



**University of  
Zurich**<sup>UZH</sup>

# Multi-sensor lake ice monitoring in the European Alps using the Google Earth Engine

GEO 610 Master's Thesis

**Author**

Michael Brechbühler  
13-930-219

**Supervised by**

Dr. Hendrik Wulf

Dr. Daniel Odermatt (EAWAG) (daniel.odermatt@eawag.ch)

**Faculty representative**

Prof. Dr. Alexander Damm

31.07.2022

Department of Geography, University of Zurich

## Abstract

Lake ice cover provides a variety of ecological and cultural ecosystem services. Therefore, the loss of lake ice cover can lead to a range of ecological consequences. With increasing pressure on account of climate-change the timing of seasonal lake ice freeze-up and break-up, also referred to as lake ice phenology (LIP), is subject to change. Recent research shows that lakes around the Northern Hemisphere have progressively earlier ice-off dates, later ice-on dates and a resulting shorter ice duration, up to the point where lakes fail to freeze. Due to its critical linkage to the characterization of the Earth's climate, lake ice phenology has been introduced by the Global Climate Observing System (GCOS) as part of the 54 Essential Climate Variables. We present a proof of concept for an extensive monitoring system for the European Alps based on a multi-sensor remote sensing approach. As part of this thesis, we created a complete processing chain implemented in the Google Earth Engine for the extraction of lake ice coverage and lake ice phenology down to lake sizes of 0.1 km<sup>2</sup> using random forest classifiers. Our accuracy assessment with webcam-based ground observations shows that we can extract phenological key dates with a high accuracy ( $R^2$  of 0.98) and a mean difference of  $-0.17 \pm 4.3$  days. Our analysis using leave-one-lake-out cross-validation showed that the performance for unseen lakes drops significantly. Thus, model generalization over the entire European Alps has to be proven in future studies with more extensive validation data before establishing an operational monitoring system.

# Contents

<b>Abstract</b>	<b>I</b>
<b>Abbreviations</b>	<b>V</b>
<b>List of Figures</b>	<b>VI</b>
<b>List of Tables</b>	<b>VIII</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Lake ice phenology . . . . .	1
1.2 Monitoring lake ice from space . . . . .	2
1.3 Research objectives . . . . .	3
<b>2 Study Area</b>	<b>5</b>
2.1 Training lakes . . . . .	7
<b>3 Data</b>	<b>8</b>
3.1 Optical data . . . . .	9
3.1.1 Landsat 7/8 . . . . .	9
3.1.2 Sentinel-2 . . . . .	10
3.1.3 Sentinel-3 . . . . .	11
3.2 Synthetic aperture radar data . . . . .	12
3.2.1 Sentinel-1 . . . . .	12
3.2.2 ALOS AW3D30 . . . . .	13
3.3 Lake data . . . . .	13
3.3.1 HydroLAKES . . . . .	13
3.3.2 JRC Global Surface Water Mapping Layers . . . . .	13
3.3.3 Webcam imagery . . . . .	14
3.3.4 Auxiliary training data . . . . .	14
3.4 Climate data . . . . .	14
3.4.1 ECMWF ERA5-Land . . . . .	14
3.4.2 MeteoSwiss Grid-Data Products . . . . .	15
<b>4 Methods</b>	<b>16</b>
4.1 Data preprocessing . . . . .	17
4.1.1 Lake masking . . . . .	17

4.1.2	Optical data . . . . .	18
4.1.2.1	Cloud masking . . . . .	18
4.1.2.2	Spectral indices . . . . .	20
4.1.3	Synthetic aperture radar data . . . . .	21
4.1.4	Temporal features . . . . .	22
4.1.5	Textures . . . . .	22
4.2	Training set generation . . . . .	23
4.2.1	Manual approach . . . . .	24
4.2.2	Automated approach . . . . .	25
4.3	Classification . . . . .	25
4.3.1	Random forest classifier . . . . .	25
4.3.2	Training procedure . . . . .	27
4.3.3	Data balancing . . . . .	27
4.3.4	Feature selection and feature importance . . . . .	27
4.3.5	Validation of classification model . . . . .	28
4.3.5.1	Validation metrics . . . . .	28
4.3.5.2	k-fold cross-validation . . . . .	29
4.3.5.3	Leave one lake out cross-validation . . . . .	30
4.3.6	Hyper-parameter tuning . . . . .	31
4.4	Information extraction . . . . .	31
4.4.1	Lake ice coverage . . . . .	32
4.4.2	Lake ice phenology . . . . .	32
4.5	Validation of lake ice phenology . . . . .	32
4.6	Correlation analysis . . . . .	32
<b>5</b>	<b>Results</b>	<b>34</b>
5.1	Spectral class profiles . . . . .	34
5.2	Training procedure . . . . .	35
5.2.1	Feature importance . . . . .	35
5.2.2	Effects of feature enhancement . . . . .	38
5.2.3	Hyper-parameter tuning results . . . . .	39
5.3	Validation results . . . . .	40
5.3.1	Binary Classification . . . . .	40
5.3.2	Leave one lake out cross-validation . . . . .	40
5.3.3	Lake ice phenology event extraction . . . . .	41
5.4	Correlation analysis . . . . .	42
5.5	Google Earth Engine Application . . . . .	44
<b>6</b>	<b>Discussion</b>	<b>45</b>
6.1	Strengths and weaknesses of our Methods . . . . .	45
6.2	Lake ice phenology event extraction . . . . .	47
6.2.1	Problems of input imagery . . . . .	48

6.3 Implications of correlation analysis . . . . .	49
<b>7 Conclusion</b>	<b>51</b>
<b>Bibliography</b>	<b>54</b>
<b>A Coverage frequency figures</b>	<b>58</b>
<b>B Leave-one-lake-out cross-validation</b>	<b>60</b>
<b>Acknowledgements</b>	<b>61</b>
<b>Personal Declaration</b>	<b>62</b>

## Abbreviations

<b>ALOS</b>	Advanced Land Observing Satellite	<b>LFR</b>	Land Full Resolution
<b>BOA</b>	Bottom-Of-Atmosphere	<b>LIC</b>	Lake Ice Coverage
<b>BSI</b>	Bare Soil Index	<b>LIP</b>	Lake Ice Phenology
<b>BUE</b>	Break-Up End	<b>LOLO</b>	Leave-One-Lake-Out
<b>BUS</b>	Break-Up Start	<b>m.a.s.l</b>	Meters Above Sea Level
<b>CFD</b>	Complete Freeze Duration	<b>ML</b>	Machine Learning
<b>CFMask</b>	C Function of Mask	<b>mNDWI</b>	Modified Normalized Difference Water Index
<b>CV</b>	Cross-Validation	<b>MODIS</b>	Moderate Resolution Imaging Spectroradiometer
<b>DEM</b>	Digital Elevation Model	<b>MSI</b>	Multi-Spectral Instrument
<b>DSM</b>	Digital Surface Model	<b>NASA</b>	National Aeronautics and Space Administration
<b>ECMF</b>	European Centre for Medium-Range Weather Forecasts	<b>NDSI</b>	Normalized Difference Snow Index
<b>ECV</b>	Essential Climate Variable	<b>NDVI</b>	Normalized Difference Vegetation Index
<b>EFR</b>	Earth Observation Full Resolution	<b>NDWI</b>	Normalized Difference Water Index
<b>ESA</b>	European Space Agency	<b>NIR</b>	Near Infrared
<b>ETM+</b>	Enhanced Thematic Mapper Plus	<b>OA</b>	Overall Accuracy
<b>FN</b>	False Negative	<b>OLCI</b>	Ocean and Land Color Instrument
<b>FP</b>	False Positive	<b>OLI</b>	Operational Land Imager
<b>FSC</b>	Fractional Snow Cover	<b>RF</b>	Random Forest
<b>FUE</b>	Freeze-Up End	<b>S-1</b>	Sentinel-1
<b>FUS</b>	Freeze-Up Start	<b>S-2</b>	Sentinel-2
<b>GCOS</b>	Global Climate Observing System	<b>S-3</b>	Sentinel-3
<b>GEE</b>	Google Earth Engine	<b>SAR</b>	Synthetic Aperture Radar
<b>GLCM</b>	Grey Level Co-occurrence Matrix	<b>SLC</b>	Single Look Complex
<b>GRD</b>	Ground Range Detected	<b>SWIR</b>	Shortwave Infrared
<b>ICD</b>	Ice Coverage Duration	<b>TIRS</b>	Thermal Infrared Sensor
<b>IW</b>	Interferometric Wide	<b>TN</b>	True Negative
<b>JAXA</b>	Japan Aerospace Exploration Agency	<b>TOA</b>	Top-Of-Atmosphere
<b>JRC</b>	Joint Research Centre	<b>TP</b>	True Positive
<b>L7</b>	Landsat 7	<b>USGS</b>	United States Geological Survey
<b>L8</b>	Landsat 8	<b>VIS</b>	Visible Spectrum

# List of Figures

1.1	Lineplot - Optical signal . . . . .	2
1.2	Histogram - SAR backscatter signal . . . . .	3
2.1	Study area . . . . .	5
2.2	Histogram study lakes - Elevation and area . . . . .	6
2.3	Histogram study lakes - Seasonality . . . . .	6
2.4	Percentiles plot study lakes - lake shading . . . . .	7
2.5	Training lakes overview map . . . . .	7
3.1	Data availability histogram . . . . .	8
3.2	Landsat 7 SLC-off scene . . . . .	9
3.3	Landsat 8 coverage frequency . . . . .	10
3.4	Sentinel-1A/B coverage frequency . . . . .	12
4.1	Methodology flowchart . . . . .	16
4.2	Lakemask example - Klöntalersee . . . . .	17
4.3	Flowchart: Optical preprocessing . . . . .	18
4.4	Sentinel-3 custom cloudmask . . . . .	19
4.5	Flowchart: SAR preprocessing . . . . .	21
4.6	Manual coarse-classification . . . . .	24
4.7	RF model scheme . . . . .	26
4.8	Data balancing . . . . .	27
4.9	k-fold Cross-Validation . . . . .	30
5.1	Sentinel-2 spectral class profiles . . . . .	34
5.2	Feature importance: Landsat 7 . . . . .	35
5.3	Feature importance: Landsat 8 . . . . .	36
5.4	Feature importance: Sentinel-2 . . . . .	36
5.5	Feature importance: Sentinel-3 . . . . .	37
5.6	Feature importance: Sentinel-1 . . . . .	37
5.7	Feature enhancement comparison: Sentinel-1 LIC . . . . .	38
5.8	Feature enhancement comparison: Sentinel-1 validation . . . . .	38
5.9	Hyper-parameter tuning - Sentinel-1 . . . . .	39
5.10	Validation results - 10-fold cross-validation . . . . .	40
5.11	LIP accuracy assessment - Linear regression analysis . . . . .	42

5.12	LIP accuracy assessment - Histogram . . . . .	42
5.13	Correlation Analysis: Lake-specific Factors . . . . .	43
5.14	Correlation Analysis: Climatic Factors . . . . .	43
5.15	GEE App Overview . . . . .	44
6.1	Sentinel-3: Geometric inaccuracy and pixel mixing . . . . .	48
6.2	Sentinel-1: Layover effects . . . . .	49
A.1	Landsat 8 coverage frequency . . . . .	58
A.2	Sentinel-2A/B coverage frequency . . . . .	58
A.3	Sentinel-3A/B coverage frequency . . . . .	59
B.1	Validation results - Leave-one-lake-out cross-validation . . . . .	60

## List of Tables

2.1	Training lakes . . . . .	7
3.1	HydroLAKES lake attributes . . . . .	13
3.2	MeteoSwiss Grid-Data Products . . . . .	15
4.1	Feature enhancement: Spectral indices . . . . .	20
4.2	Feature enhancement: Temporal features . . . . .	22
4.3	Feature enhancement: Textural features . . . . .	23
4.4	Manual coarse classification classes . . . . .	23
4.5	Manual approach: Class labels . . . . .	25
4.6	Cohens kappa interpretation . . . . .	29
4.7	Random forest model settings . . . . .	31
4.8	Lake ice phenology event definitions . . . . .	32
5.1	Hyper-parameter tuning results . . . . .	39
5.2	LOLO-CV results: Sensor-aggregated . . . . .	41
5.3	LOLO-CV results: Lake-aggregated . . . . .	41

# 1 Introduction

## 1.1 Lake ice phenology

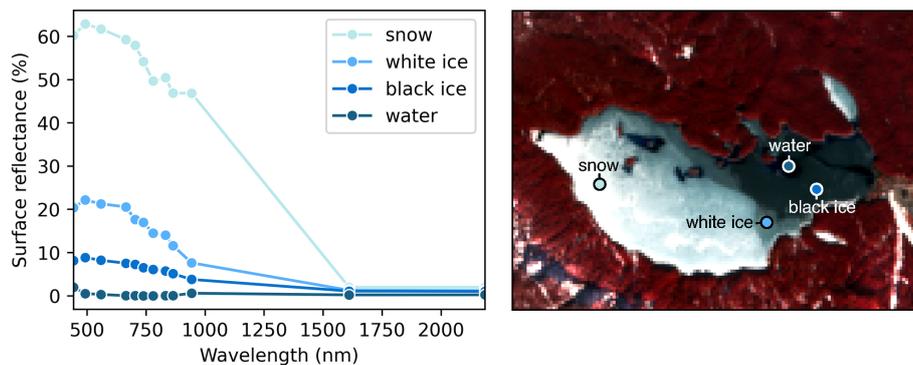
Lake ice cover provides a variety of ecological and cultural ecosystem services. Therefore, the loss of ice cover can lead to a range of ecological consequences such as reduction in water quantity, degradation of water quality, reduction in dissolved oxygen and proliferation of algal blooms (Sharma et al., 2019). Cultural services affected include the loss of important ice roads and the ability for ice fishing and recreational ice skating (Knoll et al., 2019). Furthermore the loss of freshwater lake ice has been identified as one of the earliest observed impacts of climatic warming (Walsh et al., 1998) and the identification of freezing and thawing cycles plays an important indicator in the modelling of the regional climate change and climate variability. With increasing temperatures due to global climate change, the timing of lake-ice freezing and breaking up is changing. This timing of seasonal lake ice freeze-up and break-up is also referred to as lake ice phenology (LIP). Trends in delayed freeze-up and shortened ice duration have been found to be six times faster in the last 25-year period (1992–2016) than in the previous quarter centuries (Sharma et al., 2021). Recent research also shows that lakes around the Northern Hemisphere have progressively earlier ice-off dates, later ice-on dates and a resulting shorter ice duration, up to the point where lakes fail to freeze (Sharma et al., 2020).

Due to its critical linkage to the characterization of the Earth's climate, lake ice phenology variables have been introduced by the Global Climate Observing System (GCOS) as part of the 54 Essential Climate Variables (ECV). This means that the measurement of these variables is a necessity for the thorough understanding and prediction of the Earth's climate. With the implementation plan of 2016, GCOS has established the requirements for lake ice coverage with daily measurements at < 300 m resolution and < 10% uncertainty (GCOS, 2016). Uncertainty requirements for LIP-events, that describe the timing of the freeze-up and break-up cycles, have been suggested at  $\pm 2$  days (Tom et al., 2019).

## 1.2 Monitoring lake ice from space

Ice records based on in-situ observations play a critical part in the understanding of freshwater ice cycles. In-situ records have contributed the majority of working knowledge on the topic, with some records dating back as early as 1443 (Sharma et al., 2020). However, due to its labor-intensive nature, the available records are spatially and temporally non-comprehensive. Thus satellite remote sensing is key to achieve economically feasible, frequent and also comprehensive measurements of lake ice phenology. Previous approaches have made use of **two main types** of data sources.

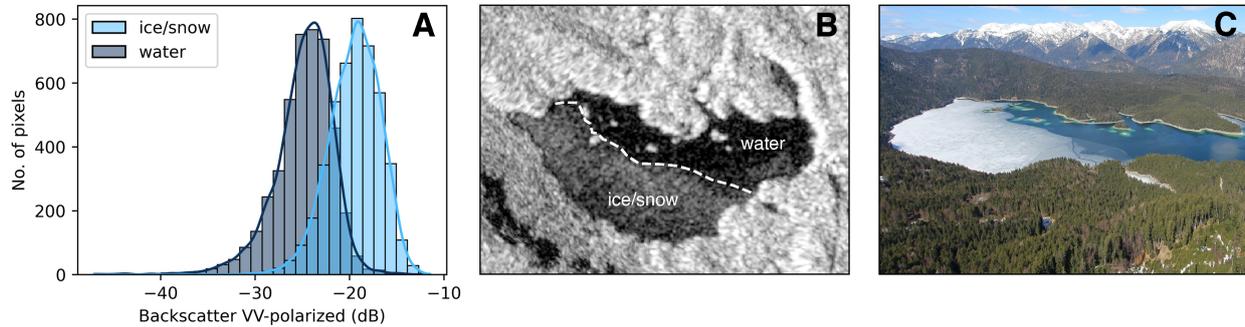
Extraction of lake ice phenology based on **optical remote sensing** makes use of the different spectral characteristics of snow, ice, and water. These surfaces can usually be distinguished by their differing reflectance (Figure 1.1) in the visible and thermal infrared (TIR) wavelengths. Depending on its formation, ice can further show a range of optical characteristics. A common distinction is made between low reflective black and highly reflective white ice (Sharma et al., 2020). Black ice forms in a downwards direction from the bottom of the ice layer as the ice continuously grows due to freezing lake water. White ice is formed from the ice surface upwards, when water is introduced into the snow matrix and the resultant slush refreezes. The introduced water can come from rain, melting snow, or the intrusion of lake water as the weight of the snow layer depresses the ice surface below the water level (Brown and Duguay, 2010). Multi-spectral optical sensors have been used in a multitude of studies to obtain reasonable results in the identification of lake ice formation cycles (Zhang et al., 2021; Yang et al., 2021; Tom et al., 2019; Latifovic and Pouliot, 2007; Nolan et al., 2003).



**Figure 1.1:** Lineplot of the optical signal (left) sampled from different surface classes based on Sentinel-2B surface reflectance over Lake Eibsee in Grainau, Germany at the 20th of March 2019. The corresponding Sentinel-2B image (right) with labeled sampling points is displayed as false color composite.

Extraction of lake ice phenology based on **active microwave remote sensing** such as synthetic aperture radar (SAR) has also been successfully demonstrated (Murfitt and Duguay, 2020; Tom et al., 2020; Hoekstra et al., 2020; Nolan et al., 2003). In this approach the detection of lake ice is based on the higher back-scatter of snow and ice relative to open water (Figure 1.2). Due to the specular reflection of water a very low return signal can be observed for open water. On the other

hand, snow and ice introduces backscatter at the uneven air-ice/snow and water-ice interfaces and additional scattering due to air bubbles and cracks within the ice/snow layer (Atwood et al., 2015; Koskinen et al., 2000).



**Figure 1.2:** SAR VV-polarized backscatter signal from Sentinel-1B of ice/snow compared against water surfaces. The histogram (A) visualizes the distribution of signal strengths of the two different surface classes with equal random samples (4,000 each) over Lake Eibsee in Grainau, Germany at the 27th of March 2019. The lines indicate the estimated kernel density. The corresponding Sentinel-1B image with annotations (B) and a high-resolution webcam image from the same day (C) are displayed for context.

In all of the mentioned studies the focus has been set on a single or a small number of lakes. To date, no study has extracted a ice phenology inventory for a large lake population over multiple years. In addition, the use of low resolution sensors such as MODIS, often limited the usage on large lakes ( $>3.2 \text{ km}^2$ ). There have been only a few exceptions where phenological lake attributes were extracted for small lakes. Even in these cases, only lakes down to a minimal size of  $0.78 \text{ km}^2$  (Tom et al., 2020; Tom et al., 2019) have been studied. In a recent review paper about remotely sensed lake ice phenology, Sharma et al. (2020) have pointed out that the potential of lake ice phenology extraction using optical remote sensing at "medium" resolution (referring to Landsat) remains largely unexplored. Due to the technical trade-offs between temporal and spatial resolution within existing sensors, previous studies have also failed to achieve the GCOS requirements of daily measurements with a resolution lower than 300 m.

This research gap highlights the need for an automated multi-sensor-based monitoring system that allows to detect freezing and thawing cycles down to small lakes ( $>0.1 \text{ km}^2$ ) with state-of-the-art high resolution and high frequency observations to achieve measurements on a daily basis.

### 1.3 Research objectives

The aim of this thesis is to establish such a monitoring system for an alpine study area and explore the feasibility of lake ice phenology monitoring for small lakes. The main working part will be an automated process for the classification of lake ice cover based on machine learning (ML) methods. The classification results will be further processed in an automated extraction of lake ice phenology. We will base our system on the Google Earth Engine (GEE) (Gorelick et al., 2017) which allows for efficient cloud-based processing of the large datasets involved. As a final product

a GEE application with high ease of use should allow untrained users to easily query and display LIP-statistics over a desired lake within our study area.

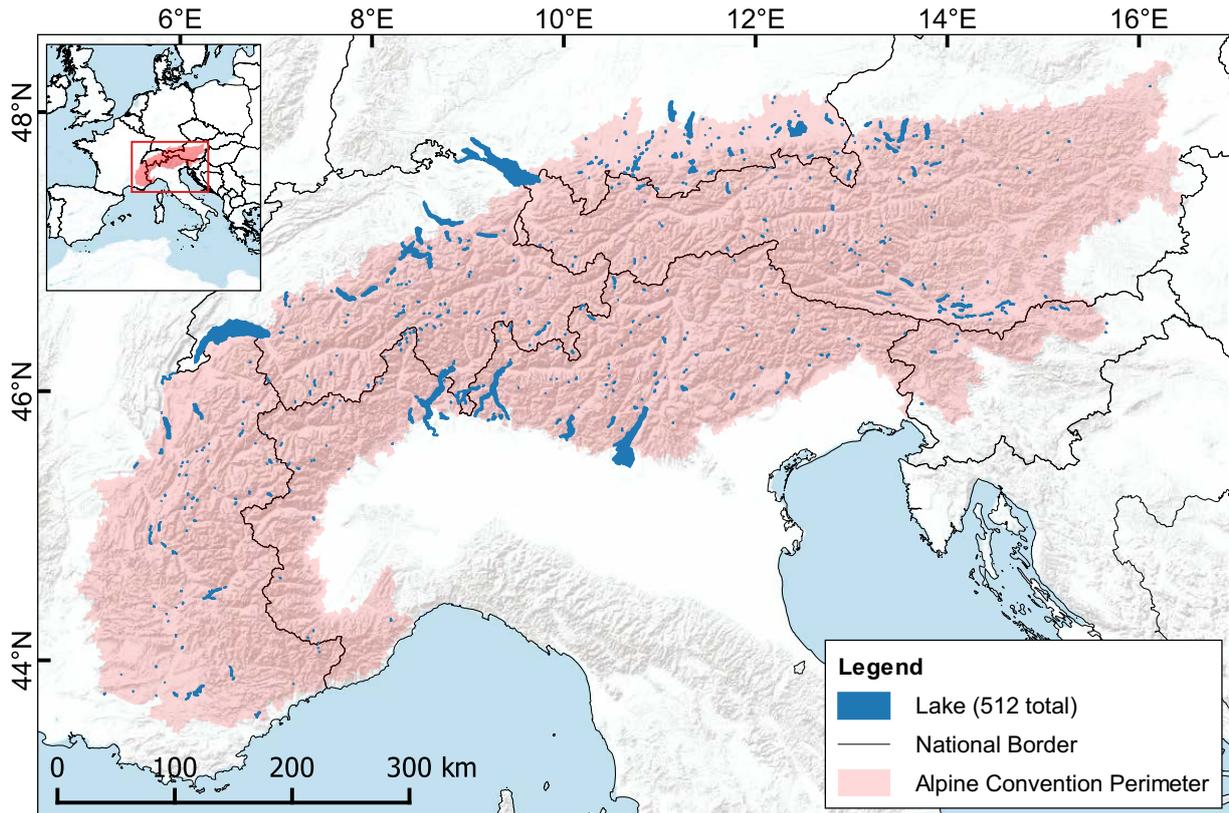
In the first part, we will explore the effectiveness of our ML-based lake ice detection approach with remotely sensed input data. Furthermore, we will try to enhance our classification with the addition of spectral, temporal and spatial features. For this part we want to answer the following research questions:

- **RQ1:** How accurately can lake ice coverage (LIC) be extracted using RS-based methods in an alpine environment?
- **RQ2:** How do the different sensors (S-1, S-2, S-3, L7, L8) compare on the ice/non-ice classification accuracy?
- **RQ3:** Which multi-spectral indices, temporal features, image textures and air temperature features can be used as input features to enhance the classification accuracy of lake ice classifiers?

