolution (at least one image per day), sufficient spectral resolution but only coarse spatial resolution. A major problem as with all optical satellite sensors is the clouds.

1.2 Work aims

The primary aim of this work is to adapt the existing MODIS-based lake ice detection methodology to VIIRS satellite images and test its performance. The images used are from the VIIRS sensor on the satellite Suomi NPP. The five high-resolution Imagery-channels (I-bands) with a ground sampling distance (GSD) of approx. 375m at nadir and daily global coverage form the back-bone of the system. Support Vector Machine (SVM)-based supervised classification and regression is performed and cross-validated. The ground truth is determined using the freely available webcams monitoring the target lakes.

The study is concentrating on four target lakes: Sihlsee, Silsersee, Silvaplanersee and St. Moritzersee. This lake selection among the six lakes treated in the Meteowiss project is done for some reasons. Firstly, many of the Swiss lakes do not freeze regularly during the winter months, especially the bigger ones that lie in the flat midland. Secondly, only lakes that are big enough with respect to the GSD of the imagery can be considered. Lastly, the restriction is due to the necessity of having validation data that is needed for such a supervised classification approach. For this project the ground truth comes mainly through webcam images that exist for these four lakes.

Table 1 shows the main characteristics of the lakes and Fig. 1 localizes them. It can be seen, that the three lakes in the valley of Engadin (Silsersee, Silvaplanersee and St. Moritzersee) lie on a remarkably higher altitude and are noticeably smaller than Sihlsee. The two different environmental settings will allow to see if the processing results are of similar quality, independent of these lake characteristics differences, and thus the method can be applicable also to other lakes.

Table 1: Overview over the characteristics of the four lakes. Source: (swisstopo, 2017)

Name	Coordinates (Lat, Lon) [°]	Height [m]	Area [km^2]
Sihlsee	47.12067, 8.78560	889	10.85
Silsersee	46.41927, 9.73294	1797	4.1
Silvaplanersee	46.44866, 9.79199	1791	3.2
St. Moritzersee	46.49345, 9.84476	1768	0.78

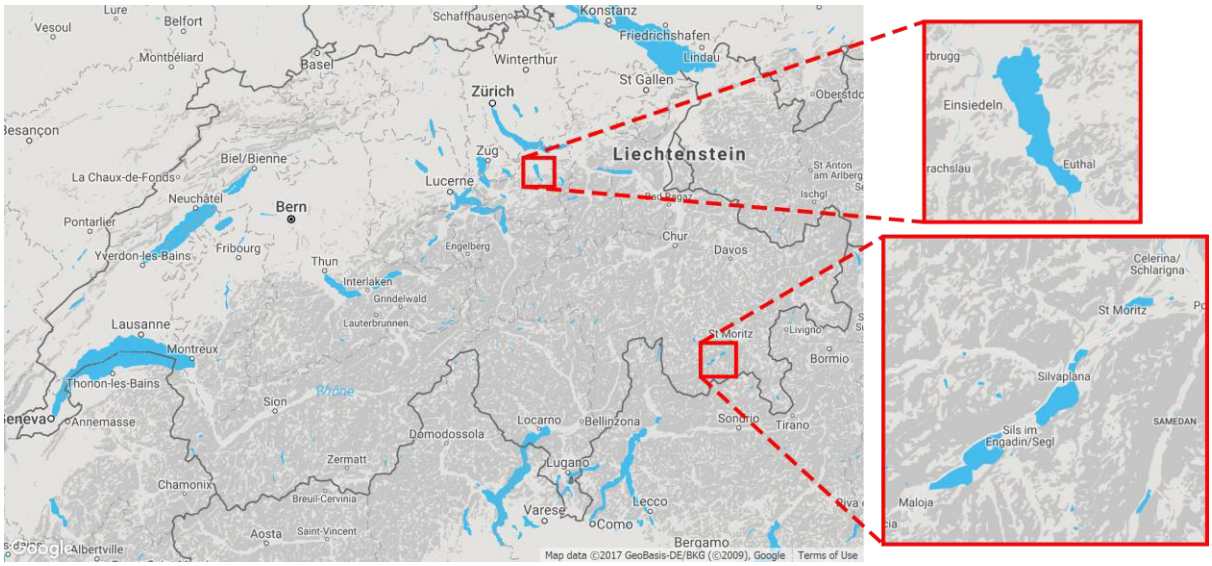


Fig. 1: Localization of the four target lakes. Zoomed-in on top right: Sihlsee. Zoomed-in on bottom right from bottom to top: Silsersee, Silvaplanersee and St. Moritzersee

The satellite data of the following three months are analyzed:

- October 2016
- December 2016
- February 2017

These three months show a variety of seasonal changes. October is still in the fall, December is at the start of the winter with first snow fall and cold temperatures and in February the accumulated cold winter days lead to frozen lakes. The limitation to only these months was necessary because the cloud-masks that are needed for the processing were at the time not yet processed for more months.

1.3 Some related work

As already mentioned, this work is part of a bigger project. Out of it also a study from Tom et al. (2017) used a similar approach with the satellite imagery of MODIS for the winter 2011-12 and summer 2012. A classification accuracy of >95% could be achieved. Another report that originates out of the overarching project is from Sütterlin et al. (2017). They used a single-channel lake surface water temperature (LSWT) algorithm to detect lake phenology. For that the I05 channel of the VIIRS sensor forms the input.

1.4 Structure of the report

After this introductory chapter, the VIIRS sensor and its data and the webcam validation data are described in chapter 2. The third chapter gives an overview of the methodology and includes for all the important processing steps an additional explanatory subchapter with intermediate results. Afterwards the final results of the classification are shown and discussed. The conclusion summarizes the work outcomes and gives an outlook on the possible future work.

2. DATA

2.1 VIIRS sensor and imagery channels

The VIIRS (Visible Infrared Imaging Radiometer Suite) sensor is on the Suomi-NPP weather satellite that was launched on 28 October 2011. For more details on Suomi-NPP and VIIRS see Suomi-NPP (2017). The instrument is a whiskbroom scanner with sixteen channels of moderate spatial resolution (0.75 km at nadir) and the five Imagery channels I01 to I05 with a spatial resolution of approx. 0.375 km at nadir. Here, only the I-channels with already calibrated data, the Sensor Data Records (SDR), are used, also to facilitate a comparison to the processing methods of the University of Bern, which were using only the I-channels. The effective bit depth of these images is 24, but is stored as 64-bit. The detailed spectral information and the main applications of these channels are described in Table 2. The channels 1-3 are in the reflective spectrum, whereas channels 4-5 are in the emissive region.

Table 2: Imagery channels with their specific spectral regions. Source: Suomi-NPP (2017)

Channel	Spectral region	Center wavelength [µm]	Bandwidth [µm]
I01	Visible and Near infrared	0.640	0.05
I02	Visible and Near infrared	0.865	0.039
I03	Short Wavelength infrared	1.61	0.06
I04	Mid wavelength infrared	3.74	0.38
I05	Thermal infrared	11.450	1.9

The satellite flies at a nominal altitude of 829 km and delivers two times daily global coverage (once during the day and once during night). Here only the daytime acquisitions are useful, but because of overlapping swaths of two consecutive orbits, our target area is often covered more than once per day. An important cause of data reduction at the acquisition level is the bow-tie effect that is caused by the whiskbroom working principle. This effect causes more than 2x bigger pixel footprints at the end of the scan (angle of +/-56.28°) and the scan width that is at nadir 11.7 km increases to 25.8 km. Fig. 2 shows in the lower part exemplarily such a swath. The orange and purple pixels are deleted and causes, as can be seen in the image on top of it, missing data stripes. In the further processing of the mapping and reprojection these deletions must be treated and have to be recovered by averaging or duplication of the neighborhood (Seaman, 2013).

