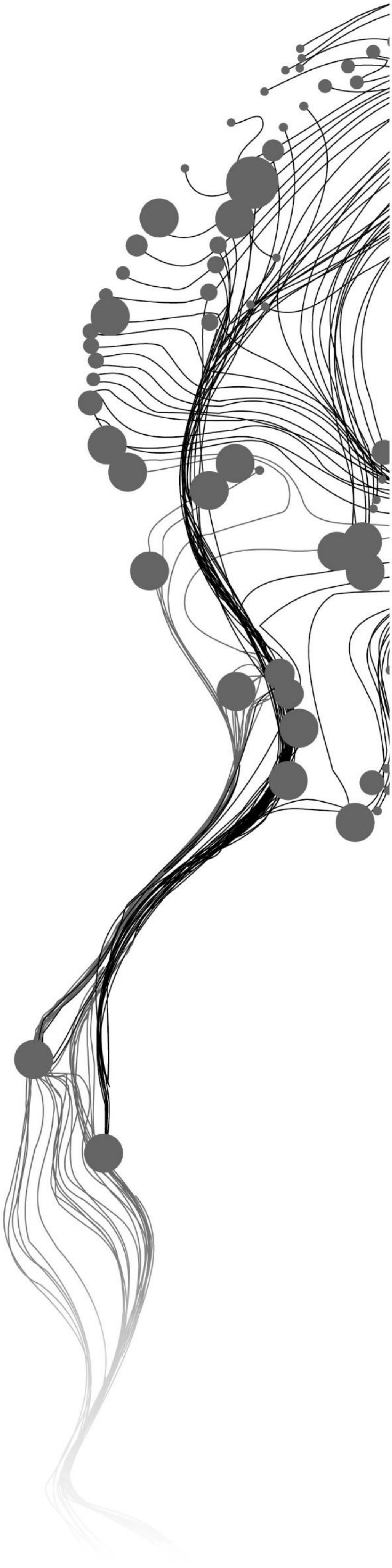
LANDSLIDE EARLY WARNING
RUNOUT MODELING
CASE STUDY IN NORWAY

DIANA COLLAZOS
August 2022

SUPERVISORS:
Dr. Bastian van den Bout
Dr. Ir. Janneke Ettema



LANDSLIDE EARLY WARNING RUNOUT MODELING CASE STUDY IN NORWAY

DIANA FERNANDA COLLAZOS CORTES
Enschede, The Netherlands, August 2022

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Natural Hazards and Disaster Risk Reduction

SUPERVISORS:
Dr. Bastian van den Bout
Dr. Ir. Janneke Ettema

THESIS ASSESSMENT BOARD:
Prof. Dr. Victor G. Jetten
Dr. Tjalling de Haas (External Examiner, Utrecht University)

ABSTRACT

A Landslide Early Warning System (LEWS) is a non-structural mitigation measure that can be applied in regions susceptible to landslides to reduce losses. In existing territorial LEWSs, rainfall thresholds are frequently used to raise warnings in administrative units as broad as municipalities or counties. However, more detailed landslide predictions might benefit these systems' response.

This thesis investigates the potential of integrating rainfall-induced failure and runout physically-based modeling in territorial LEWSs to predict local landslide impact areas using forecast precipitation. The study area is located in Norway, in the context of the Norwegian Landslide Forecasting and Warning Service. The event simulated is a rainfall-induced landslide that occurred on August 16, 2011, in Kjellberget, Stjørdal municipality.

A physically-based model was built and calibrated for the analysis using the LISEM software. It was chosen because it allows the construction of a multi-process model capable of simulating slope stability to obtain failure depth, which can be used as initial volume for runout. The calibration was based on slope stability, and the calibrated parameters were soil depth, cohesion, and Internal Friction Angle (IFA). The results gave information about the variation of Factor of Safety (FS), Failure Depth (FD), and Cohen's Kappa in the 3-dimensional space formed by the calibrated parameters. Failure and runout maps produced with the combinations that yield the highest precision, recall, and F1 were employed to examine the accuracy of the landslide simulations.

The results show that the initiation areas and runout directions were generally well replicated, but not the runout volumes. Therefore, the calibration was inadequate because it did not consider the multi-processes affecting the final landslide impact area. Based on the experiment performed, it is concluded that it is non-viable to apply the proposed methodology in territorial LEWSs. Further research could explore approaches to generate deeper failures and, thus, larger runout extents.

ACKNOWLEDGEMENTS

To my family, friends, and teachers that were there to say: 'you can do it'. To Sahara for supporting me unconditionally throughout the chaotic last months. To Enzo and David for the advice during the writing journey. To Bastian and Janneke for their supervision. To Bart for dealing with all my student affairs. And, to Graziella and Cees for making my visit to Norway possible.

Thank you very much.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

deg	Degrees
DTM	Digital Terrain Model
DTM	Digital Terrain Model
ECMWF-IFS	Integrated Forecasting System from the European Centre for Medium-Range Weather Forecasts
FD	Failure Depth
FN	False Negative
FP	False Positive
FS	Factor of Safety
GWB	Gridded Water Balance
hmax	Maximum height
IFA	Internal friction angle
km	Kilometer(s)
kPa	Kilopascal
LEWS	Landslide Early Warning System
m	Meter(s)
max sd	Maximum soil depth
MET	Norwegian Meteorological Institute
mm	Millimeter(s)
NGU	Geological Survey of Norway
NTNU	University of Science and Technology
NVE	Norwegian Water Resources and Energy Directorate
NWP	Numerical Weather Prediction
Pa	Pascal
par sd	Soil depth parameter
rad	Radians
TIF	Tagged Image File Format
TN	True Negative
TP	True Positive

1. INTRODUCTION

1.1. Background

Landslides are natural phenomena that can affect the lives of people exposed to this hazard. The Global Fatal Landslide Database, an initiative of The University of Sheffield, reports that between 2004 and 2016, landslides caused the death of 55,997 people worldwide (Froude & Petley, 2018). In response, humans have sought ways to reduce its risk by studying the underlying mechanisms that drive landslides (Holtz and Kovacs, 1981), thus seeking methods to stabilize slopes and even forecast future instability (Chae et al., 2017).

During a landslide event, part of a slope's mass dislodges and slides downhill (SGC, 2017). Two main zones can be identified in a landslide after an event: the failure surface from which the mass was removed (PMA, 2007) and the zone covered by the dislodged material known as the runout (Scheidl et al., 2013). According to van Zuidam (1986), this process can evolve with slopes greater than 8°, but it intensifies above 16°. The materials where landslides develop range between rock and soil (SGC, 2017), where the soil is the product of rock weathering and erosion (Holtz & Kovacs, 1981). A landslide occurs when the joints between the slope's materials are weakened. This slope instability can be caused by natural reasons, such as loss of cohesion during rocks' chemical disintegration (Holtz & Kovacs, 1981) or human intervention such as mining or road construction (SGC, 2017).

A landslide is considered a natural hazard if it is within proximity of people or assets that could be affected by its occurrence. A preventive approach for communities settled in such susceptible areas is to include landslide risk assessments in their urban planning (Thiebes et al., 2014) and restrict land uses accordingly (SGC, 2017). Nevertheless, it is more practical to apply mitigation measures when the hazard is unavoidable. Structural methods such as retaining walls or soil reinforcement can be adopted to stabilize slopes (Das, 2011). These measures are generally expensive, especially for smaller communities or extensive areas with multiple unstable zones (Thiebes et al., 2014). Another alternative is to relocate the population, although it is inconvenient from a social perspective as people may have developed solid cultural roots in their lands (P. de O. Melo et al., 2017).

In this context, non-structural mitigation strategies such as Landslide Early Warning Systems (LEWSs) can be implemented to protect the population and reduce losses (Calvello, 2017; Segoni et al., 2018). LEWSs gained recognition after the "Hyogo Framework for Action 2005-2015", where the importance of establishing early warning systems for natural disasters was highlighted (UNISDR, 2006), and it was further incorporated within the "Sendai Framework for Disaster Risk Reduction 2015-2030" (UNISDR, 2015).

In addition to the slope instability, a particular event called a 'trigger' is generally required to denote a landslide. The most common natural triggers of landslides are rainfall and earthquakes (SGC, 2017). Existing LEWSs are mainly built to predict weather-induced landslides because precipitation forecasts allow enough time to prepare, although forecast uncertainty affects the quality of landslide predictions (Guzzetti et al., 2020).

The water vapor condensation that originates the precipitation occurs in the atmosphere and can be monitored by satellites and other devices. Furthermore, Numerical Weather Prediction (NWP) models have been operating since the 1950s (Sene, 2016). In the last 30 years, the increase in computational power to solve numerical models and assimilate observations has allowed better weather predictions (Benjamin et al., 2019). Scientific efforts also aim to integrate weather forecasts with other related natural phenomena predictions. For instance, Schumacher et al. (2021) describe an approach for predicting excessive rainfall. This weather event is more localized and complex to estimate than others of similar nature. The proposed approach involves reanalyzing the weather model results and historical data with Machine Learning techniques. The study's motivation for detecting excessive rainfall was to predict flash floods, but it could be extended to other applications in the future, like landslides. Consequently, available data and current scientific advances in weather forecasting make the prediction of rainfall-induced landslides feasible.

LEWSs can be grouped into two categories according to their scale (Calvello, 2017; Piciullo et al., 2018). Local LEWSs focus on specific and identified slopes with monitoring devices such as extensometers, rain gauges, piezometers, and others (Pecoraro et al., 2019). For instance, in Italy, the Ancona Early Warning Center was founded in 2009 to monitor a 220 Ha landslide that caused a major disaster in December 1982 after six days of prolonged rain (Cardellini & Osimani, 2013). On the other hand, territorial LEWSs cover larger areas, from catchment to national scale. Unlike local LEWSs, the exact location where future landslides may occur is unknown (Piciullo et al., 2018). Territorial LEWSs operate over regions previously identified as susceptible to landslides (Chae et al., 2017). The Norwegian Landslide Forecasting and Warning Service is an example of an operational territorial LEWS on the national scale (Krøgli et al., 2018).

Piciullo et al. (2018) proposed a conceptual framework for territorial LEWSs composed of the following four parts represented in Figure 1:

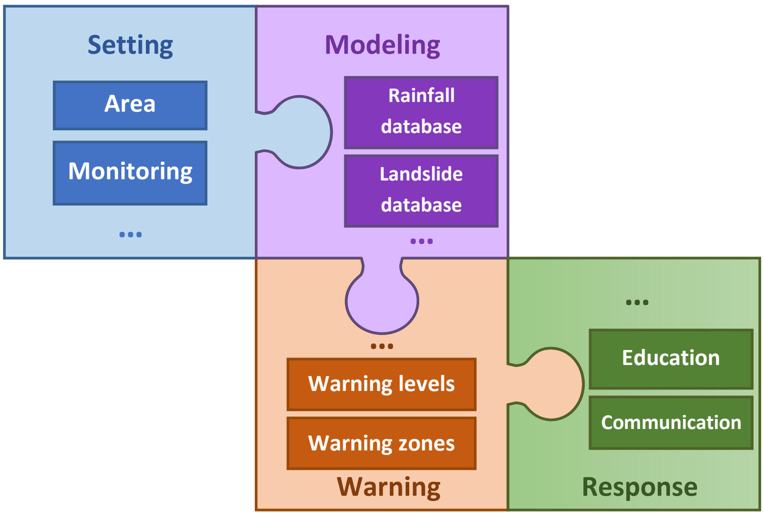


Figure 1. Conceptual framework for territorial LEWSs proposed by Piciullo et al. (2018).

1. **Setting:** Geology, geomorphology, landslides characteristics, hydrometeorological regimes, and exposure and vulnerability of the elements at risk, among other conditions, are considered for the landslide hazard and risk assessment (Calvello, 2017; Piciullo et al., 2018). It is decided which and how time-dependent variables will be monitored, together with the LEWS scale (Segoni et al., 2018). These variables will serve as inputs for a model for predicting slope instability. In territorial LEWSs, precipitation is the primary variable monitored and forecasted (Piciullo et al., 2018).

2. **Modeling:** A model is required to produce landslide forecasts. Landslides are particularly difficult to predict due to the complexity of the failure mechanisms (Xu et al., 2011; Yu et al., 2019). The model results should be accurate enough to forecast major landslide events while keeping a low rate of false alarms to prevent losing the public's credibility (Greco & Pagano, 2017; P. de O. Melo et al., 2017). There are mainly two groups of approaches employed for landslide forecasting (Chae et al., 2017; Guzzetti et al., 2020):

a. Data-driven methods use data from previous events to estimate empirical relationships or probabilities of landslide occurrence. The premise is that similar variable configurations that produced past landslides will also generate future events (Chae et al., 2017). These methods have proven efficient in numerous cases, from the most straightforward data interpolations to the advanced techniques of Machine Learning (Wang et al., 2021). Most territorial LEWSs examined by Guzzetti et al. (2020) rely on rainfall thresholds which fall in the data-driven category. The limitations are related to the change in conditions that trigger landslides, as these models cannot predict events without prior evidence (Chae et al., 2017).

b. Physically-based methods attempt to replicate natural phenomena using mathematical approximations. These models were initially designed to calculate a single slope's profile (two-dimensional) stability by breaking down the physical process into force summations (Bout et al., 2018; Verruijt, 2018a). Nowadays, physically-based models can be applied in broader areas such as catchments of a couple of square kilometers (van den Bout et al., 2022). The most sophisticated models simulate hydrology, slope stability and runout processes in succession. Additionally, pixel-wise calculations using raster layers become three-dimensional models when multiplied by elevation (Bout et al., 2018).

   An advantage of physically-based models is their ability to simulate failure and runout development (Fan et al., 2017). Runout modeling is of interest to landslide forecasting (Segoni et al., 2018), especially those of debris flow type, as it provides information about the distance traveled, the volumes released, the velocity, and the force applied when a debris flow collides with obstacles in its path (R. Melo et al., 2018). Even though accurately replicating initiation areas remains a significant challenge, in recent studies, physically-based models have shown sufficient capability to reproduce general patterns and displacement distances of mass movements (Bout et al., 2018; Fan et al., 2017). In addition, progress has been made in research for the better implementation and performance of runout models (Fan et al., 2017; van Asch et al., 2014). In the latest review found on existing territorial LEWSs by Guzzetti et al. (2020), it is mentioned that the runout is not included in the current operational models. However, it might be helpful to identify potentially affected zones far from the initiation area.

   Constraints of physically-based models are mainly related to the input data quality and availability (Canli et al., 2018; Guzzetti et al., 2020). The more heterogeneous the modeled region, which usually occurs over larger terrain extents, the more field information will be required to establish the model parameters (Chae et al., 2017). Often, numerous assumptions or approximations within a physically-based model are related to the subsoil configuration and its materials' behavior (van den Bout et al., 2021). Therefore, the uncertainty in the input data and the modeling decisions are transcendental in the quality of the results.

3. **Warning:** Every time a landslide forecasting model runs, a result is generated. It should be interpreted to determine if actions are required in response. Commonly, rainfall thresholds are used to automate the interpretation and group the response actions into warning levels (Piciullo et al., 2018). Preventive actions are required if the rainfall records are higher than determined critical values (Greco & Pagano, 2017). There should be enough time in advance since the warning is issued for preventive measures to be carried out (Segoni et al., 2018). The same warning level is raised in grouped areas that usually correspond to administrative units (Piciullo et al., 2018).

4. **Response:** A LEWS should be articulated with authorities, emergency response organizations, and, more importantly, the community. Civilians must receive information about the system's functioning to respond effectively to warnings (Piciullo et al., 2018). The public should also be aware of the system's limitations and potential false alarms (R. Melo et al., 2018).

Territorial LEWSs are relatively recent, as the majority emerged during the 21st century (Guzzetti et al., 2020; Piciullo et al., 2018). The implementation and interaction of its four parts shown in Figure 1 are relevant for a system's overall effectiveness (Piciullo et al., 2018). Evaluating the performance of a territorial LEWS remains a challenge because there are still no systematic established methods to accomplish this task (Guzzetti et al., 2020). One of the most critical aspects of the performance is the number of correctly issued warnings and false alarms, which is directly related to the predictive capacity of the landslide forecasting model. While numerous factors influence the production of accurate forecasts with enough lead time for early warning, there is still research to be explored to improve them (Piciullo et al., 2018).

The previous section introduced an overview of landslide prediction state of the art for its application in territorial LEWSs and its conceptual framework. Below is the identified research gap during the literature review that supports the investigation carried out for this thesis, followed by the study's objectives.

1.2. Research Gap

In existing territorial LEWSs, rainfall thresholds are frequently used to raise warnings in susceptible regions, usually administrative units (Piciullo et al., 2018). One of the points to improve is the scale of warnings since alerts are issued for areas as large as municipalities or counties. To that end, the model should be able to identify the probability of landslide occurrence in greater detail, which can be implemented in slope units with documented high susceptibility.

For forecasts in those specific locations, failure and runout physically-based modeling arises as an approach that could produce local warnings by predicting the impact area of landslides. Previous studies show significant improvements in runout modeling, and its implementation provides information about affected zones beyond the landslide initiation area. As the potential of runout modeling has not yet been explored in established territorial LEWSs, it still requires research regarding its ability to produce timely and accurate predictions with forecast precipitation.

1.3. Objectives and Research Questions

1.3.1. General Objective

To investigate the potential of integrating rainfall-induced failure and runout physically-based modeling in territorial LEWSs to predict local landslide impact areas using forecast precipitation.

1.3.2. Specific Objectives

1. To optimize the parameters employed for real-time physically-based modeling of failure and landslide runout using observed precipitation data.
   1.1. What are the appropriate initial hydrological conditions (initial soil saturation, groundwater height, and wetting front height) for physically-based modeling of a particular landslide event?
   1.2. What is the impact of different slope stability parameters (soil depth, cohesion, and internal friction angle) on the simulated failure depth?
   1.3. What is a suitable evaluation metric to compare observed and simulated failures?
2. To assess the accuracy of the predicted failure and runout using forecast precipitation data.
   2.1. What is the spatial overlap between the observed and predicted landslide failure and runout?
   2.2. To what extent the runout modeling can improve the applicability of a territorial LEWS?

2. STUDY AREA

2.1. Study Area Overview

This thesis study area is located in Norway, where failure and runout physically-based modeling could be additionally applied in the operational territorial LEWS to generate more localized alerts. The Norwegian Landslide Forecasting and Warning Service began operating in 2013 and emerged as an extension of the Flood Forecast Service running since 1989. It is managed by the Norwegian Water Resources and Energy Directorate (NVE) under the Ministry of Petroleum and Energy. The Norwegian territorial LEWS forecasts rainfall- and snowmelt-induced landslides such as debris flows, debris floods, debris avalanches, and others (Krøgli et al., 2018). Some essential aspects of this system are described below.