In the second part, we will use the obtained LIC results to extract lake ice phenology (LIP) and assess the accuracy of the extracted phenological key dates. Furthermore, we try to gain an insight into the relationship of lake ice cycles and climatic as well as lake-specific factors to answer the following research questions:

- **RQ4:** How accurately can lake ice phenology events be extracted using RS-based methods in an alpine environment?
- **RQ5:** Which relationship exists between the measured lake ice phenology and climatic factors (e.g. temperature)?
- **RQ6:** Which relationship exists between the measured lake ice phenology and lake-specific factors (e.g. lake area, lake depth)?

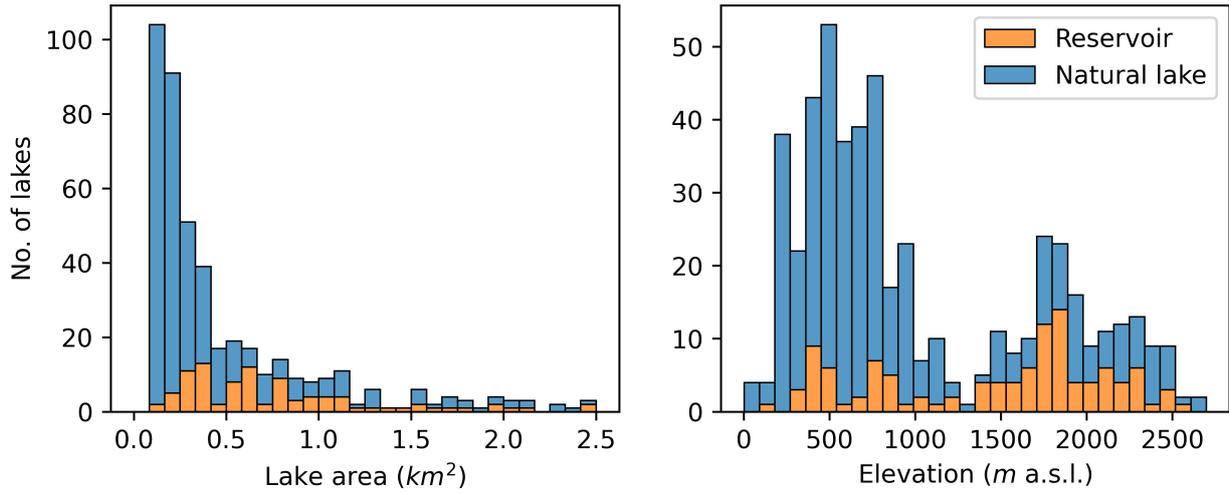
## 2 Study Area



**Figure 2.1:** Overview map of the study area showing the Alpine perimeter defined by the Alpine Convention and the 512 lakes fully or partially within as defined by the HydroLAKES dataset.

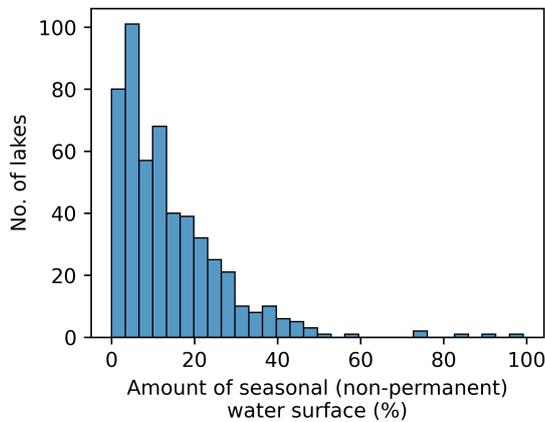
The study area is the European Alps based on the Alpine perimeter as defined by the Alpine Convention (Laner, 2018) and its lakes as defined by the HydroLAKES dataset (Messenger et al., 2016). The Alpine perimeter covers an area of approximately 190,700 km<sup>2</sup> extending over 8 countries.

All lakes fully and partially within this perimeter with an area larger than 0.1 km<sup>2</sup> are part of the studied lakes. This covers a total of 512 lakes from an area of 0.1 up to 572 km<sup>2</sup>. Situated at elevations as low as 62 up to 2,697 m above sea level. The study lakes are made up of 78% (399) natural lakes and 22% (113) reservoirs. With a median lake area of 0.35 km<sup>2</sup> and a median shore length of 3.2 km most of the Alpine lakes (68.8%) are well below the previously studied lake sizes (Figure 2.2).



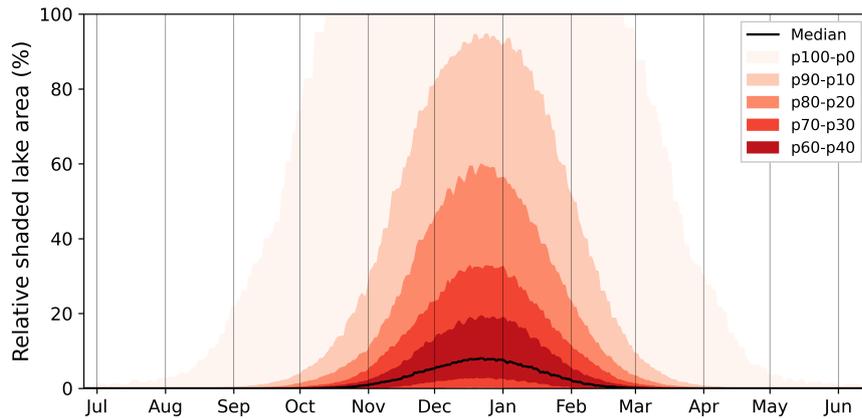
**Figure 2.2:** Histograms of the area (left) and elevation (right) of the study lakes created based on the HydroLAKES attributes. In the lake-area histogram bins above the 90th percentile were truncated to better represent the distribution of the overall data.

As a result of the strong influence of natural seasonal meltwater inflows as well as artificial discharge control in reservoirs, many of the study lakes show seasonal fluctuations in their water level leading to changes in surface area. The median amount of seasonal (non-permanent) water surface lies at 10.7% (Figure 2.3).



**Figure 2.3:** Histogram of the seasonality of the study lakes. Seasonality is presented as the relative area of the non-permanent water surfaces. Seasonality is based on the JRC Yearly Water Classification History v1.3 (Pekel et al., 2016) for the year 2019.

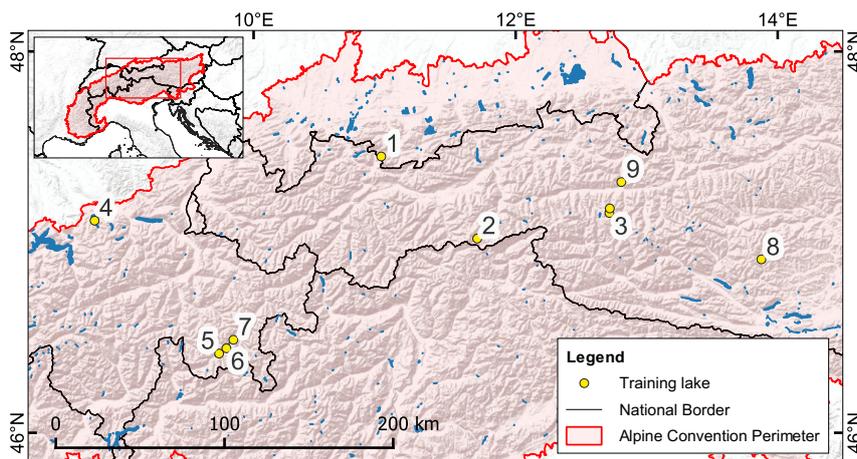
Due to their topographically embedded condition, many Alpine lakes experience topographic shading during low solar angles. This effect is amplified by the high latitude. During the winter season, this can lead to prolonged periods of shading that strongly reduce the remote sensing reflectance. The median winter coverage of shaded area reaches a maximum of 8.1% coverage for the study lakes (Figure 2.4).



**Figure 2.4:** Distribution plot of the simulated relative shaded lake area over time of the 512 study lakes. The hill-shade is simulated for the time of day during Sentinel-2A/B acquisitions in the Season 2019-2020, thus representing the expected remotely sensed shading.

## 2.1 Training lakes

The training lakes are selected lakes at locations with publicly available high-quality webcam imagery, suitable for the generation of training data. A total of 10 training lakes were chosen based on data-availability while trying to represent the highest possible range in variability of lake size, elevation and spatial distribution of the study area.



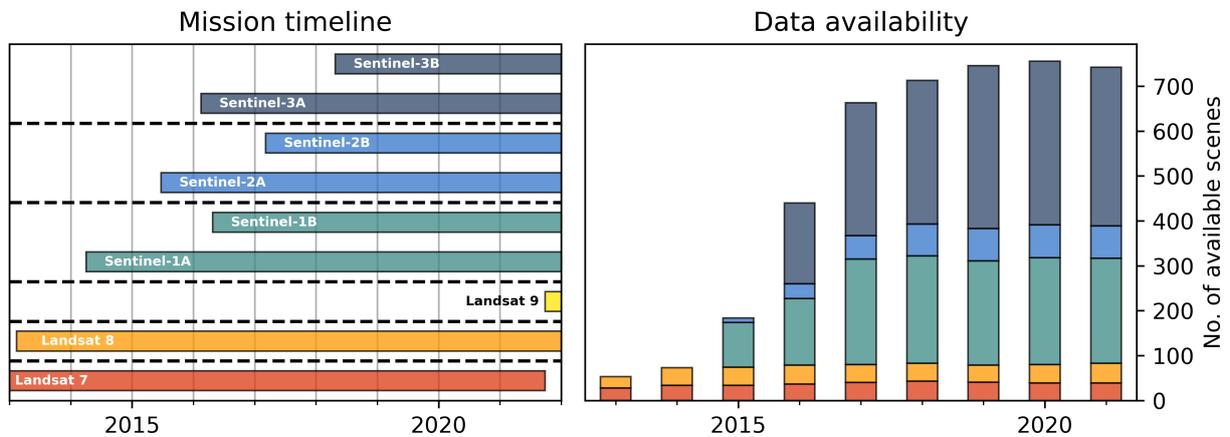
**Figure 2.5:** Overview map of the training lakes used in the generation of training data for the monitoring system. Lake IDs correspond to the lakes in Table 2.1. Note the two lakes Stausee Mooserboden and Stausee Wasserfallboden are captured on the same webcam and share the same lake ID.

**Table 2.1:** Table of the training lakes that serve as input lakes for model-training

Lake name	Lake ID	Elevation (m a.s.l.)	Area (km <sup>2</sup> )	Latitude	Longitude
Eibsee	1	973	1.7	47° 27' 24" N	10° 58' 22" E
Silsersee	5	1,797	4.1	46° 25' 20" N	9° 44' 16" E
Silvaplanersee	6	1,791	3.2	46° 27' 1" N	9° 47' 34" E
St. Moritzersee	7	1,768	0.78	46° 29' 39" N	9° 50' 42" E
Sihlsee	4	889	10.7	47° 8' 14" N	8° 46' 45" E
Schlegeisspeicher	2	1,780	2.2	47° 1' 59" N	11° 42' 17" E
Turrachsee	8	1,763	0.19	46° 55' 7" N	13° 52' 35" E
Stausee Mooserboden	3	2,038	1.6	47° 9' 32" N	12° 42' 46" E
Stausee Wasserfallboden	3	1,674	1.5	47° 11' 21" N	12° 43' 8" E
Zeller See	9	750	10.7	47° 19' 25" N	12° 48' 22" E

### 3 Data

Investments and efforts towards open data policies in Earth Observation have greatly increased in recent decades. As a result, the availability of open-access satellite imagery has experienced a strong growth. To overcome technical trade-offs between temporal and spatial resolution and achieve the highest possible temporal resolution for the monitoring of small lakes, we try to leverage this growing data source by combining imagery from multiple satellite missions. Namely, we combine the freely available data of the **Sentinel satellite suite** from the Copernicus program operated by the European Space Agency (ESA) and of the **Landsat satellite suite** operated by the National Aeronautics and Space Administration (NASA) and the United States Geological Survey (USGS). Both missions provide suitable imagery and supply a high combined temporally resolution with ensured future data availability.



**Figure 3.1:** Mission timeline (left) and data availability of Landsat 7/8 and Sentinel-1/2/3 imagery (right). The number of available scenes corresponds to the yearly scene-count for the year 2019 at Lake Sils, Switzerland.

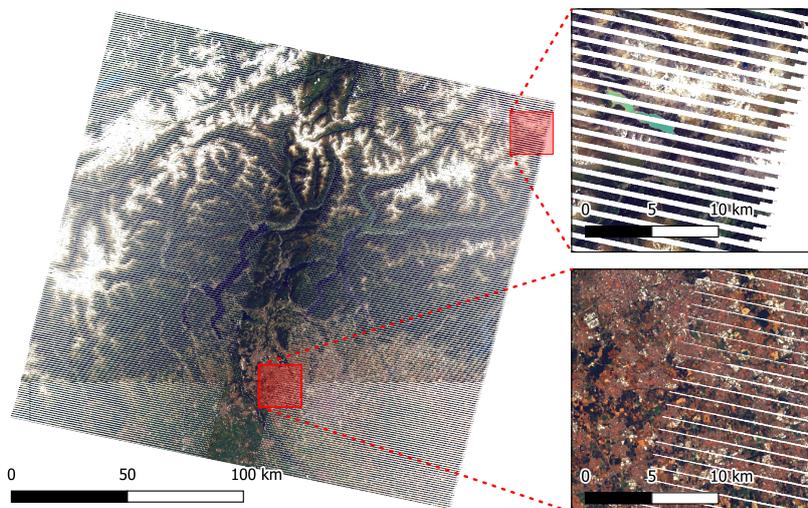
The sensors used in this study can be divided into two main categories. The Sentinel-1A/B C-SAR instrument uses active microwave remote sensing. Whereas the Landsat 7 ETM+, L8 OLI and TIRS, S-2A/B MSI and S-3A/B OLCI EFR are multi-spectral passive optical remote sensing sensors.

## 3.1 Optical data

### 3.1.1 Landsat 7/8

The **Landsat 7 (L7)** mission was launched on April, 1999 and carried the **Enhanced Thematic Mapper Plus (ETM+)** sensor. With a sun-synchronous, near-polar orbit and a repeat cycle of 16-days, L7 has delivered global imagery from its service start in April of 1999 until its recent decommission in April of 2022. The L7 ETM+ sensor imagery used in this thesis consists of data from four visible and near-infrared (NIR) bands and two short-wave infrared (SWIR) bands at a spatial resolution of 30 meters. Additionally, imagery from the thermal band with a 60 meter resolution was used (collection "LANDSAT/LE07/C02/T1\_L2" on GEE).

Most of the L7 imagery is provided with data gaps in a consistent zig-zag pattern (Figure 3.2) as a result of permanent instrument failure of the Scan Line Corrector (SLC) on May 31, 2003. Without the SLC the imagery cannot be compensated for the forward motion of the satellite, leading to duplicated image areas starting near the center of the scene with increasing width toward the edges. The duplicated areas are removed in a processing step and result in a loss of 22% of pixels. However, the incomplete scenes referenced as SLC-off still maintain the radiometric and geometric correction as before the instrument failure.

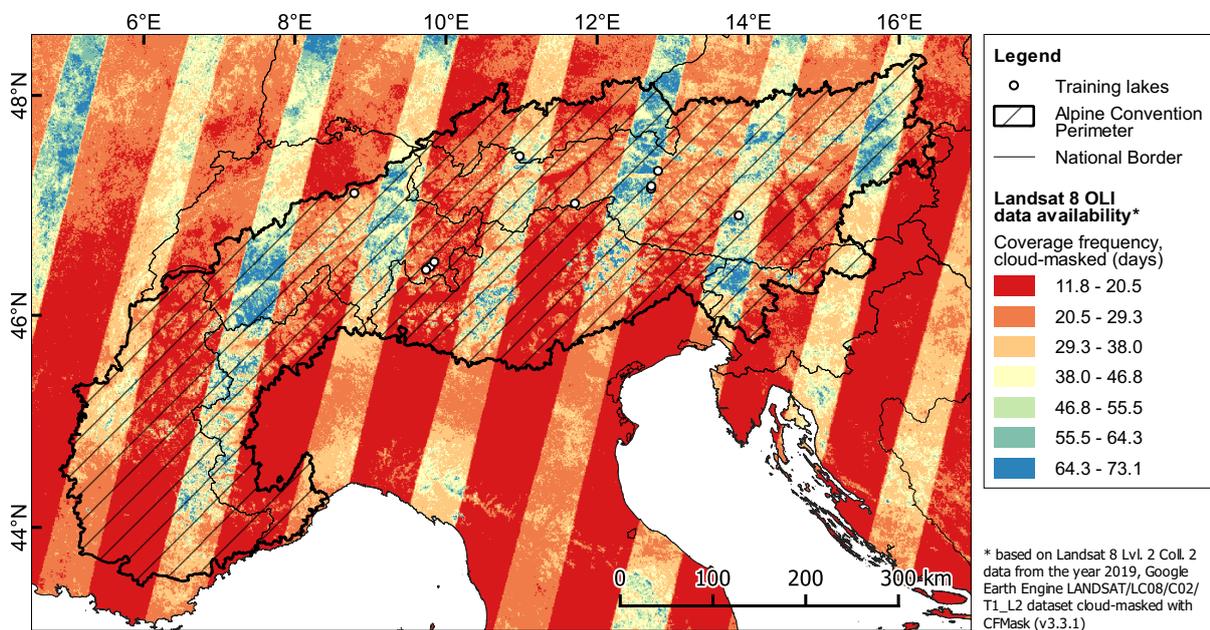


**Figure 3.2:** Landsat 7 ETM+ scene (LE07\_194028\_20170617) after the Scan Line Corrector (SLC) failure in May 31, 2003. The two close-ups demonstrate the gap near the unaffected center of the image corresponding to a single pixel (30m) in width (lower close-up), reaching 14 pixels in width (420m) at the image borders (upper close-up).

Its successor **Landsat 8 (L8)** was launched in February of 2013 with the goal of continued acquisition and availability of Landsat data and is still operational as of 2022. With identical orbit characteristics, L8 provides data with the same temporal, spatial and similar spectral resolution. The acquisition instrument was split up into two individual sensors. The **Operational Land Imager (OLI)** covers the visible, NIR and SWIR spectrum and the **Thermal Infrared Sensor (TIRS)** the thermal infrared spectrum. The L8 OLI sensor imagery used in this thesis consists of data from five visible and near-infrared (NIR) bands and two short-wave infrared (SWIR) bands at a spatial resolution of 30 meters. Additionally, thermal imagery from the TIRS-instrument with a 60 meter resolution was used (collection "LANDSAT/LC08/C02/T1\_L2" on GEE).

The available Landsat derived products have constantly evolved over the years and are available in different processing levels. For our study we used the reprocessed USGS **Collection 2 Level 2 (C2L2)** dataset accessible over GEE. These reprocessed Landsat science products make use of auxiliary data inputs and robust processing algorithms to derive atmospherically corrected **surface reflectance (SR)** and **land surface temperature (LST)** with consistency and high quality.

The data is provided as overlapping image tiles covering an approximate area of 170 km by 183 km. The high latitude of the study area leads to increased overlap between acquisitions and a consequently higher data availability than at equatorial orbit revisit period of 16 days. Considering the loss of usable data by cloud coverage, we can achieve a median coverage frequency for L7 and L8 of approximately 28.1 (see Appendix A) and 26.1 days (see Figure 3.3), respectively.



**Figure 3.3:** Map of the study area showing the spatial distribution of the coverage frequency of Landsat-8 imagery after cloud-masking in the year 2019.

### 3.1.2 Sentinel-2

The ESA Copernicus **Sentinel-2 (S-2)** mission was launched on June, 2015 with its first satellite the S-2A. It was later expanded with the launch of S-2B in March, 2017. Together the two units phased at 180 degrees build a two-satellite constellation with the same sun-synchronous, polar orbit and a combined revisit period of 5-days (equatorial). Both satellites carry the **Multi-Spectral Instrument (MSI)**, an optical multispectral sensor with a total of 13 spectral channels. The S-2 MSI imagery used in this thesis consists of data from 10 visible and near-infrared (NIR) bands and two short-wave infrared (SWIR) bands at a spatial resolution of 20 meters (collection "COPERNICUS/S2\_SR" on GEE). To get a common resolution of 20 meters, the two bands (B1 and B9) are downsampled from a resolution of 60 metres and the four bands (B2, B3, B4, B8) are upsampled from a resolution of 10 meters. The constellation's lifetime is planned to be extended with the

addition of S-2C/D in the years 2024 and 2025, respectively.

The S-2 derived products are available in different processing levels. For our study we used the reprocessed **Level-2A** products that are converted to Bottom Of Atmosphere (BOA) reflectance accessible over GEE. The atmospherically corrected Level-2A products are systematically generated over Europe since March 2018, and the production was extended to global in December 2018.

The data is provided as overlapping image tiles covering an approximate area of 100 km by 100 km. The high latitude of the study area leads to increased overlap between acquisitions and a consequently higher data availability than at the equatorial orbit revisit period of 5 days. Considering the loss of usable data by cloud coverage, we can achieve a median coverage frequency of approximately 5.1 days (see Appendix A.2).

### 3.1.3 Sentinel-3

The ESA Copernicus **Sentinel-3 (S-3)** mission was launched on June, 2015 with its first satellite the S-3A. It was later expanded with the launch of S-3B in March, 2017. Together the two units phased at 140 degrees build a two-satellite constellation with the same sun-synchronous, near-polar orbit and a combined revisit period of 1.1 days (equatorial). Both satellites carry the **Ocean and Land Color Instrument (OLCI)**, an optical multispectral sensor that has 21 spectral bands ranging from the visible to the near-infrared. The S-3 OLCI imagery used in this thesis consists of data from all 21 bands (collection "COPERNICUS/S3/OLCI" on GEE). The constellation's lifetime is planned to be extended with the addition of S-3C/D in the years 2024 and 2028, respectively.

