DGPF

ORIGINAL ARTICLE

Recent Ice Trends in Swiss Mountain Lakes: 20-year Analysis of MODIS Imagery

Manu Tom[1,2,3] · Tianyu Wu[1] · Emmanuel Baltsavias[1] · Konrad Schindler[1]

Abstract
Depleting lake ice is a climate change indicator, just like sea-level rise or glacial retreat. Monitoring Lake Ice Phenology (LIP) is useful because long-term freezing and thawing patterns serve as sentinels to understand regional and global climate change. We report a study for the Oberengadin region of Switzerland, where several small- and medium-sized mountain lakes are located. We observe the LIP events, such as freeze-up, break-up and ice cover duration, across two decades (2000–2020) from optical satellite images. We analyse the time series of MODIS imagery by estimating spatially resolved maps of lake ice for these Alpine lakes with supervised machine learning. To train the classifier we rely on reference data annotated manually based on webcam images. From the ice maps, we derive long-term LIP trends. Since the webcam data are only available for two winters, we cross-check our results against the operational MODIS and VIIRS snow products. We find a change in complete freeze duration of -0.76 and -0.89 days per annum for lakes Sils and Silvaplana, respectively. Furthermore, we observe plausible correlations of the LIP trends with climate data measured at nearby meteorological stations. We notice that mean winter air temperature has a negative correlation with the freeze duration and break-up events and a positive correlation with the freeze-up events. Additionally, we observe a strong negative correlation of sunshine during the winter months with the freeze duration and break-up events.

Zusammenfassung
Aktuelle Eisentwicklung auf Schweizer Bergseen: 20-jährige Analyse von MODIS-Daten. Abnehmende Vereisung von Seen ist, wie der Anstieg des Meeresspiegels oder das Abschmelzen von Gletschern, ein Indikator für den Klimawandel. Die Beobachtung von phänologischen Ereignissen wie Beginn und Ende der Vereisung ist nützlich, weil diese Anhaltspunkte für die Beschreibung des regionalen und globalen Klimawandels bieten. Dieser Beitrag berichtet über eine Fallstudie zu kleinen bis mittelgrossen Seen in der Region Oberengadin (Schweiz). Beobachtet werden Beginn und Ende der Vereisung sowie die Dauer der Eisbedeckung über zwei Jahrzehnte (2000–2020), anhand von optischen Satellitenbildern. Auf Basis von MODIS-Bildfolgen (sowie VIIRS-Bildfolgen, soweit verfügbar) wurden, mit Hilfe von maschinellem Lernen, explizite Karten des Seeeises erstellt. Als Trainingsdaten für den Klassifikator dienen händische Annotationen auf der Grundlage von Webcam-Bildern. Aus den Eiskarten

✉ Manu Tom
  manu.tom@eawag.ch

  Tianyu Wu
  wuti@student.ethz.ch

  Emmanuel Baltsavias
  emmanuel.baltsavias@geod.baug.ethz.ch

  Konrad Schindler
  schindler@ethz.ch

1 Photogrammetry and Remote Sensing Group, ETH Zurich, Zurich 8093, Switzerland

2 Glaciology and Geomorphodynamics Group, University of Zurich, Zurich 8057, Switzerland

3 Remote Sensing Group, Swiss Federal Institute of Aquatic Science and Technology, Dübendorf 8600, Switzerland

🖄 Springer

wurden Trends für die Seevereisung abgeleitet. Da Webcam-Daten nur für zwei Winter verfügbar sind wurden die Ergebnisse auch mit operationellen MODIS- und VIIRS-Schneeprodukten verglichen. Es zeigt sich eine Verkürzung der Vereisungsdauer um 0.76 bzw. 0.89 Tage pro Jahr für den Silsersee und den Silvaplanersee. Weiters beobachten wir plausible Korrelationen der Trends mit Klimadaten von nahe gelegenen meteorologischen Stationen. Konkret korreliert die mittlere Lufttemperatur der Wintermonate stark mit Beginn, Ende und Dauer der Vereisung. Ebenso gibt es eine starke Korrelation zwischen der Sonnenscheindauer während der Wintermonate und dem Ende sowie der Dauer der Vereisung.

1 Introduction

Lake ice cover is part of the essential climate variable: *lakes* (https://gcos.wmo.int/en/essential-climate-variables/lakes/). Many studies have reported on the response of Lake Ice Phenology (LIP) to climate variations (Brown and Duguay 2010; Duguay et al. 2006; Howell et al. 2009; Kang et al. 2012; Sharma et al. 2019; Surdu et al. 2014). Local weather patterns and lake ice formation processes are inter-connected (Brown and Duguay 2010). Hence, monitoring the long-term LIP trends can provide integral cues on the local and global climate. Increasing temperatures cause decreasing trends in the lake ice formation process. Air temperature in the vicinity of a lake affects the ice formation process within the lake and vice versa. Moreover, there are potential positive feedbacks, as frozen lakes have a higher albedo (especially when covered with snow), and thus lower absorption and evaporation (Slater et al. 2021; Wang et al. 2018). In addition to its contribution to climate studies, lake ice monitoring is also useful to organise safe transportation, especially in lakes that freeze partially, to conserve freshwater ecosystems, to trigger warnings against ice shoves caused by wind during the break-up period, and for winter tourism (Hampton and et al. 2017; Hirose et al. 2008; Knoll et al. 2019; Mullan et al. 2017).

In this study, we monitor the spatio-temporal extent[1] of ice on lakes of the Oberengadin region in the Swiss Alps (which reliably freeze every winter) daily over 20 winters. From those time series, we derive the dates of the following important LIP events: Freeze-Up Start (FUS), Freeze-Up End (FUE), Break-Up Start (BUS) and Break-Up End (BUE). Using these four dates, we also estimate the Complete Freeze Duration (CFD) and Ice Coverage Duration (ICD). For lake ice monitoring, the Global Climate Observing System (GCOS) office requirement are daily observations, and an accuracy of ± 2 days for the *ice-on/-off* dates (https://gcos.wmo.int/en/essential-climate-varia bles/lakes/ecv-requirements).

In this work, we use only image data from optical satellites[2], and provide a direct, data-driven observation not influenced by model assumptions about the ice formation process. We see satellite imagery as an independent information source and consider image analysis complementary to other methods of lake ice modelling. MODIS and VIIRS have several advantages such as wide area coverage, good spectral and fine temporal resolution (daily), and free availability. Additionally, as opposed to other optical satellites such as Landsat-8, Sentinel-2 and the like, MODIS and VIIRS offer the best spatio-temporal resolution trade-off for the application of single-sensor lake ice monitoring, even though the spatial resolution is moderate (250–1000 m Ground Sampling Distance, GSD). An important asset is the availability of long time series, e.g., MODIS data is available for the entire period since 2000, contrary to other sensor data like airborne or terrestrial photography, webcams etc. We use the linear Support Vector Machine (Cortes and Vapnik 1995, SVM) classifier to perform semantic segmentation and derive the LIP events from the resulting time series by fitting a piece-wise linear model per winter. Additionally, we perform a fourfold cross-validation experiment and assess the transferability of the learned model across space and time.

To our knowledge, the only operational lake ice product is the Climate Change Initiative Lake Ice Cover (Crétaux and et al. 2020), however, our target lakes are not included among the 250 lakes it covers. A second product, Copernicus Lake Ice Extent (LIE, https://land.copernicus.eu/global/products/lie), is still in the pre-operational stage due to accuracy issues, and coverage only starts in 2017. Though not designed for lake ice, the MODIS Snow Product (Hall and Riggs 2016, MSP) and VIIRS Snow Product (VSP, https://nsidc.org/sites/nsidc.org/ files/technical-references/VIIRS-snow-products-user-guide-final.pdf) are also reasonable proxies since lakes in the Alps are typically snow-covered for most of the frozen period. In Tom et al. (2020b), we have reported a comparison of specifications of the operational lake ice/snow products. We cross-check our results with these two snow products, see Sect. 4.1.

For Swiss lakes, a previous study (Hendricks Franssen and Scherrer 2008) has verified that the lake ice formation is strongly correlated with the surrounding air temperature. The authors deducted an empirical relationship between the sum of negative degree days (also called Accumulated Freezing Degree Days, AFDD) and the lake ice formation process, and modelled the probability of ice cover via binomial logistic regression. For that study, temperature data from 1901–2006 was gathered to study eleven lakes in the

---

[1] We measure ice cover, but not ice thickness.

[2] We have previously also used webcams (Tom et al. 2020b; Xiao et al. 2018) and Sentinel-1 Synthetic Aperture Radar (Tom et al. 2020a, SAR) for lake ice monitoring.

lower-lying Swiss plateau. However, none of the mountain lakes we target were included.

The aims of the proposed study are twofold. First, it supplements the existing databases about lake ice on mountain lakes, and further corroborates the accuracy of supervised machine learning for detecting LIP events on small lakes, using only low-resolution imagery. Second, it examines the correlation between LIP and climate data for small lakes in a topographically and climatically complex Alpine environment.

1.1 LIP Trend Analysis Studies with MODIS and VIIRS

The LIP trends of several lakes with different geographical conditions have been studied and reported in the literature. Though most of them used information from various ice databases (e.g., NSIDC), some studies directly derived the trends from radar and optical satellite data. In the following, we focus on studies that, like ours, analyse MODIS and/ or VIIRS optical satellite imagery to examine trends over multiple winters.

Latifovic and Pouliot ([2007]) used the AVHRR historical data record in addition to in-situ measurements to perform long-term (1950–2004) trend analysis of Canadian lakes with areas > 100 km^2, via an automated profile feature extraction procedure. They confirmed later freeze-up (0.12 d/a) and earlier break-up (−0.18 d/a) for the majority of lakes that were analysed and suggested that their procedure to extract the LIP events is not sensor-specific and could be applied to other satellite data, too. Murfitt and Brown ([2017]) also used MODIS data to extract lake ice trends (2001–2014) for the Canadian states of Ontario and Manitoba, and found regionally varying trends.

Zhang and Pavelsky ([2019]) proposed to threshold the red reflectance band of MODIS (threshold values determined with the help of Landsat images) to monitor ice on lakes in Maine (USA), with areas ranging from 0.13 to 305 km^2. Though they analysed data from 296 lakes over 19 years (2000–2018), significant break-up and freeze-up trends were detected only for 3%, respectively 2.1% of the lakes. The same threshold-based approach, with additional post-processing filters, was applied to a new LIP database covering 4241 Alaskan lakes (Zhang et al. [2021]) with areas > 1 km^2, over the period 2000–2019. The estimated significant LIP trends are: later freeze-up (0.29 d/a) and earlier break-up (−0.55 d/a) for 289 and 440 lakes, respectively; and earlier freeze-up (−0.33 d/a) and later break-up (0.75 d/a) for only 11 and 4 lakes, respectively.

