Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Engineering
Department of Geoscience and Petroleum

Harald Birkeland Larsen

Physically Based Modelling of Shallow Landslides with "TRIGRS"

A Review of Applications and an Implementation in the Case Study of Jølster 2019, Western Norway

Master's thesis in Geology
Supervisor: Luca Piciullo (NGI) & Bjørn Frengstad (NTNU)
Co-supervisor: Luca Schilirò (Cnr)
May 2021

The fatal "Årnes" landslide of the Jølster 2019 case study.

NTNU
Norwegian University of Science and Technology

Harald Birkeland Larsen

Physically Based Modelling of Shallow Landslides with "TRIGRS"

A Review of Applications and an Implementation in the Case Study of Jølster 2019, Western Norway

Master's thesis in Geology
Supervisor: Luca Piciullo (NGI) & Bjørn Frengstad (NTNU)
Co-supervisor: Luca Schilirò (Cnr)
May 2021

Norwegian University of Science and Technology
Faculty of Engineering
Department of Geoscience and Petroleum

NTNU
Norwegian University of
Science and Technology

Abstract

"TRIGRS" is a commonly used physically based landslide model, considering transient precipitation, runoff, infiltration, pore pressures and slope stability. This thesis presents a comprehensive literature review, identifying 42 papers applying TRIGRS in the period 2008—2020. Research goals, locations, input data and methods of performance evaluation vary substantially. In this thesis, TRIGRS is further applied to the case study of Jølster 2019, an extreme, multiple landslide event in Western Norway. Model calibration, by systematic variation of sensitive parameters, is conducted due to a data scarcity. First, back-calculation of landslides, in two sub-areas assumed representative of specific soil types, is conducted to shed light on the spatial distribution of soil parameters. It is found that estimation of both objective and relative parameters is challenging due to shortcomings of (i) TRIGRS, (ii) the calibration procedure, (iii) input data, including assumptions on initial soil conditions and (iv) uncertain landslide initiation mechanisms. Correlation (R^2 = 30%) is found between slope angles and soil depths of field measurements, improving modelling compared to uniform soil depth. Unsaturated soil analysis appears inapplicable despite unusually dry initial conditions. Further, effects of chosen evaluation criteria, including ROC, AUC, D2PC, CSI and MCC, on estimated performance are investigated. The difference between overall and threshold-specific performance (e.g., AUC is not a good proxy for the latter), and the case-specificity of performance measures (e.g., AUC should not be compared across study areas), is stressed. Two parameter combinations, optimizing different evaluation criteria, are presented. One parameter combination is subject to a novel approach, where predictions and "error magnitudes" are assessed in relation to spatial data. Both combinations identify general unstable areas (correctly predicting 25.5% – 26.6% of landslide cells) but are accompanied by numerous wrong predictions. Assumed soil property homogeneity leaves slope angle a strong predictor of estimated factor of safety, which in in combination with varying slope angles for Jølster landslides ensures their prediction and simultaneous lack of false positives are incombinable. It is indicated that alternative initiation mechanisms, not accounted for by TRIGRS, are present for a proportion of Jølster landslides. TRIGRS' practical utility is thereby limited in this and similar areas and events.

Sammendrag

"TRIGRS" er en mye brukt fysisk-basert jordskred-modell som tar hensyn til tidsvarierende nedbør, avrenning, infiltrasjon, poretrykk og skråningsstabilitet. Denne oppgaven presenterer en omfattende litteraturundersøkelse, som identifiserer 42 artikler angående TRIGRS i perioden 2008–2020. Mål, lokasjoner, inngangsdata og metoder for evaluering av resultater varierer sterkt. Videre implementeres TRIGRS i case-studien Jølster 2019, en ekstrem jordskredhendelse på Vestlandet. Kalibrering av modellen, ved systematisk variasjon av sensitive parametre, er nødvendig grunnet mangel på data. Først utforskes effekter av valg av evalueringskriterier (ROC, AUC, D2PC, CSI og MCC) på bedømmelse av modell-prestasjon. Det er viktig å skille mellom "generell" og "terskel-spesifikk" prestasjon (f.eks., AUC er ingen god inikasjon på sistnevnte), samt å være klar over stedegenheten til prestasjonsmål (f.eks., AUC bør ikke sammenlignes på tvers av studieområder). Videre tibakeregnes skred for å undersøke romlig fordeling av jordparametre. Dette foregår i to delområder antatt å representere forskjellige jordarter. Det viser seg at bestemmelse av både objektive or relative parametre er utfordrende grunnet mangler ved (i) TRIGRS, (ii) kalibreringsprosedyren, (iii) inngangsdata, inkludert antakelser om innledende jordforhold og (iv) kunnskap om skredenes utløsningsmekanismer. Korrelasjon (R^2 = 30%) mellom skråningsvinkel og observert jordmektighet forbedrer modellering sammenliknet med antakelser om konstant mektighet. Umettet analyse presterer dårlig selv under usedvanlig tørre forhold. Det presenteres to resulterende kombinasjoner av parametre, som optimaliserer forskjellige evalueringskriterier. En parameter-kombinasjon evalueres på en nyvinnende måte, hvor resultater og "feil-størrelser" sammenliknes med romlig data. Begge kombinasjonene identifiserer generelt ustabile områder (forutser 25,5% – 26,6% av utløsningsområder), men akkompagneres av mange feil. Antakelsen om homogene jordparametre gjør at sikkerhetsfaktor i stor grad bestemmes av skråningsvinkel. Samtidig varierer utløsningsområdenes skråningsvinkler stort, og de kan følgelig umulig identifiseres uten en stor andel falske positive. Variasjonen indikerere at utløsningsmekanismer som ikke hensyntas av TRIGRS er til stede, og TRIGRS' praktiske nytte begrenses deretter.

Preface

This thesis is written as part of the two-year master's degree in engineering geology at Norwegian University of Science and Technology (NTNU). It is written entirely by myself during two full semesters, and concludes five incredible years in Trondheim.

Writing the thesis has been an interesting procedure. An unknown world of papers and journals suddenly opens; you realize considerable effort is being put into topics you did not know existed, and after reading enough articles, you are convinced certain topics are of uttermost importance. I have broadened my knowledge regarding landslides and their modelling, as well as how to write (somewhat cumbersome) code in Python. But more importantly, I've learnt many lessons on the importance of planning, that if planned, things generally take longer than anticipated, and that neither research nor myself is as impressive as thought.

A number of people have helped me write, improve and enjoy this thesis. Bjørn, it has been a true pleasure to have you as my internal supervisor. Thank you for your guidance, timely metaphors and ability to motivate and inspire during every conversation. Luca P., thank you for providing a great thesis topic, steering me in the correct (and realistic) direction, as well as for making every meeting an enjoyment. Luca S., thank you for your advice and helpful revisions during the later part of writing. Thank you to HG, my partner, for your lack of interest in geology and refreshing tales about elevators and artificial intelligence. And finally, thank you to my friends and "colleagues" at PTS, having ensured lunch and coffee breaks of the highest quality.

Contents

List of Figures

List of Tables

Abbreviations

AUC	Area under the curve
CSI	Critical success index
δNDVI	Change in Normalized Difference Vegetation Index
D2PC	Distance to perfect classification
DEM	Digital elevation model
FN	False negative
FNR	False negative rate
FP	False positive
FPR	False positive rate
FS	Factor of safety
MCC	Matthew's correlation coefficient
MET	The Norwegian Meteorological Institute
NGU	Geological Survey of Norway
N	Negatives
NVE	The Norwegian Water Resources and Energy Directorate
P	Positives
PBM	Physically based model(ling)
ROC	Receiver operating characteristics
SVV	The Norwegian Public Roads Administration
T_{FS}	Threshold value for slope stability/instability
TN	True negative
TNR	True negative rate
TP	True positive
TPR	True positive rate
TRIGRS	Transient Rainfall Infiltration and Grid Based Slope Stability model

Chapter 1

Introduction

1.1 Context and research questions

Landsliding, or “the movement of rock, earth or debris down a slope” (Varnes, 1978), is a global hazard that has, as of yet, affected more than 4.7 million people during the 21st century (Guha-Sapir, 2018). Blameworthy of considerable economic losses, its frequency is expected to increase due to demographic (Froude & Petley, 2018) and climatic (Gariano & Guzzetti, 2016) changes. Landslides are generally classified in terms of material and motion, and those involving soil have brought the worst consequences (Dowling & Santi, 2014). The constituent term “rainfall-induced shallow landslides” considers rapidly initiated, often devastating events (Hungr et al., 2001), of which one method of risk mitigation is the application of physically based models (Corominas et al., 2014).

The “Transient Rainfall Infiltration and Grid-Based Slope Stability" model (TRIGRS, Baum et al., 2002; Baum et al., 2008) considers such landslides, by evaluating surface infiltration, groundwater response, pore pressures, “infinite slope stability” and runoff. The open source software has been applied in numerous case studies for different purposes (Park et al., 2013), but an overview of its usage, common input data or performance does not exist. As knowledge sharing may steer further research in a good direction, a comprehensive literature review — for the period of 2008–2020 – was conducted to throw light on the following research questions:

1. How, and under what conditions has TRIGRS been applied?
2. More specifically: How “well” can TRIGRS predict the location and timing of landslides?

Figure 1.1: Photography of the "Årnes" landslide of the Jølster event. Source: Jan Helge (SVV) (regObs.no).

The rainy, mountainous western coast of Norway is prone to rainfall-induced shallow landslides, but few attempts at implementing physically based models exist (Melchiorre and Frattini, 2012; Schilirò et al., 2021). Therefore, TRIGRS was applied in the extreme, 120-landslide case study event of Jølster, 2019. Scarcity of input data necessitated an approach of soil parameter calibration, exploring the following research question:

1. Can TRIGRS "back-calculate" absolute or relative soil parameters for the Jølster case study? If so, the case event may provide insights on the spatial distribution of soil properties which may be utilized in future risk mitigation.

2. Is the Jølster case study suitable for the application of TRIGRS in terms of e.g., topographical features, landslide characteristics and reliability of input data? This question considers the model's ability to separate stable and unstable slopes, through exploration of model results by a novel approach.

Several sub-questions are nested within the above, such as:

- How reliable is available model input data?
- Does parameter calibration in separate areas, representing different soil types, improve model performance compared to the assumption of uniform soil properties?
- Is soil depth predictable from topographical variables?
- Which approach is most suited for the Jølster case study: saturated or unsaturated soil analysis?
- What do we know about Jølster landslides' characteristics?

An additional research question emerged during the course of work. As methods of evaluating physically based model performance were found to vary greatly – which complicates comparisons and may hinder development – it was asked:

- How does the choice of evaluation criteria affect model calibration and performance evaluation?

1.2 Structure

This thesis begins (Chapter 2) by providing an overview of the phenomena of landsliding, both internationally and in Norway. This includes consequences and the distribution of risk, as well as international and Norwegian classifications. It further dives into the concept of rainfall-induced shallow landslides, examining their predisposing and triggering factors along with susceptibility and risk assessments.

Chapter 3 is devoted to TRIGRS. Model background, theory and operation is presented, as well as method and results from the literature review. Chapter 4 examines Jølster and its extreme landslide event. It then describes available input data and related considerations, both in terms of basic necessities (e.g., digital topography) and more uncertain inputs (e.g., soil properties). Lastly, the method of applying TRIGRS, which is mainly related to the calibration procedure, is introduced.

Chapter 5 presents results from a statistical analysis of Jølster landslides, attempted prediction of soil depth from topographical variables, an investigation into effects of differing evaluation criteria and soil parameter estimation by model calibration for both saturated and unsaturated versions of TRIGRS. Chapter 6 discusses priorly defined research questions by reflecting on theory and results, while chapter 7 summarizes findings in conclusions and suggestions for further research.

Chapter 2

Background

This introductory theory begins with a bird's eye view, exploring landsliding as a global hazard. Focus is then devoted to landsliding in Norway, examining their locations, types, frequencies and impacts. A closer look is given at landslide classifications, and specifically *shallow landslides* and related triggering conditions. Finally, an introduction to the potentially risk-mitigating tool of physically based landslide models is given.

2.1 Landslides

2.1.1 Global occurrence and impacts

Nature enables human flourishing, but threats of sudden violence have always been present. *Natural hazards*, or "any natural process that may cause social, economic or environmental disruption" (UNISDR, 2009, rewritten), such as floods, droughts, and volcanic activity, have since at least the Genesis flood been a part of our experience. Although the number has declined, natural disasters (causing serious disruption of societal functions), still cause about 60 000 direct fatalities per year globally (Ritchie & Roser, 2014).

Landsliding, or "the movement of rock, earth or debris down a slope" (Section 2.2 is not the most lethal hazard. Yet, in the 21st century, the phenomena affected more than of 4.7 million people and caused more than 16 000 fatalities (Guha-Sapir, 2018). These 1.2 percent of total natural hazard fatalities in the same period concentrate in large catastrophes; only two events are responsible for 50% of fatalities associated with "debris flows" (Dowling & Santi, 2014), while the most damaging event recorded, the 1920 Haiyuan landslides of China, killed 100 000 people (Tianchi, 1994).

The distribution of landslide risk

Although subject to temporal variation due to e.g., climate anomalies (Froude & Petley, 2018), the number of recorded, fatal landslide events has increased during the period 1950–2014 (Dowling & Santi, 2014; Haque et al., 2019). However, this *perceived* increase may relate to demography rather than landsliding itself (e.g., Kjekstad

& Highland, 2009). Vulnerability has increased due to population growth, urbanization and lacking slope management (Redshaw & Bottomley, 2020), especially in developing countries troubled with corruption and deficient emergency preparation (Dowling & Santi, 2014). Consequently, landslides are increasingly induced by e.g., construction, illegal mining or hill cutting, while under-reporting is gradually reduced due to technological advances (Froude & Petley, 2018).

Landslides occur where topography, geology, climate and anthropology combine in unfavorable ways. Not surprisingly, risk follows topographic relief on prior or current continental plate boundaries, with hot spots in the Rocky Mountains, Andes, Alps, Middle East and large parts of Asia (Fig. 2.1).

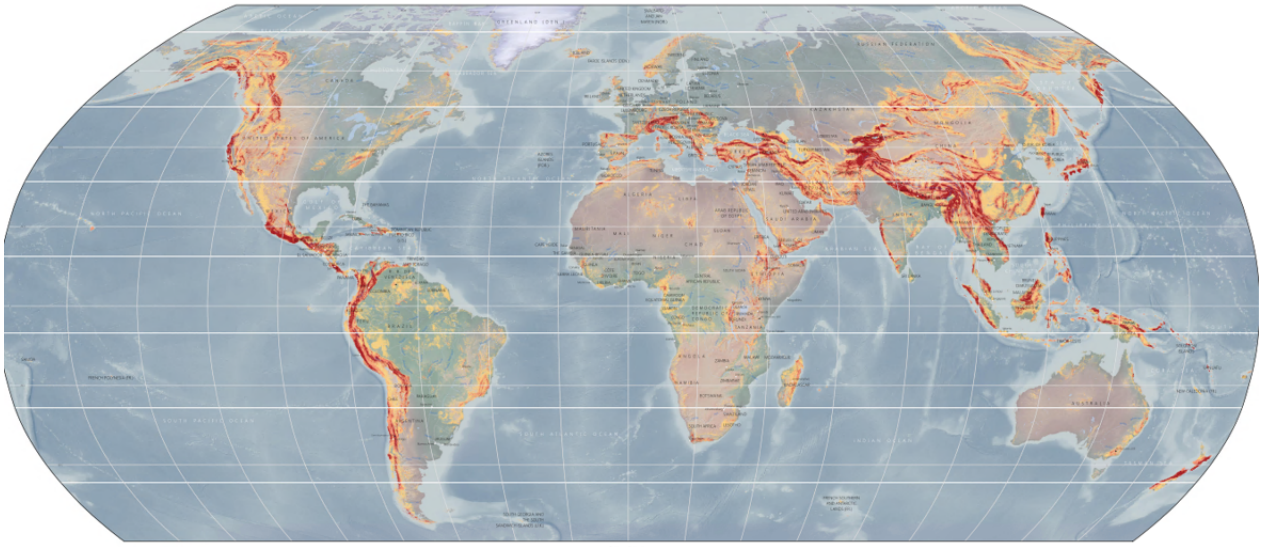


Table 7 – Categorization of landslide hazard values for ThinkHazard!