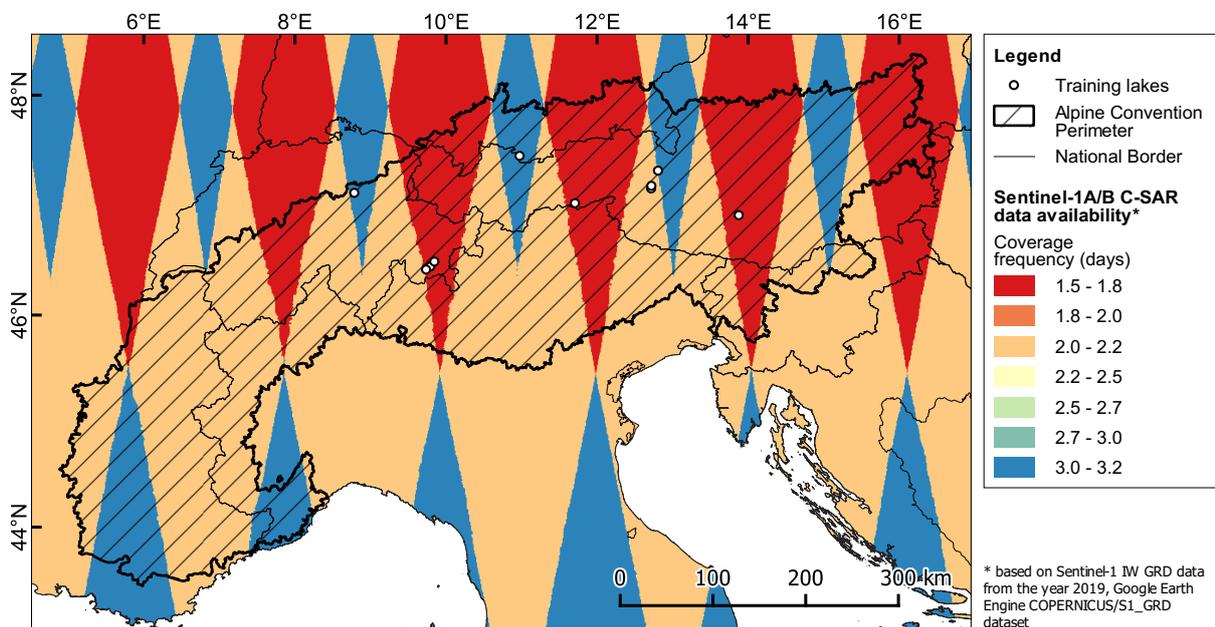
The S-3 OLCI derived products are available in different processing levels and resolutions. For our study we used the **Level-1B** Land Full Resolution (LFR) products that are provided in Top Of Atmosphere (TOA) reflectance at the full resolution of 300 metres accessible over GEE. As of the submission of this thesis none of the available S-3 products were available over GEE in atmospherically corrected Level-2 processing.

The data is provided as overlapping image tiles covering an approximate area of 1270 km by 1270 km. The high latitude of the study area leads to increased overlap between acquisitions and a consequently higher data availability than at equatorial orbit revisit period of 1.1 days. Considering the loss of usable data by cloud coverage, we can achieve a median coverage frequency of approximately 1.5 days (see Appendix A.3).

## 3.2 Synthetic aperture radar data

### 3.2.1 Sentinel-1

The ESA Copernicus **Sentinel-1 (S-1)** mission was launched on April, 2014 with its first satellite the S-1A. It was later expanded with the launch of S-1B in April, 2016. Together the two units phased at 180 degrees build a two-satellite constellation with the same sun-synchronous, near-polar orbit and a combined revisit period of 6 days (equatorial). Both satellites carry the **C-band synthetic-aperture radar (C-SAR)** instrument. The C-SAR instrument actively measures microwave back-scatter at 5.405 GHz (wavelength of 5.5 cm) and supports operation in dual polarisation (HH+HV, VV+VH) modes. The constellation's lifetime is planned to be extended with the addition of S-3C and S-3D in the year 2023.



**Figure 3.4:** Map of the study area showing the spatial distribution of the coverage frequency of Sentinel-1A/B imagery in the year 2019.

The S-1 C-SAR derived products are available in different processing levels. For our study we used the **Ground Range Detected (GRD)** products that are provided in calibrated, ortho-corrected back-scatter. Only scenes available in the **Interferometric Wide (IW)** swath mode and **VV/VH-polarized** back-scatter bands in raw power values were used (collection "COPERNICUS/S1\_GRD\_FLOAT" on GEE). The dataset is provided in a pixel-spacing of 10 metres. Since the radiometric resolution is limited to 20 metres, a down-sampling to 20 meters is applied.

The data is provided as overlapping scenes covering an approximate area of 170 km by 250 km. The high latitude of the study area leads to increased overlap between acquisitions and a consequently higher data availability than at the equatorial orbit revisit period of 6 days. Due to the active SAR technique, cloud coverage can be penetrated and does not impact the coverage fre-

quency. Thus we can achieve a median coverage frequency of approximately 3.1 days (Figure 3.4).

### 3.2.2 ALOS AW3D30

For the processing of the SAR imagery the Japan Aerospace Exploration Agency (JAXA) Advanced Land Observing Satellite (ALOS) Global Digital Surface Model (AW3D30) was used. The AW3D30 is a global digital surface model (DSM) dataset with a resolution of approximately 30 meters (1 arcsecond). The elevation of the DSM is globally calculated with stereo mapping of optical images acquired by the ALOS/Daichi mission operated by JAXA from 2006 to 2011.

## 3.3 Lake data

### 3.3.1 HydroLAKES

The **HydroLAKES v1.0** dataset (Messenger et al., 2016) is a global lake database with shoreline polygons of all global lakes with a surface area of 10 ha (0.1 km<sup>2</sup>) and higher. The dataset has been coalesced from multiple auxiliary data sources of lake polygons. In addition to the shoreline polygon, a number of geometrically derived and externally sourced attributes are available.

**Table 3.1:** List of available HydroLAKES v1.0 attributes used in this study

Attribute	Unit	Description
Lake type	[-]	Indicator for the type of lake
Lake area	km <sup>2</sup>	Lake surface area
Shore length	km	Length of shoreline
Shore deviation	[-]	Measure for complexity of shoreline
Total volume	mcm	Estimated or reported total lake reservoir volume
Avg. depth	m	Estimated lake depth derived from area and volume
Avg. discharge	m <sup>3</sup> s <sup>-1</sup>	Avg. discharge derived from modeled estimates
Avg. residence time	d	Avg. residence time derived from volume and discharge
Elevation	m a.s.l.	Elevation of lake surface

### 3.3.2 JRC Global Surface Water Mapping Layers

The **JRC Global Surface Water Mapping Layers v1.3** dataset (Millard and Richardson, 2015) contains the spatial and temporal distribution of surface water from 1984 to 2020. The existence of surface water was extracted and mapped from over 4.4 million scenes from Landsat 5, 7, and 8. Each pixel was individually classified into water/non-water using an expert system and the results were combined into a monthly history for the entire time period with a resolution of 30 m. Areas where water has never been detected are masked. For this study we make use of the yearly aggregated **JRC Yearly Water Classification History v1.3**. This dataset has been derived from the monthly history and contains a year-by-year classification into permanent and seasonal surface water.

### 3.3.3 Webcam imagery

High-resolution webcam imagery was obtained from **foto-webcam.eu** for eight of the study lakes. A total of **10'524 daily images** were acquired at 12:00 PM local time. The webcam network set in place by foto-webcam.eu utilizes repurposed single-lens reflex cameras installed in a custom enclosure instead of relying on commercially available standard webcam products. This allows them to provide high-quality and high-resolution imagery in 10 min time intervals. Access to their webcam network and image archive, as well as detailed documentation on the camera system, are freely available on their website.

Additional webcam imagery for two adjacent study lakes (Silsensee and Silvaplannersee) was sampled from one Roundshot system by **Seitz Phototechnik AG**. Their commercial Roundshot camera system produces full panoramic images which are archived in 10 min intervals and generally freely available.

### 3.3.4 Auxiliary training data

For our study we reused lake ice coarse classification results obtained by (Tom et al., 2019) for four of our study lakes. This training data consists of daily labels that give coarse estimates of the lake ice coverage for the entire lake. The training data was produced by manual visual classification of webcam imagery.

## 3.4 Climate data

### 3.4.1 ECMWF ERA5-Land

The European Centre for Medium-Range Weather Forecasts (ECMF) **ERA5-Land** dataset provides a collection of 50 global climatic land variables. This climate reanalysis dataset is provided in the form of gridded data in hourly intervals with a resolution of approximately 9 km. For our study we made use of the variables **2m air temperature** and **eastwards and northwards 10m wind component**.

### 3.4.2 MeteoSwiss Grid-Data Products

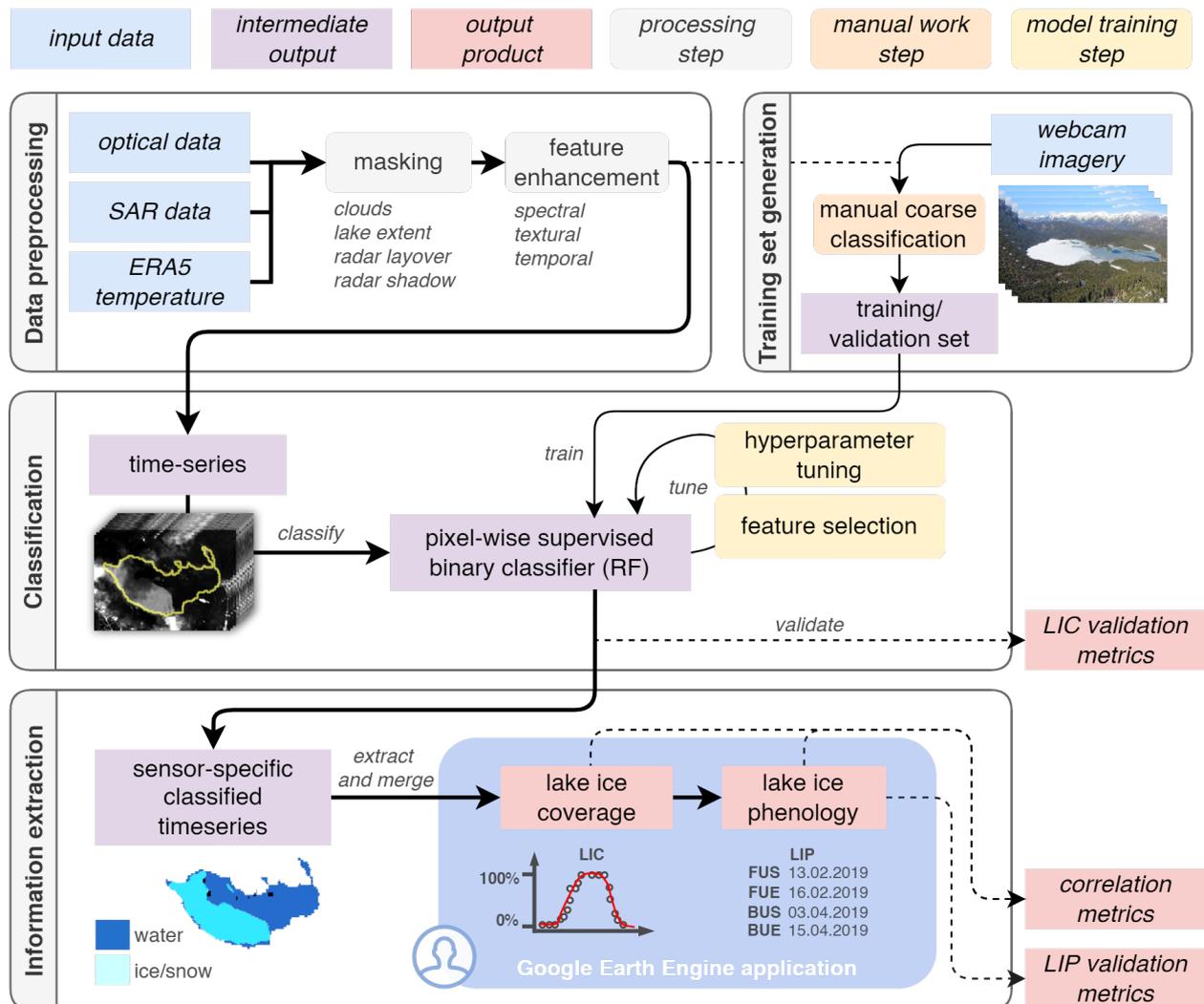
The **MeteoSwiss grid-data products TabsD, TminD, TmaxD** (MeteoSwiss, 2021) are comprised of daily mean, minimum and maximum air temperature at 2 m height above ground level. The three temperature datasets are a product of 90 homogenized long-term in-situ measurement series based on the operational station network SwissMetNet. The measurements are interpolated based on a deterministic analysis method to describe the km-scale distribution of day-to-day temperature variations in Switzerland during the past decades, dating back as far as 1961. Additionally, we used the products **RhiresD** and **SrelD**. These two products give daily mean precipitation and relative sunshine duration measurements.

**Table 3.2:** List of all MeteoSwiss Grid-Data Products used in this thesis

<b>Product</b>	<b>Description</b>
RhiresD	Daily precipitation in mm/day
TabsD	Daily mean air temperature in °C
TminD	Daily minimum air temperature in °C
TmaxD	Daily maximum air temperature in °C
SrelD	Daily relative sunshine duration in %

## 4 Methods

In this chapter we outline our methods used to extract lake ice coverage and phenology metrics from remotely sensed imagery. Figure 4.1 gives a simplified overview of the processing steps involved from data acquisition up to phenology extraction. The entire processing workflow is implemented within GEE, allowing efficient cloud-based execution. The two main output products consist of LIC and LIP metrics, which can be accessed and explored by untrained users over a GEE application.



**Figure 4.1:** Simplified flowchart of the steps applied for lake ice monitoring in this study

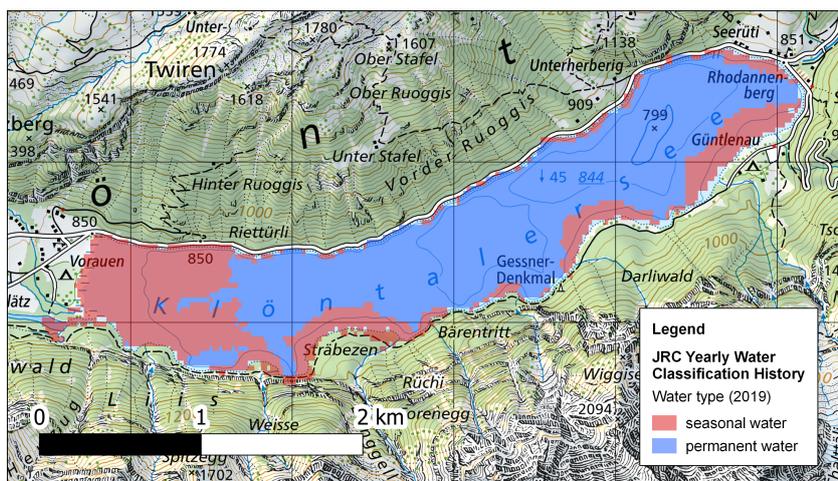
The workflow for the monitoring system is structured into four main parts: **Data preprocessing, training set generation, classification and information extraction**. With the exception of training data generation, the entire workflow is automatized and the temporal coverage could be easily extended if necessary. Furthermore, for the analysis of the outputs, validation metrics are collected in a validation step. In addition, a correlation analysis gives insight into the relationships between lake ice and climatic, as well as lake-specific factors.

## 4.1 Data preprocessing

In the data preprocessing step, the five input datasets are individually prepared for classification and training set generation. Firstly, this includes the step of masking the imagery for lake extents and unusable pixels. Pixels are classified as unusable due to either cloud cover in optical imagery or radar shadow and layover effects in SAR imagery. Secondly, in the feature enhancement step we generate and add additional spectral, textural, temporal and temperature image layers to the datasets. The goal of this step is to extend the dimensionality of the feature spaces with meaningful information content to improve classification performance. The intermediate output of the data preprocessing are five individual, masked and enhanced raster collections for each sensor.

### 4.1.1 Lake masking

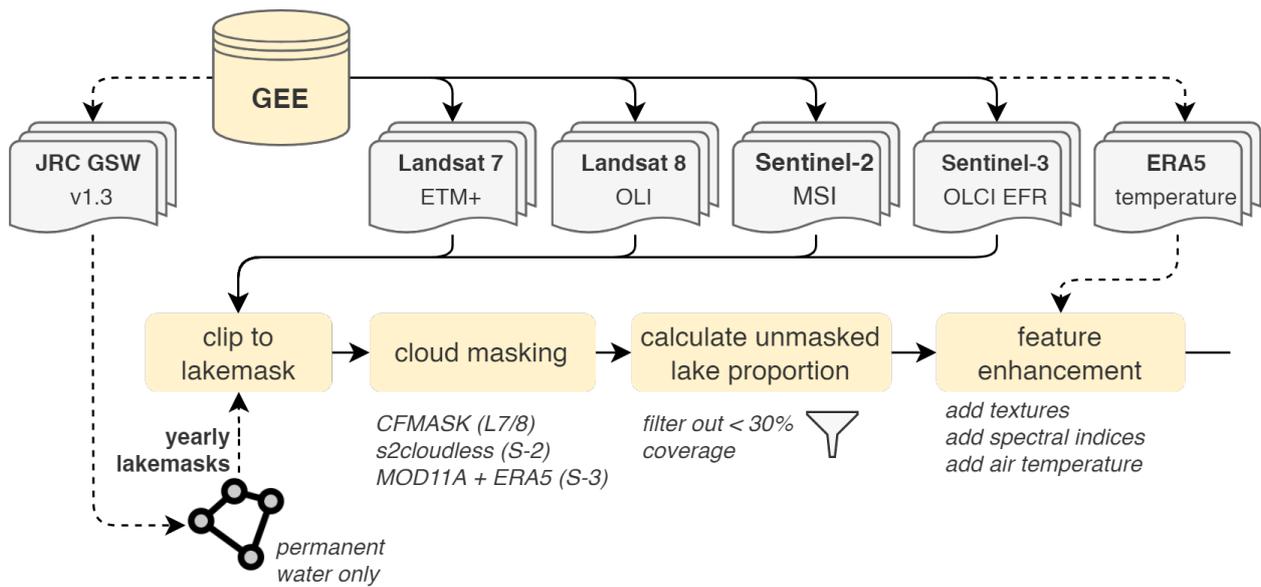
As showed in Figure 2.3, many of the alpine lakes experience seasonal fluctuations in lake extent. Because we pursue a binary classification approach that only differentiates between water and ice/snow, a static lake mask could potentially introduce land pixels in the classification process and negatively impact the classification performance. To discard land pixels we use the JRC Yearly Water Classification History v1.3 to create yearly lakemasks consisting of surfaces that contain only permanent water (see Figure 4.2).



**Figure 4.2:** Cropped extent of the JRC Yearly Water History v1.3 layer for the year 2019. The map shows the classification in seasonal and permanent water pixels derived from RS imagery for Klöntalersee, a natural lake within the study area in Glarus, Switzerland. In the year 2019 over one third of the lake’s surface was classified as seasonal water. Basemap provided by the Federal Office of Topography swisstopo.

## 4.1.2 Optical data

Due to the differences in RS methods used, preprocessing steps for optical and SAR imagery are different. Figure 4.3 gives a detailed view of the preprocessing steps we applied to the optical imagery. In a first step the full optical scenes are clipped to the area of interest using the yearly lakemasks. In a subsequent step cloudy pixels are further masked out from the images. We then apply a coverage filter to the resulting images and discard images with less than 30% of usable pixels. In a final feature enhancement step we combine the optical input imagery with layers consisting of textures, spectral indices and ERA5 2m air temperature. The individual sensor-specific raster collections are then ready for classification and training set generation.



**Figure 4.3:** Detailed flowchart of the steps applied for the preprocessing of optical imagery

### 4.1.2.1 Cloud masking

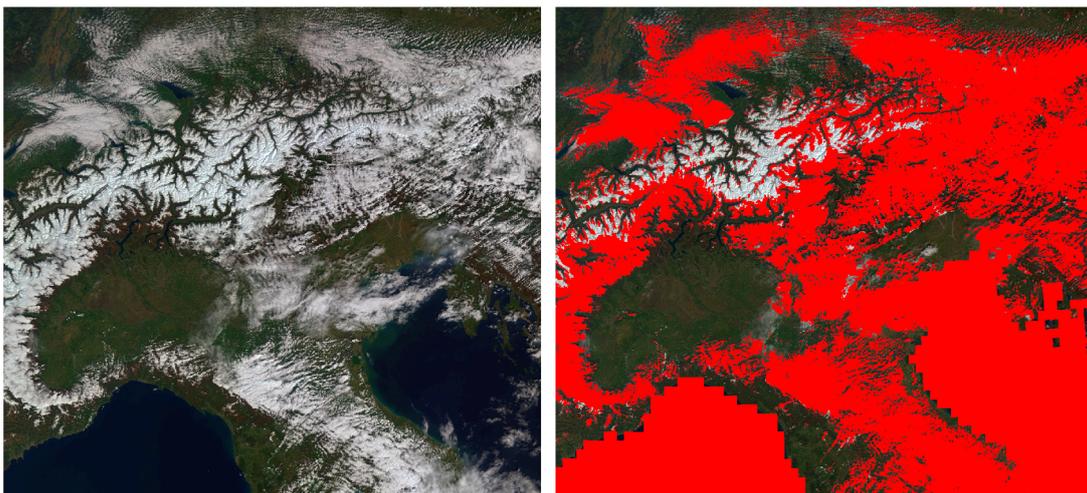
**Cloud masking** (also called cloud filtering) is a crucial step in the processing of optical RS imagery. For our purpose clouds are obstacles that partially or fully reduce visibility on the area of interest and impart atmospheric interference that affects the ability to execute a clean classification. Furthermore, shadows cast by the clouds can also affect the classification results. There are a variety of general and sensor-specific approaches and products to carry out cloud masking. In the case of L7/8 and S-2 we apply the most proven methods.

For L7/8 imagery we used the pixel flags provided by the C Function of Mask (CFMask) algorithm (Foga et al., 2017) developed specifically for Landsat imagery. CFMask is a multi-pass algorithm based on decision trees that labels pixels in different classes according to static or scene-based dynamic thresholds. In addition, CFMASK also extracts cloud shadows by iteratively estimating the height of detected clouds and projecting their shadows. The CFMASK quality flag is included in the GEE Landsat Level 2 datasets. If either of the "Cloud" or "Cloud shadow" flags were set, we discarded the pixel.

For **S-2** imagery we used the cloud probability metrics provided by *s2cloudless* (Skakun et al., 2022). *s2cloudless* is a machine learning based classifier developed specifically for S-2 imagery based on the LightGBM gradient boosting classifier. The cloud probability output of *s2cloudless*, that gives the probability that a pixel is cloudy, is available as a standalone dataset over GEE. If this cloud probability indicator was above 20% we discarded the pixel.

For the cloud masking of **S-3** imagery pixel flags are available as part of the official S-3 Synergy (S3-SYN) product. However, as of the submission of this thesis no satisfactory approach for cloud masking of the S-3 OLCI dataset was available over GEE. By using the "bright" pixel flag included in the dataset, it is possible to remove clouds, yet this approach will also remove snow/ice pixels in many cases which is unacceptable for the purpose of this study. Therefore, we developed a custom cloud masking method that makes use of the cloud detection algorithm of the Terra Moderate Resolution Imaging Spectroradiometer (MODIS) sensor.

