Precipitation and Erosion Threshold for Quick Clay Landslides Using Machine Learning Models

Jakob Brysting Kristiansen

Master Thesis
Geomorphology and Geomatics
60 credits

Department of Geosciences
The Faculty of Mathematics and Natural Sciences

UNIVERSITY OF OSLO

June / 2023

Precipitation and Erosion Threshold for Quick Clay Landslides Using Machine Learning Models

Jakob Brysting Kristiansen

Precipitation and Erosion Threshold for Quick Clay Landslides Using Machine Learning Models

http://www.duo.uio.no/

Printing: Reprosentralen, Universitetet i Oslo

Abstract

Quick clay landslides may have major consequences on human civilization and are responsible for some of the most damaging natural disasters in Norway. Quick clay hazard zones in Norway are mapped as either low-, medium-, or high-hazard, but are not updated based on any regular monitoring. The main goal of the study was to investigate if precipitation and erosion data gathered from previous quick clay landslides in Oslo and Viken in southeastern Norway could be used to create a threshold for when conditions of quick clay reach the critical levels that may trigger a landslide event. The ultimate goal is that these thresholds can be used as part of a method to periodically update the hazard zone categories. The method used to create thresholds was based on machine learning models, including two different ensemble models (RUSBoosted decision trees and Bagged decision trees) and two Support Vector Machine (SVM) models (Cubic and Quadratic kernels). The models' ability to classify landslides correctly were evaluated using area under the receiver operating characteristics curve (AUC) and confusion matrix to measure the false-negative and the false-positive rates. The results showed that the Quadratic SVM model and the Bagged decision trees had the highest AUC (0.74, 0.85) and the lowest false-negative rate (54.5 %, 63.6 %) of the models trained when the models were trained with a combination of precipitation and erosion data. Training with only precipitation data did not change the results much, though a minor improvement was seen when including erosion data. The high false-negative rates suggest that the method as used here is unsuitable as part of a monitoring system. One major problem is lack of erosion data, and an improvement of the method will probably require yearly gathering of erosion data in addition to testing of other predictor variables.

Preface

Det er flere jeg vil takke nå som oppgaven er ferdig. Først en takk til veilederne mine, Elin Skurtveit for gode tilbakemelding på oppgaven, og Jose Cepeda som hjalp med å finne ut av hva oppgaven skulle omhandle og hvordan metoden kunne utvikles.

En takk til mine medstudenter gode diskusjoner og et godt arbeidsmiljø på kontoret. Spesielt til Stewart for våre diskusjoner om alt, både faglig og ikke. Jeg vil også gi en takk til Vegard som jeg jobbet med for å lage data til oppgavene våre.

Og til slutt en stor takk til familien min for støtte under studietiden og arbeidet med masteroppgaven.

Contents

List of Figures

List of Tables

1 Introduction

Quick clay can be found in large parts of Norway and Sweden and also exists in Finland, Canada and Alaska (NGI, 2023). In Norway it is found along the coastal areas, with the main areas of quick clay in Østlandet and Trøndelag (Figure 1). Quick clay landslides can have large consequences as was recently seen in the quick clay landslide in Gjerdrum where 11 people lost their lives and more than 1600 people had to be evacuated from their homes. The landslide also caused extensive damages that cost almost 2 billion Norwegian kroner in damages and reparations after the event (NOU, 2022).

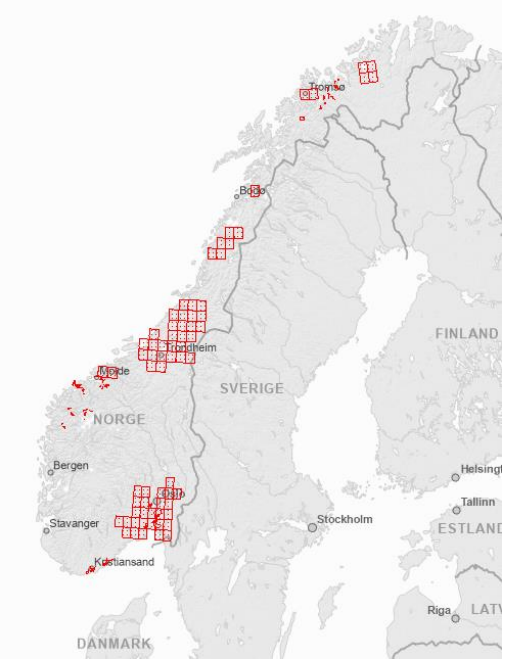


Figure 1 - Mapped quick clay areas in Norway are shown in red. Figure from NVE Atlas: https://atlas.nve.no/

Triggering factors for quick clay landslides in recent years are mainly caused by anthropic activity (L'Heureux et al., 2018) but also from natural erosion. Erosion in riverbeds is caused by wet seasons and years of high precipitation with high water flow in rivers. An increase in precipitation and runoff as a result of climate change is also expected (IEA, 2022) which could lead to an increase in erosion and erosion-triggered quick clay landslides.

Although quick clay areas are categorized and evaluated for their danger in Norway, there is no warning system similar to the one that exists for landslides (Krøgli et al., 2018) and the hazard zones categories remain static after they are set. The hazard zones are not updated based on any regular monitoring, although it can be updated based on new investigations in the hazard zone this is not part of a routine or periodic assessment (NVE, 2020b). A way to create warning systems are through rainfall thresholds, which have been used for warning of rainfall induced landslide for a while since the concept was introduced by (Endo, 1969). In later years, several types of thresholds have been suggested together with reviews of advances and issues with these thresholds (Segoni et al., 2018).

There is not much literature on rainfall thresholds related to quick clay landslides, but Gauthier & Hutchinson (2012) attempted to create a threshold based on cumulative antecedent precipitation for intervals ranging from one day to 365 days. While some of the

landslide events that they looked at had a high correlation with an antecedent precipitation interval none of the events had a high correlation with the same antecedent precipitation interval even if the events were spatially and temporarily related.

1.1 Aims

The main goal of this master thesis is to use machine learning algorithms to create a threshold based on precipitation and erosion data for when quick clay landslides release.

This will be obtained through:

- Collecting precipitation and erosion data from previous quick clay events.
- Creating a more complete quick clay landslide inventory than the one that can be downloaded from The Norwegian Water Resources and Energy Directorate (NVE).
- Developing a method with the machine learning tools in MATLAB to create thresholds for when the quick clay landslide trigger.
- Evaluating whether the thresholds can be used in a method to periodically to update the categories of the quick clay hazard zones.

1.2 Contributions

The method used in this thesis is developed together with Vegard Rogstad Kvalvær and one of our supervisors Jose Cepeda. Vegard is testing the method in Trøndelag for his thesis to get a comparison of how the method works in different areas in Norway.

The work on developing the method was split between Vegard and me, each of us focusing on one part of the method. Vegard worked on the erosion data, and I worked on the precipitation data. The work was done individually but during the process we had regular meetings with our supervisor to discuss how we wanted to complete our task and develop the method.

For the quick clay inventory, we split the work into our study areas, Vegard having the main responsibility for quality control of Trøndelag in Mid-Norway and I having the main responsibility for the quality control of the quick clay landslides in Oslo and Viken in southeastern Norway.

1.3 Software used in thesis.

ArcGIS pro

I used ArcGIS Pro v. 3.1.0 for exploratory analysis of the quick clay inventory from NVE and to create the new inventory with additional quick clay landslides from L'Heureux & Solberg, (2012) and NGI (2011). It was also used as part of the quality control of our quick clay inventory, checking for duplicates, wrongly categorized slides, wrong dates or positions, and for creating maps showing the study area and overlays of marine limit and population density maps. ArcGIS pro was also used for the erosion analyses.

MATLAB

I used MATLAB R2021a Update 7 (9.10.0.2015706) for exploratory analysis of the precipitation data and for creating scripts for statistical analysis. I used the Classification Learner App with the machine learning tools to create the quick clay landslide thresholds.

2 Theory

2.1 Characteristics of quick clay

Quick clay can be found in places that are lower than the marine limit, which is the height of the land that was depressed under water during the ice age in Norway, and because of the postglacial land uplift the marine clay is now on land. The marine limit ranges, based on where in Norway, from 0 to 220 meters above sea level (NGU, 2021).

The classification of a quick clay is in Norway based on the sensitivity of the soil. A geotechnical definition is clays that have a remoulded shear strength less than or equal to 0.5 kPa (kN/m^2). The NVE is more conservative than the geotechnical definition and classifies all brittle materials in surficial deposits, clays or silt that have a possibility of being an area landslide as materials with a remoulded shear strength less than or equal to 2 kPa (Gjerdrumutvalget, 2021).

2.1.1 Formation of Quick clay

After the ice melted and the land rose at the end of the ice age, clay deposits that were previously under sea and had formed unstable structures with salt were now above sea level. As the salt is washed out over hundreds to thousands of years, “quick clay” or “sensitive clay” is formed (L’Heureux, 2013).

The clay in saltwater forms an open card house like structure (Figure 2), where the edges and planes have a different charge and are attracted to each other. The salt contributes to the binding forces keeping the card house structure stable. As the salt is washed out, the card house structure stays the same, but the binding forces are weakened (Gjerdrumutvalget, 2021).

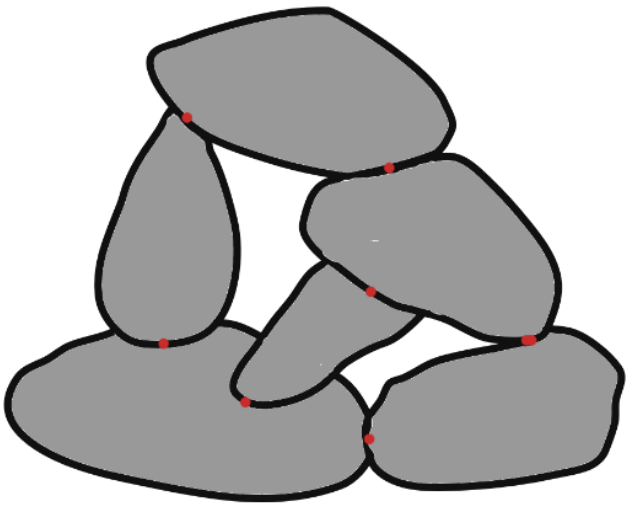


Figure 2 - Schematic illustration of quick clay card house structure. Red points indicate attraction points between clay flakes which keep the structure stable. As salt is washed out the card house structure stays stable, but the binding forces are weakened. When overloaded the structure will collapse leading to the liquification that can be seen in quick clay landslides.

2.2 Types of quick clay landslides

According to the NVE guideline safety against quick clay landslides (NVE, 2020b) the main slides for quick clay are retrogressive slides, rotational slides and flake slides.

These types of slides can happen individually but are often seen in the same quick clay landslide event, as was the case in the slide in the 1978 Rissa slide (Gregersen, 1981) where the a small initial slide developed into retrogressive slides that triggered large flake slides.

2.2.1 Retrogressive slide

Large quick clay landslides are often retrogressive (L'Heureux, 2012). The initial slide creates an unstable back scarp which can fail and then retrogressively fail until there is a stable back scarp. The retrogressively failing slope can lead to the material rapidly flowing down the slope (Nigussie, 2013). Figure 3 shows the principle of a retrogressive slide, starting from an initial slide caused by erosion in the river and retrogressively expanding up the slope.

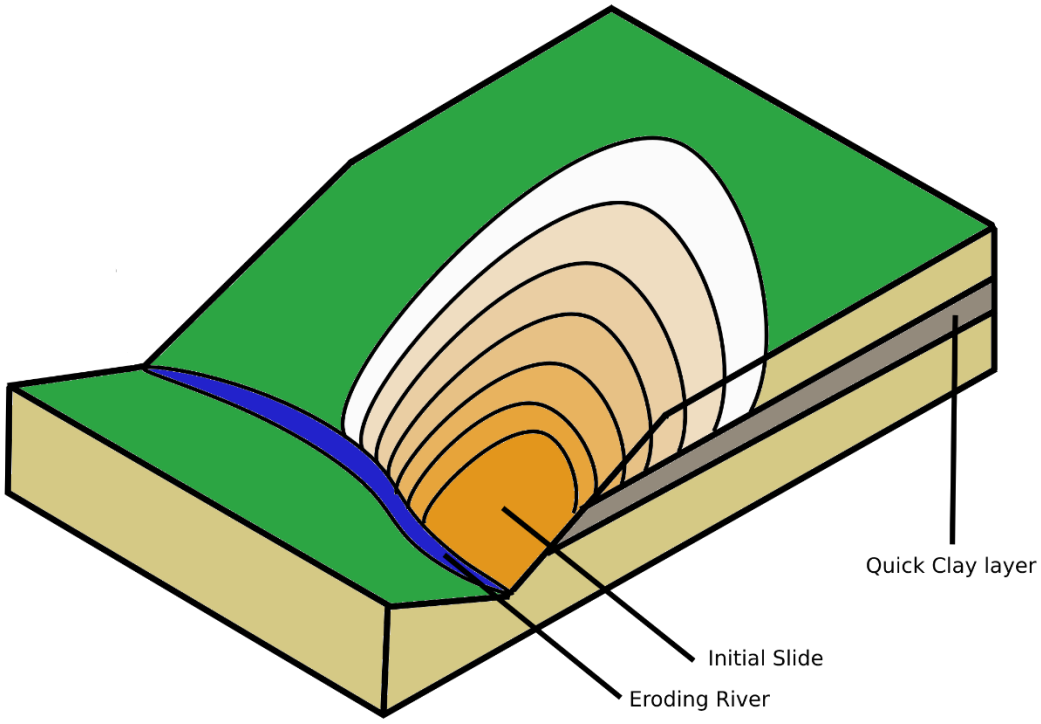


Figure 3 - Example of a retrogressive quick clay landslide where erosion in the river has weakened the hillside leading to an initial slide that retrogressively expands up the slope. Figure modified from (NVE, 2020b).

