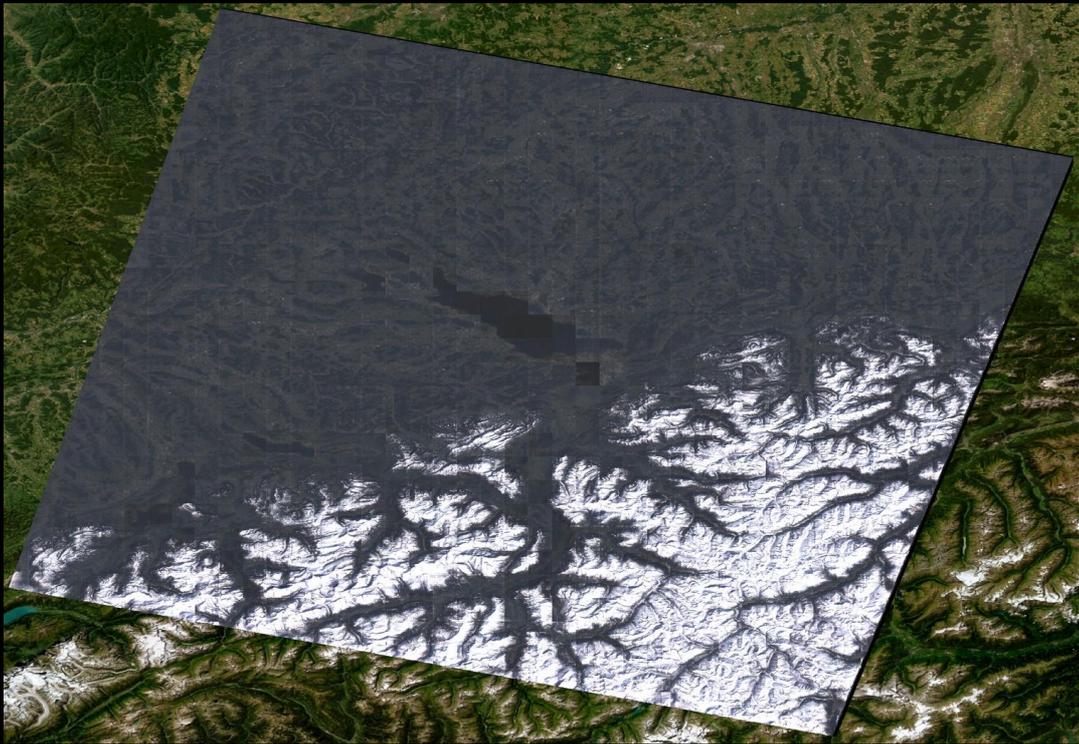


Adriaan Keurhorst

cGANs for multispectral snow analysis in the Alps



MSc Geoscience and Remote Sensing Thesis

TU Delft

May 30, 2023

Abstract

The Alps are experiencing a gradual reduction in snow cover due to rising temperatures, impacting the landscape and dependent ecosystems. While several models have been developed to study snow cover in the region, there is a lack of visual representations. This research employs a Conditional Generative Adversarial Network (cGAN) to generate a multispectral Landsat-8 image of the Alps using environmental data inputs. The study utilizes elevation, monthly precipitation, and monthly temperature data from November to March to produce an end-of-winter Landsat-8 image. The resulting multispectral image is then used to calculate the Normalised Difference Snow Index (NDSI) and determine the snow extent by counting pixels with NDSI values above 0.4. Two climate scenarios are considered, and the generated images are compared to actual Landsat-8 imagery. Findings indicate that while the generated imagery closely resembles the real imagery, the snow extent is generally underestimated in the current model configuration, and the snow reflectance is consistently overestimated across all training steps of the cGAN. Additionally, it is observed that increasing the spatial distance between the training and testing locations leads to increased error in the results. The thesis demonstrates the feasibility of using a cGAN to generate snow extent, but suggests that enhancements to the training dataset and cGAN architecture are necessary for improved accuracy. By leveraging the cGAN, future climate scenarios can be visually represented through multispectral imagery, enabling a more detailed understanding of potential future landscapes under different climate conditions.

Preface

The way we see winter has changed over the course of the last few decades. We are experiencing less and less snowfall each year, most years not experiencing any. This has become the norm in many countries, like The Netherlands, where a white Christmas is almost unheard of. This is why I chose this topic, but focussing on the Alps because they are once of the last strongholds of the cold climate in western Europe. I really think this thesis has a future in generating multispectral imagery to determine snow extents, given some changes to the input datasets. I had a good time writing this thesis because it really gave me ambition to continue in deep learning and remote sensing. It has taught me new, deeper ways of thinking and learning.

Thank you Stef and Sophie for the very helpful advice over the last eight months, while allowing me to continue experimenting and finding out new things on my own.

For the rest, I'd like to thank Anne, Karel, Alex, Noortje, Cecilia, Pepijn, Caesar and Simon for always listening to my dilemmas on how to continue. I appreciate all the support.

Enjoy the read!

Contents

Abstract.....	i
Preface	ii
1. Introduction	1
1.1. Overview	1
1.2. Conditional Generative Adversarial Networks	1
1.3. Aim of this study.....	2
1.4. Related works	3
2. Data.....	4
2.1. Study area	4
2.2. Landsat-8.....	4
2.3. Environmental data.....	5
2.4. Experiments.....	6
2.4.1. Experiment 1 – Similar training and testing locations.....	6
2.4.2. Experiment 2 – Different training and testing locations	10
3. Methodology.....	14
3.1. Creating the training and testing datasets.....	14
3.2. Training the cGAN	15
3.3. Creating the testing image	16
3.4. Evaluating the cGAN’s results	16
3.4.1. Multispectral image analysis.....	16
3.4.2. Generated snow extent.....	17
3.4.3. Generated spectral signature of snow.....	17
4. Results	19
4.1. Generated Landsat-8 images.....	19
4.1.1. Comparison of generated and true Landsat-8 imagery.....	19
4.1.2. cGAN performance per band.....	21
4.1.3. Interpretation of generated imagery results.....	24
4.2. Snow extent.....	24
4.2.1. Comparison of generated and true snow extent.....	24
4.2.2. Relationship snow extent and environmental variables	26
4.2.3. Effect of cGAN training steps on snow extent.....	27
4.2.4. Snow extent dependency on input variables	28
4.2.5. Snow extent dependency on changes in input variables	29
4.2.6. Interpretation of snow extent results.....	30
4.3. Spectral signatures.....	32
4.3.1. Spectral curve of snow	32

4.3.2.	Effect of cGAN training steps on generated snow reflectance.....	34
4.3.3.	Interpretation of spectral signature results	35
5.	Discussion.....	36
5.1.	Location transferability in training and testing.....	36
5.2.	Improving the training set	36
5.2.1.	Using higher resolution input datasets.....	36
5.2.2.	Including more variables	36
5.2.3.	Using a more accurate 'end of winter' target image.....	37
5.2.4.	Data augmentation.....	37
5.2.5.	Increasing volume of training set	37
5.3.	Improvements to the cGAN architecture.....	37
5.3.1.	Number of training steps	37
5.3.2.	Assign weight to individual months	38
5.3.3.	U-Net generator vs. ResNet generator.....	38
5.3.4.	Adaptive learning rate.....	38
5.4.	Future research	38
6.	Conclusion and outlook.....	40
7.	Appendix 1 – Conditional Generative Adversarial Network	41
8.	Appendix 2 – cGAN model summary	43
8.1.	Generator.....	43
8.2.	Discriminator	44
9.	Appendix 3 – Generated Landsat-8 images Experiment 1 Run 2018.....	46
10.	Appendix 4 – Precipitation per month in Alpine training set	48
11.	Appendix 5 – Generated reflectance when training on both Alpine and Himalayan data	49
12.	References.....	50

1. Introduction

1.1. Overview

Understanding snow characteristics and extent is vital for our understanding of the earth radiative system, and can be facilitated through remote sensing technology. Snow cover directly decreases the near-surface air temperature, increases the surface's albedo and acts as an insulator for the soil, due to its low thermal conductivity (Namias, 1985; Zhang, 2005). Snow also influences the regional fresh-water availability, river-runoff and groundwater recharge. In countries like Norway and Switzerland, snowfall plays an important role in the generation of hydroelectrical power (Beniston, 2012). With the global climate warming, the response of snow cover and snowfall is changing. Regions in China are experiencing increases in snowfall (Li et al., 2008), whereas the Alps have shown a decline in total snowfall and time on the ground over the past 80 years.

By using remote sensing technology, snowfall can be monitored over large areas with a high degree of precision, while overcoming the difficulties associated with using observing stations (Nolin, 2010; Foster et al., 1987). Remote sensing overcomes the issues relating to uneven spatial density of observing stations and the differing observation quality from these stations. It also allows large areas to be observed at once. There are several different methods to monitor snowfall using remote sensing technology. The most popular methods to monitor snow characteristics is by using either optical satellite imagery or radar data, as highlighted by studies done by Dietz et al. (2012) and Nolin et al. (2010). Optical and radar remote sensing technology can be used to determine snow cover extent, snow grain size, snow albedo, snow water equivalent, melting snow and snow depth.

Snow shows very distinct properties across the electromagnetic spectrum, making it a distinguishable surface property. Snow reflects very strongly in the visible spectrum, reflecting nearly all radiation. However, its reflectance starts to vary greatly as the wavelength increases. Pristine snow reflects up to 85% of incoming sunlight, with the remainder of the incoming solar radiation stored as heat (Zender, 2012). As snow gets older, it is affected by pollution, starting a greying process. As the snow gets darker in colour, the albedo decreases causing the snow to trap more heat in the process. Grain size also affects snow's reflectance. Smaller grains of snow scatter more radiation, and absorb less than larger grains of snow (Green, 2002). The dryness of the snow also plays a part in its reflectance, as dryer snow reflects more radiation than wetter snow. The thickness of snow also plays a part in its reflectance, as reflectance increases with increasing snow depth. These physical properties of snow are governed by climatic variables, mainly temperature and precipitation. As a result of this, the physical relationship between these climatic variables and snow reflectance can be observed with optical satellite imagery (Yang, 2013).

1.2. Conditional Generative Adversarial Networks

Recent advances in deep learning have shown great potential with remote sensing applications. Very large datasets can be processed by deep learning algorithms and applied with high precision (LeCun, 2015). Deep convolutional networks have been shown to be very powerful in image classification (Shinde and Shah, 2018). Other deep learning networks have provided breakthroughs in improving land-cover mapping, deriving vegetation parameters, deriving land surface temperatures and many more applications (Yuan, 2020). There are also deep learning methods which can generate imagery. This generated imagery can range from regular RGB pictures to multispectral satellite imagery. Generating imagery can be done through Conditional Generative Adversarial Networks, and have already been applied to a range of remote sensing applications.

A Conditional Generative Adversarial Network (cGAN) is a deep learning algorithm that can be used to generate “fake” data. cGANs are related to Generative Adversarial Nets (GANs) (Mirza & Osindero, 2014). A cGAN consists of two ‘adversarial’ models: a Generative model (G) and a Discriminative model (D). G generates data, whereas D attempts to estimate whether or not a sample came from real training data or from G. These models are trained simultaneously. The difference between a normal GAN and a conditional GAN, is that the latter can be conditioned on extra information. cGANs have shown very good results in image generation. A cGAN enables control over the output, allowing the influence of the input variables on the output image can be studied. A full explanation of how a cGAN works can be found in Appendix 1.

1.3. Aim of this study

This study is designed to determine how well a cGAN can be used to generate Landsat-8 imagery to observe the end of winter snow extent in the Alps. A cGAN is appropriate for generating multispectral imagery, from which the snow extent can be calculated. The generated image needs to be multispectral so that the reflectance values can be analysed. Reflectance values of snow provide information on the snow characteristics.

Currently, this study is focussed at generating the end of winter snow extent in the Alps for past scenarios. Given this study is successful, the cGAN can be able to be applied to generate satellite imagery to determine the end of winter snow extent, when there isn’t any imagery available. This can be applied in different scenarios. The results from this study can be used when cloud-free multispectral imagery needs to be generated over the Alps to determine the end of winter snow extent. More importantly though, the results can also be used for future climate scenarios, as defined by ScenarioMIP. An overview of ScenarioMIP can be found in O’Neill et al. (2016). These future scenarios have gridded yearly climate projections, depending on different climate scenarios. These gridded climate projections can serve as the input dataset needed for a cGAN to generate a multispectral image, to determine the end of winter snow extent in the Alps for future climate scenarios.

Important sub-questions to be answered are the following:

1. How do the true Landsat-8 reflectance values for snow compare to the generated Landsat-8 reflectance values for snow?
2. What is the importance of the number of cGAN training steps for its ability to predict snow cover in the Alps?
3. How is the performance of the cGAN affected in generating multispectral Landsat-8 imagery to determine snow extent when the training location is different that the testing location?

The first sub-question is important because the reflectance values of snow give indications on the snow properties; such as grain size, dryness, thickness and age. If the generated reflectance values of the cGAN are similar to that of the true Landsat-8 image, these snow properties can be derived from the generated image. The second sub-question is important to answer as it will provide insight to what extent the cGAN needs to be trained in order to generate a multispectral Landsat-8 image with the correct snow extent. The third sub-question is important to answer, because it will provide insights into how a cGAN should be trained when trying to estimate snowfall extent over a region. Does the location of the training data need to be similar to the testing data for determining the end of winter snow extent? Does the geography of the training data need to be similar to the testing data?

These questions are important to be answered to get a more comprehensive understanding of how cGANs can be used to predict snow extent through multispectral imagery.

1.4. Related works

A study done by Isola et al. (2017) describes image-to-image translation using a cGAN, named *pix2pix*. This outlines how a cGAN can be used to translate an input image to an output image. An example of this is generating an RGB image from label maps, or translating a satellite image of a city to a 2-dimensional map. The *pix2pix* paper is the core building block of this study, as a similar cGAN architecture is used. A detailed description of this architecture is provided in section 3.2.

Multispectral imagery has been produced through cGANs in multiple works. Requena-Mesa et al. (2019) demonstrated the cGAN's potential in generating landscapes from environmental conditions. The dataset used in this study was very large, with 10% of the earth's land surface used as the training dataset. This study showed that predicting landscapes can be done by using cGANs, highlighting that a landscape can be predicted while using environmental variables as input. They generated a multispectral image with 4 bands from a total of 32 input environmental variables. The landscapes predicted by Requena-Mesa et al. (2019) were natural landscapes, predicted using climatological data, lithological data and anthropogenic interventions. One limitation of using cGANs to predict landscapes according to this study, is that the predictive quality of a cGAN decreases as the testing location moves further away from the training location. For example, training the cGAN on Europe and then testing the cGAN on an area in North America, lead to less accurate results than testing the cGAN on an area it had already seen in training. They also noted that increasing the complexity of a cGAN only marginally improves performance, and makes the cGAN more prone to overfitting.

Rodríguez-Suárez et al. (2022) studied the effect of changing the cGAN's architecture to construct multispectral imagery from RGB imagery. They experimented with a convolutional neural network, U-Net, and ResNet as a generator to produce multispectral imagery. Their study concluded that using a ResNet is best as a generator in their application, followed by the U-net, and then a convolutional neural network. The study by Rodríguez-Suárez et al. (2022) also concluded that the testing area should already have been seen during training in order to be reconstructed accurately.

Little work has been done for full reconstruction of satellite imagery by using a cGAN. Prior to this study, no study has been found which uses a cGAN for multispectral snow analysis. This thesis looks to bridge this gap, so that multispectral snow analysis of future climate scenarios will become a possibility.

2. Data

To answer the research question, environmental data and Landsat-8 data are used. The environmental data consists of elevation data, total precipitation, minimum temperature and maximum temperature. An overview of the study area, Landsat-8 data and environmental datasets is given below.

2.1. Study area

The study area is the Alps. This is an area with consistent snowfall during winter. It has strong elevation differences across relatively small areas, ranging from around 520 metres above sea level to 4200 metres above sea level. The temperature and precipitation values are also very variable across this area (Gobiet et al., 2014). This area was chosen to determine a cGAN's ability to predict the end of winter snow extent because of these highly variable environmental factors. As a result of this high variability, the cGAN will have a greater potential to learn relationships between snow cover and the environmental variables. In addition to the high variability in environmental variables, the Alps are also a relevant study area for a large population. It is one of the most densely populated mountain areas in the world (Agrawala, 2007). This makes it an important area to study and understand.

2.2. Landsat-8

Landsat-8 was launched on February 11th, 2013. It has a temporal resolution of 8 days, and a minimum spatial resolution of 30 metres for its spectral bands. Each Landsat-8 tile is 185x185km. There are 9 spectral bands in total, and 2 thermal infrared sensors. The first 7 bands are used in this study. All bands are shown in *Table 1*.

Table 1 - Landsat-8 bands (<https://www.usgs.gov/landsat-missions/landsat-8>).

Bands	Wavelength (micrometers)	Spatial resolution (meters)
Band 1 – coastal aerosol	0.43–0.45	30
Band 2 - blue	0.45–0.51	30
Band 3 - green	0.53–0.59	30
Band 4 - red	0.64–0.67	30
Band 5 - near infrared (NIR)	0.85–0.88	30
Band 6 - SWIR 1	1.57–1.65	30
Band 7 - SWIR 2	2.11–2.29	30
Band 8 - panchromatic	0.50–0.68	15
Band 9 - cirrus	1.36–1.38	30
Band 10 - Thermal Infrared 1	10.60–11.19	100
Band 11 - Thermal Infrared 2	11.50–12.51	100

Landsat-8 imagery can be accessed through Google Earth Engine. Google Earth Engine provides a Javascript interface through which remote sensing data can be accessed. Using this interface, Landsat-8 imagery can be sorted based on cloud cover, so that (mostly) cloud free Landsat-8 tiles can be downloaded. This image needs to be as cloud free as possible, as cloud cover will make it impossible to accurately determine the snow extent using Landsat-8 data.

Since Landsat-8 data imagery is multispectral, the Normalised-Difference Snow Index (NDSI) can be computed. The NDSI is the normalised difference between green (Band 3) and the short-wave infrared (Band 6) (Yang, 2015). The formula for the NDSI, using the Landsat-8 bands, is shown below (equation 1):

$$NDSI = \frac{B3 - B6}{B3 + B6} \quad (1)$$

NDSI has been used in multiple studies to quantify snow extent (Hall, 2015; Wunderle, 2016). These studies show that a region is mapped as snow-covered if the $NDSI \geq 0.4$, and so the same threshold will be used in this study.

2.3. Environmental data

The environmental data for this project consist of elevation, temperature and precipitation data. The elevation data is from the Consortium for Spatial Information, with a spatial resolution of 0.0010 degrees (equivalent to about 72.2 metres in the Alps). The minimum temperature, maximum temperature and precipitation data are from WorldClim v2, with a spatial resolution of 4638 metres. These datasets were chosen because of their relatively high spatial resolution and ease of access. This is important because of the relatively fine resolution of Landsat-8 data compared to the environmental data.

Some sample images of the elevation, precipitation, minimum and maximum temperature are shown below over an area over the Alps (*Figure 1*).