ThinkHazard! hazard category	**Proposed ranges of annual landslide frequency / sq. km**
Very low	< 0.0001
Low	0.0001 - 0.001
Medium	0.001 - 0.01
High	> 0.01

Figure 2.1: Estimated world landslide hazard (Redshaw & Bottomley, 2020, modified). Results are based on modelling and analysis of 15.5 million events for the period 1980–2018.

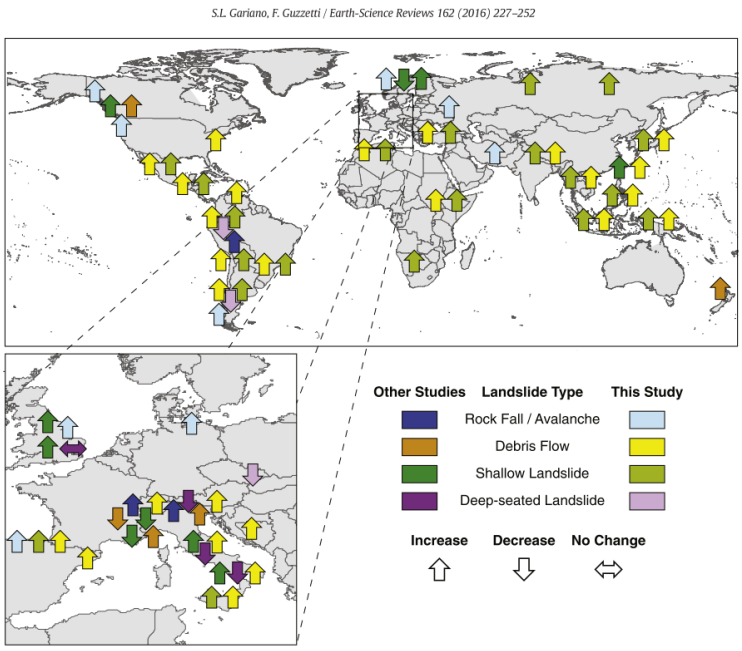


Figure 2.2: Major trends in landslide activity driven by projected climate change (Gariano & Guzzetti, 2016).

Risk is affected by anthropogenic and natural climate change, although not as strongly as by demographic changes (Froude & Petley, 2018). However, as uncertainties in emission scenarios and climate models supersede those of landslide risk's response to climate, only large scale trends are estimable (Gariano & Guzzetti, 2016). In any case, several areas may expect increased activity (Fig. 2.2), with more extreme precipitation – showing a significant spatial and temporal correlation to fatal landslides (Haque et al., 2019) – being the main factor.

Economic costs of landsliding

Besides a treat to human life, landslides cause a range of direct (e.g., loss of property, immediate engineering works) and indirect (e.g., flooding, loss of productive land, reduced real estate values, adverse ecological effects or traffic delays) losses (Kjekstad & Highland, 2009). Although Fleming and Taylor (1980) noted taxpayers' and officials' unawareness to landslide costs, few attempts have been made to assess these on a large scale. Estimates are hampered by variable methodologies, unreliable data and the common merging of landslides into larger disasters (Kjekstad & Highland, 2009). Nevertheless, certain country-wise estimates exist including annual costs in excess of 1 billion USD (3.16 billion USD converted to 2020-dollars) for both USA (Schuster, 1978) and Canada (Cruden et al., 1989) – above costs of floods, earthquakes and hurricanes. Case-specific cost estimates (e.g., Perera et al., 2018; Winter et al., 2016) also exist, both emphasizing the need for improved methodologies.

2.1.2 Landsliding in Norway

Norway is not among the highest-risk areas globally (Fig. 2.1), but have yet experienced more than 4000 historical landslide fatalities (Furseth, 2006). As is also the case globally, few but major events such as "Gauldalsraset" (year 1345: Rokoengen et al., 2001), "Verdalsraset (year 1893: Verdal Historielag, 2020), and "Loenulykkene" (year 1905, 1936 and 1950: Bryhni & Askheim, 2020) have contributed to large parts of losses.

Economic costs of landsliding in Norway

Economic aspects of landsliding in Norway is little discussed and the author knows of no national estimates. However, landslide-related insurance payments were valued at 77 million Nok (7.7 million Euro, April 2021) in 2019 alone (NNPP, 2020), reflecting costs to private property only. For casualties one may use the Norwegian Ministry of Finance's established value of statistic life (30.49 mill nok: Det Kongelige Finansdepartement, 2014) and calculate oneself. For indirect costs one may consider a geological and climatological similar country, Scotland, where Winter et al. (2016) showed how modest landslide events induced severe indirect costs due to closed transportation networks. Simultaneously, 1500–2100 landslides hit Norwegian roads each year, of which about a third lead to full or partial closure (Bjordal & Helle, 2011). The Norwegian Water Resources and Energy Directorate (NVE, responsible for landslide risk mitigation) presented a list of desirable mitigation measures valued at 2.7 billion nok in 2018 (Andersson et al., 2018).

Norway's topography and climate

Understanding landsliding in Norway requires details on the conditions in which they occur. Norway's 385 200 km^2 area stretches from the North Sea, along the Norwegian Sea to the Barents Sea with Sweden, Finland and Russia on its eastern border. During the last 60 million years, the initially uplifted and tilted Norwegian peneplain (featureless, continental plain) has been eroded by rivers and glaciers through more than 40 ice ages (Bargel et al., 2008). This created relatively gentle topography in eastern parts while fjords and mountains of more than 2000 m.a.s.l. characterize large parts of the western coast

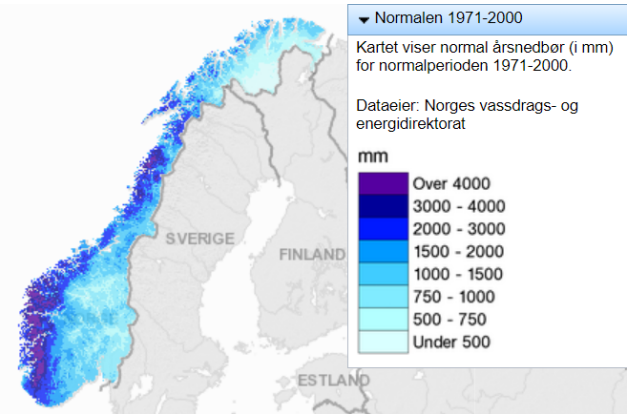


Figure 2.3: Average annual precipitation [mm] for the period 1971–2000 (http://www.senorge.no/?p=klima).

The coast is characterized by temperate climate, with especially abundant precipitation in its south-western areas (>3000 mm/year, Fig. 2.3) Orogenic uplift makes for less rain eastwards of mountains (continental climate), while high altitudes and the northernmost parts experience a polar climate. As a consequence of climatic variability, slopes in Norway's interior are more "initially unstable", requiring less precipitation than their western companions for failure (Krøgli et al., 2018; Meyer et al., 2012).

Bedrock and related landslides

Norway's bedrock is mainly grouped in a Precambrian basement, numerous overthrust sheets and local zones of special geology. The southern and western Precambrian basement includes gneiss, granites and certain metamorphosed rocks, while overthrust sheets of central and northern Norway include e.g., gabbro, schist, phyllite, metamorphosed sandstone and greenstone.

Rock falls, Norway's most widespread (but not particularly damaging) landslide type (Bjordal & Helle, 2011), unsurprisingly concentrate in western, high-relief areas. C. 50% of rock falls on roads have been registered in former Sogn of Fjordane county alone (Bjordal & Helle, 2011), and these have generally played a role in landscape development creating steep slopes and underlying talus. Also concentrated in the west are rock slides, where several sites are constantly monitored (e.g., Åkerneset, Harbitz et al., 2014).

Quaternary geology and landslides in soil

The distribution of Norway's soil deposits (Quaternary geology) is at large scale a function of numerous glacial and interglacial periods, while local-scale landforms are often attributed to the last glacial cycle ending about 12 000 years ago (Olsen et al., 2013). Glacial till of varying continuity, considered the dominant Quaternary sediment in Norway, is visible along slopes and at altitudes (Haldorsen & Jenssen, 1983), and generally locates underneath fills of glaciofluvial and fluvial origin in valley bottoms. Marine sediments cover large parts of lower-lying areas due to isostatic uplift (Ramberg et al., 2013). Norwegian landslides in soil are separated in two main types. A majority of "rainfall-induces shallow landslides and debris flows" occur in western areas (Bjordal & Helle, 2011), may take place in extreme, multiple-landslide events (Rouault, 2020), and are expected to increase in frequency due to global warming (Hanssen-Bauer & Hisdal, 2017). On the other hand, a peculiarity for Norway and Scandinavia, along with Canada, Alaska and Japan (Geertsema, 2013) is the existence of quick clay. Marine clay may through circulation of fresh water become unstable, and this extremely sensitive material (St.>50: Rankka et al., 2004) has caused some of Norway's most destructive landslides.

2.2 Landslide classifications

2.2.1 International classifications

Several works ranging back to 1862 have grouped landslides (Cruden, 2003), but the most widely used today is that of Varnes (1978). There, a landslide is defined as “the movement of a mass of rock, debris or earth down a slope”, and is thereby not limited to "sliding" nor does there have to be "land" involved. The original system covers a range of materials and modes of propagation, yielding 29 subtypes (Fig. 2.4). *Rock* refers to any solid material except ice, *debris* refers to a mixture of material with (originally) less than 80% sand and finer, while *earth* refers to plastic and cohesive material of more than 80% sand and finer (Varnes, 1978).

Movement type	Rock	Debris	Earth
Fall	1. Rock fall	2. Debris fall	3. Earth fall
Topple	4. Rock topple	5. Debris topple	6. Earth topple
Rotational sliding	7. Rock slump	8. Debris slump	9. Earth slump
Translational sliding	10. Block slide	11. Debris slide	12. Earth slide
Lateral spreading	13. Rock spread	–	14. Earth spread
Flow	15. Rock creep	16. Talus flow	21. Dry sand flow
		17. Debris flow	22. Wet sand flow
		18. Debris avalanche	23. Quick clay flow
		19. Solifluction	24. Earth flow
		20. Soil creep	25. Rapid earth flow
			26. Loess flow
Complex	27. Rock slide-debris avalanche	28. Cambering, valley bulging	29. Earth slump-earth flow

Figure 2.4: The original landslide classification of Varnes (1978), based on the combination of movement and material type. Table from Hungr et al. (2014).

Cruden and Varnes (1996) updated the original classification by introducing rate of movement as an additional criterion. Other suggested updates (e.g., Hungr et al., 2001; Hungr et al., 2014) have noted interesting issues - such as the "artificial" separation between earth and debris - and have introduced now well-established terms such as “debris-floods”. Finally, specialized classification systems exist for e.g., pure geotechnical engineering, subaqueous conditions, permafrost and quick clay (Hungr et al., 2014).

2.2.2 The Norwegian classification

Although Varnes’ classification is well-established internationally, a different, simplified system is used in Norway. Jensen et al. (2015) summarized terms, and emphasized a continuation of existing Norwegian terminology rather than connection to international classifications. Importantly for this thesis, Norwegian landslides in soil (neglecting quick clay) are described by two terms: *flomskred* (direct translation: flood-landslide) are defined as “rapid, flood-like landslides of saturated soils in steep rivers”, and include international terms debris flows and debris floods. A broader term, *jordskred* (direct translation: soil-landslide), is defined as “channelized or non-channelized

downslope-moving masses of varying water contents". It thereby includes international terms such as debris slides and debris avalanches, but also debris flows if masses are in fact channelized.

2.2.3 Shallow vs. deep-seated landslides

Varnes (1978)'s dimensions for landslide classification are those easily observable at distance. However, one commonly encounters the term (rainfall-induced) *shallow landslides*, reflecting important differences in mechanisms and impacts (Hungr et al., 2001).

Sometimes considered as failing less than three meters below ground level (WFPA, 2020), shallow landslides refers to debris slides, debris avalanches and debris flows. Shallow depths and coarse material allow pore pressures and slope stability to be rapidly and heavily affected by surface precipitation (Iverson, 2000), identifying intense, acute rainfall as their common trigger. On the other hand, deep-seated landslides are typically large, slow-moving and have the bulk of their failure plane well below root systems (WFPA, 2020). Respective shear surfaces are typically close-to circular in saturated, fine-grained clay and silt (Hungr et al., 2014), and low permeability makes failure dependent on long-term changes in groundwater conditions (on the scale of weeks or months, (Zêzere et al., 2005)).

2.3 Rainfall-induced shallow landslides

The occurrence of shallow landslides may be divided into a set of predisposing factors (e.g, slope angle, wetness, soil types) and triggering factors (here: intense rainfall) (Corominas et al., 2014).

2.3.1 Predisposing factors

Slope angle and topography

Landsliding converts potential to kinetic energy, and some slope gradient is necessary for overcoming frictional resisting forces (see Eq. 2.5). Typical debris slide failure angles are 30°–60°, while steeper slopes tend not to support soil (Hungr et al., 2014). Slope angle may also be the triggering factor if subject to sudden anthropogenic or natural modification. Additional topographical factors include sufficient total relief for sustained motion, slope aspect (reflecting variations in vegetation and/or soil moisture), and geometrical hollows or gully floors accumulating seepage (Hungr et al., 2014).

Strength properties of soil

Soil strength resists failure, and other factors kept equal, slopes of low strength are more susceptible to failure than their high-strength counterparts. The failure state is commonly (although simplified, several models exist) described by the Mohr-Coulomb failure criterion (Yu, 2002), where strength is represented by an angle of internal friction (ϕ) and cohesion (C). The former, which value depends on size, shape and friction of grains, describes how shear strength increases with effective normal stress (slope of the line in Fig. 2.5). The latter is independent of particle friction and represents electrostatic attraction, cementation and primary valence bonding and adhesion in fine grained soils (Mitchell & Soga, 2005). *Apparent* cohesion may also arise due to e.g., negative pore pressures.

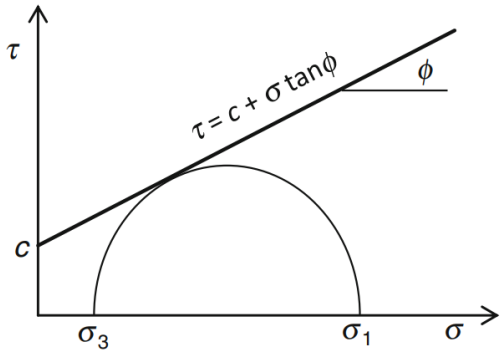


Figure 2.5: The Mohr-Coulomb stress space. The line, intersecting the vertical axis at C and which slope is equal to ϕ, represents the failure criterion. τ = shear stress, σ = normal stress while subscript 1 and 3 denotes maximum and minimum principal normal stress respectively. Figure from Zhang et al. (2015).

Hydraulic properties of soil

Infiltration and groundwater flow is largely controlled by soil hydraulic properties. Hydraulic conductivity (k) is defined through Darcy's law (Mitchell & Soga, 2005), describing the relationship between apparent water velocity (v) and hydraulic gradient (i) (Eq. 2.1). It is assumed constant in saturated soils (although affected by the viscosity and density of water to some degree), but becomes a function of water content for unsaturated soils. Its magnitude is important in determining pore pressure response, where low values favor slow, deep-seated landslides.

$$k = \frac{v}{i} \tag{2.1}$$

Soil thickness

Both driving and stabilizing forces increase with depth, while pore pressure response is delayed and reduced in magnitude (Iverson, 2000). Failure planes therefore occur at the depth of the most unfavorable conditions, depending on e.g., material properties, rainfall intensities and slope angles. Failure may also occur on bedrock sliding planes, where local high water pressure may develop (Hungr et al., 2014).