To cloud mask the S-3 OLCI scenes the cloud mask from the most timely Terra MODIS MOD11A1 v061 scene available over GEE is acquired. Using the time delay between the two scene acquisitions (based on the "day\_view\_time" band for MOD11A1) and the ERA5 hourly 10m u- and v-components of wind, the mask is further iteratively moved across the S-3 scene to create a buffer along the possible drift movement of the clouds. If a pixel was within the resulting buffered MOD11A1 cloud mask and the S-3 "bright" pixel flag was set, the pixel was discarded. From a set of non-exhaustive visual inspections, we found that this approach produced acceptable results (see Figure 4.4). The resulting time delays between the two sensors ranged from 46 to 122 minutes (5th and 95th percentile) for all applied masks.



**Figure 4.4:** Example of the custom S-3 OLCI cloud mask by combining the Terra MOD11A1 cloud-flag, S-3 bright pixel flag and ERA5 10m wind components for buffering. The images show the S-3 scene (S3A\_20190415T093540\_20190415T093840) before (left) and after (right) applying the custom cloud-filtering method. Since the Terra MODIS MOD11A product is provided for land use only, ocean pixels are masked out as well.

#### 4.1.2.2 Spectral indices

For our optical input datasets we calculated a total of **six spectral indices** listed in Table 4.1. Spectral indices are mathematical equations that combine the information of image bands from two or more wavelengths in an index value. The most popular type are Normalized Difference Indices (NDI) with the form:

$$NDI = \frac{(band_x - band_y)}{(band_x + band_y)} \quad (4.1)$$

This simple type of index represents the difference between two bands normalized by their sum. Normalized difference indices are very useful because they reduce the data dimensionality and enhance spectral features. Because they are ratio-based they can also minimize effects of illumination (topographic shadows and cloud shadows). It has been shown that the use of spectral indices to expand the feature space for classification problems can lower classification errors (Ezimand et al., 2018).

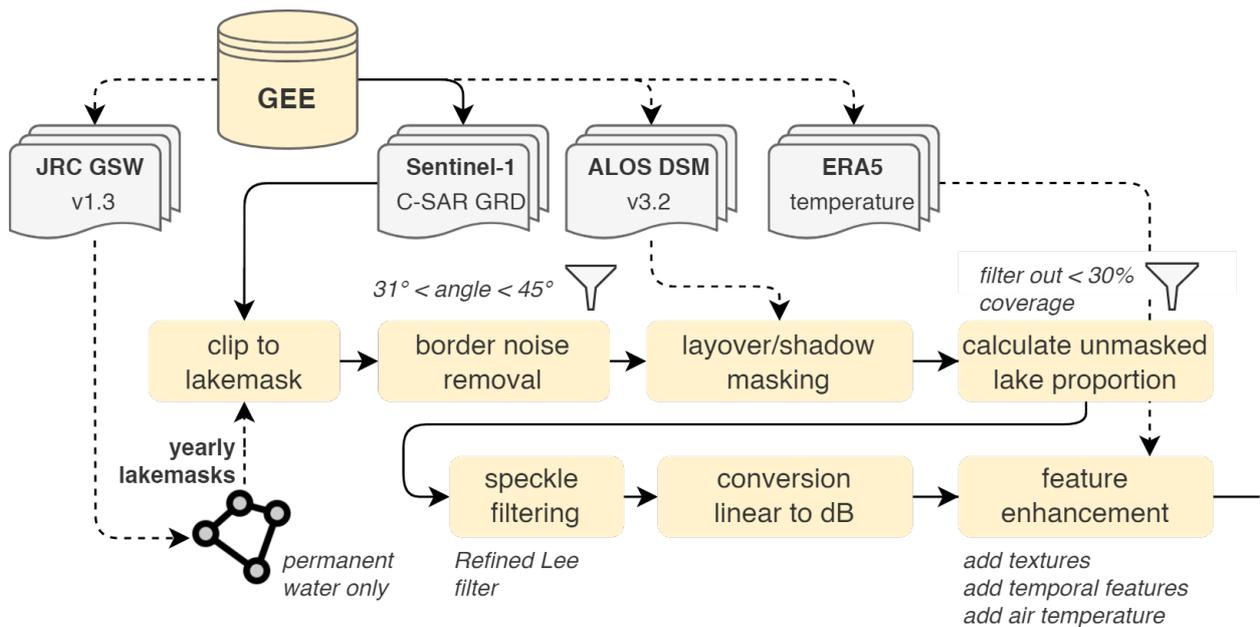
**Table 4.1:** Table of the 6 spectral indices used to enhance the classification. The rho-symbol ( $\rho$ ) denotes the reflectance in the specified spectral range. \* Fractional Snow Cover is only defined and applied for S-2 imagery.

Spectral index	Abbrev.	Equation	Reference
Normalized Difference Vegetation Index	NDVI	$\frac{\rho_{NIR} - \rho_{Red}}{\rho_{NIR} + \rho_{Red}}$	Rouse et al. (1973)
Normalized Difference Snow Index	NDSI	$\frac{\rho_{Green} - \rho_{SWIR}}{\rho_{Green} + \rho_{SWIR}}$	Dozier (1989)
Normalized Difference Water Index	NDWI	$\frac{\rho_{Green} - \rho_{NIR}}{\rho_{Green} + \rho_{NIR}}$	McFeeters (1996)
Modified Normalized Difference Water Index	mNDWI	$\frac{\rho_{Green} - \rho_{MIR}}{\rho_{Green} + \rho_{MIR}}$	Xu (2006)
Bare Soil Index	BSI	$\frac{(\rho_{SWIR} + \rho_{Red}) - (\rho_{NIR} - \rho_{Blue})}{(\rho_{SWIR} + \rho_{Red}) + (\rho_{NIR} + \rho_{Blue})}$	Rikimaru et al. (2002)
Fractional Snow Cover *	FSC	$0.5 \tanh(2.65NDSI - 1.42) + 0.5$	Gascoin et al. (2020)

### 4.1.3 Synthetic aperture radar data

The acquired S-1 C-SAR dataset from GEE consists of Level-1 Ground Range Detected (GRD) scenes processed to backscatter coefficient  $\sigma_0$ . This means that a number of important processing steps are already applied to the data. Namely the GEE processing consists of: Updating orbit metadata with restituted orbits, border noise removal, thermal noise removal, radiometric calibration (conversion to  $\sigma_0$ ) and terrain correction using the NASA SRTM 30m DEM.

Figure 4.5 gives a detailed view of the preprocessing steps we applied to the SAR imagery. In a first step the full SAR scenes are clipped to the area of interest using the yearly lakemasks. To get the S-1 dataset to an analysis-ready state, we applied some of the additional steps proposed by Mullissa et al. (2021). The additional border noise correction is applied by masking out incidence angles lower than 31.0 and higher than 45.0 degrees from the sensor's range of 29.1 to 46.0 degrees. Due to the steepness of the terrain at many of the alpine lakes there is a possibility of radar layover and shadow effects. The signal returned over these areas is not representative of the surface. To mitigate these effects, the ALOS DSM is used in combination with the masking algorithm by Mullissa et al. (2021) to create a radar layover and shadow mask. The proposed radiometric terrain flattening (conversion to  $\gamma_0$ ) is not applied. For the sole purpose of studying lakes a radiometric terrain flattening is not necessary due to the flat nature of the studied surfaces. We then apply a coverage filter to the resulting images and discard images with less than 30% of usable pixels. The mono-temporal Refined Lee filter is further applied to suppress speckle noise. Then we convert the backscatter signal from raw power values to dB. In a final feature enhancement step we combine the SAR input imagery with layers consisting of textures, temporal features and ERA5 2m air temperature. The S-1 raster collection is then ready for classification and training set generation.



**Figure 4.5:** Detailed flowchart of the steps applied for the preprocessing of SAR imagery

#### 4.1.4 Temporal features

To incorporate the **temporal context** into the feature space we used a total of **14 temporal features**. Without the addition of these features the classifier would have no temporal input data and could only perform classification based on daily measurements. The feature layers are constructed by taking the difference between small weekly and two-weekly temporal windows against a large yearly window.

We generated all of the 14 temporal features shown in Table 4.2 for the SAR input dataset. Since cloud interference reduces the temporal resolution of optical imagery and can create large gaps in the datasets, we only computed the five two weekly features for the optical datasets.

**Table 4.2:** Table with description of the temporal features used to enhance the classification

Feature suffix	Description
_w_min	weekly aggregated minimum
_w_max	weekly aggregated maximum
_w_median	weekly aggregated median
_w_mean	weekly aggregated mean
_w_pX	difference between weekly aggregated mean and yearly X-th percentile with X = [10, 30, 50, 70, 90]
_tw_pX	difference between two-weekly aggregated mean and yearly X-th percentile with X=[10, 30, 50, 70, 90]

#### 4.1.5 Textures

To incorporate the **spatial context** into the feature space we used a total of **18 image textures**. Image textures describe the distribution of gray levels within a defined image area. The two most common ways to compute image textures are from first-order or more complex second-order measures. Textures from first-order measures (e.g. mean, standard deviation) are statistical metrics that do not consider the pixel-neighbor relationships. They can also be simply derived from the histogram of an image area. Textures from second-order measures (e.g. co-occurrence matrix and variogram) take the gray levels of neighbouring pixels into account. A common approach is to use a grey level co-occurrence matrix (GLCM) in a moving window. This matrix is computed by counting the number of times a pixel of value X lies next to a pixel of value Y, in a particular direction and distance within a local image area. Then statistical measures are extracted from the GLCM and allocated to the center-pixel. First-order texture metrics are generally less effective and the use of the GLCM textures is a commonly used approach in the remote sensing community to increase model performance in classification tasks (Berberoğlu et al., 2010; Rodriguez-Galiano et al., 2012).

We generated all of the 18 textures listed in Table 4.3 and implemented within the GEE platform. This includes 13 GLCM textures proposed by Haralick et al. (1973), four GLCM textures proposed by Conners et al. (1984) and the first-order windowed entropy texture. Since image textures can be computed at different levels of resolution, the texture bands were generated at three different windows sizes of 3x3, 5x5 and 7x7 pixel neighbourhoods.

**Table 4.3:** Table of the 18 textural features used to enhance the classification

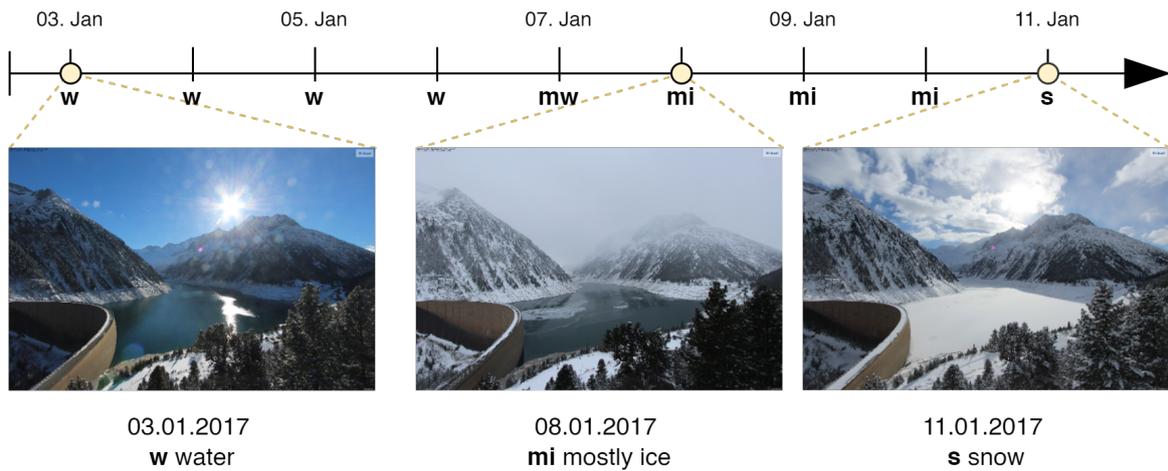
Texture	
Angular Second Moment (GLCM)	Sum Average (GLCM)
Contrast (GLCM)	Sum Variance (GLCM)
Correlation (GLCM)	Sum Entropy (GLCM)
Variance (GLCM)	Difference variance (GLCM)
Inverse Difference Moment (GLCM)	Information Measure of Corr. 1 (GLCM)
Difference entropy (GLCM)	Information Measure of Corr. 2 (GLCM)
Dissimilarity (GLCM)	Cluster prominence (GLCM)
Inertia (GLCM)	Cluster Shade (GLCM)
Entropy (GLCM)	Windowed Entropy

## 4.2 Training set generation

In the **training set generation** step, the five prepared sensor-specific datasets are used to create training sets for the each sensor’s classifier. All of the image features from pixels with a known class were extracted and annotated with a corresponding class-label. For this step pixels have to be referenced in time and space to a corresponding ground-truth label. To generate these ground-truth labels we used the acquired daily webcam imagery of our training lakes and the manual coarse-classification labeling approach proposed by Tom et al. (2019). In the manual coarse-classification step a total of 10’524 daily high-resolution webcam images were inspected by one operator and labeled according to the classes shown in Table 4.4. An example of the applied annotation can be seen in Figure 4.6.

**Table 4.4:** Manual coarse classification classes as proposed by Tom et al. (2019)

Class	Description
s	snow, when snow is on lake ice, lake frozen to ca. 90-100%
i	ice, frozen lake to ca. 90-100%
w	water, lake covered by water to ca. 90-100%
ms	more snow, ca. 60-90%, but a small part water
mi	more ice, ca. 60-90%, but a small part water
mw	more water, ca. 60-90%, but a small part frozen
c	clouds or fog covering all lake
u	unclear, when you cannot judge the lake state
n	no webcam data available



**Figure 4.6:** Example of the manual coarse classification scheme from Tom et al. (2019) applied to the webcam imagery acquired for lake Schlegeisspeicher. The time bar on top visualizes the applied labeling from the 3rd to the 11th January of 2017. The images below show the corresponding webcam images at three selected days with different classes.

#### 4.2.1 Manual approach

In the traditional approach for training set generation from optical RS imagery a single or multiple operators draw polygons and annotate them with class labels. All pixels within the polygon can then be extracted and added to the feature space for model training. This approach can yield high quality training sets with very low noise from incorrect labeling. However, the main drawbacks of this method are its inefficiency and the need for high-resolution imagery. Annotating large datasets is a very time-consuming process. From visual inspection, we found that the low-resolution imagery from S-3 is too coarse for manual labeling. In addition, the interpretation and labeling of SAR imagery, that does not show the feature characteristics humans are used to from the visual light spectrum, can be very difficult. We found that a precise annotation of lake ice signals from our collected SAR imagery was impossible for a human operator. Mainly due to a low signal-to-noise ratio and unclear boundaries compared to optical imagery (see comparison between Figure 1.1 and Figure 1.2).

As a result, we opted for an automated approach and used the traditional manual approach to create only one training set for S-2 imagery. S-2 imagery from the covered time period was manually inspected by one operator and labeled into a total of six classes listed in Table 4.5. A total of 1'193 polygons were created and labeled. In the labeling process shadowed surfaces and dark (optically thin) ice were distinguished. In case of unclear signals, the webcam imagery and the coarse classification labels were used to support the annotation process. If a surface could not be labeled with clear boundaries and high confidence it was left out. This training set serves as a benchmark and can be used to verify the performance of our automated approach. Furthermore, we can also use the spectral information from the fine class structure to study the effects of shadow and ice thickness.

**Table 4.5:** Class labels used for the training set generation in the manual approach

Class	Class description
1	ice
2	ice with shadow
3	thin ice (dark ice)
4	snow
5	snow with shadow
6	water

### 4.2.2 Automated approach

In the **automated approach** the daily labels from the coarse classification that classify the state of the lake's ice coverage based on webcam imagery are processed to extract training pixels. Days with full water or full ice/snow coverage were used to create training polygons covering the entire unmasked lake extent. Days with partial ice coverage were discarded in this approach because they contain the signal of multiple classes. This approach can be executed unsupervised and allows us to automatically label both optical and SAR imagery. Training sets for all five input datasets were generated using the automated approach.

## 4.3 Classification

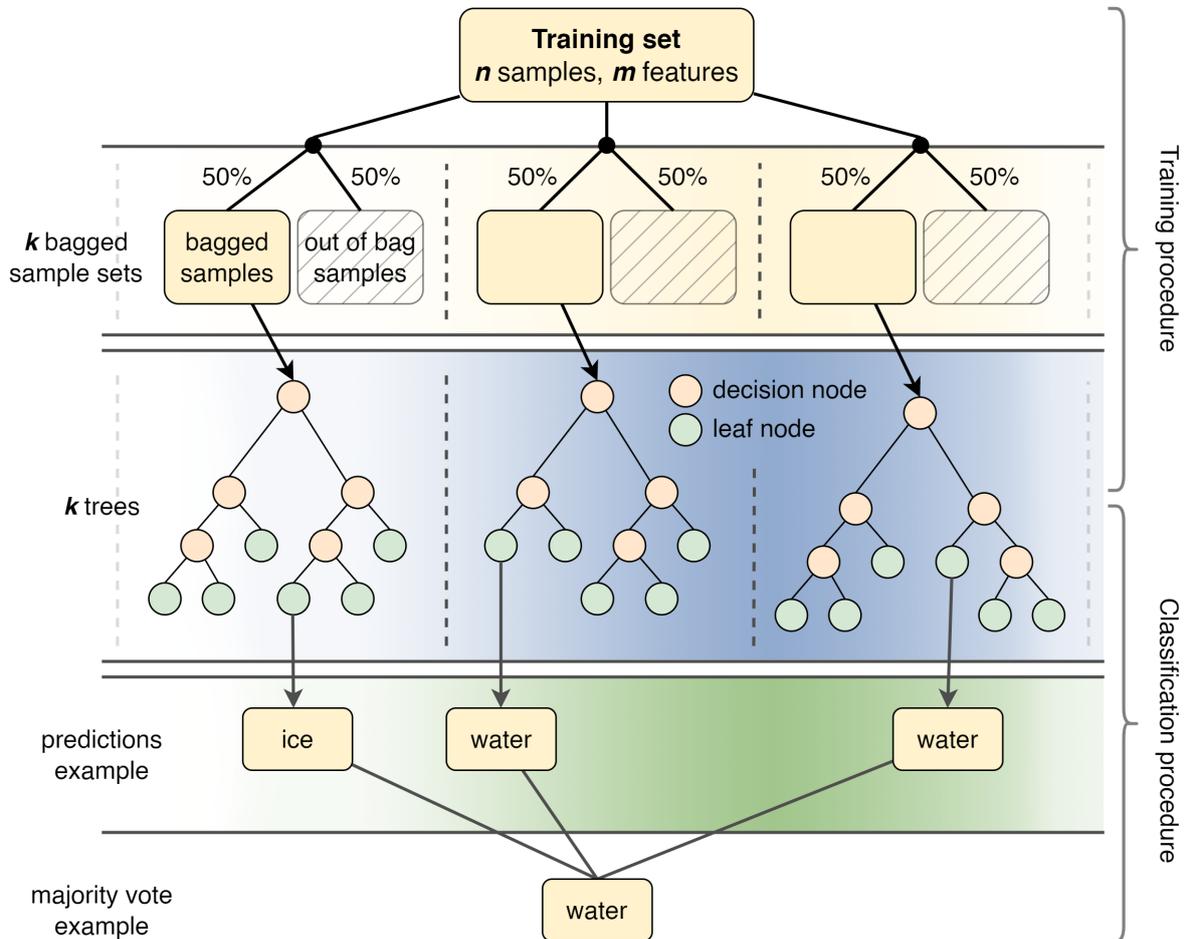
In the **classification** step the five previously generated training sets were used to train a sensor-specific supervised pixel-wise binary classifier model. Before generating the final classifiers that was used for the lake ice monitoring system, a feature selection and subsequent hyper-parameter tuning step were executed. Once the fully trained and tuned models were created they were used to classify input imagery into water and ice/snow pixels.

### 4.3.1 Random forest classifier

The supervised classification model used in this study is a **Random Forest** (RF) classifier based on the **Statistical Machine Intelligence and Learning Engine** (Smile) implemented within GEE. The RF model is a machine learning (ML) algorithm that is used widely in classification and regression problems and was first introduced by Breiman (2001).

In the training sequence of an RF model an **ensemble** in the form of a multitude of individual decision tree models is generated. The individual decisions trees are regarded as weak learners that can classify a set of input features. Each decision tree is trained using the **bagging** method (also called bootstrap aggregation), the selection of a random subset of the training set. In each individual tree, this tree-specific random training set is used to grow a decision tree with consecutive **decision nodes** that each split into two new nodes based on a threshold condition. Ultimately, each path through this tree will end in a terminal **leaf node** that defines the output of the decision tree. Decision nodes and the resulting splits are created with **random selection of features**, using only a random subset of all available input features at each node. By combining the class output

from all independent trees using the **majority vote**, the class output with the highest number of votes, the output of the ensemble is determined.



**Figure 4.7:** Simplified schematic view of a random forest model with a bagged fraction of 0.5

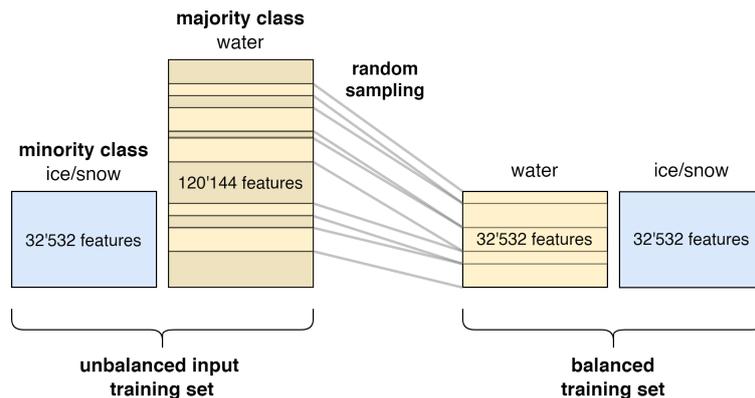
The combination of the bagging method and the random selection of features at decision nodes allows to grow an ensemble of diverse and uncorrelated decision trees. Even though the individual trees are regarded as weak learners or weak classifiers, the RF ensemble outperforms the individual trees. This will generally create a robust model that can process large datasets with a high number of input features and a low risk of overfitting. Another key benefit is that the individual decision trees are trained independently from each other and can also classify inputs independently. This allows to execute the training and classification procedure of an RF model fully parallelized and thus in an efficient manner.

### 4.3.2 Training procedure

### 4.3.3 Data balancing

Most classification models, including RF classifiers, rely on training sets with evenly distributed classes. If this is not the case, classification models are likely to be biased towards the classes with higher occurrence. Since many real-world classification problems have an **imbalanced class distribution**, multiple methods have been established to balance training sets prior to model training. The same imbalance can be found in the case of lake ice detection. The occurrence of lake ice in the Alps is limited to the winter months which results in training sets that over-represent the water class and under-represent the ice/snow class.

One of the simplest methods to deal with this imbalance is to use **under-sampling**. We can balance the training set by randomly discarding samples of the water class until the class abundances between water and ice/snow are equalized (see Figure 4.8). Since the imbalances occurring in our datasets are not too severe and we are able to collect large training sets using the automated approach, we can make use of this method to trade in training samples to reach a balanced set.



**Figure 4.8:** Simplified visualization of the data balancing. The majority class is randomly under-sampled to reach the same feature count as the minority class and balance the class abundances.