2.2.2 Rotational slide

A rotational slide is a slide with a curved slip surface that usually moves as a relatively coherent mass and shows a clear rotation (Figure 4). The size of the slide is limited by the height of the slope and limitations in the terrain can help prevent further retrogression.

In rotational slides the mass does not liquefy except for at the glide plane (Issler et al., 2012). In the case where the rotational slide triggers further slides, it is referred to as the initial slide (NVE, 2020b).

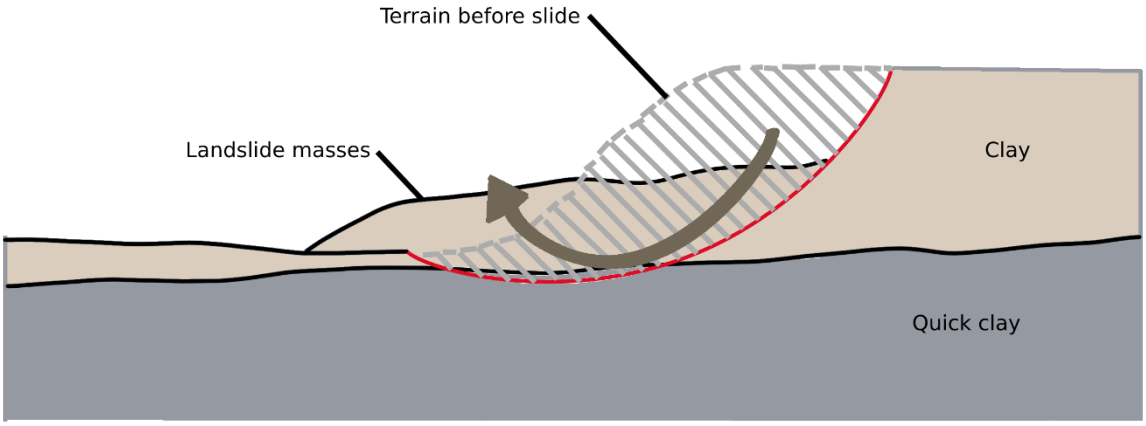


Figure 4 – Example of a rotational quick clay landslide showing rotational movement of the mass. Figure modified from NVE (2020b).

2.2.3 Flake slide

Flake slides happen when large flakes in gentle slopes release and quickly break up into large chunks or disintegrate (Torrance, 2012). This can occur in relatively thin layers of quick clay that collapse and liquefy because of loading at the top of a slope or cuts or erosion at the bottom of the slope leading to a progressive failure along the quick clay layer releasing the flake. Figure 5 shows a flake slide being triggered by addition load at the top of a slope leading to progressive failure in the quick clay layer causing a flake to release.

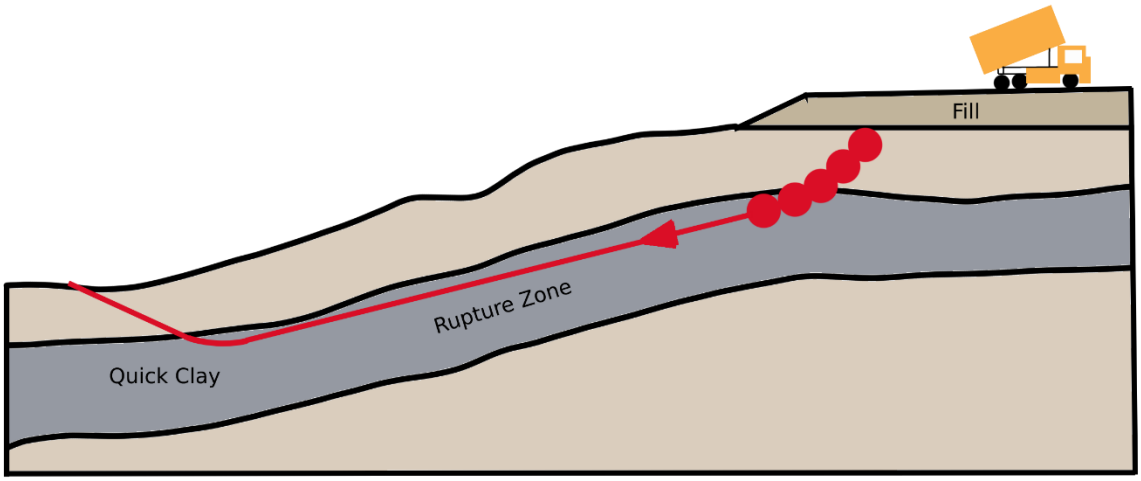


Figure 5 – Example of a flake quick clay landslide, which is triggered by loading on the top of the hill and then undergoes a progressive failure. Figure modified from (NVE, 2020b).

2.3 Triggers of quick clay landslides

The two main triggering factors of quick clay landslides can be divided into anthropic factors (filling, excavation, construction activities, urbanisation) or natural causes, mainly erosion destabilizing the slope. During the last 70 years more than 50 % of the quick clay landslides were because of anthropic factors and as much as 90 % since 2010 (L'Heureux et al., 2018). In the disaster in Gjerdrum one of the causes of increased erosion in the river was additional runoff from the urbanisation in the catchment (Gjerdrumutvalget, 2021), exemplifying how quick clay landslides can result from a mix of anthropic factors and natural factors. The autumn season in 2020 was the wettest season in Gjerdrum since autumn 2000, and part of the conclusion for why the quick clay landslide released in 2020 and not in 2000 was the effect of the erosion over the years reducing the stability of the slope.

A consideration for the future with increased extreme weather and precipitation caused by climate change (Hanssen-Bauer et al., 2016) is the increased water flow in rivers leading to an increase in erosion and whether this will again increase the amount of quick clay landslides triggered by erosion.

2.4 Hazard zones

After the 1978 Rissa quick clay landslide a national quick clay mapping project was started, the method was based on a criteria for the topography based partly on theory and partly on an analysis of earlier quick clay landslides (NVE, 2020a).

- Limited to terrain with height differences of at least 10m from bottom of rivers/sea.
- Slopes steeper than 1:15 and coastal under water slopes steeper than 1:6.
- Release area maximal length equivalent 15 x height difference on land and 6 x height coastal underwater slopes in the sea.
- Release area width is only limited by distance to more stable topography.

In local projects more consideration must be taken as there can be landslides in less critical conditions.

The hazard zones are classified in three classes, “low-hazard”, “middle-hazard” and “high-hazard” and are given a score based the on probability of a quick clay landslide according to topographical conditions, geological conditions and terrain changes including both natural and manmade triggering causes. The zones are also divided into three consequence classes “less-severe”, “severe” and “very-severe”. Based on the hazard and consequence the zones are divided into risk class 1 to 5 with lowest risk in class 1 and highest in class 5 (NVE, 2020a).

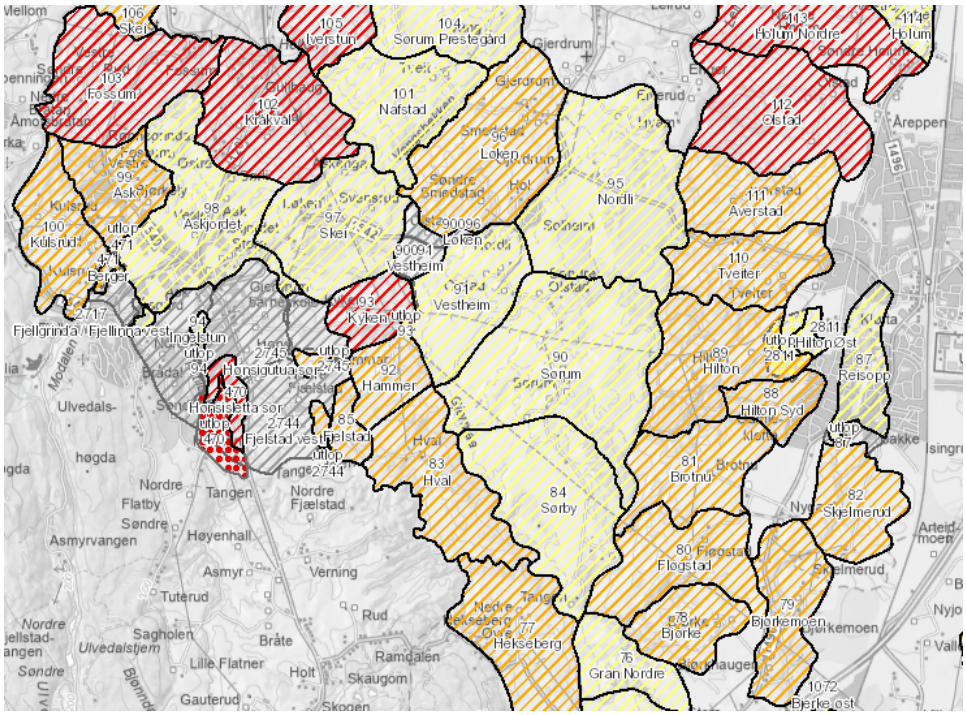


Figure 6 - Map showing examples of quick clay hazard zones in Romerike, SE Norway, from NVE Atlas (atlas.nve.no). The yellow zones are low-hazard, the orange zones are middle-hazard and the red zones are high-hazard

2.5 Machine Learning Algorithms

Machine learning methods have been used for susceptibility mapping and early warning (Collini et al., 2022; Kuradusenge et al., 2020; Li et al., 2020; Vallet et al., 2016) though more commonly used in landslide susceptibility mapping than for creating rainfall thresholds and landslide prediction.

2.5.1 Ensemble Models

RUSBoosted decision trees and Bagged decision trees are ensemble models, which combines results of several more or less successful models and not only keeping the "best" model because the less successful models might have valuable information that could improve the ability to create accurate predictions. The ensemble model also helps guard against failure in individual models (Baker & Ellison, 2008).

Random Undersampling (RUS) Boosted decision trees

The RUSBoost algorithm is a model that helps with class imbalance, through undersampling, removing examples from the majority class, and boosting which improves the performance of a weak classifier. It was introduced by Seiffert et al. (2010) based on the SMOTEBoost algorithm.

Landslide data is often imbalanced, as there are many years with no landslides, so a model like RUSBoost seems like a good fit, however this model may also create many false positives (Xiao et al., 2022) which would not work for a traditional landslide warning system.

Bagged decision trees

Bagging or Bootstrap aggregating introduced by Breiman (1996), creates an ensemble of classifiers by selecting random samples of data from the dataset with replacements which means the same data can be chosen for different samples (IBM, n.d.).

An advantage of the bagging method is that it can compensate for overfitting (Berk, 2006) however the model can be computationally expensive.

Formulas and the details of how the ensemble models work can be found in MATLAB's Classification Learner app from the help text documentation on Statistics and Machine Learning Toolbox, Classification Ensembles (The MathWorks Inc., 2023a).