The landslide forecasting method in Norway has foundations in hydrological modeling, as the LEWS arose from the Flood Forecast Service. The basis of its operation is the Gridded Water Balance (GWB) hydrological model (Beldring et al., 2003). The GWB model uses precipitation and temperature forecast inputs to generate predictions of hydrological variables employed to perform the daily landslide hazard assessment (Krøgli et al., 2018).

The weather forecasts are provided daily by the Norwegian Meteorological Institute (MET), which delivers nine days forecasts (Krøgli et al., 2018). The short-term forecasts for the first and the second day are derived from the downscaled AROME-MetCoOp model for Norway and Sweden with a horizontal resolution of 2.5 km (Müller et al., 2017). The long-term forecasts from the third to the ninth day come from the Integrated Forecasting System from the European Centre for Medium-Range Weather Forecasts (ECMWF-IFS) model with a horizontal resolution of 9 km (Krøgli et al., 2018). MET resamples these gridded data into 1 km^2 cell size during post-processing (Homleid et al., 2021b). Hence, the inputs and outputs of the GWB and other hydrological models used by NVE have a grid size of 1 km^2 (Krøgli et al., 2018). Likewise, because of the forecast 24 hours lead time, the temporal resolution of the LEWS is one day.

Müller et al. (2017) describe improvements in AROME-MetCoOp predictions compared to ECMWF-IFS, such as better estimates of the spatial extent and magnitude of precipitation events. Verifications of the NWP models performed by MET in 2021 report that both AROME-MetCoOp and ECMWF-IFS show errors when forecasting either minimum or excessive amounts of precipitation and tend to overestimate, especially in the second case (Homleid et al., 2021b, 2021a). However, Krøgli et al. (2018) account that 95% of the landslide assessments during the analyzed Norwegian LEWS operational history have been accurate. Thus, it can be considered that the weather forecasts have enough accuracy to support the current system.

The hydrological variables calculated by the GWB model and historical landslide data from Norway were analyzed by Boje et al. (2014) to establish thresholds that define danger levels that indicate the probability of landslide occurrence: low, moderate, considerable, and high. These hazard levels do not contain any information on the impact of runout as the statistical analysis only considered the potential slope failure.

The variables found to be statistically significant in landslide occurrence in Norway are (1) the water supply from rain and snowmelt and (2) the degree of soil saturation. Therefore, unlike most territorial LEWSs, the Norwegian system does not directly rely on rainfall thresholds. Instead, the probability of landslide occurrence is derived from the relationship of two variables computed from weather forecasts.

The tool used by NVE for the daily landslide hazard assessment is the website xgeo.no, where all the layers involved in this process can be consulted (Stranden et al., 2020). The layer Hydmet Geo reflects the forecasted danger levels in a 1 km^2 grid size raster covering the Norwegian territory. For each forecast day, a layer is created with the weather forecasts provided by MET. Hence, there are nine layers available for the nine forecast days. The same applies to other hydrological variables relevant to the daily landslide hazard assessment.

The warnings with the danger levels are published on the website varsom.no for three forecast days (Krøgli et al., 2018). Usually, just the first and the second forecast days are employed to raise and change warning levels. The third forecast day is also checked daily but only considered for warnings in extreme or prolonged precipitation scenarios. The following six forecast days help detect upcoming weather events or maintain warning levels. Relevant hydrological variables not included in the danger levels calculation such as runoff, snowmelt, groundwater table level, soil saturation, and soil frost are also examined to ensure the predictions match. This is one way to address forecast uncertainty, as well as compare GWB results with other hydrological models used in the Flood Forecast Service. The previous information was obtained from direct communication with Graziella Devoli, part of the NVE personnel in charge of the Norwegian Landslide Forecasting and Warning Service.

The following section focuses on the specific study area selected to perform the analysis of this thesis.

2.2. Study Area Description

To collect data for advancing research in landslide monitoring, NVE carried out a project titled KlimaDigital between 2020 and 2021 in partnership with the SINTEF Research Institute and the Norwegian University of Science and Technology (NTNU) in Trondheim, Norway. Depina et al. (2021) and Oguz et al. (2022) documented the results of this KlimaDigital project. Its objective was to select locations to install sensors for landslide monitoring in susceptible areas close to Trondheim. Criteria such as recurrent landslide reports, photographic evidence, road accessibility, and proximity to weather stations were considered to perform a multicriteria analysis. Finally, several landslide monitoring devices were installed in four preferred locations.

This thesis study area is the KlimaDigital Location 1: Kjellberget. Two main reasons support that decision: (1) it is the best-documented zone in Depina et al. (2021), and (2) there is additional information in free-access Norwegian online databases, which provided sufficient material to conduct this research. The data is presented in detail in Section 3.1: Data Preprocessing.

Figure 2 depicts the Kjellberget setting on various scales. It is located in central Norway (Figure 2A), Trøndelag county (Figure 2B), about 60 kilometers away from the city of Trondheim, in a basin along the E14 road and the Stjørdalselva river in the Stjørdal municipality (Figure 2C). In Figure 2D, the larger black rectangle named 'Location 1' encloses the Kjellberget area of almost 0.5 km^2. Within it, three geomorphological zones can be distinguished from north to south: (1) a steep slope with its summit at the top, extending until a (2) Quaternary deposit which ends in the E14 road, where (3) the Stjørdalselva river valley begins and covers the bottom of the figure. The Quaternary deposit is particularly susceptible to debris avalanches and debris flows that occasionally cause road blockages (Depina et al., 2021). The dashed black rectangle 'Location 1.1' outlines four reported landslide events.

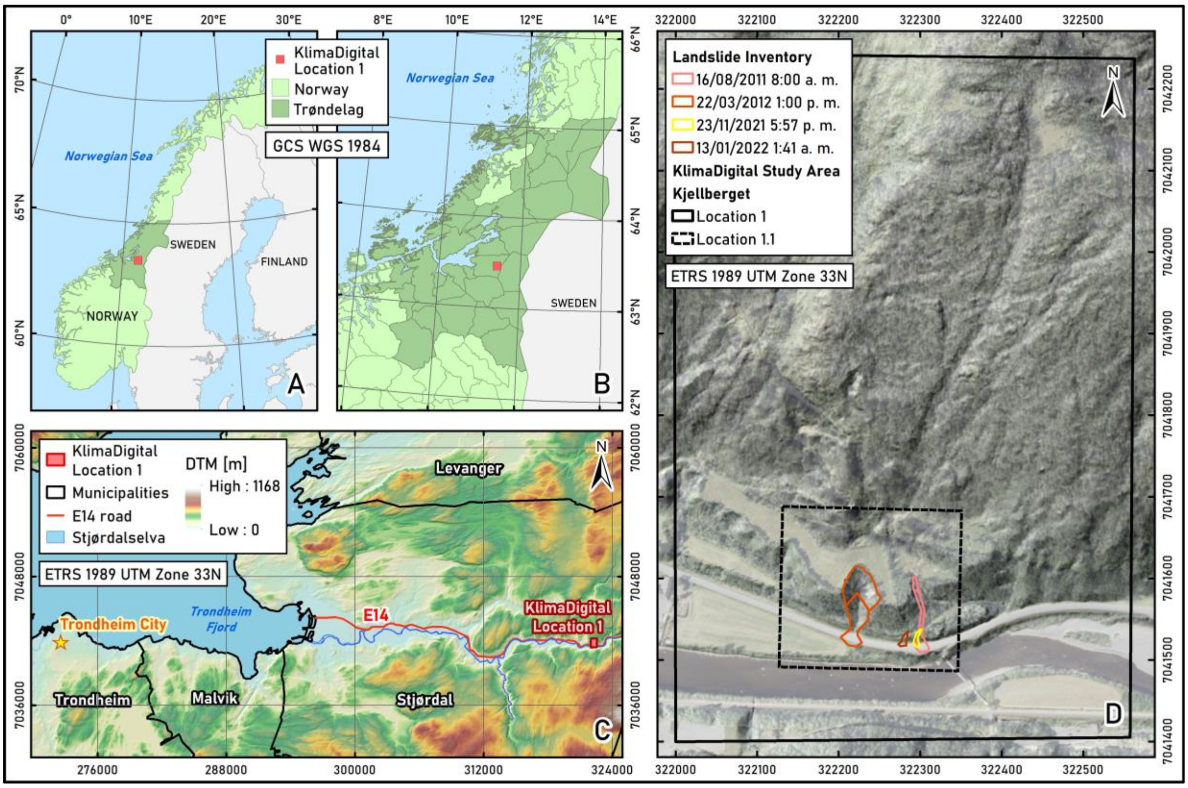


Figure 2. This thesis study area is derived from the KlimaDigital project carried out by NVE for landslide monitoring (Depina et al., 2021). It is located in central Norway (A), Trøndelag county (B), about 60 kilometers away from Trondheim city, in a basin along the E14 road and the Stjørdalselva river in the Stjørdal municipality (C). The Location 1 rectangle encloses the Kjellberget area, and Location 1.1 outlines four reported landslide events (D). Sources: shapefiles: geonorge.no, Digital Terrain Model (DTM): hoydedata.no, aerial photograph: norgeibilder.no, landslide inventory: skredregistrering.no.

The landslide inventory displayed in Figure 2D was obtained from skredregistrering.no, a Norwegian website for reporting landslides (Krøgli et al., 2018). The 2011 and 2012 events are briefly described in the KlimaDigital report (Depina et al., 2021), while the 2021 and 2022 are documented on regobs.no, an online application for observations related to natural hazards in Norway (Ekker et al., 2013). All four events are presumed to have originated from shallow landslides and can be characterized as highly saturated small-volume debris flows, which in 2011 and 2012 were enough to cover the road, but in 2021 and 2022, were minimally dispersed on the north roadside.

As the 2011 event occurred in August, during the midsummer season, there was no ice cover left from the winter because it had already melted. Hence, it can be assumed that the 2011 landslide was triggered solely by rainfall. In contrast, 2011, 2021, and 2022 landslides were reported in winter. Thus, presumably, snowmelt was involved in addition to rainfall in their triggering. Further information about weather conditions during the landslide events is presented in Subsection 3.1.6: Precipitation.

3. METHODOLOGY

This thesis explores the possibility of producing local warnings for territorial LEWS by applying physically-based failure and runout modeling to predict landslides' impact areas. An additional section in the modeling could be included in established systems where specific locations within the regions indicated by the danger thresholds are further analyzed by implementing physically-based modeling of slope failure and runout. These locations would correspond with highly susceptible slopes, such as places with recurring landslide reports. At this level of detail, running a physically-based model with forecast precipitation is feasible. The established territorial LEWS selected for this thesis is the Norwegian Landslide Forecasting and Warning Service, and the specific location is Kjellberget, an area of around 0.5 km² in central Norway.

Figure 3 is a methodology flowchart that summarizes the procedures in the analysis. Numbers 1 and 2 on the left indicate the stages to achieve this study's objectives. The first objective is to calibrate a physically-based model that can reproduce the failure and runout of landslides reported in Kjellberget. Hence, the first step was to determine related significant physical processes to build a model. Different computer applications can assist the model construction and implementation nowadays.

Several physically-based models for runout simulations have been previously tested in Norway (NGU, 2021), such as Flow-R (Fischer et al., 2012), RAMMS (Fille, 2017), and TRIGRS (Schilirò et al., 2021). However, these approaches are inconvenient in a territorial LEWS because it is necessary to indicate the initiation area and released volume by hand. Thus, the modeler's expertise and knowledge of the simulated area become a determining factor in the accuracy of the prediction.

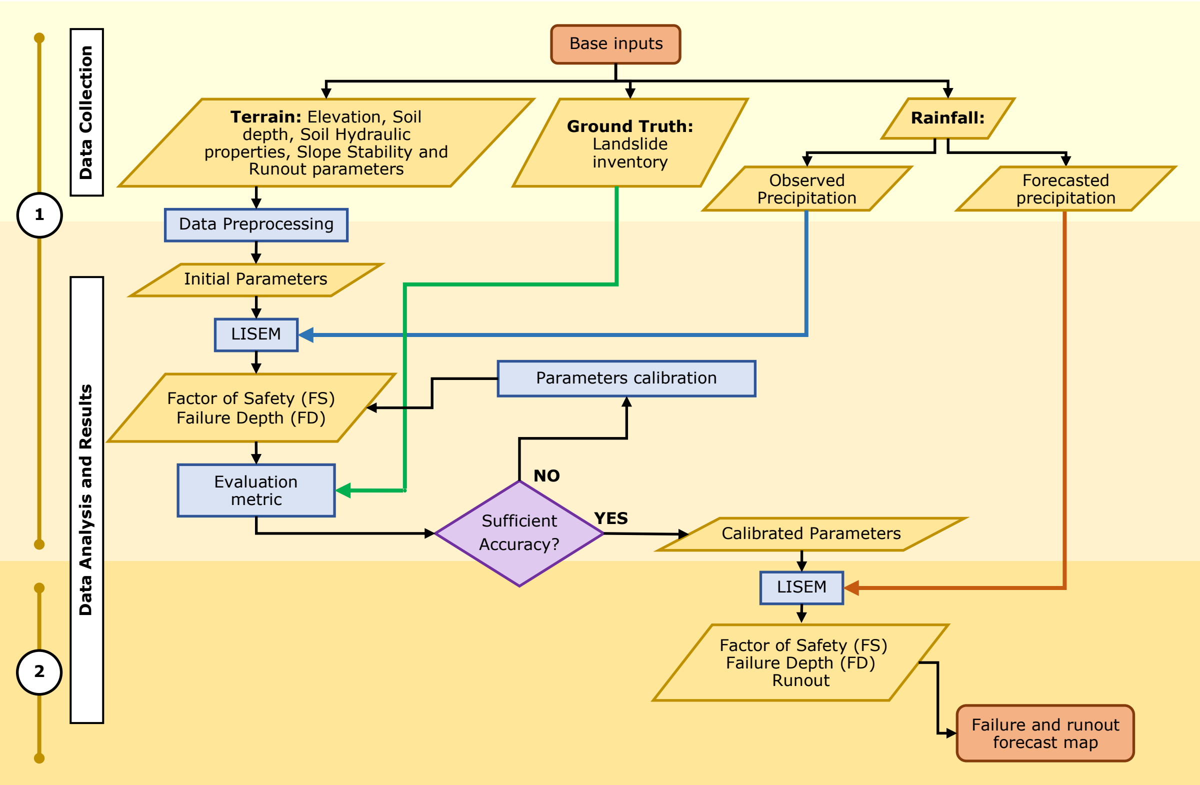


Figure 3. Methodology flowchart. Numbers 1 and 2 on the left indicate the stages to achieve this thesis' objectives. The overall process can be divided into Data Collection and Data Analysis and Results.

The LISEM software was selected to perform the modeling in this thesis analysis. This product has tool adaptations of mathematical approximations proposed by different authors related to natural hazards (Bout et al., 2018). The user can build a physically-based model through a script by including the selected tools and arranging it in the required order, allowing flexibility in the model setup. In addition, LISEM has tools available to calculate a landslide failure's location using the Factor of Safety (FS) method (Holtz & Kovacs, 1981). Moreover, within the same tool, LISEM calculates the Failure Depth (FD), which can be used to derive the initial volume for the runout. Thus, the model directly calculates the initiation area and released volume. Section 3.2: Model Setup explores the modeling procedure.

Once the necessary data to run the model is known, the second step is to collect and organize it. In Figure 3, the base inputs are clustered into three groups: Terrain, Ground Truth, and Rainfall. Terrain includes information related to elevation, soil depth, soil hydraulic properties, slope stability and runout parameters. Ground Truth refers to the field observations from skredregistrering.no (Krøgli et al., 2018), named 'landslide inventory' Figure 2D. Rainfall implies precipitation in liquid form, and it is used in the analysis for different purposes depending on whether it is observed or forecasted. The base inputs are preprocessed to convert their formats and units to those required by the LISEM tools. Section 3.1: Data Preprocessing is dedicated to information sources and data manipulation.

The LISEM box in Figure 3 represents the physically-based model built with the help of the tools in this software. With the initial parameters obtained from the preprocessing and observed precipitation, the model returns FS and FD. Calibration is expected to find a failure that matches the initiation area and can reproduce the runout in the landslide inventory. Soil depth, cohesion, and internal friction angle (IFA) were selected for the calibration because these parameters directly influence the magnitude of FS (Johari & Javadi, 2012). Different evaluation metrics are applied to assess the accuracy of the simulated failures. The calibration continues until a combination of calibrated parameters generates enough accuracy to proceed with the second objective of this thesis. More details can be found in Section 3.3: Calibration.

The second objective focuses on evaluating the failure and runout predictions produced by running the LISEM model with the calibrated parameters and forecast precipitation. Evaluation metrics can be applied again to assess the accuracy of the forecasted landslide impact area. However, that is only possible if the calibrated failure releases enough volume to reproduce the runout of the observed landslides.

Afterward, Chapter 4: Results reports the most significant outputs of all the analysis stages.

3.1. Data Preprocessing

This segment describes the data sources and preprocessing procedures to prepare the inputs for the physically-based model introduced in Section 3.2: Model Setup.

Table 1 summarizes the data sources. It is organized according to the clustering proposed in Figure 3. Relevant characteristics of the data are listed below:

- The coordinate system for all the layers is EPSG: 25833 - ETRS89 / UTM zone 33N – Projected.
- Except for the drag coefficient, the model inputs should be in raster format. In LISEM, it is recommended to work with TIF (Tagged Image File Format) because it stores the georeference.
- The elevation, a Digital Terrain Model (DTM) of 1 m^2 resolution, was used as a reference map, and all other inputs were resampled to match its cell size. This provided a good balance between the quality of the elevation, to which flow models are very sensitive (Leitão et al., 2008), and the computation time.

- Initial raster resolutions for elevation, precipitation, and aerial photographs are reported in Table 1. The rest of the data were originally punctual values, namely, a number reported in a coordinate or assumed from literature. Aside from cohesion and manning's n, other parameters were transformed into raster format by creating constant value layers.

Table 1. Data sources. The original resolution of data obtained in raster format is included.