Layering

Pre-defined failure planes are not requirements, but may enable shallow landslides either through weak layers or low-friction sliding surfaces (Hungr et al., 2014). For the former, various chemical and physical processes are responsible (e.g., pre-shearing to residual friction, weathering), while the latter may occur where blankets of loose, poorly sorted soil overly bedrock (Hungr et al., 2014) or movable soil overlies frozen zones. Such permeability contrasts between layers may enable recharge of perched water tables and locally high pore pressures (Hungr et al., 2014), which is assumed a common feature in Norway.

Initial hydrological conditions

An area's initial state affects magnitudes of possible and failure-inducing pore pressure change in response to intense precipitation. The ground water table, defined as the surface where pore pressure is equal to atmospheric pressure, varies temporally and spatially depending on e.g., long-term precipitation, soil properties, topography, geology and river networks (Dingman, 2015). As a subdued replica of topography, it follows sloping terrain more closely at higher ratios of infiltration to hydraulic conductivity.

Depending on timescale, the water table as well as water content, pore pressures and flow reaches a quasi-equilibrium (Iverson, 2000). This long-term state, governing triggering of deep-seated landslides, acts as a boundary condition upon which disturbance from intense rainfall is applied. Wet antecedent conditions generally favors shallow landsliding (see e.g., the construction of most rainfall thresholds, Segoni et al., 2018). On the other hand, dry initial conditions may still facilitate landsliding through local pore pressure increase at shallow depths (Iverson, 2000).

Vegetation and land use

Vegetation in slopes may have considerable and opposite effects on stability, of which magnitudes depend on e.g., vegetation types, densities, root networks, ages, climate and season (Kalsnes & Capobianco, 2019). Stability may increase due to root strength and the binding of soil layers, as well interception which reduces infiltration, runoff, and erosion. Interception losses may be considerable (e.g., 22% for needleleaf forests, Miralles et al., 2010), although most occurs during long-term, low-intensity events rather than those typical of shallow landsliding. Contrarily, stability may decrease due to increased slope-parallel forces (tree surcharge), increased capacity for infiltration in soil and during strong winds. Nevertheless, deforestation generally increases both the frequency and magnitude of shallow landslides , depending on the specific practice applied (e.g., Lehmann et al., 2019). For the same reason, most nature-based landslide mitigation practices involve some form of vegetation (Kalsnes & Capobianco, 2019). Purely anthropogenic factors may also increase susceptibility; infrastructure development may alter slope geometries, buildings may increase sub-surface stresses, flooding of reservoirs may increase pore pressures and river interventions may increase erosion (Froude & Petley, 2018).

2.3.2 The triggering process

Shallow landslides are triggered when shear stresses exceed shear strength (Fig. 2.5), and common triggering factors include river erosion, seismic activity, volcanic eruptions, construction and mining (Froude & Petley, 2018). However, the action of rainfall infiltration and subsequent rise in pore pressures is the most frequent trigger (Hungr et al., 2014), and the only process considered in this section.

Infiltration and pore pressure redistribution

Pore pressure response to rainfall is a complex interaction between soil properties, topography, initial conditions and water input characteristics (Dingman, 2015). For dry, idealized soils, initial infiltration rates are high due to significant contribution from a negative vertical pressure gradient. The gradient decreases with increasing water content, and infiltration rates soon cease approximating hydraulic conductivity (Dingman, 2015). The process is commonly described by the non-linear differential Richard's equation (Richards, 1931), although approximations and simplified solutions valid for certain boundary conditions are usually necessary (e.g., Green & Ampt, 1911; Iverson, 2000). In practice, surface infiltration may vary widely from idealized equations; root growth and decay, burrowing, desiccation cracks and frost action commonly produce macropores and "preferential flow" which increases infiltration. Other factors decrease infiltration rates such as swelling or hydrophobic materials, compaction and frost (Dingman, 2015).

For saturated soils (K = $K_{saturated}$ = constant), pore pressure redistribution may be described by a simplified, linear diffusion equation (Iverson, 2000):

$$\frac{\delta \psi}{\delta t} = D_0 cos^2 \alpha \frac{\delta^2 \psi}{\delta Z^2} \tag{2.2}$$

$$D = \frac{K \delta \psi}{\delta \theta} \tag{2.3}$$

In Eq. 2.2, ψ is pressure head, t is time, α is slope angle and Z is depth. Diffusivity (D) is a material parameter proportional to the slope of a material's "soil water characteristic curve" (Fig. 2.6), of which higher values increase the speed of pressure redistribution. Its maximum value (D_0) occurs in saturated conditions and ensures linearity of Eq. 2.2. For intense precipitation, the largest and most rapid pore pressure responses occur near surface level, while low-intensity rainfall increases pore pressures at low and high depths simultaneously (Iverson, 2000).

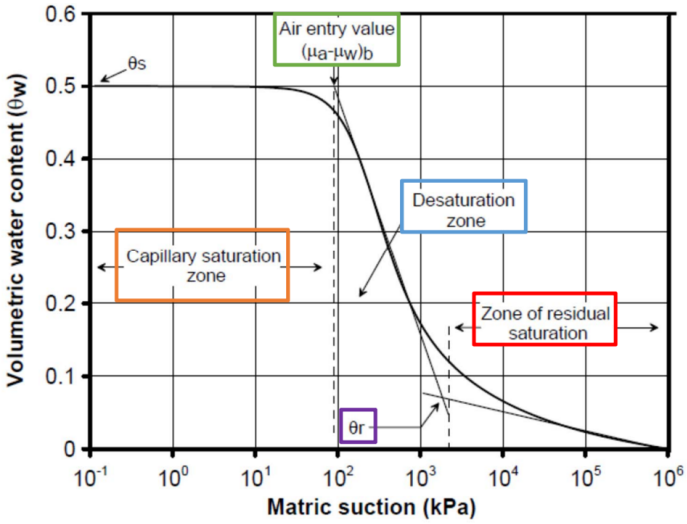


Figure 2.6: A soil water characteristic curve denotes the relationship between volumetric water content and matrix suction (negative pore pressures) in unsaturated soils, identifying zones of differing properties. Figure from Paniagua (2020).

Effective stresses and slope stability

Pore pressure (u) is the difference between total (σ) and effective (σ') stresses (Terzaghi, 1931; Eq. 2.4), and consequently, the latter is reduced through rainfall infiltration. This is equivalent of shifting the Mohr's circle of Fig. 2.5 leftwards, enabling failure if shear stresses supersede shear strength.

$$\sigma' = \sigma - u \tag{2.4}$$

A simple, commonly used model for evaluating resulting slope stability is the *infinite slope model* (Taylor, 1948). It estimates *factor of safety* (FS) as the ratio of resisting to driving forces (Eq. 2.5), where δ is slope angle, γ_w is unit weight of water, γ_s is saturated unit weight of soil and Z is depth of failure.

$$F_S = \frac{tan\phi}{tan\delta} + \frac{C - \psi\gamma_w tan\phi}{\gamma_s Z sin\delta cos\delta} \tag{2.5}$$

2.3.3 Risk and susceptibility

The presence of predisposing and possible triggering factors may call for mitigation measures, for which knowing the distribution and magnitude of *risk* is necessary. The term includes information on the probability and severity of an adverse threat (UNISDR, 2009); probability is represented by an area's *susceptibility*, while severity is a measure of likely consequences (Corominas et al., 2014).

Susceptibility, evaluated by assessing the types, frequencies and magnitudes of landslides potentially occurring in an area (Flentje et al., 2007), is static and considers inherent propensity to landsliding. As an end in itself or an initial step of risk assessment (Corominas et al., 2014), it identifies likely triggering areas and rank-orders these qualitatively or quantitatively. One may differentiate between knowledge-driven, data-driven and physically based methods, whereas choice of method depends on e.g., importance, data availability, scale of analysis, landslide types and the existence and quality of a landslide inventory (Corominas et al., 2014). Knowledge-driven methods rely on expert judgement, prior landslide activity and weighted terrain factors, but do not include extensive modelling and are often considered too subjective. *Data-driven methods* obtain, through empirical and statistical techniques such as multiple regression (e.g., Kumar et al., 2019) or neural networks (e.g., Kanungo et al., 2006), relationships between terrain factors and undesirable events. Shortcomings include lack of expert opinion, the assumption of prior failure being indicative of future failure, and the fact that effects of spatial factors may vary widely across complex terrain (Corominas et al., 2014). Rainfall thresholds in landslide early warning systems (e.g., Guzzetti et al., 2020), which magnitudes vary across different areas, may be considered a special case of data-driven assessment. Finally, physically based methods – through the application of physically based models – are described in depth throughout the next section.

2.4 Physically based models

Physically based models (henceforth, PBMs), are commonly used for assessing landslide susceptibility (Carrara et al., 2008). Unlike other methods, these explicitly consider dynamic, spatially variable factors (e.g., soil properties, precipitation) that control triggering (Crosta & Frattini, 2003), and can theoretically, by solving sets of equations representing physical processes, predict both location, magnitude and timing of landslide occurrence (Corominas et al., 2014). A "white box" approach forgoes the need of complete landslide inventories, and outputs are typically failure probabilities, factors of safety and time-development of various state-variables (Corominas et al., 2014). Commonly applied in GIS-environments, PBMs require the coupling of:

1. A hydrological model, describing groundwater conditions
2. A slope stability model, predicting the location and magnitude of failure

History, examples and challenges

Among the first PBMs applied were "LISA" (Hammond et al., 1992), "SHALSTAB" (Montgomery & Dietrich, 1994) and "SINMAP" (Pack & Goodwin, 2001). In essence, early hydrological models computed steady state groundwater conditions in equilibrium with recharge from long-term rainfall, and could therefore not take time-varying (transient) pore pressure responses into account. Slope stability calculations, by the infinite slope model (Section 2.3.2), could i.e., provide time-invariant susceptibility only. Iverson (2000) provided a necessary framework for transient analysis; various approximations of the Richards equation enabled modelling of greatly improved prediction accuracy (e.g., Crosta & Frattini, 2003), overcoming the impossibility of defining steady states representative of entire storms. This framework is also the basis for TRIGRS (Section 3.1, Baum et al., 2002).

Above models are one-dimensional, but three-dimensional models have also been developed (e.g., An et al., 2016; Simoni et al., 2008). These provides increased prediction abilities through better representation of real problems, but the non-linearity of Richard's equations and required use of numerical methods demands computational resources not necessarily viable for large scale applications (An et al., 2016).

Typical challenges for PBMs include prediction of different landslide types (Mergili et al., 2014) and uncertain performance outside calibrated case studies (Montrasio et al., 2011). Further, data acquisition over large areas is costly due to the spatial variation of soil parameters (Carrara et al., 2008), and relatively homogenous areas are required unless crude simplifications are conducted (Corominas et al., 2014). In a broader sense, as is the case for all earth-system models, an important challenge is to apply results in a manner that helps stakeholders make good decisions (Erickson et al., 2008).

Performance Evaluation

Quantifying the performance of PBMs is a difficult, but crucial task for decision-making (Corominas & Mavrouli, 2011). Evaluation may consist of verification (whether implementation is correct in terms of the conceptual frame-

work) and validation (agreement between predicted and observed results), but hereafter, the term *evaluation* is used synonymous with the latter.

Qualitative evaluation includes visualizing, comparing and classifying results (e.g., Tang et al., 2017; Al-Umar et al., 2020). The method provides an overview of spatial model performance, but does not quantify performance in a reproducible manner.

Contingency matrices and performance indexes are common quantitative evaluation methods for binary predictors in many fields (Vihinen, 2012), including landslide modelling of which output is failure/no failure. If considering landsliding a *positive* instance, each model prediction may be classified as either a true positive (TP), false positive (FP), true negative (TN) or false negative (FN), of which row-wise, column-wise and combinatory performance indexes are calculated (Fig. 2.7). Common indexes such as sensitivity (true positive rate) and specificity (true negative rate) measure different aspects of performance, and combining several indexes is therefore necessary (Vihinen, 2012). Certain indexes combine many or all contingency matrix categories, such as Matthews' correlation coefficient:

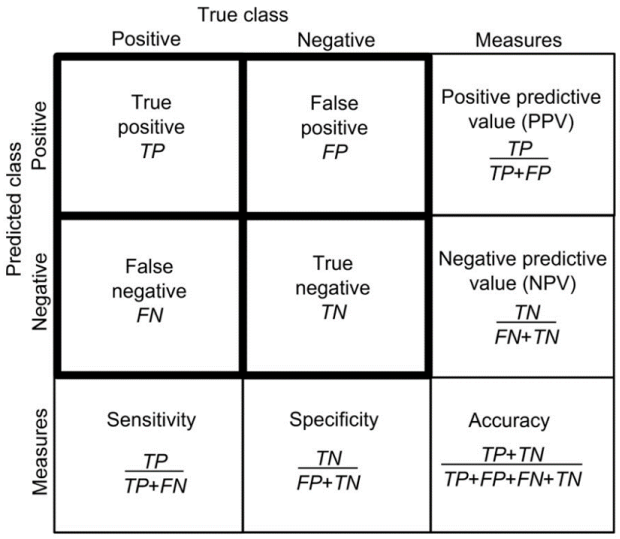


Figure 2.7: Performance indexes are calculated from the 2*2 contingency matrix. figure from Vihinen (2012).

$$\mathrm{MCC} = \frac{\mathrm{TP} \cdot \mathrm{TN} - \mathrm{FP} \cdot \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP}) \cdot (\mathrm{TP} + \mathrm{FN}) \cdot (\mathrm{TN} + \mathrm{FP}) \cdot (\mathrm{TN} + \mathrm{FN})}} \tag{2.6}$$

Receiver operating characteristics (ROC) is another common method of evaluating physically based models (e.g., Godt et al., 2008; Melchiorre & Frattini, 2012; Seefelder et al., 2017) as well as predictors in e.g., medicine, signal processing and machine learning. Providing a richer measure of performance compared to scalar indexes (Fawcett, 2006), ROC evaluates the trade-off between true and false positives across the full range of threshold values (in landslide modelling: the value of FS indicating failure/no failure (usually FS=1) is henceforth denoted T_{fs}). True positive rates and false positive rate (FPR, 1-specificity) are plotted on the x- and y-axis respectively (Fig. 2.8), while performance is plotted as either points (discrete classifiers) or curves (continuous classifiers, varying thresholds). Curves bending towards the upper left corner (or points located there) indicate good performance, while curves close to the diagonal represent classification no better than by chance (Fawcett, 2006). However, ROC-curves simply indicates a classifier's potential for rank-ordering instances, but not directly its real-world performance (Vihinen, 2012).

The *area under the curve* (AUC) is derived from ROC and considered the probability that a random positive instance will be classified as more positive than a random negative instance. Although imperfect as some classifiers may perform better at certain locations of the ROC-space only (Fawcett, 2006), it reduces curves to scalar values and is often reported and compared (e.g., Schilirò et al., 2018; Ciurleo et al., 2019).

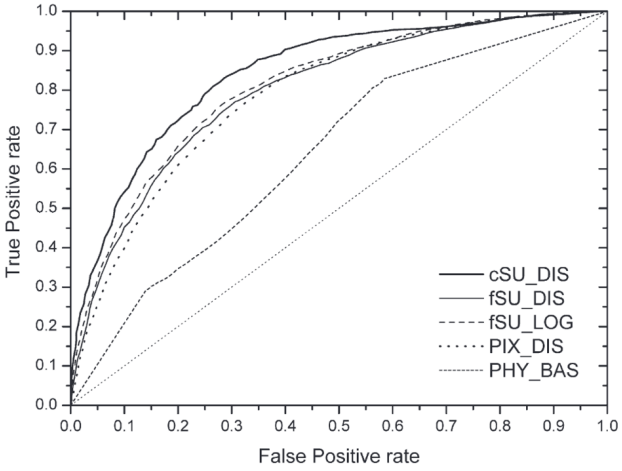


Figure 2.8: ROC-curves, of which "cSU-DIS" is generally considered the best predictor. Figure from Carrara et al. (2008).