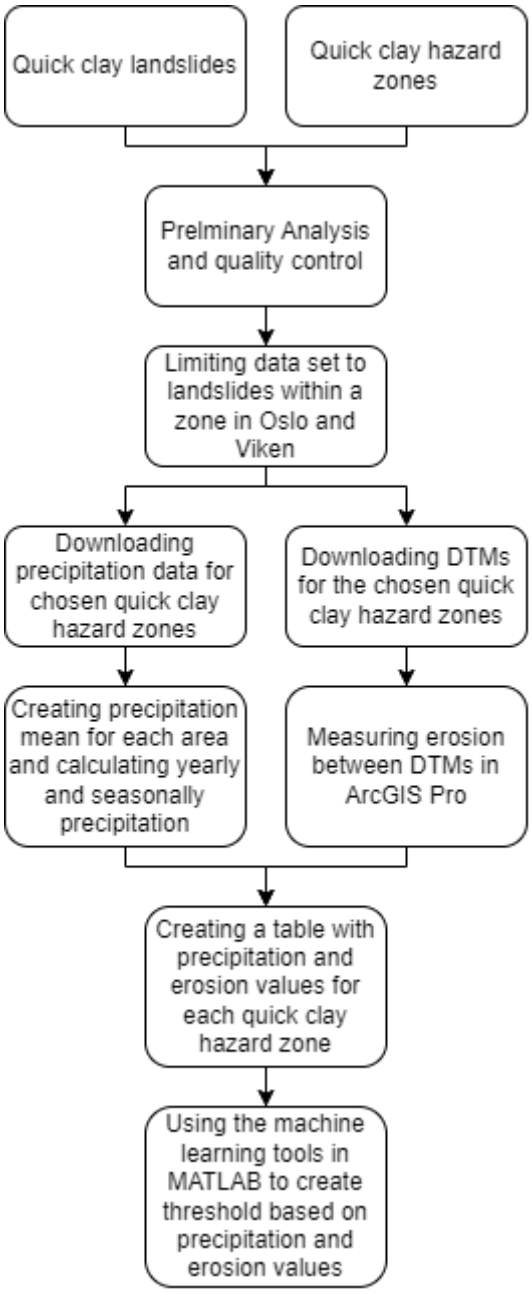
2.5.2 Support Vector Machine (SVM)

SVMs finds the best boundary or hyperplane to sperate the data into different classes. Depending on the kernel the hyperplane can be linear or non-linear. The kernel is a function that transforms the input data into a higher dimensional space where the data becomes linearly separable (Boswell, 2002).

The Linear, Cubic and Quadratic SVMs in MATLAB uses different kernels, linear for Linear, and polynomial for Cubic and Quadratic (Ekız & Erdoğmuş, 2017). The formulas and functions of the models can be found in the Classification Learner app available in the MATLAB help text for Support Vector Machine Classification (The MathWorks Inc., 2023c).

3 Method

The goal of the thesis was to develop a method to create thresholds based on precipitation and erosion for when quick clay landslides trigger, and to evaluate if that threshold could be used in a process to periodically update the hazard categorizes of the quick clay zones. The flow chart below shows the main steps of the procedure:



3.1 Study area

Oslo and Viken

In Norway over 100 000 people live on registered quick clay areas (NVE, 2022), many of which live in the counties Oslo and Viken (Figure 7) which has a population of around 2 million. Areas that have the possibility for quick clay to have been formed are all under what is called the marine limit. Large, populated areas in Oslo and Viken are below the marine limit (Figure 8).

In Oslo and Viken over 1000 quick clay hazard zones are registered, and we have seen large quick clay landslides both historically and recently. The most recent is the disaster in December 2020 in Ask, Gjerdrum where 11 people lost their lives (Gjerdrumutvalget, 2021).

The 23 events used in the exploratory precipitation analysis were the naturally triggered quick clay landslides inside hazard zones in Oslo and Viken (Table 1,Figure 7).

Table 1 - Quick clay landslide events selected for the exploratory precipitation analysis, showing the date of the event, the hazard zone and the name of the event. All selected quick clay landslide events in Oslo and Viken were within a hazard zone with a buffer of 35 m.

Date	Hazard Zone	Event Name
13.05.1823	321 Gullaug	Gullaug 2
06.09.1890	2606 Torsbekk	Utrasingen av melkefabrikken på Sannesund
20.10.1924	107 Kogstad	Kankedalen
17.04.1925	281 Årum	Gretnes
29.04.1927	1785 Brubakkveien Sør	Lodalen 2
22.12.1953	43 Borgen	Borgen
01.12.1954	353 Katterrud	Ilangskogen
03.05.1955	2435 Rolvøysund Vest	Rolvsøy
20.03.1967	77 Hekseberg	Hekseberg
10.04.1967	530 Nordby	Nordre Nordby
11.09.1984	485 Smørgrav	Strandajordet, Øvre Eiker
15.07.1999	1384 Kåbbel Nordre	Kåbøl, Våler
15.07.2000	1323 Fossnes	Hvittingfoss, Kongsberg
19.11.2000	79 Bjørkemoen	Frogner, Lillestrøm
01.01.2011	782 Foss	leirskred øst for Nesveien

01.01.2012	782 Foss	Nesveien, Skiptvedt
01.06.2012	89 Hilton	Tveiter, Gjerdrum
24.09.2012	1383 Kåbbel Søndre	Hobølelva, Våler
09.11.2012	534 Vestby	Båhus, Nannestad
20.04.2016	394 Bøler	Søndre Rotnes, Årnes
01.03.2020	517 Ånåsrud Nord	Leirbekken 1, Nannestad
17.12.2020	521 Nygård	Leirbekken 2, Nannestad
30.12.2020	470 Hønsisletta	Ask, Gjerdrum

Figure 7 - Map with the study area Oslo and Viken (Se Norway) outlined in black, with red dots indicating quick clay landslides used in the analysis. Background Map: Topografisk Norgeskart 4.

Figure 8 - Map illustrating which part of the study area (Figure 7) is below the marine limit, with red dots indicating quick clay landslides used in the analysis. Background map: Topografisk Norgeskart 4.

3.2 Data collecting and pre-processing

The data was collected from different databases, as open data available from webpages owned by organizations run by the Norwegian state (Table 2).

Table 2 - Overview of the data used for the method with information about producers and webpages from where the data were downloaded.

The Norwegian Water Resources and Energy Directorate (Norges vassdrags- og energidirektorat, NVE) • **Precipitation data from nve.api.no** • **Initial landslide database and hazard zones from nedlasting.nve.no/gis**
The Norwegian Mapping Authority (Kartverket) • **Digital Terrain Models (DTM) from hoydedata.no** • **GIS data for the analysis (background maps, county boundaries, marine limit) from geonorge.no which is a part of the "Norge Digitalt" cooperation.**

A quality check was done on the initial landslide database to identify landslides that were categorized wrong or had the wrong date and position in the database and to remove duplicates.

I filtered out landslides to only include those in Oslo and Viken, and only those that are triggered by natural causes. I chose a date for which landslides to extract for the analysis based on availability of the DTMs, and I limited the dataset to only hazard zones where landslides had occurred within the timeframe that fits with the erosion data from the DTMs.

3.3 Quick clay inventory

To ensure that all quick clay events were available for the analysis, a quick clay inventory was needed. While NVE has a quick clay data base it is not complete, missing some older events and with duplicates of reported events from different organizations. Therefore, a quality control was needed to make sure the events were not wrongly categorized or had wrong position or dates. The quality control mostly consisted of tracking down official reports, newspapers other articles and theses on the events and checking dates and position as well as whether the events had quick clay in the area and wasn't misclassified.

The new quick clay inventory was created based on the NVE existing database of quick clay events and in the process of quality control further quick clay landslides were added from L'Heureux & Solberg (2012) and NGI (2011). After a final check of the inventory for duplicate landslides another inventory was created for the analysis that only included landslides that were triggered by natural causes. This included landslides that in the inventory had missing or unknow triggering causes.

The database from NVE consist of a table with event names, place name, time of the landslide, position of the landslide, triggering cause, damage to infrastructure, death or injury, source of the registered event, and if there was a quality control of the event. The inventory is available in the appendix (Appendix 1). As the inventory covers all of Norway, a column was added to the inventory with the coordinates in UTM33.

3.4 Precipitation analysis

The goal of the precipitation analysis was to identify if there were any trends in the precipitation that could be connected to triggering of quick clay landslides. From theory (NGU, n.d.) and experience from earlier quick clay landslides, it was not expected that precipitation alone triggers quick clay landslides. However, it can cause higher pore pressure in the quick clay weakening the slope, which again means that increased precipitation can lead to increased erosion, the main natural triggering factor.

The precipitation data downloaded (Table 2) combines rain and snowmelt. The precipitation data is a grid-based dataset (1x1 km grids measuring 24-hour data) from hundreds of measuring stations in Norway. The measured precipitation data is then interpolated to the 1x1 km grids using an algorithm based on the Bayesian interpolation method "Optimal Interpolation" (Lussana et al., 2018; NVE et al., n.d.-a). The snowmelt data is added to the dataset using a snow model that calculates the amount of snow based on the precipitation and temperature in each 1x1 km grid (NVE et al., n.d.-b; Saloranta, 2014).

All the naturally triggered quick clay landslides Oslo and Viken that occurred after 1960 were inspected for daily, monthly, seasonal and yearly precipitation values using a simple code (Appendix 2) in MATLAB and compared to the maximum rain and snowmelt in the areas where quick clay landslides occurred. Seasons were defined as 3-month starting from December – February, March – May etc.

3.5 Preparing data for MATLAB classification

It was decided that only quick clay landslides that occurred inside a quick clay hazard zone would be used in the MATLAB classification, partly to reduce the time the analysis would take and partly to fit better with the availability of DTMs to calculate erosion. The dataset was further limited to the years 2000-2022, as the quick clay landslides for these years were the most accurate and properly registered quick clay events in the database. The earliest year with available DTMs was 2007. Table 3 shows the selected events for the analysis.

Table 3 - Overview of the quick clay events selected for the analysis, showing the number and name of each quick clay hazard zone, the name of the event, and the year of the event.

Hazard Zone	Event Name	Year
1323 Fossnes	Hvittingfoss, Kongsberg	2000
79 Bjørkemoen	Frogner, Lillestrøm	2000
782 Foss	Øst for Nesveien and Nesveien, Skiptvedt	2011 and 2012
89 Hilton	Tveiter, Gjerdrum	2012
1383 Kåbbel Søndre	Hobølelva, Våler	2012
534 Vestby	Båhus, Nannestad	2012
394 Bøler	Søndre Rotnes, Årnes	2016
517 Ånåsrud Nord	Leirbekken 1, Nannestad	2020
521 Nygård	Leirbekken 2, Nannestad	2020
470 Hønsisletta	Ask, Gjerdrum	2020

3.5.1 Precipitation

Taking into consideration the different precipitation amounts at different sites, a mean of the precipitation for the last 30 years, 01.01.1990 to 01.01.2023, was calculated for each quick clay landslide site. Seasonal and yearly precipitation data was compared to the mean as a percentage of yearly precipitation for that site. The 30 year mean was chosen as a baseline for comparison to avoid changing weather caused by climate change (Meteorologisk institutt, 2022) and changes in the catchment area that could affect the water supply to obtain a mean that more closely represents the current weather conditions, although going back to 1957 (Lussana et al., 2018; NVE et al., n.d.-a) was possible.

Since yearly precipitation is a measure of precipitation that fell in a particular calendar year, it is not an optimal way to measure precipitation within the year a quick clay event occurred. Therefore, I used a different measure for the years where a quick clay landslide occurred, calculating the antecedent precipitation from the date of a quick clay event. The 365 days before the quick clay event for the annual rain and snowmelt and the 90 days before an event as the seasonal rain and snowmelt to ensure that only precipitation before the event was counted and not precipitation that happened after the event but still in the same calendar year.

3.5.2 Erosion

Part of the erosion step in the method was to evaluate which quick clay zones should be included in the analysis. This was done by creating a 35m buffer in ArcGIS Pro around the quick clay hazard zones and then selecting the hazard zones where an event had occurred since year 2000 and with at least two DTMs available from separate years before the event.

The hazard zones which were selected were exported as shapefiles, which specify the area and project when downloading DTM data from hoydedata.no. The DTMs were downloaded with a resolution of 1 m and in the UTM 32 coordinate system. The DTMs were then further clipped with the ‘clip raster’ tool in ArcGIS Pro to better fit the hazard zones.

To calculate the amount of erosion in the hazard zones, the tool ‘raster calculator’ was used to subtract the oldest DTM by the newest DTM. If a zone has more than two DTMs, the erosion was calculated between every DTM in steps going from oldest to second oldest and from second oldest to the newest DTM. The result was available as different values such as mean,

min, max and std in meters. To be on the safe side when it comes to the amount of erosion for each site, the max value of erosion was chosen for the final analysis.

As DTMs were not available for every year, the erosion was set to zero in years before an available DTM. The first DTM for a given year was also set to zero erosion and then between the last and the first DTM, the erosion was interpolated up to the measured erosion in the final DTM. In years after the final DTM, the erosion was set to the same as the final measured erosion. In years after an event the erosion was again set to 0.

An erosion rate was also calculated as meters of erosion divided by the number of years between the DTMs.