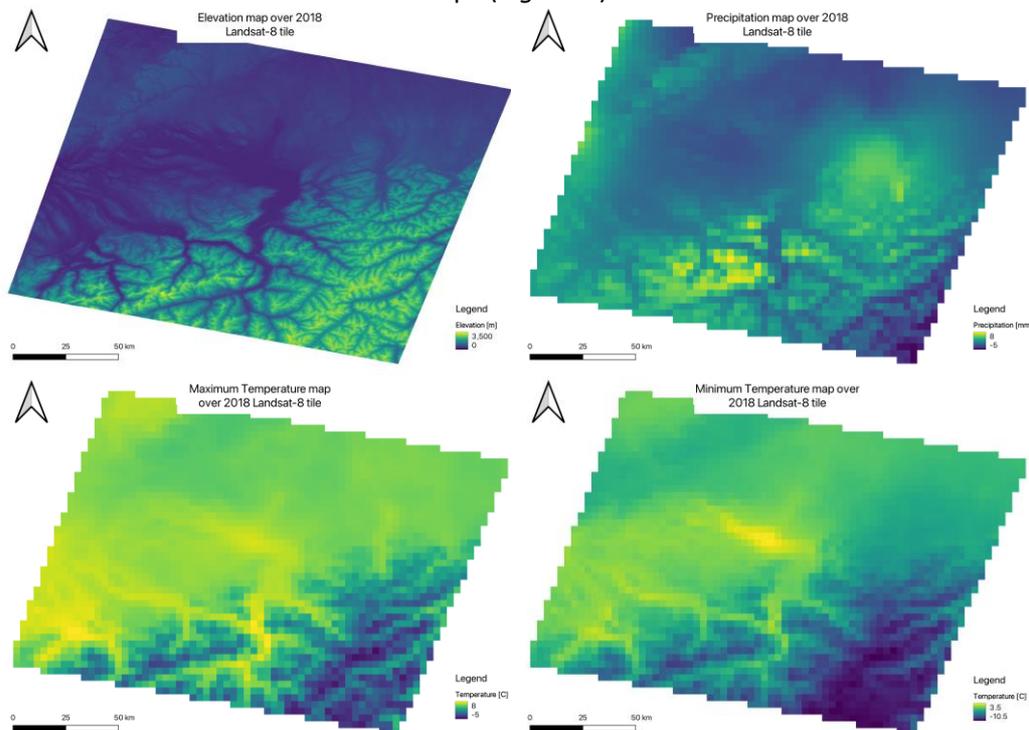


Figure 1 - Maps of environmental variables over an area in the Alps. Clockwise: elevation map, precipitation map, minimum temperature map and maximum temperature map. All shown environmental variables, apart from elevation which is constant, is for January 2018.

Figure 1 shows the different environmental variables in January over an area in the Alps. The elevation map clearly has the highest resolution, and the Alpine landscape is reflected well in the elevation differences. The same cannot be said for the precipitation and temperature datasets, although some spatial characteristics can be inferred.

Temperature, precipitation and elevation all influence snowfall (Zhang, 2019), and therefore also the snow extent at the end of winter. The snow extent at the end of winter is assumed to be the result of the environmental conditions in the winter months before that. The start of winter will be defined at November 1st, whereas the end of winter in this study will be defined as March 31st. Therefore, the monthly minimum temperature, maximum temperature and total precipitation from November 1st to March 31st determine the snow extent of a Landsat-8 image over the Alps on March 31st.

2.4. Experiments

Two experiments are designed to answer the research question, along with its sub-questions. Both experiments are designed to determine the Alpine snow extent at the end of winter. Experiment 1 uses Alpine data for its training and testing data. Experiment 2 uses Himalayan data for its training data but is tested on the Alps. This is to test the effect of moving the training area away from the testing area.

2.4.1. Experiment 1 – Similar training and testing locations

The first Experiment focusses on predicting the snow extent over the Alps using training and testing data from the Alps. The training and testing data for this Experiment are for years 2014 to 2018. First, a cloud-free Landsat-8 image over the Alps at the end of March needs to be selected. The location of these Landsat images is shown below in Figure 2.

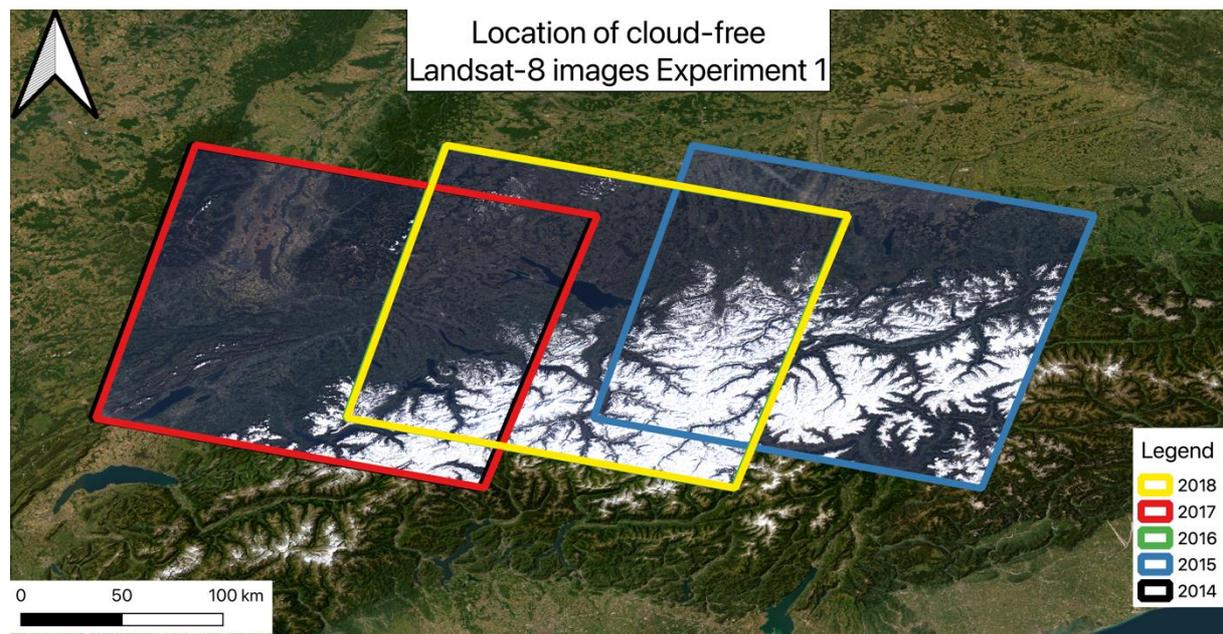
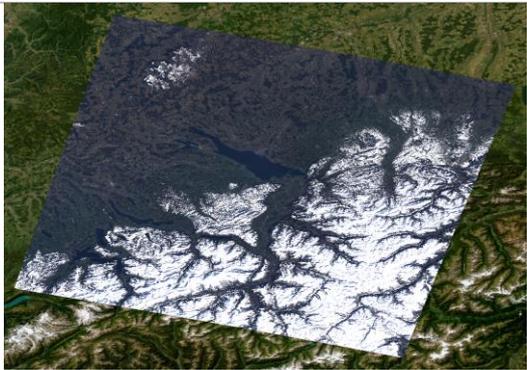


Figure 2 – Locations of the cloud-free Landsat-8 imagery for Experiment 1. Note: some of the tiles overlap and are therefore not visible in the figure.

Table 2 shows some metadata for each Landsat-8 image used in training and testing.

Table 2 - Landsat-8 training and testing metadata.

Landsat-8 metadata		
2014	Date acquired	20-03-2014
	Cloud cover [%]	2.01
	Scene ID	LC81950272014079LGN01
	RGB Image	
	Snow cover [km2]	4996
2015	Date acquired	10-04-2015
	Cloud cover	1.60
	Scene ID	LC81930272015100LGN01
	RGB Image	
	Snow cover [km2]	19843
2016	Date acquired	18-03-2016
	Cloud cover	1.03
	Scene ID	LC81940272016078LGN01

	RGB Image	
	Snow cover [km2]	18936
2017	Date acquired	28-03-2017
	Cloud cover	3.58
	Scene ID	LC81950272017087LGN00
	RGB Image	
	Snow cover [km2]	4212
2018	Date acquired	24-03-2018
	Cloud cover (%)	0.99
	Scene ID	LC81940272018083LGN00
	RGB Image	
	Snow cover [km2]	18228

The acquisition dates for the Landsat-8 imagery are not precisely at the end of March, due to cloud cover. The images were picked as close to the end of March as possible, while also taking cloud cover into consideration. Each image is over the Alps, and each image has snow cover.

Table 3 shows an overview with statistics for the environmental data of the respective Landsat-8 images, per year.

Table 3 – Alpine environmental data per year.

Year	Variable	Mean	Minimum	Maximum
2014	Max temperature (°C)	4,61	-6,82	9,90
	Min temperature (°C)	-0,40	-12,05	2,82
	Total precipitation (mm)	60,86	0,00	180,67
	Elevation (m)	521,85	0,00	4162,00
2015	Max temperature (°C)	2,45	-7,60	9,83
	Min temperature (°C)	-2,26	-12,63	1,76
	Total precipitation (mm)	43,74	0,00	136,93
	Elevation (m)	848,47	0,00	3716,00
2016	Max temperature (°C)	3,66	-3,95	9,44
	Min temperature (°C)	-1,32	-10,41	3,37
	Total precipitation (mm)	56,91	0,00	157,28
	Elevation (m)	723,38	0,00	3566,00
2017	Max temperature (°C)	3,90	-7,18	8,86
	Min temperature (°C)	-1,01	-12,24	2,56
	Total precipitation (mm)	45,64	0,00	137,33
	Elevation (m)	523,01	0,00	4162,00
2018	Max temperature (°C)	1,88	-6,69	6,90
	Min temperature (°C)	-2,60	12,31	1,93
	Total precipitation (mm)	70,63	0,00	200,40
	Elevation (m)	723,38	0,00	3566,00

Two runs are designed for Experiment 1, with each run testing on a different year. The first run uses training data from years 2014 to 2017, with 2018 as the testing image (referred to as Run 2018). The second run uses training data from 2014, 2015, 2017 and 2018 and is tested on 2016 (referred to as Run 2016). The difference between Run 2018 and Run 2016 is the values for the input training data. Year 2018 has the highest amount of precipitation, the lowest minimum temperature and the lowest maximum temperature of any year. Year 2016 lies perfectly in the middle in terms of mean precipitation, maximum temperature and minimum temperature. This will provide insight into how the cGAN's performance is influenced when presented with input variable values not seen before in training. This is why the two different runs have been selected.

Though the testing image changes between runs, the training image locations is always in the Alps. This means that the cGAN is trained and tested in the same area, which was an important factor in the performance of the cGAN in generating multispectral satellite imagery according to Requena-Mesa et al. (2019) and Rodríguez-Suárez et al. (2022).

2.4.2. Experiment 2 – Different training and testing locations

The second Experiment focusses on predicting the snow extent over the Alps, using training data from the Himalayas. The aim of including this Experiment in the study is to determine the importance of location in the training and testing datasets. The data used from the Himalayas are for the years 2014, 2015, 2017 and 2018. The number of years in the training set is the same as for the runs in Experiment 1, in order to remain consistent with the training set sizes. Below is an overview of the geographical locations of the training images of the Himalayas (*Figure 3*).

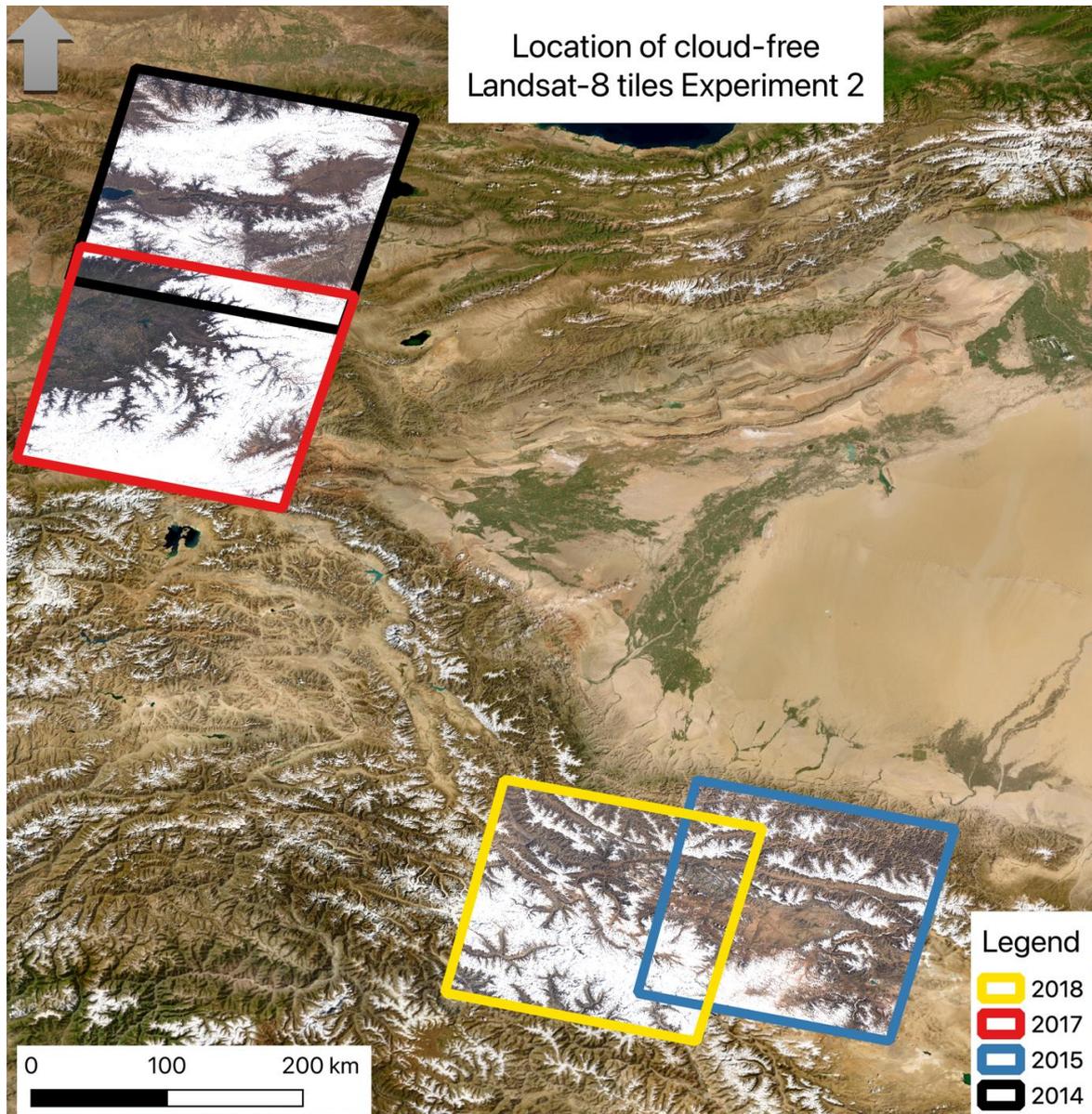
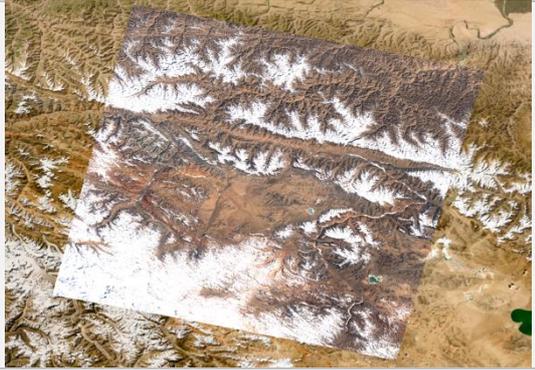
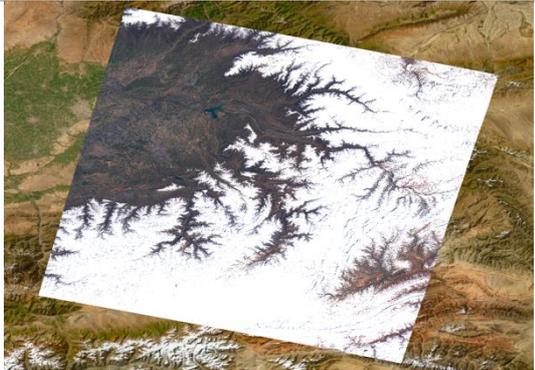


Figure 3 – Locations of the cloud-free Landsat-8 imagery for Experiment 2 (in the Himalayas). Note: some images overlap and are therefore not fully visible.

An overview of the metadata of the Landsat-8 imagery for Experiment 2 is shown below (*Table 4*):

Table 4 – Landsat-8 training and testing metadata Experiment 2

Landsat-8 metadata		
2014	Date acquired	01-04-2014
	Cloud cover	6.00
	Scene ID	LC81510312014091LGN01
	RGB Image	
	Snow cover [km2]	23839
2015	Date acquired	23-03-2015
	Cloud cover	2.14
	Scene ID	LC81470352015082LGN01
	RGB Image	
	Snow cover [km2]	14022
2017	Date acquired	09-04-2017
	Cloud cover	2.45
	Scene ID	LC81510322017099LGN00

	RGB Image	
	Snow cover [km2]	29430
2018	Date acquired	07-04-2018
	Cloud cover (%)	2.75
	Scene ID	LC81480352018097LGN00
	RGB Image	
	Snow cover [km2]	21274

Compared to the Alpine data from Experiment 1, the Himalayan imagery contains a slightly higher cloud cover. An interesting difference between the Himalayan Landsat-8 imagery and the Alpine Landsat-8 imagery is that the non-snow-covered areas have a different landcover. In the Alps, the non-snow surface is mainly covered by vegetation, whereas the non-snow surface in the Himalayas is mainly covered by bare soil. Another difference is that there is nearly twice as much snow cover in the Himalayas than in the Alps.

An overview of the Himalayan environmental data is provided in *Table 5*.

Table 5 – Himalayan environmental metadata per year.

Environmental metadata				
Year	Variable	Mean	Minimum	Maximum
2014	Max temperature (°C)	-3,95	-14,11	7,29
	Min temperature (°C)	-12,60	-26,58	0,44
	Total precipitation (mm)	21,31	0,00	66,44
	Elevation (m)	1718,42	0,00	4846,00
2015	Max temperature (°C)	-5,97	-19,78	5,05
	Min temperature (°C)	-15,09	-30,79	0,00
	Total precipitation	0,32	0,00	3,27
	Elevation	3347,38	0,00	7352,00
2017	Max temperature (°C)	-1,99	17,36	7,84
	Min temperature (°C)	-9,87	-28,18	0,60
	Total precipitation	29,86	0,00	106,12
	Elevation	1794,00	0,00	6557,00
2018	Max temperature (°C)	-5,89	-20,26	10,19
	Min temperature (°C)	-14,99	-31,69	0,40
	Total precipitation (mm)	2,06	0,00	13,60
	Elevation (m)	3395,74	0,00	8569,00

The environmental data for the Himalayas differs greatly than that from the Alps. The Himalayan temperatures are much lower, the total precipitation is much lower and the elevation is much higher. It will therefore be interesting to study the outcome of using this data to create a Landsat-8 image over the Alps.