Also derived from ROC is the *distance to perfect classification* (D2PC, Eq. 2.7), which evaluates performance at a specific threshold value. One may identify optimal threshold values by means of a minimum D2PC, and it has been used used to compare performance of distinct physically based landslide simulations (e.g., Zieher et al., 2017).

$$D2PC = \sqrt{(1 - TPR)^2 + (FPR)^2} \tag{2.7}$$

Other methods, such as the *landslide ratio of each predicted FS-class* (LR_{class}) (Park et al., 2013) are applied to varying degree, but are not described further. Two performance indexes applied in this thesis' case study, CSI and MCC, are presented in Section 4.4.1.

Chapter 3

TRIGRS: Theory, usage and literature review

3.1 Theory and usage

TRIGRS, or the Transient Rainfall Infiltration and Grid Based Slope Stability model (Baum et al., 2002, 2008), is a commonly used, open source, one-dimensional and spatially distributed model for estimating slope stability conditions in terms of factor of safety. It combines transient hydrological modelling with the infinite slope stability model, supporting both saturated and unsaturated soil conditions. For detailed descriptions of functionalities, the reader is referred to (Baum et al., 2008)

For saturated initial conditions, Iverson (2000)'s suggested approach is applied to obtain general solutions for permeable and impermeable basal layers. Pressure head is computed at various depths and times, involving both a steady and transient part. The former depends on initial water table depth, a steady infiltration rate (I_Z), saturated hydraulic conductivity (K_S) and slope angle, while for the transient part, time-varying precipitation is introduced resulting in vertical downward seepage and pressure diffusion.

For unsaturated conditions, functions describing the relationship between hydraulic conductivity, volumetric water content (θ) and matrix suction (ψ) are applied in the same framework as for the saturated solution. Estimated excess seepage volume accumulating at the base of the unsaturated zone is compared to fillable volume for estimating the water table rise. This volume is also applied as load on top of the water table, and pressure head below is computed by means of diffusion equations. In practice, the unsaturated zone smooths and delays the surface infiltration "signal" at depth, while negative pore pressures increase effective stresses.

The amount of precipitation exceeding soil infiltrability, which is variable for unsaturated but approximately equal to K_s for saturated conditions, is treated as surface runoff. This amount is transferred to adjacent cells by a simple, GIS-based runoff-routing scheme, and may infiltrate at a later stage.

The infinite slope stability model (Section 2.3.2) uses computed pressure heads to estimate factor of safety cell by cell at several depths, and the first location of $F_S < 1$ is considered the depth of failure. For the unsaturated zone, the stabilizing effect related to matrix suction is accounted for by multiplying pressure head by the effective stress

parameter (Vanapalli & Fredlund, 2000).

TRIGRS uses an initialization text file listing input data, controlling analysis options and specifying which output is saved (Table 3.1). Output is computed for chosen timesteps, and various options are available (e.g., a specific number of iterations, the potential presence of a capillary fringe or the addition of a steady background flux to prevent drying).

Table 3.1: Soil properties, spatial data and output files from TRIGRS. θ_s, θ_r and α_G are required for unsaturated analysis only. With the exception of slope angle, spatial data can assume a unique value for the entire analysis area. The production several output files must be specified by the user.

Soil properties	**Spatial data**	**Output files**
Cohesion (C)	Slope angles	Log file
Friction angle (ϕ)	Property zones	Minimum factor of safety (Fs_{min})
Unit weight of soil (γ_s)	Soil depth (Z_{max})	Depth of Fs_{min}
Diffusivity (D_0)	Initial water table depth (d)	Pore pressures at depth of Fs_{min}
Saturated hydraulic conductivity (K_s)	Background infiltration rate (I_Z)	Surface runoff
Sat. volumetric water content (θ_s)	Precipitation intensities (list)	Infiltration rate
Res. volumetric water content(θ_r)		Unsaturated zone basal flux
A_G (unsaturated parameter)		Non-convergent cells

3.2 Literature review

Aims and scope

TRIGRS is commonly utilized in physically based landslide modelling, but to my knowing, an overview of its usage does not exist. Although summaries have been provided (e.g., Park et al. 2013), these have covered only a fraction of applications. Therefore, a literature review is conducted with the following aims:

- To provide a qualitative and semi-quantitative summary of TRIGRS applications, and
- To serve as a guide for further reading - both for this thesis and for others interested in applying the model

The review examines all scientific papers concerning TRIGRS applications that have been published between 2008 (i.e. the year when the unsaturated version was introduced) and early autumn 2020. All applications are considered: from local case studies to incorporation in a large-scale warning system.

Following common practice in review articles (e.g., Segoni et al., 2018, Reichenbach et al., 2018), only English, peer-reviewed papers from journals indexed in the SCOPUS database are included. SCOPUS is ELSEVIER's (https://www.elsevier.com/) abstract and citation database, and contains a comprehensive list of independently reviewed journals (SCOPUS, 2019) whose prestige is evaluated by the Scimago Journal and Country Rank (SJR, Guerrero-Bote and Moya-Anegón, 2012). Consequently, non-indexed journals, book sections , discussions, conference proceedings, technical notes and publications part of the "grey literature" are neglected.

Information retrieval and storage

Relevant papers are registered in an excel workbook, where specific fields of information are grouped in four categories:

- *Basic information* considers identification and context, and includes year of publishing, author(s), paper title and name of journal.
- *Study area* describes where and under what conditions applications were performed. Study area sizes were registered, but climatic and geological conditions were not due to considerable variation in methods of reporting. The type(s) of investigated landslides was also indicated, together with potential considerations of snowmelt and vegetation effects.
- *Input data* includes information on type and quality of various model requirements. Landslide data was registered in terms of point features, polygon features or a combination of both, while methods for estimating soil depth were noted. DEM-resolutions were registered, along with sources of rainfall data (e.g., rain gauges, remote sensing). Estimations of initial hydraulic soil conditions were registered in terms of engineering judgement, monitoring, modelling etc., while choices of geotechnical and hydraulic parameters were registered in terms of method (e.g., field tests or literature) and spatial heterogeneity (e.g., uniform values or property zones).
- *Evaluation* considers if and how evaluation was performed, and, to some extent, the obtained results. Considering that several methods have been used simultaneously, only those with most complexity or significance were registered (simpler methods were therefore disregarded). Registration of results was difficult; only reported AUC-values and short notes describing authors' subjective opinions were stored.

Table 3.2: Papers included in this thesis' literature review.

Paper	Title
Godt et al., 2008	Transient deterministic shallow landslide modeling: Requirements for..
Kim et al., 2010	Predicting the rainfall-triggered landslides in a forested mountain region..
Sorbino et al., 2010	Susceptibility analysis of shallow landslides source areas..
Liao et al., 2011	Evaluation of TRIGRS (transient rainfall infiltration and grid-based..
Montrasio et al., 2011	Towards a real-time susceptibility assessment of rainfall-induced..
Kim et al., 2013	Modeling the contribution of trees to shallow landslide development..
Park et al., 2013	Landslide and debris flow susceptibility zonation using TRIGRS for the..
Grelle et al., 2014	Space-time prediction of rainfall-induced shallow landslides through..
Raia et al., 2014 (USA)	Improving predictive power of physically based rainfall-induced..
Raia et al., 2014 (Italy)	Improving predictive power of physically based rainfall-induced..
Alvioli et al., 2014	Scaling properties of rainfall induced landslides predicted by a..
Bordoni et al., 2015	Site-specific to local-scale shallow landslides triggering zones..
Saadatkhah et al., 2015	Hulu Kelang, Malaysia regional mapping of rainfall-induced landslides..
Chien et al., 2015	Warning model for shallow landslides induced by extreme rainfall
Schilirò et al., 2015	Evaluation of shallow landslide-triggering scenarios through..
Alvioli and Baum, 2016	Parallelization of the TRIGRS model for rainfall-induced..
Lee and Park, 2016	Assessment of shallow landslide susceptibility using the transient..
Seefelder et al., 2017	Does parameterization influence the performance of slope stability..
Saadatkhah et al., 2016	Regional modeling of rainfall-induced landslides using TRIGRS..
Peres and Cancelliere, 2016	Estimating return period of landslide triggering by Monte Carlo..
Gioia et al., 2016	Application of a process-based shallow landslide hazard model over..
Viet et al., 2017	Effect of Digital Elevation Model Resolution on Shallow Landslide..
Zhuang et al., 2017	Prediction of rainfall-induced shallow landslides in the Loess..
Tang et al., 2017	Dynamic assessment of rainfall-induced shallow landslide hazard
Wu et al., 2017	A probabilistic model for evaluating the reliability of rainfall thresholds..
Tran et al., 2017	Comparing the performance of TRIGRS and TiVaSS in spatial and..
Stancanelli et al., 2017	A combined triggering-propagation modeling approach for the..
Zieher et al. (2017)	Sensitivity analysis and calibration of a dynamic physically based..
Tran et al., 2018	Three-dimensional, time-dependent modeling of rainfall-induced..
Vieira et al., 2018	Assessing shallow landslide hazards using the TRIGRS and SHALSTAB..
Weidner et al., 2018	Regional-scale back-analysis using TRIGRS: an approach to..
Alvioli et al., 2018	Implications of climate change on landslide hazard in Central
Schilirò et al., 2018	Shallow landslide initiation on terraced slopes: Inferences from..
Ciurleo et al., 2019	Landslide susceptibility assessment by TRIGRS in a frequently..
Hsu and Liu, 2019	Combining TRIGRS and DEBRIS-2D models for the simulation of a rainfall..
Dikshit et al., 2019	Estimation of Rainfall-Induced Landslides Using the TRIGRS Model
Chen et al., 2019	Precipitation data and their uncertainty as input for..
Hidayat et al., 2019	Development of a landslide early warning system in Indonesia
Marin and Velásquez, 2020	Influence of hydraulic properties on physically modelling slope..
Al-Umar et al., 2020	GIS-based modeling of snowmelt-induced landslide susceptibility of..
Sarma et al., 2020	Influence of digital elevation models on the simulation of..
Marin and Mattos, 2020	Physically-based landslide susceptibility analysis using Monte..

Results: basic information

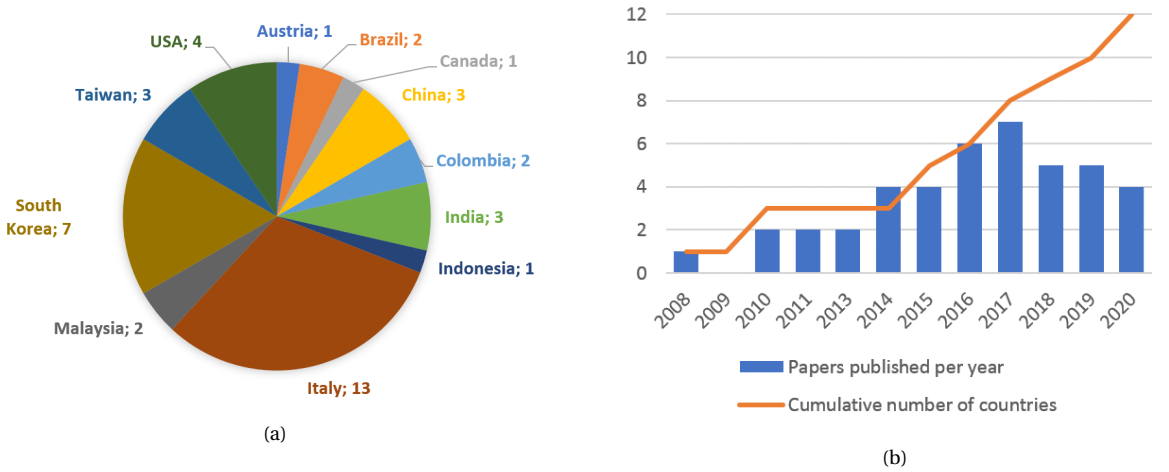
A total of 42 papers are identified (Table 3.2), published in 23 different scientific journals (Table 3.3). Journals represent various sub-fields within "Earth and Planetary Sciences" and "Environmental Sciences". Papers are published every year except for 2009, with the maximum (7 papers) in 2017 (Fig. 3.2b). An increasing activity is observed; in fact, 50% of papers are published in the last four years (neglecting papers published during late 2020).

TRIGRS is applied in a number of locations (Fig. 3.1). Studies are reported in all inhabited continents except Africa, while clusters are found in Italy (13 studies) and South Korea (7 studies) with USA, Taiwan, China and India as the following most represented countries (Fig. 3.2a). Fig. 3.2b also shows the cumulative number of countries applying TRIGRS throughout the review period. After a longer period of USA, Italy and South Korea only, a large addition of new countries has occurred from 2015 onwards.

Table 3.3: Journal titles, corresponding Scientific Journal Scores (SJR) and the number of papers associated with each journal.

Journal	SJR (2019)	Number of papers
Landslides	1.88	6
Natural Hazards and Earth System Sciences	1.005	5
Natural Hazards	0.814	3
Environmental Earth Sciences	0.604	3
Geoscientific Model Development	3.18	2
Engineering Geology	2.065	2
Journal of Hydrology	1.684	2
Geomorphology	1.384	2
Water (Switzerland)	0.657	2
Journal of Mountain Science	0.462	2
Earth Systems and Environment	.	1
Geoenvironmental Disasters	.	1
Environmental Modelling and Software	1.896	1
Georisk	1.822	1
Science of Total Environment	1.661	1
Earth Surface Processes and Landforms	1.3	1
Ecological Engineering	1.122	1
Geomatics, Natural Hazards and Risk	0.814	1
Bulletin of Engineering Geology and the Environment	0.77	1
Natural Hazards Review	0.643	1
Frontiers of Earth Science	0.479	1
Geosciences (Switzerland)	0.474	1
Arabian Journal of Geosciences	0.404	1

Figure 3.1: Global spatial distribution of TRIGRS applications.

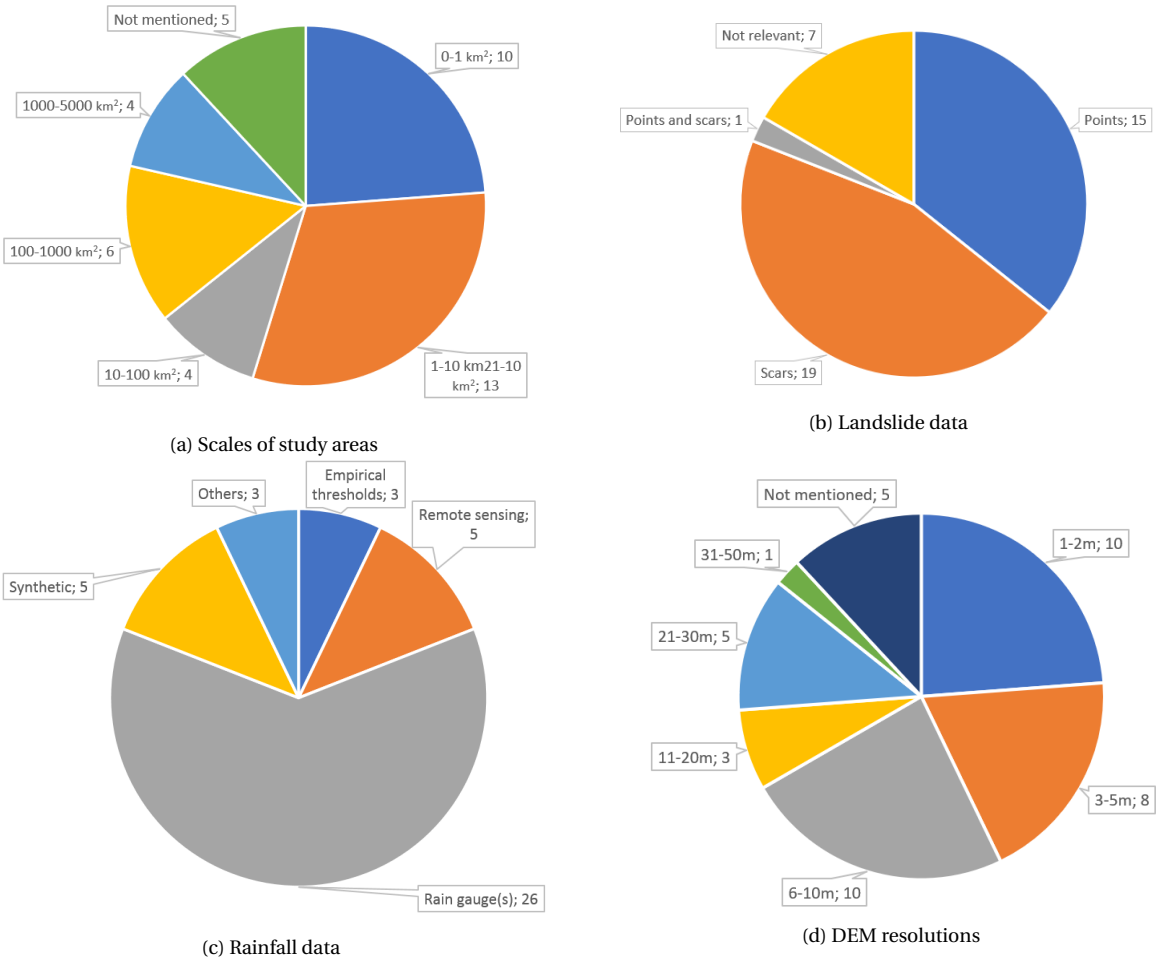


(a) (b)

Figure 3.2: (a) Number of papers published for each country. (b) Temporal distribution of applications, along with the cumulative number of identified countries.