Table 4 includes the DTMs used in the project.

Table 4 - The Digital Terrain Models (DTM) from hoydenorge.no used to calculate erosion.

Name of DTM Projects
Romerike 07pkt 2007
Romeriksåsene 2013
NDH Akershus 2pkt 2015
Gjerdrum Ullensaker Nannestad 5pkt 2020
Ullensaker Nes Nannestad 2010

3.5.3 Antecedent precipitation method

To test the same method with a different approach to the formatting of the data, an antecedent dataset was created, with daily, 30- 60- and 90-day intervals antecedent rain and snowmelt. The same timeframe was chosen 2000-2022. The table (Appendix 3) consists of the antecedent intervals, the hazard zone name, the date and response which indicates whether there was a quick clay landslide that day.

3.6 Training models in MATLAB

To create thresholds through model training in MATLAB's Classification Learner app, three different datasets were used: 1) only precipitation, 2) precipitation and erosion, and 3) antecedent precipitation. For all three datasets, the response variable (class) was Landslide/Not Landslide, dependent on whether there was a quick clay landslide that year. Predictor variables are shown in Table 5. The full table for dataset 1 and 2 is found in Appendix 4 and the table for dataset 3 is found in Appendix 3. Datasets 1 and 2 are class imbalanced with only 11 out of 230 datapoints belonging to the Landslide class. This is even more severe for the antecedent precipitation dataset with daily precipitation values leading to a class imbalance with still only 11 datapoints within the Landslide class but over 80,000 within the Not landslide class.

Table 5 - Predictor variables (with explanation) for the three datasets, which were used to train in MATLAB models in classification of quick clay landslide event.

Predictor variable	Explanation	Dataset 1: only precipitation	Dataset 2: precipitation and erosion	Dataset 3: antecedent precipitation
Hazard zone	Name of the quick clay hazard zone	X	X	X
Year/Date	Ranging from 2000-2022	X	X	X
Highest seasonal rain and snowmelt %	Season with highest rain and snowmelt for that year	X	X	
Annual rain and snowmelt %	The annual rain and snowmelt for that year	X	X	
Erosion between DTMs	Erosion between the two or more DTMs		X	
Erosion Rate	Calculated as the erosion that occurred between the DTMs		X	
DTM for year	yes or no of whether there was a DTM available for that year		X	
DTM Date	Day and month of when the DTM was created for that specific year		X	
Daily antecedent precipitation and snowmelt				X
30-day antecedent precipitation and snowmelt				X

60-day antecedent precipitation and snowmelt		X
90-day antecedent precipitation and snowmelt		X

3.6.1 Model selection in MATLAB

The model results can differ from each session and therefore, several sessions of each model were done to find the "best" classification models. A session starts with defining the models that you wish to use and then training them on the imported dataset. After the model is done training, the results can be evaluated in MATLAB by using either the Validation Confusion matrix (see 3.6.2), which gives information about the amount of correct and wrong classifications or using the ROC curve graph showing the relationship between true positives and false positives (see 3.6.3).

The dataset was small enough for all available models in MATLAB to be tested in a short amount of time. The final decision of which models to use was mainly based on results from the testing in MATLAB. The models selected to perform the analysis in the end were two SVM models, Cubic and Quadratic, and the two ensemble models RUSBoosted- and Bagged decision trees. Before training a model some advanced options can be set depending on the type of model, for example for the ensemble models the number of learners can be set. For the final analysis, the models were left on default values (Table 6), but they were all tested with different options and no significant differences in results were found.

Table 6 – The advanced options selected for the Support Vector machine (SVM) and the ensemble models RUSBoosted- and Bagged decision trees in MATLAB. All options were left on default values after initial tests showed significant differences with different settings.

Model	Advanced options
SVM	Box constraint: 1 Kernel Scale Mode: Auto Standardize data: On
RUSBoosted	Number of splits: 20 Number of learners: 30 Learning rate: 0.1
Bagged	Number of splits: 6 Number of learners: 30 Learning rate: 0.1

3.6.2 Confusion matrix

A confusion matrix can be used to define the performance of a classification method. Figure 9 Figure 9illustrates the matrix for a binary classification, where P is positive, N is negative, T is true and F stands for False. True Positive (TP) would be an observation that is true and is predicted to be true, that is predicting a landslide when there was one. False Positive (FP) would be predicting a landslide when there wasn't a landslide.

In MATLAB's Classification Learner app, the confusion matrix has two options, the True Positive Rates (TPR) and False Negatives Rates (FNR) option and the Positive Predictive Values (PPV) and False Discovery Rates (FDR) option. TPR is the proportion of correctly classified observations per true class and FNR is the proportions of incorrectly classified observations per true class. PPV and FDR work well for displaying the amount of false positives in the classifier, PPV the proportion of correctly classified observations per class and FDR is the proportion of incorrectly classified observations per class (The MathWorks Inc., 2023d).

The last option can be important if the data set has imbalanced classes like landslide data where there are many years without an event. The false positives will not be as visible in the first option.

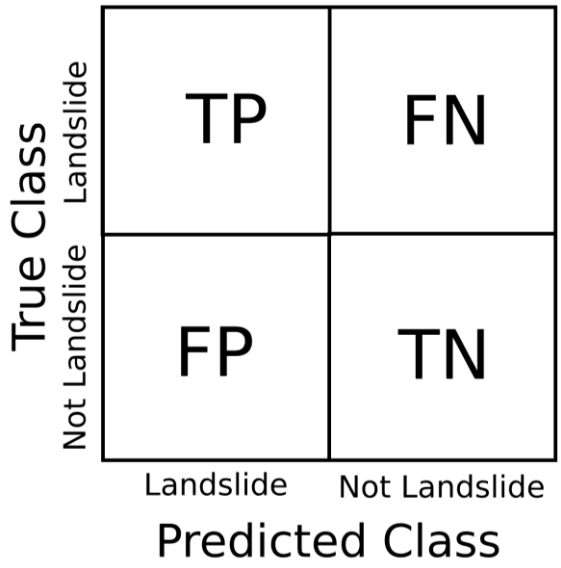


Figure 9 – Example of a Confusion Matrix for the classes Landslide and Not Landslide. P- positive, N- negative, F – false, T- true.

3.6.3 Receiver operating characteristics (ROC) curve

ROC is a method to evaluate classifiers based on their performance, with a graph visualizing the trade-off between hit rates and false alarm rates of classifiers (Fawcett, 2006). Using two parameters, true positive rate (TPR) and False positive rate (FPR), the ROC curve plots the TPR and FPR parameters at different thresholds.

True Positive Rate:

$$TPR = \frac{TP}{TP + FN}$$

False Positive Rate:

$$FPR = \frac{FP}{FP + TN}$$

The ROC results can be evaluated by calculating the Area Under the ROC Curve (AUC). The value of the AUC will be between 0.0 and 1.0 (Fawcett, 2006). A random classifier would have the same values for the FRP and the TPR with an AUC value of 0.5 and a perfect classifier would have a value of 1.0 (The MathWorks Inc., 2023b) as shown in Figure 10.

What is considered an acceptable AUC value can change depending on the field of study and the aim of the analysis. One way to quantify the AUC values is 0.9-1 as an excellent model, 0.9-0.8 as a good model and 0.8-0.7 as an acceptable model (Kikuchi et al., 2023).

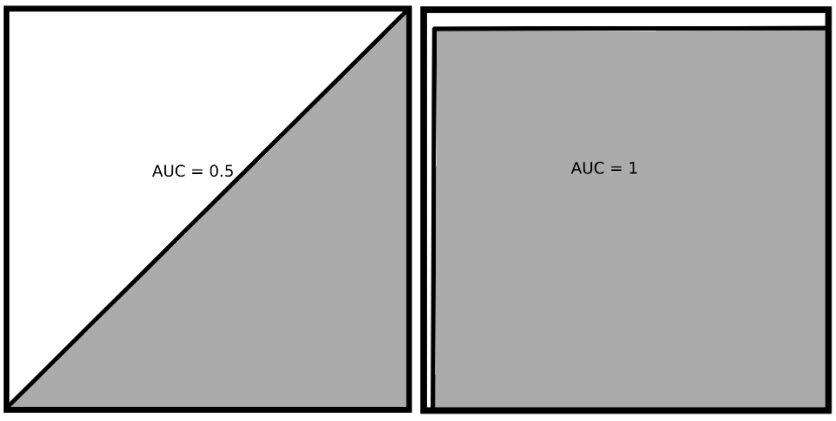


Figure 10 - Illustration of the principle behind Area Under the Curve (AUC), showing an example of a random classifier to the left (0.5) and a perfect classifier to the right (1.0).

4 Results

4.1 Precipitation analysis

The exploratory analysis shows no clear correlation in the data that points to precipitation being the main triggering factor in any of the quick clay events. There was, however, a high amount of precipitation both yearly and seasonally during the years where the events occur, but this does not represent the peak precipitation events for the area. For some events there was no precipitation at the day of the event.

In the hazard zone 1323 Fossnes and the quick clay event called Hvittingfoss in year 2000, the yearly precipitation and snowmelt was the peak precipitation and snowmelt for that area during the last 70 years, however the landslide released on the 15th of July 2000 and the precipitation events that created the highest precipitation came during the autumn season after the landslide had released (Figure 11).

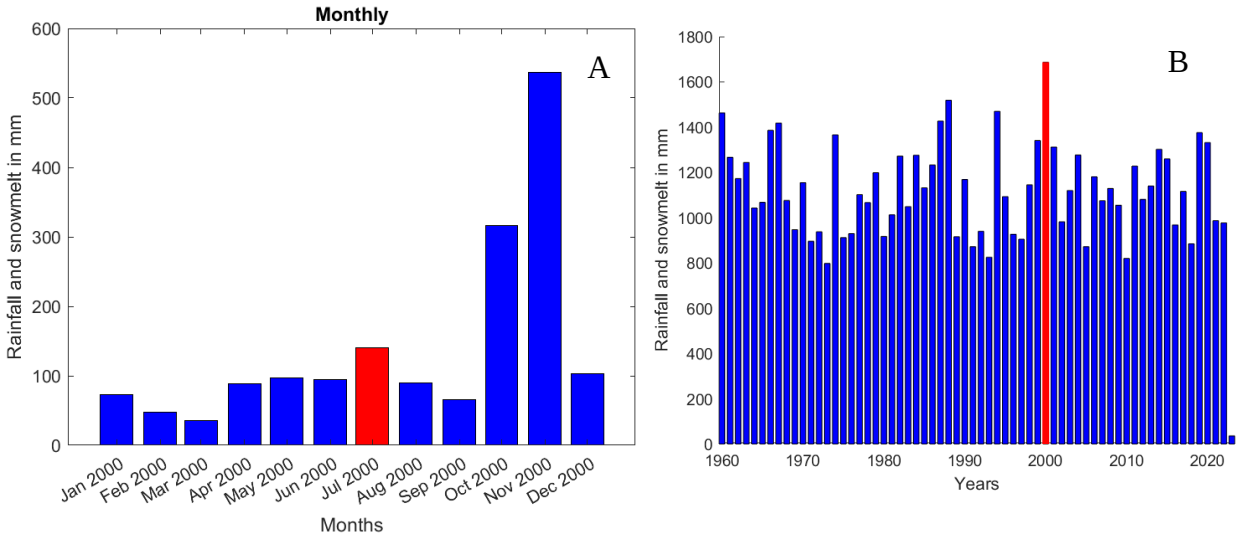


Figure 11 - Monthly and yearly precipitation for the quick clay event Hvittingfoss in 2000, in the hazard zone 1323 Fossnes. Monthly precipitation of the month when the event happened (marked in red) is low compared to precipitation in the autumn season (A), which is causing the peak rainfall and snowmelt for the area in the year 2000 (B).

4.2 Models trained with only precipitation data

The Cubic SVM has a 27.3 % True Positive Rate (TPR) and a 72.7 % False Negative Rate (FNR) for the landslide Class, classifying three of the 11 quick clay landslides correctly and classifying three non-landslides as landslide with a False Discovery Rate (FDR) of 50 % for the landslide class (Table 7). The Cubic SVM achieved an AUC value of 0.67 (Table 8).

The Quadratic SVM has a 36.4 % TPR and a 63.6 % FNR for the landslide Class, classifying four of the 11 quick clay landslides correctly and no false positives with a 0 % FDR for the landslide class (Table 7). The Quadratic SVM achieved an AUC value of 0.71 (Table 8).

The RUSBoosted decision tree ensemble model has a 45.5 % TPR and a 54.5 % FNR for the landslide Class, classifying five of the 11 quick clay landslides correctly and classifying 60 non-landslides as landslide with a FDR of 92.3 % for the landslide class (Table 7). The RUSBoosted decision tree ensemble model achieved an AUC value of 0.57 (Table 8).