	Input Data	Source	
Terrain	Elevation	Digital Terrain Model (DTM): hoydedata.no • NDH Malvik-Stjørdal 2pkt Prosjektrapport (Terratec AS, 2016). ○ DTM resolution: 1 m^2	
Terrain	Soil depth	NADAG Database (ngu.no) • Geoteknisk Datarapport E14 TS-tiltak Nr. Vd1399A (Statens vegvesen, 2016)	
Terrain	**Soil Hydraulic Properties**		
Terrain	Saturated Hydraulic Conductivity	Derived from SPAW statistical correlations (Saxton & Rawls, 2006): • Granulometry obtained from Statens vegvesen (2016) • Organic matter: SoilGrids250m 2.0 (Hengl et al., 2014) ○ Resolution: 250 m	
Terrain	Porosity		
Terrain	Field Capacity		
Terrain	Pressure head	Assumed from Klimadigital suction measurements (Depina et al., 2021)	
Terrain	**Slope Stability Parameters**		
Terrain	Cohesion	Direct shear tests performed during the KlimaDigital project (Depina et al., 2021), rock strength properties (Wyllie & Norrish, 2017)	
Terrain	Internal Friction Angle (IFA)		
Terrain	Particle density	Assumed from soil particle density common values (Haan et al., 1994)	
Terrain	**Runout Parameters**		
Terrain	Manning's n	Aerial photographs for landcover map creation: Norge i bilder (Geovekst, 2020) ○ Meråker 2010, resolution: 0.1 m ○ Trøndelag 2014, resolution: 0.25 m	Manning's n values table: orst.edu (Chow, 1959)
Terrain	Dominant grain size	Granulometry obtained from Statens vegvesen (2016)	
Terrain	Drag coefficient	Assumed from Pudasaini (2012)	
Ground Truth	Landslide records	Norwegian mass movements database: skredregistrering.no (Krøgli et al., 2018) regobs.no (Ekker et al., 2013)	NVE landslide inventory: Provided by NVE personnel Graziella Devoli (gde@nve.no)
Rainfall	Precipitation	Observed: xgeo.no (Stranden et al., 2020) ○ Resolution: 1 km^2	Forecasted: Provided by NVE personnel Jess Andersen (jea@nve.no) ○ Resolution: 1 km^2

The most significant aspects in the preparation of each data group are mentioned in the following subsections.

3.1.1. Elevation

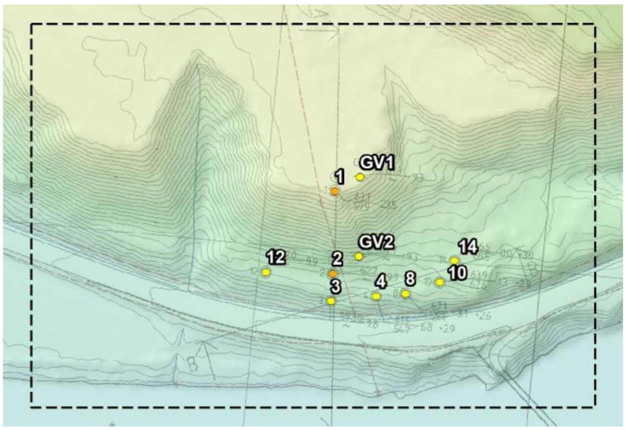
The DTM employed in this analysis is a 1 m^2 resolution raster created from laser scans as part of a Norwegian project to establish a detailed nationwide elevation model (Terratec AS, 2016). It is freely available on the website hoydedata.no. No corrections were made to the DTM since no issues were found when checking the flow direction and flow accumulation.

3.1.2. Soil Depth

For this analysis, a soil depth model was built using the SteadyStateSoil LISEM tool adapted from von Ruette et al. (2013). This model is based on the dynamics of erosion and deposition governed by elevation and slope. On high ground ridges, erosion is active, and the soil depth is shallow. On the contrary, in low-lying flat terrain, deposition is prevailing, and the soil layer is thick. The approach proposed by von Ruette et al. (2013) is a useful approximation for computing a soil depth layer for physically-based modeling. Nevertheless, as soil production depends on many other factors besides physics, constructing an accurate soil depth model remains difficult.

The soil depth model was calibrated with drilling depths measurements found in a geotechnical report from Statens vegvesen (2016), which is the Norwegian Public Roads Administration. It is available in the NADAG Database at ngu.no, the Norwegian Geological Survey (NGU) website.

Figure 4 presents the location of the 10 boreholes drilled at Kjellberget reported by Statens vegvesen (2016). The boreholes 1, 4, 8 and 10 reached the bedrock, and the 14 was directly drilled into outcropping rock. All drilling depths listed in Figure 4 were used to calibrate the soil model, even if the perforation did not reach the maximum soil depth because there is no other information available.



Borehole number	Drilling depth [m]	Reached bedrock	Sieve analysis
1	23.5	**Yes**	**Yes**
GV1	9.3	No	No
2	9.7	No	**Yes**
GV2	9.3	No	No
3	9.8	No	No
4	6.8	**Yes**	No
8	3.1	**Yes**	No
10	1.3	**Yes**	No
12	9.9	No	No
14	0	**Yes**	No

Figure 4. Boreholes performed in Kellberget reported in Statens vegvesen (2016). The background contour lines are from the map in the original report. The table on the right summarizes the soil drilling depth, if the borehole reached the bedrock and if a sieve analysis was carried out with samples from the site.

The calibration process consisted of varying the soil depth model by iterating two parameters within the *SteadyStateSoil* tool: (1) the soil production from rock weathering, (2) the speed of soil movement, and (3) a constant value outside the tool which is proportionally multiplied by the soil depth. The model with the minimum Root-Mean-Square Error (RMSE) between the soil depth values generated by the model and the drilling depths observations in Figure 4 was selected. The resulting soil depth model is presented in Figure 12, Subsection 4.1.1: Soil Depth Model.

3.1.3. Soil Hydraulic Properties

This group includes the variables that determine the water behavior inside the soil: saturated hydraulic conductivity, porosity, field capacity, and pressure head. While the first three depend on the configuration and size of the soil particles (Saxton & Rawls, 2006), the fourth varies with the water content, making its precise estimation a complex task beyond this thesis's scope (Zhang et al., 2018). Instead, a simple approach was applied to establish the initial soil moisture and pressure head in the model.

The geotechnical report from Statens vegvesen (2016) includes the results of two sieve analyses performed with samples collected in boreholes 1 and 2, depicted in Figure 4. Using the SPAW hydrology tool, the granulometry obtained from these tests was employed to derive values for saturated hydraulic conductivity, porosity, and field capacity. This product was created by Saxton and Rawls (2006) and relies on pedotransfer functions that consider grain size and organic matter content to compute soil properties. The organic matter was adopted from the SoilGrids database documented in Hengl et al. (2014).

From the physical perspective, the amount of water that flows through the soil is constrained by the porosity and field capacity. The first is the maximum available space that can be occupied by water within the soil (Holtz & Kovacs, 1981), and the second indicates the moisture level below which capillary forces retain the water (Butcher, 2004). Therefore, the initial soil moisture for water to flow through the soil column was set between porosity and field capacity.

Statens vegvesen (2016) also includes measurements of moisture content in boreholes 1 and 2, which were employed to estimate the initial pressure head with SPAW hydrology. Records from monitoring devices installed in Kellberget are also presented in the KlimaDigital document and depicted in Figure 5 (Depina et al., 2021). The pressure head value obtained with SPAW hydrology is within the range of reported suction, although it is a reasonably wide range.

Figure 5. Location of monitoring devices (D1.1 and D1.2) and weather station installed in Kjellberget during the KlimaDigital project (Depina et al., 2021). Also featured in the image is a trial pit site excavated to obtain soil samples. Background image: Stjørdal 2015 – norgeibilder.no.

Annex 0 details the magnitudes of the variables mentioned in this subsection.

3.1.4. Soil Stability Parameters

This group comprises cohesion, IFA, and particle density, soil parameters required for the slope stability estimation.

In section 2.2: Study Area Description, three zones are distinguished in Kjellberget according to the geomorphology: (1) a steep slope, (2) a Quaternary deposit, and (3) the Stjørdalselva river valley. The KlimaDigital project mentions that the steep slope is covered by a thin moraine layer, between 0.5 and 1 meter thick (Depina et al., 2021). During the project, direct shear laboratory tests were performed to obtain cohesion and IFA values of the Quaternary deposit and the moraine. The Quaternary deposit was sampled in Kjellberget, and the trial pit site is depicted in Figure 5. On the other hand, the moraine sample was collected in another KlimaDigital study area located two kilometers down the E14 road to the east of Kjellberget.

The values in Table 2 derived from KlimaDigital indicate that the moraine has less cohesion than the Quaternary deposit. However, in aerial photographs from norgeibilder.no (Geovekst, 2020), such as the background image in Figure 5, erosion is not visible on the steep slope but in the Quaternary deposit. Also, when checking Street View images from Google Maps (2010), the bedrock is observed to outcrop along some places on the E14 road northside. Consequently, in this thesis analysis, the moraine was not considered. Instead, the steep slope composition was assumed as bedrock material, and its cohesion was assigned according to Wyllie and Norrish (2017) 'hard rock with discontinuities' reference. Oguz et al. (2022) performed a calibration for landslide modeling at Kjellberget with data from KlimaDigital and the results obtained for cohesion and IFA are included in Table 2.

Table 2. Reported values for cohesion and IFA for the moraine and Quaternary deposit materials in the study area. The sources are mentioned in the table.

Material	Creation Method	Cohesion	IFA	Source
Moraine	Direct shear test	5.5 kPa	38.2°	KlimaDigital (Depina et al., 2021)
Quaternary deposit	Direct shear test	9.5 kPa	41.4°	KlimaDigital (Depina et al., 2021)
Quaternary deposit	Calibration	6 kPa	40°	Oguz et al. (2022)

According to the above, a cohesion map was created with two zones: the Quaternary deposit and the steep slope. In any case, the first has less cohesion than the second. The same value for the steep slope was also assigned to locations on the road where rock outcropping was identified with Google's Street View. The final cohesion map is presented in Subsection 4.1.2: Cohesion Map.

On the other hand, the IFA value reported on KlimaDigital (Depina et al., 2021) was applied to the entire study area, same as particle density that was assumed from common values (Haan et al., 1994). Efforts have focused on cohesion zonation because it can be modified intuitively to alter the slope stability of a specific location (Johari & Javadi, 2012). Namely, with more cohesion there is a greater probability of stability.

Thus, the model can be adjusted to produce failures where the landslide inventory suggests higher susceptibility or prevent them where there is no evidence.

3.1.5. Runout Parameters

The runout parameters considered in this analysis are Manning's n, dominant grain size, and drag coefficient. Besides, a basal friction angle is also required to simulate the runout (Pudasaini, 2012). Unlike the IFA, which is used for static situations, the basal represents the friction in motion. Still, the IFA was used as the initial approximation for the basal friction angle.

Manning's n is a coefficient used to relate a fluid's velocity to the surface's roughness (Arcement & Schneider, 1989). To estimate it, a land cover map was first created using aerial photographs from norgeibilder.no (Geovekst, 2020). Then, each cover type was assigned a value 'n' following the tables from Chow (1959) via orst.edu. The final map is presented in Subsection 4.1.3: Manning's n Map.

The dominant grain size was derived from the granulometry reported in Statens vegvesen (2016).

The drag coefficient indicates how solids and liquids move together in fluids such as debris flows. In the simulations, a value indicating liquids and solids moving together was employed based on descriptions in skredregistrering.no (Krøgli et al., 2018) and the photographs in regobs.no (Ekker et al., 2013). The drag coefficient magnitude is discussed in detail in Pudasaini (2012).

3.1.6. Precipitation

Precipitation records for the four events are presented below to examine their behavior before, during, and shortly after each landslide. The estimate of rain and snowmelt, onwards represented as 'rain+snowmelt', is added for the three events that occurred in winter. These records were obtained from xgeo.no, where the observed precipitation corresponds to an interpolation between captured values in weather stations (Stranden et al., 2020). At the same time, the rain+snowmelt layer is calculated by NVE with observed precipitation and temperature using a hydrological model (Stranden et al., 2020). Both variables report the millimeters of accumulated water in 24 hours starting at 6:00 am, which is the time when the rainfall measurements are recorded. Therefore, the temporal resolution of precipitation is 24 hours. At the same time, the spatial resolution is 1 km², as mentioned in Section 2.1: Study Area Overview. Figure 2D shows the inventory graphically, and the legend includes the time each landslide was triggered.

The study area falls between two grids. The upper cell covers a third part of the Kjellberget perimeter, matching the steep slope's summit. Thus, the values of the top and bottom grids are plotted separately. Precipitation forecasts provided by NVE personnel are also included for some dates, but this information is limited due to time constraints for additional data requests. The label on the forecasts' series corresponds to the issue date. The first record in the forecast series is observed, and the following days are forecasts. This arrangement was made to show the transition between the observed and forecasted values.

- **Landslide on August 16, 2011:** In Figure 6, the rainfall is essentially the same at the top and bottom grids for this case. Two small precipitation events occurred a few days before the landslide but probably did not influence the triggering much as the rainfall recorded on August 16. The landslide was registered at 8:00 am. As a result, when the landslide occurred, 69 mm of rain had fallen over the area since the previous day. In the following days, the rainfall lessened substantially.

  The precipitation forecasts issued on August 13, 14, and 15 show that the estimations improved as the lead time decreased. However, the forecast the day before the triggering underestimated the observed rainfall with a significant difference of almost 20 mm. The forecasts in Figure 6 were calculated using an earlier version of the weather models currently employed by NVE in daily landslide hazard assessment.

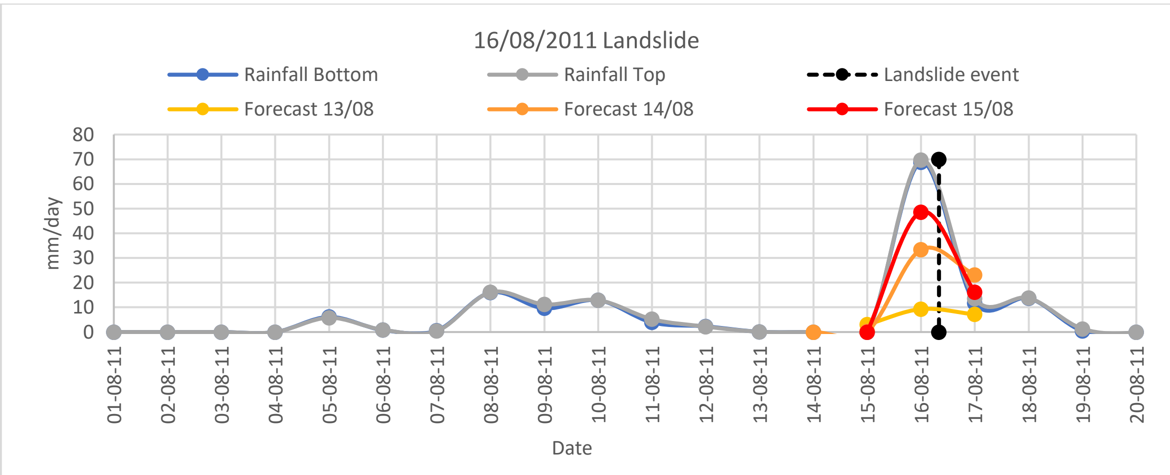


Figure 6. Rainfall records for the Landslide on August 16, 2011. The precipitation forecasts issued on August 13, 14, and 15, 2011, show that the estimations improved as the lead time decreased but still understated the triggering rainfall significantly.

- **Landslide on March 22, 2012:** The temperature increases at the end of March in Norway, and the ice cap left over from winter is melting. In Figure 7, the rain+snowmelt from the top grid is higher than the precipitation on both the top and bottom grids during rainfall events. It confirms that snowmelt added significant moisture to the soil in addition to rainfall, contributing to the landslide triggering.

  In this case, there was considerable precipitation between March 11 and 14, which decreased in the following days, but raised again on March 20, and even more on March 22. Due to the antecedent rain, the soil saturation was probably high before the March 22 precipitation that triggered the landslide. This event was reported at 1 pm, but since the rain is recorded at 6 am, it is unclear how much precipitation had fallen at the time of the triggering. The March 21 forecast completely failed to predict the triggering rainfall.

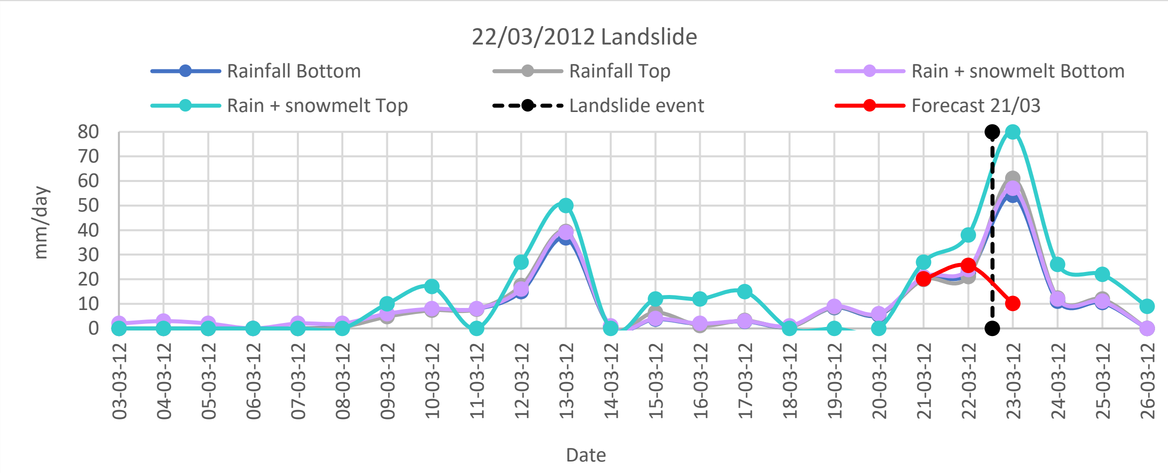


Figure 7. Rainfall records for the Landslide on March 22, 2012. The forecast precipitation issued on March 21, 2012, is included. Source: xgeo.no.

- **Landslide on November 23, 2021:** Figure 8 shows that in both top and bottom grids, rainfall and rain+snowmelt fairly coincide. This concordance is because November's low winter temperatures in Norway cause the water to freeze instead of melt. There were only small rains during the month's first half, but two significant rainfall events later occurred in the second half. This landslide was reported at 5:57 pm. On the landslide day, the rain+snowmelt was slightly higher than the rainfall. However, the latter seems more significant in the triggering. In addition, the soil may still have been saturated from the August 19 to 21 precipitation.

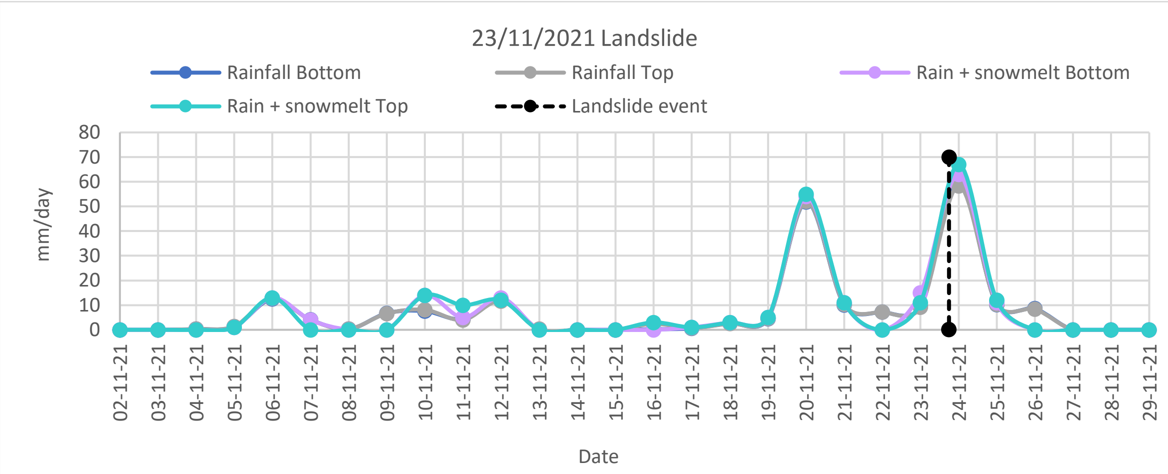


Figure 8. Rainfall records for the Landslide on November 23, 2021. Source: xgeo.no.