3. Methodology

3.1. Creating the training and testing datasets

Training and testing a cGAN requires sufficient training data. To train a cGAN, input imagery and target imagery is needed, which in this study is environmental data and Landsat-8 data, respectively. *Figure 4* shows how the input and target datasets are designed:

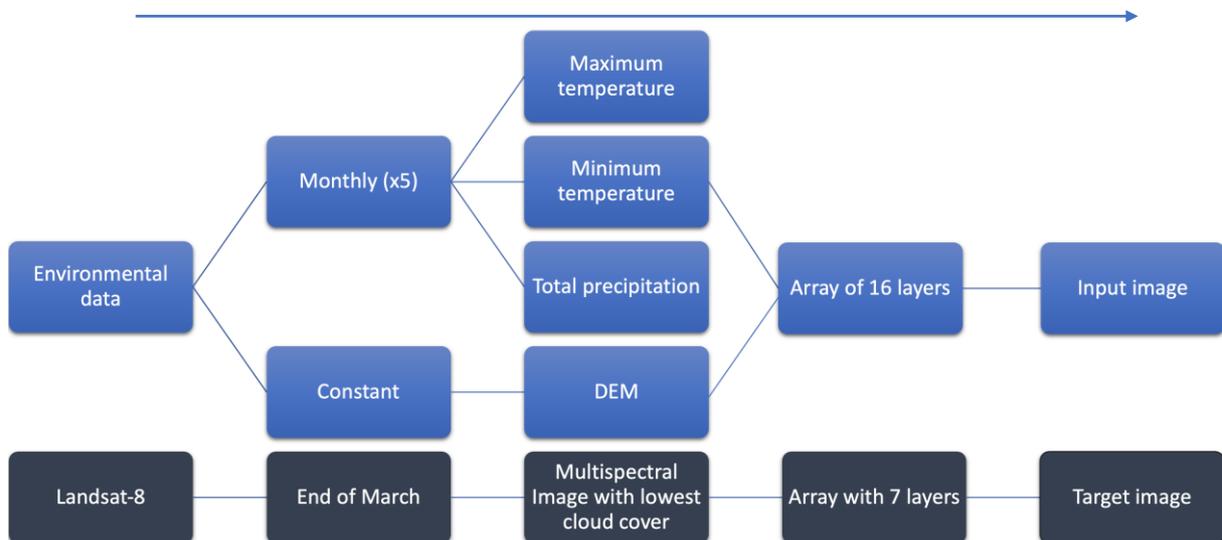


Figure 4 – Input and target datasets for any year. The environmental layer consists of a DEM, and monthly maximum temperature, minimum temperature and precipitation data. This array has 16 layers, corresponding to the number of environmental variables over the period of 5 months. The target image array consists of a multispectral cloud-free Landsat-8 image at the end of March. The number of layers in the target image array (7) corresponds to the number of bands of a Landsat-8 image.

Figure 5 shows how the training and testing datasets are created, for Experiment 1 Run 2018.

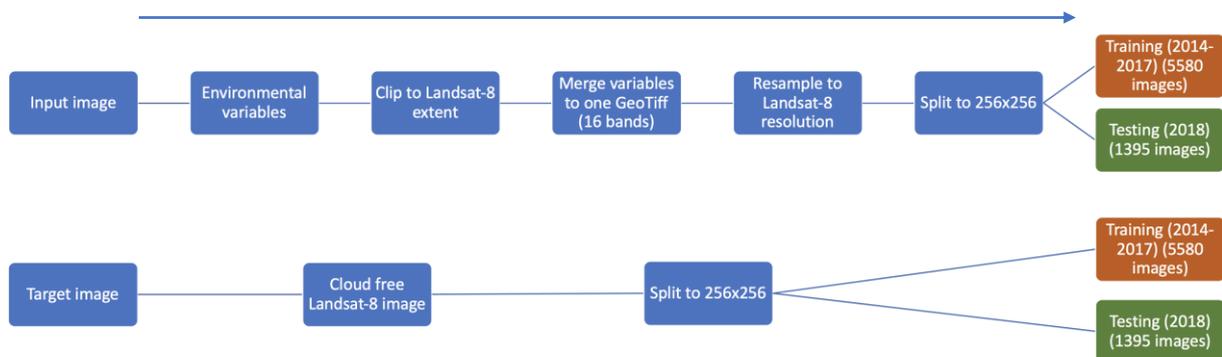


Figure 5 – Workflow for pre-processing the training and testing datasets for Experiment 1 Run 2018. The input image consists of the environmental variables, which are clipped to the Landsat-8 image extent of the same year. These environmental variables are merged into one file and resampled to the Landsat-8 resolution. After this, each year’s environmental image with 16 bands is split into 1395 images of 256x256 pixels. The target image dataset are the cloud free Landsat-8 images, which are also split into 1395 images of 256x256 pixels per year.

The training and testing datasets are pre-processed according to *Figure 5*. The workflow shown in *Figure 5* is for Experiment 1 Run 2018. *Table 6* shows an overview of the training and testing years per Experiment and Run.

Table 6 – Years included in the training and testing datasets, per Experiment and Run. The "Run" refers to the testing year.

Experiment	Run	Training	Testing
1	2018	2014, 2015, 2016, 2017 (Alps)	2018 (Alps)
1	2016	2014, 2015, 2017, 2018 (Alps)	2016 (Alps)
2	2018	2014, 2015, 2017, 2018 (Himalayas)	2018 (Alps)

The internal structure of the datasets within Experiment 2 is the same as in Experiment 1. The number of training samples differ slightly between Experiments, because the Landsat-8 tiles cover a slightly smaller area in the Himalayas. Experiment 2 is tested on 2018 so that the results of this Experiment could be compared to the results of Experiment 1 Run 2018.

The training and testing datasets are converted to compressed NumPy arrays. There is one compressed NumPy array for the training dataset, and one for the testing dataset. The training dataset therefore consists of two layers. The first layer contains the input image arrays, with 16 bands per image array. The second layer, which are the target image arrays, has 7 bands per image array. The testing dataset also consists of two layers. Like the training dataset, the first layer are the input image arrays, and the second layer are the target image arrays. The first layer has 16 bands per image array, whereas the second layer has 7 bands per image array. The training set contains data of 4 years, whereas the testing set contains data of 1 year, resulting in an 80-20 distribution for the training/testing sets. An 80-20 distribution for the training/testing sets is a commonly used distribution (Mohanty et al., 2016; Luongo et al., 2021; Rozet et al., 2019). Both the training and testing datasets need to be normalised. They are normalised as entire datasets to values between -1 and 1. This needs to be done because the cGAN uses a tanh activation function for the output layer of G in the range of -1 and 1. Therefore, the discriminator also needs to receive images with pixel values between -1 and 1. The tanh activation is used for the cGAN because it is able to generate higher gradients in training, which in turn results in higher updates in the weights of the network. This allows the cGAN to have larger updates of the weights during training compared to using a sigmoid activation (Antoniadis, 2023), reducing computing time. A risk associated with the larger learning rate is that the loss of the model is less likely to converge to exactly 0. For this, it would need a smaller learning rate.

3.2. Training the cGAN

The training datasets are used to train the Generator (G) and the Discriminator (D). The cGAN algorithm used in this study is based on the Tensorflow implementation pix2pix paper by Isola et al. (2017). This algorithm contains a generator with a U-Net based architecture. The U-Net is an encoder-decoder, but with the possibility of skip connections between mirrored layers. The U-Net is used for this implementation because it allows low-level information from the input to be directly implemented into the output, without the risk of losing these "details" in the bottle-neck of the encoder-decoder (Ronneberger et al., 2015). According to this paper, using a U-Net yields higher quality results than using a regular encoder-decoder. The performance U-Net was also examined in the study by Rodríguez-Suárez et al. (2022), who concluded that a U-Net can generate satisfactory results for generating multispectral imagery. D is represented by a convolutional PatchGAN. A PatchGAN uses fewer parameters than a PixelGAN, has a lower runtime and can be applied to larger images. Instead of classifying each individual pixel as real or fake, a PatchGAN is able to classify a $N \times N$ patch as real or fake.

Both G and D use an Adam optimiser. The Adam optimiser “method is straightforward to implement, is computationally efficient, has little memory requirements, is invariant to diagonal rescaling of the gradients, and is well suited for problems that are large in terms of data and/or parameters” (Kingma & Ba, 2014). The learning rate of G and D is also the same, at 2e-4, as per the pix2pix paper.

G in this study has 54’436’487 trainable parameters, and 10’880 non-trainable parameters. There are 14 sequential layers, and 1 convolutional layer. There are 6 concatenating layers. These are for the skip-connections. The full model summary is shown in Appendix 2.

D has 2’786’049 trainable parameters. The number non-trainable parameters for D is 1’972 parameters. The reason it has less parameters than G is because it has many less layers. D consists of 3 sequential layers, 2 convolutional layers, 2 zero-padding layers, a batch normalisation layer and a leaky ReLU layer. The model summary of D is also shown in Appendix 2.

The optimal number of training steps needed depends greatly on its application. 100’000 training steps will be used to answer the research question. This number was chosen as a result of hardware limitations, and looking at the Tensorflow pix2pix implementation, where 80’000 steps were deemed satisfactory to generate an accurate image. Critically, 80’000 training steps chosen in the Tensorflow pix2pix implementation was for a 1 channel to 3 channel translation. This choice of training steps is evaluated in the Discussion.

3.3. Creating the testing image

The output of the cGAN are subsets of the Landsat-8 image, in the form of 7-dimensional arrays. These arrays are individually converted to a GeoTiff format through python. They are converted to GeoTiffs by using the same CRS and geo-transform as the test images which the arrays are generated to mimic. Finally, the 1395 GeoTiff images of size 256x256 pixels are merged together to create one single generated multi-spectral Landsat-8 tile. This tile is the generated image on which the testing results are based.

3.4. Evaluating the cGAN’s results

The Results are divided into three sections. In the first section, the cGAN’s ability to predict multi-spectral imagery with the provided environmental variables is assessed, with snow cover the primary issue. The second section presents the results related to the snow extent. The final section of the results compares the generated reflectance values of snow to the true snow reflectance values. The exact contents of the Results section are detailed below.

3.4.1. Multispectral image analysis

To quantify the resemblance of the generated imagery to the true imagery, the True Error (TE) per band is calculated, after 100’000 steps. The formula for the TE is shown below (equation 2):

$$TE = \hat{y} - y \quad (2)$$

The generated reflectance values should be similar to the true reflectance values, so that the snow extent can be estimated through NDSI calculations. The generated reflectance values per band are examined and compared per Experiment. The RGB composites of the generated Landsat-8 images are presented per Experiment and qualitatively analysed and compared to the true Landsat-8 images.

3.4.2. Generated snow extent

Following the assessment of the generated reflectance values, the results section focusses on the generated snow extent. The snow extent is computed by looking at the number of pixels with an NDSI value greater than or equal to 0.4. To determine the performance of the cGAN in predicting the snow extent, a visual analysis is done by looking at where the snow extent is either over- or underestimated, and the reason for this. The over-/underestimations of the snow extent are compared to the input environmental variables. This is done through violin plots, relating the true error in snow extent to the environmental variables. The over-/underestimations of the snow extent are also visualised using 2-dimensional TE plots over the testing area.

The effect of the number of training steps on the performance of the cGAN to predict the snow extent is also examined. The snow extent is computed for every 5'000 steps and compared to the true snow extent.

To get a better understanding the effect of the individual environmental variables on the snow extent, two types of sensitivity analyses are done. The first sensitivity analysis is done by adding values to each of the individual environmental variables across all winter months. This is done to determine what environmental variable has the highest influence on the end of winter snow extent. The sensitivity analyses are performed on Experiment 1 Run 2018, and Experiment 2. For elevation, the range of values added to each pixel ranges from -600 metres to +600 metres. For precipitation, the range of added values range from -150 millimetres to +150 millimetres. The range of values added to the maximum and minimum temperature range from -15°C to +15°C. This creates new imagery, from which the snow extent can be computed. The snow extent is then plotted on the y-axis against the environmental variable on the x-axis, to give a visual representation of the sensitivity analysis. It is expected that the cGAN will have learned the positive relationships between elevation and snow cover, and precipitation and snow cover. It is also expected that the cGAN will have learned the negative relationship between temperature and snow-cover. The importance of the features is also assessed. Using the dry adiabatic lapse rate, which says that the temperature decreases with 9.8C/km, the elevation and temperature can be related. This can give an indication on what the cGAN has learned to be a more important feature for predicting snow extent. Temperature and elevation both affect precipitation, and there is no simple relationship between these variables.

The second sensitivity analysis is done to determine if the cGAN gives higher weights to certain months. This is done by adding a value to a single month in the testing input dataset, and then determining the end of winter snow extent. The desired result from this sensitivity analysis is to find out which winter month a change in environment causes the biggest change in the end of winter snow extent. The process is as follows: One variable in the first month (November) is changed. The cGAN then generates a new multispectral image from this new input image. From this new generated image, the snow extent is computed. Next, one variable in the second month (December) is changed. The cGAN then creates another multispectral image from this new input image, from which the snow extent is computed. This process is repeated for every month, for precipitation and temperature. 100 millimetres is the value added to the individual months to the precipitation runs, and 10°C is added to the temperature runs. This second sensitivity analysis is only done on Experiment 1, as it is expected to perform best in the multispectral image generation.

3.4.3. Generated spectral signature of snow

The last section of the results is focussed on the generated spectral signature of snow-covered areas, compared to the true spectral signature of snow. The spectral signature of snow is computed by looking at pixels where snow is present in both the generated imagery and the true imagery. The mean reflectance values are then computed per band over this snow area to generate the spectral

signature of snow. Aside from the spectral signature of snow, the reflectance histograms of each band is also presented and compared for each Experiment. Finally, the true error of the reflectance of snow for the generated versus the true image is computed per 5'000 training steps, to assess if the accuracy of the generated snow reflectance changes over the training steps.

4. Results

The results are presented in the same order as they were presented in the methodology. An interpretation of the results of a sub-section is presented at the end of the sub-section.

4.1. Generated Landsat-8 images

4.1.1. Comparison of generated and true Landsat-8 imagery

The generated Landsat-8 images from Experiment 1 and Experiment 2, compared to the true Landsat-8 images, are shown below (Figure 6).

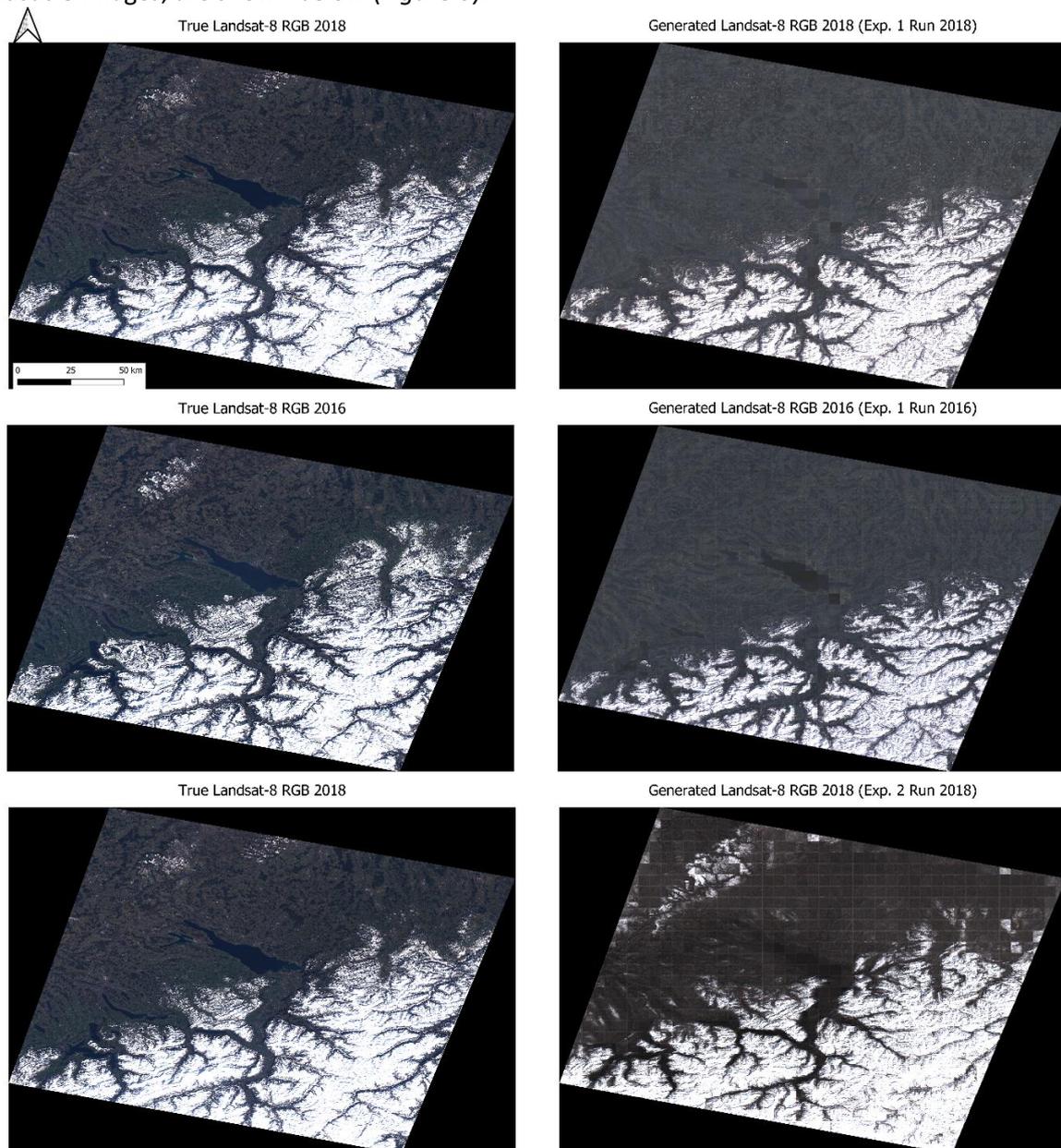


Figure 6 - RGB comparison between the true Landsat-8 imagery (left) and the generated Landsat-8 imagery (right) after 100'000 training steps.

All generated images shown above are after 100'000 training steps. This is because the quality of the generated imagery appears to increase with more training steps. RGB imagery for Experiment 1 Run 2018 after 5'000 steps, 50'000 steps and 75'000 steps are shown in Appendix 3. The generated imagery for both runs for Experiment 1, in RGB, shows a very high resemblance to the true imagery. The mountain range in the south-eastern corner of the image is mostly snow covered, whereas the lower lying areas on the north-western half is not snow covered. In addition to this, the lake in the middle of the true Landsat-8 image is also present in the generated imagery of Experiment 1, despite the cGAN not being given any data on water bodies. The RGB colours of the lower lying areas on the north-western half are very similar in the generated of Experiment 1.

The bottom row of *Figure 6* shows that using Himalayan training data to predict an Alpine landscape results a poorer representation of the landscape. The non-snow-covered landscape is grey in RGB, whereas it should be green. The lake in the centre of the generated image is unrecognisable. Tiling also appears to be an issue in the generated image of Experiment 2, as individual 256x256 pixel tiles can be easily seen. The north-west corner of the generated image also contains snow, when there shouldn't be any snow. In general, the generated image of Experiment 2 does still resemble a Landsat-8 image, just not as well as the images generated in Experiment 1.

4.1.2. cGAN performance per band

Below are the true error plots for the bands of the generated Landsat-8 bands compared to the true Landsat-8 bands of Experiment 1 Run 2018 after 100'000 training steps (*Figure 7*).