The Bagged decision trees ensemble model has a 18.2 % TPR and a 81.8 % FNR for the landslide Class, classifying two of the 11 quick clay landslides correctly no false positives with a 0 % FDR for the landslide class (Table 7). The Bagged decision tree ensemble model achieved an AUC value of 0.73 (Table 8).

Table 7 - Comparison of the four models trained with only precipitation data, showing the True Positive Rate (TPR), the False Negative Rate (FNR), the Positive Predictive Value (PPV) and the False Discovery Rate (FDR) for the Landslide and Not Landslide classes.

	Landslide		Not Landslide		Landslide		Not Landslide	
Models	**TPR**	**FNR**	**TPR**	**FNR**	**PPV**	**FDR**	**PPV**	**FDR**
Cubic	27.3 %	72.7 %	98.6 %	1.4 %	50.0 %	50.0 %	96.4 %	3.6 %
Quadratic	36.4 %	63.6 %	100 %	0.0 %	100 %	0.0 %	96.9 %	3.1 %
RUSB	45.5 %	54.5 %	72.6 %	27.4 %	7.7 %	92.3 %	96.4 %	3.6 %
Bagged	18.2 %	81.5 %	100 %	0 %	100 %	0 %	96.1 %	3.9 %

Table 8 - Comparison of the four models trained with only precipitation data, showing the number of True and False predictions for the Landslide and Not Landslide classes, and the Area Under the Curve (AUC) values.

	Landslide		Not Landslide		AUC
Models	**True**	**False**	**True**	**False**	
Cubic	3	3	216	8	0.67
Quadratic	4	0	219	7	0.71
RUSB	5	60	159	6	0.57
Bagged	2	0	219	9	0.73

4.3 Models trained with precipitation and erosion

The Cubic SVM has a 36.4 % True Positive Rate (TPR) and a 63.6 % False Negative Rate (FNR) for the landslide Class, classifying four of the 11 quick clay landslides correctly and no false positives with a 0 % FDR (Table 9). The Cubic SVM achieved an AUC value of 0.72 (Table 10).

The Quadratic SVM has a 45.4 % TPR and a 54.4 % FNR for the landslide Class, classifying five of the 11 quick clay landslides correctly and no false positives with a 0 % FDR (Table 9). The Quadratic SVM achieved an AUC value of 0.74 (Table 10).

The RUSBoosted decision tree ensemble model has a 54.4 % TPR and a 45.5 % FNR for the landslide Class, classifying six of the 11 quick clay landslides correctly and classifying 63 non-landslides as landslide with a FDR of 91.3 % for the landslide class (Table 9). The RUSBoosted decision tree ensemble model achieved an AUC value of 0.61 (Table 10).

The Bagged decision trees ensemble model has a has a 36.4 % TPR and a 63.3 % FNR for the landslide Class, classifying four of the 11 quick clay landslides correctly no false positives with a 0 % FDR for the landslide class (Table 9). The Bagged decision tree ensemble model achieved an AUC value of 0.85 (Table 10).

Table 9 - Comparison of the four models trained with precipitation and erosion data, showing the True Positive Rate (TPR), the False Negative Rate (FNR), the Positive Predictive Value (PPV) and the False Discovery Rate (FDR) for the Landslide and Not Landslide classes.

	Landslide		Not Landslide		Landslide		Not Landslide	
Models	**TPR**	**FNR**	**TPR**	**FNR**	**PPV**	**FDR**	**PPV**	**FDR**
Cubic	36.4 %	63.6 %	100 %	0 %	100 %	0 %	96.9 %	3.1 %
Quadratic	45.4 %	54.5 %	100 %	0 %	100 %	0 %	97.3 %	2.7 %
RUSB	54.5 %	45.5 %	71.2 %	28.8 %	8.7 %	91.3 %	96.9 %	3.1 %
Bagged	36.4 %	63.6 %	100 %	0 %	100 %	0 %	96.5 %	3.5 %

Table 10 - Comparison of the four models trained with precipitation and erosion data, showing the number of True and False predictions for the Landslide and Not Landslide classes, and the Area Under the Curve (AUC) values.

	Landslide		Not Landslide		AUC
Models	**True**	**False**	**True**	**False**	
Cubic	4	0	219	7	0.72
Quadratic	5	0	219	6	0.74
RUSB	6	63	156	5	0.61
Bagged	4	0	219	7	0.85

4.4 Comparison between models trained with precipitation versus precipitation/erosion

Comparing the SVM models there was an increase in performance with the addition of erosion data for all models. The Cubic SVM model correctly classifies one more event and three fewer false positives, and the Quadratic model also classifies one more event correctly. The models also have a small increase in AUC by 0.05 and 0.03 respectively (Table 11).

Of the ensemble models the RUSBoosted decision trees model has an increase in correctly classified quick clay landslides, correctly identifying one more event, however there is also an increase in the false positive with three more non-events classified as quick clay landslides. And the Bagged decision trees model also has an increase in correctly classified events with two more events correctly classified. The models also have an increase in AUC by 0.04 and 0.12 respectively (Table 11).

Table 11 - Difference in classification of each of the four models when trained with only precipitation data versus precipitation and erosion data in combination. The difference in number of True and False predictions for the Landslide and Not Landslide classes and the Area Under the Curve (AUC) values are calculated from Table 8 and Table 10.

	Landslide		Not Landslide		AUC
Models	**True**	**False**	**True**	**False**	
Cubic	1	-3	3	-1	0.05
Quadratic	1	0	0	-1	0.03
RUSB	1	3	-3	-1	0.04
Bagged	2	0	0	-2	0.12

4.5 Models trained with antecedent precipitation data

The only model to classify any quick clay landslides correctly with the antecedent data was the RUSBoosted decision trees ensemble model it had a 36.4 % TPR and a 63.6 % FNR for the landslide Class, classifying four of the 11 quick clay landslides correctly and classifying 30272 non-landslides as landslide with a FDR of 100 % for the landslide class. The RUSBoosted decision tree ensemble model achieved an AUC value of 0.49 (Table 12).

Table 12 - Results from the antecedent precipitation method with intervals of 1, 30, 60 and 90 days with the RUSBoosted decision trees ensemble model. Showing the True Positive Rate (TPR) and the False Negative Rate (FNR), the Positive Predictive Value (PPV) and the False Discovery Rate (FDR), the number of True and False predictions and the Area Under the Curve (AUC) values for the Landslide and Not Landslide classes.

Landslide		**Not Landslide**		**AUC**
TPR	**FNR**	**TPR**	**FNR**	0.49
36.4 %	63.6 %	64 %	36 %	
PPV	**FDR**	**PPV**	**FDR**	
0 %	100 %	100 %	0 %	
True	**False**	**True**	**False**	
4	30272	53727	7	

5 Discussion

Quick clay landslides may have major consequences resulting in large damages for society. The triggering factors are often a combination of anthropogenic and natural causes, including natural erosion following periods of high precipitation. With the expected increase in precipitation and extreme weather events as a result of the ongoing climate change (IEA, 2022; Meteorologisk institutt, 2022), together with the increase in urbanisation which might also contribute to increased water flow as was seen in the Gjerdrum event (Gjerdrumutvalget, 2021), there is a possibility that there will be an increase in erosion trigged quick clay landslides.

The method used in this thesis was developed with the aim to use machine learning techniques to create a threshold for when quick clay landslides trigger based on precipitation and erosion data gathered from previous quick clay landslide events. If successful, the result would be used to evaluate whether this kind of data and method can be used to periodically update the categories of quick clay hazard zones. To obtain this, selected models were trained with three different datasets (only precipitation data, precipitation data and erosion data combined, and antecedent precipitation data) and evaluated for their ability to correctly classifying quick clay landslides.

5.1 Model performance with regard to classifying quick clay landslides

Evaluating the models trained with only the precipitation dataset, the ensemble model Bagged decision trees performed best according to AUC with a value of 0.73 (Table 8), however it only classified two of the quick clay landslides. The Quadratic SVM model had an AUC value of 0.71 (Table 8), only 0.02 less than the Bagged decision trees model but classified four quick clay landslides correctly. In numbers of correctly classified quick clay landslides, the ensemble model RUSBoosted decision trees classified most (five) quick clay landslides correctly. However, as mentioned by Xiao et al. (2022) the RUSBoosted model is unsuitable for warning systems because of the high amount of false positives it creates, as was also the case in my study with 60 false positives (Table 8).

When adding erosion data, all models showed a small improvement in performance, but the overall relationship between the models stayed the same, with the Bagged decision trees model classifying less quick clay landslides than the Quadratic SVM model but with a higher AUC value. While all the models had an increase in AUC value with the addition of erosion data, only the Bagged decision trees model had a significant increase from 0.73 to 0.85 (Table 11). An AUC value of 0.85 is an indication that it can be a good model (Kikuchi et al., 2023), but since the quick clay landslide dataset is so imbalanced with only 11 of the 230 data points representing the landslide class, the AUC value can be misleading. The Bagged decision trees model only had a true positive rate of 36.4 % (Table 9) classifying four out of the 11 quick clay landslides.

It can be hard to evaluate models based on AUC values, as what is considered a good result can change depending on the field. As a comparison, Kuradusenge et al. (2020) used machine learning models to predict rainfall-induced landslides in Rwanda. They used the random forest and logistic regression models, which achieved AUC values of 0.995 and 0.997 and had a false negative rate around 10 %. The best performing models in my results only classified five or six of the 11 quick clay landslides correctly. If the method was to be used as a first step in a warning system for quick clay landslides, a model with a false negative rate of around 40-50 % would certainly not be ideal.

A positive result from the addition of erosion data was a reduction in false positives for the models. Of the selected models only the Cubic SVM had false positives in the final result, but

in some of the sessions during the training of the models several of the initially tested models had more false positives when trained only on precipitation data.

The dataset based on antecedent precipitation intervals was included as an alternative way to represent precipitation data. When training the models with antecedent precipitation in the intervals 1 day, 30 days, 60 days and 90 days, only the RUSBoosted decision trees model managed to classify anything, but had a large number of false positives (Table 12). Using daily precipitation data resulted in a larger dataset, which in turn created even more false positives with the RUSBoosted decision trees model.

Gauthier & Hutchinson (2012) concluded that while some antecedent intervals had high correlation with landslides, they were unique for each site, and that precipitation alone could not trigger quick clay landslides. This compares well with my results with antecedent precipitation, where I don't see a clear relationship between any specific interval of antecedent precipitation and the release of the quick clay landslides that were correctly classified. Besides, most models trained in MATLAB could not classify any quick clay landslides.

According to the Gjerdrum report (Gjerdrumutvalget, 2021), the measured erosion in Tisilbekken was 2.5 m between the years 2007 and 2015. This fits well with the erosion measured in my study, which showed a max erosion of 2.78 m erosion between 2007 and 2015. While the Gjerdrum report considers more factors when it comes to changes to the topography in the Gjerdrum area, such as fills and accumulation, or changes in the path of the river, this nevertheless shows that the method applied here can be used to measure erosion provided that DTMs of a good enough quality are available (see discussion below with regard to quality of DTMs).

5.2 Uncertainty in data

There are several points of uncertainty with regard to the data used in this study, one important point being the fact that the precipitation data is obtained from 1x1 km grids and interpolated based on a set of measuring stations (Lussana et al., 2018; NVE et al., n.d.-a). This means that the precipitation values used might not be the actual precipitation in the quick clay zones. A solution to this problem could be to use only zones with a measuring station to obtain values closer to reality. However, this would reduce the number of areas that could be included in the method as not every hazard zone has a measuring station.

The precipitation data used for the analyses include rain and snowmelt to obtain the complete water supply. This might add another uncertainty as the snowmelt data are not actually measured but based on precipitation and temperature data, and because the snow is not directly measured this can lead to snowmelt data that differ from the real snow conditions especially when the temperature is around zero degrees Celsius (NVE et al., n.d.-b; Saloranta, 2014).

There are also two points of uncertainty related to the DTMs. Firstly, there are not enough DTMs to cover several years for every quick clay hazard zone, leading to some hazard zones with no way to measure the erosion and other hazard zones with just one measure of erosion. Secondly, with initially few DTMs available, it might be problematic to remove low resolution DTMs, without losing information. For the Gjerdrum 2020 quick clay event, the 2013 DTM had to be removed because of poor resolution. In the report after the disaster (Gjerdrumutvalget, 2021), they came to the same conclusion, that the 2013 DTM was of too poor quality to be used for the erosion analysis. For the Gjerdrum site, this was ok as there were other DTMs available, but some sites had only two DTMs and removing one would lead to no erosion being measured at all.