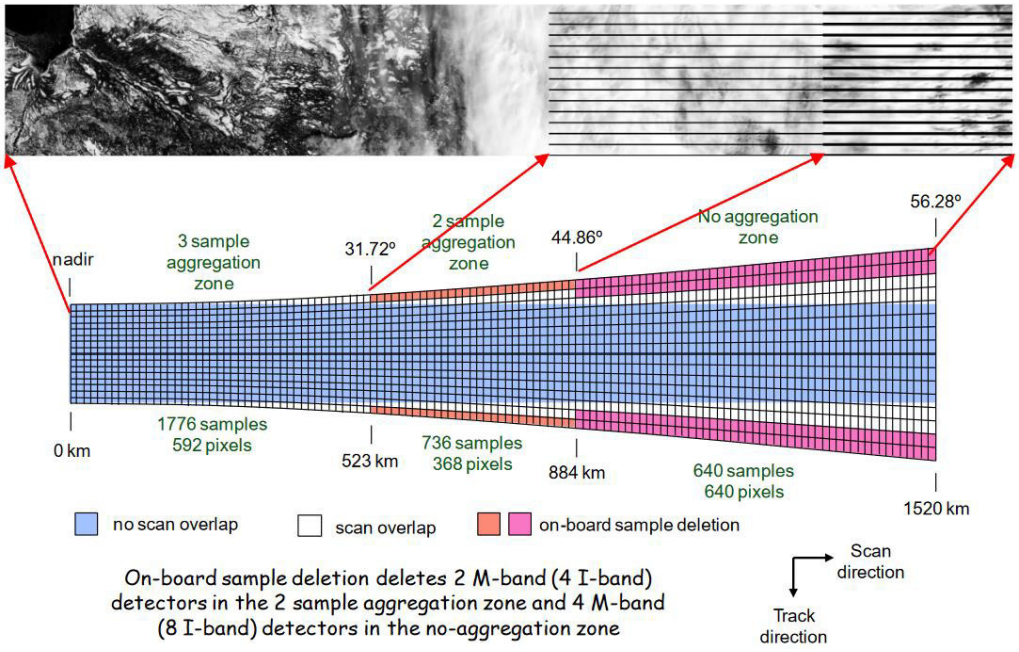


Fig. 2: Bottom: schematics of bow-tie effect, bow-tie deletion and aggregation scheme for single-gain M-bands (scale is exaggerated in the track direction. Top: example of bow-tie deletion effect when the raw data is displayed in sample space. Source: Cao et al. (2013)

Some attention has also to be paid on the geolocation of the satellite images. The information about how to geolocate the images into any reference system is independently stored in geolocation files. For the I-bands two different such geolocation files exist. The GIMGO projects everything onto a smooth ellipsoid, whereas GITCO also corrects for the terrain. In Switzerland the ground surface elevations change strongly, therefore it is necessary to use the GITCO-product to locate the images with high accuracy.
For more detailed explanations and background the *VIIRS Sensor Data Record (SDR) User's Guide* by Cao et al. (2013) is a recommended reading. The information in this chapter depends on this source, if not otherwise stated.

2.2 Webcam images

Freely available webcam images can be automatically downloaded from the Internet in real-time and can provide ground truth via manual interpretation, which is however not always easy, and thus sometimes uncertain. They deliver multiple images per day and give a good base for labeling frozen and non-frozen lake areas.
For the period of the VIIRS data the automatic download wasn't activated in October, but there we can confidently assume that the lakes were not frozen. From some webcams an only partial time coverage is given. For the following dates images are completely missing due to some technical problems:

- 01. – 03. December
- 24. – 28. December
- 17. February
- 22. – 25. February

Fig. 3 shows for each lake webcam footage of two cameras and how their view changes for frozen and non-frozen dates. The problematics that arise by determining the ground truth out of the webcam images are explained in Chapter 3.6.

Fig. 3: Example images of two webcams per lake. From upper left corner in clockwise direction: Sihlsee, Silsersee, St. Moritzersee, Silvaplanersee. The upper images in the blocks are from 4. Dezember 2016, except for the two at the very left are from 16.12 (their first date of recording). All lower images in the blocks are from 15. February 2017.

3. METHODOLODY AND DATA PROCESSING

3.1 Overview of the processing chain

The flowchart in Fig. 4 gives an overview of the processing chain with examples of the input data. In the following subchapters the procedure and the intermediate results of the processing steps are described.

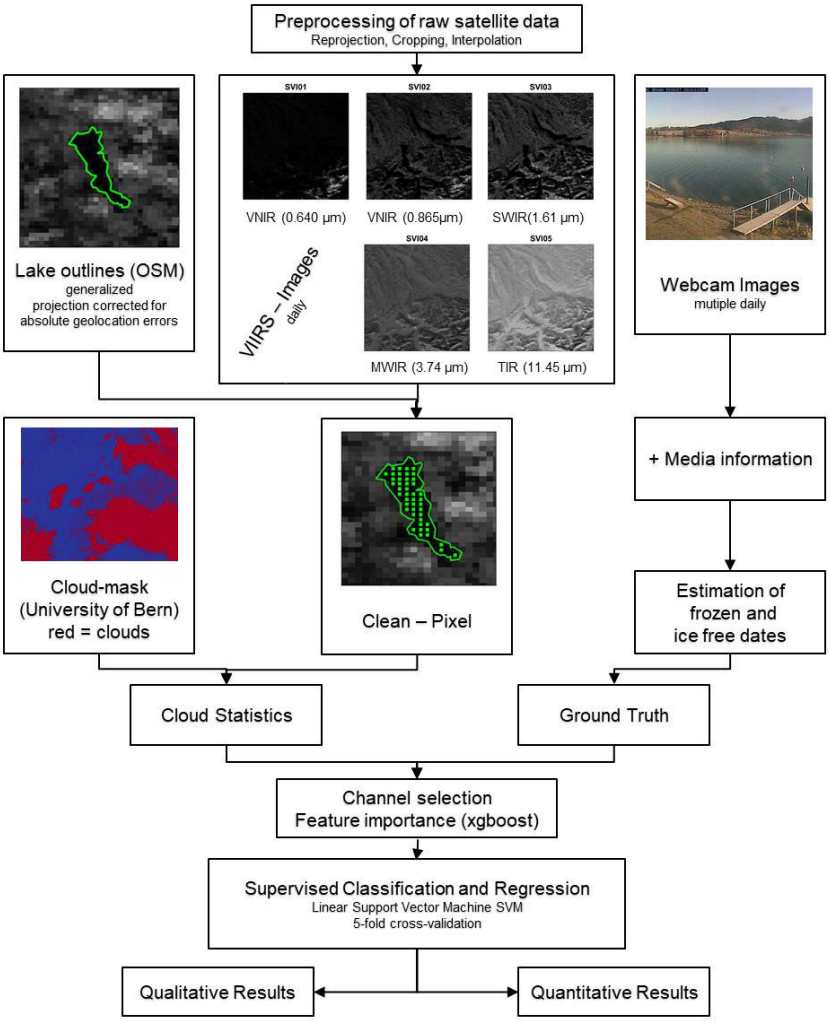


Fig. 4 Flowchart showing the pipeline that is set up for lake ice detection. OSM stands for OpenStreetMap.

In Chapter 3.2 the preprocessing of the raw satellite data is discussed, the product of it are images that show all the same extent. The lake outlines that are originally extracted from OpenStreetMap are generalized, projected onto the images using the georefencing information and corrected for absolute geolocation errors, this step being explained in Chapter 3.3. The combination of the backprojected and for absolute geolocation errors corrected lake outlines and satellite images define the clean pixels (Chapter 3.4). Using the cloud-masks for excluding cloudy clean pixels results in a cloud statistic whose results are described in Chapter 3.5. For the classification only acquisitions where more than 30% of all lake pixels are cloud-free are used.

Regarding the ground truth determination (Chapter 3.6), webcam images and some, rather scarce, information from media in the Internet are combined to estimate frozen and non-frozen dates.

Chapter 3.7 describes the characteristics of the grey-values of all clean and non-cloudy pixels, labeled as frozen and not-frozen. For getting an idea, which channels should be used for the classification we also compute the relative importance of the single channels for the differentiation of the two classes with the boosting-software xgboost from Chen et al. (2016). The last subchapter 3.8 explains the working principle of the SVM and shows the chosen parameters.

3.2 Preprocessing Satellite Images

All the preprocessing steps of the raw VIIRS-imagery is done by the University of Bern. This includes the download of the data from the Comprehensive Large Array-data Stewardship System (CLASS), reprojecting and resampling the images into the spatial reference system WGS84-UTM 32N, cropping the images so that they all show the same extent and save it in the *.tif* data format.
Additionally, the University of Bern also provided the cloud-masks that are used for the filtering out of cloudy pixel. The original cloud-mask data includes four levels of cloud confidence. To get a binary cloud-mask a conservative approach is chosen, so that a unclear pixel is rather classified as cloudy than not. More information about their procedure can be found in the First Year Report of the project (ETH Zürich et al., 2017).
Resulting are per acquisition five images (one per channel) and one cloud-mask, all the size of 1945 x1763 pixel. Fig. 5 shows an I01-channel image with the related cloud-mask.
The provided images that are not entirely covered by observations are manually discarded from the pipeline. This results in total 114 acquisitions with minimal one acquisition per day.