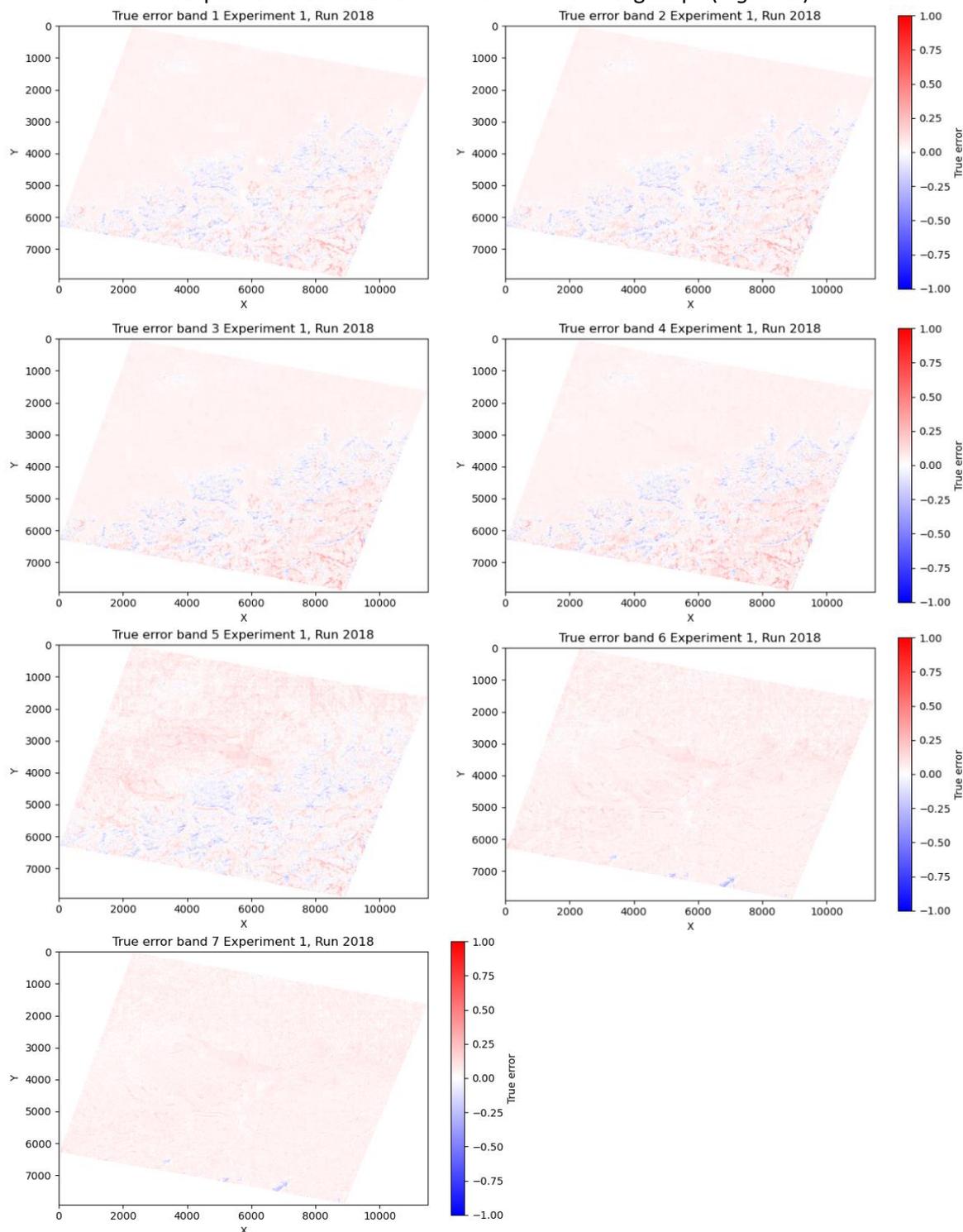


Figure 7 – 2-dimensional True Error (TE) plots for all bands of Experiment 1 Run 2018 after 100'000 training steps.

The 2-dimensional error plots show that most of the error occurs in bands 1, 2, 3, 4 and 5. The areas with the highest error appear to be in the mountainous areas (south-east of images). The mountainous areas are also the areas where the snow cover can be observed, meaning that the errors are likely to be related to the reflectance of the generated snow-cover. The bands in which the error is the highest, are also the bands where the reflectance of snow is the highest. In addition

to this, the cGAN appears to consistently underestimate the reflectance in the valleys in the first 5 bands. As snow reflects strongly in these bands, it is likely that the cGAN is predicting a bare surface rather than a snow-covered surface. The error is very small in the non-mountainous areas of the image. The north-western part of the image is light pink. This indicates that the cGAN consistently overestimates the reflectance of the lower lying landscape.

Below are the 2-dimensional true error plots for Experiment 1, Run 2016 after 100'000 training steps (Figure 8).

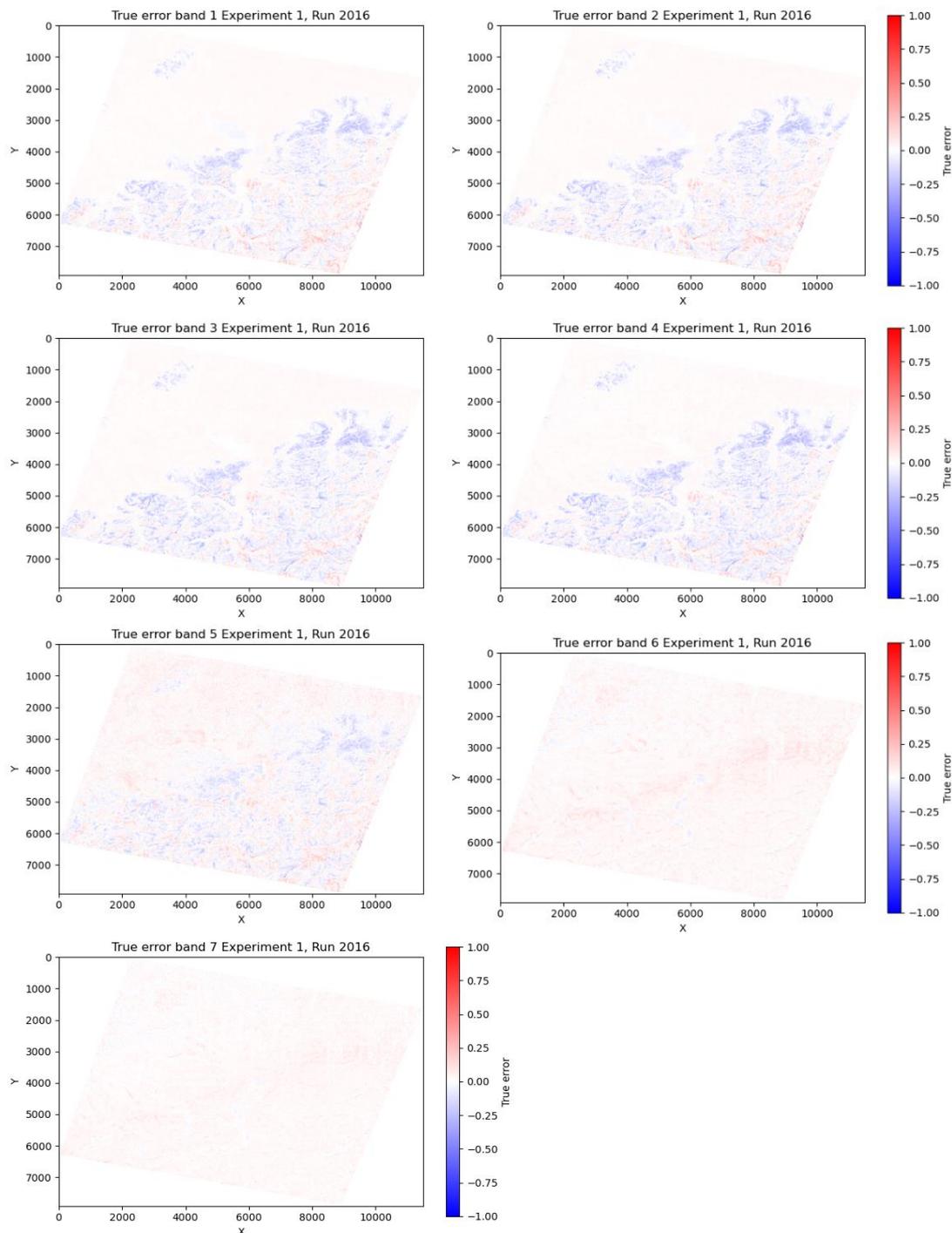


Figure 8 – 2-dimensional True Error (TE) plots for all bands of Experiment 1 Run 2016 after 100'000 training steps.

Clear differences can be seen compared to the 2-dimensional true error plots of the 2018 run of Experiment 1, and the 2016 run. This run shows a lot more underestimation of the first five bands in the lower lying areas than the 2018 run. These underestimations occur in mostly the same locations for the first 5 bands, and are very likely to be related to snow cover. One big difference with the true error plots of the 2018 run, is that there is much less overestimation of reflectance in the lower-lying, non-snow-covered north-western half of the image. Furthermore, the error in Band 6 and Band 7 are lower in this run compared to the 2018 run. One similarity between the runs, is that the reflectance is overestimated on the mountain peaks.

Lastly are the 2-dimensional true error plots for Experiment 2, which is tested on 2018 (*Figure 9*).

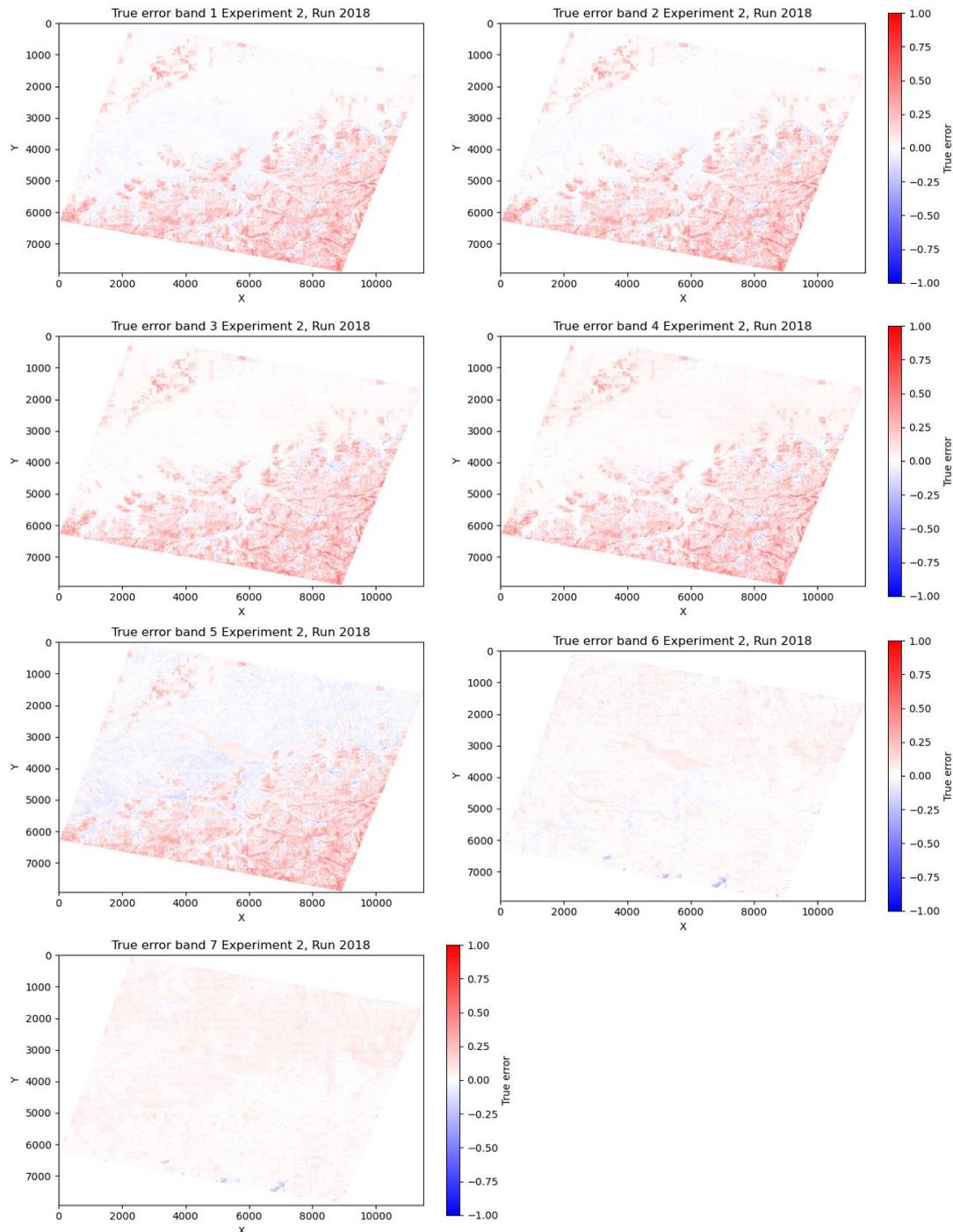


Figure 9 – 2-dimensional True Error (TE) plots for all bands of Experiment 2 Run 2018 after 100'000 training steps.

Figure 9 shows considerably different error plots than those of Experiment 1. Though band 6 and band 7 also appear to have the lowest true error over the entire image, just like the Runs within Experiment 1, the magnitude of the error is larger. The snow-covered areas show consistent overestimations of reflectance in the first 5 bands. This indicates that even though the generated image does predict snow, the reflectance is too high. By looking at these 2-dimensional error plots, it can be deduced that the performance of the cGAN decreases when including Himalayan data in generating a multispectral image for the Alps.

4.1.3. Interpretation of generated imagery results

One of the main areas of focus of this investigation, is to determine the effect of moving the training location away from the testing location. This was the aim of including Experiment 2 in the study, as it included 4 years of Himalayan training data. This Himalayan data has very different environmental and spectral conditions compared to the Alps.

It is clear from the results that the training location has a significant influence on the generated multispectral image. The generated multispectral image of Experiment 2 has a much larger true error per band than the generated imagery of Experiment 1. The generated reflectance values no longer closely correspond to the true imagery, unlike in Experiment 1. In addition to this, the snow extent estimation shows a much higher variance and error than in Experiment 1. The spectral signature of snow has a larger error in Experiment 2, and its variance per training step is also much larger than in Experiment 1.

The choice of using the Himalayas as training data for Experiment 2 was made due to its vastly different environmental conditions. As a result of this, the cGAN of Experiment 2 had to completely extrapolate the relationships between surface reflectance and environmental conditions when tested on the Alps. The effect of this has been shown above. These results agree with the results obtained by the study performed by Requena-Mesa et al. (2019), who concluded that the training location greatly influences the cGAN's performance at test time. They provided an example where the cGAN was trained on Europe, Asia and Africa and tested on the Americas. The error of their results increased.

4.2. Snow extent

4.2.1. Comparison of generated and true snow extent

NDSI is a normalised difference index. The advantage of this is that it is insensitive to constant bias in the generated image, meaning that even if the entire image's reflectance is overestimated, the snow extent can still be computed accurately. *Figure 10* below shows the computed snow extents and 2-dimensional true error plots of the true Landsat-8 images and of the corresponding Runs.

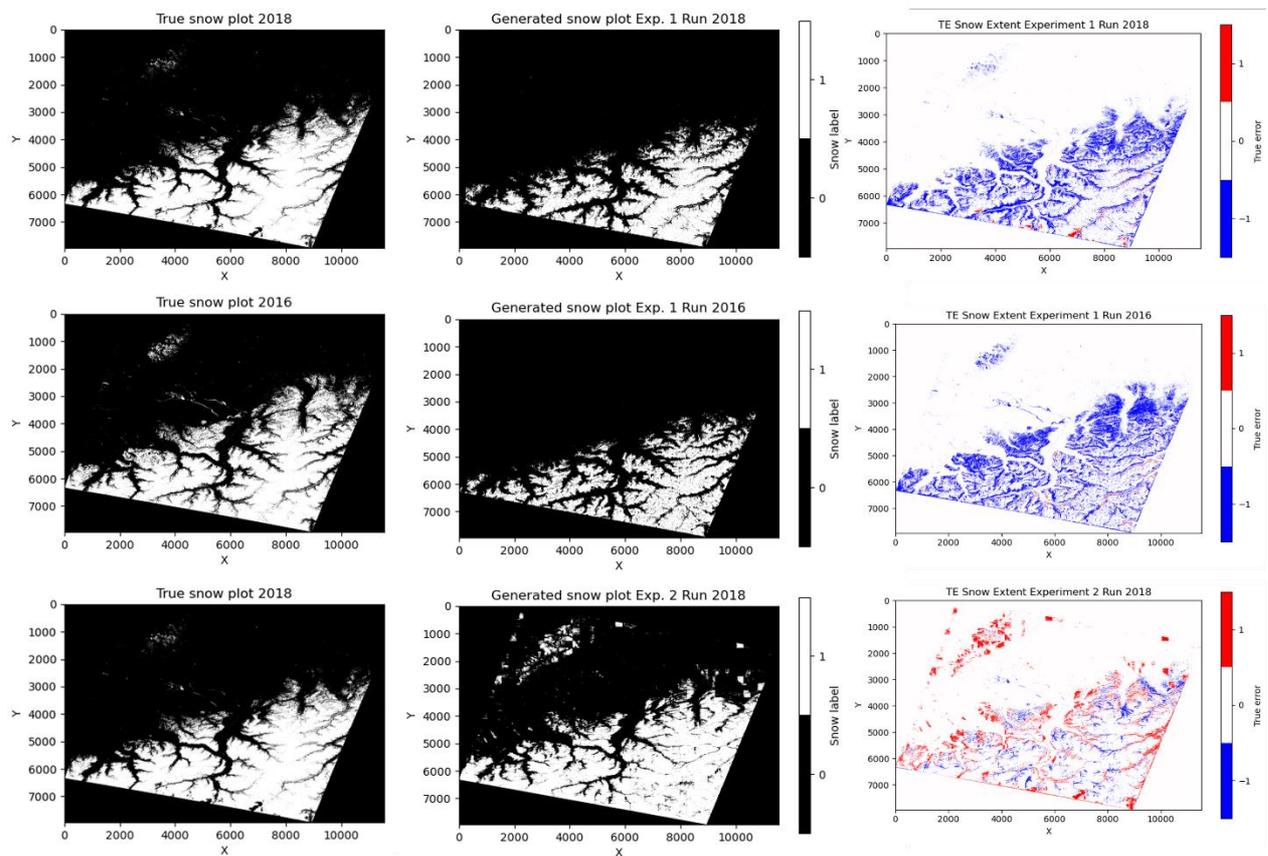


Figure 10 – Comparison of true Landsat-8 snow plots versus the generated snow plots for each Experiments, after 100'000 steps. Snow pixels = 1, non-snow pixels = 0. The True Error is the difference between the generated and the true imagery, as defined in (2).

Figure 10 shows several interesting results. Firstly, it appears that the snow extent is underestimated (blue colour) in every Run of Experiment 1. The valleys in the mountainous areas all appear to underestimate the snow extent, although the snow extent appears to be correctly estimated in higher elevation areas. In addition to the valleys inside the mountains, the lower elevation areas outside the mountains also lack snow cover in the generated image.

The 2018 run from Experiment 1 shows that the snow extent is correctly estimated in the higher elevation areas. These areas are all in white. However, the correctly estimated snow extents are mostly surrounded by blue areas. These are areas where the snow extent is underestimated in the generated image. These blue areas are generally in the valleys and low elevation areas. In addition to the blue, there are also red significant areas on the south side of the image. This indicates an overestimation of snow in the generated image compared to the true Landsat-8 image. However, when looking at the true Landsat-8 image for 2018 in Figure 6, these areas do appear to be snow covered.

The 2-dimensional TE plot for the 2016 run of Experiment 1 shows very similar results to that of the 2018 run. The snow extent appears to be correctly estimated in the higher elevated areas, where the TE is 0. However, the valleys and lower elevated areas are also mostly blue in this plot, signifying that the generated image did not generate snow-covered valleys. It is unclear if the red areas are on south-facing slopes or north-facing slopes.

Experiment 2 shows that the snow extent is over- and underestimated over the entire image. The snow extent appears to be overestimated greatly in the north-west corner of the image. Unlike the results from Experiment 1, Experiment 2 shows both over- and underestimations of snow extent in

the valleys, with no discernible pattern evident. The mountain tops appear to have no error in snow extent, similar to the plots of Experiment 1. As mentioned in the Results section of 4.1., tiling is also present in the snow extent plot of Experiment 2. For example, there is one 256x256 pixel tile completely overestimated in the north-eastern corner of the image, even though the neighbouring tiles show no error.

4.2.2. Relationship snow extent and environmental variables

Violin plots have been created to determine if there is a relationship between the over-/underestimations of snow extent with the environmental variables. A violin plot is a mix between a boxplot and a kernel density plot, highlighting the distribution of the individual variables. The plots are shown below (Figure 11).