5.3 Uncertainty in method

One large issue with the method developed using machine learning models is the risk of overfitting because the dataset is too small to be split into both training data and testing data. By testing on the trained dataset the models can perform better or even perfectly on the trained data, but when used on an unknown dataset the model is inaccurate (Ying, 2019). This is an important issue if the method is to be used as a type of warning system. However, splitting the available data into a training and a testing dataset would lead to models being trained on too little data and creating poor models because the initial training dataset would be too small for the models to find the parameters that would lead to a landslide.

According to Berk (2006) the bagging model can compensate for the issue of overfitting. However, Xiao et al. (2022) had issues with overfitting with the training set almost achieving an AUC value of 1.00, but the testing set an AUC value of only 0.64 and suggested that the bagging model was not suitable for class-imbalanced problems.

An improvement to the method could be adding additional predictors. Adding more predictors that could increase the correlation between precipitation and erosion would address the lack of DTMs and erosion values. Adding predictors that measure water saturation in the soil for the month or season of the quick clay landslide, or measuring the peaks of the water flow rate in rivers as an additional measure of increase in erosion, could potentially help the models to find thresholds for erosion triggered landslides. Gauthier & Hutchinson (2012) found that ground frost conditions and thawing coincided with several large landslides in Eastern Canada, and that these are important factors to take into consideration when looking for triggering factors. While the conditions in Norway might be somewhat different than those in Canada, in the Gjerdrum report (Gjerdrumutvalget, 2021) they also noted that it was an unusually mild December with little ground frost.

Collini et al. (2022) did a review of previous work using machine learning for susceptibility mapping and predicting landslides and compared to their own work to find the “best” predictors. They found that for rainfall-induced landslides, features such as 3-day antecedent rainfall, max temperature the previous day and the level of water in the rivers were the most relevant predictors. This could indicate that more predictors focusing on water levels or river flow rates also might increase the performance of detecting quick clay events.

As part of the method, an updated quick clay inventory had to be produced. This inventory is more complete than the data that can be downloaded from NVE, with some additional quick clay landslides from both L'Heureux & Solberg (2012) and NGI (2011). Although a quality control was done, checking for wrongly categorized landslides, wrong coordinates or dates, there are still uncertainties in the inventory, as for some landslides I was unable to find any reports or articles about the events.

6 Conclusion

The main goal of the thesis was to develop a method using machine learning to create thresholds for when quick clay landslides release based on precipitation and erosion data collected from previous events. For this I used machine learning models in MATLAB which included Cubic and Quadratic SVM and RUSBoosted and Bagged decision tree models. To evaluate the models AUC and the Confusion matrix tools in MATLAB was used.

The main conclusion from the study was that even the best performing machine learning models, the Quadratic SVM model and the Bagged decision trees model had high false negative rates at 54.4 % and 63.6 % and thus evaluated as unsuitable for the purpose of updating the categories of the quick clay hazard zones.

The models did show an improvement in AUC when trained with precipitation and erosion data in combination compared to only using precipitation data, 0.03 for the Quadratic SVM and 0.12 for the Bagged decision trees. However, overall the low amount of available erosion data is a major problem with the method.

6.1 Future studies

A suggestion for future studies would be to investigate if an improvement in performance of the models could be obtained with the addition of more erosion data. This would, however, require a method for gathering erosion measurements from selected quick clay sites, perhaps on a yearly basis.

Another point for further study would be to add more predictors such as water flow rate or water levels in rivers, measurement of ground frost, water supply or other similar predictors, which might better explain or detect the possible correlation between precipitation events and increase in erosion.

Bibliography

Berk, R. A. (2006). An Introduction to Ensemble Methods for Data Analysis. *Sociological Methods & Research*, *34*(3), 263–295. https://doi.org/10.1177/0049124105283119

Boswell, D. (2002). Introduction to Support Vector Machines. *Department of Computer Science and Engineering University of California San Diego*.

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(2), 123–140. https://doi.org/10.1007/BF00058655

Collini, E., Palesi, L. A. I., Nesi, P., Pantaleo, G., Nocentini, N., & Rosi, A. (2022). Predicting and Understanding Landslide Events With Explainable AI. *IEEE Access*, *10*, 31175–31189. https://doi.org/10.1109/ACCESS.2022.3158328

Ekız, S., & Erdoğmuş, P. (2017). Comparative study of heart disease classification. *2017 Electric Electronics, Computer Science, Biomedical Engineerings' Meeting (EBBT)*, 1–4. https://doi.org/10.1109/EBBT.2017.7956761

Endo, T. (1969). *Probable distribution of the amount of rainfall causing landslides* (Annual Report 1968), Hokkaido Branch, Govern. Forest Experiment Station, Sapporo, 123–136.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognition Letters*, *27*(8), 861–874. https://doi.org/10.1016/j.patrec.2005.10.010

Gauthier, D., & Hutchinson, D. J. (2012). Evaluation of potential meteorological triggers of large landslides in sensitive glaciomarine clay, eastern Canada. *Natural Hazards and Earth System Sciences*, *12*(11), 3359–3375. https://doi.org/10.5194/nhess-12-3359-2012

Gjerdrumutvalget. (2021). *Årsakene til kvikkleireskredet i Gjerdrum 2020*. https://www.regjeringen.no/no/dokumenter/arsakene-til-kvikkleireskredet-i-gjerdrum-2020/id2872948/

Gregersen, O. (1981). The quick Clay Landslide in Rissa, Norway. Contribution to the Tenth International Conference on Soil Mechanics and Foundation Engineering, Stockholm, Sweden, 15-19 June 1981. Norwegian Geotechnical Institute, Oslo, Norway. Publication No. 135, pp. 421–426.

Hanssen-Bauer, I., Førland, E., Haddeland, I., Hisdal, H., Lawrence, D., Mayer, S., Nesje, A., Nilsen, J. E., Sandven, S., Sandø, A., Sorteberg, A., & Ådlandsvik, B. (2016). *Klima i Norge 2100—Kunnskapsgrunnlag for klimatilpasning oppdatert i 2015* (No. 2/2015). Miljø-Direktoratet. https://www.miljodirektoratet.no/globalassets/publikasjoner/m406/m406.pdf

IBM. (n.d.). *What is Bagging? | IBM*. Retrieved 11 June 2023, from https://www.ibm.com/topics/bagging

IEA. (2022). *Norway Climate Resilience Policy Indicator – Analysis* (Climate Resilience Policy Indicator). IEA. https://www.iea.org/articles/norway-climate-resilience-policy-indicator

Issler, D., Cepeda, J. M., Luna, B. Q., & Venditti, V. (2012). *NIFS-N1 Q-Bing—Utløpsmodell for kvikkleireskred: Back-analyses of run-out for Norwegian quick-clay landslides.* [Publication No. 135] Norwegian Geotechnical Institute. https://publikasjoner.nve.no/rapport/2013/rapport2013_46.pdf

Kikuchi, T., Sakita, K., Nishiyama, S., & Takahashi, K. (2023). Landslide susceptibility mapping using automatically constructed CNN architectures with pre-slide topographic DEM of deep-seated catastrophic landslides caused by Typhoon Talas. *Natural Hazards*, *117*(1), 339–364. https://doi.org/10.1007/s11069-023-05862-w

Krøgli, I. K., Devoli, G., Colleuille, H., Boje, S., Sund, M., & Engen, I. K. (2018). The Norwegian forecasting and warning service for rainfall- and snowmelt-induced

landslides. *Natural Hazards and Earth System Sciences*, *18*(5), 1427–1450. https://doi.org/10.5194/nhess-18-1427-2018

Kuradusenge, M., Kumaran, S., & Zennaro, M. (2020). Rainfall-Induced Landslide Prediction Using Machine Learning Models: The Case of Ngororero District, Rwanda. *International Journal of Environmental Research and Public Health*, *17*(11), 4147. https://doi.org/10.3390/ijerph17114147

L'Heureux, J.-S. (2013). *Karakterisering av historiske kvikkleireskred og input parametere for Q-BING* [NIFS report no 38/2013]. The Norwegian Water Resources and Energy Directorate. https://nve.brage.unit.no/nve-xmlui/handle/11250/2497008

L'Heureux, J.-S. (2012). *A study of the retrogressive behavior and mobility of Norwegian quick clay landslides*. In: Proceedings of the 11th International & 2nd North American Symposium on Landslides, Banff, Canada.

L'Heureux, J.-S., Høydal, Ø. A., Paniagua Lopez, A. P., & Lacasse, S. (2018). *Impact of climate change and human activity on quick clay landslide occurrence in Norway.* Second JTC1 Workshops on Triggering and Propagation of Rapid Flow-like Landslides. https://www.issmge.org/filemanager/joint_committees/1/2nd_JTC1_Workshop_E-proceedings_20190628.pdf

L'Heureux, J.-S., & Solberg, I.-L. (2012). *Utstrekning og utløpsdistanse for kvikkleireskred basert på katalog over skredhendelser i Norge*. [NIFS report no. 21/213]. The Norwegian Water Resources and Energy Directorate. https://publikasjoner.nve.no/rapport/2013/rapport2013_21.pdf

Li, M., Zhu, M., Hec, Y., He, Z., Wang, N., Zheng, Z., & Zhou, G. (2020). Warning of Rainfall-Induced Landslide in Bazhou District. *IGARSS 2020 - 2020 IEEE*

International Geoscience and Remote Sensing Symposium, 6879–6882. https://doi.org/10.1109/IGARSS39084.2020.9324416

Lussana, C., Saloranta, T., Skaugen, T., Magnusson, J., Tveito, O. E., & Andersen, J. (2018). SeNorge2 daily precipitation, an observational gridded dataset over Norway from 1957 to the present day. *Earth System Science Data*, *10*(1), 235–249. https://doi.org/10.5194/essd-10-235-2018

Meteorologisk institutt. (2022, June 27). *Klima fra 1900 til i dag*. Meteorologisk institutt. https://www.met.no/vaer-og-klima/klima-siste-150-ar

NGI. (2011). *Fare- og risikokartlegging av kvikkleireområder, Oslo kommune* (No. 20081717-00-1-R; Risiko for Kvikkleireskred). Norwegian Geotechnical Institute. https://webfileservice.nve.no/API/PublishedFiles/Download/201600907/1866567

NGI. (2023, February 8). *What is quick clay?* What is quick clay? https://www.ngi.no/en/research-and-consulting/natural-hazards-container/avalanches-and-slides/quick-clay-landslides//What-is-quick-clay

NGU. (n.d.). *Hva utløser kvikkleireskred?* Hva Utløser Kvikkleireskred? Retrieved 26 May 2023, from https://www.ngu.no/geologi-og-risiko/hva-utloser-kvikkleireskred

NGU. (2021, February 10). *Marin grense | Norges geologiske undersøkelse*. https://www.ngu.no/emne/marin-grense Jeg er ikke sikker at man kommer på riktig side med denne lenke- sjekk

Nigussie, D. G. (2013). *Numerical modelling of run-out of sensitive clay slide debris*. Master thesis Norwegian University of Science and Technology, Trondheim. https://vegvesen.brage.unit.no/vegvesen-xmlui/handle/11250/190812

NOU. (2022). *På trygg grunn. Bedre håndtering av kvikkleirerisiko.* (NOU 2022: 3; Norges offentlige utredninger 2022). Olje- og energidepartementet. https://www.regjeringen.no/no/dokumenter/nou-2022-3/id2905694/

NVE. (2020a). *Oversiktskartlegging og klassifisering av faregrad, konsekvens og risiko for kvikkleireskred* (No. 9/2020; NVE Ekstern rapport). Norwegian Geotechnical Institute.