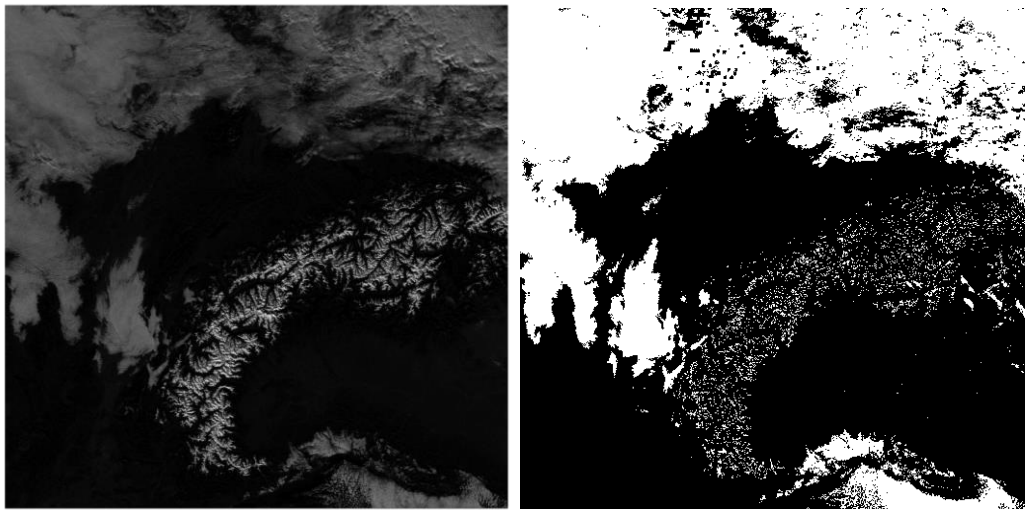


Fig. 5: Left: I01 - channel on the 2. December, where the snow-covered Alps are clearly visible. Right: The corresponding cloud-mask (white = cloud, black = no cloud).

3.3 Lake outlines

The lake outlines are the same as they are used in Tom et al. (2017) for the classification with MODIS. They are originating from OpenStreetMap and dowloaded via overpass-turbo.eu. They are generalized with help of the Douglas-Peucker-algorithm and projected into the image coordinates.
To be sure that the lakes in the satellite images correspond to each other an absolute geolocation correction is done with the approach from Aksakal (2013). For that mostly cloud-free images for October and December 2016 of the second channel are used because of the strong lake contrast in this channel. The lake outlines included 15 bigger Swiss lakes.
The results of this algorithm are shown in Table 3. Because no big difference for the four evaluated images appear, it is assumed that the shifts are stable, and the averaged shift is used for all images.

Table 3: Results of the absolute geolocation correction.

Image (day_time)	Weighted mean x-shift	Weighted mean y-shift	Standard deviation x	Standard deviation y	Median x	Median y
d20161004_t1203164	0.0	-0.1	0.2	0.1	0	-0.2
d20161030_t1214511	0.1	-0.3	0.1	0.2	0	-0.3
d20161202_t1200406	0.0	-0.4	0.4	0.3	0	-0.5
d20161228_t1212151	0.0	-0.4	0.1	0.3	0	-0.35
Average	0.0	-0.3				

3.4 Clean pixels

After implementing the shift, it is computed which pixels lie completely within the lake (clean pixels) and which are crossing the lake outlines (mixed pixels). The result can be seen in Fig. 6 and the numerical values in Table 4. Unfortunately for St. Moritzersee no clean pixel exists. A possible, additional processing step could be to try a classification also with mixed pixels.

Table 4: The number of clean and mixed pixels per acquisition and lake

Lake	Number of clean pixels	Number of mixed pixels
Sihlsee	45	63
Silsersee	11	37
Silvaplanersee	9	24
St. Moritzersee	0	11

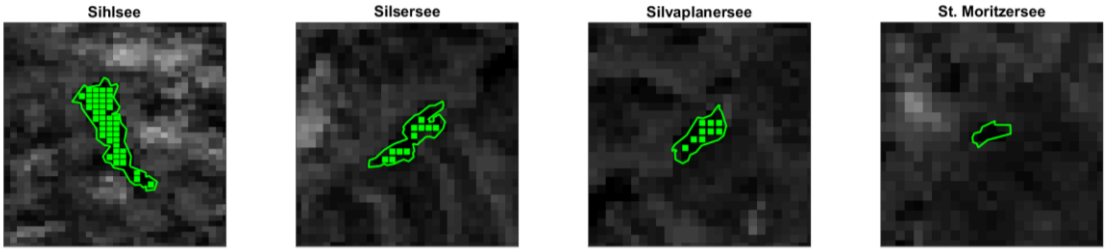


Fig. 6: The lake outlines projected on the images with the clean pixels illustrated as green squares.

3.5 Cloud-free clean pixels

Clouds must be filtered out of the data to reach a good result in the classification, because they get easily confused with ice and snow. In Switzerland especially in the winter months the cloud cover is substantial over longer periods. This leads to a significant reduction of the information and could cause long observation gaps. Fig. 7 shows how high the percentages of clouded pixels are per acquisition. The red line is the defined threshold of 70%, that is adapted from the work of Tom et al. (2017). If the cloud percentage is higher than the threshold, the whole acquisition is excluded from the classification. The raw data for Fig. 7 can be found in the Appendix Table 9. The resulting number of now remaining acquisitions and their total amount of pixels per channel that we will use for the classification are listed in Table 5.

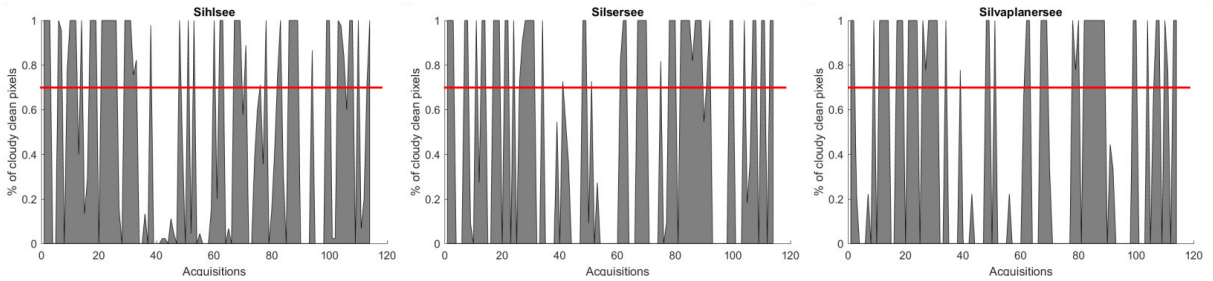


Fig. 7: Timeline of the cloud-cover for the three lakes. The red line is the defined threshold over which the acquisition is excluded.

Table 5: Total acquisitions and pixels per channel that are used for the classification.

Lake	Number of acquisitions < 70% cloudy	Total number of non-cloudy pixels per channel
Sihlsee	60	2431
Silsersee	59	612
Silvaplanersee	64	546

3.6 Ground truth

As already mentioned in Chapter 2.2 the ground truth determination is manually done with the help of webcam images. For each day the labels *frozen*, *not frozen* or *partially frozen* were assigned. Dates that are assigned with *partially frozen*, meaning that only parts of the lake are frozen, would not be used for the classification. To the best of my knowledge, in October and December the three lakes were not frozen whereas February was completely frozen for the three remaining lakes. A uncertainty remains at the end of December, where from 24. – 28. the webcam images are missing and otherwise the images are ambiguous. It is probable that on some days, especially during the morning hours, some thin layer of ice occurred. The changing light conditions between the webcam images makes it difficult to get to a conclusion. In the cases of doubt the decision fell to the label *non-frozen*, because the satellite images were never taken in the early morning but between 11 – 14 o'clock and therefore the thin ice sheet is considered as already leave melted. Fig. 8 shows such a questionable case. Another big uncertainty using only the webcam images is their incomplete coverage of the lake and therefore missing data for the hidden lake parts. Because of all these uncertainties, the period of 16. – 31. December will in future works not be used as input for the classification. It was not possible to include that change in this work due to time constraints.

As additional sources Internet media were searched. The most useful but sparse information was delivered by the archive of the newspaper *Engadiner Post* and the Facebook site from the Kitesurfing school in Silvaplana.

Bringing the labeling together with the previously estimated clean cloud-free pixels gives the final quantification of the input data for the classification (see Table 6).