### 4.3.4 Feature selection and feature importance

The additional temporal, textural and spectral features added to the training sets in the feature enhancement step strongly increase the dimensionality and size of the training sets. In the case of the S-2 input imagery, we would reach a total of 835 input features if we included all additional features from the feature enhancement step. Even though RF models can generally handle large datasets with high dimensionality due to the bagging and random feature selection methods that are applied, the speed of the training procedure would be strongly reduced. In addition, memory limitations set within the GEE framework can be easily reached with training sets of large sizes. This problem makes a **feature selection** step necessary, which removes features with low information content before training and tuning of the final models. In this step prior to the hyperparameter tuning, we generate RF models for all four optical datasets with **default parameter settings** and using only the original bands. This allows us to find the two most important bands for each optical sensor. The feature enhancement is then limited to these bands and the presented

temporal and textural features are only generated for these two selected bands. This step is omitted for the S-1 imagery because the input imagery consists of only two image bands (VV and VH polarization).

To determine the **feature importance** of the input features and control the feature selection process we make use of the **Gini-importance** measure. The Smile RF model used in this study is generating the optimal splits at its decision nodes based on the the **Gini-impurity**. This measurement is defined as the probability of misclassification of a new random observation, ranging from 0 to 0.5. The lower the Gini-impurity, the lower the likelihood of misclassification and the better the split. Once the Gini-impurity is at zero the terminal leaf node is reached, since only one class is remaining. The Gini-importance is then computed by adding up the decreases for each individual input feature over all trees in the RF ensemble. This gives us an aggregated impurity that represents the feature importance within the trained RF model.

Since the Gini-impurity metric is already computed as part of the RF model, the Gini-importance is widely used as a measure of feature importance (Chen et al., 2021; Nabil et al., 2022) and is one of the advantages RF classifiers have over alternative ML-based classification algorithms. The use of Gini-importance as a feature selection measure has been demonstrated to eliminate unimportant variables and outperformed other non RF-based approaches (Chen et al., 2020).

#### 4.3.5 Validation of classification model

One of the most important steps in the generation of a classification model is the evaluation of its performance. In our case, we want to see how well our sensor-specific binary classifiers can differentiate between water and snow/ice pixels.

##### 4.3.5.1 Validation metrics

To assess the performance of a binary classifier a variety of validation metrics are available. For the validation step we made use of the **Overall Accuracy (OA)**. OA is defined as the probability that an individual sample will be correctly classified. For this measurement we count the number of each possible outcome.

- **TN** (true negative) - count of water pixels correctly classified as water
- **FN** (false negative) - count of ice/snow pixels incorrectly classified as water
- **TP** (true positive) - count of ice/snow pixels correctly classified as ice/snow
- **FP** (false positive) - count of water pixels incorrectly classified as ice/snow

With all outcomes defined we can now define the validation metrics.

$$OA = \frac{TP + TN}{TP + FP + TN + FN} \quad (4.2)$$

As explained in Section 4.3.3, the distribution of water and ice/snow pixels in our study area is inherently imbalanced. The data-balancing procedure can only be applied to the training set. Balancing the test set would produce misleading validation metrics that do not represent the real model performance. Since the OA metric does not take this class imbalance in the test set into account, it will be skewed towards the classifier’s performance on the majority class. Therefore, we also calculate the **sensitivity** (also called true positive rate) and **specificity** (also called true negative rate or recall). The sensitivity describes the model’s ability to predict true positives – correctly classified ice/snow pixels. Whereas specificity describes the model’s ability to predict true negatives – correctly classified water pixels.

$$Sensitivity = \frac{TP}{TP + FN} \quad Specificity = \frac{TN}{TN + FP} \quad (4.3)$$

Furthermore, we also calculate the **Cohens Kappa** ( $\kappa$ ) that is well known in the geographical and medical sciences. The Cohens Kappa measures the inter-rater reliability for categorical items. The resulting kappa values are in a range from 0 to 1 and can be interpreted according to Table 4.6.

**Table 4.6:** Cohens kappa ranges and their interpretation according to Landis and Koch (1977)

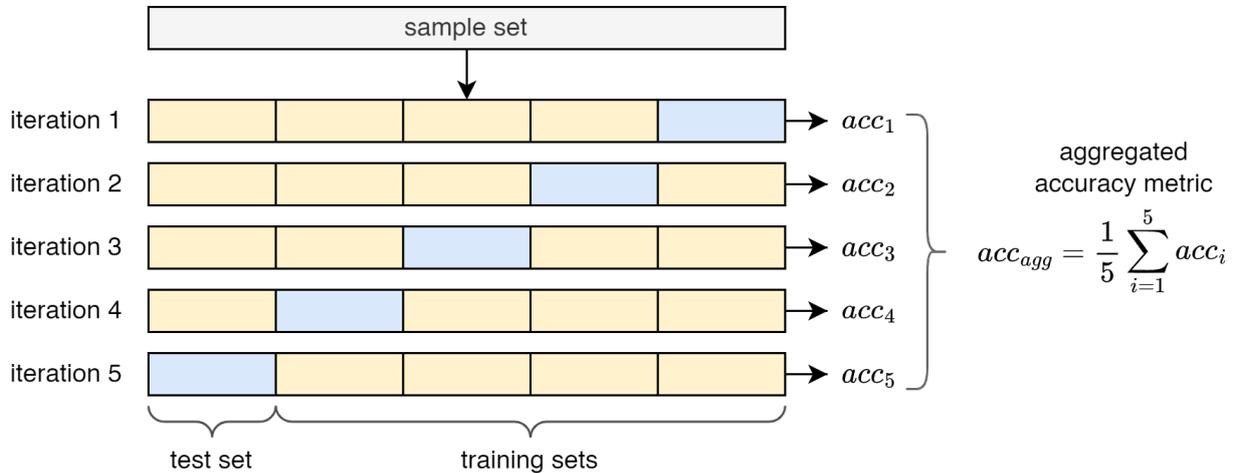
Cohen’s Kappa statistic ( $\kappa$ )	Strength of agreement
<0.00	Poor
0.00 - 0.20	Slight
0.20 - 0.40	Fair
0.41 - 0.60	Moderate
0.61 - 0.80	Substantial
0.81 - 1.00	Almost perfect

#### 4.3.5.2 k-fold cross-validation

The classical approach to test a classification model is to use a **train-test-split** that splits the input data into a training set for model-generation and the test set for model testing. Generally this split is applied randomly to generate around 70% training data and 30% test data with the same class proportions. The model is then trained with the large set and tested on the smaller set of unseen data. Validation metrics can finally be retrieved to evaluate the model’s classification performance.

A more robust approach is to use a **k-fold cross-validation** (k-fold CV). In this approach instead of a single split,  $k$  number of random splits with equal class distributions are first created. Then

the model is evaluated in  $k$ -iterations while holding out a single set for testing and  $k - 1$  sets for training at each iteration. The validation metrics retrieved are finally aggregated. Because  $k$ -fold CV is using all training data available, the risk for a biased model evaluation due to a poorly chosen split is reduced.



**Figure 4.9:** Schematic visualization of the  $k$ -fold cross-validation with 5 folds

For our model validation procedure we made use of a **10-fold CV** ( $k = 10$ ) to examine the **overall performance** of our binary sensor-specific classifiers.

#### 4.3.5.3 Leave one lake out cross-validation

In the  $k$ -fold CV the splits are applied randomly. Therefore, it is likely that all training lakes are available within the train and test set. Even though a single pixel is not allowed to be repeated in both sets, the chance is still very high that close neighbouring pixels from the same acquisition are present in both sets. Due to the high positive spatial auto-correlation that occurs in neighbouring pixels, the classifier is already familiar with the presented feature values, even though it is not trained on the specific sample itself. In case of a monitoring system without extensive training data the generalization of the classifier, the performance on unseen lakes is more important than the performance of classifying already seen lakes and values.

We can give an estimate on the generalization of the model using the **Leave-one-lake-out cross-validation** (LOLO-CV). Instead of applying random splits for the test and train sets, we can move all but one lakes inside the train set. The validation metrics will then tell us how well our classifier performs on the completely unknown held out lake. Analogous to the  $k$ -fold CV, we can repeat this for all lakes and get information about how well the classifier handles unseen lakes in the final monitoring system.

In the validation step we applied LOLO-CV for all sensors to examine the **generalization power** of our binary sensor-specific classifiers.

### 4.3.6 Hyper-parameter tuning

After the **feature-selection** step the final classifier models for all five sensors are generated using the balanced and feature-enhanced training sets. For the training procedure hyper-parameters that control the learning process have to be set by the user before starting the model generation. For this we apply a **hyper-parameter tuning** step. In this step four hyper-parameters (see Table 4.7) are tuned and set at values to optimally solve the classification problem.

**Table 4.7:** RF model settings and their definitions

Hyper-parameter	Definition	Default setting
<b>numberOfTrees</b>	number of decision trees to create as part of the ensemble	-
<b>variablesPerSplit</b>	number of variables to randomly sample per decision node/split	square root of input features
<b>minLeafPopulation</b>	only create nodes whose training set contains at least this many points	1

A widely used method for the hyper-parameter tuning is to use a **grid search**. A grid of manually set hyper-parameter values is used to train multiple models and assess their performance. The combination set of hyper-parameters that produces the best classification results is then used to train the final model.

A grid tuning with all three hyper-parameters and 100 settings each would require a total of 1 million ( $100^3$  combinations) models to be trained for each sensor. To reduce this number and make the tuning procedure feasible, we first extracted the optimal values for `numberOfTrees` (100 settings). With this parameter set, a grid-search using `minLeafPopulation` and `variablesPerSplit` is executed with 100 settings for each hyper-parameter ( $100^2$  combinations). This reduces the total amount of models to be trained from 1 million to only  $10 \cdot 100$  per sensor. The models are each evaluated using the LOLO-CV method to maximize the generalisation performance of the final sensor-specific models.

## 4.4 Information extraction

With the final models prepared, all input data for days with at least 30% of available lake coverage were processed and classified into water and ice/snow pixels. From the resulting classified time-series we then extracted the lake ice coverage (LIC) as a daily percentage of relative covered lake area. Subsequently this signal was processed in a lake ice phenology (LIP) event detection to find the start and end dates of the lake freezing and thawing periods.

#### 4.4.1 Lake ice coverage

The daily LIC values build one of the main outputs of this study. After generating the classified maps of water and ice/snow, we merged the LIC data retrieved from all sensors. This resulted in a time-series of daily datapoints which were then fitted with a curve using a sliding window aggregation with a window size of 10 days and a weighted mean. The weights were chosen according to the available unmasked lake coverage for each datapoint.

#### 4.4.2 Lake ice phenology

To define the phenology of a lake's ice coverage, we extract four key events that describe the ice cycles from the start of freezing to the end of thawing. We used the LIP event definitions proposed by Tom et al. (2020) shown in Table 4.8. For the extraction we iterated through the LIC curve and checked each day for the necessary LIP event conditions. An event was recorded if the conditions were fulfilled and the preceding events had been detected.

**Table 4.8:** Lake ice phenology event definitions as proposed by Tom et al. (2020)

Event	Definition
Freeze-Up Start (FUS)	30% or more of the non-cloudy portion of the lake is frozen and the just previous non-cloudy day should be <30% frozen
Freeze-Up End (FUE)	70% or more of the non-cloudy portion of the lake is frozen and the just previous non-cloudy day should be <70% frozen
Break-Up Start (BUS)	30% or more of the non-cloudy portion of the lake is non-frozen and the just previous non-cloudy day should be <30% non-frozen
Break-Up End (BUE)	70% or more of the non-cloudy portion of the lake is non-frozen and the just previous non-cloudy day should be <70% non-frozen
Ice Coverage Duration (ICD)	BUE - FUS
Complete Freeze Duration (CFD)	BUS - FUE

### 4.5 Validation of lake ice phenology

To assess the accuracy of the LIP event extraction, we compared the extracted events against events detected based on webcam imagery. A total of **86 LIP events** have been manually detected and extracted by one operator using collected webcam imagery. Only days with good visibility were processed using the same LIP event definitions used for the RS extraction. In an accuracy assessment all matching events available in both sets were compared using a **linear regression analysis**.

### 4.6 Correlation analysis

To understand the relationships between the extracted phenology and climatic and lake-specific factors we executed a correlation analysis. In this step we created a correlation matrix using the Pearson correlation coefficient (PCC). The PCC is calculated as the ratio between the covariance of two variables and the product of their standard deviations. Thus, it results in a normalized measurement ranging from -1 (perfect negative correlation) to 1 (perfect positive correlation).

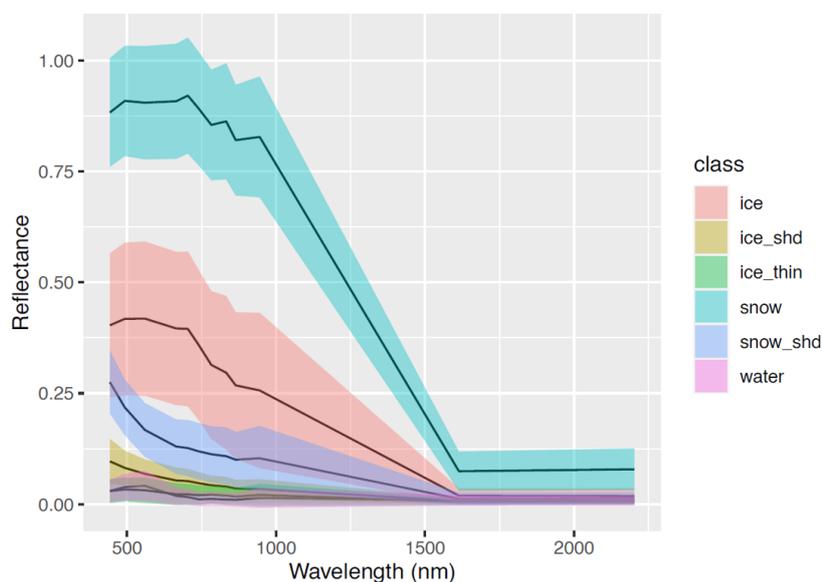
We calculated the PCC by combining the retrieved phenology with climatic data from the MeteoSwiss Grid-Data Products (see Table 3.2) and lake-specific attributes from the HydroLAKES v1.0 dataset (see Table 3.1). Since the MeteoSwiss product availability is limited to Switzerland, the climatic relationships were only calculated for 112 lakes. The lake-specific relationships were calculated for all lakes within the study area.

## 5 Results

### 5.1 Spectral class profiles

To get insight into the **class separability** of water and ice/snow covered surfaces, we extracted the mean spectral profiles of the fine class structure resulting from the **S-2 manual polygon classification**. The spectral profiles show a high separability between the snow, shadowed snow, ice and water surfaces. In comparison, the spectral differences between the classes water, thin ice and shadowed ice are much smaller and their spectra show a strong overlap. This overlap could potentially lead to high signal ambiguity and wrong classification results.

Generally, the spectral separability of the classes is high within the visual (400 to 700 nm) and near-infrared (700 to 1400 nm) spectrum and is strongly reduced in the short wave infrared (1400 to 3000nm) spectrum.



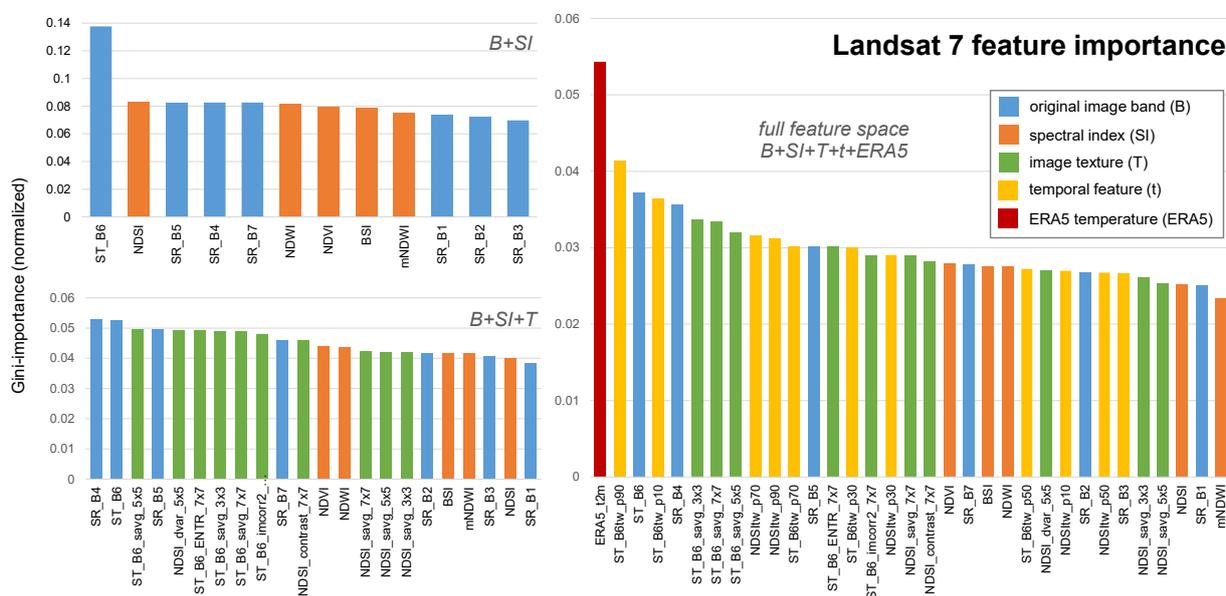
**Figure 5.1:** Mean spectral profiles retrieved from all training samples from the Sentinel-2 imagery created with the manual approach. The shaded areas corresponds to  $\pm 1$  standard deviation per band calculated from all available samples.

## 5.2 Training procedure

### 5.2.1 Feature importance

Feature importance was extracted using the normalized **Gini-importance** as explained in Section 4.3.4. The results were used to assess the importance of the input features and for the feature selection in the generation of textural and temporal features. The extracted importance metrics are only valid for the model and feature space they were extracted from. Therefore, executing the importance analysis with multiple combinations was necessary. We generated RF classifiers using different feature combinations for all sensors. We also extracted the importance metrics for the full feature spaces used for the final classifier models.

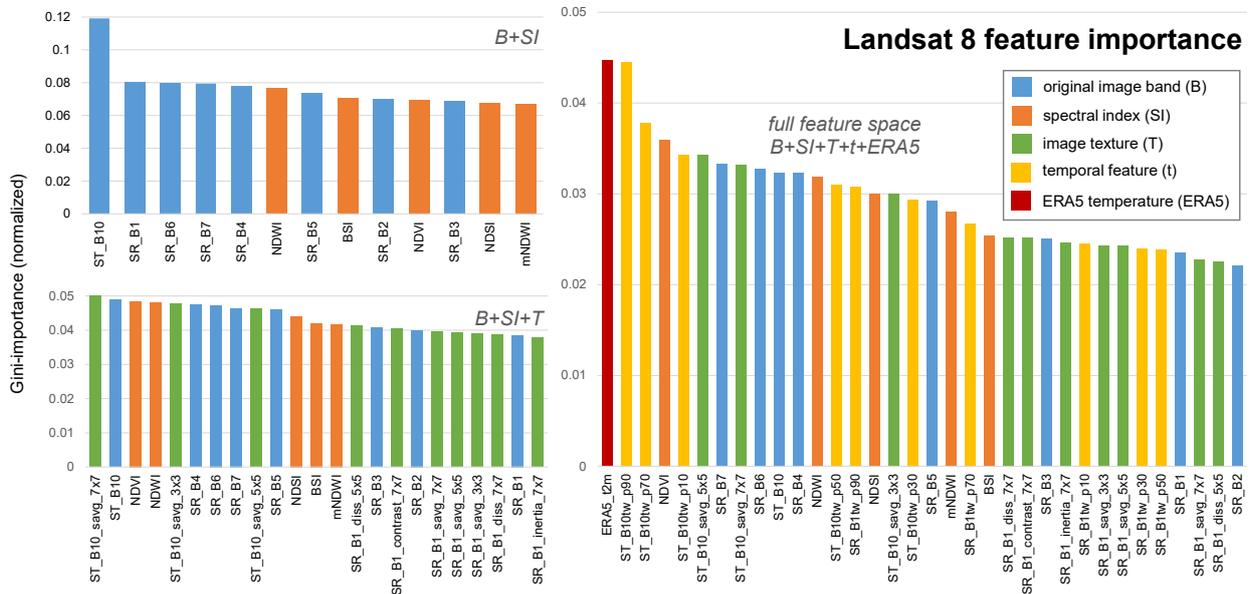
The feature importance results for the **L7 classifiers** (see Figure 5.2), combining all original image bands and spectral indices (B+SI), show that the surface temperature band **ST\_B6** is clearly the most important band for the classification. At the second place with much lower importance lies the **NDSI** feature. These two bands were further processed in the texture generation to generate the 10 most important texture features at different window sizes (3x3, 5x5, 7x7) and all temporal features. Overall the texture bands showed a higher importance compared to the spectral indices, once added to the feature space. In the full feature space the ERA5 air temperature feature and the surface temperature band, as well as temporal and textural information retrieved from it, show the highest importance values.



**Figure 5.2:** Feature importance of Landsat 7 RF classifiers with different feature combinations computed with normalized Gini-impurity outputs

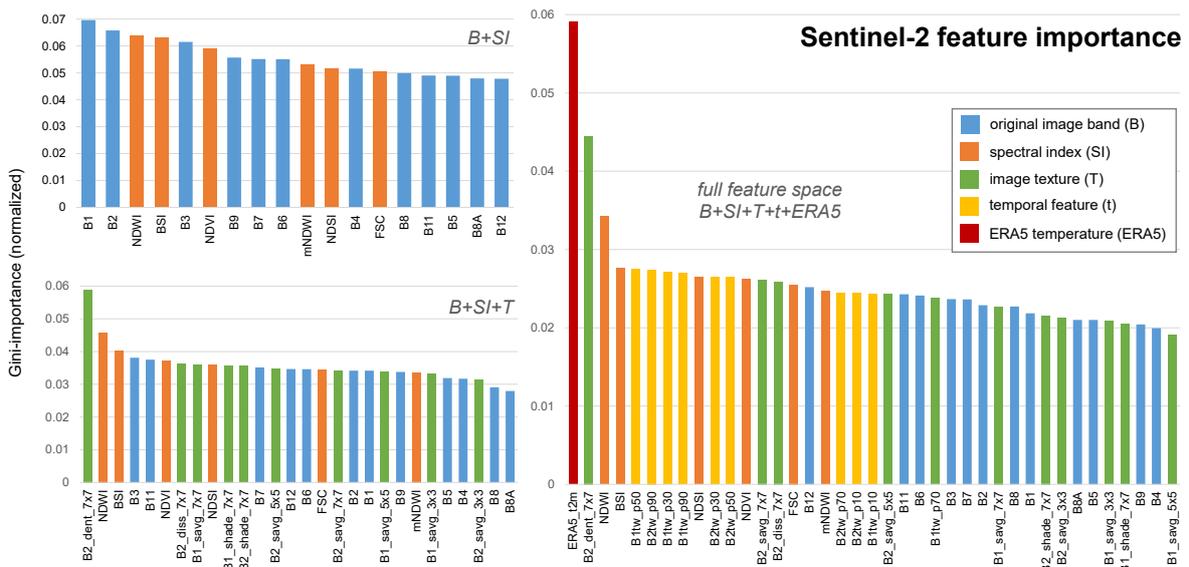
The results extracted for the **L8 classifiers** (see Figure 5.3) are similar to the ones from L7. They equally show that the highest importance for the B+SI combination is given to the surface temperature band **ST\_B10**. The second place with much lower importance, is given to the **SR\_B1** (blue) band. These two features were processed in the textural and temporal feature generation. In the

full feature space the ERA5 air temperature feature takes the highest place again. However, it is nearly matched by temporal features retrieved from surface temperature.