Šmejkalová et al. ([2016]) extracted the LIP trends (2000–2013) for 13,300 Arctic lakes (area >1 km^2) using MODIS imagery, and observed a trend towards earlier break-up. They reported a mean shift in BUS in the range: −0.10 d/a (Northern Europe) to −1.05 d/a (central Siberia), and BUE in the range: −0.14 d/a to −0.72 d/a. Kropáček et al. ([2013]) studied the LIP trends of 59 lakes (area > 100 km^2) on the Tibetan Plateau from 2001 to 2010 using MODIS data. However, the estimated LIP trends varied across the target lakes and it was concluded that the 10-year time span is too short to draw a firm conclusion about LIP trends. Gou et al. ([2015]) analysed the ice formation trends (2000–2013) in lake Nam Co (Tibet, area 1920 km^2) using MODIS and in-situ data and found strong correlations with air temperature and wind speed patterns. This study found that high wind speeds during winter time could potentially expedite the freeze-up process. Additionally, this work reported a significant reduction in the total freeze duration. Gou et al. ([2017]) later analysed Nam Co for the period 2000 till 2015 using multiple MODIS products and reported delayed FUS (0.58 d/a) as well as BUS (0.09 d/a), and reduced ice duration (−0.49 d/a) trends. Another study (Yao et al. [2016]) also noted an increasingly shorter freeze duration during the period 2000–2011 when investigating the lakes in the Hoh Xil region (Tibet, 22 lakes with area > 100 km^2), using MODIS, Landsat TM/ETM+, and meteorological data. In addition, that work estimated later freeze-up and earlier break-up trends. They reported that the FUS, FUE, BUS, BUE, CFD and ICD shifted on average by 0.73, 0.34, −1.66, −0.81, −1.91, −2.21 d/a, respectively. Cai et al. ([2019]) also analysed 58 lakes (area > 41 km^2) located on the Tibetan Plateau during the period from 2001 till 2017 using both Terra and Aqua MODIS imagery. For 47 lakes, a later FUS was noticed (0.55 d/a) while for the remaining 11 lakes an earlier FUS was observed (−0.44 d/a). For 50% of the target lakes, an earlier BUE (−0.69 d/a) was noted, however, for the other half a later BUE (0.39 d/a) was observed. Additionally, they reported a reduced ice cover duration for 40 lakes (−0.8 d/a), while for 18 lakes an increase was noted (1.11 d/a).

Yang et al. ([2019]) used MODIS to estimate the LIP trends for 8 large lakes (106 to 3461 km^2) in Northeastern China from 2003 to 2016. Later FUS (0.65 d/a), earlier BUE (−0.19 d/a) and shorter freeze duration (−0.84 d/a) trends were noticed. Qi et al. ([2020]) used AVHRR, MODIS, and Landsat data to extract the LIP of Qinghai lake (China, area of 4294 km^2) for the period 1980–2018. They estimated a shift of 0.16, 0.19, −0.36, and −0.42 d/a for FUS, FUE, BUS and BUE respectively, also pointing towards progressively later freeze-up and earlier break-up. Additionally, they computed the decreasing patterns in ICD (−0.58 d/a) and CFD (−0.52 d/a). That study also identified correlations between the LIP and climate indicators like the AFDD, wind speed, precipitation, etc. during the winter season. Cai et al. ([2020]) used a threshold-based method to extract LIP trends from MODIS snow product for 23 lakes (2001–2018, Xinjiang Uygur Region, area: 11 to 1004 km^2) in China. They found that the ICD decreased (−1.08 d/a) in 16 out of the 23 lakes and

increased (1.18 d/a) for the rest. In addition, they reported later freeze-up (0.52 d/a) and earlier break-up (−0.51 d/a) in 17 and 18 lakes, respectively. Additionally, they found that the freeze-up events are more affected by lake-specific factors such as area and mineralisation; while climatic factors like Lake Surface Water Temperature (LSWT) have more influence on the break-up events. That work also emphasised that LSWT has a stronger influence on the LIP events than the air temperature.

In this work, we deal with small- and medium-sized lakes, whereas most trend studies that used MODIS (Gou et al. 2015; Qi et al. 2019, 2020; Yao et al. 2016) focused on large lakes. Notable exceptions are a study by Šmejkalová et al. (2016), concerning only break-up trends; and by Zhang and Pavelsky (2019), who reported results for lakes as small as 0.13 km^2, but with a limited accuracy of 5–8 days mean absolute error for the ice-on/-off dates, and with no significant trend. In Zhang et al. (2021), the same team processed lakes down to 1 km^2 with an accuracy of 5–11 days. It is interesting to note that, unlike the present study, both these works (Zhang and Pavelsky 2019; Zhang et al. 2021) did not perform any correction for absolute geo-location error when handling small lakes of only a few, often mixed, MODIS pixels. They also found it necessary to tune the thresholds separately for each lake and even use different thresholds for the same lake in different years. On the contrary, for practical reasons we derive a fixed set of parameters for all winters and the whole (admittedly, much smaller) set of target lakes.

Compared to MODIS, the literature on lake ice monitoring with VIIRS is limited. Sütterlin et al. (2017) estimated the LIP dates for winter 2016–2017 in selected Swiss lakes using the LSWT derived from visible and near-infrared reflectances, and VIIRS thermal infrared band (I_5). Later, for winter 2016–2017, Tom et al. (2020b) estimated the LIP dates of lakes Sihl, Sils, Silvaplana, and St. Moritz from VIIRS and MODIS data. To our knowledge, no multi-winter VIIRS-based LIP trend analysis has been reported yet. To summarise, most related works reviewed so far have found trends towards later freeze-up, earlier break-up and declining freeze duration. The prevalent methods are physics-inspired models based on empirical indices and thresholds. To our knowledge, none of the earlier trend studies applied Machine Learning (ML) methods to identify lake ice. Kropáček et al. (2013) used k-means clustering to group the target lakes but not to detect lake ice.

1.2 Lake Ice Observation with Machine Learning

The last decades have seen the rise of ML in remote sensing and the Earth sciences. That is, large-scale statistical data analysis is used to capture the complex input-output relationships in a data-driven manner. In Tom et al. (2020b), we investigated pixel-wise classification of the

Table 1 Key LIP events: Freeze-Up Start (FUS), Freeze-Up End (FUE), Break-Up Start (BUS), Break-Up End (BUE), Ice Coverage Duration (ICD) and Complete Freeze Duration (CFD)

Event	Definition
FUS	30% or more of the non-cloudy lake portion is frozen and the just previous non-cloudy day should be < 30% frozen
FUE	70% or more of the non-cloudy lake portion is frozen and the just previous non-cloudy day should be < 70% frozen
BUS	30% or more of the non-cloudy lake portion is non-frozen and the just previous non-cloudy day should be < 30% non-frozen
BUE	70% or more of the non-cloudy lake portion is non-frozen and the just previous non-cloudy day should be < 70% non-frozen
ICD	BUE—FUS
CFD	BUS—FUE

spatio-temporal extent of lake ice from MODIS and VIIRS imagery with SVM. Each pixel was classified as either *frozen* or *non-frozen* in a supervised manner. We presented extensive experiments on data from two full winters and confirmed the efficacy of SVM for lake ice monitoring with MODIS and VIIRS. In this study, we apply SVM to quantify the 20-year lake ice trends. Xiao et al. (2018) and Prabha et al. (2020) explored the potential of convolutional neural networks for lake ice detection in terrestrial webcam images (RGB). They performed a supervised classification of the lake pixels using the Tiramisu (Jégou et al. 2016), respectively Deeplab v3+ (Chen et al. 2018) networks, into the four classes: *water, ice, snow* and *clutter*. Recently, Hoekstra et al. (2020) proposed an automated approach for ice vs. water classification in RADARSAT-2 data, combining unsupervised iterative region growing using semantics and supervised random forest labelling. A deep learning approach to lake ice detection in Sentinel-1 SAR imagery has been described in Tom et al. (2020a), and achieved promising results, including transferability across lakes and winters. Very recently, Wu et al. (2021) compared the capabilities of four different ML methodologies: multinomial logistic regression, SVM, random forest, and gradient boosting trees for lake ice observation using the MODIS data. They modelled lake ice monitoring as a 3-class (*ice, water, cloud*) supervised classification problem. The four classifiers were tested on 17 large lakes from North America and Europe with areas > 1040 km^2, and achieved >94% accuracy. Random forest and gradient boosting trees showed better generalisation performance on this dataset of large lakes.

Table 2 Details of the lakes (primary source: Wikipedia)

	Sihl	Sils	Silvaplana	St. Moritz
Latitude (°N)	47.14	46.42	46.45	46.49
Longitude (°E)	8.78	9.74	9.79	9.85
Altitude (m)	889	1797	1791	1768
Max. depth (m)	23	71	77	42
Avg. depth (m)	17	35	48	26
Area (km^2)	11.3	4.1	2.7	0.78
Volume (Mm^3)	96	137	140	20
Meteo station	EIN	SIA	SIA	SAM
Latitude (°N)	47.13	46.43	46.43	46.53
Longitude (°E)	8.75	9.77	9.77	9.88
Altitude (m)	910	1804	1804	1708

The last four rows display information about the nearest meteorological stations

1.3 Definitions Used

The pixels that lie completely inside a lake are termed as *clean pixels*. *Non-transition dates* are the days when a lake is either completely frozen or completely non-frozen, the remaining days in a winter season are termed *transition dates*. For each winter, we process all the dates from the beginning of September until the end of May. Definitions of the key LIP events are shown in Table 1.

2 Study Area and Data

2.1 Study Area

We process four small- to medium-sized Swiss Alpine lakes: Sihl (11.3 km^2), Sils (4.1 km^2), Silvaplana (2.7 km^2) and St. Moritz (0.8 km^2), see Fig. 1 and Table 2. For the three small lakes in the region Oberengadin (Sils, Silvaplana, St. Moritz), located at an altitude > 1750 m, there are long in-situ observation series (important for climate studies), and they are also included in the NSIDC lake ice database (https://nsidc.org), although not updated recently. The fourth lake (Sihl) from the region Einsiedeln is relatively larger, lies at a lower altitude on the North part of the Alps, and has different environmental conditions.
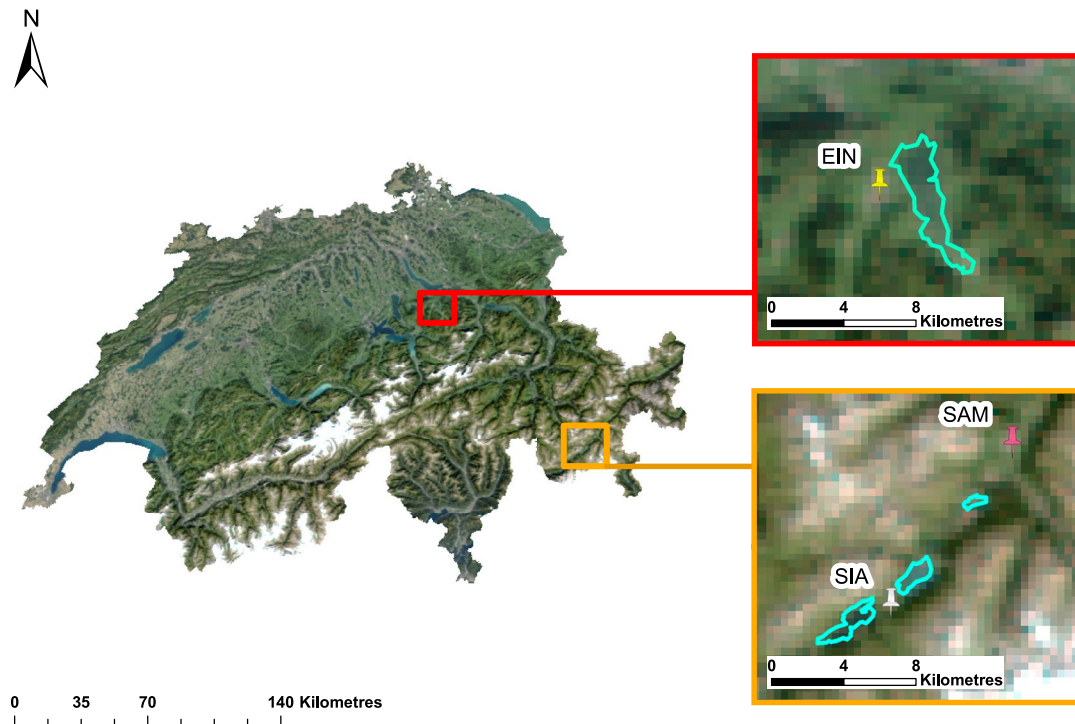