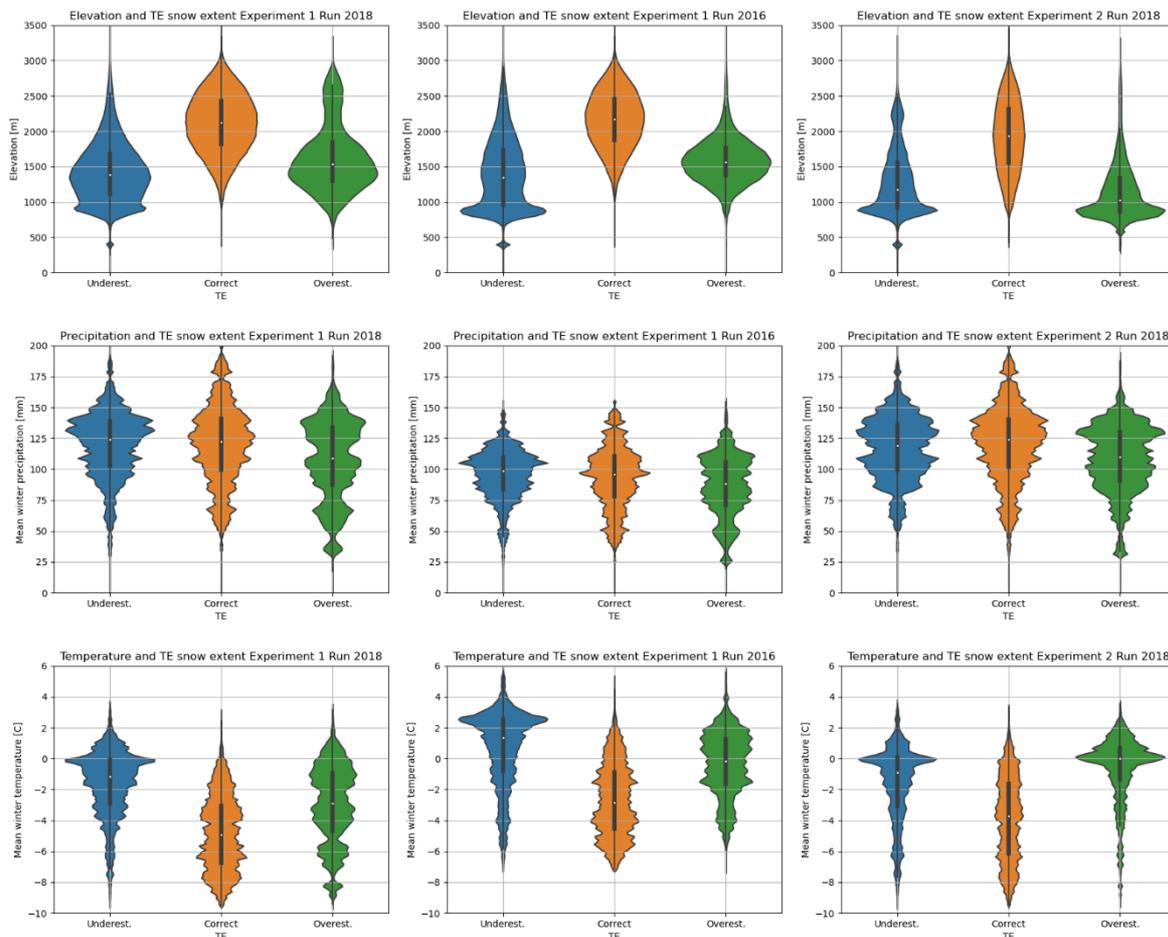


Figure 11 – Violin plots relating the environmental variables to the true error (TE) of the generated snow extent. The first row shows the elevation and TE of snow extent relationship. The second row shows the precipitation and TE of snow relationship. The third row shows the temperature and TE of snow relationship.

The violin plots provide a better understanding of where the over-/underestimation of snow extent is for the generated imagery. The first row of violin plots, which are for elevation, show that all Experiments best generate the snow extent at higher elevations. The overestimations of snow extent occur at higher elevations for Experiment 1, whereas the overestimations of snow extent occur at lower elevations for Experiment 2. The reason this is different can be seen in the north-west of Figure 10, where the cGAN is predicting snow at a lower elevation area even though there is no snow in the true image. The distribution of the over-/underestimation of snow extent in relation to elevation is very similar in Experiment 2. This coincides with the conclusion drawn from Figure 10, where it was stated that no discernible pattern could be recognised in relation to elevation and the error in generated snow extent.

The next row of the violin plots shows the TE in relation to the mean precipitation over the course of the winter period. It shows that the errors are made irrespective of the precipitation, as the kernel distributions of the true errors are clearly overlapping.

The results from the last row of the violin plots, the relationship with temperature, show a comparable result to the violin plots regarding elevation. This is reasonable, because temperature depends on elevation. The two Runs within Experiment 1 show very similar behaviour in the errors in generating the snow extent, whereas Experiment 2 shows a different behaviour. Experiment 1 underestimates the snow extent at warmer temperatures and generates the correct snow extent at lower temperatures. Experiment 2 over- and underestimates the snow extent at warmer temperatures. The snow extent is correctly generated at colder temperatures.

4.2.3. Effect of cGAN training steps on snow extent

To understand the effect of the training steps on the cGAN’s ability to predict the snow extent, the following plot is made (Figure 12):

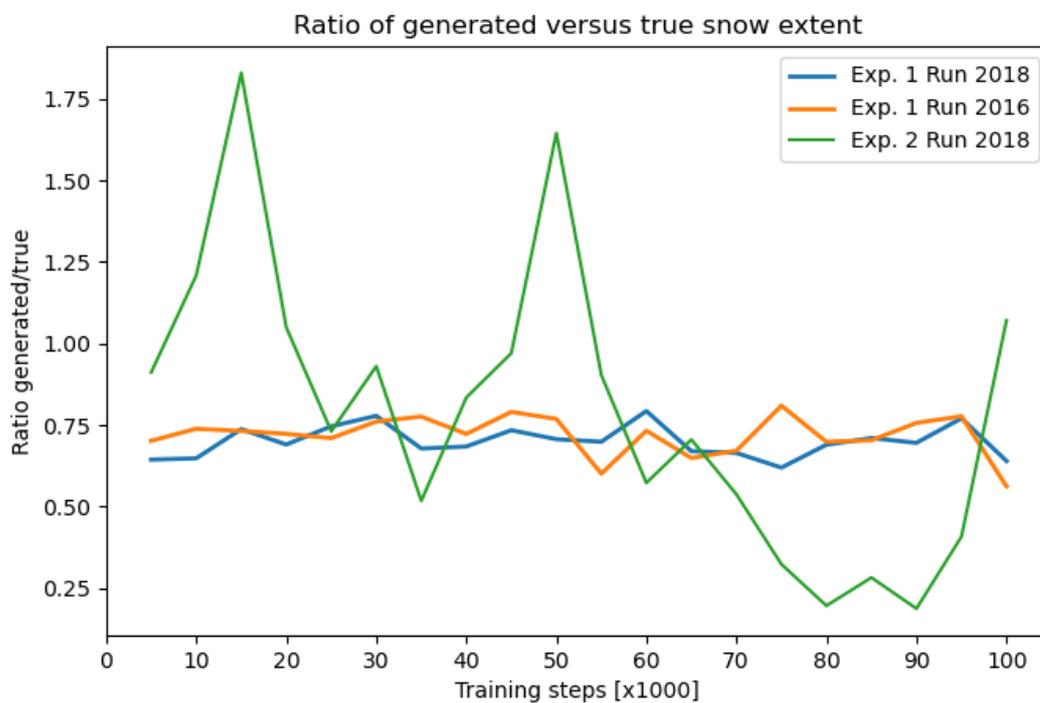


Figure 12 – Plot showing the training step dependency of the ratio of the generated/true snow extent. The ratio of the generated versus true snow extent is defined as $ratio = \frac{snow\ extent_{generated}}{snow\ extent_{true}}$.

Figure 12 shows, that ratio of the generated snow extent to the true snow extent is very similar for both runs within Experiment 1. These two runs consistently underestimate the snow extent, regardless of the training steps. They show very similar means and variance.

Experiment 2 shows very different results it shows a decreasing snow extent per training step, with a much higher variance compared to Experiment 1. The curve of the ratio fluctuates above and below 1, and does not appear to be stable. Despite this, the ratio of the generated over true snow extent is close to 1 at 100’000 training steps. By only looking at this plot, it may appear that Experiment 2 does generate a better snow extent than Experiment 1. This is not the case though, as the snow extent is not generated in the correct areas. This is evident in the True Error plots of the snow extent (Figure 10).

Some relevant absolute statistics are shown in *Table 7*.

Table 7 – Statistics relating to the true snow extent and the generated snow extent, after 100'000 training steps.

Experiment	Run	True snow extent [km ²]	Mean generated snow extent [km ²]	Standard deviation of generated snow extent [km ²]
1	2018	18790	13158	878
	2016	18960	13642	1145
2	2018	18790	14865	8133

These statistics are a confirmation of the conclusions drawn from *Figure 10*. It shows that the mean of the ratio of the generated versus the true snow extent, is very similar for the two Runs in Experiment 1, across all training steps. Furthermore, *Table 7* shows that the standard deviation of Experiment 2 is at least 4 times larger than the standard deviations reported from Experiment 1. The results from *Figure 12* show no real improvement in the snow extent generation as a function of training steps for Experiment 1. This could suggest that there is little reason to continue training the model past 20'000 training steps, if the generated snow extent does not really change anymore. However, the quality of the generated Landsat-8 imagery does appear to increase with increasing training steps. With an increasing number of training steps, the generated image contains more detail. One pixel is able to contain more landscape features after 100'000 training steps than after 10'000 training steps. This is shown for a small area in *Figure 13*.

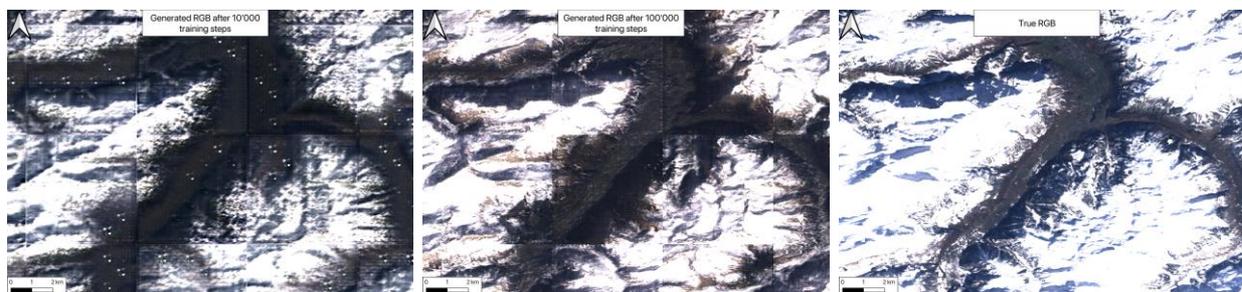


Figure 13 - Comparison of generated RGB images for Experiment 1 Run 2018, after 10'000 training steps (left) and 100'000 training steps (middle). The true Landsat-8 RGB image is shown right.

The two generated images have a very similar snow extent. One big difference which can be noted between the generated images, is that the image after 100'000 training steps shows more snow in the valleys. This is what is meant with more training steps creating more detail in the predicted image. The training steps may not have a very large effect for determining the total snow extent within this study, but it would be very important for a study focussing on a smaller area, where more detail is needed.

4.2.4. Snow extent dependency on input variables

To reiterate, the difference between Run 2018 and Run 2016 from Experiment 1 are the environmental variables. 2018 was an extreme year, with never seen before temperature and precipitation values. 2016 was an average year, with all environmental values lying in the middle of the training range. From the snow extent analysis, no clear conclusion can be drawn to what improves or worsens the performance of the cGAN. Following this, a sensitivity analysis is done (*Figure 14*). This sensitivity analysis is done for the 2018 run of Experiment 1 and for Experiment 2.

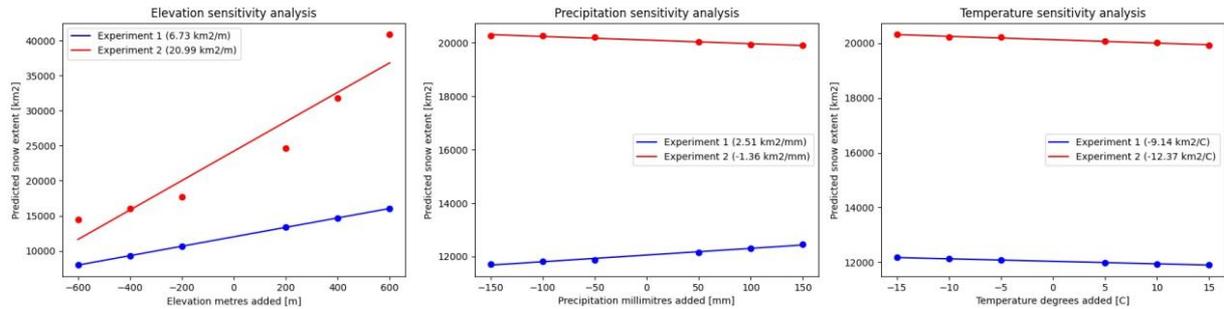


Figure 14 – Results of sensitivity analysis. From left to right: elevation and snow extent relationship, precipitation and snow extent relationship, temperature and snow extent relationship. The linear relationships are given in the legend.

The plots above show the generated snow extent on the y-axis for differing environmental settings, which are shown on the x-axis. What the plots indicate is that the cGAN correctly learned the relationships between the environmental variables and the snow extent. The relationships learned for Experiment 1 coincide with literature, with elevation and precipitation exhibiting a positive relationship in relation to snow cover, and temperature a negative relationship. The relationships are not the same for Experiment 2. Experiment 2 shows a negative relationship between precipitation and snow cover. This is counterintuitive, as more precipitation should mean more snow. This is discussed in 4.2.6. The other relationships that the cGAN of Experiment 2 has learned are roughly similar to the ones learned in Experiment 1. An interesting difference in the elevation sensitivity analysis can be seen between the two Experiments. Experiment 1 shows a perfectly linear relationship, whereas the data points of Experiment 2 deviate from the linear least squares regression line. The results from the Experiment 2 elevation sensitivity analysis show that the slope is very comparable to that of Experiment 1 before 0, but then increases sharply after 0. This is also discussed in 4.2.6.

Using this sensitivity analysis as a means to do a feature importance test may be trivial when considering that the variables cannot be related deterministically. However, some feature importance analysis can still be done relating to these sensitivity plots. Assuming the dry adiabatic lapse rate, the influence of elevation versus temperature on predicting snow extent can be determined. As shown in the plots, the linear relationship of elevation and snow extent is $6.73 \text{ km}^2/\text{m}$ for Experiment 1. Dividing this by the dry adiabatic lapse rate equates to $-687 \text{ km}^2/\text{C}$. As this value is 10^2 larger than the linear relationship between temperature and snow area of Experiment 1, it can be deduced that elevation is more important for the cGAN's snow extent prediction than temperature. Assuming that precipitation and elevation share a similar but positive relationship as temperature and elevation, it could be assumed that elevation is the most important feature for the cGAN when predicting the snow extent through generated Landsat-8 imagery.

4.2.5. Snow extent dependency on changes in input variables

The second sensitivity analysis determines if the cGAN treats all months as equal, or if certain months are treated as more important than others. This is done by adding values for precipitation and temperature, per month. This is not done for elevation because it is constant over such a short period of time. The results of this sensitivity analysis are shown below (Figure 15):

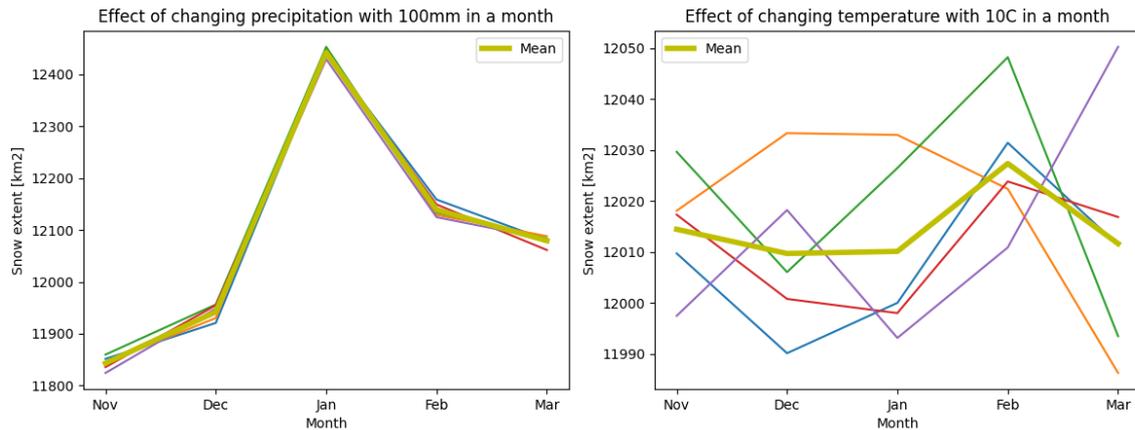


Figure 15 – Graphical results of sensitivity analysis 2, after 5 individual runs. The left figure shows the effect of adding 100mm of precipitation per month. The right figure shows the results of adding 10C of temperature per month.

The plots show that changing the precipitation in January affects snow cover most. This is because the training set of the Alps, on average, saw January as the month with the highest precipitation. This plot can be seen in Appendix 4. An increase in temperature shows no effect depending of month.

4.2.6. Interpretation of snow extent results

The achieved results show that cGAN's can be used to determine the end of winter snow extent through generated Landsat-8 imagery. The generated reflectance values are similar to the true reflectance values of the Landsat-8 imagery, allowing the NDSI calculations to be made. The NDSI calculations could then be used to determine the snow extent over the testing area.

There are several discussion points that can be drawn from the snow extent analysis:

1. The snow extent is underestimated in both Experiments. Experiment 1 yields more accurate and precise estimations of the snow extent, whereas the results from Experiment 2 show a much larger variation in estimating snow extent depending on the training step, and also underestimates the snow extent on average over all training steps.
2. For Experiment 1, the underestimations correlate to the elevation, as lower elevation areas are generally snow-covered in the true imagery but not in the generated imagery. The snow extent is both overestimated and underestimated at lower elevations for Experiment 2.
3. The cGAN has also learned relationships between the environmental variables and the snow extent. It has learned linear relationships for Experiment 1; positive relationships between elevation, precipitation and snow extent and a negative relationship between temperature and snow extent. For Experiment 2, the results are different.
4. The cGAN has learned that a change in precipitation in January is most important for generating the end of winter snow extent.

These points are addressed below.

For Experiment 1, the underestimations of the snow extent mainly occur at lower elevations and in valleys. The underestimation of snow extent in the valleys is due to the coarse resolution of the precipitation and temperature data. To reiterate, the spatial resolution of the precipitation and temperature data is 4638 metres, and is shown in Figure 1. This data is being used to predict the snow extent at pixel level, which has a spatial resolution of 30 metres. The coarseness of the environmental dataset means that large areas are generalised under a single pixel. This prevents the

cGAN from being able to make detailed predictions over an area. This in turn leads to the cGAN from predicting no snow cover in the valleys. The valleys between mountains are much smaller than the resolution of the precipitation and temperature datasets, causing the cGAN then becomes overly reliant on the elevation dataset. During training, it has been seen that the lower elevation areas are generally not snow covered (north-west of 2016 and 2018 Landsat-8 images, for example), and so the cGAN will not predict snow in these lower elevation valleys.