- **Landslide on January 13, 2022:** Figure 9 shows low rainfall from late December to mid-January. Thus, the rainfall reported on January 13 likely triggered this landslide. The record was registered at 1:41 am. In the following days, it continued to rain but with less intensity. The forecast issued on December 12 accurately predicts the rainfall on January 13 and its subsequent decrease on January 14. In this case, the rain+snowmelt was less than the rainfall alone. Hence, it is assumed that this event was entirely rainfall-induced.

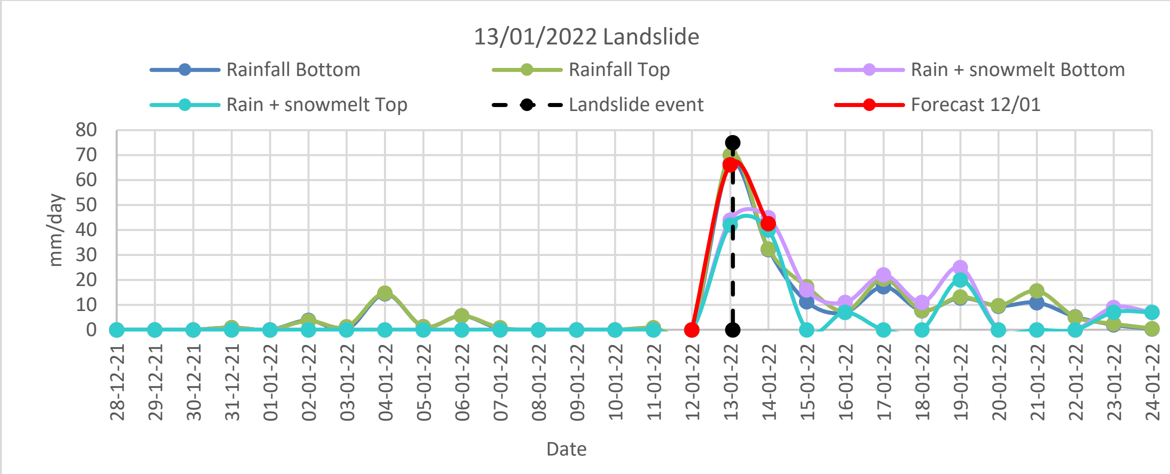


Figure 9. Rainfall records for the Landslide on January 13, 2022. The forecast issued on January 12, 2022, is included. Source: xgeo.no.

This subsection described how rain and snowmelt influence the occurrence of landslides depending on the season. It is worth mentioning that in Figure 7, Figure 8, and Figure 9, there are dates in which the value of rain+snowmelt is less than rainfall. It is a possible error of the hydrological model used for rain+snowmelt computation since it should be greater than or equal to rainfall. Nevertheless, rain alone seems to play a more critical role in triggering the analyzed landslides since all events coincide with days when the recorded rainfall was considerably high, around 60 and 70 mm.

3.2. Model Setup

The physically-based model used to simulate landslide failure and runout is described in this section. This model was built during an Internship at NVE and is further explained in Collazos (2022). Its most relevant aspects are summarized below.

The workflow diagram in Figure 10 is divided into three components according to the main processes simulated: hydrology, slope stability, and runout. The letters in italics in the blue boxes indicate the method employed to simulate the stated process. Each blue box has an adaptation in LISEM, and the authors will be mentioned when describing the tools. The flow is important because the tools' outputs are inputs for subsequent processes. The main modeling assumptions are included in the diagram.

Although the study area comes from the Location 1: Kjellberget polygon in Figure 2D, it is slightly different for the modeling. The hydrology and slope stability simulation area is around 120 m larger in the north and 80 m on the east side. It was expanded to capture better the water descending from the slope's summit towards the Quaternary deposit where the landslides have been reported. At the same time, it is approximately 80 m shorter on the south side, with the new limit set along the deepest part of the river. This configuration allows the water to flow from the Quaternary deposit to the lower boundary without stagnating in river depressions. The final hydrology and slope stability simulation area is 0.59 km² and can be observed in the soil depth and cohesion maps in Section 4.1: Preprocessing Results. The south bank of the river is included in the visualization to provide a context of the entire Kjellberget.

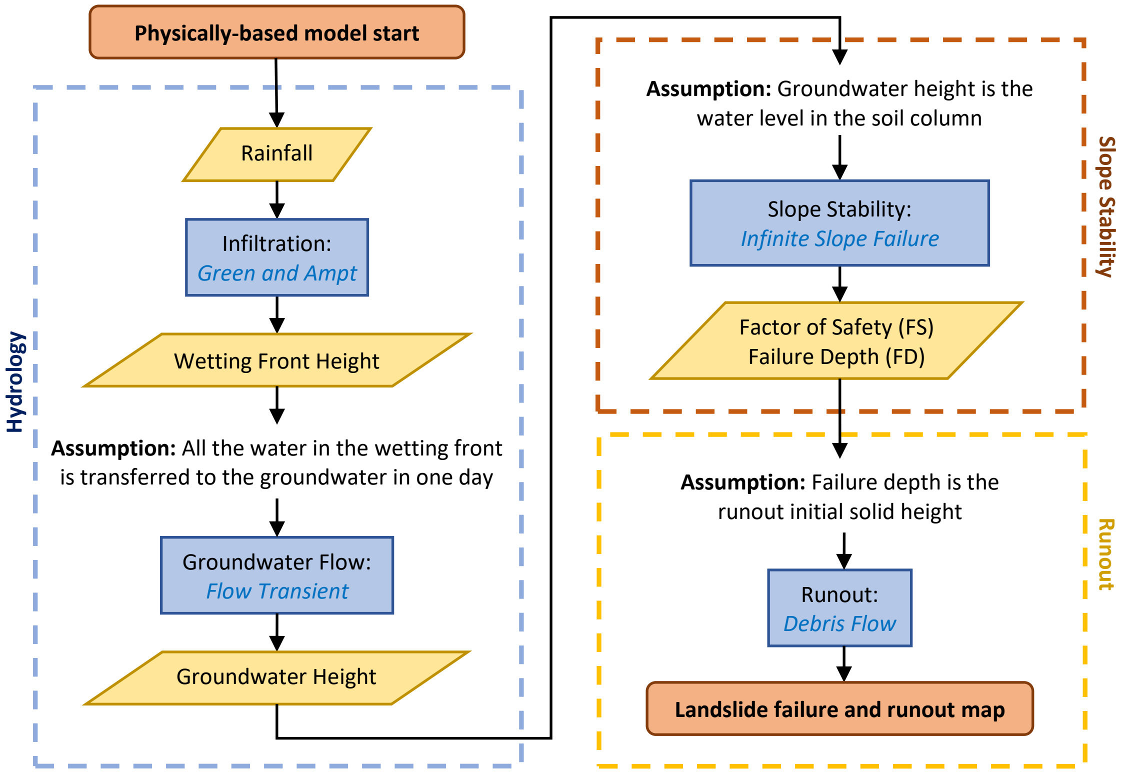


Figure 10. Physically-based landslide failure and runout model workflow. It was developed during an internship at NVE (Collazos, 2022).

3.2.1. Hydrology

The model depicted in Figure 10 starts by simulating infiltration with the GreenandAmpt tool named after the authors Green and Ampt (1911). This method was designed to replicate the diffusive percolation of water over an unsaturated zone for one infiltration event, meaning the tool needs to be restarted to simulate the next event. The infiltration produces a wetting front in the upper part of the soil column with higher saturation than the lower parts, where recent moisture has not reached yet. The observed and forecasted precipitation data have a temporal resolution of 24 hours. Thus, the accumulated rainfall in one day is modeled as an infiltration event.

Since there is no information on the subsoil structure or the time it takes for infiltrated water to descend through the soil column, it is assumed that the water in the wetting front is transferred to the groundwater at the end of one simulated day. Other processes in Figure 10 also stick to one-day timesteps. Likewise, daily timesteps are frequently used in hydrological models such as SWAT (Adla et al., 2019).

The wetting front resets to an unsaturated state after the infiltrated water is transferred to the groundwater. The tool employed for groundwater flow is FlowTransient, adapted from the depth-averaged Darcy method described by van Beek (2002).

The non-infiltrated water may be considered runoff, but it is removed from the model because this analysis focuses on runout generated by dislodged material from the slope. Hence, the effects of erosion by surface water are not included in this study.

Some days of antecedent rain are included in the simulations besides the precipitation of the triggering day in order to replicate precursor soil moisture conditions. It is expected that the soil will have a high saturation level if it has rained on the previous days and vice-versa.

In the Norwegian LEWS, the probability of landslide occurrence is usually identified with leading times of +24h and +48h (Krøgli et al., 2018). The proposed approach states that the physically-based model would run in specific locations once detected by the danger levels. It is assumed that the model in Figure 10 should be able to complete in no more than three hours because it is the current time between AROME-MetCoOp weather forecast runs (Homleid et al., 2021a). Thus, the model execution would be completed when new forecasts are available. Therefore, the antecedent rainfall to be included in the model should be sufficient to reflect the precedent soil moisture conditions but be able to run in the shortest time possible.

3.2.2. Slope Stability

After one simulated day, the groundwater height is assumed to be the saturated level in the soil column. The soil depth determines the soil column height. The Infinite Slope Failure is a method widely used in geotechnics (Verruijt, 2018b). It was selected to model slope stability in this analysis because it is suitable for reproducing shallow landslides (Chae et al., 2017). As mentioned in Section 2.2: Study Area Description, the landslide events examined in this study fall in the shallow category.

The tool employed to model the Infinite Slope Failure method was SlopeStabilityIS. Its implementation in LISEM is described in van den Bout et al. (2021). The tool calculates the FS per pixel over the simulated area. If FS is less than 1 in a pixel, it can be interpreted that the slope is unstable at that location. Consequently, the tool additionally computes the FD in that pixel. This approach produces FS and FD maps for each simulated day. If FS is above 1, meaning the terrain is stable, FD is zero.

3.2.3. Runout

For the runout, the simulation area was narrowed to a similar extent to Location 1.1 in Figure 2D. The extension is slightly smaller on the upper limit because it was not necessary to include the Quaternary deposit top as the landslide initiation areas are located on the south-facing slopes. Reducing the modeled region is meant to save computation time. In addition, the model is built with data and assumptions from the Quaternary deposit, and thus, it is worthless to include adjacent zones because the model may not be valid there. The final runout simulation area is 0.03 km^2 which can be observed in the extension of Subsection 4.1.3: Manning's n Map.

The runout was simulated with the FlowDebris tool adapted from the Two-Phase flow equations proposed by Pudasaini (2012), which can reproduce flows with different solid contents and flow properties. The initial solid height is the FD obtained from the slope stability minus the pore's voids characterized by porosity. The solid is assumed to be totally saturated, as descriptions from skredregistering.no (Krøgli et al., 2018) and images from regobs.no (Ekker et al., 2013) suggest. Hence, the height of the initial fluid is the porosity that is the available space in the solids.

The FlowDebris tool's timesteps are in seconds. The simulated time depends on the runout velocity to reach its final state. One hour is generally sufficient for solids to get their final position, but fluids may take longer. The resulting runout maps presented in Section 4.3: Model Results Using Observed Precipitation show the maximum height of the solids and fluids mixture in the modeled debris flow.

3.3. Calibration

This section describes the procedures for calibrating the physically-based model described in Section 3.2: Model Setup.

Calibration can be approached from different perspectives depending on the purpose of the study. In this thesis, the aim is for the proposed model to be capable of reproducing landslide impact areas. The model generates the runout volume from the FD. An accurate failure is expected to generate a runout that can replicate the landslide. Thus, the calibration was performed based on the FS and failure area obtained from the SlopeStabilityIS tool. Additionally, more simulations can be run in the available time for the thesis analysis by calibrating only the failure because it takes less computational time than runout.

The parameters selected for the calibration are soil depth, cohesion, and IFA. These three parameters are included inside the SlopeStabilityIS tool, which returns FS and FD. Johari and Javadi (2012) studied the impact of cohesion, IFA, and unit weight, determined by the soil depth, on the FS and demonstrated that the variation in these parameters directly influences the magnitude of FS.

Some additional remarks about data uncertainty regarding soil depth, cohesion, and IFA are mentioned below.

The soil depth model was obtained with the SteadyStateSoil LISEM tool adapted from von Ruette et al. (2013) and calibrated with field data from Statens vegvesen (2016). Nevertheless, five of the ten drilling depths did not report the maximum soil thickness, as explained in Subsection 3.1.2: Soil Depth. Furthermore, from the two zones identified in Kjellberget in Section 2.2: Study Area Description involved in hydrology and slope stability calculation, the drilling was performed only in the Quaternary deposit. Hence, information on the moraine depth covering the bedrock of the steep slope was not included in the calibration of the soil depth model.

Likewise, cohesion and IFA are also uncertain because only one soil sample from the Quaternary deposit was tested in the laboratory. Therefore, the entire study area has a single measured value for these two parameters.

The event selected for calibrating the model is the August 16, 2011's landslide. It is the only record in the inventory solely induced by rainfall, which is probably less uncertain than rain+snowmelt. When snowmelt is involved in the triggering, the rain+snowmelt layer can be integrated into the model, just like rainfall. However, rain+snowmelt is computed with a hydrological model, and rainfall is one of the main inputs (Stranden et al., 2020). In Subsection 3.1.6: Precipitation, it can be observed that in Figure 7, Figure 8, and Figure 9, rain+snowmelt values are occasionally lower than rainfall. Nevertheless, Rain+snowmelt is expected to be at least equal to rainfall.

Consequently, the observed rainfall is considered more reliable because the processes within the hydrological model applied by NVE can add uncertainty when computing rain+snowmelt. Furthermore, rainfall interpolations include more ground truth data than rain+snowmelt. MET reports that weather stations for rainfall recording have a denser network than those for measuring ice thickness (Homleid et al., 2021a). For all the above reasons, selecting a rainfall-induced landslide is intended to prevent additional uncertainties related to rain+snowmelt data.

For the antecedent rain, as mentioned in Subsection 3.2.1: Hydrology, it was decided to include the two small events depicted in Figure 6 that occurred at the beginning of August. The model simulates the hydrology and slope stability from August 4 until August 16. From here on, the word 'timestep' will refer to the modeled day. Thus, timestep 1 is August 4, and timestep 13 is August 16. Table 3 relates the date with the timestep and the daily rainfall.

Table 3. Daily rainfall simulated in the model. The initial timestep 1 corresponds with August 4, 2011, and the final timestep 13 with August 16, 2011. Source: xgeo.no.

Date	Timestep	Rainfall [mm/day]
4/08/2011	1	0
5/08/2011	2	6.1
6/08/2011	3	0.8
7/08/2011	4	0.6
8/08/2011	5	16
9/08/2011	6	9.6
10/08/2011	7	12.9
11/08/2011	8	3.9
12/08/2011	9	2.3
13/08/2011	10	0.1
14/08/2011	11	0
15/08/2011	12	0
16/08/2011	**13**	**68.8**

The soil depth model was calibrated by multiplying the raster layer by a constant value. Therefore, it is modified proportionally to its original shape. This constant is varied between 50% and 200% of the initial value because more failures are expected to occur in heavier and, thus, deeper soils (Johari & Javadi, 2012).

Subsection 3.1.4: Soil Stability Parameters explained that the study area was divided into the steep slope and the Quaternary deposit regions to assign cohesion magnitudes. The steep slope was given a constant value throughout the calibration, a number consistent with the bedrock behavior, set to prevent the triggering of failures (Wyllie & Norrish, 2017). The cohesion in the Quaternary deposit was iterated through a range based on the reported values in Table 2 and trial and error checks to narrow an interval potentially producing failures.

The IFA layer in this study is a constant value raster. The calibration range was decided according to values with similar texture classes reported by the Association of Swiss Road and Traffic Engineers (1999) via geotechdata.info, and the field information in Table 2. The texture was derived from the granulometry in Statens vegvesen (2016).

Table 4 lists the values iterated during the calibration, each parameter having 12 values. Spacing the range by 12 values was considered a good sampling between the minimum and maximum numbers intended to be tested. The model was run in a nested for-loop with all possible combinations: 12*12*12 = 1728 iterations. The abbreviation 'par sd' refers to the parameter (par) multiplied by the soil depth (sd) layer. The 'max sd' is the maximum thickness of the resulting soil model and is used to characterize the soil depth variable.

Table 4. Range of values iterated during calibration. The column 'n' enumerates the parameters. The soil depth layer is multiplied by the value in the 'par sd' column, and the resulting layer's maximum soil depth (max sd) is used to characterize the soil depth variable. The LISEM tool SlopeStabilityIS requires IFA in radians (rad), but it is also presented in degrees (deg).

n	par sd	max sd [m]	cohesion [Pa]	IFA [rad]	IFA [deg]
1	0.65	5.88	3450	0.506	29.0
2	0.80	7.24	4005	0.533	30.6
3	0.95	8.59	4560	0.560	32.1
4	1.10	9.95	5115	0.587	33.6
5	1.25	11.31	5670	0.614	35.2
6	1.40	12.67	6225	0.641	36.7
7	1.55	14.02	6780	0.668	38.3
8	1.70	15.38	7335	0.695	39.8
9	1.85	16.74	7890	0.722	41.4
10	2.00	18.09	8445	0.749	42.9
11	2.15	19.45	9000	0.776	44.5
12	2.30	20.81	9555	0.803	46.0

Evaluation metrics are intended to find an optimal set of parameters among all the combinations in the calibration that can successfully reproduce the landslide event on August 16, 2011. The metrics were employed to compare the landslide initiation area with the simulated failures. The initiation area depicted in Figure 11 was inferred from the landslide inventory (Depina et al., 2021; Krøgli et al., 2018) and multitemporal aerial photographs from norgeibilder.no (Geovekst, 2020). It is larger than the failure indicated in the inventory polygon to cover the adjacent erosion identified in the orthophotos.