Fig. 1 MODIS orthophoto map (RGB composite, red: B_1, green: B_4, blue: B_3) of Switzerland (left) captured on 7 September 2016. Red and amber rectangles show the regions Einsiedeln (around lake Sihl) and Oberengadin (with lakes Sils, Silvaplana and St. Moritz, from left to right), respectively. Inside each zoomed rectangle on the right, the respective lake outlines are shown in light green and the nearest meteorological stations (EIN:Einsiedeln, SIA: Segl Maria, SAM: Samedan) are marked using pins
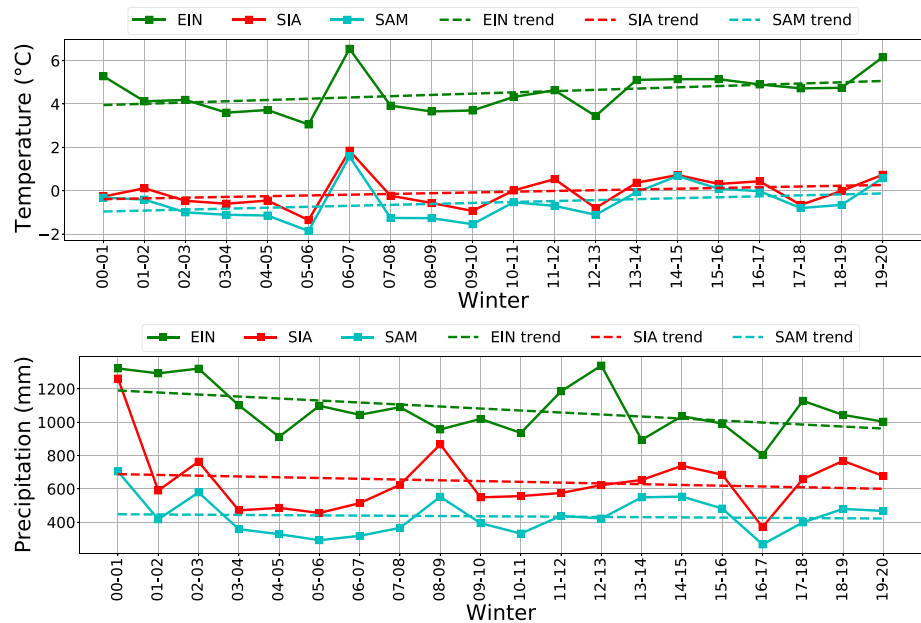
For these four lakes, there is no reference freeze/thaw data available from the past two decades. Hence, we study the weather patterns in the regions near the lakes. For each lake, the temperature and precipitation data recorded at the nearest meteorological stations are shown in Fig. 2. It can be seen that during the past 20 winters, at all the three meteorological stations, the mean temperature follows an increasing trend. On the other hand, precipitation has a decreasing pattern. While Meteoswiss has reported a significant trend of temperature increase in the Swiss Alps since 1864, they have so far not confirmed a notable precipitation trend (https://www.meteoswiss.admin.ch/home/climate/climate-change-in-switzerland/temperature-and-precipitation-trends.html). Over the shorter period of the past 20 winters, precipitation has been slightly declining. Warmer winters at higher altitudes in Switzerland could be linked to a decrease in precipitation, see Rebetez (1996). The pattern of precipitation over the 20 years differs somewhat between the station EIN and the two other (similar) stations, e.g., see the winters 2008–2009, 2012–2013.

2.2 Data

In our analysis, we use the data from Terra MODIS (https://terra.nasa.gov/about/terra-instruments/modis) and Suomi NPP VIIRS (https://ncc.nesdis.noaa.gov/VIIRS/) satellites downloaded from the LAADS (https://ladsweb.modaps.eosdis.nasa.gov) and NOAA (https://www.avl.class.noaa.gov/) databases, respectively. For MODIS processing, we downloaded the MOD02 (geolocated and calibrated radiance, level 1b, Top Of Atmosphere), MOD03 (geolocation) and MOD35_L2 (cloud mask) products and pre-processed

using *MRTSWATH* (https://lpdaac.usgs.gov/tools/modis_reprojection_tool_swath/, re-projection and re-sampling) and *LDOPE* (Roy et al. 2002, cloud mask) software. For VIIRS, we downloaded the Scientific Data Record data for the imagery bands, IICMO and VICMO products for the cloud masks, and GITCO (for image bands) and GMTCO (for cloud masks) for terrain corrected geolocation. VIIRS pre-processing is done using the following software packages: *SatPy* (https://satpy.readthedocs.io/) for assembling the data granules, mapping and re-sampling, *H5py* (https://www.h5py.org) for cloud mask extraction, *PyResample* (https://resample.readthedocs.io) and *GDAL* (https://gdal.org) for re-sampling of cloud masks.

Figure 3 displays more details of the data that we use as a stacked bar chart (one colour per lake). For all target lakes, the total number of pixels in each winter is shown on the *y*-axis, against the winters in chronological order on the *x*-axis. In Fig. 3, note that some winters are relatively less cloudy and hence the number of cloud-free pixels varies across winters, even for the same lake and sensor. Due to its small size there exist no clean pixel for St. Moritz in VIIRS imagery bands (Tom et al. 2020b), hence we exclude it from the VIIRS analysis. It can be inferred from Fig. 3 that, contrary to lake Sihl located at a lower altitude with different surrounding topography, the cloud patterns of lakes Sils, Silvaplana and St. Moritz are quite similar, due to geographical proximity (see also Fig. 1). However, minor differences exist (in a few winters) between the two very nearby lakes Sils and Silvaplana due to cloud mask errors. Different acquisition times of MODIS and VIIRS within a day can also result in varying cloud masks. Figure 3c shows that approximately 40 to 60% of all observations per winter are covered by clouds,

Fig. 3 Row 1 displays the *clean, cloud-free* pixels (*transition* and *non-transition* dates) from the four target lakes. Data from both MODIS (20 winters, 4 lakes) and VIIRS (8 winters, 3 lakes) is displayed. Row 2 shows the percentage of at least 30% Non-Cloudy (NC) days during each winter
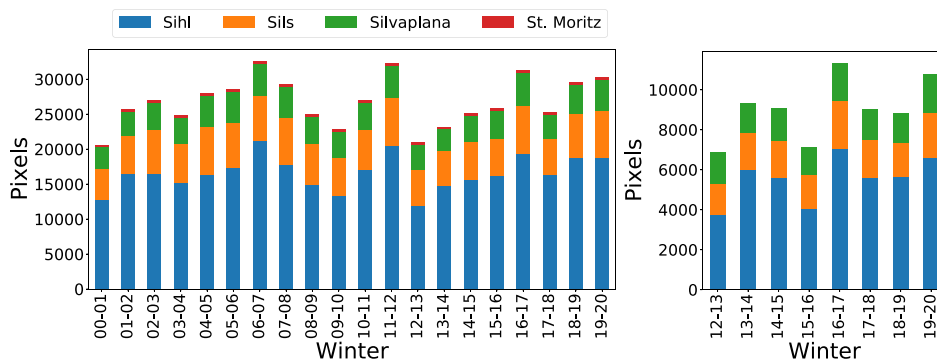
(a) MODIS pixels

(b) VIIRS pixels
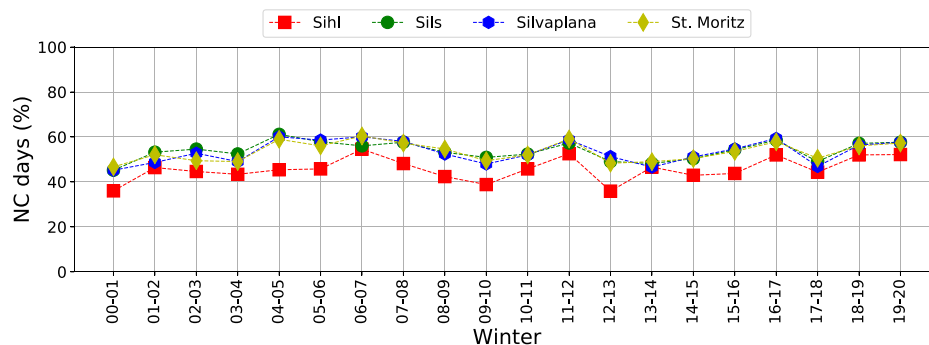
(c) MODIS clouds (20 winters)

strongly reducing the effective temporal resolution of the time series.

Ground truth. Our ground truth is based on the visual interpretation of freely available high-resolution images from webcams monitoring the target lakes. One label (*fully frozen, fully non-frozen, partially frozen*) per day is assigned. Two different operators looked at each image, i.e., a second expert verified the judgement of the first operator to minimise interpretation errors. When deciphering a webcam image was difficult, additional images were used from other webcams viewing the same lake (if available), images from the same webcam but at other acquisition times on the same day, and images of the same webcam for the days before and after the given observation day. We also improved the webcam-based ground truth using sporadic information from media reports, and by visually interpreting Sentinel-2 images, whenever available and cloud-free. No webcam data is available from the winters before 2016–2017. Moreover, the manual interpretation process is labour intensive. Thus, ground truth is available only for winters 2016–2017 and 2017–2018. Even though visual interpretation is the standard practice, a certain level of label

noise inevitably remains in the ground truth, due to factors such as interpretation errors, image compression artefacts, large distance and flat viewing angle on the lake, etc. Furthermore, the webcams used are not optimally mounted for lake ice monitoring and hence do not always cover the full lake area (or even a major
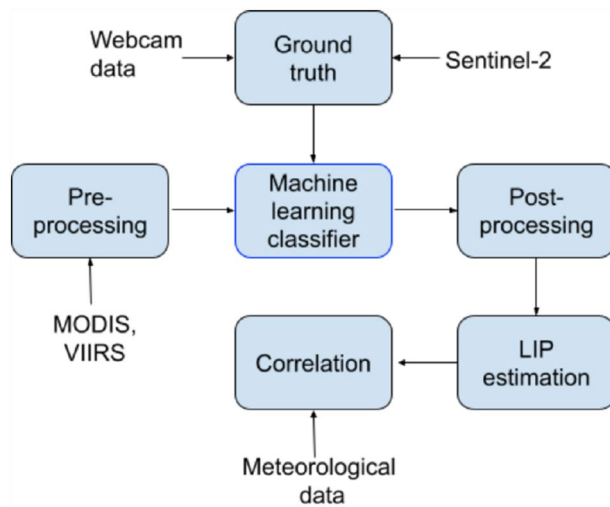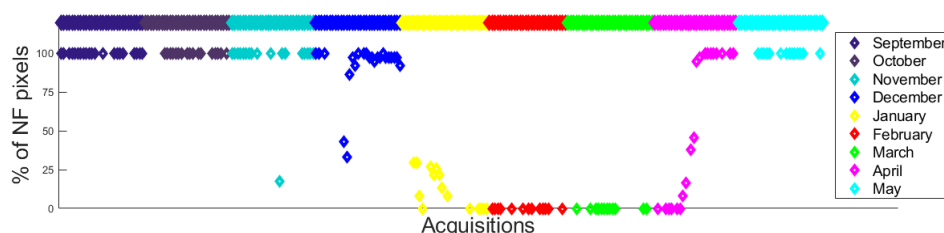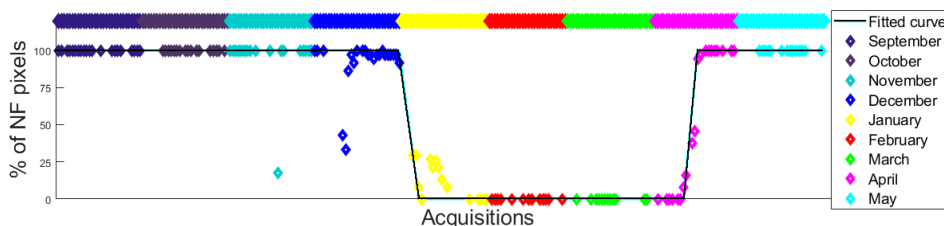

Fig. 4 Flow diagram of the proposed methodology

Fig. 5 Piece-wise linear ("U with wings") curve fitting example. NF indicates Non-Frozen. Results are displayed only for the cloud-free acquisitions

(a) Results timeline before curve fitting

(b) After curve fitting

portion of it), even for the smallest lake St. Moritz. Still, the ground truth serves the purpose, in the sense that it has significantly fewer wrongly labelled pixels than the automatic prediction results. For the winters 2016–2017 and 2017–2018, we see no possibility to obtain a more accurate, spatially explicit ground truth for our task.