NVE. (2020b). *Veileder nr. 1/2019 Sikkerhet mot kvikkleireskred: Vurdering av områdestabilitet ved arealplanlegging og utbygging i områder med kvikkleire og andre jordarter med sprøbruddegenskaper* (Veileder No. 1/2019). The Norwegian Water Resources and Energy Directorate. https://publikasjoner.nve.no/veileder/2019/veileder2019_01.pdf

NVE. (2022, October 6). *Er det farlig å bo på kvikkleire?* - The Norwegian Water Resources and Energy Directorate. https://www.nve.no/om-nve/spoer-nve/om-kvikkleire/er-det-farlig-aa-bo-paa-kvikkleire/

NVE, Meteorologisk institutt, & Kartverket. (n.d.-a). *SeNorge—Nedbør og temperaturkart*. Retrieved 3 June 2023, from https://www.senorge.no/PrecTempMap

NVE, Meteorologisk institutt, & Kartverket. (n.d.-b). *SeNorge—Snøkart*. Retrieved 3 June 2023, from https://www.senorge.no/Snowmap

Saloranta, T. (2014). *New version (v.1.1.1) of the seNorge snow model and snow maps for Norway* (No. 6). The Norwegian Water Resources and Energy Directorate.. https://publikasjoner.nve.no/rapport/2014/rapport2014_06.pdf

Segoni, S., Piciullo, L., & Gariano, S. L. (2018). A review of the recent literature on rainfall thresholds for landslide occurrence. *Landslides*, *15*(8), 1483–1501. https://doi.org/10.1007/s10346-018-0966-4

Seiffert, C., Khoshgoftaar, T. M., Van Hulse, J., & Napolitano, A. (2010). RUSBoost: A Hybrid Approach to Alleviating Class Imbalance. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *40*(1), 185–197. https://doi.org/10.1109/TSMCA.2009.2029559

The MathWorks Inc. (2023a). *Classification Ensembles*. Classification Ensembles Boosting, Random Forest, Bagging, Random Subspace, and ECOC Ensembles for Multiclass Learning. https://se.mathworks.com/help/stats/classification-ensembles.html?s_tid=CRUX_lftnav

The MathWorks Inc. (2023b). *ROC Curve and Performance Metrics—MATLAB & Simulink*. ROC Curve and Performance Metrics. https://se.mathworks.com/help/deeplearning/ug/performance-curves.html

The MathWorks Inc. (2023c). *Support Vector Machine Classification*. Support Vector Machine Classification Support Vector Machines for Binary or Multiclass Classification. https://se.mathworks.com/help/stats/support-vector-machine-classification.html?s_tid=CRUX_lftnav

The MathWorks Inc. (2023d). *Visualize and Assess Classifier Performance in Classification Learner—MATLAB & Simulink*. Visualize and Assess Classifier Performance in Classification Learner. https://se.mathworks.com/help/stats/assess-classifier-performance.html?s_tid=srchtitle_Check%20Performance%20Per%20Class%20in%20the%20Confusion%20Matrix_1

Torrance, J. K. (2012). Landslides in quick clay. In D. Stead & J. J. Clague (Eds.), *Landslides: Types, Mechanisms and Modeling* (pp. 83–94). Cambridge University Press. https://doi.org/10.1017/CBO9780511740367.009

Vallet, A., Varron, D., Bertrand, C., Fabbri, O., & Mudry, J. (2016). A multi-dimensional statistical rainfall threshold for deep landslides based on groundwater recharge and support vector machines. *Natural Hazards*, *84*(2), 821–849. https://doi.org/10.1007/s11069-016-2453-3

Xiao, T., Zhang, L. M., Cheung, R. W. M., & Lacasse, S. (2022). Predicting spatio-temporal man-made slope failures induced by rainfall in Hong Kong using machine learning techniques. *Géotechnique*, *1-17*. https://doi.org/10.1680/jgeot.21.00160

Ying, X. (2019). An Overview of Overfitting and its Solutions. *Journal of Physics: Conference Series*, *1168*(2), 022022. https://doi.org/10.1088/1742-6596/1168/2/022022

Appendix

Appendix 1: Quick clay inventory

Appendix 2: MATLAB Code

Appendix 3: Antecedent precipitation table

Appendix 4: Table imported to MATLAB

Github Link: https://github.com/JakobBGit/MasterThesisAppendix

Quick clay Inventory

Appendix 1

Only chosen columns. Excel file with the full table with all Columns available from:

https://github.com/JakobBGit/MasterThesisAppendix/blob/main/Final_Inventory_Thesis.xlsx

skredNavn	stedsnavn	skredTidspunkt	utlosningArsak	kilde	POINT_X	POINT_Y
Heimstad	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	269773	7032958
Langørjan	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	257825	7041212
Leiffossen	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	271783	7036019
Lund	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	265680	7032848
Olderdalen	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	275966	7039457
Othilienborg	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	271873	7038763
Sjetnemarka	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	270109	7034849
Stavset	Trondheim	Forhistorisk		(L'Heureux & Solberg, 2012)	267243	7036904
Groruddalen	Oslo	Forhistorisk		(NGI, 2011)	271070	6654022
Telthusbakken	Telthusbakken	15.07.1100		NGU	262423	6650472
Duedalen	Duedalen	28.07.1625 12:00:00	Erosjon	NGU	270817.9	7041294
Bakklandet	Bakklandet	20.11.1634 12:00:00	Erosjon	NGU	270728.8	7041510
Leirfallsgata	Leirfallsgata	26.07.1705		NGU	262837.9	6650062
Litl-Amdal		01.09.1723 12:00:00	Naturlig utlost		353816.3	7157878

Kvam-Auglaraset	Kvam	21.09.1726 21:00:00	Naturlig utlost	NGU	328007.5	7082495
Dæli	Rissa	09.03.1760 16:30:00	Naturlig utlost	NGU	259347.3	7067259
Lille-Amdal		01.08.1763 12:00:00	Naturlig utlost		353856.5	7157915
Tesenfallet	Tesenfallet	21.10.1795 10:00:00	Naturlig utlost	NGU	300069.3	6674774
Stubergfallet	Værnes	15.10.1807 23:00:00	Naturlig utlost	NGU	298659.4	7040749
Tiller	Tiller	07.03.1816 17:30:00	Naturlig utlost	NGU	271243.1	7032371
Gullaug 2		13.05.1823 12:00:00	Hoyt poretrykk		235649.4	6631991
Egga	Eggem	14.01.1825 12:00:00	Naturlig utlost	NGU	256330.7	7037403
Brådalen	Brådalen	20.05.1831 12:00:00	Naturlig utlost	NGU	258818.1	7033839
Oppdal 3	Oppedal 3	30.08.1854 14:00:00	Naturlig utlost	NGU	351856.4	7150409
Gløymem	Gløymem	21.03.1857 12:00:00	Naturlig utlost	NGU	359589.2	7158867
Nypan	Nypan	10.11.1867 08:00:00	Naturlig utlost	NGU	266530.7	7028743
Leirfallet 2	Leirfallet 2	26.03.1869 09:00:00	Naturlig utlost	NGU	262488.1	7032800
Kvidal	Kvidal	17.12.1871 12:00:00	Naturlig utlost	NGU	252312	7052844
Holem	Holem	07.04.1874 12:00:00	Naturlig utlost	NGU	361878.8	7126307
Svendengen		23.03.1878 19:00:00	Graving		264181.8	6648402
Utrasingen av melkefabrikken på Sannesund		06.09.1890 08:00:00	Naturlig utlost		277321.3	6577239
Elvberg		04.12.1892 22:00:00	Naturlig utlost		311864.9	7067688
Verdalsraset	Verdalsraset	19.05.1893 00:30:00	Erosjon	NGU	333414.2	7077461

Bislet	Bislet	02.07.1895 16:30:00	Ikke gitt	NGU	261664.7	6650531
Grubbåsen	Grubbåsen	1900-05-05 10:00:00	Ikke gitt	NGU	303775.5	7058112
Gullsmedvika	Gullsmedvika	1902-11-11 18:00:00		NGU	461150	7355600
Haugan, Vuku	Haugan, Vuku	1906-09-19 8:30:00	Naturlig utlost	NGU	345978.2	7079858
Smestadbanen	Smestadbanen	1913-07-03 12:00:00	Ikke gitt	NGU	260726.2	6651149
Skarpsno-Skøyen	Skarpsno-Skøyen	1913-10-24 12:00:00		NGU	259061.2	6650295
Chr. Kroghsgt	Chr. Kroghsgt	1914-06-05 14:00:00	Ikke gitt	NGU	262995.5	6649586
Lånke		1918-03-17 3:00:00	Naturlig utlost		301399.3	7039962
Meråker		1919-08-19 13:00:00	Naturlig utlost		334797.2	7038594
Kankedalen	Kankedalen	1924-10-20 22:00:00	Infiltrasjon av vann	NGU	279667.2	6668809
Lodalen	Lodalen	1925-01-29 0:00:00		NGU	264354.3	6647955
Gretnes	Gretnes	1925-04-17 18:30:00	Erosjon	NGU	276901.5	6575772
Lodalen 2	Lodalen 2	1927-04-29 5:00:00		NGU	269671.9	6652403
Brådalen 5	Brådalen 5	1928-04-24 12:00:00	Naturlig utlost	NGU	258996.4	7033823
Brådalen-Gaulosen	Brådalen-Gaulosen	1928-05-01 12:00:00	Naturlig utlost	NGU	259084.4	7033755
Ner-Tingstad	Ner-Tingstad	1932-01-27 12:00:00	Naturlig utlost	NGU	318688.5	7072624
Grungstadvatnet	Grungstadvatnet	1932-03-31 10:00:00	Naturlig utlost	NGU	366300	7162700
Nydalen	Nydalen	1934-12-12 13:00:00	Ikke gitt	NGU	263366.9	6653671
Merradalen	Merradalen	1936-03-31 0:00:00		NGU	254559.9	6658140

Ness	Ness	1936-04-03 12:00:00	Naturlig utlost	NGU	381350	7163300
Arnebråtveien	Arnebråtveien	1936-05-29 12:00:00	Graving	NGU	258173	6653424
Holand	Holand	1942-01-12 19:00:00	Naturlig utlost	NGU	362665.4	7153704
	Byneset, Nedre Mule	1943-05-18 0:00:00		NGI	258001	7033162
Lade	Lade	1944-04-11 16:30:00	Naturlig utlost	NGU	272281.5	7043298
	Kverne	1944-10-09 0:00:00		NGI	230500	6581492
	Tune / Isebakke	1944-11-10 0:00:00		NGI	288733	6558764
Leirådalen	Leirådalen	1950-07-28 12:00:00	Naturlig utlost	NGU	373729.3	7128825
	Bøler	1950-10-10 0:00:00		NGI	285271	6627170
Østersem	Østersem	1951-07-27 12:00:00	Naturlig utlost	NGU	371000	7124300
Gaustadbekken	Oslo	1953-04-01 0:00:00			260811.5	6652500
Bekkelaget	Bekkelaget	1953-10-07 7:37:00	Utfylling	NGU	263373.9	6645806
Borgen	Borgen	1953-12-22 23:00:00	Naturlig utlost	NGU	288861.7	6668415
Lodalen 3	Lodalen 3	1954-10-06 0:00:00		NGU	264010.1	6648147
	Ilangskogen	1954-12-01 0:00:00		NGI	303674.2	6676020
Bragernes 8	Viktoriatomten	1955-01-06 12:00:00	Ikke gitt	NGU	230218.1	6632768
	Rolvsøy	1955-05-03 0:00:00		NGI	273137	6576335
Ingedalsbekken		1956-06-01 12:00:00	Ikke gitt		283660	6566792
Økern	Økern	1957-06-04 19:00:00	Ikke gitt	NGU	267057.5	6650338

Urstad		1958-07-08 5:30:00	Naturlig utlost		376135.5	7158946
Vibstad	Vibstad	1959-02-22 22:30:00	Erosjon	NGU	351904.8	7153681
Furre	Furu	1959-04-14 8:50:00	Erosjon	NGU	343983.4	7151475
Bergsmoen	Bergsmoen	1961-04-15 3:00:00	Naturlig utlost	NGU	365230	7151810
Hovraset	Hovraset	1962-09-14 23:00:00	Naturlig utlost	NGU	303982.1	7048739
Bergsmoen 2	Bergsmoen 2	1965-04-06 22:00:00	Naturlig utlost	NGU	365440	7151680
Selnes	Selnes	1965-04-18 15:15:00		NGU	328136.4	7149493
	Heksberg	1967-03-19 0:00:00		NGI	290275.3	6655947
Hekseberg	Hekseberg	1967-03-20 12:00:00	Erosjon	NGU	282144.1	6663108
	Nordre Nordby	1967-04-10 0:00:00		NGI	276819	6682009
	Hovin	1968-04-30 0:00:00		NGI	251576.1	6983635
	Drammen travbane	1970-12-01 0:00:00		NGI	225706.3	6634286
	G. Dahle	1971-07-13 0:00:00		NGI	279445.5	6661451
	Trevarebyen	1972-05-21 0:00:00		NGI	270650.8	6664447
	Våle	1972-07-12 0:00:00		NGI	223526.6	6583595
	Halse - Nordmøre	1972-09-07 0:00:00		NGI	162921.8	7016436
	Flatner	1973-06-16 0:00:00		NGI	279480.8	6661516
Gullaug 1		1974-11-29 10:02:00	Utfylling		235172.5	6631753
Båstad	Båstad	1974-12-05 16:30:00	Graving	NGU	290598.2	6624195