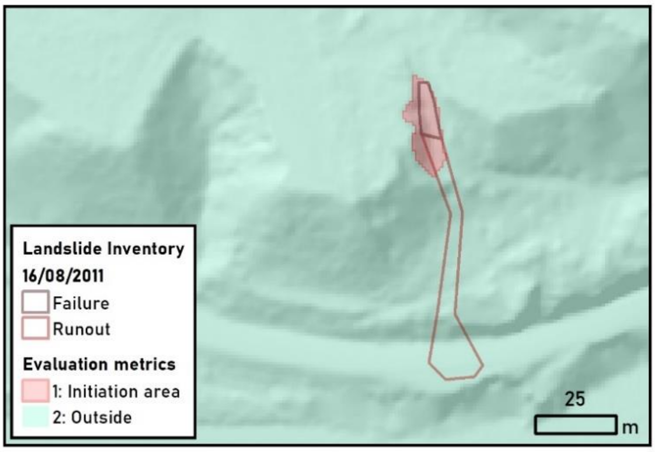


Figure 11. Initiation area employed in evaluation metrics calculations.

The evaluation metrics computation started by constructing the confusion matrix in Table 5. The matrix relates the location of the initiation area presented in Figure 11, referred to as observed, and the simulated failures in the calibration labeled as predicted. The SlopeStabilityIS tool returns FS and FD. As mentioned in Subsection 3.2.2: Slope Stability, when FS is less than 1 in a pixel, it means instability, and the tool automatically calculates the FD. If FS is above 1, indicating stability, FD is set to 0. Thereby, for each parameter combination, there is a resulting map with values per pixel for FS and another for FD. In this case, the FD was chosen to compare the observed and predicted values in the confusion matrix. Hence, it classifies according to the predicted FD of each pixel and the two regions portrayed in Figure 11: (1) the initiation area and (2) the outside, where no failures are expected during the simulation period.

Finally, the number of pixels classified with the confusion matrix as True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are counted to calculate the evaluation metrics.

Table 5. Confusion matrix based on the areas depicted in Figure 11. The number 1 refers to the initiation area of the landslide on August 16, 2011, and the number 2 is the outside, where it is assumed that no landslide occurred during the simulated period.

		Predicted	
		Positive	Negative
Observed	Positive	True Positive (TP): FD > 0 in 1	True Negative (TN): FD = 0 in 2
	Negative	False Positive (FP): FD > 0 in 2	False Negative (FN): FD = 0 in 1

Cohen's Kappa, precision, recall, and F1 were the selected evaluation metrics to assess the accuracy of the simulated failures. These metrics allow comparing the accuracy based on different criteria and are commonly employed for quantitative evaluations (Dalianis, 2018). The metrics equations employed in this analysis are presented in Table 6 in terms of TP, TN, FP, and FN.

Table 6. Evaluation metrics equations in terms of the TP, TN, FP, and FN defined in Table 5. Sources: Precision, Recall, and F1: Dalianis (2018), Cohen's Kappa: Chicco et al. (2021).

Precision (P)	**Recall (R)**	**F1**
$P = \frac{TP}{TP + FP}$	$R = \frac{TP}{TP + FN}$	$F1 = 2 * \frac{P * R}{P + R}$
Cohen's Kappa (K)		
$K = \frac{2 * (TP * TN - FP * FN)}{(TP + FP) * (FP + TN) + (TP + FN) * (FN + TN)}$		

Previously, it was mentioned that there are resulting maps for each parameter combination. In the code employed for the modeling, FS and FD are temporal variables that change with each iteration. Choosing not to save all the resulting maps to disk helps to reduce computational time. Cohen's Kappa was calculated during the calibration with the MapCohensKappa tool available in LISEM. However, for precision, recall, and F1, it was decided to select a sample of combinations based on the FS, FD, and Cohen's Kappa calibration results. In this way, the computational time is also optimized by avoiding the calculation of these three metrics for combinations that, for example, do not produce failures.

Graphs with 3-dimensions were selected for plotting the resulting 1728 combinations. Each point can be conceived as a coordinate in the 3-dimensional space formed by the values of max sd, cohesion, and IFA as listed in Table 4. Since the points are equally distributed on all three axes, the 1728 combinations span the space in a cube shape. The colors of the points are used as the 4^{th}-dimension to analyze the relationship of the parameter combinations against FS, FD, and Cohen's Kappa.

An iteration number is associated with each failure and runout map resulting from the evaluation metrics analysis. This number represents the code's execution order and ranges from 1 to 1728. The runout was simulated with 7200 timesteps, representing the displacement of the modeled debris flows after 2 hours.

The main outcomes of this study are presented in Chapter 4: Results below.

4. RESULTS

4.1. Preprocessing Results

The following subsections present three maps: soil depth, cohesion, and Manning's n, generated during the preprocessing of data for the physically-based model introduced in Section 3.2: Model Setup to simulate failure and runout in Kjellberget, Stjørdal, Norway.

4.1.1. Soil Depth Model

Figure 12 presents the soil depth model resulting from the calibration of the SteadyStateSoil LISEM tool based on von Ruette et al. (2013) with field measurements from Statens vegvesen (2016). The model proposed by the authors is based on the dynamics of erosion and deposition governed by elevation and slope. Figure 12 shows that the absence of soil is exhibited at protruding elevations, more commonly in the upper part of the figure. At the same time, the areas with the deepest soil coincide with terrain depressions where the water flow concentrates.

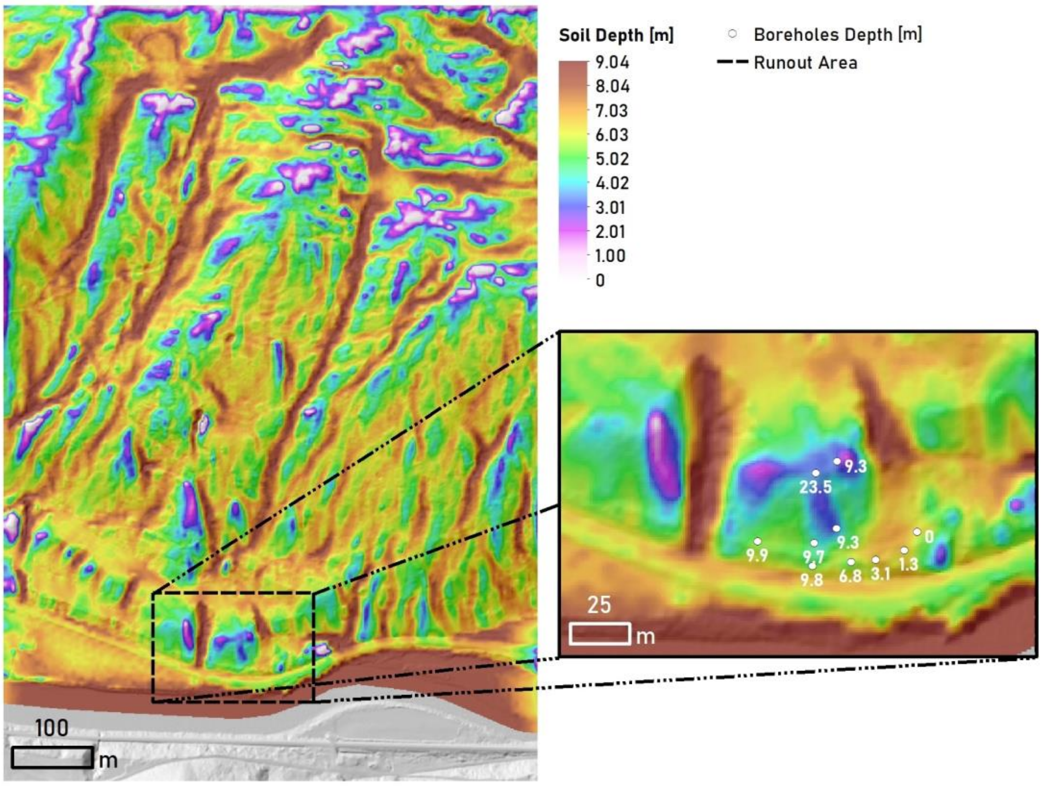


Figure 12. Soil depth model created after calibrating the SteadyStateSoil LISEM tool adapted from (von Ruette et al., 2013) with the drilling depths listed in Figure 4 from the Statens vegvesen (2016) geotechnical report.

The zoomed map in Figure 12 depicts the borehole depths from Statens vegvesen (2016). As mentioned in Section 2.2: Study Area Description, Kjellberget can be divided into three regions from top to bottom: steep slope, Quaternary deposit, and Stjørdalselva river valley. The drilling was performed in the Quaternary deposit. At its deepest, the deposit is reported to be 23.5 m, but in the soil depth model, the same location is 3.4 m deep, resulting in a difference of 20.1 m. The rest of the measurements also show differences between 1.4 and 6.2 m with the soil depth model. The Quaternary deposit originated from the Stjørdalselva river fluvial dynamics that cannot be captured at this level of detail. As a matter of fact, to run the SteadyStateSoil tool, it was necessary to resample the DTM from 1m² to 5m² because the model did not show good results at a lower resolution. Later, the resulting soil depth raster was resampled again to 1 m^2 to match the rest of the input data.

Despite the above, the SteadyStateSoil tool allowed the creation of a soil depth layer in a suitable format to run the physically-based model in LISEM, which is the best approximation that could be obtained for this analysis.

4.1.2. Cohesion Map

As stated in Subsection 3.1.4: Soil Stability Parameters, the study area at Kjellberget was divided into two regions according to its geomorphology and composition for modeling purposes. These regions were determined with descriptions included in the KlimaDigital report (Depina et al., 2021), the slope derived from the DTM, aerial photographs from norgeibilder.no (Geovekst, 2020), and Google Street View images (Google, 2010). The steep slope zone was assigned a high cohesion value according to Wyllie & Norrish (2017). Since no erosion is observed in this zone in the aerial photographs, no considerable failures are expected to appear during the simulations. It was achieved by setting a constant high cohesion value to the steep slope throughout the calibration process.

In the Quaternary deposit, on the other hand, erosion is visible in all available orthophotos, which date from 2010, 2014, 2015, and 2019. Accordingly, the cohesion in the Quaternary deposit varied during the calibration, as it is where landslides have been reported. Besides, some places near the road were assigned the same high cohesion as the steep slope because rock outcropping was observed in Street View. The rock is a more cohesive material that does not fail like the soil. Again, assigning a higher value is a way to prevent failures from appearing in unrealistic places. The resulting cohesion map is shown in Figure 13.

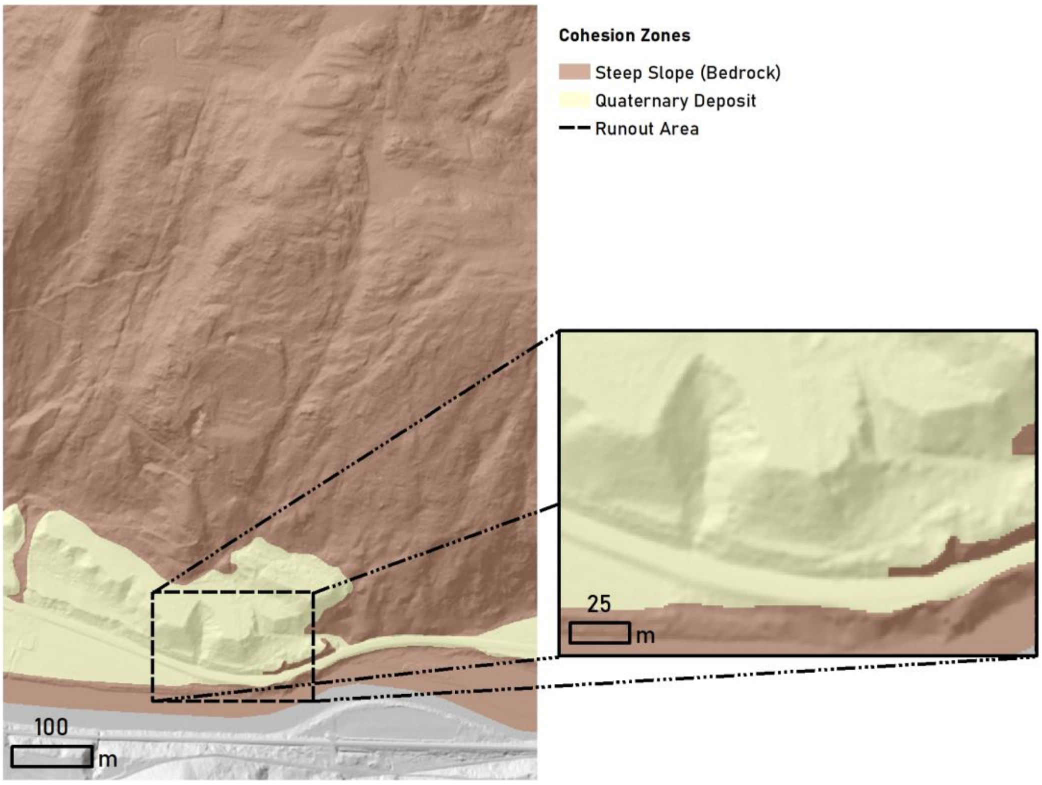


Figure 13. Cohesion Map. Two cohesion values were assigned according to the regions presented on this map. The Quaternary deposit cohesion varied during calibration while the Steep Slope (bedrock) remained constant.

4.1.3. Manning's n Map

Manning's n is a coefficient used to relate a fluid's velocity to the surface's roughness (Arcement & Schneider, 1989). It was introduced in Subsection 3.1.5: Runout Parameters. The landcover and subsequent Manning's n values assigned to the runout area are displayed in Figure 14.

Figure 14. Manning's n map showing the landcover categories from Chow (1959) via orst.edu employed to assign the Manning's n values. The background photograph was obtained from norgeibilder.no, Stjørdal 2015.

4.2. Calibration Results

This unit reports the results of the procedures described in Section 3.3: Calibration. The outcomes of the 1728 simulations in relation to FS, FD, and Cohen's Kappa are presented below.

It was found that the obtained values for FS inside the initiation area, depicted in Figure 11, can be related to the timesteps listed in Table 3. The hydrology and slope stability of the August 16, 2011 landslide were simulated with daily timesteps from August 4 until August 16. Therefore, the simulation started in timestep 1 and ended in timestep 13, as indicated in Table 3. Since no failures were reported during the days before the landslide, the FS was assumed above 1 for timesteps 1 to 12. On the contrary, the FS was expected to be less than 1 on timestep 13, the day of the triggering. However, among the combinations, some cases showed failures once or several times between timesteps 1 to 13, and others stayed stable during all timesteps. Figure 15 presents the 3-dimensional space formed by max sd, cohesion, and IFA, with colors related to FS and timesteps. The combinations were classified into three cases:

- **Premature failure (792 yellow dots):** FS is less than 1 for at least one timestep between 1 and 12.
- **Correct-day failure (53 red dots):** FS is above 1 for all timesteps between 1 and 12, and is less than 1 when timestep equals 13.
- **No failure (883 blue dots):** FS is above 1 for all timesteps between 1 and 13.

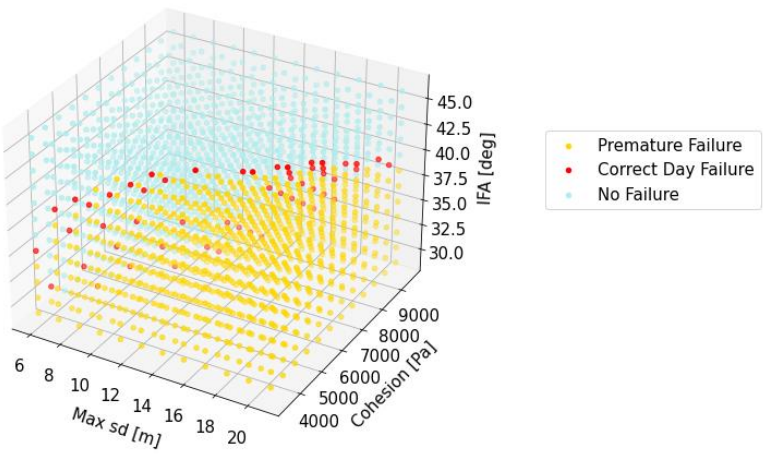


Figure 15. Results classified according to the obtained values of FS during the calibration and the timesteps listed in Table 3. The yellow dots are combinations that showed failures prematurely, between timesteps 1 and 12. The red dots exhibited failures exclusively on timestep 13, the correct triggering day. The blue dots are combinations where no failures occurred in the entire simulation.

The correct day failure red dots in Figure 15 can be interpreted as an interface between the underlying yellow dots indicating premature failure, and the overlying blue dots representing simulations where no failure was triggered at any timestep.

The number of simulations that showed failures at timestep 13 was 845: 792 from premature failure and 53 from correct-day failure. The failure frequency of the iterated values was calculated to examine the relationship between max sd, cohesion, and IFA with slope stability. It is presented in the three bar charts in Figure 16. From left to right, it is observed that heavier soils with less cohesion and lower IFA exhibited failures more frequently.

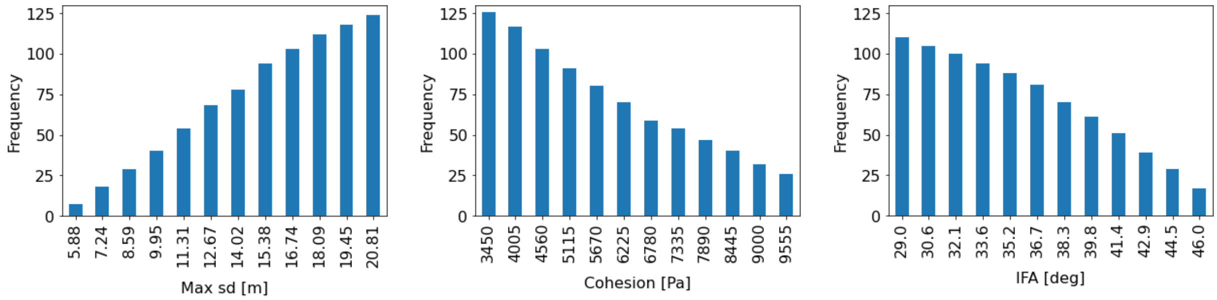


Figure 16. Failure frequency of the iterated values per parameter during the calibration. From left to right, it is observed that heavier soils with less cohesion and lower IFA exhibited failures more frequently.

Similarly, the failure frequency among the correct-day class combinations was computed and presented in Figure 17. In this case, none of the bar charts show a clear trend like in Figure 16. In general, the three graphs in Figure 17 exhibit two peaks with a minimum in the middle of the iterated range. The peaks do not match or show similar behavior. Hence, a combination with higher failure frequency cannot be identified in the correct-day failure class.