3 Methodology

A flowchart of the method is shown in Fig. 4. We perform the pre-processing steps as in Tom et al. (2020b). First, the absolute geolocation error for both sensors (0.75, respectively 0.85 pixels *x*- and *y*-shifts for MODIS; 0.0, respectively 0.3 pixels *x*- and *y*-shifts for VIIRS) are corrected. The generalised (Douglas and Peucker 1973) lake outlines are then back-projected onto the images to extract the clean pixels. Mixed pixels are discarded from the analysis. Binary cloud masks are derived from the respective cloud mask products to limit the analysis only to cloud-free pixels. We super-resolved all low-resolution MODIS bands (500 m, 1000 m) to 250 m using bilinear interpolation prior to the analysis. This step is not required for VIIRS as all used bands have the same GSD (≈ 375 m).

3.1 Machine Learning for Lake Ice Extraction

We model lake ice detection in optical satellite images as a per-pixel 2-class (*frozen, non-frozen*) supervised classification problem and employ a linear SVM (Cortes and Vapnik 1995) classifier. As in Tom et al. (2020b), for each pixel, the feature vector is formed by directly stacking the 12 (5) bands

of MODIS (VIIRS). The bands that offer maximum separability for the task of lake ice monitoring were automatically chosen by the supervised XGBoost feature selection algorithm (Chen and Guestrin 2016). We treat *snow-on-ice* and *snow-free-ice* as a single class: *frozen*. Class *non-frozen* denotes the open water pixels.

While there recently has been a strong interest in deep learning for remote sensing tasks (Camps-Valls et al. 2021), deep neural networks are not suitable for our particular application, due to the scarcity of pixels with reliable ground truth. The lakes that we monitor are small and ground truth is available only for two winters (see Sect. 2.2), which is too little to train data-hungry neural networks. Also, given the large GSD and limited need for spatial context, we do not expect deep models to greatly outperform shallower ones.

3.2 LIP Estimation

Each winter, using the trained ML model, we process all available non-cloudy acquisitions and generate pixel-wise classification maps (one per acquisition). To recover the temporal evolution (per winter), the percentage of non-frozen pixels is computed from each classification map and is plotted on the *y*-axis against the acquisition time on the *x*-axis. Then, as in Tom et al. (2020b), multi-temporal smoothing is performed using a Gaussian kernel with a standard deviation of 0.6 days and window width of 3 days. An example MODIS results (timeline) for lake Sils from winter 2006–2007 is shown in Fig. 5a. Results from different months are displayed in different colours, see the legend.

After multi-temporal smoothing, we find all the potential candidates for the following four critical dates: FUS, FUE,

BUS and BUE, see Table 1 for the corresponding definitions. Within a winter, it is possible that > 1 candidates exist per critical date which all satisfy the respective definition. To weed out some spurious candidates, we enforce the constraint that the four dates must occur in the following chronological order: FUS→FUE→BUS→BUE. Then we exhaustively search for the optimal set of four dates among the remaining candidates. To that end, we fit a continuous, piece-wise linear *"U with wings"* shape to the per-day values of percentage of non-frozen pixels, such that the fitting residuals z are minimised (see example fit in Fig. 5b, shown in black colour). In detail, the loss function for the fit is defined as:

$$L_{LIP} = \frac{1}{P} \cdot \sum_{i=1}^{N} H_\phi(z) \tag{1}$$

where N is the total number of cloud-free acquisitions.

$$H_\phi(z) = \begin{cases} z^2 & |z| \leq \phi \\ 2\phi|z| - \phi^2 & |z| > \phi \end{cases} \tag{2}$$

is the Huber norm of the residual. For the shape parameter ϕ, we use a constant value of 1.35 which offers a good trade-off between the robust l_1-norm for large residuals and the statistically efficient l_2-norm for small residuals (Owen 2006).

Per lake, we assume that each critical date occurs only once per winter, which is always true in Oberengadin. Lake Sihl does not always fully freeze. As it lies outside of the target region and is included mostly to ensure transferability of the ice classifier, we do not extract the LIP events for Sihl. Moreover, we decide to exclude lake St. Moritz since it is too small for the GSD of MODIS (only 4 clean pixels), making the fraction of frozen pixels overly susceptible to noise. We thus prefer to study only the two main lakes in Oberengadin, Sils and Silvaplana, in terms of long-term lake ice trends. These two lakes fully freeze every year and typically have a single freeze-up and break-up period. To further stabilise the LIP estimates we include a weak prior probability for each phenological date, in the form of a diffuse Gaussian distribution.

The prior probability (P) is given by:

$$P = P_{fus} \cdot P_{fue} \cdot P_{bus} \cdot P_{bue} \tag{3}$$

where P_{fus}, P_{fue}, P_{bus}, and P_{bue} are Gaussian normal distributions for the events FUS, FUE, BUS and BUE, respectively. The prior formalises the knowledge that freeze-up normally occurs around the end of December and takes around three days, and break-up occurs around the end of April over a similar period, for both target lakes. To not bias the estimation, but only to minimise the risk of implausible results, we choose very wide Gaussians ($\sigma = 1$ month). Furthermore, we impose a constraint that the duration of freeze-up

Table 3 Test-train split for the MODIS data from 20 winters. Aux set refers to auxiliary set

Test set	Training set
2016–2017	2017–2018
2017–2018	2016–2017
Remaining winters	2016–2017, 2017–2018, aux set

(FUE-FUS) and break-up (BUE-BUS) is not more than two weeks.

We use 30% as the threshold to estimate the four dates. For example, a date is considered a FUS candidate if 30% or more of the non-cloudy portion of the lake is frozen. Some studies based on MODIS (Qi et al. 2020; Reed et al. 2009; Yao et al. 2016) have used 10% as the threshold, while another approach (Kropáček et al. 2013) even employed 5%. All of them monitored larger lakes (45 to 4294 km^2 area). We empirically found that for our rather tiny lakes the above thresholds are too strict and a threshold of 30% is needed to ensure reliable decisions. To see why, consider that in the best case (Sils, cloud-free) a lake has 33 clean pixels, but that number can go down to as few as 7 (Silvaplana, 70% cloud cover). Note also that on such small lakes a large portion of all pixels is very close to the lake's shoreline, where the geo-location error (in the worst case 0.5 pixel) may have a significant impact.

4 Results

In addition to the overall classification accuracy, we report also a stricter measure, mean intersection-over-union (mIoU), which better depicts the performance when the class distribution is imbalanced. Overall classification accuracy is given by:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{4}$$

where *TP, TN, FP*, and *FN* represent true positive pixels, true negatives, false positives and false negatives, respectively. For each class, the Intersection-over-Union score (IoU or Jaccard Index) is given by:

$$IoU = \frac{TP}{TP + FP + FN} \tag{5}$$
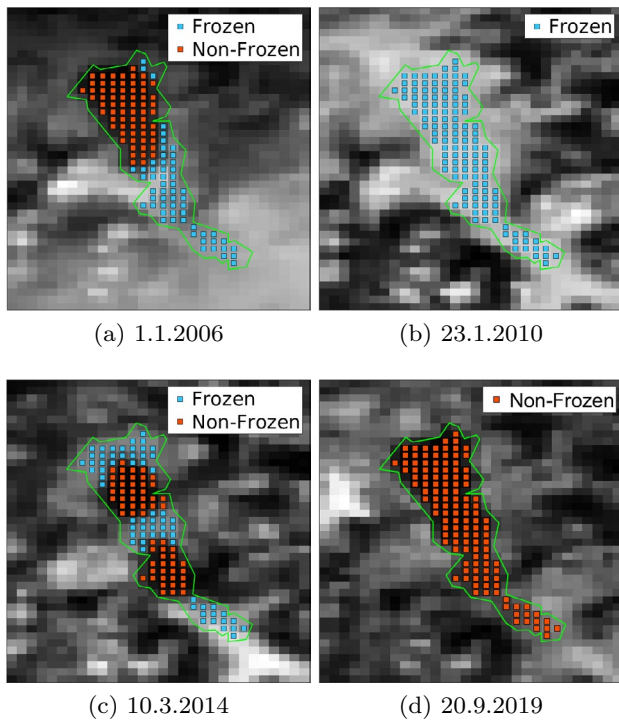
mIoU is the average of the per-class IoUs.

(a) 1.1.2006   (b) 23.1.2010

(c) 10.3.2014   (d) 20.9.2019

Fig. 6 MODIS classification results (lake Sihl, overlaid on band B_1) on selected dates from the past 20 winters. blue and red squares are overlaid on the pixels detected as frozen and non-frozen respectively

4.1 Experiments on MODIS Data from 20 Winters

Test-train split. We process MODIS data from all the 20 winters since 2000–2001 (inclusive). Details of the training set for each tested winter are shown in Table 3. To avoid systematic biases in the estimated ice maps due to overfitting to a particular year, we proceed as follows: we train the linear SVM model on all non-transition dates of 2016–2017 and use it to estimate lake ice coverage for all days in 2017–2018 (including transition dates). We repeat that procedure in the opposite direction, i.e., we train on all non-transition days of 2017–2018 and perform inference for all dates of 2016–2017. Then, we merge all non-transition dates from both winters into a new, larger *two-winter* training set, which we further augment with an *auxiliary dataset*. The latter contains all acquisitions of lakes Sils and Silvaplana captured during the remaining 18 years in September (when the lakes are never frozen) and in February (when the lakes are always frozen). The purpose of the auxiliary dataset is to cover a wider range of weather and lighting conditions that might not have been encountered in the two winters with annotated ground truth, for better transferability. Data of lake Sihl is not included in the auxiliary set, as it does not freeze reliably, St. Moritz is ignored due to its negligible number of pixels. The
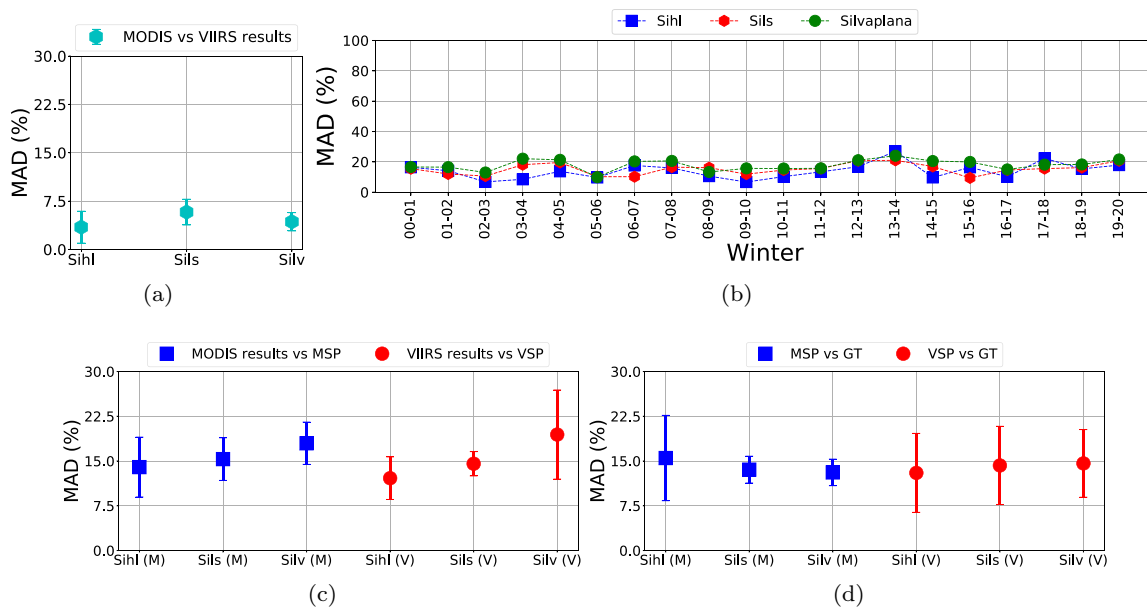