Results: study areas and input data

Study area extents range from the 0.014 km^2 Bongwha catchment in Korea (Kim et al., 2013) to the 4098 km^2 Upper Tiber River Basin of Central Italy (Alvioli et al., 2014; Alvioli et al., 2018). While median area size is 5 km^2, the average size 362 km^2 due to large-scale outliers. The most common study area extent within 1–10 km^2 (Fig. 3.3a), while areas at slope scale are not represented except from papers focused in ground water monitoring (Bordoni et al., 2015). For 10 papers, areal extent was not specifically mentioned but could be estimated from scaled figures. Unfortunately, for 4 papers areal extent was neither mentioned nor estimable. One paper considered a range of undescribed area extents due to TRIGRS being incorporated in a large-scale landslide early warning system (Hidayat et al., 2019).



(a) Scales of study areas

(b) Landslide data

(c) Rainfall data

(d) DEM resolutions

Figure 3.3: Identified spatial data for TRIGRS applications.

All papers consider prediction of *shallow landslides* (Section 2.3), either directly or through empirical correlations between unstable cells and expected frequency of debris flows (e.g., Stancanelli et al., 2017). In fact, 9 papers evaluate debris flow susceptibility only, assuming triggering by pore pressure increase. Landslide data is typically compiled from field studies, remote sensing or a combinations of both (e.g., Park et al., 2013; Zieher et al., 2017). Most papers (19) utilize landslide scars (source area polygons) for comparative evaluations (Fig. 3.3b: among these, one paper estimates scars as the upper third of total paths (Seefelder et al., 2017). Several papers (15) instead rely on landslide point data, while one paper (Montrasio et al., 2011) uses buffers around points as an approximation of scars. For a number of papers (7) landslide data is irrelevant, as issues such as climate change implications on slope stability (Alvioli et al., 2018) and model parameterization (e.g., Marin and Velásquez, 2020) are explored.

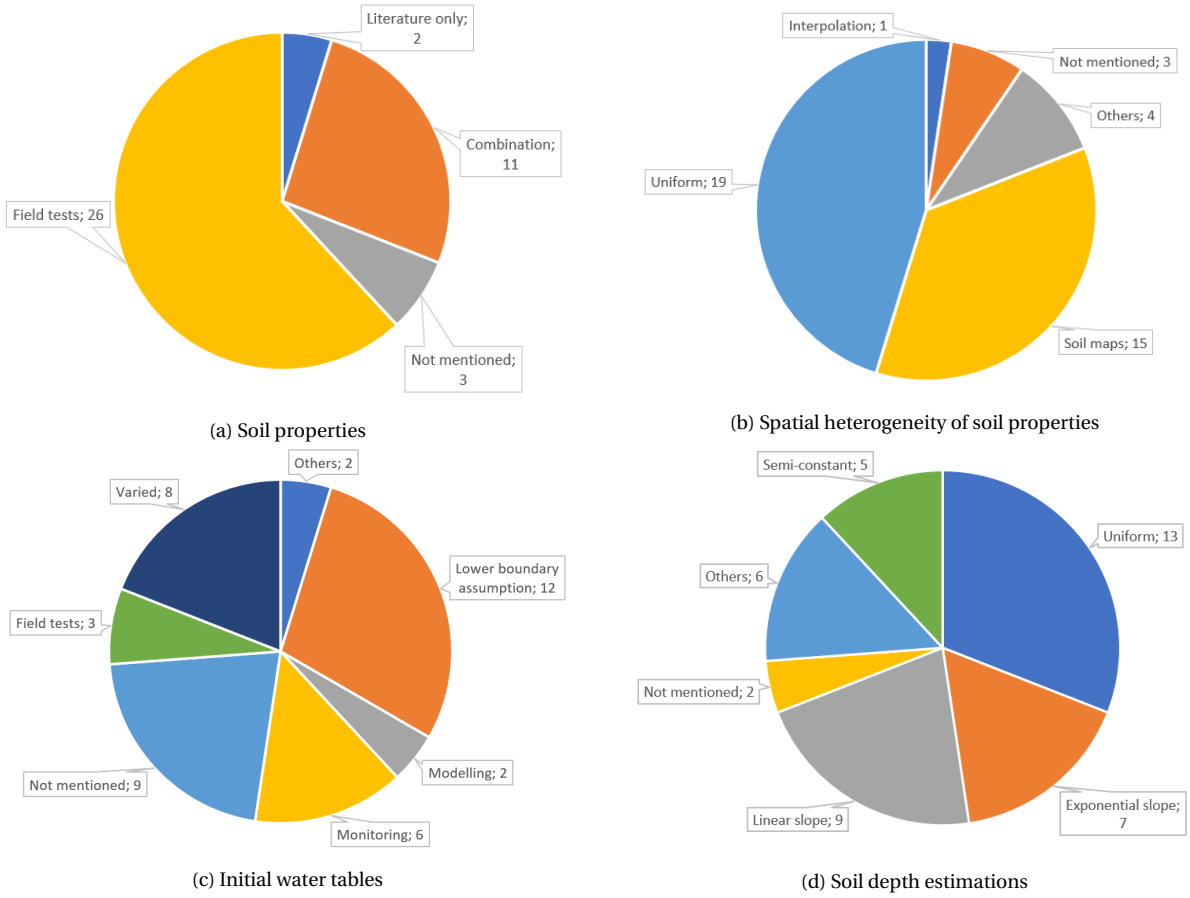
Rainfall data is gathered from various sources (Fig. 3.3c), but the majority of papers (26) rely on one or more rain gauges. Single gauges are used in the absence of other rainfall data or if considered representable for the entire study area, but data from several gauges is also combined through some interpolation technique (e.g., Tran et al., 2017; Zieher et al., 2017; Alvioli et al., 2018). Data from remote sensing (typically radar) is applied in 5 papers, although commonly verified by rain gauge measurements. Finally, rainfall data is not always employed; empirical

rainfall thresholds known to induce slope failure are used to e.g., evaluate a probabilistic version of TRIGRS (Raia et al., 2014), while "synthetic" (hypothetical) rainfall scenarios are used to investigate area response to a range of situations (e.g., Peres and Cancelliere, 2016).

Different digital elevation model (DEM)-resolutions are applied (Fig. 3.3d), of which the most used area 10m*10m (9 papers) and 2m*2m (7 papers). Maximum reported resolution is 1m*1m (Tran et al., 2018), while the lowest is 40m*40m (Chien et al., 2015). 3 papers consider different resolutions to evaluate corresponding effects on prediction accuracy, while two papers leave resolution unmentioned.

Fig. 3.4a shows the most common methods employed for estimating geotechnical and hydraulic parameters. A majority (26 papers) rely on field tests, 2 papers use literature values only while 11 papers utilize a combination of field tests, literature and calibration. Representativeness of tests is variable, but the number of field tests or test types were not registered. Unfortunately, 3 papers do not mention how parameters were obtained. Note that a number of papers placed in the “field tests”-category for soil properties used literature values for hydraulic diffusivity (D_0) and background infiltration rate (I_z).

Most papers 19 papers apply homogenous soil properties, while the most common method for defining soil property zones relies on soil maps (15 papers) (Fig. 3.4b). Map accuracy and resolution is seldom addressed, but typical resolutions are 1:10 000 (Montrasio et al., 2011) and 1:100 000 (Saadatkhah et al., 2016). The “others”-category includes differentiation based on vegetation-land use (Schilirò et al., 2018) or more or less loosely defined areas (e.g., Hidayat et al., 2019), while only one paper makes use of interpolation techniques (Lee & Park, 2016).



(a) Soil properties

(b) Spatial heterogeneity of soil properties

(c) Initial water tables

(d) Soil depth estimations

Figure 3.4: Identified methods for obtaining geoetechnical and hydraulic data for TRIGRS applications.

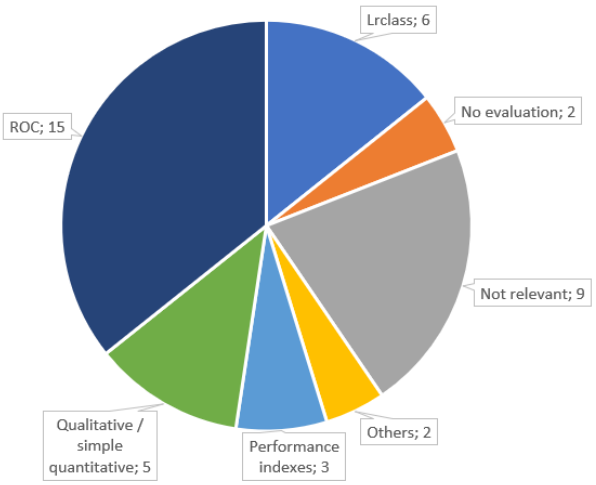
Different approaches are identified for estimating initial hydrological conditions (Fig. 3.4c). A number of papers (12) assume the water table is located at the soil's lower boundary; either due to distinctly dry conditions or a lack of data. Field tests (3 papers) - or in more extended form, field monitoring - (6 papers) provide increased detail. Ground water levels are also mathematically modelled (3 papers), varied through sensitivity analyses or calibration procedures (8 papers) or not mentioned at all (9 papers).

Approaches for estimating soil depth range from uniform values to identification of complex relationships with topographical factors (Fig. 5.2). The former (18 papers) is generally the case when required level of detail is low or uniform/almost uniform soil depth can reasonably be assumed (e.g., Sorbino et al., 2010; Ciurleo et al., 2019; Park et al., 2013). On the other hand, 21 papers model spatially distributed soil depth, while among these, 16 papers utilize either a linear or exponential relationship to slope angle. The "others"-category of Fig. 5.2 includes 5 papers utilizing additional explanatory variables such as contributing area (Alvioli & Baum, 2016), curvature (Tang et al., 2017) and relative position on slope (Schilirò et al., 2018). Also, it is worth noting that Kim et al. (2013) perform interpolation based 109 individual tests, although their study area was barely 100m across.

Vegetation effects are rarely considered. Only 5 papers apply root cohesion and tree surcharge (e.g. Park et al., 2013), while generally, the issue is not discussed at all.

Results: evaluation

Main identified evaluation methods are shown in Fig. 3.5a. "Qualitative/simple quantitative" methods are used in 5 papers, and include e.g., purely qualitative map comparisons, simple reporting of FS-values, and percentages of cells unstable at various times. The most common approach is ROC, both as curves (e.g., Zieher et al., 2017) and one or more point-values (e.g., Lee and Park, 2016). A subset of papers evaluated performance in terms of AUC, whose values are reported in Fig. 3.5b). Since its introduction (Park et al., 2013), LR_{class} has been used by 6 papers, although interestingly, only in Asia. A few papers (3) use various other performance indexes. Finally, evaluation is neither performed nor desirable for 9 papers, while for 2 papers only the former is true.



(a)

Paper	Highest reported AUC
Montrasio et al. (2011)	0.754
Raia et al. (2014) (USA)	0.73
Raia et al. (2014) (Italy)	0.65
Schiliro et al. (2015)	0.795
Seefelder et al. (2017)	0.737
Zhuang et al. (2017)	0.787
Zieher et al. (2017)	0.878
Schiliro et al. (2018)	0.748
Ciurleo et al. (2018)	0.863

(b)

Figure 3.5: (a) Main identified evaluation methods. (b) Maximum reported AUC-values.

Chapter 4

Case study: Jølster, Western Norway

4.1 The case study

4.1.1 Jølster, Western Norway

Jølster of Indre Sunnfjord municipality is located in Vestland County, western Norway (Fig. 4.1b). About 3000 inhabitants (SSB, 2016) live relatively spread across rural 620 km^2, with main villages being Skei, Vassenden and Langhaugane. Main sectors are agriculture, tourism, public administrations and service, while limited industry is present except hydropower (Thorsnæs & Askheim, 2020).

This thesis' study area (Fig. 4.1A, 235.356 km^2) is centered around Vassenden at the western shore of lake Jølstravatnet. It covers the area subject to landsliding during the case event (Section 4.2.1) and precipitation analysis by Sandvoll (2020).

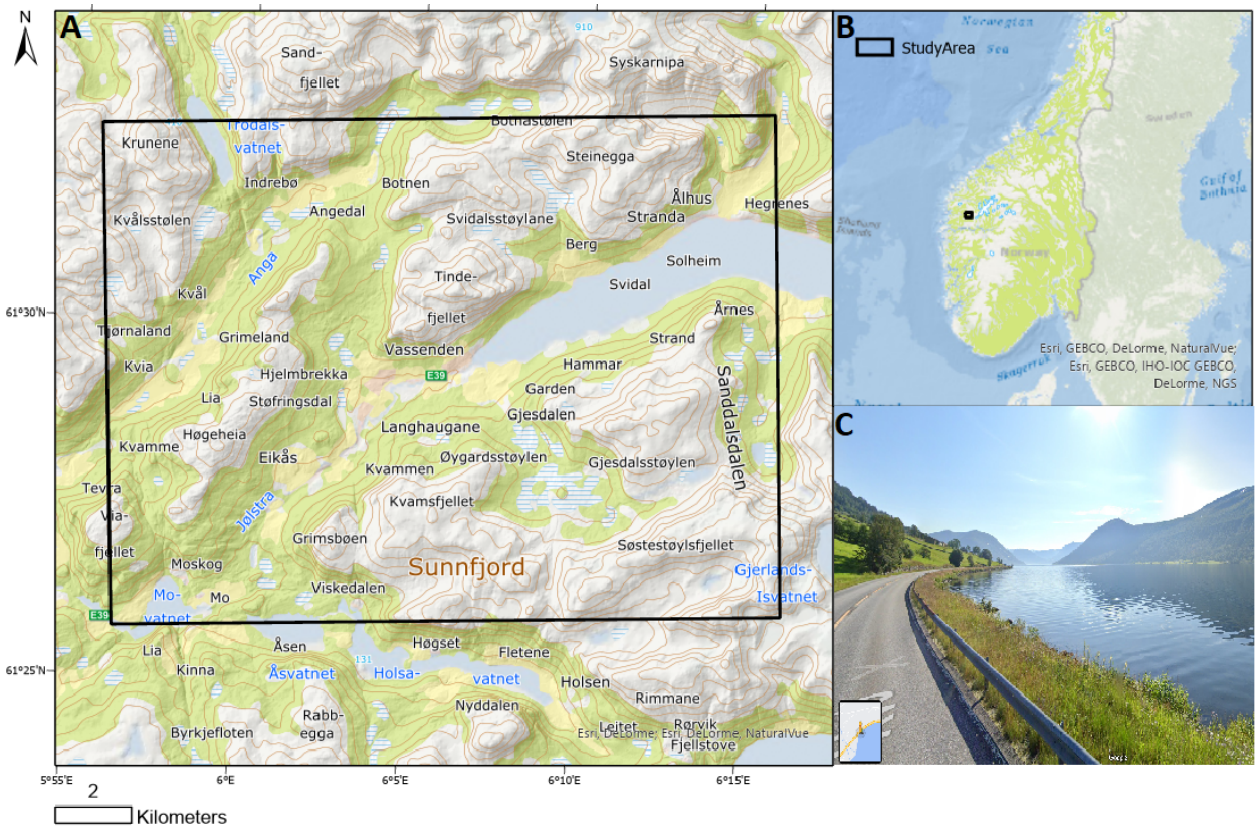
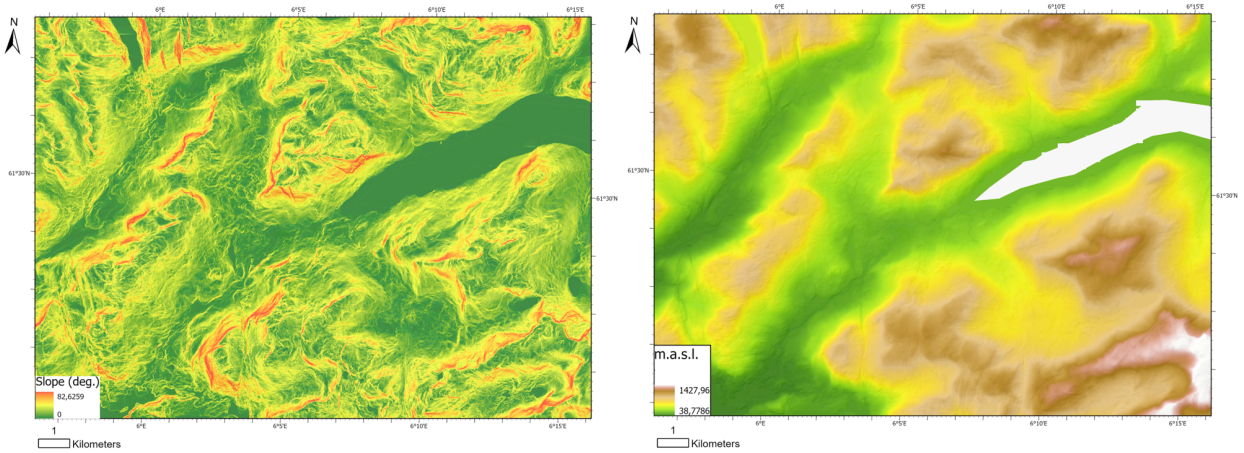


Figure 4.1: (A) The study area, centered around the western shore of lake Jølstravatnet. (B) Jølster's location in western Norway. (C) Eastwards view from the northern shore of lake Jølstravatnet. From Google Street View (https://www.google.no/maps/).