Table 6: Number of acquisitions and pixels per class

Lake	Acquisitions Frozen	Pixels Frozen	Acquisitions not Frozen	Pixels not Frozen
Sihlsee	19	774	41	1657
Silsersee	15	153	44	459
Silvaplanersee	17	140	47	406

The webcam images also show that the frozen lake can be subclassified. Either the lake is only covered by ice, or on top of the ice a layer of snow is accumulated. Fig. 9 shows a comparison of the appearance with two webcam images. These two states behave differently in the spectral characteristics and could complicate the classification. For the study period only Sihlsee had for the days 1. – 5. February no solid snow cover on top of the ice.

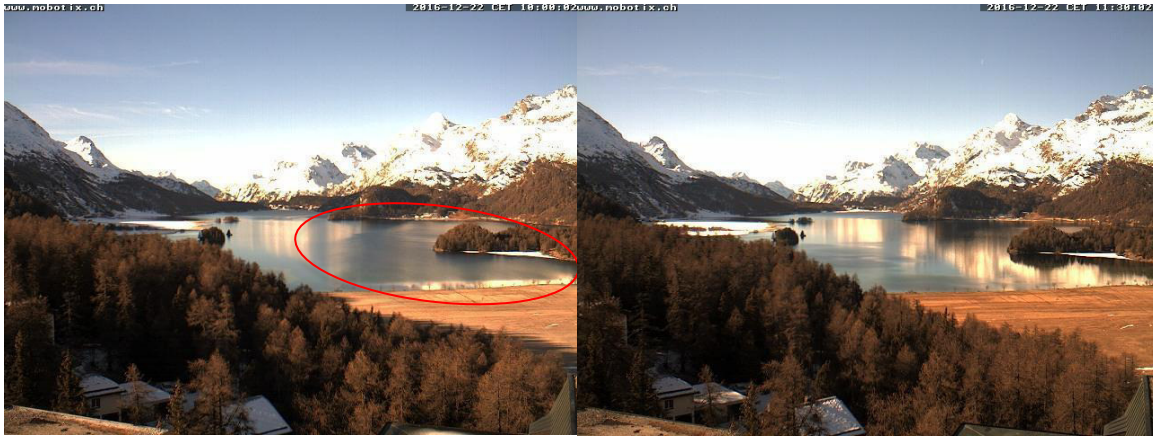


Fig. 8: A thin ice sheet circled in red on the 22. December at 10.00 (left) that is no longer visible at 11.30 o'clock (right)

Fig. 9: Ice on 03 February (left) and snow cover on the 06. February (right) on Sihlsee.

3.7 Grey-value statistics and channel selection

The distribution of the grey-values per channel are shown in Fig. 10. They show that the separability of the classes is good for channels 1 and 2 whereas the grey-values overlap strongly for the channels 3-5. Significant differences between the lakes are not visible. The non-frozen distribution for the 4th and 5th channels show two peaks. By splitting the non-frozen grey-values in those from October and December it is proved, that the two peaks coincide with the months (Fig. 15). The right peak is mainly formed from October values whereas December values overlap stronger with the frozen values.

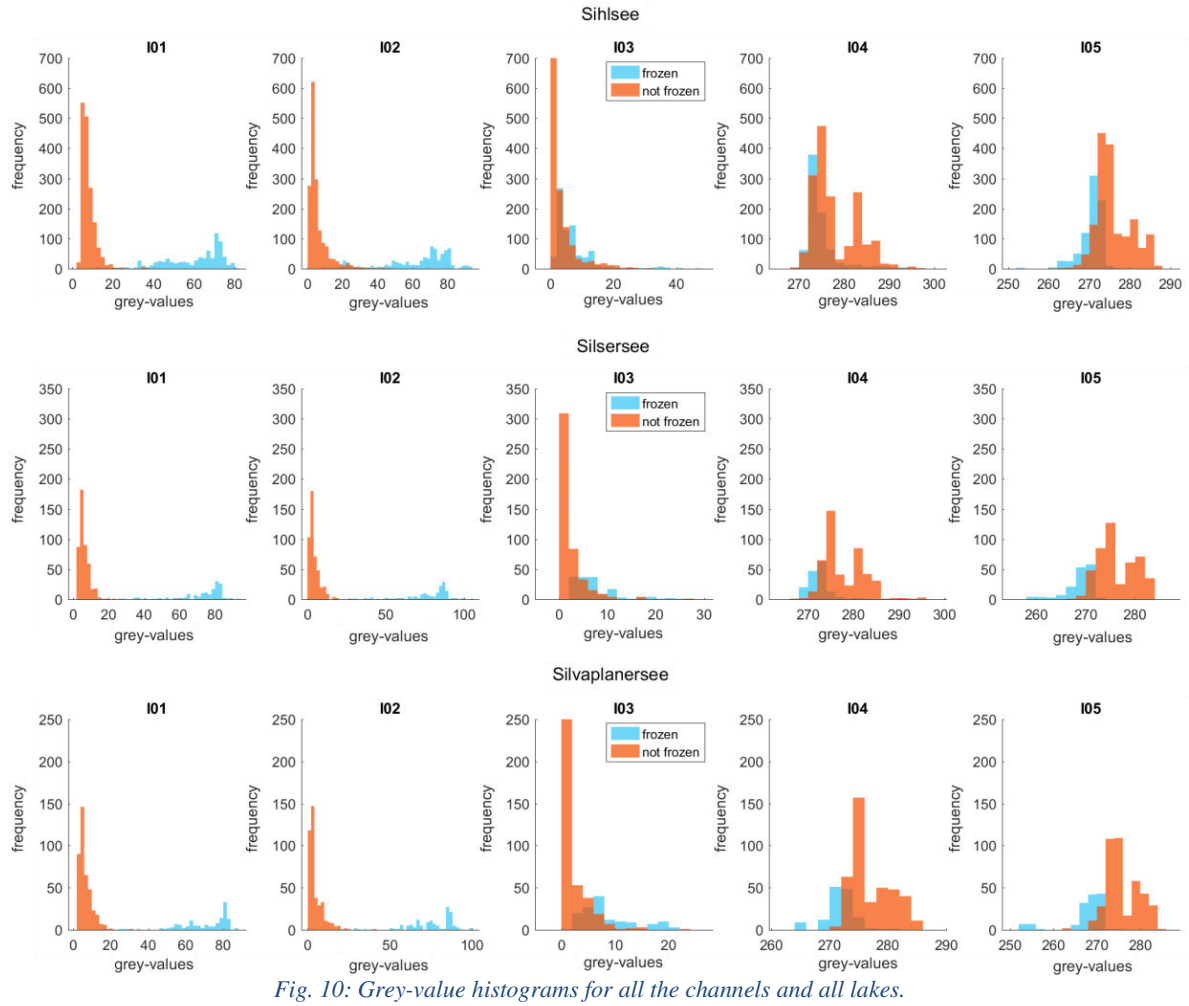


Fig. 10: Grey-value histograms for all the channels and all lakes.

Another way to look at the significance of the channels for the classification is to use them in the boosting software xgboost and determine the feature importance. The relative F-scores in Fig. 11 show how often one channel (feature) is responsible for the differentiation of the classes compared to all the other channels. After looking at the the grey-value histograms, it is surprising that channel 2 reaches only small F-scores. The explanation is that channel 1 and 2 have a very similar behavior but channel 1 differentiates most of the time slightly better than channel 2 and therefore gets all the F-scores for itself. Silsersee is located next to Silvaplanersee behaves less similar that Silvaplanersee to Sihlsee. Therefore, environmental and lake characteristics seem not to influence the channel importance.

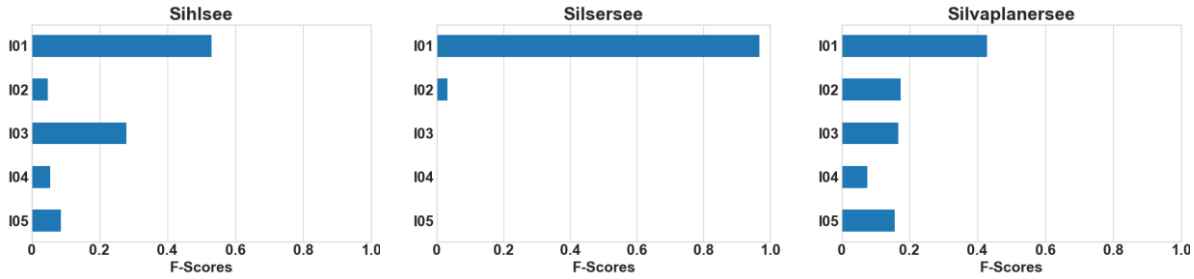


Fig. 11: Relative F-scores for the channels computed with xgboost

3.8 Classification and regression with a SVM

Support Vector Machines can be used for a supervised learning of classification problems (in this case binary, as only two target classes exist). The SVM searches for the hyperplane with a maximal margin between the classes. The data points that are the closest to the boundary form the support vectors after which the technique is named. A classification assigns the output strictly to a class (a hard classifier), whereas the regression returns a set of continuous values of probabilities (soft classifier).