A point of discussion are the number of training steps chosen for the cGAN. The results presented in Figure 12, showing the ratio of the generated snow extent to the true snow extent as a function of the training steps, indicate that the training steps have little effect on the cGAN's predictive ability of the snow extent. However, this is only true for the range of 0 to 100'000 training steps performed in the training of the cGAN, and possibly more training steps are needed for the cGAN's snow extent prediction to stabilise. Though the variance is high, the ratio of the generated/true snow extent of Experiment 2 shows a downward trend towards a True error of 0, but not enough training steps were performed to generate a clear trend. The downward trend with high variance is very different to the trend observed from Experiment 1, where the snow extent prediction is fairly constant. This shows that using Alpine data to predict the Alpine snow extent yields more constant predictions over the training steps than using Himalayan training data. Due to hardware restraints, more than 100'000 training steps could not be performed.

Both Experiments show a considerable variance in estimating the snow extent per training step. One of the reasons why the snow extent calculations are not constant can be due to the learning rate of the model. This learning rate is constant at $2e-4$ for both the Generator and the Discriminator. If this value is too high, the Generator and the Discriminator will not be able to converge to the true optimum of the model, whereas a too small learning rate will require the model to use a lot of computation time. To get around this, the learning rate should decay over training steps. This can be specified in Tensorflow, where it is possible to program how the learning rate will decay over training time. Using this method may help achieve a more stable snow extent calculation quicker.

For Experiment 1, the learned relationships are all linear. The cGAN has not been provided with any empirical relationships between the environmental variables and snow extent, so it has learned these relationships using only the input and target training data. These relationships are correct, to an extent. Previous studies have also found linear relationships between elevation, temperature, precipitation and snow cover (Jain et al., 2009; Rebetz, 1996; Foster et al., 1983). They found a linear negative relationship between temperature and snow cover, but with much lower temperatures. In this study's sensitivity analysis, there is 15°C added to the temperature variables, and the cGAN still produced a partially snow-covered image as it has learned a linear relationship between temperature and snow extent. However, adding 15°C to each of the winter months of the testing dataset would mean that there is never a negative temperature. Realistically, much less snow is expected as the winter months would only contain melt days. However, melt days were not included in the training dataset, and is a recommendation for future work.

The first sensitivity analysis has shown that the cGAN of Experiment 1 is most dependent on elevation for generating the snow extent. The sensitivity analysis also showed that temperature and precipitation have a lower importance in predicting the snow extent. One reason the temperature and precipitation have a such a significantly lower impact on the prediction of the snow extent is a result of the resolution of the input dataset, for these variables. The pixel sizes are too large to make detailed predictions on snow extent. This large pixel size of temperature and precipitation means that there are less relationships that the cGAN can learn between these variables and the target image. Despite the coarse resolution, the cGAN still learned the expected relationships between temperature, precipitation and snow extent for Experiment 1.

The first sensitivity analysis also shows that the cGAN of Experiment 2 has learned a negative relationship between precipitation and snow cover, contrary to Experiment 1 and literature. This can be attributed to the training data of the Himalayas. The training data of the Himalayas contained very little precipitation data. Years 2015 and 2018 barely received any precipitation, while still having a large snow extent. The range of the mean precipitation training data in the Himalayas is between [0.32, 29.86] mm. However, since it is cold enough in the Himalayas, the mountains are still snow covered even if there is no precipitation. This means that the training data presented to the cGAN shows no real relationship between precipitation and snow cover.

The first sensitivity analysis also shows that the gradients between elevation and snow extent are much greater in Experiment 2 than Experiment 1. This is especially visible when the elevation change is greater than 0. This is because the training set from the Himalayas contains a much higher snow extent, at high elevations. As soon as the elevations are increased in the sensitivity analysis of Experiment 2, the elevation conditions become closer to the Himalayan situation. The snow extent in the Himalayas is around twice as much as in the Alps. It could be that as soon as the elevation increases in the testing image, that the cGAN tries to emulate the Himalayan situation, causing the steep increase in the relationship between elevation and snow cover. It is unclear why this slope does not occur when the elevation change is less than 0.

4.3. Spectral signatures

4.3.1. Spectral curve of snow

Figure 16 shows the true and generated spectral signatures of snow, for both Experiment 1 and Experiment 2, after 100'000 training steps.

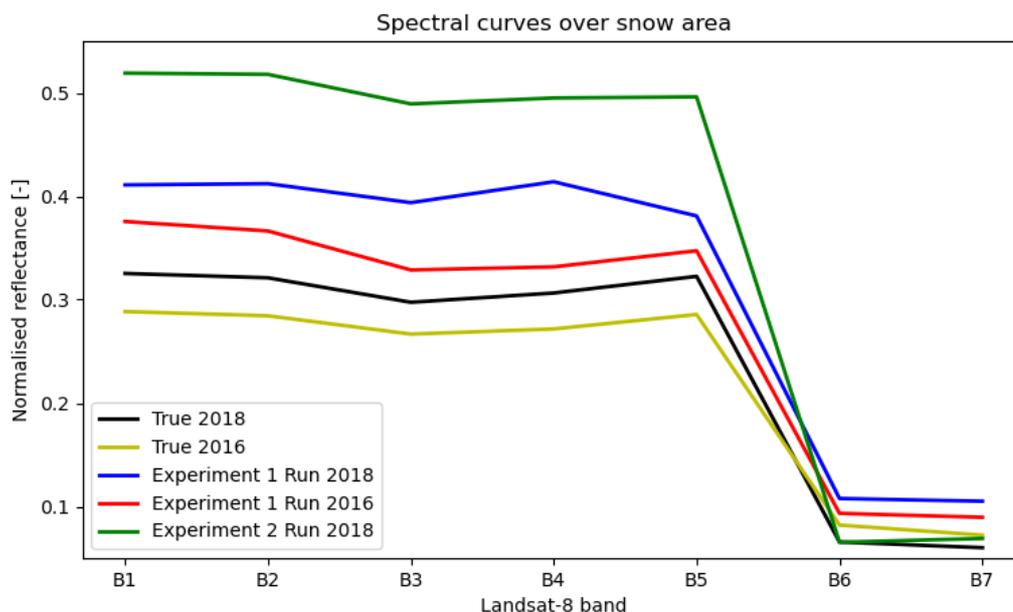


Figure 16 - Spectral curves of snow for each Experiment after 100'000 training steps.

The snow spectral curves show that the generated reflectance of snow is generally higher than the true reflectance. This is true for both runs within Experiment 1, and for Experiment 2. Within Experiment 1, Figure 16 shows that the 2016 run is slightly more similar to the true snow spectrum than the 2018 run. Experiment 2 shows a much larger overestimation of the snow reflectance compared to the Runs of Experiment 1. The general shape of each spectral curve is similar.

Band histograms for snow-covered areas show interesting characteristics, and aid the explanation of why the generated snow reflectance is too high. These are shown below (*Figure 17*):

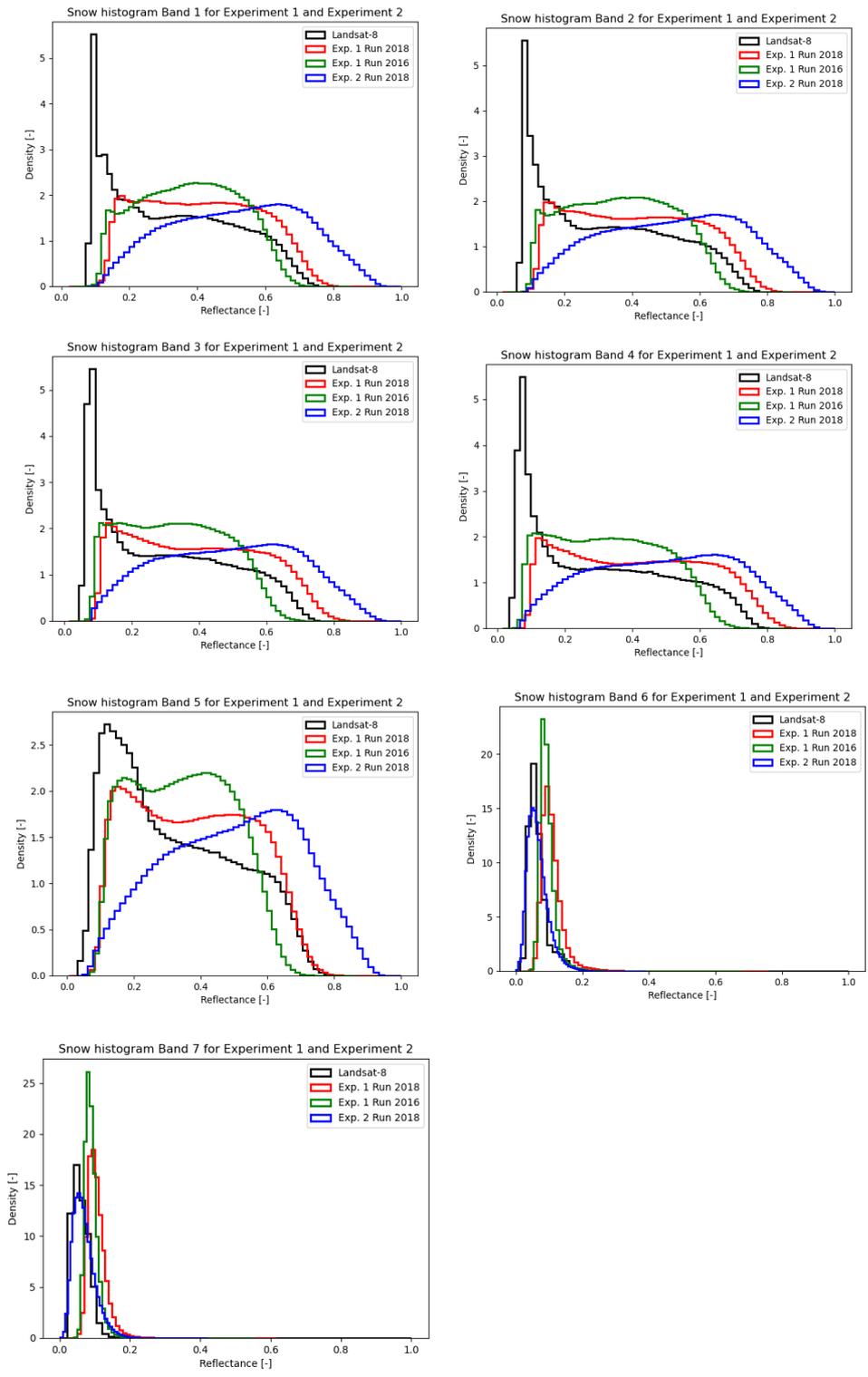


Figure 17 - Band histograms of snow-covered areas, for the true Landsat-8 image and each Experiment after 100'000 training steps.

The band histograms bolster the conclusions which could be drawn from the spectral signatures of snow. Once again, band 6 and band 7 show the lowest reflectance. The first 5 bands, show very distinct characteristics, but are different for the two Experiments. In Experiment 1, the peak of the reflectance is at a lower reflectance than the peak of the reflectance in Experiment 2. In addition to

this, the width of the histogram curves for the first 5 bands are much narrower than for Experiment 2. The more narrow curve of Experiment 1 better represents the true histogram curve of snow, so it can be concluded that Experiment 1 also performs better than Experiment 2 in these plots.

4.3.2. Effect of cGAN training steps on generated snow reflectance

The effect of the training steps on the generated reflectance for snow is also analysed, and shown in Figure 18.

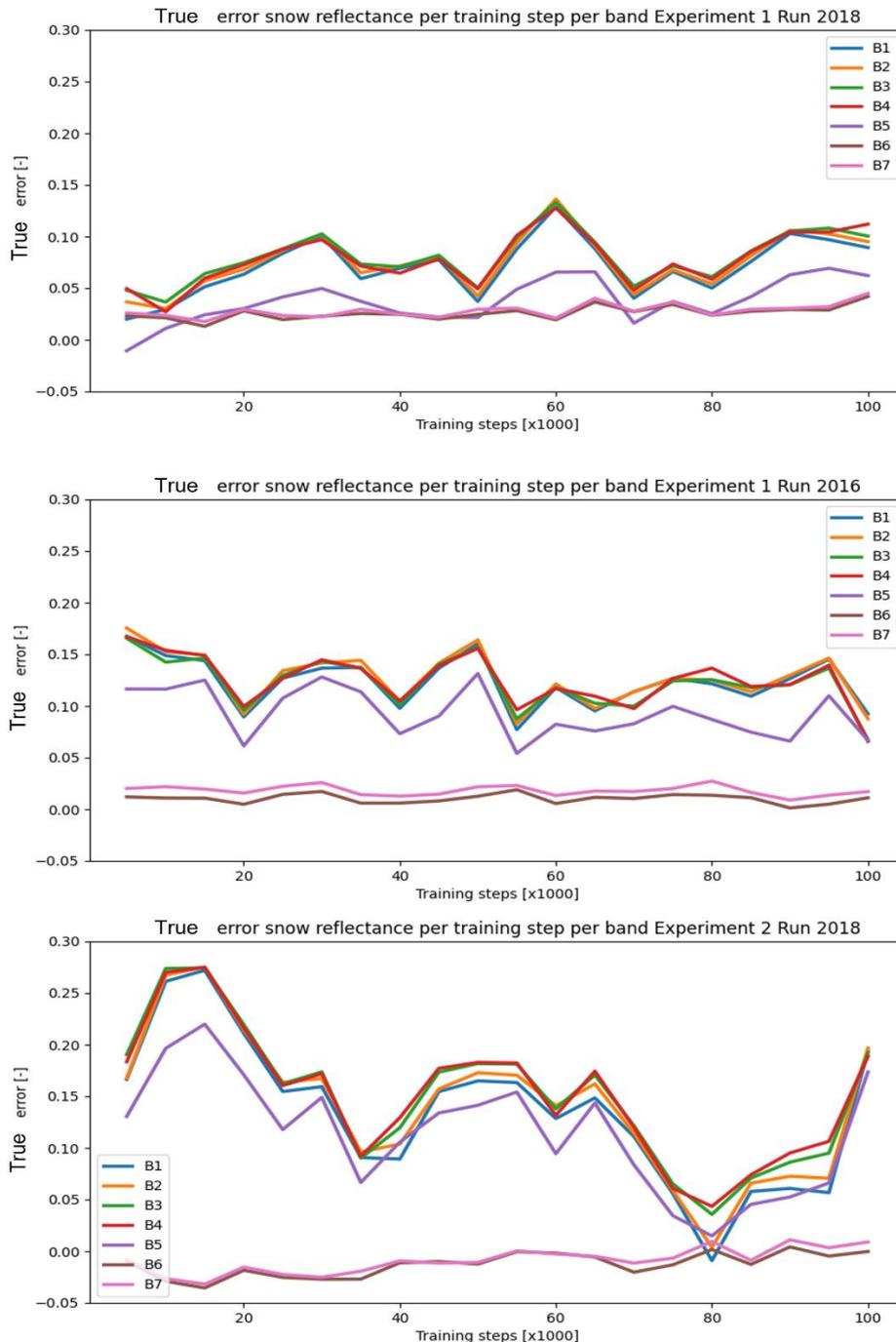


Figure 18 - Training step dependency of the generated snow reflectance for each Experiment.

The plots above show different characteristics for the two Experiments. The results for the first Experiment show that the number of training steps has a very marginal effect on the cGAN’s ability to predict the snow reflectance in the Alps. The TE for the snow reflectance does not converge to 0

between 0 and 100'000 training steps. The variance in the generated snow reflectance is higher in Experiment 2 than in Experiment 1, highlighting yet again the importance of the training location in relation to the testing location. One similarity between each Experiment is that Band 6 and Band 7 have the lowest error, but this is because snow reflects very little in these bands.

4.3.3. Interpretation of spectral signature results

If generating the snow extent was the only objective, the target image could have been a snow map instead of Landsat-8 imagery. In this way, the cGAN would only need to learn the relationship between the environmental variables and the snow extent. Now, it has learned the relationship between the environmental variables and the 7 Landsat-8 bands, from which the snow extent is computed. However, generating a multi-spectral Landsat-8 image does provide more information regarding the quality of the snow. It can provide some information about the snow's characteristics, given the generated snow spectral curves resemble the true snow spectral curves.

These results show that the generated spectral curves of snow have a similar shape across all Experiments to the true spectral curve of the snow. For Experiment 1 and Experiment 2, the main issue is the consistent overestimation of the snow reflectance across all bands. The results from Experiment 2 show that the error of the spectral signature is higher when using Himalayan data, supporting the conclusion that using Himalayan training data to predict an Alpine landscape increases the error in the results.

Interestingly, combining Alpine and Himalayan data results in a better generation of the spectral signature of snow. These results are shown in the Appendix 4, as they are not a direct study objective. One conclusion which may be drawn from this, is that increasing the amount of training data will aid the cGAN in being able to predict a more correct spectral signature of snow. Increasing the size of the training set typically yields more accurate results in machine- and deep learning (Millard & Richardson, 2015; Sun et al., 2017). This study could have benefitted from a larger range of training data over the Alps. Currently, Experiment 1 uses 4 end of winter Landsat-8 images in training. Experiment 2 uses 8 end of winter Landsat-8 images in training. The result of doubling the training dataset is considerable, as the spectral curve of snow was immediately better estimated in Experiment 2, despite the training data being geographically less relevant. Increasing the training dataset with Alpine data will most likely decrease the True error in the reflectance of snow.

The effect of the trainings steps needs to be discussed. The TE of the reflectance of snow is consistently positive for all bands in Experiment 1 and in Experiment 2. The true error does not appear to converge to 0 for either Experiment in the 100'000 training steps. This may be due to not being provided enough training data. Appendix 5 shows that the true error does converge close to 0 if both Alpine and Himalayan data is used. To generate less error in the reflectance of snow, the solution would be to add more Alpine training data.

5. Discussion

5.1. Location transferability in training and testing

It is clear from all the sections in the results that the training location has a significant effect on the performance of the cGAN in generating multispectral imagery. This study used the Himalayas to test the cGAN's transferability. However, to what extent was choosing the Himalayas a good starting point to test the cGAN's transferability from training to testing location? Results from previous studies have indicated that a different training location has a negative effect on the cGAN's performance at test time. This is now also shown to be true in relation to snow cover. Using the Himalayas as training location to then test on the Alps, would in any case lead to a poor result as the environmental conditions are too different. A more similar mountain range to the Alps, such as the Pyrenees, would most likely have resulted in better results when tested on the Alps because the environmental conditions are more similar. This was not investigated in this study, and is recommended for future studies. It may give a better understanding of to what extent the training location can be different to the testing location, and still maintain good model performance.

The generalisability of the cGAN can be increased by broadening the training set (Adnan & Umer, 2022). Broadening the training set can be done training the cGAN on a mountain ranges such as the Alps, Himalayas, Andes, Rockies and Pyrenees. The effect of this is that the cGAN's performance is likely to decrease less when it is faced with an unseen mountain range. A similar result has been found in a cGAN study by do Lago et al. (2023), who used different training locations to improve the cGAN's generalisability to unseen testing locations. However, the specific aim of this study was not to make a cGAN which could predict the snow extent accurately on different mountain ranges, and it would be an extension of the scope of the current setup.

5.2. Improving the training set

5.2.1. Using higher resolution input datasets

Improving the training dataset is vital for improving the model performance, as reported by Millard and Richardson (2015). Using Himalayan training data deteriorates the results in comparison to using geographically relevant data, showing the cGAN performs better when the location of the training dataset is similar to the location of the testing dataset. To improve the training dataset for a more accurate snow extent calculation, several changes and additions should be made. Firstly, it would be advantageous to use a dataset with a higher resolution for precipitation and temperature. The only higher resolution dataset is the elevation dataset. As mentioned in the Data section, the precipitation and temperature datasets have a coarse resolution compared to both the elevation and Landsat-8 datasets. This causes the cGAN to generate very strong relationships between elevation and snow cover. This strong relationship can be seen very well in the first sensitivity analysis of Experiment 1 and Experiment 2. Using a higher resolution precipitation and temperature dataset will allow the cGAN to learn more detailed relationships between these variables and snow extent.