(a) (b)

Figure 4.2: The study area's topography: (a) slope angles (1m*1m DEM) and (b) elevation (1*1m DEM).

Jølstravatnet (206.1–207.5 m.a.s.l.) and the river Jølstra are facilitated by a main, east–west, glacially eroded U-valley. To the north is another significant northeast—-southwest valley, while numerous, typically north–south sub-valleys are also present. Steep slopes (Figs. 4.1C 4.2a) rise from the lake in alpine fashion, reaching peaks of more than 1400 m.a.s.l. (Fig. 4.2b). Glacial erosion has created bands of steep cliffs on west-facing slopes, while

topography is more rounded in the opposite direction.

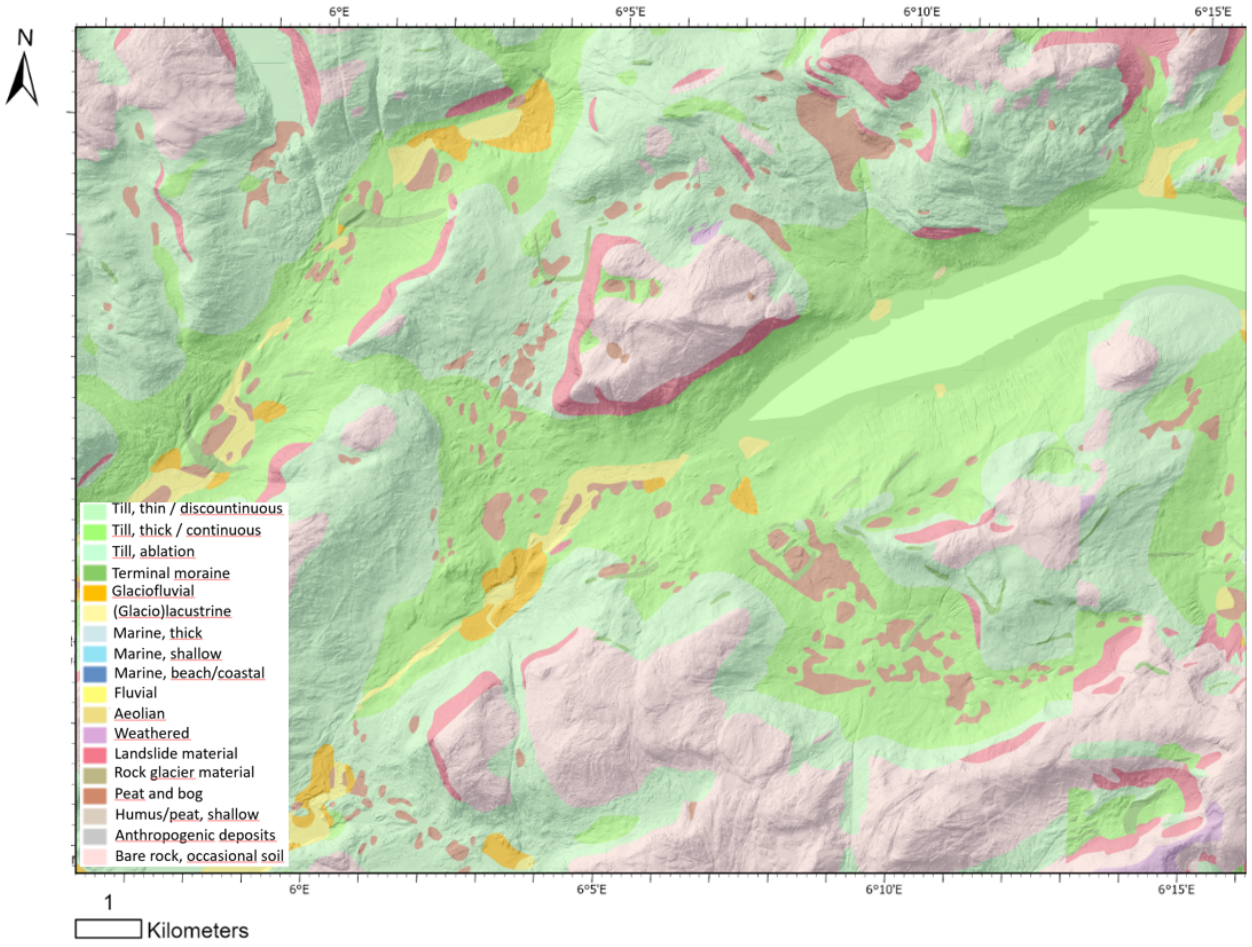


Figure 4.3: Soil types for the study area. Soil maps are described in Section 4.2.3.

Quaternary geology is dominated by glacial basal till of decreasing depth and continuity with altitude (Fig. 4.3). Landslide material is deposited proximal to cliffs, while little or no soil is found above 600-800 m.a.s.l. Limited zones of other types of terminal and ablation till, peat, bog, fluvial and marine sediments are also present. Bedrock mainly consists of gneiss, with sparse regions of quartz-monzonite / monzonite, meta-sandstone and amphibolite. Climate is maritime, although affected by topography. Precipitation, which is typically above 2000 mm annually (Hefre et al., 2019), decreases eastwards and shows considerable seasonal variation (Fig. 4.4).

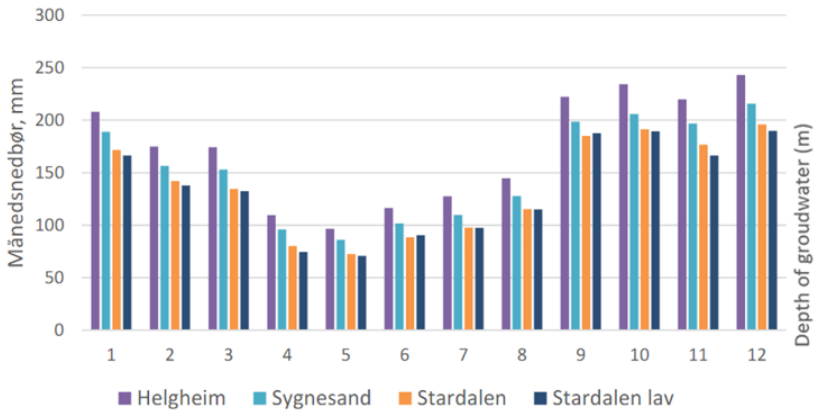


Figure 4.4: Average precipitation for rain gauges in / close to Jølster. Figure from Hefre et al. (2019.

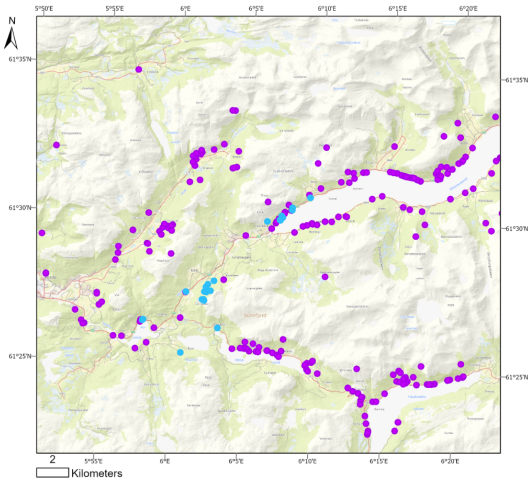


Figure 4.5: Historical (purple) and case event (blue) landslides registered in the national landslide database (Section 4.2.1).

Historically, Jølster has been severely prone to landsliding (Fig. 4.5). Detailed hazard mapping has been performed in certain prioritized Norwegian areas (NVE, 2011), and Hefre et al. (2019) mapped parts Jølster not long before the case event. Debris flows and slush flows were identified as dominating landslide types, while avalanches and rockfalls also posed risk. Additionally, after the event, Sandersen and Høydal (2019) performed supplementary mapping and gave suggestions for mitigation measures for one affected area.

4.1.2 The event: July 30, 2019

July of 2019 was unusually dry (Fig. 4.6). On Monday July 29, an east-moving low-pressure front mixed with warm air and created cumulonimbus-clouds through convection (Sandvoll, 2020). Weather forecasts for the following day indicated 50–90mm total rainfall and peak intensities of 15-20mm/h (Sandvoll, 2020), and corresponding warning level issued from MET (The Norwegian Meteorological Institute) and NVE (The Norwegian Water Resources and Energy Directorate) read “yellow” or “severe thunderstorms with relatively low consequences” (Agersten et al., 2019). No landslide warning was issued due to dry initial conditions and relatively unextreme forecasts.

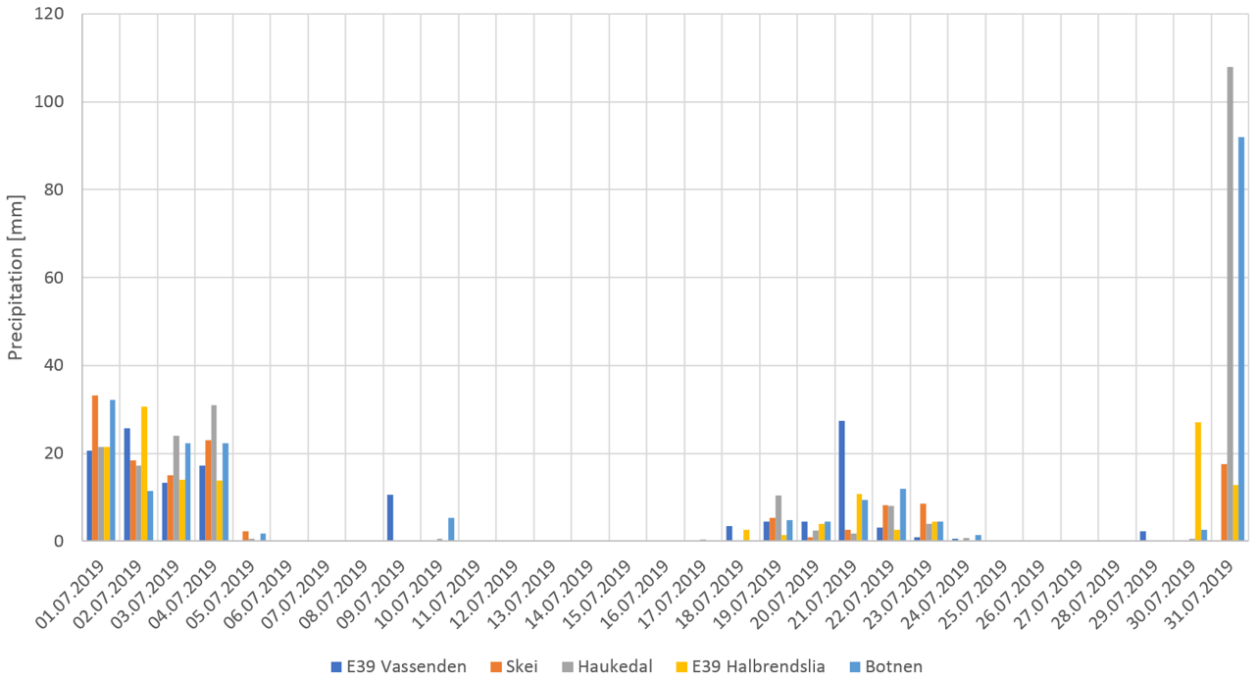
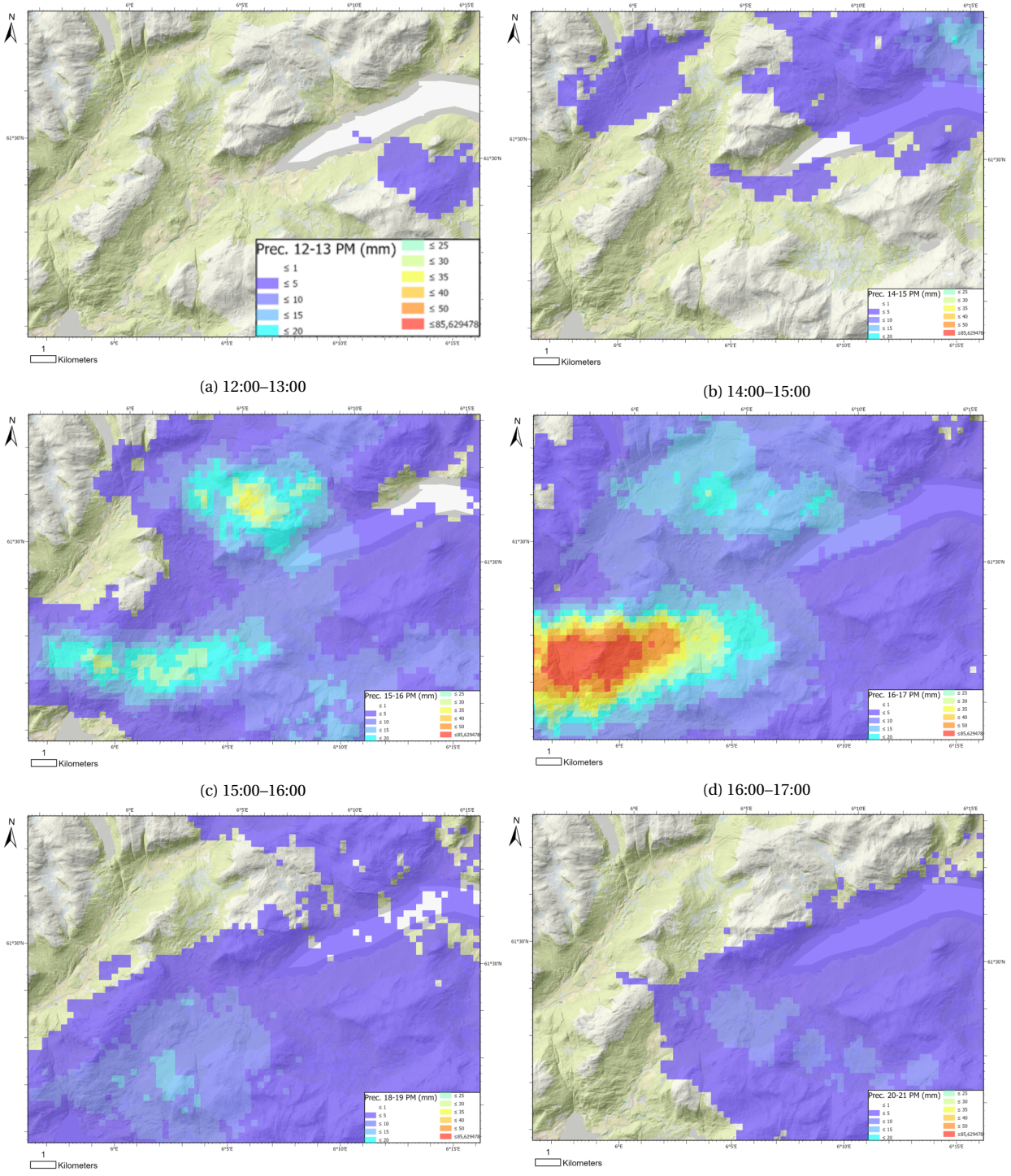


Figure 4.6: Daily precipitation for July, 2019 from rain gauges within or close to the study area (MET, 2021). Note that the closest rain gauge, "E39 Vassenden", did not record at the day of the event due to technical issues.