(a)   (b)

(c)   (d)

Fig. 7 (**a**) The comparison of our Machine Learning (ML) products (MODIS and VIIRS) for the 8 common winters. MAD stands for Mean Absolute Difference. Silv represents lake Silvaplana. (**b**) Our (MODIS) ML product (per-winter MAD) vs. MODIS Snow Product (MSP) for 20 winters. (**c**) Comparison of our MODIS (M, 20 winters) and VIIRS (V, 8 winters) ML products with the respective snow products (MSP and VIIRS Snow Product, VSP). (**d**) Deviations between the two snow products and our webcam-based ground truth (GT)

two-winter and auxiliary datasets are merged and used to train a linear SVM model, which is then used to predict ice cover maps for the 18 remaining winters.

Qualitative results. Exemplary qualitative results of lake Sihl are shown in Fig. 6. The respective dates are displayed below each sub-figure. The lake outline overlaid on the MODIS band B_1 is shown in green. Pixels detected as frozen and non-frozen are shown as blue and red squares, respectively. The results include fully-frozen, fully non-frozen and partially frozen days.

Additional check using VIIRS. Direct quantitative analysis is not possible, since no ground truth is available for 18 out of the 20 winters. To validate our ML product (MODIS), we additionally process the VIIRS data from 8 winters (since winter 2012–2013, inclusive) and compare the results. Since a pixel-to-pixel comparison is not straightforward due to different GSDs, we fit the timelines per winter for each lake as described before (Fig. 5a) and compute absolute differences (AD) between the daily estimates for the percentage of frozen pixels. The AD is computed only on dates when both MODIS and VIIRS acquisitions are present, and when the lake is at least 30% cloud-free. The ADs are then aggregated to obtain a Mean Absolute Difference (MAD) per winter. Figure 7a shows, for each lake, the mean and standard deviation of the MAD across the 8 common winters. The low mean values (3.5, 5.8 and 4.3 percent respectively for Sihl, Sils and Silvaplana) show that our MODIS and VIIRS ML products are in good agreement, especially considering that a part of the MAD is due to the difference in GSD between MODIS (250 m) and VIIRS ($\approx$ 375 m). Note also that the acquisition times during the day (and hence the cloud masks) can differ; and that, although the absolute geolocation has been corrected for both sensors, errors up to 0.5 pixels can still remain (Aksakal 2013) and affect the selection of clean pixels near the lake shore.

Comparison with operational snow products. We compare our MODIS (20 winters) and VIIRS (8 winters) ML products to the respective snow products: MODIS snow product (collection 6, MOD10A1), VIIRS snow product (collection 1, VNP10A1F). For the regions of interest, the VIIRS snow product has some data gaps, hence the comparison is done whenever it is available. For actual snow cover mapping, errors of 7–13% have been reported for MODIS snow product (Hall and Riggs 2016). Our findings are in line with this: for the two winters 2016–2017 and 2017–2018 (non-transition days only) we observe an error of 14% w.r.t. our ground truth, see Fig. 7c.

For each lake, we first estimate the percentage of frozen pixels per day using our MODIS and VIIRS ML products. Since a pixel-to-pixel registration is difficult in the presence

of absolute geolocation shifts and/or GSD differences, the daily percentage of frozen pixels is also computed from the snow products and the MAD is estimated for each winter. See Fig. 7d for the comparison of our ML product (MODIS) and the MODIS snow product. For the three lakes, the per-winter MAD is shown on the y-axis against the winters on the x-axis.

Overall, the 20-year time series inter-comparison (per-lake mean and standard deviation of MAD, Fig. 7b) does not suggest large, systematic inconsistencies. On average, our MODIS and VIIRS ML products deviate by mean MAD values of 14–18% and 12–19% respectively. These deviations are only a little higher than the estimated error of the snow products and are relatively stable across different years.

It is important to point out that the snow products are an imperfect proxy for lake ice because a lake can be frozen but not snow-covered, especially near freeze-up when it has not yet snowed onto the ice. Also, mixed ice and water cases go undetected in the MODIS snow product (Hall and Riggs 2016). Figure 7c shows that the snow products are less consistent with the manually annotated ground truth than our ice maps. Most deviations between our estimates and the snow products occur around the transition dates, mostly freeze-up. Additionally, MODIS and VIIRS snow products use a less conservative cloud mask than we do (accepting not only *confident clear* and *probably clear*, but also *uncertain clear* as cloud-free). Despite these issues, the inter-comparison provides a second check for our ML products. For completeness, we note that our algorithm has a similar issue and thin ice is sometimes confused with open water: firstly, snow-free ice is rare and underrepresented in the training set. Secondly, it appears predominantly near the transition dates (especially freeze-up) when we do not have pixel-accurate ground truth. Thirdly, thin ice and open water are difficult to distinguish, we observed that even human interpreters at times confused them when interpreting webcam images.

It is interesting to note that, for both sensors, the mean MAD is inversely proportional to the lake area (see Fig. 7b). This hints at residual errors in the products' geolocation, which would affect smaller lakes more due to the larger fraction of pixels near the lake outline. Besides the < 0.5 -pixel inaccuracy of our maps, inaccurate geolocation of the snow products has been reported (more for MODIS, less for VIIRS) especially for freshwater bodies, due to uncertainties in gridding, reprojection etc. (Hall and Riggs 2016).

LIP trends using MODIS data. As discussed in Sect. 3.2, we fit the "U with wings" polygon to each winter to estimate the four critical dates: FUS, FUE, BUS and BUE. Sometimes, these phenological dates are defined such that a second, consecutive day with similar ice conditions is required to confirm the event. We do not enforce this constraint,
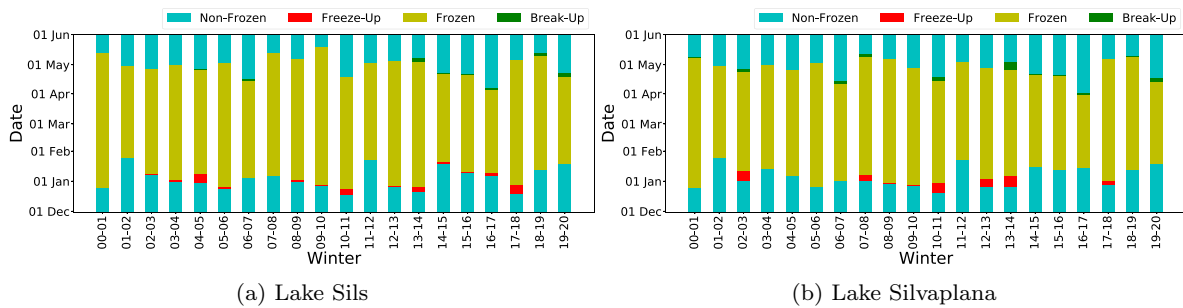
(a) Lake Sils

(b) Lake Silvaplana

Fig. 8 Temporal LIP characteristics estimated from MODIS using linear SVM classifier (20 winters)

because, quite often, the days after a potential freeze-up or break-up date are cloudy, and looking further ahead runs the risk of pruning the correct candidates.

Using the estimated LIP dates from 20 winters, we plot their temporal evolution for lakes Sils and Silvaplana in Fig. 8. On the *y*-axis, all the dates from 1 December to 1 June (we skip September till November since no LIP events were detected during these months), while on the *x*-axis we show the winters in chronological order. In each winter, the non-frozen, freeze-up, frozen and break-up periods are displayed in blue, red, blue and dark green colours, respectively.

It can be seen from Fig. 8 that the freeze-thaw patterns of both lakes vary considerably across winters. For lake Sils (Silvaplana), on average, the FUS occurred on 3 January (5 January) followed by a freeze-up period of 3 (3) days until FUE on 6 January (8 January). Additionally, on average, the lake remained fully frozen (CFD) for 113 (108) days until BUS on 29 April (26 April) and the break-up period lasted 1 (1) day until BUE on 30 April (27 April). The average number of days from FUS to BUE is 117 (112).

The Oberengadin region with lakes Sils and Silvaplana is a single valley (Fig. 1) and hence the two have similar weather conditions. Silvaplana is relatively deeper but has a smaller area than Sils, making them comparable in terms of volume, too. So similar LIP patterns can be expected. However, the clouds above the lakes (especially on partly cloudy days), and the associated cloud mask errors can cause small differences. In winter 2016–2017, the ice-on date of the two lakes, confirmed by visual interpretation of webcams, differ by 7 (low confidence) to 10 (medium confidence) days, see also Tom et al. (2020b).

In most winters, the LIP characteristics of these lakes derived using our approach are in agreement, see Fig. 8. However, there are some outliers too (> 10 days deviation). A notable outlier is the break-up period in winter 2009–2010. For Sils (Silvaplana), BUS and BUE were both estimated as 19 May (28 April). This drift primarily happened because of a huge data gap due to clouds and cloud mask errors. During the period from 28 April till 20 May, Silvaplana had > 30% cloud-free MODIS acquisitions only

on 28 April, 29 April, 8 May and 20 May, and the lake was detected as non-frozen on all these dates. However, Sils had MODIS acquisitions on 28 April, 29 April, 5 May and 19 May. On 5 May the lake was detected as 100% frozen due to a false negative cloud mask, although break-up had started on the two earlier dates (75%, respectively 60% frozen) and the lake was ice-free on May 19. We also checked the Landsat-7 acquisitions on 20 April 2010 and 22 May 2010 and found that both lakes were fully covered by snow on the former date and fully non-frozen on the latter date. No cloud-free Landsat-7 data is available between these two dates. For Sils, the actual BUS probably happened on 29 April (> 30% non-frozen) and BUE soon after (likely on 30 April, since the BUE of Silvaplana was detected on 28 April and Sils was detected < 70% non-frozen on 29 April). However, both dates went undetected until 19 May, because of the clouds in combination with the maximum allowed duration of 2 weeks for the break-up.

In winter 2003–2004, the freeze-up periods of Sils (FUS on 1 January, FUE on 2 January) and Silvaplana (FUS and FUE on 14 January) were also detected far apart, again due to a data gap because of clouds. Sils was estimated 68% and 90% frozen on 1 and 2 January, respectively, so they were chosen as FUS and FUE. On lake Silvaplana, the sequence for 1-5 January was 4%→13%→0%→21%→0% frozen. Then 14 January and 21 January were both found 100% frozen, so the fitting chose 14 January as both FUS and FUE. No cloud-free MODIS data exist on the intermediate dates 6–13 January and 15–20 January, and we could also not find any cloud-free Landsat-7 images between 21 December 2003 and 29 January 2004 (both inclusive) to check but could confirm 0% ice cover on 20 December and 100% cover on 30 January. Connecting all the dots, we speculate that the FUS and FUE of Silvaplana occurred soon after 5 January.