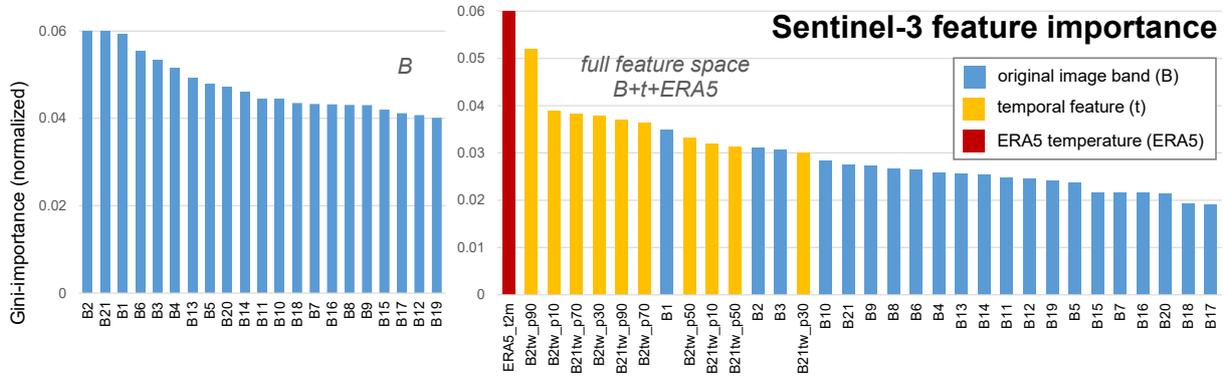
**Figure 5.3:** Feature importance of Landsat 8 RF classifiers with different feature combinations computed with normalized Gini-impurity outputs

The results for the **S-2 classifiers** (see Figure 5.4) are based on the manual training set. The B+SI combination show that the bands B1 (aerosol) and B2 (blue) were the most prominent features. These two features were selected for feature generation and further processed. In the full feature space the ERA5 air temperature feature takes over the first place followed by a B2-based texture feature.



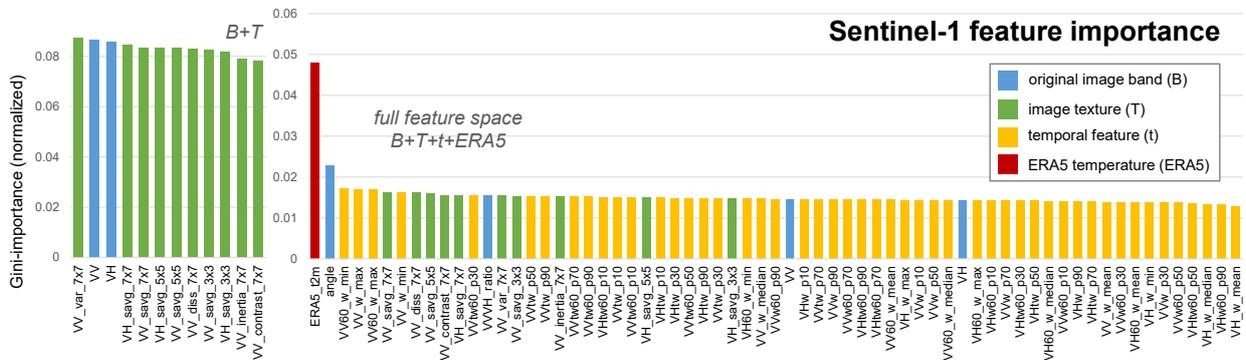
**Figure 5.4:** Feature importance of Sentinel-2 RF classifiers (manual approach) with different feature combinations computed with normalized Gini-impurity outputs

For the **S-3 classifiers** (see Figure 5.5) no textural features were calculated due to the low resolution of the sensor. The two most important bands were B2 (Yellow substance and detrital pigments) and B21 (atmospheric/aerosol correction). In the full feature space the ERA5 air temperature feature takes the lead again followed by temporal features based on the B2 band.



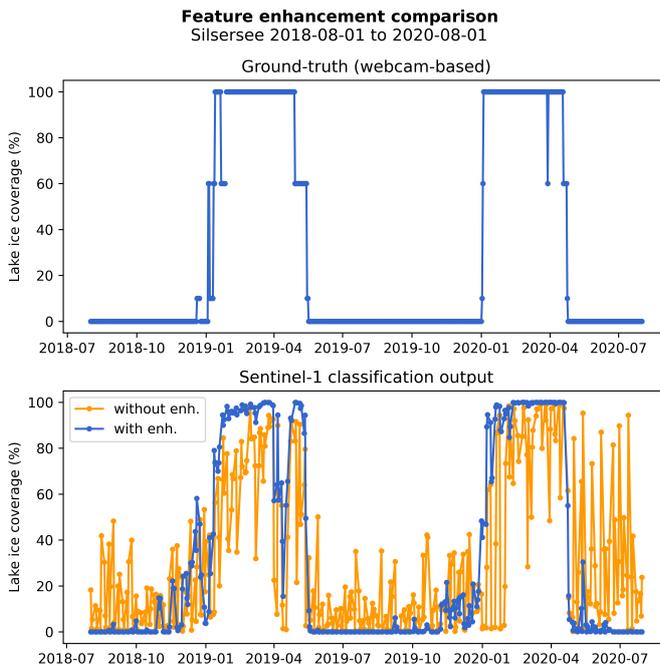
**Figure 5.5:** Feature importance of Sentinel-3 RF classifiers with different feature combinations computed with normalized Gini-impurity outputs

Finally, the feature importance results for the **S-1 classifiers** (see Figure 5.6) and the B+T combination show a slight preference for the VV-polarized over the VH-polarized band. For the full feature space the incidence angle measurements (angle) and a ratio based on both main input bands (VVVH\_ratio) were added. Here, the ERA5 air temperature feature is clearly the most important input feature, followed by the incidence angle. Generally, VV-based textures showed higher importance than VH-based ones.



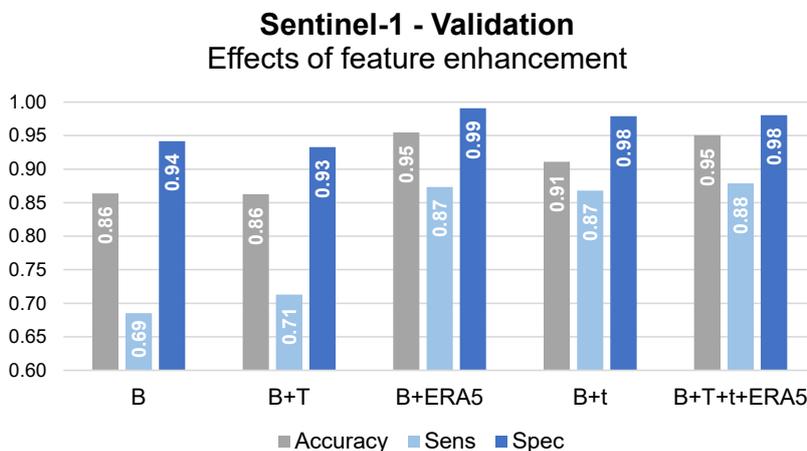
## 5.2.2 Effects of feature enhancement

To assess the effects of our feature enhancement step, we tested the performance of classifiers with different input combinations. The performance impacts with and without feature enhancement are best visible for the S-1 classifiers. In the extracted LIC curve we can see that the added features can greatly increase the stability of the retrieved ice signal (see Figure 5.7). Less notable improvements were observed for the optical classifiers.



**Figure 5.7:** Feature enhancement comparison of Sentinel-1 LIC output from lake Silsersee during the period from 1st of August 2018 to the 1st of August 2020. The upper plot shows the estimated LIC using the webcam-based coarse-classification. The lower plot shows the classification output of two Sentinel-1 classifiers with application of feature enhancement and without.

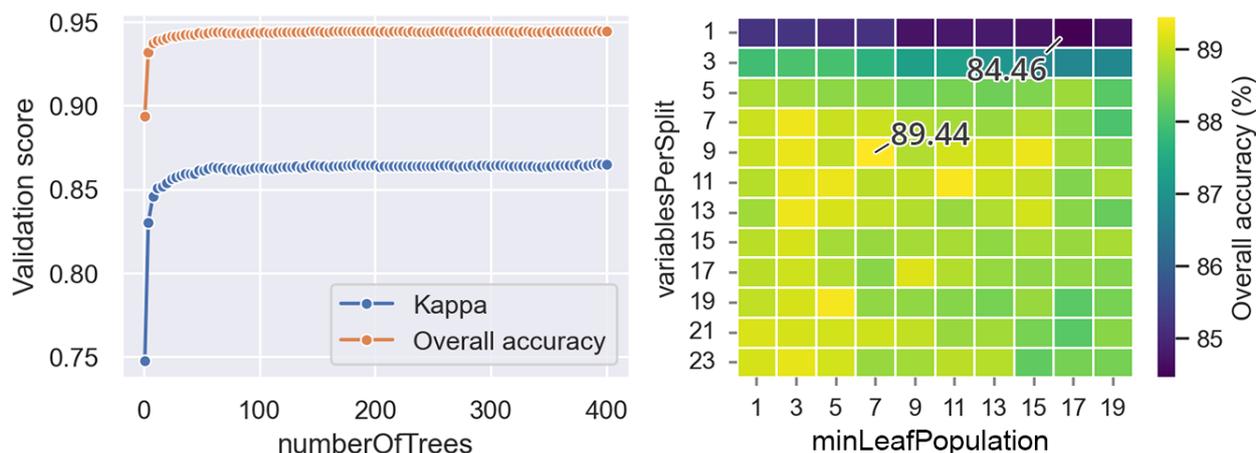
By adding each feature enhancement set individually and running a 10-fold cross-validation, we can compare the impacts of the added textural, temporal and temperature features (see Figure 5.8). The greatest impact on performance can be observed by adding the ERA5 air temperature features. Slightly less notable, but still high impacts on performance are obtained by adding the temporal features. In comparison, the performance increase from the addition of textural features is low. Overall the results show that the feature enhancement can greatly increase the sensitivity, the model’s ability to correctly classify ice pixels.



**Figure 5.8:** Comparison of validation metrics retrieved from Sentinel-1 classifiers with different feature combinations including the final full feature space.

### 5.2.3 Hyper-parameter tuning results

In the hyper-parameter tuning step we estimated the optimal model settings for three hyper-parameters as explained in Section 4.3.6. We used the 10-fold-CV for the hyper-parameters `numberOfTrees` and the LOLO-CV for `minLeafPopulation` and `variablesPerSplit`. The resulting optimal settings are displayed in Table 5.1 together with the suggested Smile RF default settings.



**Figure 5.9:** Hyper-parameter tuning results for Sentinel-1. The left plot shows the tuning results of the hyper-parameter `numberOfTrees` using 10-fold cross-validation. The right plots shows the grid-tuning results of the hyper-parameters `minLeafPopulation` and `variablesPerSplit` using leave-one-lake-out cross-validation.

We found that the performance gains from an increase of `numberOfTrees` stagnate at zero after reaching a maximum of approximately 200 trees. This was observed for all trained models. At this size, adding more trees to the model would only increase the training time.

The hyper-parameters `minLeafPopulation` and `variablesPerSplit` were tuned together in a grid-tuning to optimize model generalization. Contrary to the number of trees, these two hyper-parameters showed peak model performance at a maximum followed by a decrease.

**Table 5.1:** Optimal model settings from the hyper-parameter tuning step

Hyper-parameter	Default settings	Tuning results				
		S-1	S-2	S-3	L7	L8
<code>numberOfTrees</code>	-	200	200	200	200	200
<code>variablesPerSplit</code>	square root of input variables	9	11	5	3	5
<code>minLeafPopulation</code>	1	7	1	17	15	3

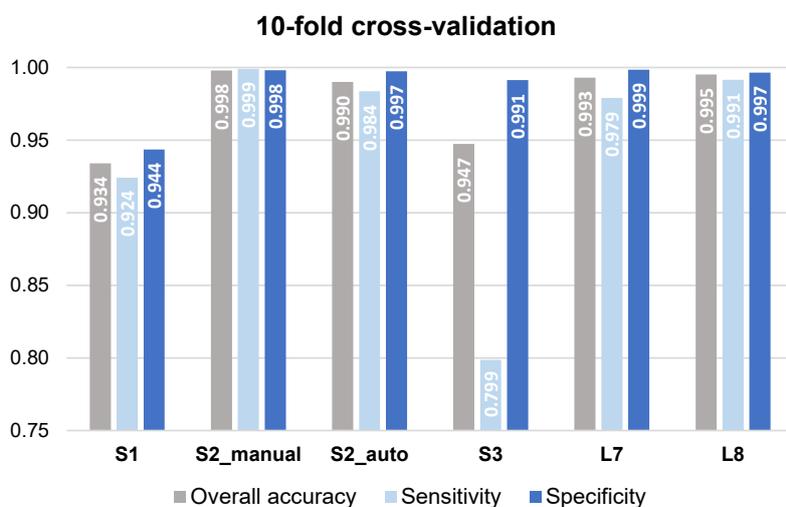
## 5.3 Validation results

### 5.3.1 Binary Classification

For the **performance assessment** of the tuned and enhanced final classifier models we applied a **10-fold CV** (see Figure 5.10). The retrieved validation result from the high-resolution optical L7, L8 and S-2 classifier all show excellent performance at similar scores for all three applied metrics.

For the S-2 input imagery the performance of both models, trained using the manual approach and using the automated approach, were extracted to allow a comparison of the two approaches. The classifier trained using the manual approach showed a higher score in all three metrics and was therefore chosen for the LIC and LIP extraction.

The S-1 classifier shows the lowest overall accuracy (OA) score of 93.4%. However, even though the S-3 classifier performed slightly better on the OA with 94.7% it shows the worst performance in sensitivity with only 79.9% of correctly classified ice pixels. This is possible due to the inherent class imbalance in the validation set and indicates that the S-3 classifier performs the worst in the distinction of water and ice/snow pixels.



**Figure 5.10:** Validation results from the 10-fold cross-validation of the final classifier models. Validation metrics displayed are computed as the aggregated mean of all folds.

### 5.3.2 Leave one lake out cross-validation

To assess the **generalization performance** of the tuned and enhanced final classifier models, we further applied a **LOLO-CV**. This validation method tells us how well the classifiers can perform on samples from unseen lakes. A full overview of all values can be found in Appendix B. The sensor-aggregated results are shown in Table 5.2 and the lake aggregated results in Table 5.3.

The reached mean overall accuracy is within excellent and moderate levels, ranging from a maximum of 96.3% with S-2 (manual) to a minimum of 83.8% with S-1. Due to the imbalanced class distribution the metrics sensitivity and specificity are more meaningful than OA. For Specificity, the ability to correctly classify water pixels, the mean values range from a maximum of 98.6% with

S-2 (manual) to a minimum of 85.1% with S-1. For Sensitivity, the ability to correctly classify ice pixels, the mean values range from a maximum of 97.8% with S-2 (auto) to a minimum of 74.1% with S-3. With the lowest mean sensitivity values and the lowest minimal sensitivity values the S-3 classifier show the overall worst generalization performance.

**Table 5.2:** Sensor-aggregated results from the Leave-one-lake-out cross-validation

Sensor	Overall accuracy		Sensitivity		Specificity	
	Mean	Min	Mean	Min	Mean	Min
S-1	83.8	67.6	74.4	53.5	85.1	64.5
S-2 (manual)	96.3	91.3	94.4	84.2	98.6	95.2
S-2 (auto)	94.7	89.1	97.8	92.6	89.4	68.9
S-3	86.8	79.6	74.1	27.0	89.7	42.9
L7	94.5	87.7	89.2	72.3	97.9	89.6
L8	93.8	67.9	93.5	67.8	98.0	90.0

The lowest per lake mean OA, mean specificity and mean sensitivity were all reached for St.Moritzersee with values of 77.8%, 78.0% and 76.9% respectively. The lowest sensitivity was reached for Eibsee with a value of 27.0% (S-3).

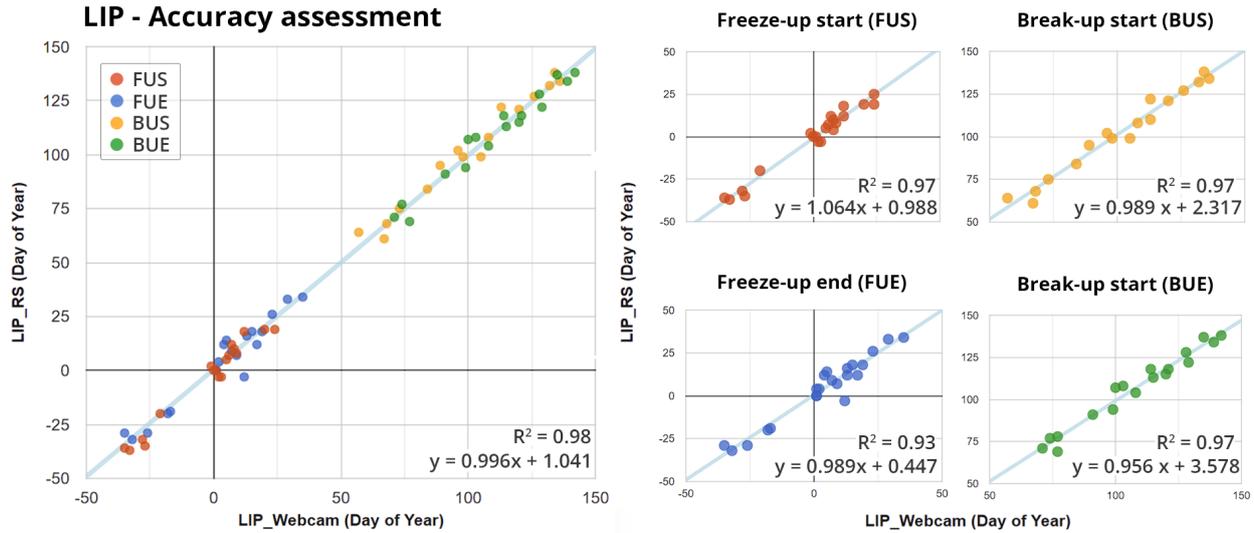
**Table 5.3:** Lake-aggregated results from the Leave-one-lake-out cross-validation

Lake	Overall accuracy		Sensitivity		Specificity	
	Mean	Min	Mean	Min	Mean	Min
Eibsee	90.3	81.1	79.8	27.0	94.0	68.9
Schlegeis	92.8	82.4	91.8	66.7	90.6	74.7
Schwaigerhaus	92.0	88.8	94.7	75.8	92.2	88.5
Silsersee	91.3	87.0	88.5	78.4	94.0	88.3
Turrachersee	93.6	84.2	90.7	64.5	97.4	88.7
Zellersee	96.0	85.6	78.8	53.5	97.7	88.8
Sihlsee	92.8	84.9	79.9	55.2	98.0	93.5
Silvaplannersee	89.5	80.8	89.1	81.5	92.5	70.0
Silvaplannersee	77.8	67.6	78.0	67.8	76.9	42.9

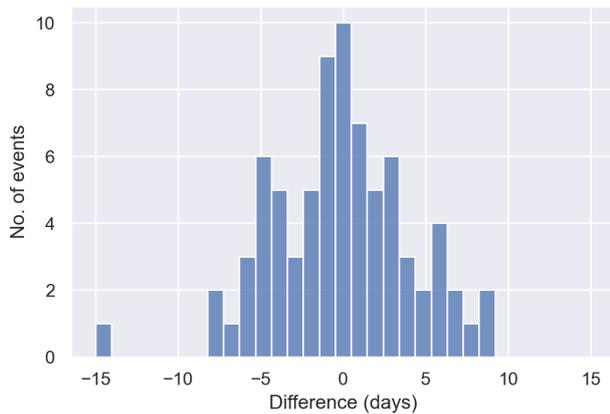
### 5.3.3 Lake ice phenology event extraction

To assess the accuracy of the LIP events extracted from LIC data we applied a **linear regression analysis**. A total of 78 LIP events (LIP\_RS) extracted from our remotely sensed data were matched up against manually extracted events (LIP\_GT) retrieved from webcam imagery. We were not able to match up 8 events due to missing cycle detections resulting in a LIP event omission error of 9.3%. The matched up events were fitted with a regression line resulting in an  $R^2$  value of 0.98 (see Figure 5.11).

The achieved mean difference of the matched events lies at  $-0.17 \pm 4.3$  days. Figure 5.12 shows the distribution of difference between LIP\_GT and LIP\_RS. The highest difference was found for a FUE event which was detected 15 days too early using the RS-based LIP algorithm.



**Figure 5.11:** Linear regression analysis using remotely sensed LIP events as predictor for observations from webcam-based LIP events

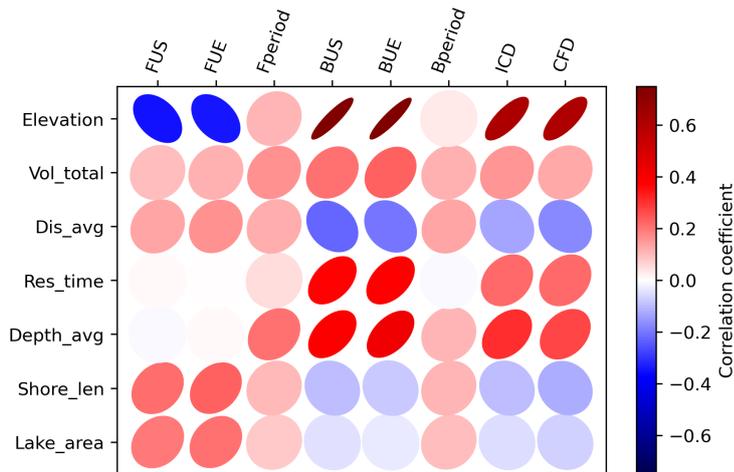


**Figure 5.12:** Histogram of time differences between remotely sensed LIP event dates and webcam-based events. Negative values indicate earlier remotely sensed detection of date.

## 5.4 Correlation analysis

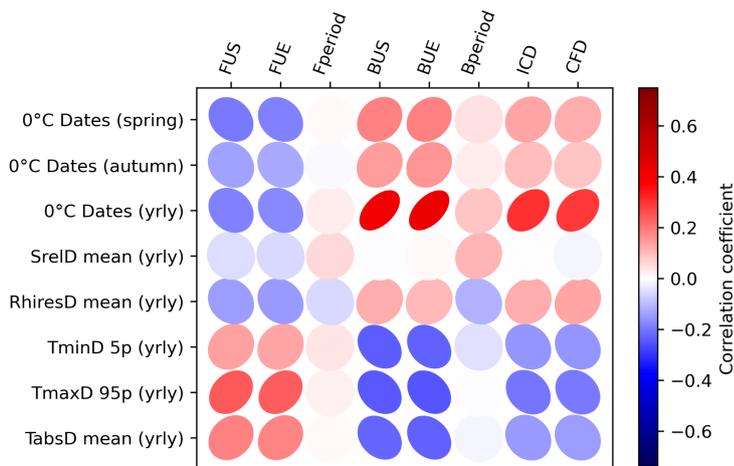
To explore the relationships between extracted phenological events and lake-specific as well as climatic factors, we applied a **correlation analysis**. As explained in Section 4.6, full LIP cycles extracted from the dataset were extended with lake attributes from the HydroLAKES dataset (see 3.3.1) and climatic variables from the MeteoSwiss Grid-Data Products (see 3.4.2) to build two correlation matrices. For the entire study area 387 lakes (75.6% of all study lakes) were found to have at least one full ice cycle from the observed period from the year 2016 to 2021.