An additional low-pressure front joined the mix (Agersten et al., 2019), and the combination of two high- and low-pressure fronts led to a stationary weather situation with unexpectedly high levels of precipitation (Sandvoll, 2020). Locality of rainfall is reflected in (Fig. 4.6).



(a) 12:00–13:00

(b) 14:00–15:00

(c) 15:00–16:00

(d) 16:00–17:00

(e) 18:00–19::00

(f) 20:00–21:00

Figure 4.7: Precipitation radar estimates for July 30, 2019, local time. The legend, enlarged in (a), is similar for all sub-figures. Radar data is described in Section 4.2.2.

First rainfall occurred at 12 AM in eastern parts of the area (Fig. 4.7a). Although precipitation intensities and storm extent increased during the following hours (Fig. 4.7b), authorities evaluated the situation as safe based on prior experience (Sandvoll, 2020). Intensities increased at about 3 PM as the storm moved westwards, while between 15 and 16 PM, 20–30 mm/h poured down at central parts of the study area (Fig. 4.7c). Then, towards 5PM, extreme intensities of up to 85 mm/h were in central and western parts (Fig. 4.7).

Extreme precipitation resulted in numerous landslides, first reported at 16:26. Roads on both lake sides closed, power supply and communication lines were cut and evacuation began. County road 5690 (country road 451 earlier), at the south side of the lake, was opened for traffic after 7 PM due to calming decreasing intensities (Figs. 4.7e, f). However, well after maximum precipitation, two additional landslides, of which one had fatal consequences, hit the road at approximately 8:45 PM.

In total, 120 landslides initiated within the study area, 150 people were evacuated, one person lost its life and severe damage occurred on private property and infrastructure. The scale of consequences was due to limited warning and the sheer number of landslides (Rouault, 2020), and the event is regarded among Norway's most severe shallow landslide catastrophes.

4.2 Input data

4.2.1 Landslides

The compiled landslide inventory of July 30, 2019 is unusually rich, originating from several sources. Following is a description of data sources, selection of relevant landslides and method of preparation.

Data sources

A national landslide database (henceforth, NLDB) is available in Norway (https://temakart.nve.no/tema/SkredHendelser), including more than 3000 landslides of various subtypes. Although duplicates, spatial bias towards road networks and inaccuracies in classification, location and time of occurrence are present (Rouault, 2020), it provides an overview of landslide activity and is used in e.g., hazard mapping. NLDB lists 16 landslides for the case event (Fig. 4.8), typically in close proximity to people and thereby reported in the news. Landslide polygons represent full landslide areas (sources, tracks and deposits), while types are registered as either “flomskred” or “jordskred”. Of importance are initiation times with accuracies of ±15–30 minutes, but other information is limited.

Landslides are also mapped as part of a PhD-project (Rouault, 2020; Lindsay et al., in prep.). "Change in Normalized Difference Vegetation Index” (δNDVI) from satellite Sentinel-2 was utilized, and 104 *additional* landslides (full landslide scars) where identified and classified as either debris flows, debris avalanches or debris slides. Reliability was verified by GPS, helicopter, drones and field surveys, and the method was found valuable in reducing shortcomings of landslide databases (Lindsay et al., in prep.). As opposed to the NLDB, the δNDVI dataset contains small, remote soil slips barely visible one year after the event, which would typically go unnoticed with conventional mapping. However, the dataset lacks information on e.g., occurrence times, initiation mechanisms and geology.

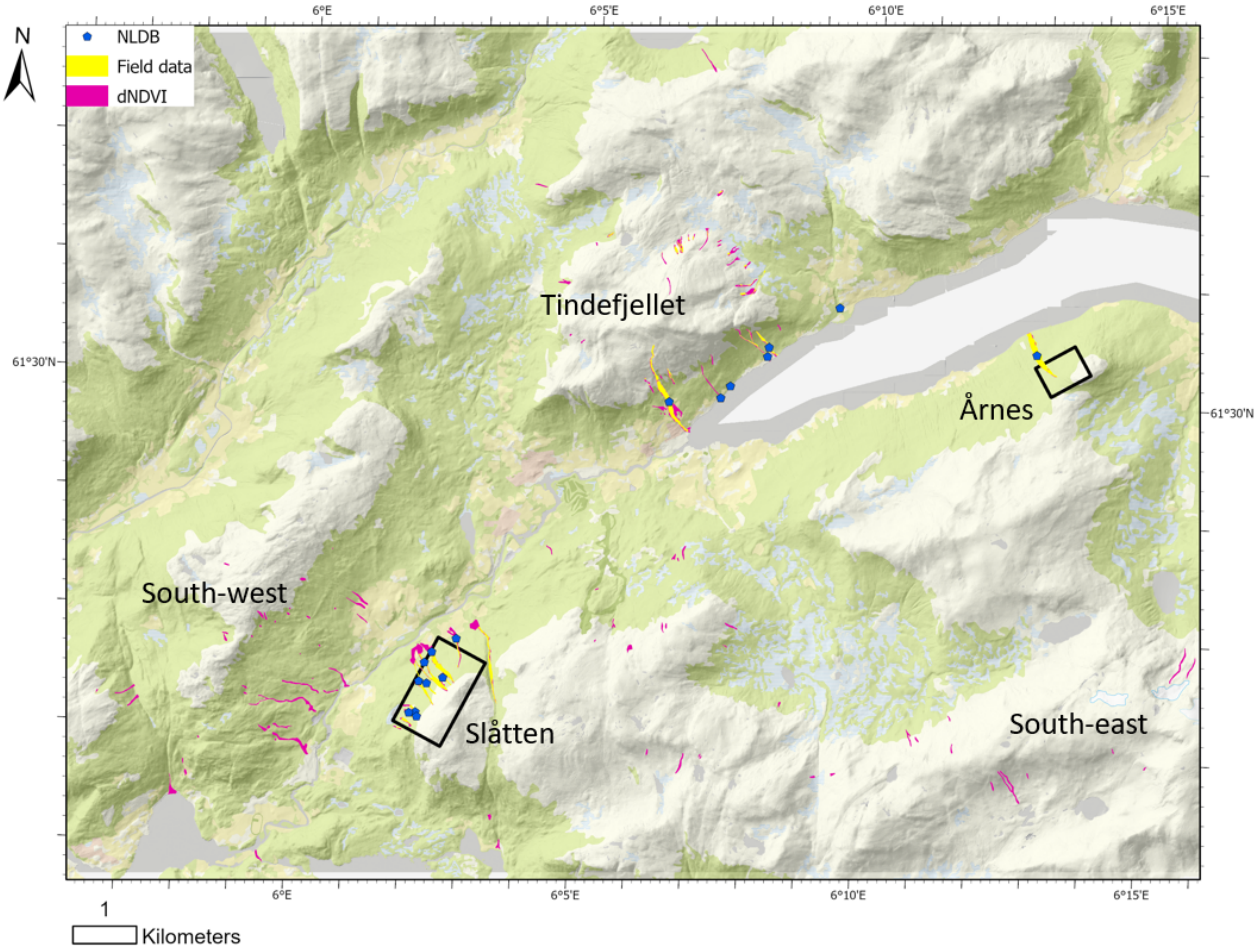


Figure 4.8: All identified landslides for July 30, 2019. Note that considerable overlap between datasets is present.

A subset (51) of landslides already mentioned have been subject to detailed field investigations (Rüther et al., in prep.). These (henceforth, field data) provide increased spatial accuracy by delineating source areas, tracks and deposits and classifying landslides in terms of "jordskred" and "flomskred", and include qualitative information on geological conditions and vegetation, as well as selected source area geometries, slope angles and soil depths. Unfortunately, geotechnical and geological parameters are not accounted for.

In summary, a total of 120 landslides of high spatial accuracy are identified for an extensive study area. Some include spatial data only, while others have been subject to field studies and are delineated in source area, track and deposit. Time of occurrence is largely unknown, except for landslides registered in the NLDB.

Data preparation

TRIGRS considers source areas initiated by pore pressure increase only. To utilize landslide data in calibration and performance evaluation, the following preparation was performed:

1. Selection of relevant landslides (29 landslides excluded from δNDVI and field data)
2. Manual delineation of source areas (δNDVI)

3. Preparation of failure grids

The δNDVI dataset contains various landslide types, classified based on whether propagation continued for a short or long distance, within or outside preexisting channels (Erin Lindsay, personal communication). Therefore, landslides described as “river erosion” were excluded, along with "debris flow" source areas located within distinct rivers. Also excluded were landslides overlapping field data, as the latter was considered more reliable.

Similarly, field data includes various landslide types (Denise Rüther, personal communication). Excluded were debris floods (recognized by no specific source area) and landslides which source areas were described as “eroded", "washed away" etc. A number of landslide tracks follow streams and were described as having “transformed from “jordskred” into “flomskred”, but these were generally included.

Further, for the δNDVI dataset, manual delineation of source areas (Fig. 4.9) was performed under the assumption that source area extents were similar in size and geometry to those identified by Rüther et al. (in prep.). The latter generally showed quadratic or rectangular (longer than wide) shapes, with typical lengths of 10-30m.

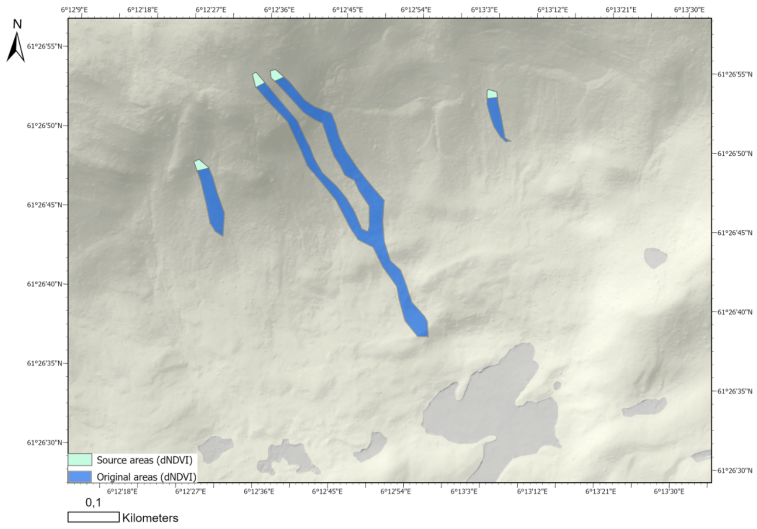
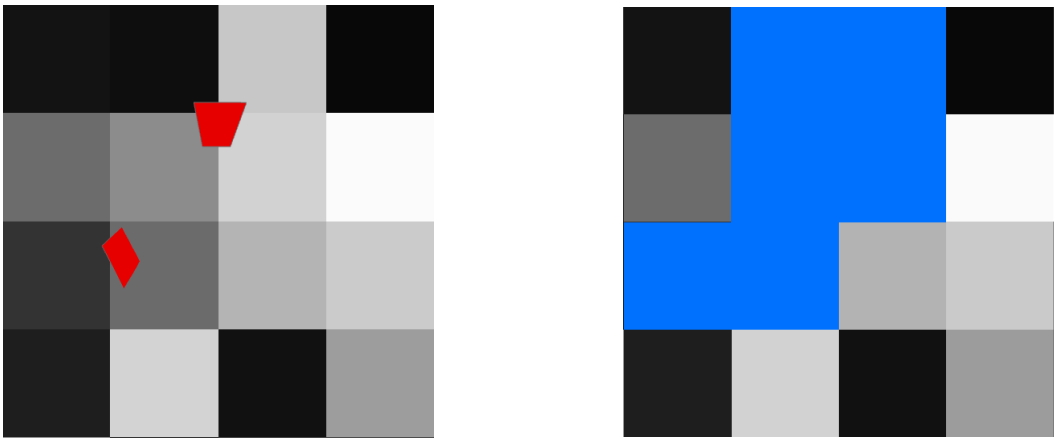


Figure 4.9: An example of manual source area delineation for the δNDVI dataset.

The definition of *correct predictions* was necessary to prepare failure grids (rasterized maps of source areas). As mismatches in resolution between model output and source areas are unavoidable, these were, following the approach of Zieher et al., (2017), defined as if "any part of a source area intersected a model cell of FS<1". Consequently, due to requirements of ArcGIS Pro’s "polygon to raster"-tool, source area polygons were expanded in order to overlap raster cell midpoints (Fig. 4.10).



(a) (b)

Figure 4.10: The definition of of correct predictions in 50*50m resolution: (a) original source areas and (b) the resulting failure grid. Source areas increase in size; the uppermost area, originally ca. 20m across, is effectively treated as 100*100m.

Finally, to provide an overview of landslide characteristics, a simple statistical analysis was performed on the 91 source areas identified as relevant.

4.2.2 Precipitation and hydraulic conditions

Precipitation

Precipitation estimates from weather radar (mm/5min, 250 meters spatial resolution) were originally prepared for Sandvoll (2020) by Christoffer Elo at The Norwegian Meteorological Institute (MET). In that thesis the dataset was, through comparisons to nearby rain gauges, landslide locations and occurrence times, as well as interviews with local residents, considered superior to other data sources in capturing local intensities and is therefore also used in this thesis.

Intensities were summarized over each storm storm hour, resulting in 14 separate data files ranging from 12:00 to 02:00 the day after the event (Table 4.1). For compatibility, a factor of 2.778e-7 was applied to convert intensities from mm/h to m/s. Additional daily rain gauge measurements, used for evaluating antecedent conditions, were downloaded from "SeKlima" (MET, 2021).

Table 4.1: Precipitation periods and hourly intensities for sub-areas Årnes and Slåtten. Note the different storm behaviours.

Period #	Time (UTC)	Time (local)	End time (s)	mm/h: Årnes	mm/h: Slåtten
1	10-11	12-13	3600	1.0	0
2	11-12	13-14	7200	0.3	0
3	12-13	14-15	10800	1.2	0
4	13-14	15-16	14400	2.0	24.0
5	14-15	16-17	18000	4.1	22.0
6	15-16	17-18	21600	0.6	7.2
7	16-17	18-19	25200	1.1	9.3
8	17-18	19-20	28800	3.5	2.4
9	18-19	20-21	32400	3.2	1.1
10	19-20	21-22	36000	0.2	2.3
11	20-21	22-23	39600	0.2	0.8
12	21-22	23-24	43200	0.0	1.2
13	22-23	24-25	46800	1.8	2.1
14	23-24	25-26	50400	0.6	0

Initial hydraulic conditions

Initial hydraulic conditions include depth of groundwater table, background infiltration rate and a general assessment of whether saturated or unsaturated analysis is most suitable. Groundwater levels from measuring station “Førde/Moskog rør 3” were downloaded from "SeNorge" (http://www.senorge.no/index.html?p=senorgeny&st=water), while distributed groundwater levels leading up to the event, estimated by NVE through a series of HBV-models (Sælthun, 1996), were downloaded from "Xgeo" (http://www.xgeo.no/index.html?p=fag).