The result of an SVM-classification does not have to correspond to the result of the regression, where the probability of 0.5 defines the class borders. Where the maximal margin lies, depends on the data distribution and not the probability values of the regression.

MATLAB has implemented both, an SVM classification and regression. For the input data (feature vector) all 31 possible channel combinations of the clean cloud-free pixels are tried out and the reached accuracies compared to each other.

For the parameters the default values are used except for the following parameter. “Standardize” is set to *True* because the data sample is small enough to include a standardization which is always recommended. The standardization removes the influence of different scaled input vectors. The kernel function that is on default *linear* was not changed because it leads to sufficient accuracies. The implementation of a non-linear kernel will be done in future work for the comparison of the classification results of VIIRS to the ones of MODIS. For all the default parameter definitions see the MATLAB documentation of the *fitcsvm* function (MathWorks, 2017).

A 5-fold cross-validation is performed for quantifiying of the accuracy of the model. With that approach, the data gets randomly split in 5 subparts. Four of them are used for the training of the model that is used to predict the fifth subpart (fold). This is done iteratively so that every fold is once predicted. The random assignment of the observations to the folds leads to slightly varying results for each new iteration.

4. RESULTS AND DISCUSSION

For all possible channel combinations and the results of the cross-validation a confusion matrix computed up and the overall accuracies and the kappa values are computed. (Richards, 2012)
Exemplarily the confusion matrix from the classification with the channels I01/I02/I03 for Sihlsee is shown in Table 7. From the 2431 pixels only 12 are wrongly classified with the same amount of six for both classes.

Table 7: Confusion-matrix for Sihlsee with the channel combination of I01, I02, I03

		actual class			
		non-frozen	frozen	sum	user's accuracies
predicted class	non-frozen	**768**	**6**	774	0.9922
	frozen	**6**	**1651**	1657	0.9963
	sum	774	1657	2431	
	producer's accuracies	0.9922	0.9963		

Overall Accuracy	0.995
Kappa Value	0.988

The definition of the accuracy measures that are listed in Table 8 are:

$$Overall\ Accuracy = \frac{True\ Positive + True\ Negative}{Total\ Number\ of\ Observations} \quad \text{(Eq. 1)}$$

$$Kappa\ Value = \frac{Overall\ Accuracy - probability\ of\ chance\ agreement}{1 - probability\ of\ chance\ agreement} \quad \text{(Eq. 2)}$$

The quality measures for the example above are therefore:

$$Overall\ Accuracy = \frac{768 + 1651}{2431} = 0.995 \quad \text{(Eq. 3)}$$

$$Prob.of\ chance\ agreement = \frac{774}{2431} * \frac{774}{2431} + \frac{1657}{2431} * \frac{1657}{2431} = 0.5716 \quad \text{(Eq. 4)}$$

$$Kappa\ Value = \frac{0.9951 - 0.5716}{1 - 0.5716} = 0.988 \quad \text{(Eq. 5)}$$

The classification is highly successful for many channel combinations. Overall accuracies higher than 0.99 and kappa values higher than 0.98 are often achieved. Table 8 gives an overview of most successful channel combinations. Already with the sole usage of channel 1 the results are not much worse than the best combinations, whose accuracies are highlighted in red for Sihlsee and Silvaplanersee. For Silsersee 17 channel combinations lead to a perfect classification.

Table 8: Overall Accuracies and Kappa Values for different channel combinations. The results change slightly with each new iteration of SVM.

Channel combinations	**Sihlsee**		**Silsersee**		**Silvaplanersee**	
	Acc.	Kappa	Acc.	Kappa	Acc.	Kappa
I01	0.990	0.977	1	1	0.991	0.976
I01, I02, I03	0.995	0.988	1	1	0.993	0.981
I01, I02, I03, I04	0.995	0.988	1	1	0.995	0.985
I01, I02, I03, I05	0.995	0.988	1	1	0.993	0.981
I01, I02, I04, I05	0.993	0.985	1	1	0.995	0.985
I01 – I05	0.995	0.988	1	1	0.995	0.985

The accuracy differences for the best combinations are not significant and have to be tested further on with bigger data sets to get more detailed performance insights.

The misclassifications from the example in Table 7 occur in only three different acquisitions that are shown in Fig. 12. The reason why is clear for the 11. October (on the left): A veil of cloud lies across the lake and the cloud-mask did mask out not enough pixels or the pixels are influenced by the neighboring clouds. For the misclassifications in February no conclusive explanation can be given. A possibility is that the ground truth is wrong. The boundary of the SVM-classification lies in the orange region of the regression results and therefore around the probability value of 0.75. This is the reason why the yellow and green pixels are also classified to the class *frozen.*

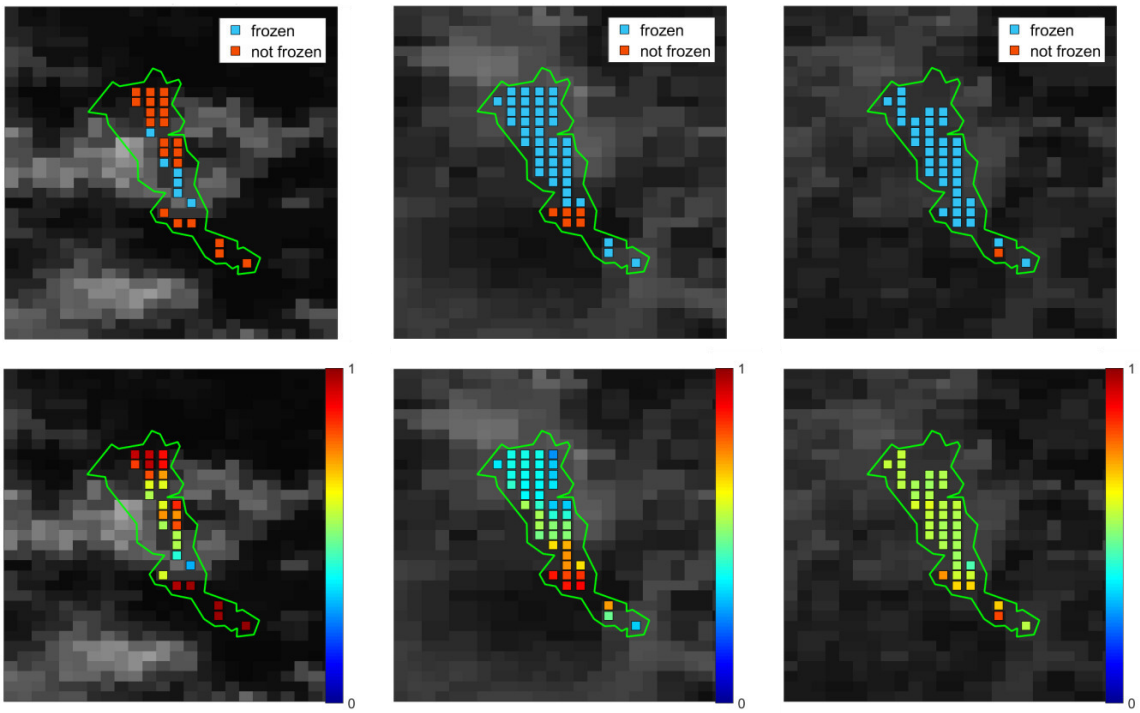


Fig. 12: Dates with misclassifications for Sihlsee from left to right: 11.10.16 (ground truth: not frozen), 01.02.17 (ground truth.: frozen), 02.02.17 (ground truth.: frozen). The upper row shows the classification, the lower the regression results. For the regression: Red (1) = non-frozen, blue (0) = frozen. Where colored pixels inside the lakes are missing, the cloud-mask assigned them as cloudy.

A timeline of the pixelwise results of the regression for Sihlsee, as it is shown in Fig. 13, gives an additional insight to the results. Notable is also how noisy the regression results are during the frozen period compared to the non-frozen one. This can be explained through the fact that the frozen state can appear in many different facets and that the reflecting characteristics of the snow cover changes steadily because of environmental influences. The respective timelines for Silsersee and Silvaplanersee can be found in the Appendix (Fig. 16 and Fig. 17).