5.2.2. Including more variables

The number of input variables to generate multispectral imagery should be increased to achieve better results. This was shown by Requena-Mesa et al. (2019), who used 32 unique environmental variables to predict 4 spectral bands, as opposed to the 3 unique environmental variables used to predict 7 spectral bands in this study.

Several additions should also be made to the training set to improve the cGAN's ability to predict the snow extent through generated Landsat-8 imagery in the current setup of the model. One variable which should be added to the training set is the mean temperature. Currently, only the minimum and maximum temperature over a month is used. These variables were chosen based on the available variables from WorldClim v2. However, the minimum and maximum temperatures are highly influenced by extreme days inside a month. The whole month could be cold, but one warm day will greatly increase the maximum temperature value of a pixel over the month. Adding the mean temperature as a variable will allow the cGAN to not be influenced too heavily by extremes, while still taking temperature extremes into account. Another improvement would be to include melt days in the training set. This will provide information on the number of days where there has been net ablation/accumulation. Another improvement would be including the aspect of slopes into the training set. As the Alps are in the northern hemisphere, south-facing slopes should be less snow-covered than northern-facing slopes (Sharma, 2014). Aspect is most influential on snow cover in lower elevation areas (Jain, 2009), which is exactly where most of the snow extent errors were made.

In addition to more environmental data, it could be beneficial to include a "beginning of winter" snow extent map. This was considered but not done because the ultimate goal of this study is to apply the methodology for future scenarios. In future scenarios, a "beginning of winter" snow extent map is not available.

5.2.3. Using a more accurate 'end of winter' target image

Another improvement of the training set is to use Landsat-8 data that is exactly from the end of March. In the current setup, the "end of winter" Landsat-8 images range from March 18th to April 10th. The dates are not precisely at the end of March due to lack of cloud free imagery. However, choosing more temporally consistent end of winter images may lead to more accurate generated images from the cGAN. In general terms, the performance of a machine learning algorithm is highly dependent on the quality of the training set (Elmes et al., 2020). By using Landsat-8 data which is not precisely at the end of winter, the training set quality is lower than it could be.

5.2.4. Data augmentation

Data augmentation is another way to improve the training set (Wang et al., 2019). Data augmentation consists of flipping, translating, rotating and mirroring training images to create "more" images for the training set, and has been reported to improve the performance of neural networks (Taylor & Nitschke, 2018). It allows the creation of more data without going through the laborious process of collecting more data.

5.2.5. Increasing volume of training set

Finally, with more computing power than currently used, a much longer training dataset should be used. Ideally, the training dataset should include more than 4 years. In addition to adding years to the training set, daily precipitation and temperature measurements can be used instead of monthly.

5.3. Improvements to the cGAN architecture

5.3.1. Number of training steps

Another point of discussion is the number of training steps. 100'000 training steps were chosen for both Experiments. This was chosen on the basis that 80'000 were deemed sufficient in the Tensorflow pix2pix tutorial, on which this cGAN architecture is based. 100'000 training steps were chosen as the datasets of this study contain more layers (channels), so more parameters need to be learned. However, a case could be made for increasing the number of training steps. Increasing the

number of training steps could allow the cGAN to obtain a smaller loss, thereby increasing the accuracy of its predictions.

5.3.2. Assign weight to individual months

To achieve more accurate results for the snow cover at the end of March, less weight should be given to the early winter months like November and December, and more weight to February and March. The reason for this is as follows: if December and January are exceptionally warm, most of the snow which may have fallen in November will be melted. If then the temperature in February and March is exceptionally low, and the precipitation high, it would be expected that snow extent at the end of March is relatively high. The months of December and January are of less influence in this scenario. It would therefore be reasonable to add smaller weights to the early winter months, and larger weights to the later winter months.

If the climate scenario described in the previous paragraph would be inverted, i.e., a cold and snowy early winter and a warm late winter, the weights would still be useful as the later months would still have the largest influence on the end of winter snow extent. The main risk associated with adding weights to the winter months, is how to set these weights. If the early winter months receive too small weights, they will have too little influence on the end of winter snow extent. Likewise, if the late winter months receive too large weights, they will become too important and the cGAN will disregard the early winter months.

5.3.3. U-Net generator vs. ResNet generator

The cGAN used in this study uses a U-Net in the generator. Rodriguez-Suarez et al. (2022) evaluated the use of a U-Net compared to other network models, such as convolutional neural networks and ResNet. Their study found ResNet to be the best choice for reconstructing multispectral imagery from RGB photos. As the aim of their study is not the same as the aim of this study, their results are not entirely relevant. However, their results do indicate that using a U-Net architecture in the generator is not always the best way forward. It would have been beneficial to do an evaluation of different generator architectures for this study.

5.3.4. Adaptive learning rate

The cGAN architecture can further be improved by using an adaptive learning rate. Using an adaptive learning rate in the generator and discriminator would allow the model to converge more accurately (Zeiler, 2012). Using an adaptive learning rate also resolves the issue related to how to choose a learning rate. An adaptive learning rate does not need a user-specified constant learning rate, which is the case for the cGAN used in this study.

5.4. Future research

The goal of this research was to show that the snow extent can be calculated using environmental data over the Alps using generated multispectral imagery. At the moment, this study has only focussed on the past, with images generated for 2016 and 2018 using different locations of the training sets.

Multispectral imagery can also be created for the future over the Alps using future environmental conditions. These are provided by the IPCC emission scenarios from ScenarioMIP. ScenarioMIP contains monthly gridded temperature and precipitation projections. Their gridded projections are freely available to download from WorldClim v2, for example. Generating multispectral imagery for future climate can provide more information on how the snow extent will change under different

emission scenarios. In addition to this, the multispectral imagery can give indications on the snow characteristics in a future climate.

By understanding the pitfalls and limitations of the current model, the cGAN can be improved to become more accurate in its snow extent predictions. Questions that need to be answered in future research, is when the optimal number of training steps is reached, what other climate variables need to be included for a better snow extent prediction, and how to improve the spectral signature of the generated snow. With these questions answered, the currently inexistent Landsat-8 imagery can be generated over the Alps using the ScenarioMIP variables to create snow maps of the Alps in the future. This will provide valuable insights into the future snow characteristics over the Alps.

6. Conclusion and outlook

The research question of this study asked the following: To what extent can a cGAN be used to determine the end of winter snow extent in the Alps through generated Landsat-8 imagery and environmental data? The environmental data used in this study was an elevation map, monthly precipitation data, monthly minimum temperature data and monthly maximum temperature data. The Landsat-8 data was from the end of winter, which has been defined as being the end of March. The results show that the snow extent is consistently underestimated in the current setup when using only Alpine data, and very variable when using Himalayan training data. The snow extent is typically underestimated at lower elevations. The reflectance values of the snow are overestimated when using Alpine training data, and even more overestimated when using Himalayan training data. It has also been shown that increasing the training steps have little influence in predicting a better snow extent, but do improve the quality of the generated Landsat-8 imagery. The generated imagery after 100'000 training steps has much more detail than the generated imagery after 10'000 training steps. Lastly, this study showed that moving the training location away from the testing location increases error in the results, both in relation to the multispectral Landsat-8 image generation and the snow extent estimation. This corresponds to the results found in previous studies, but have now been shown to also be true in relation to generating the snow extent.

A cGAN is a very valuable tool for multispectral image generation. It provides an accurate representation of what a landscape looks like, while only being given environmental data. Once the model is trained, no ground truth data needs to be collected for image generation, making a cGAN a very viable alternative for snow extent analysis in future scenarios. It is recommended to continue research into this, because the true potential of the cGAN is yet to be reached in relation to multispectral snow extent analysis in the Alps.

The full code is available on GitHub. This code includes the preparation of the input and target datasets, and then the full code of the cGAN. The order of the scripts is detailed and the workings are briefly described. The link is: <https://github.com/adriaanzico/cGAN-for-snow>.

7. Appendix 1 – Conditional Generative Adversarial Network

The prediction of Landsat-8 imagery through environmental data is done using cGANs. A cGAN is an extension of a regular GAN, because the model can be conditioned on additional data for both the Generator (G) and the Discriminator (D). Isola et al. (2017) designed a cGAN for image-to-image translation (pix2pix). The network first learns how to map an output image from an input image, but also incorporate loss functions to maximise the training efficiency. G is trained in such a way that it will produce outputs which cannot be distinguished from the “ground truth” by D. D is trained on “real” and “fake” data to better distinguish between the two. The performance of G and D generally improve as the training steps increase. The pix2pix methodology, for training G, is highlighted in the *Figure 19*:

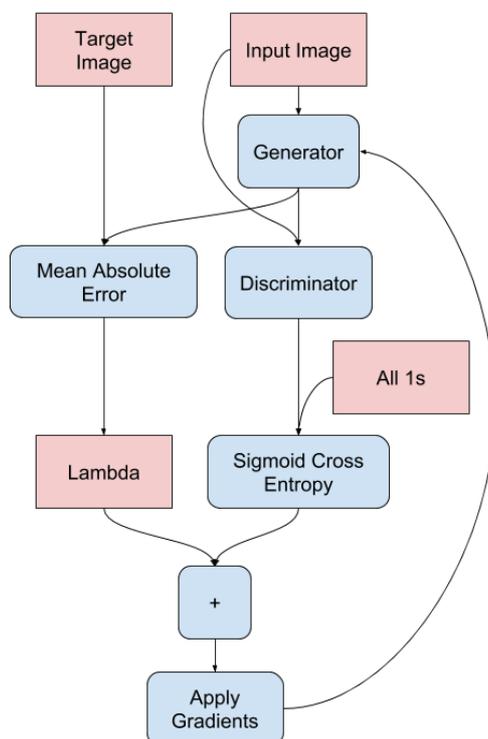


Figure 19 - Training procedure for generator (G). G is provided with an input image, in this case environmental statistics, and generated an image. This image is passed onto the discriminator (D), which determines the quality of G. Simultaneously, the mean absolute error (MAE) is calculated between the generated image and the true image. Lambda is set to 100 as recommended by the authors of the pix2pix paper. The generator loss is a sigmoid cross-entropy loss of the generated images and an array of ones. The lambda and the generator loss are added and then applied to the generator in the next training step. Image is from <https://www.tensorflow.org/tutorials/generative/pix2pix>.

The training of G and D occur simultaneously. The training of D, derived from the pix2pix paper, is shown in *Figure 20*:

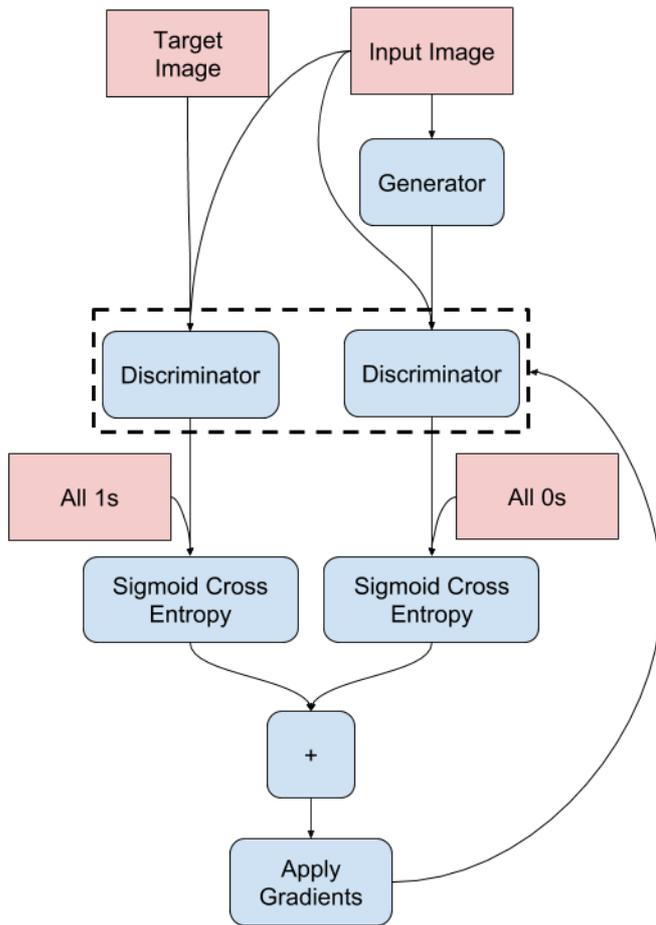


Figure 20 -Training procedure for discriminator (D). D is trained on two sets of images, the “true” image and the generated image, from generator (G). D then computes two loss functions: one for the “true” image and one for the “fake” image. the these sigmoid cross-entropy losses are then summed and returned to the generator for the next training step. Image is from <https://www.tensorflow.org/tutorials/generative/pix2pix>.

8. Appendix 2 – cGAN model summary

8.1. Generator

Model: "model"

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	[(None, 256, 256, 16)]	0	[]
sequential (Sequential)	(None, 128, 128, 64)	16384	['input_1[0][0]']
sequential_1 (Sequential)	(None, 64, 64, 128)	131584	['sequential[0][0]']
sequential_2 (Sequential)	(None, 32, 32, 256)	525312	['sequential_1[0][0]']
sequential_3 (Sequential)	(None, 16, 16, 512)	2099200	['sequential_2[0][0]']
sequential_4 (Sequential)	(None, 8, 8, 512)	4196352	['sequential_3[0][0]']
sequential_5 (Sequential)	(None, 4, 4, 512)	4196352	['sequential_4[0][0]']
sequential_6 (Sequential)	(None, 2, 2, 512)	4196352	['sequential_5[0][0]']
sequential_7 (Sequential)	(None, 1, 1, 512)	4196352	['sequential_6[0][0]']
sequential_8 (Sequential)	(None, 2, 2, 512)	4196352	['sequential_7[0][0]']
concatenate (Concatenate)	(None, 2, 2, 1024)	0	['sequential_8[0][0]', 'sequential_6[0][0]']
sequential_9 (Sequential)	(None, 4, 4, 512)	8390656	['concatenate[0][0]']
concatenate_1 (Concatenate)	(None, 4, 4, 1024)	0	['sequential_9[0][0]', 'sequential_5[0][0]']
sequential_10 (Sequential)	(None, 8, 8, 512)	8390656	['concatenate_1[0][0]']
concatenate_2 (Concatenate)	(None, 8, 8, 1024)	0	['sequential_10[0][0]', 'sequential_4[0][0]']

```

sequential_11 (Sequential) (None, 16, 16, 512) 8390656 ['concatenate_2[0][0]']
concatenate_3 (Concatenate) (None, 16, 16, 1024) 0 ['sequential_11[0][0]',
) 'sequential_3[0][0]']
sequential_12 (Sequential) (None, 32, 32, 256) 4195328 ['concatenate_3[0][0]']
concatenate_4 (Concatenate) (None, 32, 32, 512) 0 ['sequential_12[0][0]',
) 'sequential_2[0][0]']
sequential_13 (Sequential) (None, 64, 64, 128) 1049088 ['concatenate_4[0][0]']
concatenate_5 (Concatenate) (None, 64, 64, 256) 0 ['sequential_13[0][0]',
) 'sequential_1[0][0]']
sequential_14 (Sequential) (None, 128, 128, 64) 262400 ['concatenate_5[0][0]']
)
concatenate_6 (Concatenate) (None, 128, 128, 128) 0 ['sequential_14[0][0]',
) 'sequential[0][0]']
conv2d_transpose_7 (Conv2DTranspose) (None, 256, 256, 7) 14343 ['concatenate_6[0][0]']
)

```

```

=====
Total params: 54,447,367
Trainable params: 54,436,487
Non-trainable params: 10,880
=====