In winter 2013–2014, our method asserts FUE of Sils on 26 December and of Silvaplana on 6 January. Between those dates, there were a number of partially frozen dates, but with more ice cover for Sils than Silvaplana. Additionally, 2–5 January were cloudy, leading the fitting to choose the earlier date for the former, but the later one for the latter.
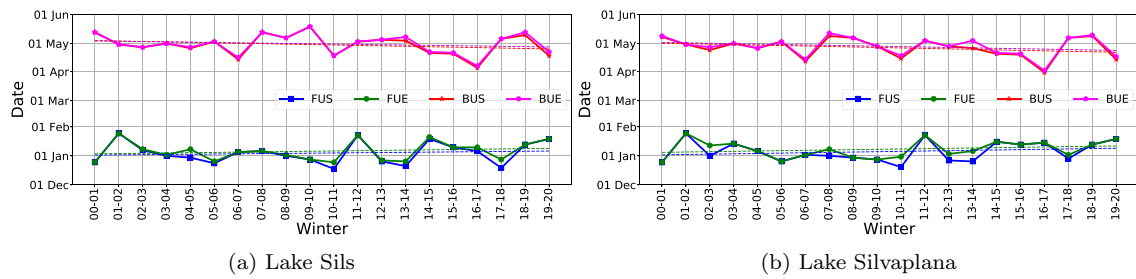
(a) Lake Sils

(b) Lake Silvaplana

Fig. 9 Ice freeze-up (FUS, FUE) and break-up (BUS, BUE) trends (20 winters)
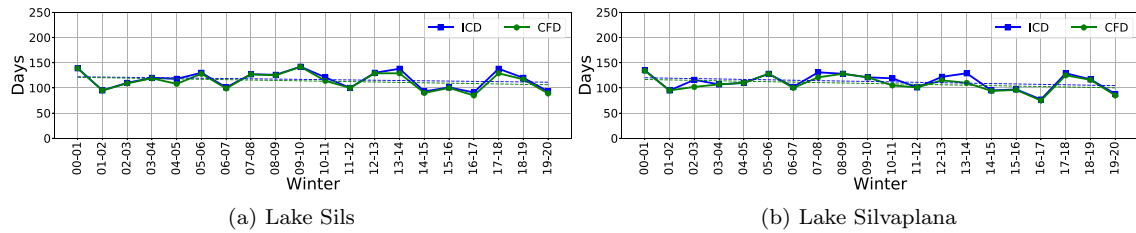

(a) Lake Sils

(b) Lake Silvaplana

Fig. 10 Freeze duration (ICD, CFD) trends (20 winters)

Table 4 Estimated LIP trends (black) and results before manual correction of the automatic results (bold)

Lake	FUS	FUE	BUS	BUE	ICD	CFD
Sils	0.23	0.31	−0.46/**0.47**	−0.32/**0.34**	−0.55/**0.57**	−0.76/**0.78**
Silvaplana	0.45/**0.37**	0.38/**0.36**	−0.51	−0.45	−0.9/**0.82**	−0.89/**0.87**

We again checked with Landsat-7 that on 15 December both lakes were fully non-frozen, whereas on 25 January both lakes were fully snow-covered. There exist no cloud-free Landsat-7 image in between these two dates to pin down the dates more accurately.

In some winters, there is almost no freeze-up and/or break-up period detected by our algorithm. This is partly a byproduct of the relatively loose threshold needed to estimate the initial candidates for our small lakes (see Sect. 3.2), bringing the start and end dates of the transition closer together; and also influenced by frequent cloud cover during the critical transition dates (often more than half of all days c.f. Sect. 3c). For instance, if a couple of adjacent dates are cloudy during break-up (and the real BUS occurred during one of these dates) and on the next non-cloudy day, the lake is estimated 70.1% non-frozen, then our fitting will choose this date as both BUS and BUE.

We go on to analyse the freeze-up and break-up patterns, by plotting the time series of the four critical dates over the past 20 winters for the same two lakes, see Fig. 9. Additionally, per phenological date, we fit a linear trend. Progressively later freeze-up and earlier break-up are apparent for both lakes.

In each winter, we also derive the remaining LIP events (ICD, CFD) listed in Table 1. Their trends are shown in Fig. 10, with the duration in days on the y-axis and the winters on the x-axis. Obviously, ICD and CFD are decreasing for both lakes.

Quantitative trend values that we estimated for all the LIP events of Sils and Silvaplana are shown in Table 4. As explained above, we correct obvious failures of the automatic analysis, and set the following corrections for lake Sils: BUS and BUE occurred on 29 April and 30 April, respectively, in winter 2009–2010. Similarly, for Silvaplana: FUS and FUE occurred on 6 January in winter 2003–2004 and FUE occurred on 26 December in winter 2013–2014. For completeness, we also fit trends without the correction – these differ only slightly and confirm that the corrections hardly impact the overall picture. The trend towards earlier break-up is more pronounced than the one towards later freeze-up, for both Sils and Silvaplana. It is interesting to note that the decrease in freeze duration is stronger for the slightly smaller lake Silvaplana.

Correlation: LIP events and meteorological data. We have also studied the (centred and normalised) cross-correlation
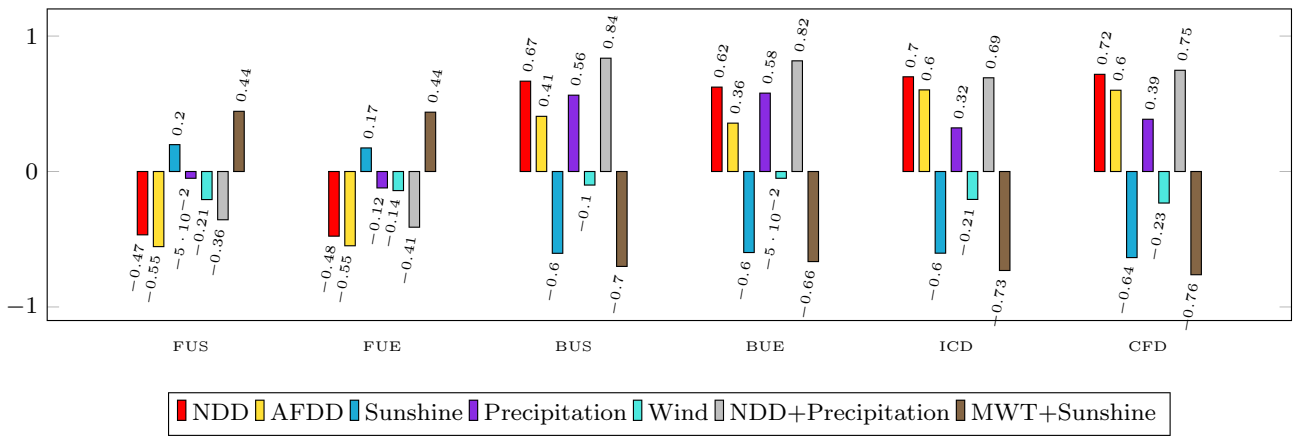
Fig. 11 Bar graphs showing the 20 winter correlation (*y*-axis) of the LIP events with climate variables (*x*-axis) for lake Sils. NDD and AFDD represent Negative Degree Days and Accumulated Freezing Degree Days respectively
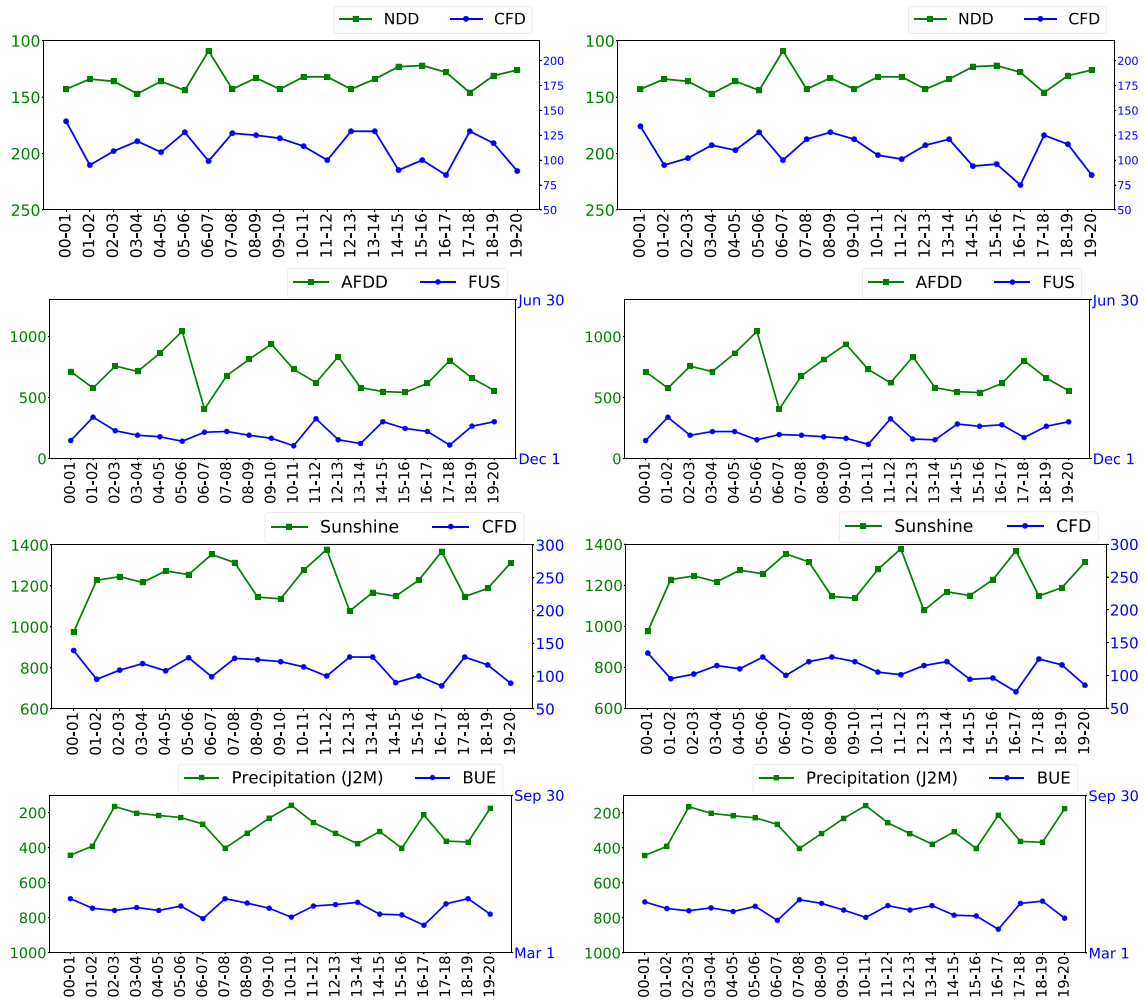

Fig. 12 Relationship between the LIP events and weather variables: NDD (Negative Degree Days, days) and CFD (days) are shown in the first row, AFDD (Accumulated Freezing Degree Days, °C) and FUS in the second row, total winter sunshine (hours) and CFD in the third row, total precipitation (mm) in the months January to May (J2M) and BUE in the last row. Results for lakes Sils and Silvaplana are displayed in left and right columns respectively. In all the sub-figures, *x*-axis displays the winters from 2000–2001 (00–01) till 2019–2020 (19–20)

∈ [−1, 1] between the LIP events (corrected version) and climate variables such as temperature, sunshine duration, precipitation and wind during the 20 winters. The results are shown in Fig. 11 for lake Sils. We do not display the results for lake Silvaplana due to space reasons. Air temperature (2m above ground) and precipitation data were collected from the nearest meteorological station SIA. However, we used the sunshine and wind measurements at station SAM, since these were not available for the complete 20 winter time span at SIA. We did not use the cloud information (number of non-cloudy pixels) from MODIS data as a measure of sunshine duration, since that would ignore the evolution throughout the day, and suffers from a non-negligible amount of cloud mask errors.