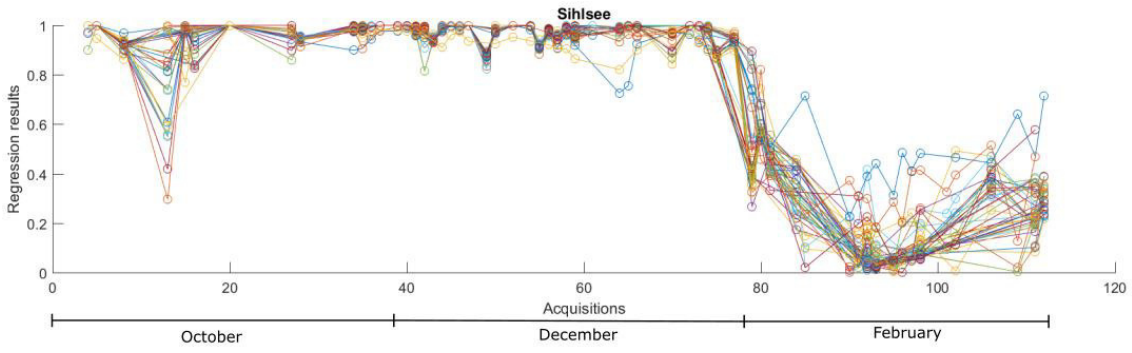


Fig. 13: The timeline shows the evolution of the regression values for each clean pixel of Sihlsee. Non-frozen = 1, Frozen = 0. Dates the pixels are cloud-free are indicated with small circles and a line connects them to each other. The outliers on the 11 October on the left part of the graph are well visible.

For Silvaplanersee the best classification result misclassifies three pixels wrongly as non-frozen in February. They are illustrated in Fig. 14. The reasons are not known.
An unrelated striking aspect that can be seen in Fig. 14 (middle and right) is that some pixels have the exact same grey-values as their neighbors. It is possible that those are the leftover of the bow-tie effect (see Chapter 2.1).

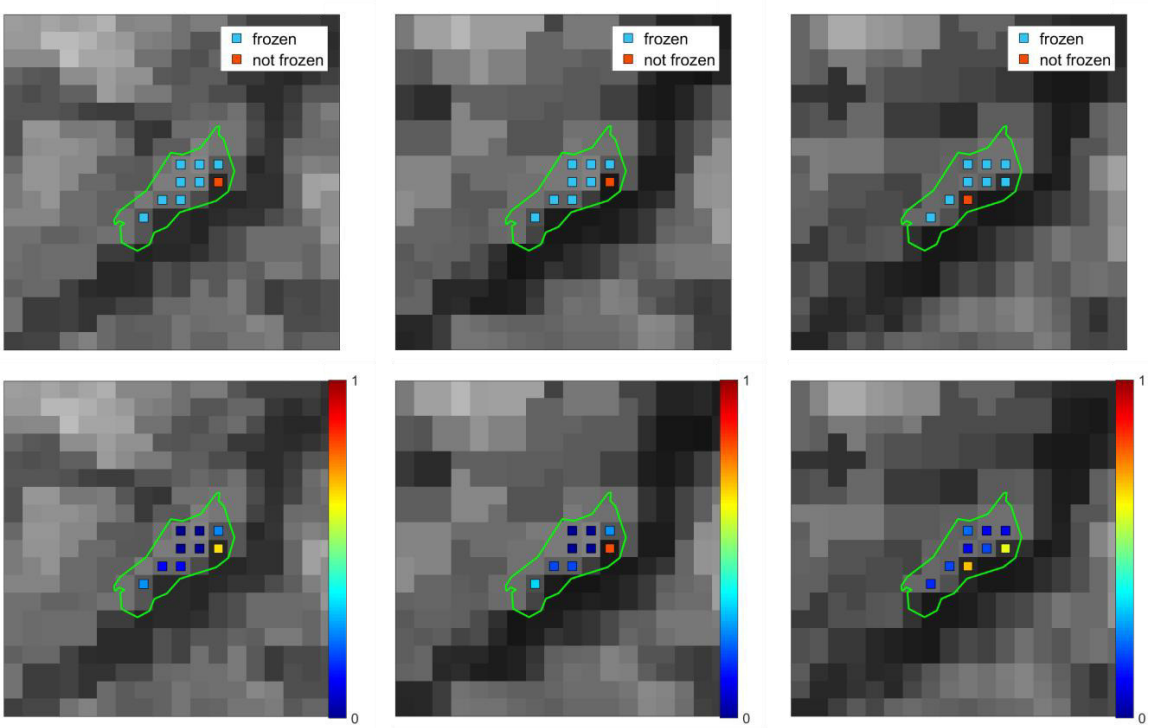


Fig. 14: Dates with misclassifications for Silvaplanersee from left to right: 12.02.2017, 16.02.17, 27.02.17 (ground truth for all: frozen). The upper row shows the classification, the lower the regression results. For the regression: Red (1) = non-frozen, blue (0) = frozen.

The classification for Silsersee delivers for many channel combinations a perfect result without any misclassifications.

5. CONCLUSION AND OUTLOOK

The results of the supervised classification and regression show that the lake ice detection with VIIRS-imagery performs very well with accuracies pf over 98%. To achieve such good accuracies, it is necessary to have very accurate cloud-masks. The case of a defective cloud-mask, that caused misclassifications, highlights the importance.
An intrinsic problem of optical satellite data is the missing information on cloudy days. This can lead to significant information gaps and a big reduction of usable data. An interpolation for these gaps will be especially challenging for periods of freezing and thawing.
An important following task is to verify the results with more data, because the here used sample is rather small. A complete record of a winter that includes the thawing and freezing periods and a comparison of it to other years has to prove the robustness of the results. With bigger samples also the different sub-characteristics of the *frozen* class, ice or snow-cover on top of it, could be evaluated.
In connection to the project *Integrated Monitoring of Ice in Selected Swiss Lakes* this work gives a basis for the comparison with the other implemented methods. To detect ice on small lakes like St. Moritzersee another approach is needed, or the classification of mixed pixels must be considered.
For the future an integration of further data sources to the classification as temperature records could help in the interpolation of gaps. The inclusion of a temporal or spatial interpolation as a way to filter out outliers could leverage the already high accuracy.

ACKNOWLEDGMENTS

This work is part of the project *Integrated Monitoring of Ice in Selected Swiss Lakes* funded by Swiss Federal Office of Meteorology and Climatology MeteoSwiss in the framework of GCOS Switzerland.
I thank Melanie Sütterlin of the University of Bern for providing the satellite data.
And last but not least I would also like to thank Manu Tom, my supervisor who supported this work with his know-how and experience.

REFERENCES

Aksakal, S. K., 2013. Geometric accuracy investigations of SEVIRI high resolution visible (HRV) level 1.5 imagery. Remote Sensing 5(5), pp. 2475–2491. Accessible at: http://www.mdpi.com/2072-4292/5/5/2475 Accessed: 27.12.2017.

Cao, C., Xiong, X., Wolfe, R., DeLuccia, F., Liu, Q., Blonski, S., Lin, G., Nishihama, M., Pogorzala, D., Oudrari, H., & Hillger, D., 2013. Visible Infrared Imaging Radiometer Suite (VIIRS) Sensor Data Record (SDR) User's Guide Version 1.2. NOAA Technical Report NESDIS 142. Accessible at: https://lpdaac.usgs.gov/sites/default/files/public/product_documentation/vnp_v1.2_user_guide.pdf Accessed: 27.12.2017.

Chen, T., Guestrin, C., 2016. Xgboost: A scalable tree boosting system. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, California, USA, pp. 785–794. Accessible at: https://arxiv.org/abs/1603.02754 Accessed: 27.12.2017.

Engadiner Post, 2017. Archiv der Engadiner Post. Accessible at: http://dsb.engadinerpost.ch/DSB_Anbieter_03_Shw.asp?KundenID=Engadinerpost&VKundenID=724norm&S=1&AnbieterID=28742 Accessed: 27.12.2017.

ETH Zürich PRS, 2017. Accessible at: https://www.ethz.ch/content/specialinterest/baug/institute-igp/photogrammetry-and-remote-sensing/en/research/current_projects/integrated-monitoring-of-ice-swiss-lakes.html Accessed: 27.12.2017.

ETH Zürich, University of Bern, EAWAG, 2017. Integrated monitoring of ice in selected Swiss lakes - First Year Report.