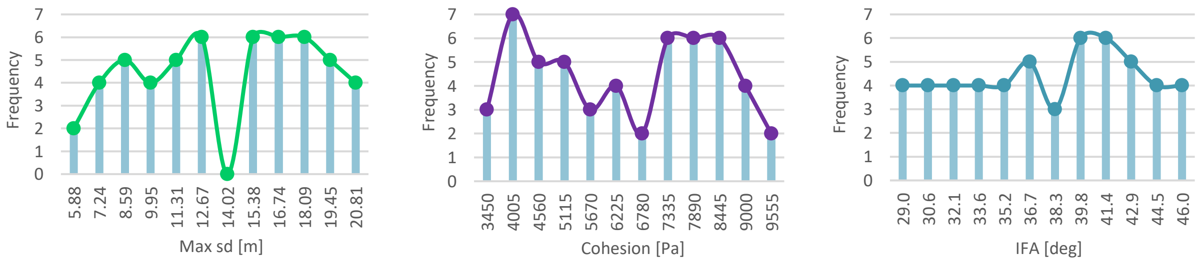


Figure 17. Failure frequency of the iterated values per parameter in the 'correct day failure' class from Figure 15. The three graphs exhibit two peaks with a minimum in the middle of the iterated range. A single optimal combination of parameters cannot be identified because the failure frequencies in the three graphs do not match.

The calibration also gave outcomes for FD and Kappa. Figure 18 presents two graphs with information related to the FD magnitudes obtained in the calibration within the initiation area depicted in Figure 11.

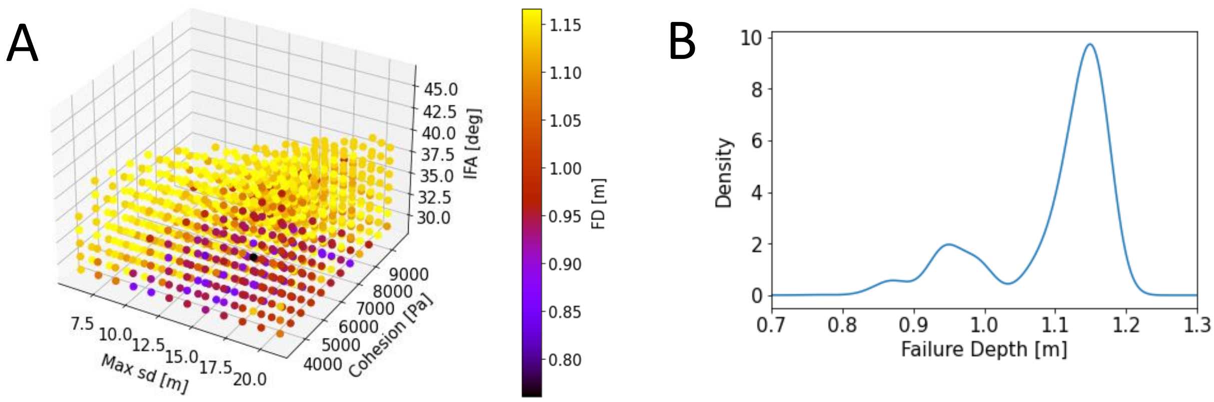


Figure 18. Failure Depth (FD) obtained in the calibration within the initiation area depicted in Figure 11. Graph A is a 3-dimensional scatter plot where the dots colors represent the FD magnitude. Graph B is a density plot showing the FD distribution.

In Figure 18A, the color gradient in the 3-dimensional scatter plot is used to represent the FD magnitude on a continuous scale. It can be observed that heavier soils with less cohesion and lower IFA exhibited lower FD magnitudes. On the other hand, a density plot built with the FD magnitudes is presented in Figure 18B. The minimum and maximum depths are 0.76 and 1.17 m, but most failures, 620 out of 845 (73%), fall between 1.10 and 1.17 m. The remaining 225 (27%) are between 0.76 and 1.09 m.

The Cohen's Kappa coefficients obtained in the calibration are presented in Figure 19 for the 845 simulations that showed failures at timestep 13. In Figure 19A, the color gradient in the 3-dimensional scatter plot is used to represent the Kappa values on a continuous scale. The highest Kappa scores, which indicate greater agreement between the observed and predicted failures, are evidenced in parameter combinations with low max sd and low cohesion. The relationship is unclear for IFA, but the higher scores are around 32° and 35°. In Figure 19B, the density plot shows the distribution of the resulting Kappa values, which can be divided into three parts: (1) [0.00 to 0.04]: 542 combinations (64%), (2) (0.04 to 0.08]: 261 combinations (31%), and higher than 0.08: 42 combinations (5%). The maximum Kappa obtained was 0.105.

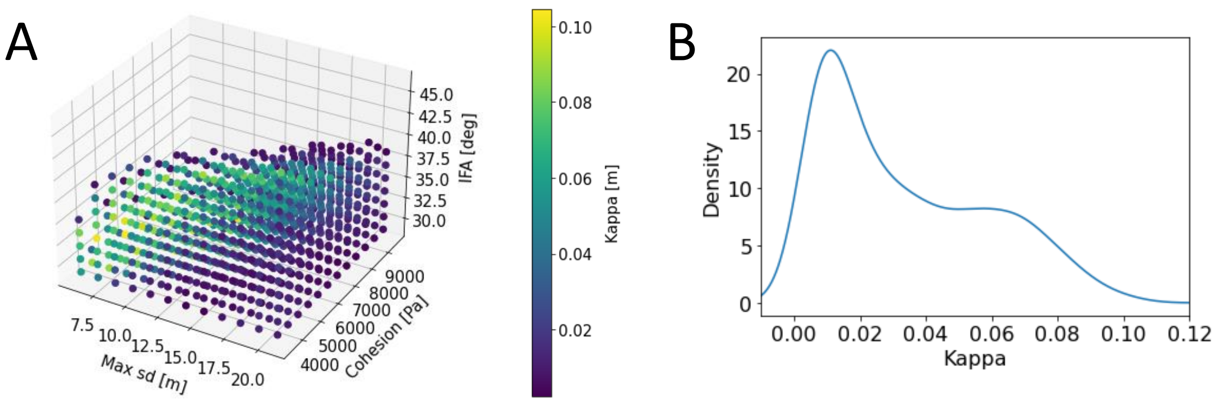


Figure 19. Cohen's Kappa coefficients obtained for the 845 parameter combinations that generated failures during the calibration. Graph A is a 3-dimensional scatter plot where the dot colors represent the Kappa values. Graph B is a density plot showing the Kappa distribution.

As explained in Section 3.3: Calibration, a sample of combinations is selected to continue calculating the evaluation metrics precision, recall, and F1. It was decided to take the combinations that exhibited the maximum Kappa scores since the FS and FD outcomes are inconclusive. The cutoff was set at Kappa greater than 0.08, resulting in a sample of 42 combinations. The resulting maps with the highest precision, recall, and F1 are introduced in the next section.

4.3. Model Results Using Observed Precipitation

This section presents the maps generated with the observed precipitation in Table 3 and the parameter combinations that produced the highest Cohen's Kappa, precision, recall, and F1 in the calibration.

Figure 20 presents the failure frequency map created from the sum of pixels with FD above zero among the 42 combinations selected with the maximum Kappa. Figure 20A displays the entire simulated area for hydrology and slope stability, as described in Section 3.2: Model Setup. Figure 20B zooms to the Quaternary deposit, where the landslides were reported. In Figure 20B is observed that the most frequent failure locations in the sample, up to 35 pixels in red color, coincide with the source areas in the landslide inventory. Also, the failures match the erosion zones identified in aerial photographs from norgeibilder.no. As an illustration, an orthophoto from 2015 is displayed in Figure 20C. A landslide scar in the left gully and another active erosion zone on top of the right earthen road can be detected.

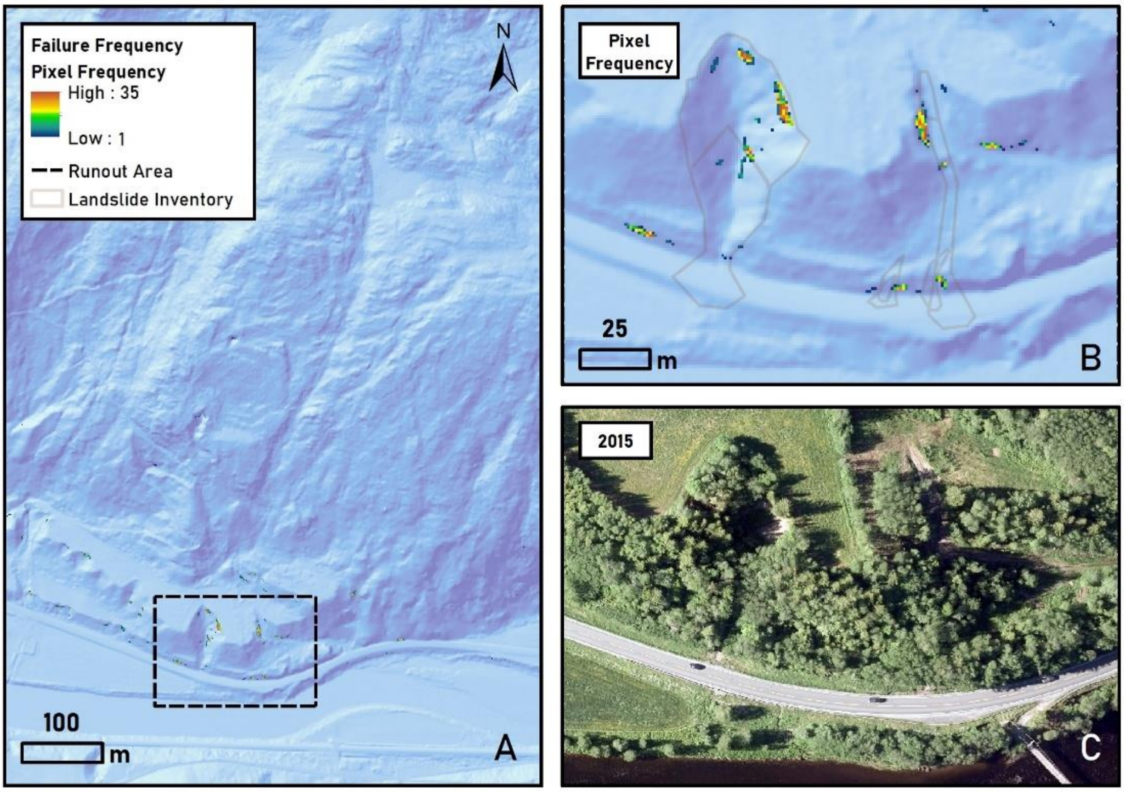


Figure 20. Failure frequency map created from the sum of pixels that exhibited failure among the 42 combinations with the highest Kappa. A covers the entire simulated area. B and C zoom to the reported landslides' location.

The failure and runout maps obtained with the combinations that yielded the highest precision, recall, and F1 are presented in Figure 21, Figure 22, and Figure 23. In this case, the combination with the maximum F1 score is the same for Kappa.

Similarly to Figure 20, all three figures are composed of maps A, B, and C. A shows the failures in the entire hydrology and slope stability simulation area, and B and C zoom to the runout region, as explained in Section 3.2: Model Setup. The legend in A characterizes the FD, while in C reports the maximum height (hmax) of the solids and fluids mixture in the debris flow. The figures are accompanied by tables with information about the calibrated parameters employed in each case.

Figure 21 depicts the simulated failure and runout with the highest precision in the calibration using the parameters summarized in Table 7. In this case, the deepest failure and the maximum runout height have the same magnitude: 1.38 m. Figure 21B shows that the failures for the highest precision are relatively few and small. These are exhibited inside the initiation area of the March 22, 2012 (left) and August 16, 2011 (right) landslides, although it was only calibrated based on the latter. Also, no failures are detected outside the landslide boundaries. It is observed in Figure 21C that the runout for both landslides follows the same direction as the polygons in the inventory. However, the released material is not enough to reach the road as expected for the aim of reproducing the August 16, 2011 (right) landslide.

Table 7. Combination of parameters that reproduced the August 16, 2011 landslide failure with the best precision.

iteration	precision	par sd	max sd [m]	cohesion [Pa]	IFA [rad]	IFA [deg]
4	0.480	0.65	5.88	3450	0.587	33.6

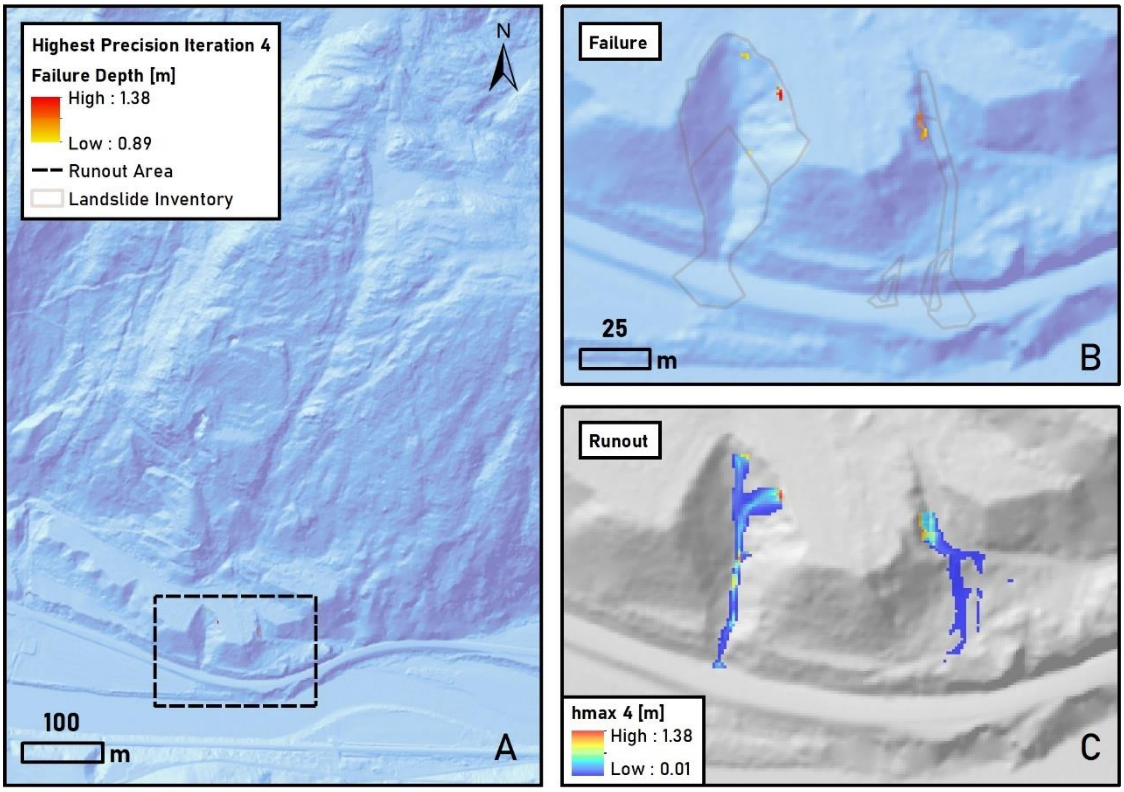


Figure 21. Failure and runout maps obtained with the combination that yielded the highest precision in the calibration. A covers the entire hydrology and slope stability simulated area. B and C zoom to runout region. C displays the runout derived from the failures presented in B.

Figure 22 depicts the simulated failure and runout with the highest recall in the calibration using the parameters summarized in Table 8. The deepest failure exhibited was 1.66 m, and the maximum runout height was 2.22 m. In this case, the debris flow material was accumulated upon reaching the road in the outlet of the March 22, 2012 (left) landslide, and thus, its height is above the maximum FD. Figure 22B shows that the failures with the highest recall, unlike the ones with the highest precision in Figure 21B, are scattered in the initiation areas, especially in the March 22, 2012 (left) landslide. Additionally, several failures are observed outside the polygons' boundaries, some quite close to the road.

Table 8. Combination of parameters that reproduced the August 16, 2011 landslide failure with the best recall.

iteration	recall	par sd	max sd [m]	cohesion [Pa]	IFA [rad]	IFA [deg]
625	0.106	1.25	11.31	5670	0.506	29.0

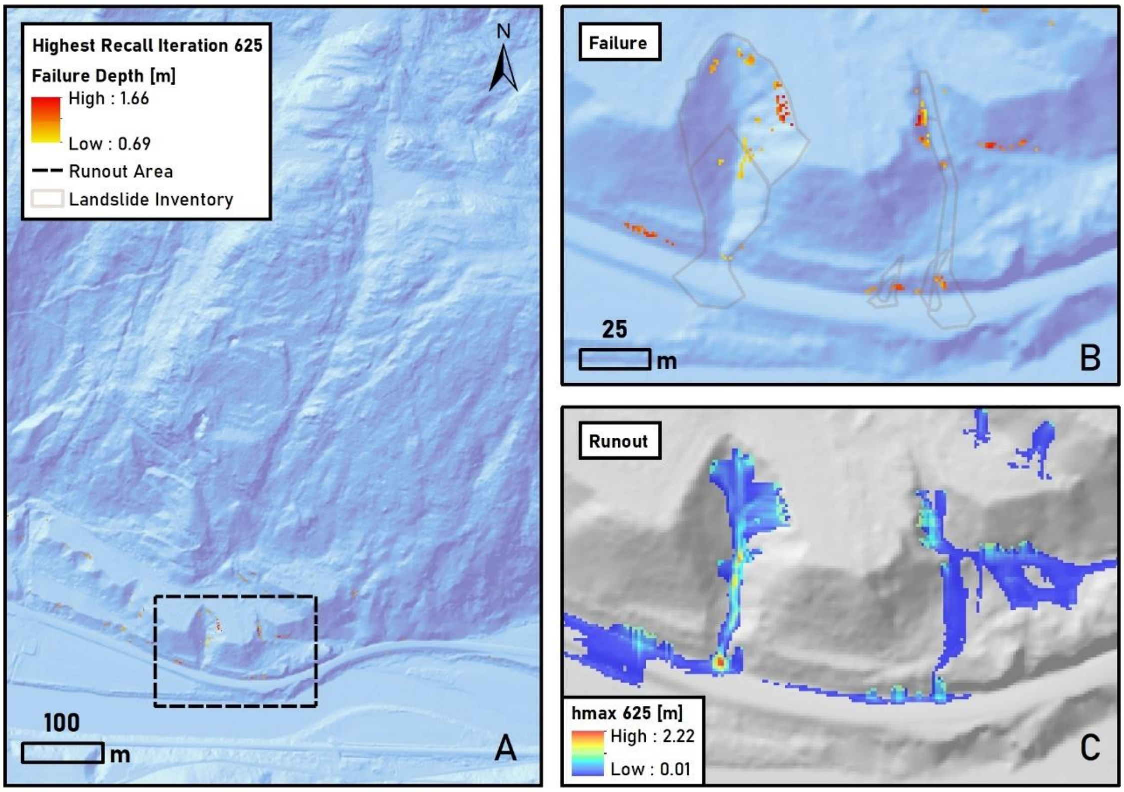


Figure 22. Failure and runout maps obtained with the combination that yielded the highest recall in the calibration. A covers the entire hydrology and slope stability simulated area. B and C zoom to runout region. C displays the runout derived from the failures presented in B.