For the correlation analysis between **LIP events and lake-specific factors** a total of **1775 ice cycles** were extracted from the LIC coverage over the entire study area. 85 ice cycles shorter than 4 days (based on CFD) were filtered out. We then selected meaningful lake-specific attributes that estimate the hydrological and geomorphological properties of the lakes. The retrieved coefficients reach a maximum positive correlation of 0.76 between BUS and elevation and a minimum negative correlation of -0.35 between FUS and elevation.



**Figure 5.13:** Correlation matrix showing the correlation between LIP events and HydroLAKES attributes based on the Pearson's coefficient. Blue ellipses show a negative correlation and red ellipses a positive correlation. The ellipticity of the plotted datapoint indicates the strength of correlation.

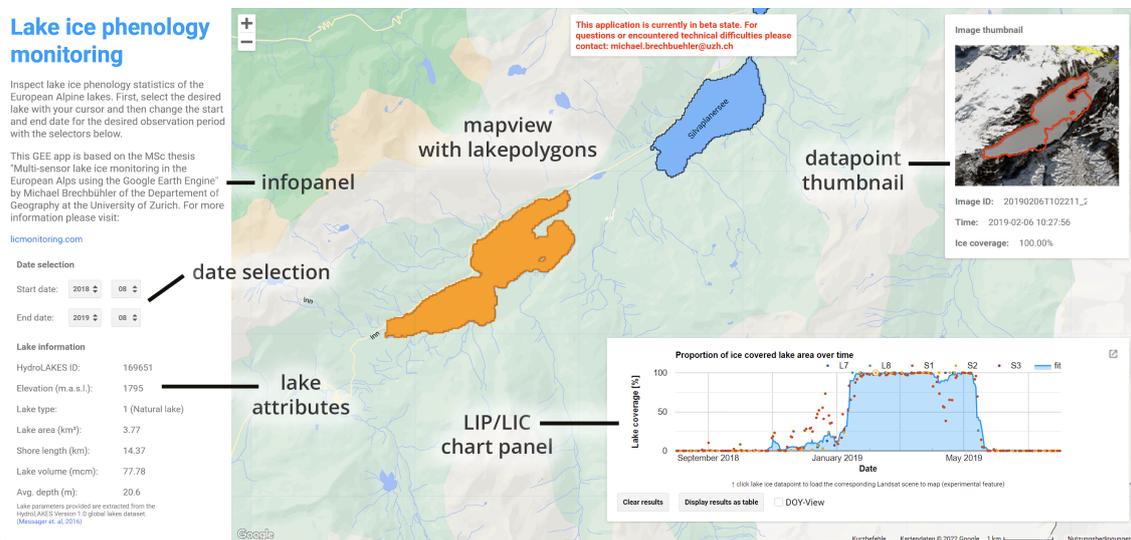
For the correlation analysis of **LIP events and climatic factors** only lakes within Switzerland were extracted due to the spatial availability of the MeteoSwiss Grid-Data Products. We applied the same filtering condition to retrieve **502 ice cycles**. The retrieved coefficients reach a maximum positive correlation of 0.43 between BUE and mean yearly 0 °C days. A minimum negative correlation of -0.25 was reached between BUE and the 95-percentile yearly maximum air temperature.



**Figure 5.14:** Correlation matrix showing the correlation between LIP events and climatic factors based on the MeteoSwiss Grid-Data Products. Blue ellipses show a negative correlation and red ellipses a positive correlation. The ellipticity of the plotted datapoint indicates the strength of correlation. The 0 °C dates (autumn/spring) represent the seasonal number of days when air temperature reached 0 °C.

## 5.5 Google Earth Engine Application

To visualize the LIC and LIP data extracted for this thesis and to make the data accessible for potentially interested users, we have developed a **GEE application**. Figure 5.15 shows an overview of the application with annotations for all main working parts. Within the application a user can query LIC data and LIP event dates for all of the study lakes within the Alpine Convention Perimeter. After selecting the lake of interest, the user can specify a time range in the infopanel. The estimated daily LIC values are then displayed with the fitted curve in a chart panel. With a button the chart can be converted to a table with detected LIP events within the selected time range. Furthermore, individual datapoints can be clicked to get a thumbnail preview of the scene used for LIC extraction. With this feature it is possible to quickly identify outliers. The application can be accessed at [licmonitoring.com](http://licmonitoring.com).



**Figure 5.15:** Overview of the GEE lake ice monitoring application with annotations of the main working parts

## 6 Discussion

### 6.1 Strengths and weaknesses of our Methods

The main focus of this thesis has been the implementation of a lake ice monitoring system using RS imagery and pixel-wise supervised classification models. From the outset, we have approached this problem using a multi-sensor approach. Due to the trade-offs between temporal and spatial resolution of current RS sensor technology, this is the only approach to achieve high-frequency observations of lake ice down to lake sizes of  $0.1 \text{ km}^2$  and the resolution requirement of 300m posed by GCOS. Furthermore, due to the cloud effects in optical sensors, we had to expand the input imagery with the active cloud-penetrating S-1 C-SAR products to fulfill the requirement of high temporal resolution.

One of the key steps of this process was the generation of five sensor-specific training sets. The manual interpretation and classification of imagery from multiple sensors is a labour intensive task with many pitfalls. Generally, it is not possible to create a perfect and completely unbiased training set using this approach. Furthermore, the interpretation and labeling of SAR imagery can be very difficult. We used the traditional manual approach to create a S-2 training dataset with a fine class structure. Firstly, we used this training set to show, that water and ice/snow classes can be spectrally distinguished (see Figure 1.1) even with the effects of topographic shadowing that is affecting a large proportion of the studied alpine lakes. Secondly, we were able to use it to create a classifier that showed very good classification results for the binary classification of water and snow/ice pixels. To automatically generate the remaining training sets we adopted the method put forward by Tom et al. (2019). In the automated approach webcam-imagery is used to classify fully ice-covered and ice-free days. For these days, samples are then taken from the entire lake area. Using this approach, the manual interpretation is shifted from RS imagery to webcam imagery that can be more accurately interpreted. Another benefit is that we can use this method to generate training data for SAR imagery. In our validation we were able to compare the manual and automated approach using two differently created S-2 training sets. The comparison (see Figure 5.10) has shown that the automated approach is able to deliver satisfactory results comparable to the results achieved with the traditional manual method used as a benchmark.

Nonetheless, there are some clear disadvantages to this approach. The labeling of webcam imagery can be difficult. In our case the webcams that were used did not always cover the entire

lake extent. Thus, it is possible that days with partial coverage were processed and produced wrong sample labels. Even though the interpretation of webcam imagery is a much easier task than the interpretation of RS imagery, there are still ambiguous signals that can be encountered. For example, thin ice is very hard to differentiate from open water. Thus, it is likely that the resulting labels are inconsistent. Another big disadvantage is that days with partial coverage, that build a key element of lake ice phenology and define the transition state, are ignored. This leads to less training samples and more importantly, none of the retrieved samples contain information from the partial state. Overall, we also found that the coverage of high-quality webcam imagery that is freely available is still very limited. We were only able to source freely available imagery from two providers. As a result, the selected training lakes are of small number and additionally clustered together in the central part of the study area. For this reason, it is possible that the generated training sets are not representative of all lakes within the study area.

With the training sets fully prepared we applied a supervised pixel-wise binary classification. For this step a multitude of classification methods are available. We opted for Random Forest models due to many reasons. RF models are generally regarded as less prone to overfitting than other classification models available (Fernández-Delgado et al., 2014). Breiman (2001) has shown that due to the bagging and random feature selection, the number of decision trees in a RF model can be increased without the risk of overfitting. Fernández-Delgado et al. (2014) has further shown that RF models outperform other ML-based models in real world application. A major advantage is that the training and classification procedure in RF models can be easily parallelized. In comparison, linearly executed models such as the often used Support Vector Machine (SVM) are much less memory- and time-efficient. This is especially important for a fully-parallelized environment such as the Google Earth Engine. However, studies with more contemporary Convolutional Neural Networks (CNN) have shown to achieve higher classification performances for lake ice (Tom et al., 2020) classification than our models have reached.

A clear benefit of the RF model is the supplementary output of importance metrics. For our thesis the Gini-importance was a key tool in the identification of important input features and feature selection. The feature selection was necessary to allow a model computation within the memory constraints posed by the GEE platform. With the extracted importance metrics we were able to show that the most important input features were the ERA5 air temperature feature and the surface temperature bands if available (L7 and L8). The feature importance analysis also showed that the added textural and temporal features were of high importance for all classifiers. For the input bands of optical classifiers the results show that bands in the lower visual spectrum (blue) are of higher importance. This finding fits the spectral profiles retrieved for S-2 which show the highest class separability in the lower part of the visual spectrum.

To finish the model generation, we assessed each model's performance using 10-fold cross-validation (10-fold CV), a well-established method to estimate model performance. The results showed high overall accuracy (OA), sensitivity and specificity scores ranging from 92.4% up to 99.8% for all

models except for S-3 which has shown a clear underestimation of ice pixels with a sensitivity of only 79.9%. However, we argue that this validation method is not fitting for our purpose. Due to the spatial auto-correlation that occurs in neighbouring pixels, the classifier is already familiar with the presented feature values even though it is not trained on the specific sample itself. This leads to over-optimistic validation results. Since our intention is to build an extensive monitoring system, the performance on data of unseen lakes is much more important. Therefore, we subsequently applied a leave-one-lake-out cross-validation (LOLO-CV) to assess the generalization performance of our classifiers. In the LOLO-CV results we can see that all validation metrics dropped in value. Most notable, the sensitivity, the ability to correctly classify ice pixels dropped strongly for S-1 and S-3 to a mean sensitivity of 74.4% and 74.1%, respectively. We believe that the main cause for the drop in S-1 performance are untrained incidence angles. A lake is usually covered by four relative orbits with specific incidence angles for lake pixels. The retrieved backscatter signal of thin ice surfaces can differ with changing angles Atwood et al. (2015). If the model is not trained on the matching relative orbits and specific angles it is likely to perform worse in the signal classification. The LOLO-CV results for S-3 show overall bad generalization performance and high ice underestimation. We believe that the reasons for this are mainly geometric errors in the georectification process and spectral mixing (see 6.2.1).

To finish the lake ice coverage (LIC) extraction we merged the datapoints extracted from all individual sensors. We applied a moving window aggregation with a weighted mean. By using weights based on the available lake coverage our method takes partial lake coverage into account and gives higher preference to non-cloudy images. With this approach it possible to smooth out the retrieved signal, remove outliers and provide daily LIC datapoints. We decided on a window size of 10 days for the aggregation. This value showed good results in our tests and is in accordance with the window size tuning results from (Tom et al., 2019). However, due to the high impact on the final LIC results, we recommend that a more thorough analysis of the chosen aggregation parameter and aggregation metric should be carried out.

## 6.2 Lake ice phenology event extraction

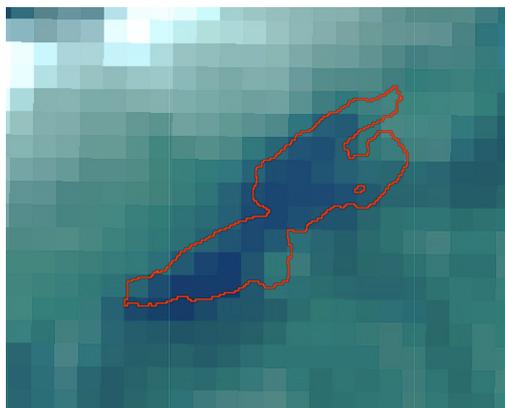
To extract phenological key dates that describe the freezing (FUS and FUE) and thawing (BUS and BUE) cycles, we further applied a LIP event detection algorithm. Currently, there are no widely agreed-upon threshold values to clearly define these transitional states between the fully-unfrozen and fully-frozen state. A multitude of LIC threshold values have been used in lake ice studies (Tom et al., 2020). We have chosen to define the transitional states based on a 30% and 70% thresholds as proposed by Tom et al. (2020) (see Table 4.8). To assess the accuracy of the extracted LIP event dates we matched up 78 LIP events labeled from webcam imagery against the remotely sensed dates in a linear regression analysis. We were able to show that our algorithm can identify LIP event dates with a high  $R^2$  value of 0.98. However, the achieved mean difference of  $-0.17 \pm 4.3$  days, between predicted and independent variable, shows a large spread. The resulting distribution of the mean difference (see Figure 5.12) shows that in most cases we cannot exactly

identify the LIP event date. In turn, this indicates that the daily LIC uncertainty of  $\pm 10\%$  posed by GCOS is likely not reached by our algorithm. We argue that the cause for this relatively high spread partially lies in the inaccuracy of the webcam-based LIP dates. In most cases it is very hard or impossible to clearly define the relative coverage from webcam imagery within a 10% margin.

### 6.2.1 Problems of input imagery

During the data-processing steps in our study we came across several problems regarding the used input imagery. Most problems were encountered with the low resolution (300 meter) S-3 imagery. A crucial difference from S-3 to the other optical datasets used in this study is the processing level. Whereas the S-2, L7 and L8 datasets are all processed to level-2 surface reflectance, the S-3 dataset is currently only available as level-1 top-of-atmosphere (TOA) radiance on GEE. This means that the used imagery is uncorrected for atmospheric effects, which is likely contributing to the low achieved performance. In addition, due to the lack of processing, no supplementary cloud flagging was available. Thus, we had to implement a custom cloud masking algorithm based on the cloud masks of MODIS imagery (see Section 3.1.3). Within this thesis we were not able to thoroughly assess the performance of this cloud masking algorithm. We are therefore not confident that the resulting images are cloud-masked with a high accuracy.

Another major problem lies in the low-resolution of S-3 imagery. Our screening has shown that many images have geometric errors and do not align well with lake polygons. Figure 6.1 illustrates this problem with the example of Silsersee, one of the medium-sized training lakes at 4.1 km<sup>2</sup>. Additionally, due to the coarse resolution, pixels along the shoreline have a high tendency for spectral mixing. Thus, the measured signal is a result of overlapping signals originating from both land and lake. Due to the small lake sizes studied, we were not able to buffer the lakemasks and discard shoreline pixels, as this would have resulted in only a few to no remaining pixels for the smallest training lakes. We infer that S-3 is not well suited for the analysis of lake ice cycles at the small lake sizes studied in this thesis ( $< 10.7 \text{ km}^2$ ).

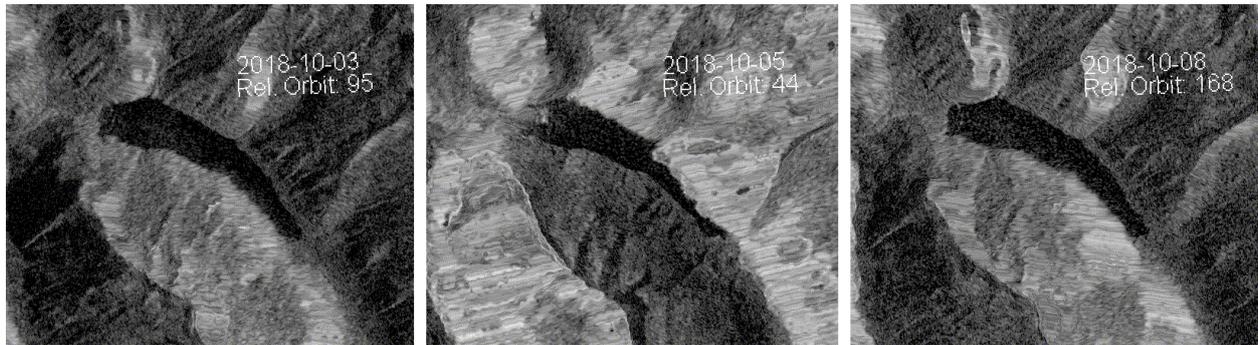


**Figure 6.1:** Cropped S-3 scene of lake Silsersee with overlapped lake polygon. Due to geometric misalignment and spectral mixing the signal retrieved over the lake is untrustworthy.

For the S-1 imagery, our validation has also shown lower performance in comparison with the high-resolution optical sensors. The main cause for this is low sensitivity for ice/snow (see Figure 5.10). Since the SAR imagery is based on microwave backscatter, moisture content and surface roughness have to be considered. Conditions affecting the roughness of the measured surfaces

will alter the retrieved backscatter. This can cause signal ambiguities between ice and water when the ice is wet and smooth. Additionally, windy conditions can increase the surface roughness of water, increasing the backscatter and leading to difficulties in ice detection (Sobiech and Dierking, 2013).

The screening of S-1 imagery has also shown that for lakes with steep surrounding slopes severe layover effects can cause geometric distortions. These effects are not corrected in the terrain-correction steps applied in the GEE processing. Figure 6.2 illustrates this problem with a timeseries retrieved from Speicher Zillergründel in Austria. As a result, the backscatter retrieved from lake pixels can be altered, leading to drops in classification performance.



**Figure 6.2:** Cropped Sentinel-1 (VV-polarized) time-series of Speicher Zillergründel, Austria. The consecutive days are imaged from different relative orbits. The oblique viewing angle of the acquisition on the 5th of October 2018 leads to pronounced layover effects over the lake, affecting the retrieved backscatter signal.

### 6.3 Implications of correlation analysis

To explore the relationships between lake ice phenology and lake-specific and climatic factors we applied a correlation analysis. As a part of this analysis we were able to show that the majority of the alpine lakes within the study area (387 lakes, 75.6% of all study lakes) show at least one ice cycle during the period of 2016 to 2021.

For the correlation between **lake-specific factors** and LIP events (see Figure 5.13) our main findings are:

- Elevation shows the highest correlation with LIP events. Elevation and freeze-up patterns show a moderate negative correlation (-0.34). Break-up patterns show a strong positive correlation (0.77). This indicates earlier freeze-up, later break-up and overall increased ice duration at higher elevations. From the strength of correlation, we conclude that break-up patterns are more affected by elevation.
- Mean average depth and residence time mainly show a moderate positive correlation with break-up pattern (0.40 and 0.37), leading to later ice break-up with increasing depth and

higher residence time.

- Lake area and shoreline length have a moderate positive correlation with freeze-up (0.19 and 0.22) and slight negative correlation with break-up patterns (-0.04 and 0.08). This indicates later freeze-up, earlier break-up and overall decreased ice duration for increases in lake size and shoreline length.

For the correlation between **climatic factors** and LIP events (see Figure 5.14) our main findings are:

- The strongest correlation can be found between the mean number of yearly below 0°C dates and ice cycles with a moderate negative correlation for freeze-up (-0.18) and a moderate positive correlation for break-up patterns (0.42). This indicates earlier freeze-up, later break-up and overall increased ice duration with increase in days below 0°C.
- The mean yearly air temperature, 5th percentile minimum and 95th percentile maximum air temperature all show a moderate positive correlation with freeze-up (0.19) and a moderate negative correlation with break-up patterns (-0.24). This shows that higher air temperatures reduce the length of ice cycles across the studied lakes.
- Precipitation shows a moderate negative correlation with freeze-up (-0.15) and a moderate positive correlation with break-up patterns (0.11). This indicates that ice cycles increase in duration with higher experienced precipitation.

## 7 Conclusion

As we have now shown and discussed our results, we would like to address our initial research questions.

- **RQ1:** How accurately can lake ice coverage (LIC) be extracted using RS-based methods in an alpine environment?

Our model validation has shown that our generated classifiers can retrieve lake ice coverage with a mean overall accuracy of 96.1%. We reached a mean sensitivity (ability to correctly classify ice) of 93.8% and a mean specificity (ability to correctly classify water) of 98.4%. The leave-one-lake-out cross-validation has shown that the generalization performance, the accuracy to classify pixels of unseen lakes, is significantly lower with mean overall accuracy, mean sensitivity and mean specificity values of 91.0%, 85.1% and 93.9% respectively

- **RQ2:** How do the different sensors (S-1, S-2, S-3, L7, L8) compare on the ice/non-ice classification accuracy?

Our model validation using a 10-fold cross-validation has shown that the high resolution optical sensor S-2, L7 and L8 used can retrieve lake ice coverage with a high mean overall accuracy across sensors of 96.2%. We reached a mean sensitivity (ability to correctly classify ice) of 98.8% and a mean specificity (ability to correctly classify water) of 99.8% across these sensors. In comparison S-1 and S-3 have shown worse performances and underestimation of ice with overall accuracy values of 93.4% and 94.7% and sensitivity values of 92.4% and 79.9%, respectively.

- **RQ3:** Which multi-spectral indices, temporal features, image textures and air temperature features can be used as input features to enhance the classification accuracy of lake ice classifiers?

With our feature importance analysis and model evaluation results we have shown that the feature enhancement using ERA5 air temperature data has the highest potential to improve model performance for ice classification. The most prominent increase in performance was

reached for S-1 with OA, sensitivity and specificity increases of 9.1%, 4.9% and 18.8%, respectively. At the second place, temporal features such as the proposed weekly aggregated difference to the yearly mean, have shown a similarly high feature importance and potential to improve model performance. GLCM-based textures have also shown a small improvement with moderate to high feature importance. Finally, the addition of spectral features for optical sensor has shown only a small increase in model performance. We advise their use mainly as a substitute for image bands.

- **RQ4:** How accurately can lake ice phenology (LIP) events be extracted using RS-based methods in an alpine environment?

Our algorithm was able to extract key lake ice phenology dates (FUS, FUE, BUS, BUE) with a  $R^2$  value of 0.98 and a difference of  $-0.17 \pm 4.3$  days between remotely sensed and ground-based measurements. This shows that LIP dates can be accurately derived from remotely sensed imagery.

- **RQ5:** Which relationship exists between the measured lake ice phenology and climatic factors (e.g. temperature)?

The strongest correlation can be found between the mean number of yearly below  $0^\circ\text{C}$  dates and ice cycles with a moderate negative correlation for freeze-up (-0.18) and a moderate positive correlation for break-up patterns (0.42). This indicates earlier freeze-up, later break-up and overall increased ice duration with increase in days below  $0^\circ\text{C}$ .

- **RQ6:** Which relationship exists between the measured lake ice phenology and lake-specific variables (e.g. lake area, lake depth)?

Elevation shows the highest correlation with LIP events. Elevation and freeze-up patterns show a moderate negative correlation (-0.34). Whereas, break-up patterns show a strong positive correlation (0.77). This indicates earlier freeze-up, later break-up and overall increased ice duration at higher elevations. From the strength of correlation, break-up patterns are more affected by elevation.

As part of this thesis, we created a complete processing chain for the extraction of lake ice coverage and lake ice phenology based on RS imagery in the European Alps. Since the entire workflow is implemented as part of the Google Earth Engine it allows us to further expand the retrieved information with future imagery. In addition, the processing chain could be adjusted for near real-time production without the need of much additional work. With our results and findings we were able to answer all the initially posed research questions. Furthermore, we could show that a

multi-sensor approach can give a good estimation of lake ice coverage and phenology that fulfills the GCOS requirements of daily temporal resolution. The reached accuracy in LIP estimation shows that we likely did not fulfill the ice coverage uncertainty requirements of  $\pm 10\%$ . Due to the low validation scores and generalization performance reached by S-3 imagery and the observed problems with geometric accuracy, we advise against the use of S-3 imagery for lake ice detection for small lakes ( $< 10 \text{ km}^2$ ).