MathWorks, 2017. MATLAB Documentation - fitcsvm. Accessible at: https://ch.mathworks.com/help/stats/fitcsvm.html? Accessed: 27.12.2017.

MeteoSwiss, 2017. Accessible at: http://www.meteoschweiz.admin.ch/home/forschung-und-zusammenarbeit/internationale-zusammenarbeit/gcos.html. Accessed: 27.12.2017

NOAA, 2017. Accessible at: https://www.class.ngdc.noaa.gov/. Accessed: 27.12.2017

Richards, J. A., 2013. *Remote Sensing Digital Image Analysis An Introduction.* Springer-Verlag Berlin Heidelberg

Seaman, C., 2013. Beginner's Guide to VIIRS Imagery Data. Accessible at: http://rammb.cira.colostate.edu/projects/npp/Beginner_Guide_to_VIIRS_Imagery_Data.pdf Accessed: 27.12.2017

Suomi-NPP, 2017. Accessible at: https://eoportal.org/web/eoportal/satellite-missions/s/suomi-npp. Accessed: 27.12.2017

Sütterlin, M., Duguay-Tetzlaff, A., & Wunderle, S., 2017. Toward a lake ice phenology derived from VIIRS data. In: Proceedings of the 19th EGU General Assembly, Vienna, Austria., p.8907

Swiss Kitesurf GmbH., 2017. Accessible at: https://www.facebook.com/SwissKitesurf/ Accessed: 27.12.2017.

Swisstopo, 2017. Accessible at. https://map.geo.admin.ch/?lang=de&topic=ech&bgLayer=ch.swisstopo.pixelkarte-farbe&layers=ch.bafu.vec25-seen Accessed: 27.12.2017.

Tom, M., Lanaras, C., Baltsavias, E., & Schindler, K., 2017. Ice Detection in Swiss Lakes using MODIS Data. In Asian Conference on Remote Sensing, 2017, New Delhi, India. Accessible at: https://www.ethz.ch/content/dam/ethz/special-interest/baug/igp/photogrammetry-remote-sensing-dam/documents/pdf/Papers/Tom_ETHZ_ACRS17.pdf Accessed: 27.12.2017.

World Meteorological Organization. (2017). Accessible at: https://public.wmo.int/en/programmes/global-climate-observing-system Accessed: 27.12.2017

APPENDIX

Table 9: Cloud statistics for all acquisitions.

	Sihlsee			Silsersee			Silvaplanersee				Sihlsee			Silsersee			Silvaplanersee		
	45 clean pixels			11 clean pixels			9 clean pixels												
	cloud-free	cloudy	% clouds	cloud-free	cloudy	% clouds	cloud-free	cloudy	% clouds		cloud-free	cloudy	% clouds	cloud-free	cloudy	% clouds	cloud-free	cloudy	% clouds
d20161001_t1119031	0	45	100	0	11	100	0	9	100	d20161216_t1234282	45	0	0	11	0	0	9	0	0
d20161001_t1301287	0	45	100	0	11	100	0	9	100	d20161217_t1216580	38	7	16	11	0	0	9	0	0
d20161002_t1243584	0	45	100	0	11	100	7	2	22	d20161218_t1159296	0	45	100	11	0	0	9	0	0
d20161003_t1226282	45	0	0	11	0	0	9	0	0	d20161219_t1141594	36	9	20	2	9	82	3	6	67
d20161004_t1203164	45	0	0	11	0	0	9	0	0	d20161220_t1118475	0	45	100	0	11	100	0	9	100
d20161005_t1145461	0	45	100	11	0	0	9	0	0	d20161220_t1301131	0	45	100	0	11	100	0	9	100
d20161006_t1128159	2	43	96	0	11	100	7	2	22	d20161221_t1243424	45	0	0	11	0	0	9	0	0
d20161006_t1310415	45	0	0	0	11	100	9	0	0	d20161222_t1226122	42	3	7	11	0	0	9	0	0
d20161007_t1247296	10	35	78	10	1	9	0	9	100	d20161223_t1203004	45	0	0	11	0	0	9	0	0
d20161008_t1230012	0	45	100	11	0	0	9	0	0	d20161224_t1145320	0	45	100	0	11	100	0	9	100
d20161009_t1212310	0	45	100	0	11	100	0	9	100	d20161225_t1128018	0	45	100	0	11	100	0	9	100
d20161010_t1155008	0	45	100	8	3	27	0	9	100	d20161225_t1310256	0	45	100	0	11	100	0	9	100
d20161011_t1131489	27	18	40	0	11	100	0	9	100	d20161226_t1247155	19	26	58	0	11	100	6	3	33
d20161011_t1314145	0	45	100	0	11	100	0	9	100	d20161227_t1229453	5	40	89	11	0	0	9	0	0
d20161012_t1114187	39	6	13	11	0	0	9	0	0	d20161228_t1212151	45	0	0	11	0	0	9	0	0
d20161012_t1256443	32	13	29	11	0	0	9	0	0	d20161229_t1154449	45	0	0	11	0	0	9	0	0
d20161013_t1239141	0	45	100	0	11	100	0	9	100	d20161230_t1131331	28	17	38	11	0	0	9	0	0
d20161014_t1216023	0	45	100	0	11	100	0	9	100	d20161230_t1313587	18	27	60	2	9	82	9	0	0
d20161015_t1158320	0	45	100	0	11	100	0	9	100	d20161231_t1114029	13	32	71	11	0	0	9	0	0
d20161016_t1141036	45	0	0	11	0	0	9	0	0	d20161231_t1256285	29	16	36	10	1	9	9	0	0
d20161017_t1123334	0	45	100	0	11	100	0	9	100	d20170201_t1111393	0	45	100	0	11	100	0	9	100
d20161017_t1300174	0	45	100	0	11	100	0	9	100	d20170201_t1254049	45	0	0	0	11	100	2	7	78
d20161018_t1242471	0	45	100	11	0	0	0	9	100	d20170202_t1236347	38	7	16	0	11	100	0	9	100
d20161019_t1225169	0	45	100	0	11	100	0	9	100	d20170203_t1219045	26	19	42	11	0	0	9	0	0
d20161020_t1202051	0	45	100	11	0	0	9	0	0	d20170204_t1155526	11	34	76	0	11	100	0	9	100
d20161021_t1144349	0	45	100	3	8	73	0	9	100	d20170205_t1138242	0	45	100	0	11	100	0	9	100
d20161022_t1127047	38	7	16	1	10	91	2	7	78	d20170206_t1120540	31	14	31	0	11	100	0	9	100
d20161022_t1309303	45	0	0	0	11	100	0	9	100	d20170206_t1303196	45	0	0	0	11	100	0	9	100
d20161024_t1228500	0	45	100	0	11	100	0	9	100	d20170207_t1240077	0	45	100	2	9	82	0	9	100
d20161025_t1211198	0	45	100	0	11	100	0	9	100	d20170208_t1222375	0	45	100	0	11	100	0	9	100
d20161026_t1153495	0	45	100	0	11	100	0	9	100	d20170209_t1205072	0	45	100	0	11	100	0	9	100
d20161027_t1136193	11	34	76	11	0	0	9	0	0	d20170210_t1147370	0	45	100	0	11	100	0	9	100
d20161027_t1313033	8	37	82	11	0	0	9	0	0	d20170211_t1124252	45	0	0	5	6	55	9	0	0
d20161028_t1113075	45	0	0	0	11	100	0	9	100	d20170211_t1306508	45	0	0	3	8	73	5	4	44
d20161028_t1255331	45	0	0	11	0	0	9	0	0	d20170212_t1106567	45	0	0	0	11	100	6	3	33
d20161029_t1238029	39	6	13	11	0	0	9	0	0	d20170212_t1249205	45	0	0	11	0	0	9	0	0
d20161030_t1214511	45	0	0	11	0	0	9	0	0	d20170213_t1231503	6	39	87	11	0	0	9	0	0
d20161031_t1157209	1	44	98	11	0	0	9	0	0	d20170214_t1208402	45	0	0	11	0	0	9	0	0
d20161201_t1218108	45	0	0	5	6	55	2	7	78	d20170215_t1151099	45	0	0	11	0	0	9	0	0
d20161202_t1200406	45	0	0	11	0	0	9	0	0	d20170216_t1133397	45	0	0	11	0	0	9	0	0
d20161203_t1137288	45	0	0	3	8	73	9	0	0	d20170216_t1316053	45	0	0	11	0	0	9	0	0
d20161204_t1119586	44	1	2	5	6	55	9	0	0	d20170217_t1116095	0	45	100	0	11	100	0	9	100
d20161204_t1302242	44	1	2	7	4	36	7	2	22	d20170217_t1252534	0	45	100	0	11	100	0	9	100
d20161205_t1244540	45	0	0	11	0	0	9	0	0	d20170218_t1235232	44	1	2	11	0	0	9	0	0
d20161206_t1221421	40	5	11	11	0	0	9	0	0	d20170219_t1217529	44	1	2	11	0	0	9	0	0
d20161207_t1204119	43	2	4	11	0	0	9	0	0	d20170220_t1200227	0	45	100	11	0	0	9	0	0
d20161208_t1146417	45	0	0	11	0	0	9	0	0	d20170221_t1137126	1	44	98	0	11	100	0	9	100
d20161209_t1129133	0	45	100	0	11	100	0	9	100	d20170222_t1119423	7	38	84	9	2	18	9	0	0
d20161209_t1305555	25	20	44	0	11	100	0	9	100	d20170222_t1302079	18	27	60	7	4	36	3	6	67
d20161210_t1248271	45	0	0	10	1	9	9	0	0	d20170223_t1244377	0	45	100	0	11	100	0	9	100
d20161211_t1230569	0	45	100	3	8	73	0	9	100	d20170224_t1221258	0	45	100	0	11	100	0	9	100
d20161212_t1213267	43	2	4	11	0	0	9	0	0	d20170225_t1203556	45	0	0	11	0	0	9	0	0
d20161213_t1150148	0	45	100	8	3	27	9	0	0	d20170226_t1146253	0	45	100	0	11	100	0	9	100
d20161214_t1132446	45	0	0	11	0	0	9	0	0	d20170227_t1128551	42	3	7	0	11	100	2	7	78
d20161214_t1315102	43	2	4	11	0	0	9	0	0	d20170227_t1305390	36	9	20	11	0	0	9	0	0
d20161215_t1115144	45	0	0	11	0	0	7	2	22	d20170228_t1105450	13	32	71	0	11	100	0	9	100
d20161215_t1257400	45	0	0	11	0	0	9	0	0	d20170228_t1248106	0	45	100	0	11	100	0	9	100