```

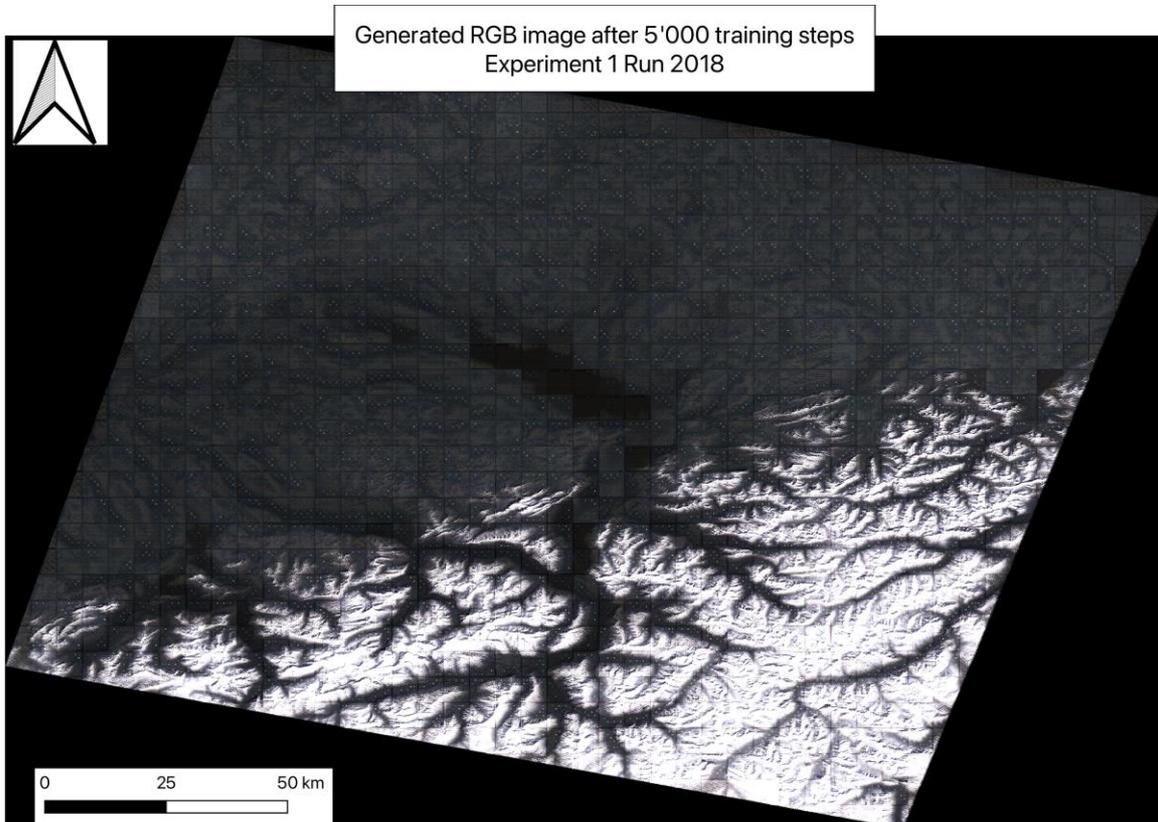
8.2. Discriminator

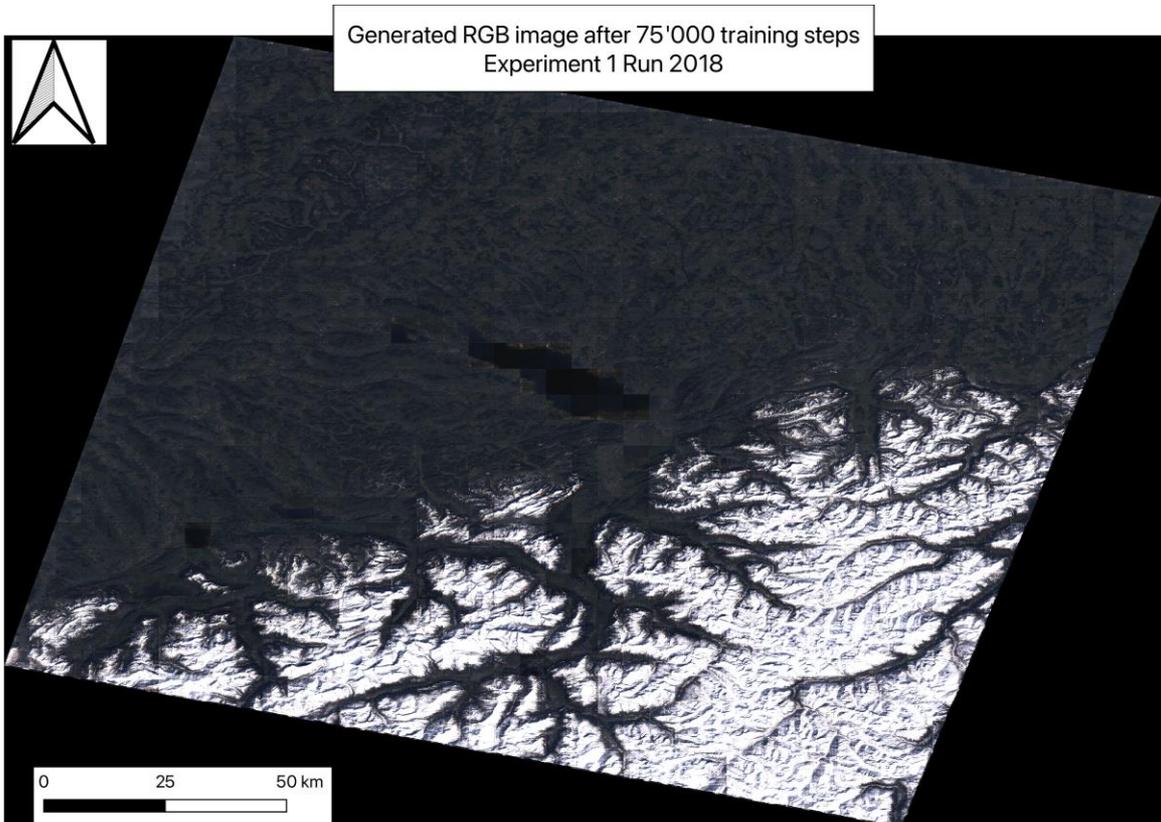
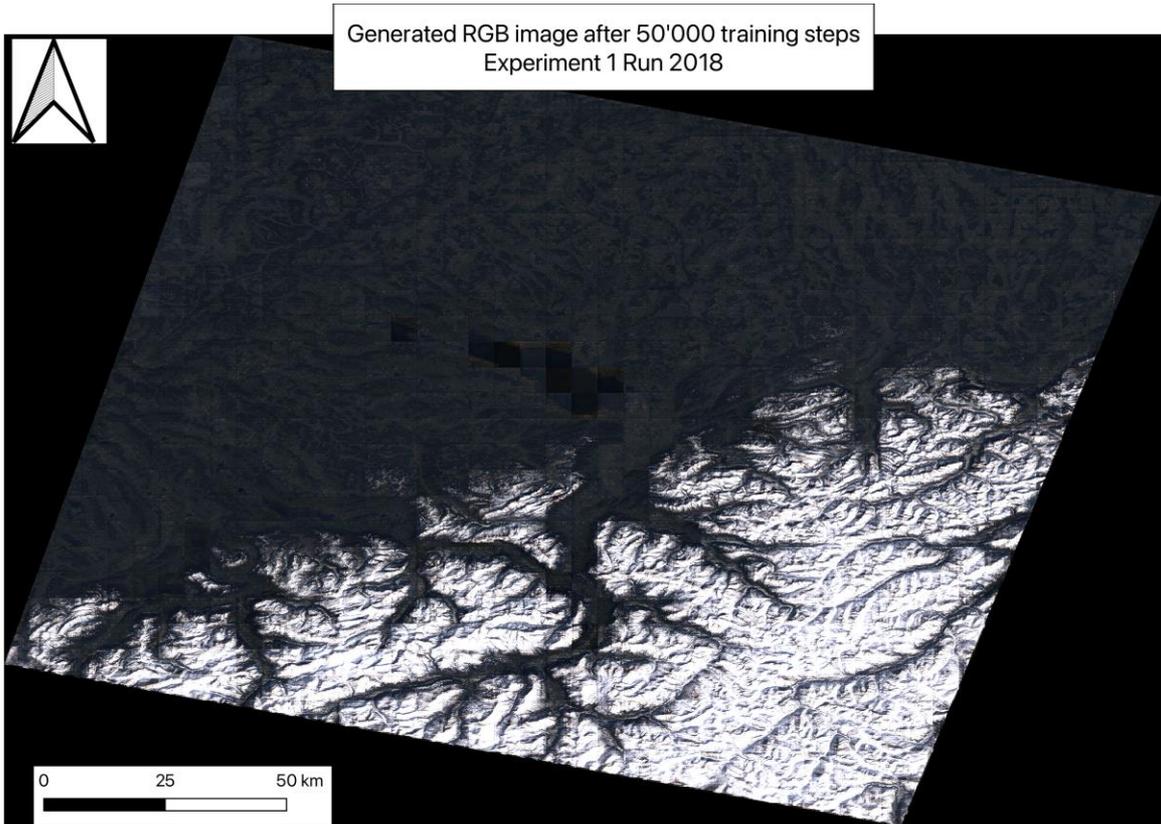
Layer (type)	Output Shape	Param #	Connected to
input_image (InputLayer)	[(None, 256, 256, 16)]	0	[]
target_image (InputLayer)	[(None, 256, 256, 7)]	0	[]
concatenate_7 (Concatenate)	(None, 256, 256, 23)	0	['input_image[0][0]', 'target_image[0][0]']
sequential_15 (Sequential)	(None, 128, 128, 64)	23552	['concatenate_7[0][0]']
sequential_16 (Sequential)	(None, 64, 64, 128)	131584	['sequential_15[0][0]']
sequential_17 (Sequential)	(None, 32, 32, 256)	525312	['sequential_16[0][0]']
zero_padding2d (ZeroPadding2D)	(None, 34, 34, 256)	0	['sequential_17[0][0]']

conv2d_11 (Conv2D) (None, 31, 31, 512) 2097152 ['zero_padding2d[0][0]']
batch_normalization_16 (Batch Normalization) (None, 31, 31, 512) 2048 ['conv2d_11[0][0]']
leaky_re_lu_11 (LeakyReLU) (None, 31, 31, 512) 0 ['batch_normalization_16[0][0]']
zero_padding2d_1 (ZeroPadding2D) (None, 33, 33, 512) 0 ['leaky_re_lu_11[0][0]']
conv2d_12 (Conv2D) (None, 30, 30, 1) 8193 ['zero_padding2d_1[0][0]']

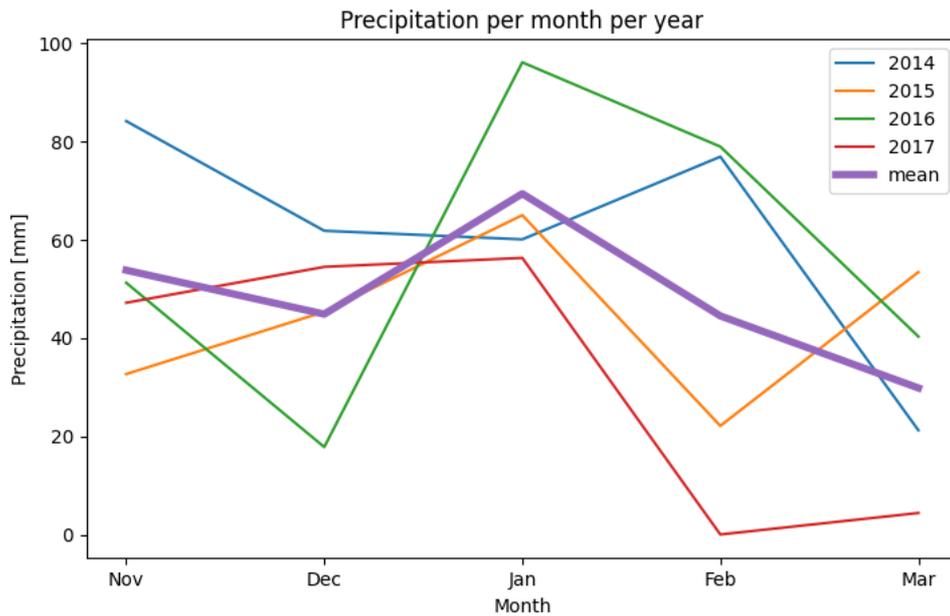
=====
Total params: 2,787,841
Trainable params: 2,786,049
Non-trainable params: 1,792

9. Appendix 3 – Generated Landsat-8 images Experiment 1 Run 2018



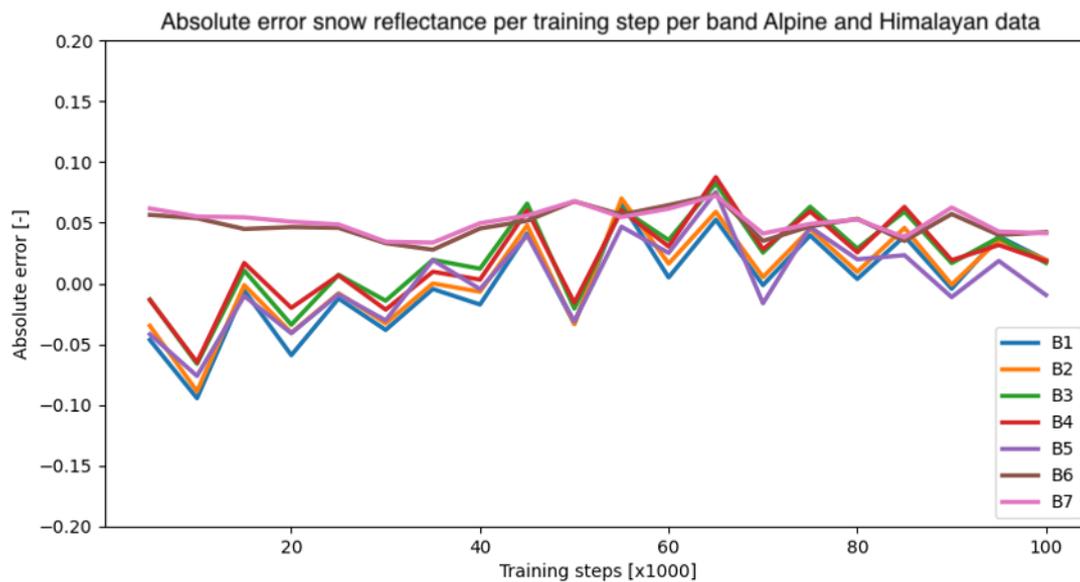


10. Appendix 4 – Precipitation per month in Alpine training set



This figure shows that January received the most precipitation, on average, in the training set. This may be the reason why the cGAN is biased towards giving January the most weight when adding precipitation to a certain month.

11. Appendix 5 – Generated reflectance when training on both Alpine and Himalayan data



The figure above shows the effect of using both Alpine and Himalayan training data. The True error of the snow reflectance per training step converges to 0 after 100'000 training steps. The exact reason for this is unknown, but it may be attributed to using more training data in the training set (double the amount of data than what has been used in Experiment 1 or Experiment 2).

12. References

- Adnan, N., & Umer, F. (2022). Understanding deep learning-challenges and prospects.
- Agrawala, S. (2007). Climate change in the European Alps: adapting winter tourism and natural hazards management. Organisation for Economic Cooperation and Development (OECD).
- Andreas Juergen Dietz, Claudia Kuenzer, Ursula Gessner & Stefan Dech (2012) Remote sensing of snow – a review of available methods, *International Journal of Remote Sensing*, 33:13, 4094-4134, DOI: 10.1080/01431161.2011.640964
- Antoniadis, P. (2023, March 16). Activation Functions: Sigmoid vs Tanh. Baeldung. <https://www.baeldung.com/cs/sigmoid-vs-tanh-functions>
- Beniston, M. (2012), Is snow in the Alps receding or disappearing?. *WIREs Clim Change*, 3: 349-358. <https://doi.org/10.1002/wcc.179>
- do Lago, C. A., Giacomoni, M. H., Bentivoglio, R., Taormina, R., Junior, M. N. G., & Mendiondo, E. M. (2023). Generalizing rapid flood predictions to unseen urban catchments with conditional generative adversarial networks. *Journal of Hydrology*, 618, 129276.
- Elmes, A., Alemohammad, H., Avery, R., Caylor, K., Eastman, J. R., Fishgold, L., ... & Estes, L. (2020). Accounting for training data error in machine learning applied to Earth observations. *Remote Sensing*, 12(6), 1034.
- Foster, J., Owe, M., & Rango, A. (1983). Snow cover and temperature relationships in North America and Eurasia. *Journal of Applied Meteorology and Climatology*, 22(3), 460-469.
- Foster, J. L., Hall, D. K., & Chang, A. T. C. (1987). Remote sensing of snow. *Eos, Transactions American Geophysical Union*, 68(32), 682-684.
- Gobiet, A., Kotlarski, S., Beniston, M., Heinrich, G., Rajczak, J., & Stoffel, M. (2014). 21st century climate change in the European Alps—A review. *Science of the total environment*, 493, 1138-1151.
- Green, R., Dozier, J., Roberts, D., & Painter, T. (2002). Spectral snow-reflectance models for grain-size and liquid-water fraction in melting snow for the solar-reflected spectrum. *Annals of Glaciology*, 34, 71-73. doi:10.3189/172756402781817987
- Hall, D.K., Frei, A. and Déry, S.J. (2015). Remote sensing of snow extent. In *Remote Sensing of the Cryosphere*, M. Tedesco (Ed.). <https://doi.org/10.1002/9781118368909.ch3>
- Isola, P., Zhu, J. Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1125-1134).
- Jain, S. K., Goswami, A., & Saraf, A. K. (2009). Role of elevation and aspect in snow distribution in Western Himalaya. *Water resources management*, 23, 71-83.
- Jain, S. K., Goswami, A., & Saraf, A. K. (2009). Role of elevation and aspect in snow distribution in Western Himalaya. *Water resources management*, 23, 71-83.
- Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
- L. Taylor and G. Nitschke, "Improving Deep Learning with Generic Data Augmentation," 2018 IEEE Symposium Series on Computational Intelligence (SSCI), Bangalore, India, 2018, pp. 1542-1547, doi: 10.1109/SSCI.2018.8628742.
- Landsat 8. (n.d.). U.S. Geological Survey. <https://www.usgs.gov/landsat-missions/landsat-8>

- LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* 521, 436–444 (2015).
<https://doi.org/10.1038/nature14539>
- Li, X., Cheng, G., Jin, H., Kang, E., Che, T., Jin, R., ... & Shen, Y. (2008). Cryospheric change in China. *Global and Planetary Change*, 62(3-4), 210-218.
- Luongo, F., Hakim, R., Nguyen, J. H., Anandkumar, A., & Hung, A. J. (2021). Deep learning-based computer vision to recognize and classify suturing gestures in robot-assisted surgery. *Surgery*, 169(5), 1240-1244.
- Millard, K., & Richardson, M. (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), 8489-8515.
- Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.
- Mohanty, S. P., Hughes, D. P., & Salathé, M. (2016). Using deep learning for image-based plant disease detection. *Frontiers in plant science*, 7, 1419.
- Namias, J. (1985). Some Empirical Evidence for the Influence of Snow Cover on Temperature and Precipitation, *Monthly Weather Review*, 113(9), 1542-1553. Retrieved Feb 12, 2023, from https://journals.ametsoc.org/view/journals/mwre/113/9/1520-0493_1985_113_1542_seefti_2_0_co_2.xml
- Nolin, A. (2010). Recent advances in remote sensing of seasonal snow. *Journal of Glaciology*, 56(200), 1141-1150. doi:10.3189/002214311796406077
- O'Neill, B. C., Tebaldi, C., Van Vuuren, D. P., Eyring, V., Friedlingstein, P., Hurtt, G., ... & Sanderson, B. M. (2016). The scenario model intercomparison project (ScenarioMIP) for CMIP6. *Geoscientific Model Development*, 9(9), 3461-3482.
- P. P. Shinde and S. Shah, "A Review of Machine Learning and Deep Learning Applications," 2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA), Pune, India, 2018, pp. 1-6, doi: 10.1109/ICCUBEA.2018.8697857
- pix2pix: Image-to-image translation with a conditional GAN. (2022, December 20). TensorFlow Core. <https://www.tensorflow.org/tutorials/generative/pix2pix>
- Rebetz, M. (1996). Seasonal relationship between temperature, precipitation and snow cover in a mountainous region. *Theoretical and applied climatology*, 54, 99-106.
- Requena-Mesa, C., Reichstein, M., Mahecha, M., Kraft, B., Denzler, J. (2019). Predicting Landscapes from Environmental Conditions Using Generative Networks. In: Fink, G., Frintrop, S., Jiang, X. (eds) *Pattern Recognition. DAGM GCPR 2019. Lecture Notes in Computer Science()*, vol 11824. Springer, Cham. https://doi.org/10.1007/978-3-030-33676-9_14
- Rodríguez-Suárez B, Quesada-Barriuso P, Argüello F. Design of CGAN Models for Multispectral Reconstruction in Remote Sensing. *Remote Sensing*. 2022; 14(4):816.
<https://doi.org/10.3390/rs14040816>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18* (pp. 234-241). Springer International Publishing.
- Rozet, A., Kronish, I. M., Schwartz, J. E., & Davidson, K. W. (2019). Using machine learning to derive just-in-time and personalized predictors of stress: observational study bridging the gap between nomothetic and ideographic approaches. *Journal of medical Internet research*, 21(4), e12910.

- Sharma, V., Mishra, V. D., & Joshi, P. K. (2014). Topographic controls on spatio-temporal snow cover distribution in Northwest Himalaya. *International Journal of Remote Sensing*, 35(9), 3036-3056.
- Sun, C., Shrivastava, A., Singh, S., & Gupta, A. (2017). Revisiting unreasonable effectiveness of data in deep learning era. In *Proceedings of the IEEE international conference on computer vision* (pp. 843-852).
- tf.keras.optimizers.Adam. (n.d.). TensorFlow v2.12.0.
https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam
- Veyret, P. , Diem, . Aubrey and Poulsen, . Thomas M. (2023, February 5). Alps. *Encyclopedia Britannica*. <https://www.britannica.com/place/Alps>
- Wang, Y., Pan, X., Song, S., Zhang, H., Huang, G., & Wu, C. (2019). Implicit semantic data augmentation for deep networks. *Advances in Neural Information Processing Systems*, 32.
- Wunderle S, Gross T, Hüsler F. Snow Extent Variability in Lesotho Derived from MODIS Data (2000–2014). *Remote Sensing*. 2016; 8(6):448. <https://doi.org/10.3390/rs8060448>
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., ... & Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature climate change*, 3(10), 875-883.
- Yang, J., Jiang, L., Ménard, C. B., Luo, J., Lemmetyinen, J., & Pulliainen, J. (2015). Evaluation of snow products over the Tibetan Plateau. *Hydrological Processes*, 29(15), 3247-3260.
- Yuan, Q., Shen, H., Li, T., Li, Z., Li, S., Jiang, Y., ... & Zhang, L. (2020). Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sensing of Environment*, 241, 111716.
- Zeiler, M. D. (2012). Adadelta: an adaptive learning rate method. *arXiv preprint arXiv:1212.5701*.
- Zender, C. Snowfall brightens Antarctic future. *Nature Clim Change* 2, 770–771 (2012).
<https://doi.org/10.1038/nclimate1730>
- Zhang, T. (2005), Influence of the seasonal snow cover on the ground thermal regime: An overview, *Rev. Geophys.*, 43, RG4002, doi:10.1029/2004RG000157.
- Zhang, X., Li, X., Li, L. et al. Environmental factors influencing snowfall and snowfall prediction in the Tianshan Mountains, Northwest China. *J. Arid Land* 11, 15–28 (2019). <https://doi.org/10.1007/s>