Negative Degree Days (NDD) corresponds to the total number of days in a winter with sub-zero air temperature (daily mean, °C). As expected, Fig. 11 shows that NDD has a strong positive correlation with the freeze durations, and break-up events, and a negative correlation with the freeze-up events. We conclude that, indeed, as winters got warmer over the past 20 years the lakes froze later and broke up earlier. The relationship of NDD with CFD is shown in Fig. 12.

AFDD represents the cumulative sum (of daily mean temperature) on the days with average air temperature below the freezing point (0°C) in a winter season. AFDD is a popular proxy for ice thickness (Beyene and Jain 2018; Qi et al. 2020). For both Sils and Silvaplana, AFDD has strong positive correlations with ICD and CFD, a strong negative correlation with the freeze-up events, and a moderate positive correlation with ice break-up events, see Fig. 11, again indicating that in colder winters (higher AFDD) the freeze-up occurs earlier and the break-up later, leading to a longer freeze duration. The relatively weaker correlation for the break-up indicates that freeze-up played a larger role in that event. As an example, the relationship with FUS is shown Fig. 12.

To study the effect of sunshine on LIP events, we correlate the total winter sunshine (hours) with the freeze length events ICD and CFD, total sunshine in the months of September to December (S2D) with the freeze-up events, and the total sunshine from January to May (J2M) with the break-up events. Here, we assume that the sunshine in the months after freeze-up has no connection with freeze-up events. Similarly, we assume that the sunshine in the early winter months (September till December) does not affect the break-up events. We notice a strong negative correlation of the total winter sunshine with ICD, CFD and break-up events. The more sunshine in the months near break-up, the earlier the ice/snow melts, which also reduces the total freeze duration. An example relationship with CFD is visualised in Fig. 12.

We also check the relationship between the LIP events and total precipitation during the winter months. Similar to sunshine analysis, we correlate the total precipitation during the months from September till December, January till May and September till May to the freeze-up, break-up and freeze duration events respectively, see Fig. 11. Notable are the break-up events with a good positive correlation. More precipitation in the months January to May (likely to be predominantly snow), favours later break-up, and vice-versa. The trend for BUE is shown in Fig. 12.

Inspired by Gou et al. (2015), we also looked at the effect of wind on the LIP events, which may also influence lake freezing. We correlated the mean winter wind speed (km/h) with CFD and ICD, mean wind speed from September to December with FUE and FUS, and mean wind speed from January to May with BUS and BUE. However, we did not find any significant correlations, see Fig. 11.

Finally, we explore the correlation of LIP events with the weighted combination of variables (after standardisation by mean and standard deviation) such as NDD, sunshine, precipitation and Mean Winter Temperature (MWT), see Fig. 11 for results. MWT corresponds to the air temperature (°C) averaged over the full winter. It can be noted (Fig. 11) that the weighted combination of NDD and precipitation (equal weights) has a strong positive correlation with the break-up and freeze duration events indicating that during the colder winters (high NDD) with higher precipitation (probably more snow) the break-up occurs later, essentially leading to a higher freeze duration. On the other hand, the weighted combination of MWT and Sunshine (equal weights) has a strong negative correlation with the break-up and freeze-duration events. Hotter winters (high MWT) with more sunshine speed up the ice break-up, thus reducing the freeze duration and vice-versa.

4.2 Ablation Studies with MODIS Data from 2 Winters

As a first ablation study, we combine the data (independently for MODIS and VIIRS) of all the available lakes from winters 2016–2017 and 2017–2018 and perform 4-fold cross-validation and report the overall accuracy and mIoU, see Table 5.

To study how well the classifier generalise across space and time, we train a model on all except one lake (respectively, winter) and test on the held-out lake (winter).
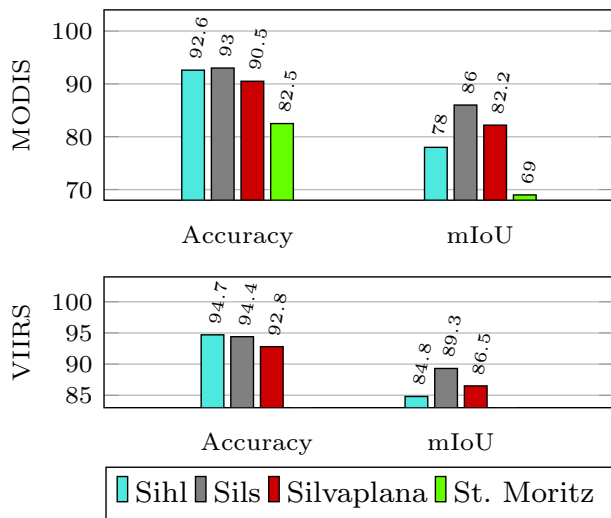
Table 5 Four-fold cross-validation results (in percent).

Sensor	Feature vector	Method	Acc	mIoU
MODIS	12 bands	Linear SVM	93.4	83.9
VIIRS	5 bands	Linear SVM	95.1	88.4

Overall classification accuracy (Acc) and mean intersection-over-union (mIoU) scores are shown

Fig. 13 Transferability study (across lakes) results
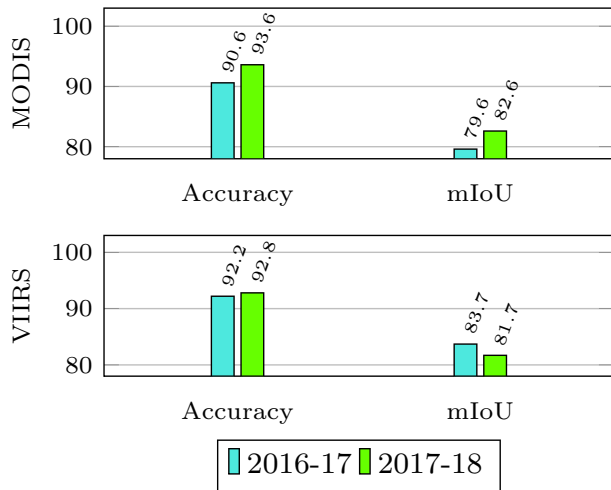

Fig. 14 Transferability study (across winters) results

Figure 13 displays the results (bar graphs showing overall accuracy and mIoU) of the classifier for *leave-one-lake-out* setting on MODIS (top) and VIIRS (bottom) data. It can be seen that the performance varies across lakes and sensors.

On both sensors, the best performance (especially in terms of mIoU) is achieved for the lakes Sils and Silvaplana. This is likely due to them having the most similar characteristics and imaging conditions. I.e., pixels from one of them are representative also of the other one, such that the classifier trained in one of the two generalises well to the other. Lake St. Moritz (only for MODIS) has too few clean pixels per acquisition to draw any conclusions about transferability. However, we still include it in our processing to study how

Table 6 Transferability loss encountered by our classifier.

Sensor	Loss type	Linear SVM
MODIS	Across lakes	3.7/**5.1**
MODIS	Across winters	1.3/**2.8**
VIIRS	Across lakes	1.1/**1.5**
VIIRS	Across winters	2.6/**5.7**

Drop (in percent) for overall accuracy and mIoU are shown in normal and bold fonts, respectively

far lake ice monitoring with MODIS can be pushed (in terms of lake area)—indeed, the classification is > 82.5% correct. Lake Sihl from the region Einsiedeln is different compared to the other three lakes from the region Engadin in terms of area, weather, surrounding topography etc., c.f. Sect. 2.1. Hence, the performance on lake Sihl is interesting to assess geographical transferability over longer distances.

As a second transferability study, more important for our time series analysis, we check how well the trained classifier can be transferred across different winters. We train on one winter and test the model on the held-out winter (*leave-one-winter-out*), see Fig. 14. We only have data from two consecutive winters (2016–2017, 2017–2018) to perform this analysis. Still, we believe that the experiment is representative for transferability to unseen years, since the weather conditions in different years are largely uncorrelated (c.f. Fig. 2). In particular for the two available winters, 2017–2018 was markedly colder than the previous year, see Fig. 2.

The classifier exhibits a certain performance drop when having to generalise beyond the exact training conditions. Table 6 shows the detailed performance drops compared to Table 5. Note that in the ablation studies we must hold out some ground truth for evaluation and therefore have a smaller training set.

4.3 Discussion

In any ML-based system, the variety in the training dataset has a critical influence on the model being learnt. Our dataset consists of small lakes and has a significant class imbalance since we include all cloud-free dates from September till May, of which only a minority is frozen. This is a biased, but realistic scenario, representative of mountain lakes in sub-Arctic and temperate climate zones.

Our MODIS and VIIRS products validate each other in a relative sense (< 5.6% MAD in the worst case, lake Sils), but could be subject to a common bias. In the absence of ground truth, there is no way to assess our absolute accuracy, but as an external check against a methodologically different mapping scheme we inter-compared our ML products against the

respective operational snow products. The deviations were in the expected range (MAD < 20%).

Some errors exist in both MODIS and VIIRS cloud masks. The most critical ones are false negatives, where an actually cloudy pixel goes undetected. Such cases can corrupt model learning and inference and introduce errors in the predicted ice maps. The trade-off between spatial and temporal resolution makes it difficult to monitor smaller lakes—with 21 MODIS (9 VIIRS pixels) for lake Silvaplana and only 4 MODIS pixels for lake St. Moritz, our study goes to the limit in that respect. A further, often-named obstacle for optical satellite observation is occlusions due to clouds, which significantly reduce the effective temporal resolution and also cause irregular gaps in the time series. These unpredictable data gaps are particularly troublesome for ice phenology because the critical events occur over a short time and at times of the year when clouds are frequent in sub-Arctic and mid-latitudes. Such gaps are the main source of error in our LIP estimation, besides cloud mask errors, confusion between open water and thin/floating snow-free-ice, and quantisation effects around hard thresholds. This makes phenological observations challenging—in particular, the uncertainties of our predictions are largest during freeze-up, because of the frequent, but short-lived presence of snow-free ice. Still, it appears that our classifier copes better with the ice reflectance than simple index-based snow products. ML is a powerful tool to recognise the underlying patterns where mechanistic models are lacking or too complicated. Furthermore, unlike the traditional threshold-based approaches which needed separate tuning for each lake and winter, our ML methodology relies on a fixed set of parameters applicable for all our target lakes and winters.

For small lakes, even small geolocalisation errors have a large effect. Our work is also on the challenging ends of the spectrum in terms of local weather conditions: in a drier climate the observations would be less affected by clouds (we process lakes with as little as 30% cloud-free area to obtain sufficient temporal coverage), and fewer clouds mean fewer cloud-mask errors.

5 Conclusion

In this paper, we reported results for selected lakes in Southeastern Switzerland, where we have retrieved lake ice phenology based on MODIS optical image time series. On the one hand, we have shown that, even for small high-Alpine lakes, ice cover can be derived from low spatial resolution MODIS data; and that lake ice phenology retrieved in this manner over 20 years exhibits meaningful correlations with climate data. On the other hand, we have confirmed that a dedicated machine learning scheme maps lake ice more accurately than the classical index- and threshold-based approaches.