With the limited training data collected for this study, we were only able to cover a very limited extent of the study area. Thus, we were not able to make an extensive generalization assessment that proves a good model performance over the entire study area. We argue that more training data is necessary to create classifiers for an extensive and operational monitoring system. If possible, the used training data should also include information from the partial states of ice coverage. Therefore we would like to present our work mainly as a proof of concept. We are excited and optimistic for future implementations using more contemporary approaches such as Convolutional Neural Networks. Finally, we encourage the reader to inspect the obtained results at [licmonitoring.com](http://licmonitoring.com).

## Bibliography

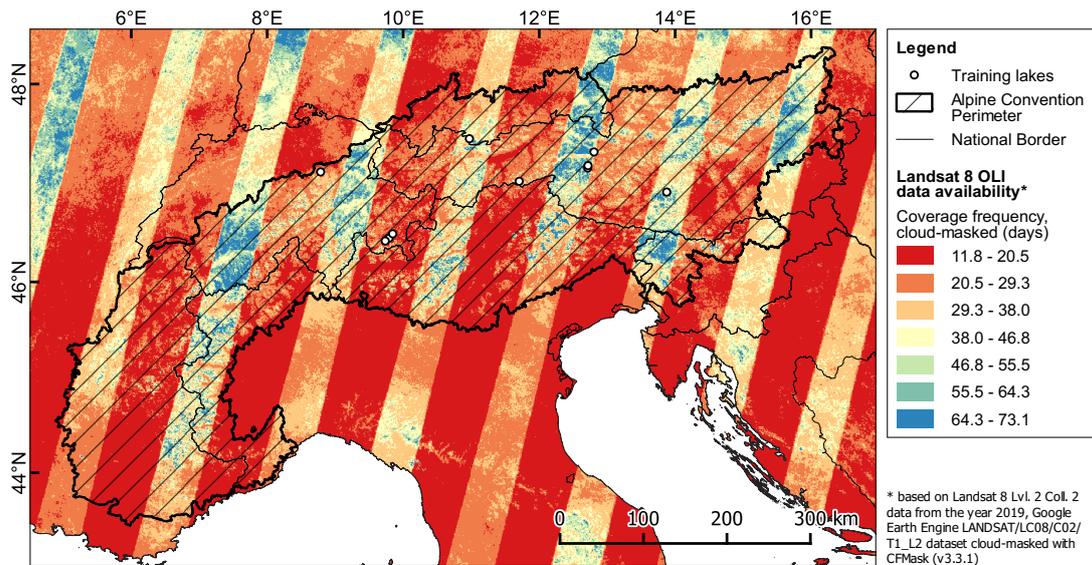
- Atwood, D. K., G. E. Gunn, C. Roussi, J. Wu, C. Duguay, and K. Sarabandi (Nov. 2015). "Microwave Backscatter from Arctic Lake Ice and Polarimetric Implications." *IEEE Transactions on Geoscience and Remote Sensing* 53.11, pp. 5972–5982.
- Berberoğlu, S., A. Akin, P. M. Atkinson, and P. J. Curran (2010). "Utilizing image texture to detect land-cover change in Mediterranean coastal wetlands." *International Journal of Remote Sensing* 31.11, pp. 2793–2815.
- Breiman, L. (2001). *Random Forests*. Tech. rep., pp. 5–32.
- Brown, L. C. and C. R. Duguay (2010). "The response and role of ice cover in lake-climate interactions." *Progress in Physical Geography* 34.5, pp. 671–704.
- Chen, C., H. Bagan, X. Xie, Y. La, and Y. Yamagata (May 2021). "Combination of sentinel-2 and palsar-2 for local climate zone classification: A case study of nanchang, China." *Remote Sensing* 13.10.
- Chen, R. C., C. Dewi, S. W. Huang, and R. E. Caraka (Dec. 2020). "Selecting critical features for data classification based on machine learning methods." *Journal of Big Data* 7.1.
- Connors, R. W., M. M. Trivedi, and C. A. Harlow (Mar. 1984). "Segmentation of a high-resolution urban scene using texture operators." *Computer Vision, Graphics, and Image Processing* 25.3, pp. 273–310.
- Dozier, J. (Apr. 1989). "Spectral signature of alpine snow cover from the landsat thematic mapper." *Remote Sensing of Environment* 28, pp. 9–22.
- Ezimand, K., A. A. Kakroodi, and M. Kiavarz (Dec. 2018). "The development of spectral indices for detecting built-up land areas and their relationship with land-surface temperature." *International Journal of Remote Sensing* 39.23, pp. 8428–8449.
- Fernández-Delgado, M., E. Cernadas, S. Barro, D. Amorim, and A. Fernández-Delgado (2014). *Do we Need Hundreds of Classifiers to Solve Real World Classification Problems?* Tech. rep., pp. 3133–3181.
- Foga, S., P. L. Scaramuzza, S. Guo, Z. Zhu, R. D. Dilley, T. Beckmann, G. L. Schmidt, J. L. Dwyer, M. Joseph Hughes, and B. Laue (June 2017). "Cloud detection algorithm comparison and validation for operational Landsat data products." *Remote Sensing of Environment* 194, pp. 379–390.
- Gascoin, S., Z. B. Dumont, C. Deschamps-Berger, F. Marti, G. Salgues, J. I. López-Moreno, J. Revuelto, T. Michon, P. Schattan, and O. Hagolle (Sept. 2020). "Estimating fractional snow cover in open terrain from Sentinel-2 using the normalized difference snow index." *Remote Sensing* 12.18.

- GCOS (2016). *The Global Observing System for Climate: Implementation Needs GCOS-200 (GOOS-214)*. Tech. rep. Global Climate Observation System.
- Gorelick, N., M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore (Dec. 2017). "Google Earth Engine: Planetary-scale geospatial analysis for everyone." *Remote Sensing of Environment* 202.
- Haralick, R. M., I. Dinstein, and K. Shanmugam (1973). "Textural Features for Image Classification." *IEEE Transactions on Systems, Man and Cybernetics* SMC-3.6, pp. 610–621.
- Hoekstra, M., M. Jiang, D. A. Clausi, and C. Duguay (May 2020). "Lake ice-water classification of RADARSAT-2 images by integrating IRGS Segmentation with pixel-based random forest labeling." *Remote Sensing* 12.9.
- Knoll, L. B., S. Sharma, B. A. Denfeld, G. Flaim, Y. Hori, J. J. Magnuson, D. Straile, and G. A. Weyhenmeyer (Oct. 2019). "Consequences of lake and river ice loss on cultural ecosystem services." *Limnology and Oceanography Letters* 4.5, pp. 119–131.
- Koskinen, J., J. Pulliainen, and M. Hallikainen (2000). "Effect of snow wetness to C-band backscatter - a modeling approach." *International Geoscience and Remote Sensing Symposium (IGARSS)*. Vol. 4. IEEE, pp. 1754–1756.
- Landis, J. R. and G. G. Koch (Mar. 1977). "The Measurement of Observer Agreement for Categorical Data." *Biometrics* 33.1, p. 159.
- Laner, P. (2018). *The Alps in 25 maps*. Tech. rep. Permanent Secretariat of the Alpine Convention.
- Latifovic, R. and D. Pouliot (Feb. 2007). "Analysis of climate change impacts on lake ice phenology in Canada using the historical satellite data record." *Remote Sensing of Environment* 106.4, pp. 492–507.
- McFeeters, S (1996). "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." *International Journal of Remote Sensing* 17.7, pp. 1425–1432.
- Messenger, M. L., B. Lehner, G. Grill, I. Nedeva, and O. Schmitt (Dec. 2016). "Estimating the volume and age of water stored in global lakes using a geo-statistical approach." *Nature Communications* 7.
- MeteoSwiss (2021). *MeteoSwiss: Documentation of MeteoSwiss Grid-Data Products: Daily Mean, Minimum and Maximum Temperature: TabsD, TminD, TmaxD*. URL: [https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-temperatur/doc/ProdDoc\\_TabsD.pdf](https://www.meteoswiss.admin.ch/content/dam/meteoswiss/de/service-und-publikationen/produkt/raeumliche-daten-temperatur/doc/ProdDoc_TabsD.pdf).
- Millard, K. and M. Richardson (2015). "On the importance of training data sample selection in Random Forest image classification: A case study in peatland ecosystem mapping." *Remote Sensing* 7.7, pp. 8489–8515.
- Mullissa, A., A. Vollrath, C. Odongo-Braun, B. Slagter, J. Balling, Y. Gou, N. Gorelick, and J. Reiche (May 2021). "Sentinel-1 SAR Backscatter Analysis Ready Data Preparation in Google Earth Engine." *Remote Sensing* 13.10, p. 1954.
- Murfitt, J. and C. R. Duguay (Feb. 2020). "Assessing the performance of methods for monitoring ice phenology of the world's largest high arctic lake using high-density time series analysis of Sentinel-1 data." *Remote Sensing* 12.3.

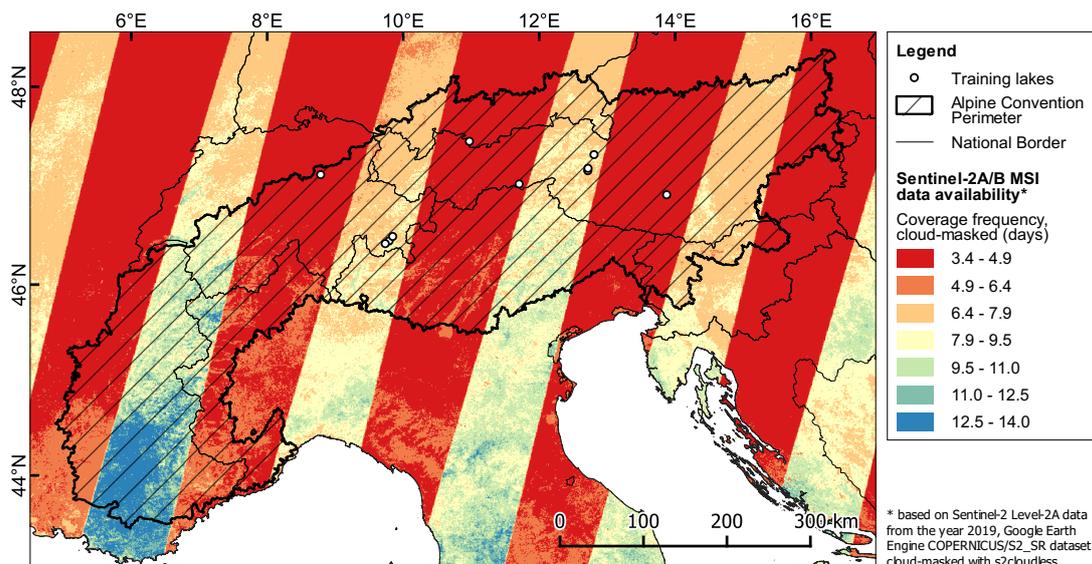
- Nabil, M., E. Farg, S. M. Arafat, M. Aboelghar, N. M. Afify, and M. M. Elsharkawy (Aug. 2022). "Tree-fruits crop type mapping from Sentinel-1 and Sentinel-2 data integration in Egypt's New Delta project." *Remote Sensing Applications: Society and Environment* 27, p. 100776.
- Nolan, M., G. Liston, P. Prokein, J. Brigham-Grette, V. L. Sharpton, and R. Huntzinger (Jan. 2003). "Analysis of lake ice dynamics and morphology on Lake El'gygytgyn, NE Siberia, using synthetic aperture radar (SAR) and Landsat." *Journal of Geophysical Research: Atmospheres* 108.2.
- Pekel, J.-F., A. Cottam, N. Gorelick, and A. S. Belward (Dec. 2016). "High-resolution mapping of global surface water and its long-term changes." *Nature* 540.7633, pp. 418–422.
- Rikimaru, A, P Roy, and S Miyatake (2002). "Tropical forest cover density mapping." *Tropical Ecology* 41.1, pp. 39–47.
- Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan (2012). "Random Forest classification of Mediterranean land cover using multi-seasonal imagery and multi-seasonal texture." *Remote Sensing of Environment* 121, pp. 93–107.
- Rouse, J, R Haas, J Schell, and D Deering (1973). *Monitoring vegetation systems in the Great Plains with ERTS*. Tech. rep. 2.
- Sharma, S., K. Blagrove, J. J. Magnuson, C. M. O'Reilly, S. Oliver, R. D. Batt, M. R. Magee, D. Straile, G. A. Weyhenmeyer, L. Winslow, and R. I. Woolway (Mar. 2019). "Widespread loss of lake ice around the Northern Hemisphere in a warming world." *Nature Climate Change* 9.3, pp. 227–231.
- Sharma, S., M. F. Meyer, J. Culpepper, X. Yang, S. Hampton, S. A. Berger, M. R. Brousil, S. C. Fradkin, S. N. Higgins, K. J. Jankowski, G. Kirillin, A. P. Smits, E. C. Whitaker, F. Yousef, and S. Zhang (Aug. 2020). "Integrating Perspectives to Understand Lake Ice Dynamics in a Changing World." *Journal of Geophysical Research: Biogeosciences* 125.8.
- Sharma, S., D. C. Richardson, R. I. Woolway, M. A. Imrit, D. Bouffard, K. Blagrove, J. Daly, A. Filazzola, N. Granin, J. Korhonen, J. Magnuson, W. Marszelewski, S. I. S. Matsuzaki, W. Perry, D. M. Robertson, L. G. Rudstam, G. A. Weyhenmeyer, and H. Yao (Oct. 2021). "Loss of Ice Cover, Shifting Phenology, and More Extreme Events in Northern Hemisphere Lakes." *Journal of Geophysical Research: Biogeosciences* 126.10.
- Skakun, S., J. Wevers, C. Brockmann, G. Doxani, M. Aleksandrov, M. Batič, D. Frantz, F. Gascon, L. Gómez-Chova, O. Hagolle, D. López-Puigdollers, J. Louis, M. Lubej, G. Mateo-García, J. Osman, D. Peressutti, B. Pflug, J. Puc, R. Richter, J.-C. Roger, P. Scaramuzza, E. Vermote, N. Vesel, A. Zupanc, and L. Žust (June 2022). "Cloud Mask Intercomparison eXercise (CMIX): An evaluation of cloud masking algorithms for Landsat 8 and Sentinel-2." *Remote Sensing of Environment* 274, p. 112990.
- Sobiech, J. and W. Dierking (July 2013). "Observing lake- and river-ice decay with SAR: advantages and limitations of the unsupervised k-means classification approach." *Annals of Glaciology* 54.62, pp. 65–72.
- Tom, M., R. Aguilar, P. Imhof, S. Leinss, E. Baltsavias, and K. Schindler (Feb. 2020). "Lake Ice Detection from Sentinel-1 SAR with Deep Learning."
- Tom, M., M. Suetterlin, D. Bouffard, M. Rothermel, S. Wunderle, and E. Baltsavias (Jan. 2019). *Project Integrated monitoring of ice in selected Swiss lakes - Final Report*. Tech. rep. ETH Zürich, EAWAG, University of Bern.

- Walsh, S. E., S. J. Vavrus, J. A. Foley, V. A. Fisher, R. H. Wynne, and J. D. Lenters (Nov. 1998). "Global patterns of lake ice phenology and climate: Model simulations and observations." *Journal of Geophysical Research Atmospheres* 103.D22, pp. 28825–28837.
- Xu, H. (2006). "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery." *International Journal of Remote Sensing* 27.14.
- Yang, X., T. M. Pavelsky, L. P. Bendezu, and S. Zhang (2021). "Simple Method to Extract Lake Ice Condition From Landsat Images." *IEEE Transactions on Geoscience and Remote Sensing*.
- Zhang, X., K. Wang, and G. Kirillin (July 2021). "An automatic method to detect lake ice phenology using MODIS daily temperature imagery." *Remote Sensing* 13.14.

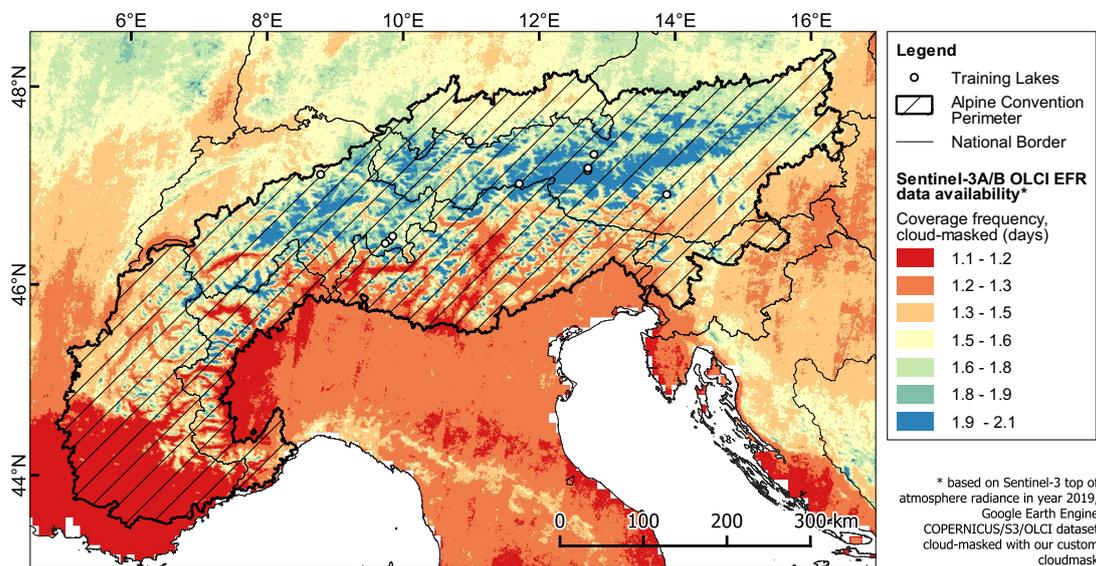
## A Coverage frequency figures



**Figure A.1:** Map of the study area showing the spatial distribution of the coverage frequency of Landsat-8 imagery after cloud-masking in the year 2019

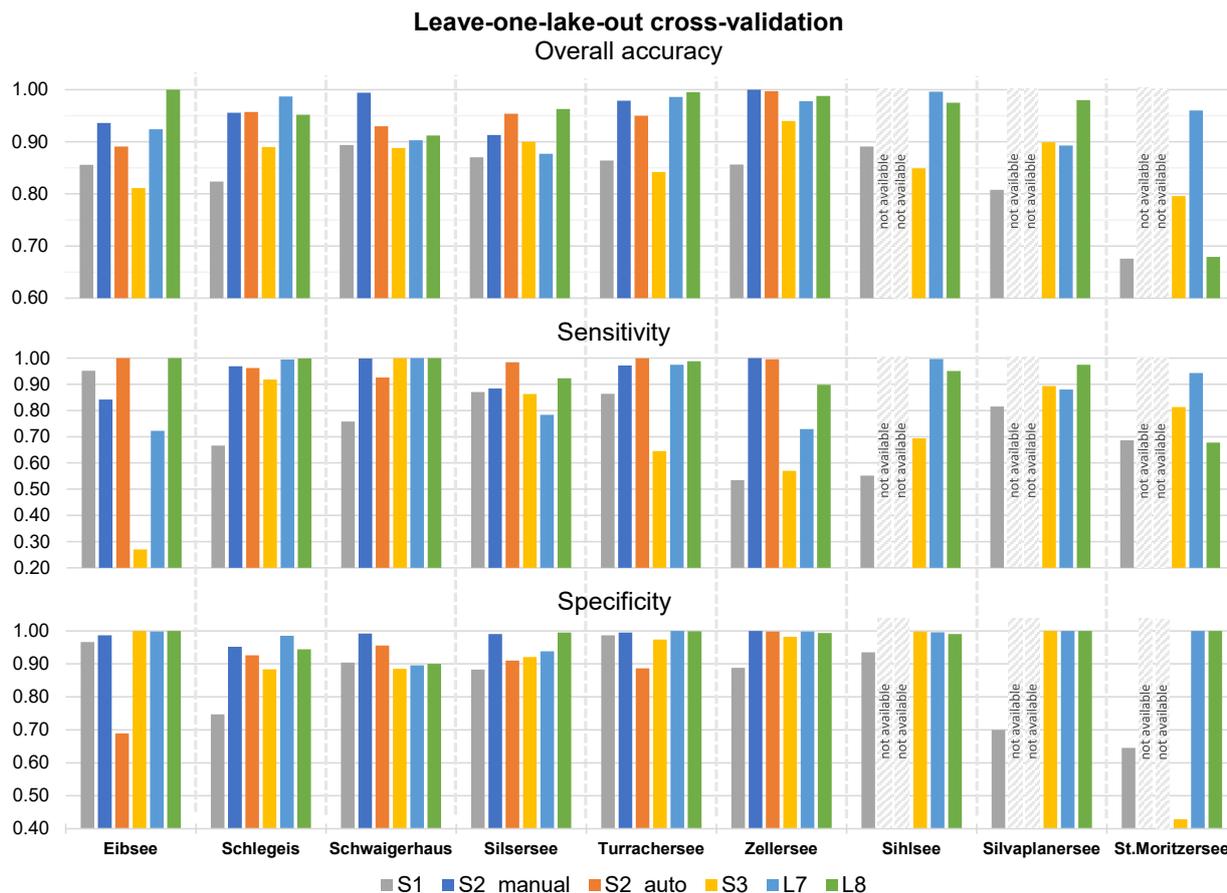


**Figure A.2:** Map of the study area showing the spatial distribution of the coverage frequency of Sentinel-2A/B imagery after cloud-masking in the year 2019



**Figure A.3:** Map of the study area showing the spatial distribution of the coverage frequency of Sentinel-3A/B imagery after cloud-masking in the year 2019

## B Leave-one-lake-out cross-validation



**Figure B.1:** Validation results from the leave-one-lake-out cross-validation (LOLO-CV). Sentinel-2 acquisitions do not overlap with the temporal range of the ground-truth labels of Sihlsee, Silvaplanersee and St.Moritzersee repurposed from the work of (Tom et al., 2019) and are therefore not available.

## Acknowledgements

I would like to thank my two thesis advisors **Dr. Hendrik Wulf** of the Department of Geography at the University of Zurich and **Dr. Daniel Odermatt** from the Swiss Federal Institute of Aquatic Science and Technology (Eawag). The open and constructive exchange and their expertise, interest and motivation helped me greatly with the work on this thesis. Furthermore, I also thank **Florian Radlherr** from the foto-webcam.eu team, that allowed me quick and straightforward access to all of their high-resolution webcam imagery. And lastly, I like to thank **Prof. Dr. Peter Molnar** of the Institute of Environmental Engineering at ETH Zurich for giving me access to the MeteoSwiss grid-data products. This thesis would not have been possible without their support. Thank you!

Stay healthy and safe.

Michael Brechbühler

## Personal Declaration

I hereby declare that the submitted thesis is the result of my own, independent work. All external sources are explicitly acknowledged in the thesis.

Date

July 31, 2022

Signature

Michael Brechbühler

A handwritten signature in black ink, appearing to read "Michael Brechbühler", written over a horizontal line.