Figure 22C depicts the runout generated from the failures in Figure 22B. It overflows more than the runout produced with the highest precision in Figure 21C. The failures on the left of the zoomed figures generated a runout that covered the road with a thin layer of dislodged material. However, this is not reflected in the inventory, meaning that (1) it may never have happened or (2) it was too minor to report. On the other hand, the highest recall reproduced the debris flow direction of the August 16, 2011 (right) landslide better than the highest precision. It spilled over the north roadside but did not cover the road as expected. Like the highest precision simulation, the volume released in the case of the highest recall remained insufficient for replicating the landslide impact area.

Figure 23 depicts the simulated failure and runout with the highest F1 and Kappa in the calibration using the parameters summarized in Table 9. In this case, the deepest failure was 1.38 m, and the maximum runout height was 1.87 m. The highest F1 for failure and runout generally exhibited an intermediate behavior compared to that obtained using the highest precision and recall cases. The failures in Figure 23B are more abundant than the highest precision in Figure 21B but not as scattered as the highest recall in Figure 22B. In Figure 23C, some roadside failures produced a runout that spread over the pavement, although not as wide as the highest recall in Figure 22C.

Table 9. Combination of parameters that reproduced the August 16, 2011 landslide failure with the best F1 and Kappa.

iteration	F1	Kappa	par sd	max sd [m]	cohesion [Pa]	IFA [rad]	IFA [deg]
304	0.1047	0.1045	0.95	8.59	4005	0.587	33.6

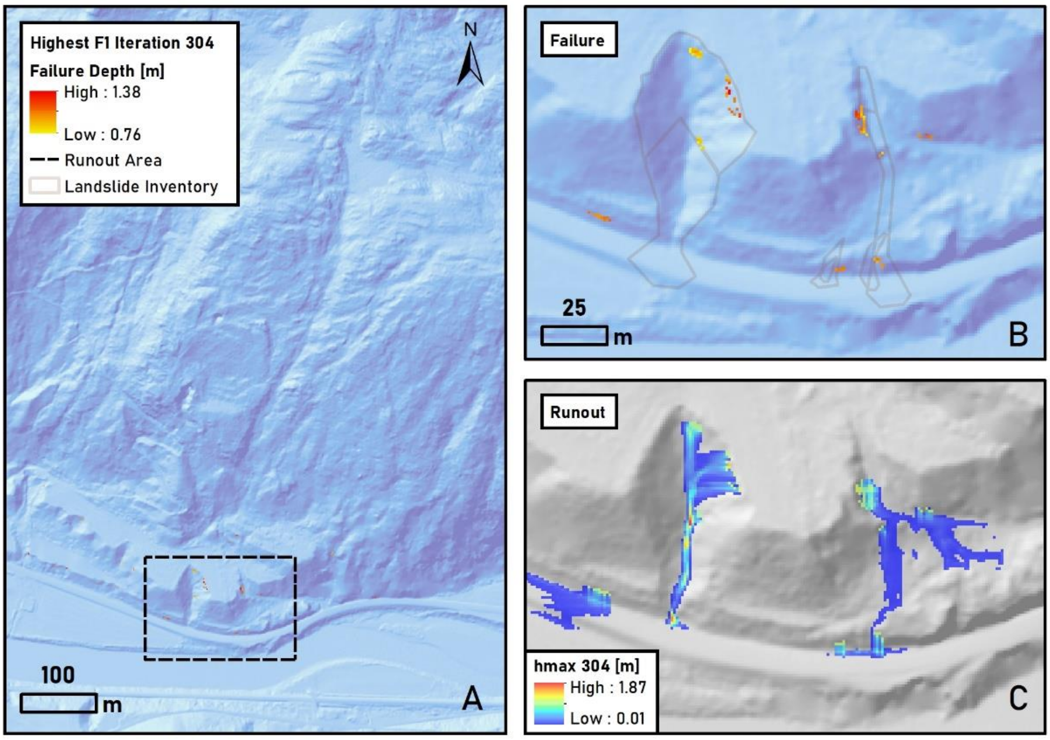


Figure 23. Failure and runout maps obtained with the combination that yielded the highest F1 and Kappa in the calibration. A covers the entire hydrology and slope stability simulated area. B and C zoom to runout region. C displays the runout derived from the failures presented in B.

Like the previous cases, the highest F1 reproduced the runout of the March 22, 2012 (left) and August 16, 2011 (right) landslides, even though it was calibrated based only on the latter. The failure locations and runout directions were relatively well reproduced, but the released material was not enough to cover the road as the August 16, 2011 (right) polygon indicates.

5. DISCUSSIONS AND CONCLUSIONS

5.1. Discussions

In this thesis, an analysis was performed to explore the potential of integrating landslide failure and runout physically-based modeling in territorial LEWSs. A physically-based model built with LISEM was calibrated using a rainfall-induced landslide event on August 16, 2011.

After running 1728 simulations in the calibration, it was found that the obtained FS could be related to the daily timesteps. The results were classified according to the stability before and during the August 16, 2011 landslide in three cases: premature failure (792), correct day failure (53), and no failure (883), as depicted in Figure 15.

The parameters calibrated were soil depth, cohesion, and IFA, which were selected because these directly influence the magnitude of FS and, therefore, the occurrence of failures, as demonstrated by Johari & Javadi (2012). The failure frequency of the iterated values was calculated for all the combinations of parameters that produced failure in the calibration, which includes the premature and correct day cases (845). Figure 16 shows that the failure frequency changed almost linearly as these three parameters varied. The relationship was proportional in the case of soil depth, meaning that the deeper and heavier soils exhibited failures more frequently. The relationship was inverse for cohesion and IFA, as failures were more frequent with lower magnitudes of these two parameters. This behavior was expected as soil depth, cohesion, and IFA are included in the equation of the infinite slope approach employed in the model, as well as it agrees with Johari & Javadi's (2012) findings.

It was also expected that an optimal combination of soil depth, cohesion, and IFA could be found among the 53 correct-day failure cases for replicating the selected landslide. In this case, the failure frequencies presented in Figure 17 did not show a clear trend but instead fluctuated in the range of the iterated values. As no match could be identified in the three graphs, it was concluded that the calibrated parameters in the correct-day failure class do not converge to a single optimal value based on FS.

Subsequently, the FD results were analyzed in all the 845 combinations that produced failure since identifying an optimal combination was not possible with FS in the correct-day failure class. The color gradient in the 3-dimensional scatter plot in Figure 18A indicates that FD increases as the combinations approach the correct-day failure interface. However, the depth magnitude and its variation among the simulations were relatively small. From the density plot in Figure 18B, it is translated that 73% of the simulations showed between 1.10 and 1.17 m depth, with a minimum variation of 0.07 m. The remaining 27% corresponded to even lower depths, between 0.76 and 1.09 m, but with a more significant variation of 0.33 m. Overall, the obtained FD magnitudes were shallow, meaning that the simulated failures would release little material for potential runout. Additionally, a single combination where the calibration converges to a greater depth value could not be identified because the deepest failures were scattered in the space, as shown in Figure 18A.

The only available information for comparing the predicted and observed failures were the landslide inventory polygons from skredregistrering.no. The evaluation metrics Cohen's Kappa, precision, recall, and F1 were performed because these allow assessing the accuracy based on the area. As the calibration focuses on the failure, the area considered as observed was the initiation of the August 16, 2011 landslide depicted in Figure 11. Adjacent zones to the inventory polygon were included in the observed area to incorporate erosion identified in multitemporal aerial photographs from norgeibilder.no. From direct communication with Graziella Devoli, part of the NVE personnel, it was known that some landslides are drawn with uncertainty. For example, the March 22, 2012 polygon presented in Figure 2D encloses the entire gully where erosion is active and not just the zone affected by that particular landslide. Hence, in this study, the inventory was not treated rigorously as ground truth but more like a reference. Thus, the evaluation metrics were employed to identify the combinations that produced a greater agreement between the predicted and observed failures, more than absolute accuracy measurements.

Cohen's Kappa was run inside the calibration code, and the resulting scores for all the 845 combinations that generated failure are shown in Figure 19. A critical finding when comparing the results for FS (Figure 15), FD (Figure 18), and Cohen's Kappa (Figure 19) was that Kappa and FD did not show their maximum values in the same combinations as the correct-day interface indicated by FS. Therefore, it was concluded that with the performed calibration, it was unachievable to find a single combination that (1) produces failures only on the timestep of the triggering day, (2) that has the maximum FD, and (3) that the initiation area coincides with the inventory.

In the methodology, it was proposed to calculate the remaining evaluation metrics according to the previous interpretation of the calibration results to save computational time. Hence, it was decided to take a sample of combinations with the highest Kappa. This selection ensures the most attainable agreement between the observed and predicted failures among the calibration results, which is essential to replicate the landslide accurately. Thus, precision, recall, and F1 were computed for 42 combinations with Kappa values above 0.08. It was considered a good number to compare the failures because the variation in the agreement was already evident in this range. Additionally, a failure frequency map was created with this sample to examine the initiation areas predicted by the physically-based model.

The maximum Kappa obtained in the calibration was 0.105. Thereby, the overall Kappa results showed poor to slight agreement, according to Sim & Wright (2005). When examining the failure frequency map in Figure 20B, it was noticed that most failures related to the August 16, 2011 landslide occurred towards the left of the area indicated as observed. Since the predicted failures only partially matched the observation, the evaluation metrics scores were consistently low. Nevertheless, based on visual comparison with aerial photographs like in Figure 20C, it can be concluded that the simulations with higher Kappa could reproduce the initiation areas of the reported landslides in Kjellberget reasonably well.

The failure and runout maps obtained with the combinations that yielded the highest precision, recall, and F1 were presented in Figure 21, Figure 22, and Figure 23. As the combination with the maximum F1 score was the same for Kappa, both were represented in the last figure. The tables accompanying the maps report the soil depth, cohesion, and IFA values used in each case. It is evidenced that the optimal calibrated parameters are not identical for the different evaluation metrics. Although the values are concentrated in the first five rows of Table 4, the part of the range where the lowest numbers are found, it was not possible to determine their convergence towards a single optimal combination of parameters. Moreover, there is also no coincidence with the field measurements in Table 2.

In the failure and runout maps, it is observed that the August 16, 2011 and March 22, 2012 landslides were replicated simultaneously in the three cases, even though the calibration was based only on the first. This fact shows the uncertainty in the model because it simulates two events when it should only be one. On the other hand, the failure locations and runout directions were replicated reasonably well. Nevertheless, the released volume was not enough to cover the road as the August 16, 2011 polygon indicates, which was suspected since the FD magnitudes were found to be shallow. Therefore, it was unattainable to reproduce the landslide impact area with the performed calibration.

For the Norwegian LEWS, the proposed methodology in this thesis would be relevant if it can provide accurate predictions at a lower spatial resolution than what is already available. In theory, the statistical model currently used by NVE is reliable at the scale of administrative units such as municipalities (Krøgli et al., 2018). In practice, empirical knowledge plays a fundamental role in the local response, according to Graziella Devoli. In the case of the August 16, 2011 event, the E14 road was closed in advance to prevent accidents because NVE is aware that landslides regularly occur on that road. However, the E14 runs approximately 67 km across Norway. The ability to detect where there is more probability of landslides along the road for forecasted rainfall events would be of an added value to improve the response mechanisms.

In this study, it was not possible to replicate the landslide impact area, but the failure locations and runout directions showed promising results. Validation with landslides from different dates and experiments in separated areas would be necessary to determine if the reproduced failure and runout are reliable enough to raise warnings. Also, a less detailed scale may be considered for interpreting the model results, distinguishing the slopes by geomorphological features or geological materials. For instance, the sole runout development exhibited in the simulations may indicate that landslides could be potentially triggered in the Kjellberget Quaternary deposit.

The first specific objective of this thesis was to find an optimal combination of parameters to replicate the August 16, 2011 landslide impact area with observed precipitation. However, achieving this objective with the obtained results was concluded to be not possible. It was expected that failures in accurate locations would release enough volume to reproduce the observed runout. Thus, the calibration only considered the slope stability process. After the analysis, the outcomes proved that this assumption was wrong. For a multi-process model like the one in this study, basing the calibration on a single process led to unsuccessful results.

The second specific objective was to assess the accuracy of the landslide prediction with forecast precipitation. Nevertheless, a simulation with sufficient accuracy to reproduce the August 16, 2011 landslide impact area was not achieved. In this case, the forecasted rainfall depicted in Figure 6 was considerably lower than the observed. As the water entering the model would be less with the forecast, the added weight to the soil would also be less, and the stability would not be affected as much as with the observed precipitation. Therefore, running the model with the forecast would not generate a more extensive runout that could improve the results. However, this information is not enough to conclude the overall performance of the forecast, which may even have improved in recent years, considering that in 2011 the weather model was an older version of the one currently employed. Hence, more experiments will be necessary to determine the accuracy of landslide predictions with physically-based models using forecast precipitation.

LISEM was chosen for the modeling because it allows the construction of a multi-process model capable of simulating slope stability to obtain failure depth, which can be used as initial volume for runout. It is an advantage over other physically-based models tested in Norway, where inputs for initiation area and released volume are required. However, the runout obtained with LISEM was not enough to reproduce the August 16, 2011 landslide impact area. On the contrary, previous research with physically-based models like Flow-R (Fischer et al., 2012) and TRIGRS (Schilirò et al., 2021) could replicate the displacement and spatial distribution of tested landslides successfully. In the context of territorial LEWSs, a detailed physically-based model would be of interest if it can provide accurate predictions automatically. Nevertheless, all the approaches, either consulted or tested, showed drawbacks in simultaneously achieving the model automatization and the generation of accurate predictions.

Finally, the general objective of this thesis was to explore the potential of integrating physically-based failure and runout modeling in territorial LEWSs. Because of the high uncertainty, dependence on data availability, and complex calibration settings, it was found that the studied physically-based modeling would not improve the applicability of these systems.

5.2. Limitations

Several aspects of the datasets employed, and the methodology applied in this study might be subject to limitations.

The physically-based approach implemented a relatively simple setup, meaning numerous assumptions were adopted, such as the simplified groundwater representation or the infinite slope hypothesis. The hydrology was not individually calibrated and lacked slower groundwater response from the unsaturated layer, which can be particularly relevant for replicating the water behavior in deeper soils. In the infinite slope method, horizontal forces are not considered, and the failures are limited to a local equilibrium. Therefore, small local failures might be predicted unrealistically, and more complex types of failures might not even be predicted. Additionally, data constraints exacerbate these limitations. For instance, several significant subsurface properties like spatial strength parameters and groundwater permeability are unknown. The simplified model setup allows running a model despite the limited circumstances, but important processes might be missed.

On the other hand, including multiple processes in the calibration, such as slope stability and runout, would have been beyond the scope of this study in terms of time availability and computational power. Moreover, the failure and runout simulations showed that the different accuracy metrics resulted in distinct calibrated parameters. While it was not possible to find a single combination that converged towards an optimal landslide replication, it was also found that there is no straightforward way to determine it. Therefore, analyzing the convergence of the parameters in a multi-process calibration could be even more challenging. Likewise, using additional observed information, such as precipitation with a higher temporal resolution, groundwater stages, failure time, released volume, runout evolution, and others, might benefit the model but also leads to a highly complex setup.

In this work, the simulations were performed at 1 m^2 spatial resolution. For larger areas, a lower resolution would be required to run the model in a reasonable time, leading to even further limitations related to the performance of the model. Indeed, an advantage of placing the study in Norway was the access to a high-quality DTM model and other data obtained from online databases, in addition to the information provided by NVE. Therefore, the transferability of this study to other regions with less documentation is a major limitation for further research.

5.3. Recommendations

For NVE:

- The information collected during the KlimaDigital project was essential for conducting this research. A continuous gathering of data, as well as accurately monitoring and systematically registering landslides, allows studies like the one presented in this thesis to continue being performed.
- When plotting the time series in subsection 3.1.6: Precipitation, it was noticed that the triggering rainfall was between 60 and 70 mm for the four analyzed landslide events. Localized rainfall thresholds derived from empirical relationships could be a more practical approach to explore in predicting local landslide triggering.

For future research:

- The main impediment for replicating the August 16, 2011 landslide event with the proposed methodology was that the failures did not released enough volume for the runout. Two alternatives can still be investigated: (1) to apply a multi-process calibration including a factor multiplying the initial runout height to increase the released volume, and (2) to explore other failure methods that might generate deeper failures like the ellipsoid sampling (Hovland, 1977) or the fiber bundle method (Cohen et al., 2009).
- Approaches such as the one proposed in Zhang et al. (2018) for calibrating coupled hydrology and slope stability processes could be considered for developing methods for accurate simultaneous calibration of slope stability and runout.

5.4. Conclusions

In this thesis, an analysis was performed to explore the potential of integrating landslide failure and runout physically-based modeling in territorial LEWSs. The Norwegian Landslide Forecasting and Warning Service currently operates at the national level in Norway, which allowed to structure the methodology of this study within the framework of an established system. A physically-based model was built and calibrated using the LISEM software. The study area was Kjellberget, a location in the Stjørdal municipality in central Norway. The calibration of the model was based on a rainfall-induced landslide on August 16, 2011.

When plotting the results in the 3-dimensional space formed by the calibrated parameters soil depth, cohesion, and IFA, it was found that FS could be related to the simulated daily timesteps. The combinations of parameters that exhibited instability only on the correct day of the triggering represented an interface between the underlying cases with premature failure and the overlaying combinations that generated stability during the entire simulation. However, identifying a single combination converging towards an accurate replication of the August 16, 2011 landslide was not possible based either on FS, FD, or Cohen's Kappa.

According to the obtained failure and runout maps with the highest precision, recall, and F1 in the calibration, it is concluded that the model was able to replicate the initiation areas and runout directions reasonably well. Nevertheless, the released volume derived from the FD was insufficient to replicate the landslide impact area. Only slope stability was considered in the calibration, and the evaluation metrics were calculated by comparing the observed and predicted initiation areas. However, not including the runout in the calibration of the multi-process model employed in this study led to unsuccessful results.

Based on the experiment performed, it is concluded that it is non-viable to apply the proposed methodology in territorial LEWSs because of the high uncertainty, dependence on data availability, and complex calibration settings of multi-process physically-based models. Nevertheless, approaches for increasing the released volume to produce more accurate runout simulations can still be explored in further research to improve their applicability.

LIST OF REFERENCES

Adla, S., Tripathi, S., & Disse, M. (2019). Can we calibrate a daily time-step hydrological model using monthly time-step discharge data? *Water (Switzerland)*, *11*(9). https://doi.org/10.3390/w11091750

Arcement, G. J., & Schneider, V. R. (1989). *Guide for Selecting Manning's Roughness Coefficients for Natural Channels and Flood Plains*. https://pubs.usgs.gov/wsp/2339/report.pdf

Association of Swiss Road and Traffic Engineers. (1999). *Swiss Standard SN 670 010b: Characteristic Coefficients of soils*.