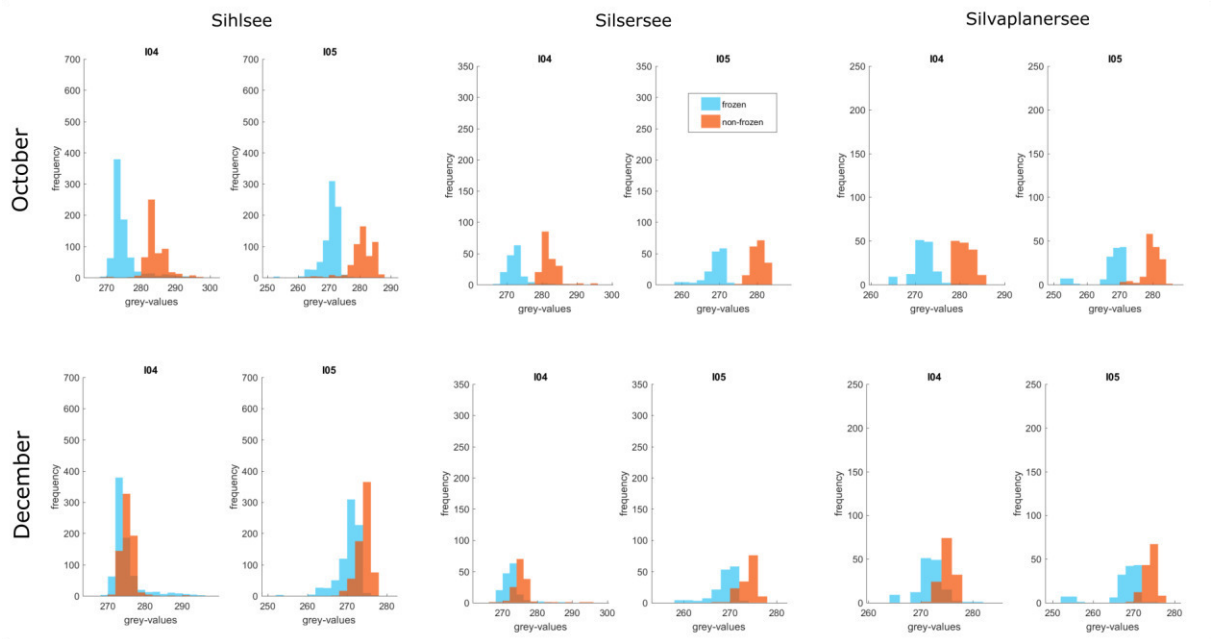


Fig. 15: Grey-value histograms for channels I04 and I05 for all lakes. Top: months October and February. Bottom: months December and February.

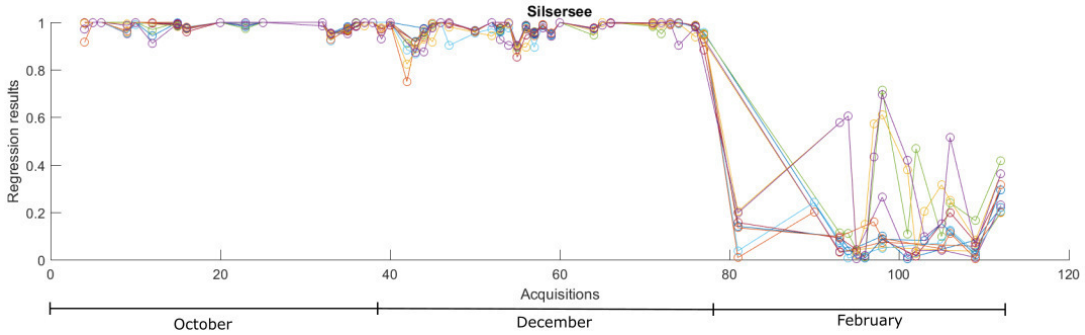


Fig. 16: The timeline shows the evolution of the regression values for each clean pixel of Silsersee. Non-frozen = 1, Frozen = 0. Dates the pixels are cloud-free are indicated with small circles and a line connects them to each other.

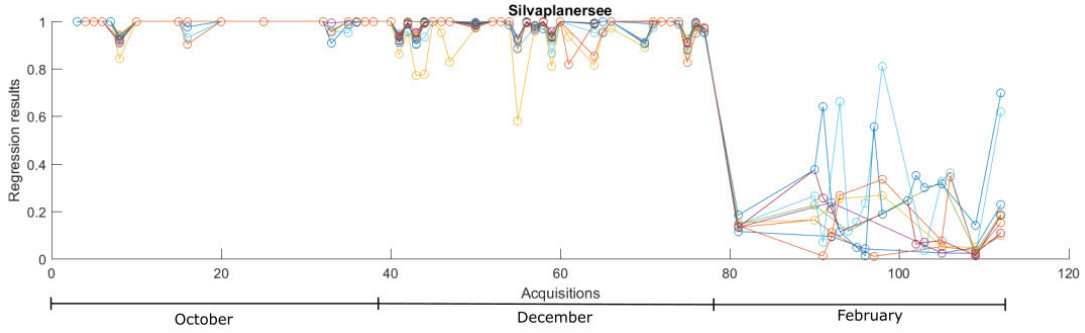


Fig. 17: The timeline shows the evolution of the regression values for each clean pixel of Silvaplanersee. Non-frozen = 1, Frozen = 0. Dates the pixels are cloud-free are indicated with small circles and a line connects them to each other.

Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Declaration of originality

The signed declaration of originality is a component of every semester paper, Bachelor's thesis, Master's thesis and any other degree paper undertaken during the course of studies, including the respective electronic versions.

Lecturers may also require a declaration of originality for other written papers compiled for their courses.

I hereby confirm that I am the sole author of the written work here enclosed and that I have compiled it in my own words. Parts excepted are corrections of form and content by the supervisor.

Title of work (in block letters):

Lake ice detection using low spatial resolution satellite images

Authored by (in block letters):

For papers written by groups the names of all authors are required.

Name(s):	**First name(s):**
Kälin	Ursula

With my signature I confirm that

- I have committed none of the forms of plagiarism described in the 'Citation etiquette' information sheet.
- I have documented all methods, data and processes truthfully.
- I have not manipulated any data.
- I have mentioned all persons who were significant facilitators of the work.

I am aware that the work may be screened electronically for plagiarism.

Place, date	**Signature(s)**
Zürich 12.01.2018	U. Kälin

For papers written by groups the names of all authors are required. Their signatures collectively guarantee the entire content of the written paper.