As expected, our results point towards later freeze-up (*freeze-up start* at a rate of 0.23 d/a for lake Sils, respectively 0.45 d/a for Silvaplana and *freeze-up end* at a rate of 0.31 d/a for lake Sils, respectively 0.38 d/a for Silvaplana), earlier break-up (*break-up start*: −0.46 d/a for lake Sils, respectively −0.51 d/a for Silvaplana and *break-up end*: −0.32 d/a for lake Sils, respectively −0.45 d/a for Silvaplana) and decreasing freeze duration (*ice coverage duration*: −0.55 d/a for lake Sils, respectively −0.90 d/a for Silvaplana and *complete freeze duration*: −0.76 d/a for lake Sils, respectively −0.89 d/a for Silvaplana). We also observed significant (but not surprising) correlations with climate indicators such as temperature, sunshine and precipitation.

Our approach is generic and easy to apply to other sensors beyond MODIS and VIIRS (given training data). Importantly, the VIIRS sensor is projected to ensure continuity well into the future, opening up the possibility to establish an even longer time series. One solution for the cloud issues of optical satellites is to complement/replace them with radar observations, e.g., Sentinel-1 SAR. We have done preliminary research in this direction (Tom et al. 2020a). SAR-optical data fusion holds great promise, particularly in view of the GCOS requirement to monitor lake ice at daily temporal resolution. We expect that machine learning-based ice detection itself could be further improved with pixel-accurate annotations during transition dates, as well as for more winters and a wider variety of lakes. Unfortunately, gathering such data is not only a considerable, tedious effort, but also poses its own challenges. In most locations and for older data, no corresponding webcam data (or similar regular photography) is available; even when available, its coverage is almost invariably incomplete; and even with usable webcams and satellite imagery, manual annotation is not trivial and prone to mistakes exactly in the situations that are most critical also for computational analysis (such as thin, black ice). We speculate that, given the enormous archive of unlabelled satellite data, approaches such as unsupervised, semi-supervised or active learning may be applicable and could improve the lake ice detector.

Declarations

Conflict of interest The authors declare no conflict of interest.

References

Aksakal SK (2013) Geometric accuracy investigations of SEVIRI high resolution visible (HRV) level 1.5 imagery. Remote Sens 5:2475–2491

Beyene MT, Jain S (2018) Freezing degree-day thresholds and Lake ice-out dates: Understanding the role of El Niño conditions. Int J Climatol 38:4335–4344

Brown LC, Duguay CR (2010) The response and role of ice cover in lake-climate interactions. Prog Phys Geogr 34:671–704

Cai Y, Ke CQ, Li X, Zhang G, Duan Z, Lee H (2019) Variations of lake ice phenology on the Tibetan Plateau from 2001 to 2017 based on MODIS data. J Geophys Res Atmos 124:825–843

Cai Y, Ke CQ, Yao G, Shen X (2020) MODIS-observed variations of lake ice phenology in Xinjiang, China. Clim Change 158:575–592

Camps-Valls G, Tuia D, Zhu XX, Reichstein M (eds) (2021) Deep learning for the Earth Sciences: A comprehensive approach to remote sensing, climate science and geosciences. Wiley and Sons, New York

Chen T, Guestrin C (2016) XGBoost: A Scalable Tree Boosting System. In: International Conference on Knowledge Discovery and Data Mining, San Francisco, USA, 13-17 August

Chen LC, Zhu Y, Papandreou G, Schroff F, Adam H (2018) Encoder-decoder with atrous separable convolution for semantic image segmentation. In: European Conference on Computer Vision, Munich, Germany

Cortes C, Vapnik V (1995) Support-Vector Networks. Mach Learn 20:273–297

Crétaux JF, et al (2020) ESA Lakes Climate Change Initiative (Lakes_cci): Lake products, Version 1.0. Centre for Environmental Data Analysis. Available online: https://doi.org/10.5285/3c324bb4ee394d0d876fe2e1db217378 (accessed 3 August 2021)

Douglas DH, Peucker TK (1973) Algorithms for the reduction of the number of points required to represent a digitized line or its caricature. Cartographica 10:112–122

Duguay C, Prowse T, Bonsal B, Brown R, Lacroix M, Menard P (2006) Recent trends in Canadian lake ice cover. Hydrol Process 20:781–801

Gou P, Ye Q, Wei Q (2015) Lake ice change at the Nam Co Lake on the Tibetan Plateau during 2000–2013 and influencing factors. Prog Geogr 34:1241–1249

Gou P, Ye Q, Che T, Feng Q, Ding B, Lin C, Zong J (2017) Lake ice phenology of Nam Co, Central Tibetan Plateau, China,

derived from multiple MODIS data products. J Great Lakes Res 43:989–998

Hall DK, Riggs GA (2016) MODIS/Terra Snow Cover Daily L3 Global 500m SIN Grid, Version 6. NASA National Snow and Ice Data Center Distributed Active Archive Center Available online. https://doi.org/10.5067/MODIS/MOD10A1.006(accessed 3 August 2021)

Hampton SE et al (2017) Ecology under lake ice. Ecol Lett 20:98–111

Hendricks Franssen HJ, Scherrer SC (2008) Freezing of lakes on the Swiss plateau in the period 1901–2006. Int J Climatol 28:421–433

Hirose T, Kapfer M, Bennett J, Cott P, Manson G, Solomon S (2008) Bottomfast ice mapping and the measurement of ice thickness on tundra lakes using c-band synthetic aperture radar remote sensing. J Am Water Resour Assoc 44:285–292

Hoekstra M, Jiang M, Clausi DA, Duguay C (2020) Lake Ice-Water Classification of RADARSAT-2 Images by Integrating IRGS Segmentation with Pixel-Based Random Forest Labeling. Remote Sens 12:1425

Howell S, Brown L, Kang K, Duguay C (2009) Variability in ice phenology on Great Bear Lake and Great Slave Lake, Northwest Territories, Canada, from SeaWinds/QuikSCAT: 2000–2006. Remote Sens Environ 113:816–834

Jégou S, Drozdzal M, Vázquez D, Romero A, Bengio Y (2016) The One Hundred Layers Tiramisu: Fully Convolutional DenseNets for Semantic Segmentation. In: International Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, USA, 26 June–1 July

Kang K, Duguay C, Howell S (2012) Estimating ice phenology on large northern lakes from AMSR-E: algorithm development and application to Great Bear Lake and Great Slave Lake, Canada. Cryosphere 6:235–254

Knoll L, Sharma S, Denfeld B, Flaim G, Hori Y, Magnuson J, Straile D, Weyhenmeyer G (2019) Consequences of lake and river ice loss on cultural ecosystem services. Limnol Oceanogr Lett 4:119–131

Kropáček J, Maussion F, Chen F, Hoerz S, Hochschild V (2013) Analysis of ice phenology of lakes on the Tibetan Plateau from MODIS data. Cryosphere 7:287–301

Latifovic R, Pouliot D (2007) Analysis of climate change impacts on lake ice phenology in Canada using the historical satellite data record. Remote Sens Environ 106:492–507

Mullan D, Swindles G, Patterson T, Galloway J, Macumber A, Falck H, Crossley L, Chen J, Pisaric M (2017) Climate change and the long-term viability of the world's busiest heavy haul ice road. Theor Appl Climatol 129:1089–1108

Murfitt J, Brown L (2017) Lake ice and temperature trends for Ontario and Manitoba: 2001 to 2014. Hydrol Process 31:3596–3609

Owen AB (2006) A robust hybrid of lasso and ridge regression, Technical Report, Stanford University. Available online: https://statweb.stanford.edu/~owen/reports/hhu.pdf (accessed 3 August 2021)

Prabha R, Tom M, Rothermel M, Baltsavias E, Leal-Taixe L, Schindler K (2020) Lake ice monitoring with webcams and crowd-sourced images. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 2–2020:549–556

Qi M, Yao X, Li X, Duan H, Gao Y, Liu J (2019) Spatiotemporal characteristics of Qinghai Lake ice phenology between 2000 and 2016. J Geogr Sci 29:115–130

Qi M, Liu S, Yao X, Xie F, Gao Y (2020) Monitoring the ice phenology of qinghai lake from 1980 to 2018 using multisource remote sensing data and google earth engine. Remote Sens 12:2217

Rebetez M (1996) Seasonal relationship between temperature, precipitation and snow cover in a mountainous region. Theor Appl Climatol 54:99–106

Reed B, Budde M, Spencer P, Miller AE (2009) Integration of MODIS-derived metrics to assess interannual variability in

snowpack, lake ice, and NDVI in southwest Alaska. Remote Sens Environ 113:1443–1452

Roy D, Borak J, Devadiga S, Wolfe R, Zheng M, Descloitres J (2002) The MODIS land product quality assessment approach. Remote Sens Environ 83:62–76

Slater T, Lawrence RI, Otosaka NI, Shepherd A, Gourmelen N, Jakob L, Tepes P, Gilbert L, Nienow P (2021) Review article: Earth's ice imbalance. Cryosphere 15:233–246

Sharma S, Blagrave K, Magnuson JJ, O's Reilly CM, Oliver S, Batt RD, Magee MR, Straile D, Weyhenmeyer GA, Winslow L, Woolway RL (2019) Widespread loss of lake ice around the Northern Hemisphere in a warming world. Nat Clim Chang 9:227–231

Šmejkalová T, Edwards M, Dash J (2016) Arctic lakes show strong decadal trend in earlier spring ice-out. Sci Rep 6:1–8

Surdu CM, Duguay CR, Brown LC, Fernández Prieto D (2014) Response of ice cover on shallow lakes of the North Slope of Alaska to contemporary climate conditions (1950–2011): Radar remote-sensing and numerical modeling data analysis. Cryosphere 8:167–180

Sütterlin M, Duguay-Tetzlaff A, Wunderle S (2017) Toward a Lake Ice Phenology Derived from VIIRS Data. In: EGU General Assembly, Vienna, Austria, 23-28 April

Tom M, Aguilar R, Imhof P, Leinss S, Baltsavias E, Schindler K (2020) Lake ice detection from Sentinel-1 SAR with deep learning. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 3–2020:409–416

Tom M, Prabha R, Wu T, Baltsavias E, Leal-Taixe L, Schindler K (2020) Ice monitoring in swiss lakes from optical satellites and webcams using machine learning. Remote Sens 12:3555

Wang W, Lee X, Xiao W, Liu S, Schultz N, Wang Y, Zhang M, Zhao L (2018) Global lake evaporation accelerated by changes in surface energy allocation in a warmer climate. Nat Geosci 11:410–414

Wu Y, Duguay CR, Xu L (2021) Assessment of machine learning classifiers for global lake ice cover mapping from MODIS TOA reflectance data. Remote Sens Environ 253:112206

Xiao M, Rothermel M, Tom M, Galliani S, Baltsavias E, Schindler K (2018) Lake ice monitoring with webcams. ISPRS Ann Photogramm Remote Sens Spatial Inf Sci 2:311–317

Yang Q, Song K, Wen Z, Hao X, Fang C (2019) Recent trends of ice phenology for eight large lakes using MODIS products in Northeast China. Int J Remote Sens 40:5388–5410

Yao X, Li L, Zhao J, Sun M, Li J, Gong P, An L (2016) Spatial-temporal variations of lake ice phenology in the Hoh Xil region from 2000 to 2011. J Geogr Sci 26:70–82

Zhang S, Pavelsky TM (2019) Remote sensing of lake ice phenology across a range of lakes sizes, ME, USA. Remote Sens 11:1718

Zhang S, Pavelsky TM, Arp CD, Yang X (2021) Remote sensing of lake ice phenology in Alaska. Environ Res Lett 16:064007