Beldring, S., Engeland, K., Roald, L. A., Sælthun, N. R., & Voksø, A. (2003). Estimation of parameters in a distributed precipitation-runoff model for Norway. *Hydrology and Earth System Sciences*, *7*(3), 304–316. https://doi.org/10.5194/hess-7-304-2003

Benjamin, S. G., Brown, J. M., Brunet, G., Lynch, P., Saito, K., & Schlatter, T. W. (2019). 100 Years of Progress in Forecasting and NWP Applications. *Meteorological Monographs*, *59*, 13.1-13.67. https://doi.org/10.1175/AMSMONOGRAPHS-D-18-0020.1

Boje, S., Devoli, G., & Colleuille, H. (2014). *Terskelstudier for utløsning av jordskred i Norge*.

Bout, B., Lombardo, L., van Westen, C. J., & Jetten, V. G. (2018). Integration of two-phase solid fluid equations in a catchment model for flashfloods, debris flows and shallow slope failures. *Environmental Modelling & Software*, *105*, 1–16. https://doi.org/10.1016/j.envsoft.2018.03.017

Butcher, K. J. (2004). 9.2.4 Infiltration Rate. In *CIBSE Guide G - Public Health Engineering (2nd Edition)* (pp. 1–2). The Chartered Institution of Building Services Engineers (CIBSE). https://app.knovel.com/hotlink/khtml/id:kt00C9CDL1/cibse-guide-g-public/infiltration-rate

Calvello, M. (2017). Early warning strategies to cope with landslide risk. *Rivista Italiana Di Geotecnica*, *51*(2), 63–91. https://doi.org/10.19199/2017.2.0557-1405.063

Canli, E., Mergili, M., Thiebes, B., & Glade, T. (2018). Probabilistic landslide ensemble prediction systems: Lessons to be learned from hydrology. *Natural Hazards and Earth System Sciences*, *18*(8), 2183–2202. https://doi.org/10.5194/nhess-18-2183-2018

Cardellini, S., & Osimani, P. (2013). The Ancona Early Warning Centre, Instrumentation and Continuous Monitoring of the Landslides. In *Landslide Science and Practice* (Vol. 2, pp. 57–65). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31445-2_7

Chae, B. G., Park, H. J., Catani, F., Simoni, A., & Berti, M. (2017). Landslide prediction, monitoring and early warning: a concise review of state-of-the-art. *Geosciences Journal*, *21*(6), 1033–1070. https://doi.org/10.1007/s12303-017-0034-4

Chicco, D., Warrens, M. J., & Jurman, G. (2021). The Matthews Correlation Coefficient (MCC) is More Informative Than Cohen's Kappa and Brier Score in Binary Classification Assessment. *IEEE Access*, *9*, 78368–78381. https://doi.org/10.1109/ACCESS.2021.3084050

Chow, V. te. (1959). *Open-channel hydraulics* . McGraw-Hill (McGraw-Hill civil engineering series).

Collazos, D. (2022). *Data required for modeling landslides with LISEM in Norway - Internship Report*.

Dalianis, H. (2018). Evaluation Metrics and Evaluation. In *Clinical Text Mining* (pp. 45–53). Springer International Publishing. https://doi.org/10.1007/978-3-319-78503-5_6

Das, B. M. (2011). Slope Stabilization Methods. In *Geotechnical Engineering Handbook*. J. Ross Publishing, Inc. https://app.knovel.com/hotlink/khtml/id:kt00C811G1/geotechnical-engineering/slope-stabilization-methods

Depina, I., Devoli, G., & Oguz, E. A. (2021). *KlimaDigital Case Study*.

Ekker, R., Kværne, K., Os, A., Humstad, T., Wartiainen, A., Eide, V., & Hansen, R. K. (2013). regObs-public database for submitting and sharing observations. In *International Snow Science Workshop*. https://arc.lib.montana.edu/snow-science/objects/ISSW13_paper_P5-42.pdf

Fan, L., Lehmann, P., McArdell, B., & Or, D. (2017). Linking rainfall-induced landslides with debris flows runout patterns towards catchment scale hazard assessment. *Geomorphology*, *280*, 1–15. https://doi.org/10.1016/j.geomorph.2016.10.007

Fille, L. (2017). *Back calculation of Debris flow Run-Out & Entrainment Using the Voellmy Rheology* [Master thesis]. Norwegian University of Science and Technology (NTNU).

Fischer, L., Rubensdotter, L., Sletten, K., Stalsberg, K., Melchiorre, C., Melchiorre, P., & Jaboyedoff, M. (2012). Debris flow modeling for susceptibility mapping at regional to national scale in Norway. *Landslides and Engineered Slopes: Protecting Society through Improved Understanding*, 723–729.

Froude, M. J., & Petley, D. N. (2018). Global fatal landslide occurrence from 2004 to 2016. *Natural Hazards and Earth System Sciences*, *18*(8), 2161–2181. https://doi.org/10.5194/nhess-18-2161-2018

Geovekst. (2020). *Produktspesifikasjon for ortofoto*. https://norgeibilder.no/

Google. (2010, April). *Mellomriksvegen - E14*. Street View. https://www.google.com/maps/@63.4575107,11.4342983,3a,90y,11.93h,89.22t/data=!3m9!1e1!3m7!1sM3u0_Aqbr0DvkLz0EYZVyg!2e0!7i13312!8i6656!9m2!1b1!2i18

Greco, R., & Pagano, L. (2017). Basic features of the predictive tools of early warning systems for water-related natural hazards: Examples for shallow landslides. *Natural Hazards and Earth System Sciences*, *17*(12), 2213–2227. https://doi.org/10.5194/nhess-17-2213-2017

Green, W. H., & Ampt, G. A. (1911). Studies on Soil Phyics. *The Journal of Agricultural Science*, *4*(1), 1–24. https://doi.org/10.1017/S0021859600001441

Guzzetti, F., Gariano, S. L., Peruccacci, S., Brunetti, M. T., Marchesini, I., Rossi, M., & Melillo, M. (2020). Geographical landslide early warning systems. *Earth-Science Reviews*, *200*(October 2019), 102973. https://doi.org/10.1016/j.earscirev.2019.102973

Haan, C. T., Barfield, B. J., & Hayes, J. C. (1994). Bulk Density. In *Design Hydrology and Sedimentology for Small Catchments* (pp. 212–213). Elsevier. https://app.knovel.com/hotlink/khtml/id:kt00BFK7N1/design-hydrology-sedimentology/bulk-density

Hengl, T., de Jesus, J. M., MacMillan, R. A., Batjes, N. H., Heuvelink, G. B. M., Ribeiro, E., Samuel-Rosa, A., Kempen, B., Leenaars, J. G. B., Walsh, M. G., & Gonzalez, M. R. (2014). SoilGrids1km — Global Soil Information Based on Automated Mapping. *PLoS ONE*, *9*(8), e105992. https://doi.org/10.1371/journal.pone.0105992

Holtz, R. D., & Kovacs, W. D. (1981). *An introduction to Geotechnical Engineering* (K. Skrable, Ed.). Prentice-Hall.

Homleid, M., Noer, G., Tveter, F. T., & Østvand, L. (2021a). *Verification of Operational Weather Prediction Models June to August 2021*. https://www.met.no/publikasjoner/met-info/met-info-2021

Homleid, M., Noer, G., Tveter, F. T., & Østvand, L. (2021b). *Verification of Operational Weather Prediction Models September to November 2021*. https://www.met.no/publikasjoner/met-info/met-info-2021

Hovland, H. J. (1977). Three-Dimensional Slope Stability Analysis Method. *Journal of the Geotechnical Engineering Division*, *103*(9), 971–986. https://doi.org/10.1061/AJGEB6.0000493

Johari, A., & Javadi, A. A. (2012). Reliability assessment of infinite slope stability using the jointly distributed random variables method. *Scientia Iranica*, *19*(3), 423–429. https://doi.org/10.1016/j.scient.2012.04.006

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

Leitão, J. P., Boonya-Aroonnet, S., Prodanović, D., & Maksimović, Č. (2008). *Influence of DEM resolution on surface flow network for pluvial urban flooding and simulations of integrated system*.

Melo, P. de O., de Britto, R. M., Fontainha, T. C., Leiras, A., & Bandeira, R. A. de M. (2017). Evaluation of community leaders' perception regarding Alerta Rio, the warning system for landslides caused by heavy rains in Rio de Janeiro. *Natural Hazards*, *89*(3), 1343–1368. https://doi.org/10.1007/s11069-017-3023-z

Melo, R., van Asch, T., & Zêzere, J. L. (2018). Debris flow run-out simulation and analysis using a dynamic model. *Natural Hazards and Earth System Sciences*, *18*(2), 555–570. https://doi.org/10.5194/nhess-18-555-2018

Müller, M., Homleid, M., Ivarsson, K. I., Køltzow, M. A. Ø., Lindskog, M., Midtbø, K. H., Andrae, U., Aspelien, T., Berggren, L., Bjørge, D., Dahlgren, P., Kristiansen, J., Randriamampianina, R., Ridal, M., & Vignes, O. (2017). AROME-MetCoOp: A nordic convective-scale operational weather prediction model. *Weather and Forecasting*, *32*(2), 609–627. https://doi.org/10.1175/WAF-D-16-0099.1

NGU. (2021, September 6). *Run-out modelling of landslides*. Norges Geologiske Undersøkelse. https://www.ngu.no/en/topic/run-out-modelling-landslides

Oguz, E. A., Depina, I., Myhre, B., Devoli, G., Rustad, H., & Thakur, V. (2022). IoT-based hydrological monitoring of water-induced landslides: a case study in central Norway. *Bulletin of Engineering Geology and the Environment*, *81*(5), 217. https://doi.org/10.1007/s10064-022-02721-z

Pecoraro, G., Calvello, M., & Piciullo, L. (2019). Monitoring strategies for local landslide early warning systems. *Landslides*, *16*(2), 213–231. https://doi.org/10.1007/s10346-018-1068-z

Piciullo, L., Calvello, M., & Cepeda, J. M. (2018). Territorial early warning systems for rainfall-induced landslides. *Earth-Science Reviews*, *179*, 228–247. https://doi.org/10.1016/j.earscirev.2018.02.013

Proyecto Multinacional Andino (PMA): Geociencias para las Comunidades Andinas. (2007). Movimientos en masa en la región andina: una guía para la evaluación de amenazas. *Publicación Geológica Multinacional*, *4*. https://hdl.handle.net/20.500.12544/2830

Pudasaini, S. P. (2012). A general two-phase debris flow model. *Journal of Geophysical Research: Earth Surface*, *117*(3), 1–28. https://doi.org/10.1029/2011JF002186

Saxton, K. E., & Rawls, W. J. (2006). Soil Water Characteristic Estimates by Texture and Organic Matter for Hydrologic Solutions. *Soil Science Society of America Journal*, *70*(5), 1569–1578. https://doi.org/10.2136/sssaj2005.0117

Scheidl, C., Rickenmann, D., & McArdell, B. W. (2013). Runout Prediction of Debris Flows and Similar Mass Movements. In *Landslide Science and Practice* (Vol. 3, pp. 221–229). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-31310-3_30

Schilirò, L., Cepeda, J., Devoli, G., & Piciullo, L. (2021). Regional Analyses of Rainfall-Induced Landslide Initiation in Upper Gudbrandsdalen (South-Eastern Norway) Using TRIGRS Model. *Geosciences*, *11*(1), 35. https://doi.org/10.3390/geosciences11010035

Schumacher, R. S., Hill, A. J., Klein, M., Nelson, J. A., Erickson, M. J., Trojniak, S. M., & Herman, G. R. (2021). From Random Forests to Flood Forecasts: A Research to Operations Success Story. *Bulletin of the American Meteorological Society*, *102*(9), E1742–E1755. https://doi.org/10.1175/BAMS-D-20-0186.1

Segoni, S., Piciullo, L., & Gariano, S. L. (2018). Preface: Landslide early warning systems: monitoring systems, rainfall thresholds, warning models, performance evaluation and risk perception. *Natural Hazards and Earth System Sciences*, *18*(12), 3179–3186. https://doi.org/10.5194/nhess-18-3179-2018

Sene, K. (2016). Introduction. In *Hydrometeorology: Forecasting and Applications* (pp. 1–34). Springer International Publishing. https://doi.org/10.1007/978-3-319-23546-2

SGC. (2017). *Guía metodológica para la zonificación de amenaza por movimientos en masa escala 1: 25.000* (Servicio Geológico Colombiano, Ed.; 1st ed.). Servicio Geológico Colombiano. https://doi.org/10.32685/9789585978225

Sim, J., & Wright, C. C. (2005). The Kappa Statistic in Reliability Studies: Use, Interpretation, and Sample Size Requirements. *Physical Therapy*, *85*(3), 257–268. https://doi.org/10.1093/ptj/85.3.257

Statens vegvesen. (2016). *Geoteknisk Datarapport E14 TS-tiltak Nr. Vd1399A*. https://geo.ngu.no/kart/nadag/

Stranden, H. B., Devoli, G., Engeset, R., Saloranta, T., Andersen, J., Melvold, K., Humstad, T., Orset, K. I., Orthe, N. K., Ekker, R., Assessment, R., Flows, D., Making, D., Science, A., & Assessment, R. (2020). *14th Congress INTERPRAEVENT 2020*.

Terratec AS. (2016). *NDH Malvik-Stjørdal 2pkt 2016 Prosjektrapport*. www.hoydedata.no

Thiebes, B., Bell, R., Glade, T., Jäger, S., Mayer, J., Anderson, M., & Holcombe, L. (2014). Integration of a limit-equilibrium model into a landslide early warning system. *Landslides*, *11*(5), 859–875. https://doi.org/10.1007/s10346-013-0416-2

UNISDR. (2006). Developing Early Warning Systems: A Checklist. *Third International Conference on Early Warning*, *March*, 1–13.

UNISDR. (2015). *Sendai Framework for Disaster Risk Reduction 2015 - 2030*. https://www.preventionweb.net/files/43291_sendaiframeworkfordrren.pdf

van Asch, Th. W. J., Tang, C., Alkema, D., Zhu, J., & Zhou, W. (2014). An integrated model to assess critical rainfall thresholds for run-out distances of debris flows. *Natural Hazards*, *70*(1), 299–311. https://doi.org/10.1007/s11069-013-0810-z

van Beek, L. P. H. (2002). *Assessment of the influence of changes in land use and climate on landslide activity in a Mediterranean environment*.

van den Bout, B., Lombardo, L., Chiyang, M., van Westen, C., & Jetten, V. (2021). Physically-based catchment-scale prediction of slope failure volume and geometry. *Engineering Geology*, *284*(November 2020), 105942. https://doi.org/10.1016/j.enggeo.2020.105942

van den Bout, B., van Westen, C. J., & Jetten, V. G. (2022). Non-Monotonic Relationships between Return Periods of Precipitation Surface Hazard Intensity. *Water*, *14*(9), 1348. https://doi.org/10.3390/w14091348

van Zuidam, R. A. (1986). *Aerial photo-interpretation in terrain analysis and geomorphologic mapping*. Smits Publishers.

Verruijt, A. (2018a). Slope Stability. In *An Introduction to Soil Mechanics* (pp. 343–348). Springer, Cham. https://doi.org/10.1007/978-3-319-61185-3_45

Verruijt, A. (2018b). Stability of Infinite Slope. *Theory and Applications of Transport in Porous Media*, *30*, 335–342. https://doi.org/10.1007/978-3-319-61185-3_44

von Ruette, J., Lehmann, P., & Or, D. (2013a). Rainfall-triggered shallow landslides at catchment scale: Threshold mechanics-based modeling for abruptness and localization. *Water Resources Research*, *49*(10), 6266–6285. https://doi.org/10.1002/wrcr.20418

von Ruette, J., Lehmann, P., & Or, D. (2013b). Rainfall-triggered shallow landslides at catchment scale: Threshold mechanics-based modeling for abruptness and localization. *Water Resources Research*, *49*(10), 6266–6285. https://doi.org/10.1002/wrcr.20418

Wang, N., Lombardo, L., Luigi, S., Cheng, W., Liu, C., Xiong, J., & Wang, R. (2021). Using satellite rainfall products to assess the triggering conditions for hydro-morphological processes in different geomorphological settings in China. *International Journal of Applied Earth Observations and Geoinformation*, *102*(June), 102350. https://doi.org/10.1016/j.jag.2021.102350

Wyllie, D. C., & Norrish, N. I. (2017). ROCK STRENGTH PROPERTIES AMD THEIR MEASUREMENT. In *ROCK SLOPE ENGINEERING* (p. 386). https://onlinepubs.trb.org/Onlinepubs/sr/sr247/sr247-014.pdf

Xu, Q., Yuan, Y., Zeng, Y., & Hack, R. (2011). Some new pre-warning criteria for creep slope failure. *Science China Technological Sciences*, *54*(1 SUPPL.), 210–220. https://doi.org/10.1007/s11431-011-4640-5

Yu, F., Hunt, A., Egli, M., & Raab, G. (2019). Comparison and Contrast in Soil Depth Evolution for Steady State and Stochastic Erosion Processes: Possible Implications for Landslide Prediction. *Geochemistry, Geophysics, Geosystems*, *20*(6), 2886–2906. https://doi.org/10.1029/2018GC008125

Zhang, L., Wu, F., Zheng, Y., Chen, L., Zhang, J., & Li, X. (2018). Probabilistic calibration of a coupled hydromechanical slope stability model with integration of multiple observations. *Georisk*, *12*(3), 169–182. https://doi.org/10.1080/17499518.2018.1440317

ANNEX

Granulometry and moisture content for soil hydraulic properties estimation.

Table 10. Granulometry of boreholes 1 and 2 in Figure 4 Statens vegvesen (2016) geotechnical report.

Borehole	X	Y	Date	Depth [m]	Water content (%) ★	Clay (%)	Sand (%)	Gravel (%)
1	7038921.5	621324.35	26/08/2015	1.5	3.2	3.5	33.3	59.1
1	7038921.5	621324.35	26/08/2015	3.5	7.1	4.1	54.4	21.2
2	7038892.1	621326.43	26/08/2015	1.5	22	11.2	57.3	12.2
2	7038892.1	621326.43	26/08/2015	3.5	13.9	11.2	45.4	32.7
2	7038892.1	621326.43	26/08/2015	5.5	12.8	21	47	19.2

Table 11. Derived soil hydraulic properties from SPAW hydrology (Saxton & Rawls, 2006) using granulometry from Table 10.

Organic Wt (%) Assumed	SPAW Porosity (%) thetas	SPAW Field Capacity (%) thetar	SPAW Saturated Hydraulic Conductivity (Ksat) [mm/hr]	SPAW Saturated Hydraulic Conductivity (Ksat) [m/s]	Matric Potential [kPa]	Pressure head [m] with ★
1.5	42.7	21.9	13.77	3.83E-06	1500	0.153
0.5	39.5	14.8	37.97	1.05E-05	185	0.019
1.5	42.4	18.8	29.12	8.09E-06	29	0.003
0.5	39.6	20.3	12.4	3.44E-06	134	0.014
0	39.2	24.3	6.56	1.82E-06	1414	0.144
Average	40.68	20.02		5.55E-06		0.078