	Holtet Gård, Tune	1976-04-07 0:00:00		NGI	329667.2	6733981
Lånke	Lånke	1976-04-20 12:00:00	Naturlig utlost	NGU	298429.6	7041395
Songe	Songe, Tore Olsens Mek Verksted	1976-11-15 0:00:00		NGI	153197.6	6520554
Terråk	Terråk	1977-03-15 15:00:00	Graving	NGU	377220	7221020
Hyggen	Hyggen	1978-01-25 8:30:00	Utfylling	NGU	239362.4	6628818
Rissaraset	Rissa	1978-04-29 14:10:00	Utfylling	NGU	248900.3	7057320
Lahelle, Fredrikstad	Fredrikstad	1980-08-17 23:00:00	Utfylling	NGU	269571.1	6571051
Vibe, Steinkjer	Vibe	1982-10-04 6:00:00	Ikke gitt	NGU	332359	7102817
Bjørnåga, Vefsn	Øver-Bjørnåga	1984-07-08 3:00:00	Naturlig utlost	NGU	421700.7	7294350
Hilton, Ullensaker	Hilton	1984-07-19 14:00:00	Infiltrasjon av vann	NGU	284261.3	6665359
Strandajordet, Øvre Eiker	Strandajordet	1984-09-11 14:45:00	Ikke gitt	NGU	211999.2	6633015
Imsen, Åfjord	Imsen	1984-12-01 13:30:00	Utfylling	NGU	266474.3	7105600
Fosshaugen, Målselv	Fosshaugen	1985-05-02 11:00:00	Naturlig utlost	NGU	645464.3	7663619
Lundsbekken, Indre Østfold	Lundsbekken	1988-04-15 0:00:00	Infiltrasjon av vann	NGU	278073.2	6612909
Sandbukta, Balsfjord	Sandbukta	1988-08-24 6:35:00	Graving	NGU	671816.4	7686713
Nordsetrønningen, Klæbu	Nordsetrønningen	1988-11-24 0:00:00	Infiltrasjon av vann	NGU	273016.2	7031240
Leinstranda, Trondheim	Stokkaunet	1988-12-23 0:00:00	Hoyt poretrykk	NGI	266115.4	7029155
Lersbakken, Heimdal	Lersbakken	1988-12-31 0:00:00	Infiltrasjon av vann	NGU	266541.3	7031104
Jørstad, Snåsa	Jørstad	1989-03-12 0:00:00	Ikke gitt	NGU	365882.7	7122248

Sørkedalen, Trondheim	Sørkedalen	1989-11-22 0:00:00		NGU	266993.3	7032067
Finneidfjord, Hemnes	Finneidfjord	1996-06-20 0:30:00	Utfylling	NGU	445775.3	7340787
Åby		1999-06-01 15:48:00	Erosjon		192146.3	6551536
Kåbøl, Våler	Kåbøl	1999-07-15 0:00:00	Erosjon	NGU	264251.8	6602960
Hvittingfoss, Kongsberg	Hvittingfosss	2000-07-15 0:00:00		NGU	217353.4	6603179
Frogner, Lillestrøm	Frogner	2000-11-19 12:00:00		NGU	284003.1	6663352
Malvik, Trondheim	Malvik	2002-04-24 4:00:00	Utfylling	NGU	281896.8	7039994
Beitstad, Steinkjer	Beitstad	2002-04-27 12:00:00	Erosjon	NGU	325991.5	7115547
Kamperud, Ørje	Kamprud	2002-05-06 18:30:00	Graving	NGU	306400	6603526
Sjåenget, Overhalla	Sjåenget	2007-03-23 12:00:00	Kunstig utlost	NGU	340415.1	7154928
Reina, Overhalla	Reina	2007-05-16 8:15:00	Erosjon	NGU	354658.6	7158038
Tortenlia, Fauske	Tortenlia	2008-01-17 12:00:00	Utfylling	NGU	517134.2	7464176
Skred vest for Vidnesveien		2009-01-01 12:00:00	Erosjon		283152.7	6596585
Kattmarka	Kattmarka	2009-03-13 11:50:00	Kunstig utlost	NGU	328655.6	7154577
Lyngseidet, Lyngen		2010-09-03 15:40:00	Utfylling	NGU	703213	7725273
leirskred øst for Nesveien		2011-01-01 12:00:00	Erosjon		286836.1	6601167
Setnes, Rauma	Setnes	2011-07-07 0:00:00		NGU	123951.6	6957444
Åby		2011-09-02 12:00:00	Erosjon		192198.5	6551496
Esp, Trondheim	Esp	2012-01-01 0:00:00	Erosjon	NGU	257211.1	7038563

Nesveien, Skiptvedt		2012-01-01 12:00:00	Erosjon		286065.9	6601668
Svensrud, Gjerdrum	Svensrud, Gjerdrum kommune	2012-05-20 8:30:00	Graving	regObs	280847.6	6666536
Tveiter, Gjerdrum		2012-06-01 0:00:00	Naturlig utlost		283374.8	6665845
Hobølelva, Våler	Kåbbel	2012-09-24 1:00:00	Naturlig utlost	NGU	264443.1	6602504
Båhus, Nannestad	Båhus, Nannestad	2012-11-09 21:00:00	Erosjon	regObs	278486.8	6683078
Hauga, Suldal	Garaneset	2014-02-17 12:00:00	Utfylling	regObs	6898.383	6625678
Eskalerende erosjon vest for Riukveien		2015-01-01 12:00:00	Erosjon		283291.4	6604838
Skjeggestad, Holmestrand		2015-02-02 14:40:00	Utfylling		233810.6	6601761
Ogndal, Steinkjer	Ogndal	2015-03-17 16:30:00	Graving	regObs	339018.5	7103205
Elstad, Grong		2015-04-24 0:00:00	Naturlig utlost		376969.5	7157646
Ingedal, Sarpsborg		2015-05-21 14:00:00	Naturlig utlost		283984.3	6567332
Fv 76 ved Bjørnstokkvika, Tosbotn	Fv 76 ved Tosbotn	2016-04-01 7:28:00	Infiltrasjon av vann	regObs	404204.4	7247123
Fv 76 ved Bekkevoll, Tosbotn	Fv 76 ved Bekkevoll, Tosbotn	2016-04-02 12:50:00	Infiltrasjon av vann	regObs	403957.1	7246982
Søndre Rotnes, Årnes		2016-04-20 8:00:00	Erosjon		302735	6670249
Asakveien, Sørum		2016-11-10 15:55:00	Utfylling		287163.1	6661159
Pallaunet	Pallaunet	2016-12-08 17:05:00	Naturlig utlost	regObs	261192	7068456
Styggneset, Overhalla		2018-10-31 0:00:00	Ikke gitt	regObs	352013.2	7157255
Bommamyra, Steinkjer		2018-12-24 7:00:00	Naturlig utlost	regObs	329691.8	7105003
		2019-03-05 17:04:38		regObs	599967	7593000

Li, Nittedal		2019-09-16 9:37:00	Graving		273197.4	6659118
Steinselva, Jøa		2019-09-26 8:00:00	Erosjon	regObs	321940.6	7173456
		2019-10-02 16:05:33		regObs	322297	7172924
Østerøyveien 41, Sandefjord		2020-01-06 14:00:00	Utfylling		229107.5	6563753
Leirbekken 1, Nannestad	Leirsked leirbekken, Nannestad kommune	2020-03-01 12:00:00	Naturlig utlost	regObs	279120.8	6679165
Brennstadmoen, Rana	Rana, brennstadmoen boligfelt	2020-04-15 6:00:00	Erosjon	regObs	460366.3	7358446
Jonsrud, Vefsn		2020-05-13 6:30:00	Naturlig utlost	regObs	415074	7315426
Kråknes, Alta	Kråknes	2020-06-03 15:30:00	Utfylling	regObs	806608.2	7789315
Leirbekken 2, Nannestad	Leirbekken, Nannestad	2020-12-17 12:00:00	Naturlig utlost	regObs	279217.7	6679305
Ask, Gjerdrum		2020-12-30 3:55:00	Erosjon	regObs	279459.7	6665226
Solvang, Balsfjord	Balsfjord / Troms	2021-05-13 10:00:00	Naturlig utlost	regObs	667922.1	7702600
Søndre Vestby, Aurskog-Høland		2021-07-05 8:00:00	Erosjon	regObs	299530.8	6629140
Skred ved Ombekken i Overhalla		2022-07-24 17:30:00	Naturlig utlost		354753.5	7159296
Mikkeløra ved Fredriksberg	Mikkeløra ved Fredriksberg	2022-10-16 7:19:00	Erosjon	regObs	644123	7669144

Appendix

Appendix 2

Code on Github:

https://github.com/JakobBGit/MasterThesisAppendix/blob/main/Code_Thesis

MATLAB code for exploratory analysis of precipitation data.

```
clear all
clc

Fname = 'Example';
TT=readtimetable([Fname,'.csv'],'NumHeaderLines', 1);  % read the file; convert to timetable
TT.Properties.VariableNames{1} = 'R_S_mm'; %renaming the precipitation variable
R = standardizeMissing(TT, 65535); % identifying the cells with 65535 as nodata and removing them
TT = rmmissing(R);
TR1 = timerange('01.01.1990','01.01.2023'); %setting timerange for the dataset
TT = TT(TR1,:);

TTmS = retime(TT,'monthly','sum');
TTyS = retime(TT,'yearly','sum');
TTyS = retime(TT,'yearly','sum');

TT2 = [datetime(1989,12,31):calmonths(3):datetime(2023,04,01)]'; %defining the seasons, dec-feb(winter), mar-may, jun-aug, sep-nov
TTsS = retime(TT,TT2,'sum'); %seasonal Sum of rain and snowmelt
TTsSM = retime(TTsS, 'yearly', 'max');%max of the seasonal Sum

[~, mon, ~] = ymd(TTsS.Date);
WinterSeason = TTsS(mon == 12, :);
SpringSeason = TTsS(mon == 3, :);
SummerSeason = TTsS(mon == 6, :);
AutumnSeason = TTsS(mon == 9, :);


yearlymean = mean(TTyS.R_S_mm);

percentofyear = ((TTyS.R_S_mm / yearlymean)*100);

TTyS = addvars(TTyS, percentofyear);

percentofyearSeason = ((TTsS.R_S_mm / yearlymean)*100);

TTsS = addvars(TTsS, percentofyearSeason);

TTsSM = retime(TTsS, 'yearly', 'max');%max of the seasonal Sum
```

```
%%
TRS = timerange('2019','years');
OneYearS = TTsM(TRS,:);
max(OneYearS.WaterSupplyPer)

%%

TR = timerange('2019','years');
OneYear = TTmM(TR,:);
b = bar(OneYear.Date,OneYear.WaterSupplyPer, 'b');
b.FaceColor = 'flat';
b.CData(9,:) = [1,0,0];
title('Monthly')
ylabel('Mean Water Supply %')

exportgraphics(gca,strcat(Fname,'Monthly_mean.png'))

%%
hold on
b = bar(TTyM.Date,TTyM.WaterSupplyPer,0.7, 'b');
TTyMD = TTyM('01.01.2019',:);
b2 = bar(TTyMD.Date,TTyMD.WaterSupplyPer, 310, 'r','EdgeColor', 'none');
title('Yearly Mean Water Supply')
ylabel('Mean Water Supply in %')
xlabel('Years')
%ylim([2 4])
hold off
exportgraphics(gca,strcat(Fname,'Yearly_mean.png'))

%%
hold on
b = bar(TTsM.Date,TTsM.WaterSupplyPer, 'b');
TTsMD = TTsM('30-Jun-2000',:);
b2 = bar(TTsMD.Date,TTsMD.WaterSupplyPer, 75, 'r','EdgeColor', 'none');
title('3 Month Seasons')
ylabel('Mean Water Supply in %')
xlabel('Years')
%ylim([0 8])
hold off

exportgraphics(gca,strcat(Fname,'Season_mean.png'))

%%

TR = timerange('15.07.2000','months');
OneMonth = TT(TR,:);
TR2 = OneMonth('15.07.2000 06:00',:);
hold on
b = bar(OneMonth.Date,OneMonth.WaterSupplyPer, 'b');
b2 = bar(TR2.Date,TR2.WaterSupplyPer, 'r');
%ax = gca;
%ax.XTick = OneMonth.Date;
title('Daily Water Supply')
ylabel('Mean Water Supply in %')
hold off

exportgraphics(gca,strcat(Fname,'Daily_mean.png'))
```

Antecedent Precipitation

Appendix 3

The table is available on Github.

https://github.com/JakobBGit/MasterThesisAppendix/blob/main/AntecedentPrecipitationOsloAndViken

Final Table Imported to MATLAB

Appendix 4

The table is available on Github.

https://github.com/JakobBGit/MasterThesisAppendix/blob/main/ImportMatlabOsloViken