4.2.3 Soil

Types and properties

The spatial distribution of soil types (Fig. 4.3) was represented through Quaternary geology maps, prepared by the Geological Survey of Norway (NGU, 2017). Considering landslide-prone areas, three soil types were identified as relevant: till (both continuous and discontinuous), landslide material and “bare rock”. Only brief comments on soil depth were included in maps, such as "more than 0.5m soil depth". Soil properties, including geotechnical and hydraulic parameters, were estimated in part from literature (Section 4.3) and in part through calibration (Section 4.4.2).

Soil depth

The spatial pattern of soil depth is a function of several interplaying factors, such as lithology, climate, slope, terrain curvature, upslope contributing area, and vegetation cover (Catani et al., 2010). To reduce uncertainties, various approaches for estimating soil depth have been suggested and applied (Fig. 5.2, both within landslide modelling and other disciplines. Linear (e.g., Saulnier et al., 1997) and exponential (e.g., Delmonaco et al., 2003) relationships to slope are commonly applied, while other explanatory variables (e.g., curvature and relative location on slope) have also been successfully used (Godt et al., 2008).

For this thesis, 27 in situ source area soil depth observations were available (Rüther et al., in prep.). Plausible explanatory variables (altitude and slope angle) were tested for correlation with soil depth, using statistical software "Minitab"'s (**Minitab**) linear and exponential regression. Additionally, observations were grouped on the basis of soil type and sub-area affiliation for possibly increased correlation. Finally, as DEM elevation errors propagate through each subsequent map (Haneberg, 2006), effects of DEM inaccuracies on final soil depth predictions were assessed. All predictions were evaluated using the coefficient of determination (R^2).

Additional soil depth data was downloaded from the National Groundwater Database (GRANADA, https://www.ngu.no/prosjekter/granada). Slope gradients at well locations from 1m and 10m resolution DEMs were tested as explanatory variables similarly to field measurements, and soil depths for soil types "continuous till" and "thin or discontinuous till" were evaluated separately. Numerous wells had either zero or no number in their soil depth attribute field; these were excluded, as it was uncertain whether these represented "no soil" or “no measurements”.

4.2.4 Vegetation

Vegetation maps were obtained through NIBIO’s (Norwegian Institute of Bioeconomy Researc)’ "AR50" (NIBIO, 2016) and "Sat-Skog" (Gjertsen & Nilsen, 2012). The former is a simplification of the more detailed areal resources maps “AR5”, which due to low resolution (1:50 000) is unsuited for quantitative analysis but suitable for visualization. The latter is generated through a combination of satellite data and reference points, providing increase resolution.

4.2.5 Digital topography

Digital elevation models (DEMs) of 1m, 10m and 50m resolution were downloaded from Høydedata (https://hoydedata.no/LaserInns a governmental public platform with data from various cooperators. DEMs form the basis for other data, and slope angles, flow directions and hillshade maps were created for each resolution in ArcGis Pro (Esri Inc., 2021)). Airborne laser data (2 points per m^2) was also available (Terratec, 2016), but remained unused due to computational requirements.

4.3 Model parameterization

Model input parameters not subject to calibration were estimated on the basis of available data, and were assumed as spatially uniform across the study area.

4.3.1 Unit weight of soil

Unit weight of soil was set equal to 20 kN/m^3, similar to the value used by Melchiorre and Frattini (2012) in their application of a physically based model in Otta, Norway.

4.3.2 Unsaturated parameters

Parameters relevant for unsaturated analysis only, i.e., saturated volumetric water content (θ_s), residual volumetric water content (θ_s) and Gardner's alpha (α_g), were estimated following the approach from Schilirò et al. (2015). Although unsaturated parameters are optimally obtained through field tests (Godt et al., 2008), it was instead relied on *pedotransfer functions*; these are empirical relationships between easily (e.g., grain size distribution) and less easily (e.g., α) measured soil properties. Bulk density ($1.62 g/cm^3$), grain size distribution and certain soil mechanical parameters for one sample of Jølster till were available (Fig. 4.11; Opsal, 2017; Opsal and Langeland, 2019). Although fine material (silt and clay, 4.2% of total sample) was not separated, it was assumed that the majority was silt. Quantities 1.65% clay, 8.24% silt and 90.1% sand (percentages of clay, silt and sand excluding coarser material) were therefore along with bulk density used as input in the "Rosetta Lite" pedotransfer module (Schaap et al., 2001), encoded in the Hydrus 1-D hydrological model (Simunek et al., 2008). The module consists of a hierarchical, increasingly complex set of functions, which based on neural networks estimate Van Genuchten-Mualem hydraulic parameters (Mualem, 1976; M. T. van Genuchten, 1980) with "reasonable" performance. To convert Rosetta Lites's alpha (α_{vg}) to α_g, formulas suggested by Ghezzehei et al. (2007) were applied. The "capillary drive" (1) and "capillary-length" (2) approaches yielded different results (Table 4.2), both of which were later tested in simulations. Rosetta Lite also estimates saturated hydraulic conductivity, but this value was applied only in unsaturated analysis.

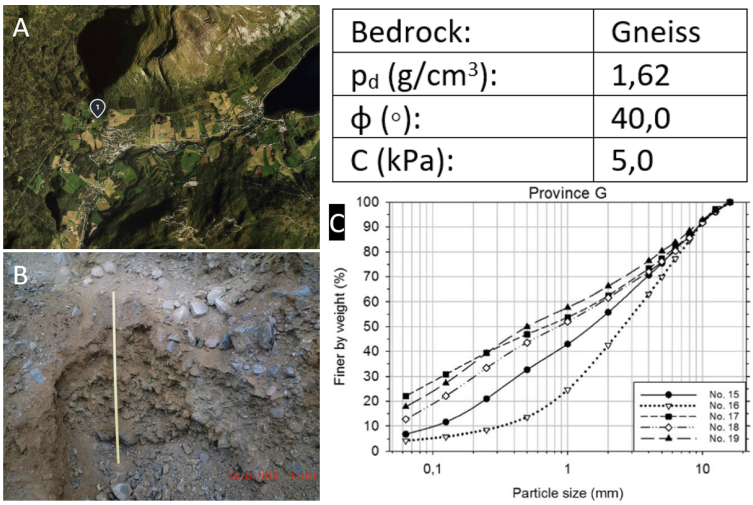


Figure 4.11: The Jølster till sample (grain size curve no. 16) from Opsal and Langeland (2019). (A) Location. (B) Field picture. (C) Grain sizes. The relevant sample is no. 16, while "province G" refers to a particular source rock province. The upper table denotes reported properties.

Table 4.2: Unsaturated parameters, their abbreviation, estimation method and predicted values. N is an intermediate parameter used in the conversion of α-values.

Parameter	Abbreviation	Obtainment	Value
Residual volumetric water content	θ_r	Rosetta Lite	0.0443 [cm^3/cm^3]
Saturated volumetric water content	θ_s	Rosetta Lite	0.5345 [cm^3/cm^3]
Saturated hydraulic conductivity	K_s	Rosetta Lite	6.287e-5 [m/s]
Alpha (Van Genucthen-Mualem)	α_{vg}	Rosetta Lite	5.83 [1/m]
N (Van Genucthen-Mualem)	n	Rosetta Lite	1.778 [-]
Alpha $(Gardner)_1$	α_{g1}	Eq. 17 of (Ghezzehei et al., 2007)	7.58 [1/m]
Alpha $(Gardner)_2$	α_{g2}	Eq 23 of (Ghezzehei et al., 2007)	15.077 [1/m]

4.3.3 Initial groundwater table

The initial groundwater table was assumed uniformly located at the intersection between soil and bedrock. This is a common assumption in literature (Fig. 3.4c), and if we compare recorded precipitation for July 2019 (Fig. 4.6) to average precipitation for July (Fig. 4.12b), the month emerges unusually dry after July 4^{th}. Further, observed groundwater level in the valley bottom was more than 1.5m below the 25th percentile (Fig. 4.12a, while modelled (Figs. 4.12c, d) levels indicate low or very low levels depending on location.

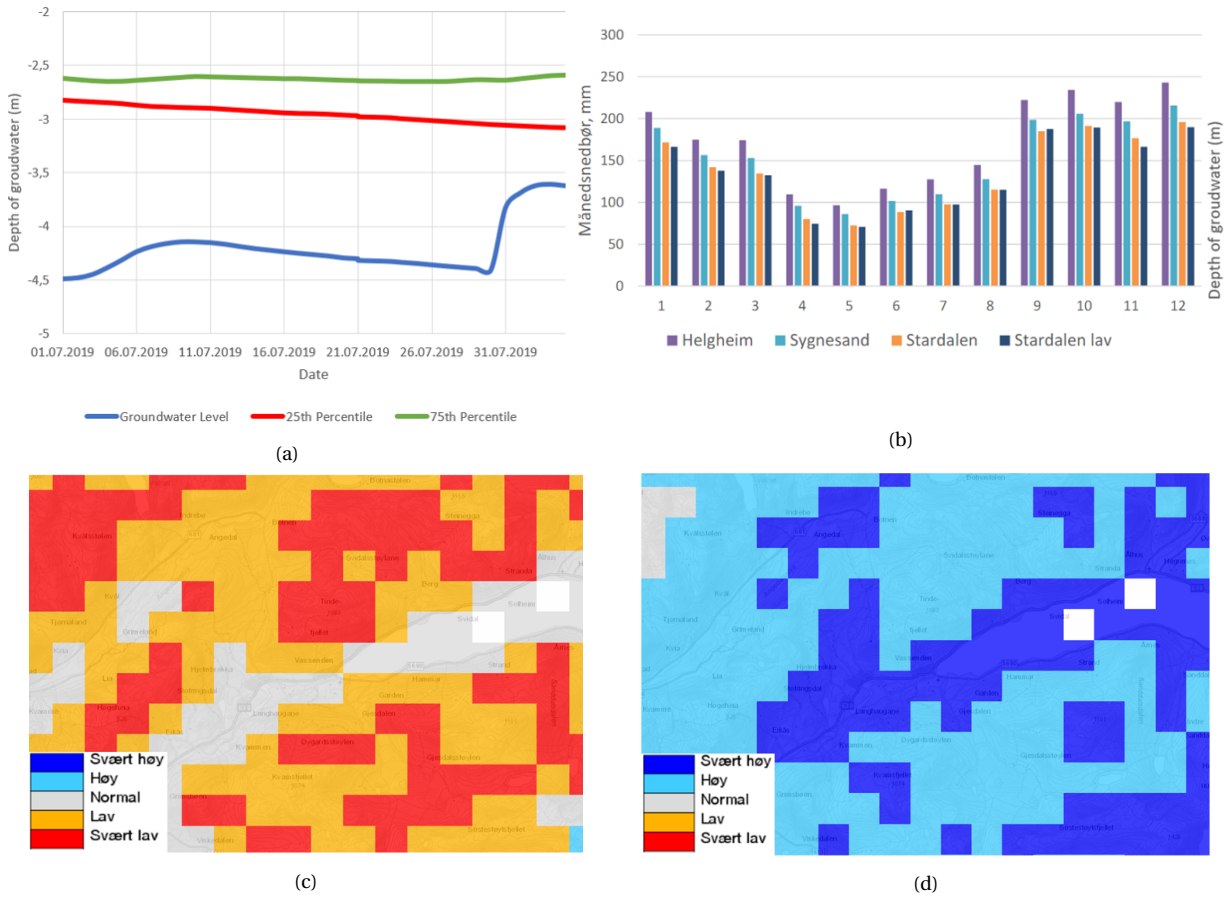


Figure 4.12: Basis for assuming initial groundwater levels. (a) Monthly precipitation normals for a subset of rain gauges in close proximity to the study area. Figure from Hefre et al. (2019). (b) Monthly precipitation normals for a subset of rain gauges in close proximity to the study area. Figure from Hefre et al. (2019). (c) and (d) Modeled, distributed groundwater conditions for Jølster at July, 30 (before the event) and July 31 (after the event), in correspondence to the study area extent. Legend in Norwegian; red = very low/deep, blue = very high/shallow.

4.3.4 Background infiltration rate

Various methods for estimating background infiltration rate (I_z) are identified in literature. The approach used by e.g., Baum et al. (2008), which simply utilizes average precipitation for two antecedent weeks, yields values of I_z according to Table 4.3 for Jølster. However, simpler methods such as $I_z = 0.01 * K_s$ (e.g., Liu & Wu, 2008) also exist, which would result in I_z = 6.29e-7 m/s if using K_s estimated by Rosetta Lite (Table 4.2). Due to uncertainties in the theoretical basis and desired conservativity, a rather high value if I_Z = 1e-7 m/s was applied throughout all simulations.

Table 4.3: Average two-week precipitation rates antecedent to July 30, 2019 for rain gauges close to the study area.

Rain gauge	P_{avg} [mm/d]	P_{avg} [m/s]
E39 Vassenden	3,3	3.82E-08
Skei	1,9	2.14E-08
Haukedal	2,1	2.40E-08
E39 Halbrendslia	1,9	2.16E-08
Botnen	2,8	3.27E-08

4.4 Modelling procedure

4.4.1 Performance evaluation

Evaluation criteria are necessary for evaluating performance and identifying optimal parameter combinations, but a platitude of methods are identified in literature (Fig. 3.5a). For this thesis, four criteria were applied:

- D2PC (Eq. 2.7), which evaluates performance at $T_{fs} = 1$, is applied due to its relationship with ROC and usage in calibration by Zieher et al. (2017).
- As an attempt to avoid the adverse effects induced by unconditionally stable cells, I introduce the term "critical cells". These are cells that may *potentially* fail; these are here defined by upper and lower slope angle limits of 50° and 20° respectively, but should be specifically evaluated for each independent case. Cells of slope angles above or below these limits are considered unconditionally stable due to little soil and too low slope angles for initiation respectively. "Critical negatives" (N*), defined as the number of critical cells minus the number of positives, may thus be notably fewer than initially considered. Now, a modified false positive rate (FPR*), defined as FP/N*, may be plotted in the ROC-space and yield the second evaluation criterion, D2PC*.
- Critical success index (CSI), initially applied in evaluation of a landslide early warning system (Cheung et al., 2006), is applied due to its neglection of true negative cells. Defined as TP/(TP+FP+FN), it may (almost) be thought of as "the number of correctly predicted landslides per wrong prediction".
- As an alternative measure, MCC (Eq. 2.6) considers performance in all four contingency matrix categories (Eq. 2.6). Regarded as more robust than many other measures (Chicco & Jurman, 2020; Vihinen, 2012), it is not commonly used in landslide modelling.

For investigation of criterion effects, a similar calibration procedure was applied as for parameter back-calculation (Section 4.4.2), although a 50*50m DEM was used across the entire study area for computational speed. Attention was given to performance differences, and relationships between true positives, false positives and various criteria across a range of parameters were explored. The negative to positive rate (N/P) was considered to explain true positive bias in ROC, while N* was considered in relation to differences between D2PC and D2PC*. I distinguish between *overall* performance (the amalgamated performance through a range of cut-off values/T_{FS}, e.g., AUC) vs. *threshold-specific* performance (at one specific T_{FS}, e.g., D2PC, CSI). The former indicates a general ability to rank-order cells, while the latter considers actual results at one instance. I also distinguish between *objective* performance (absolute results) and *relative* performance (to other results).

4.4.2 Calibration procedure

Calibration for parameter back-calculation was performed, with 10*10m DEM-resolution, for both saturated and unsaturated analysis. To automatize and improve the procedure, scripts were written in Python with the ability to generate TRIGRS initiation files, repeatedly execute TRIGRS and evaluate model performance through a desired set of measures (Table 4.4, full scripts in Appendix A).

Two distinct sub-areas were chosen for independent calibration, in such way that parameters would later be applied over the entire study are. This significantly reduced computation time and provided the ability to explore whether local-scale calibration was sufficient for setting up a large-scale model. "Årnes" and "Slåtten" (see areas of Fig. 4.8) were chosen on the basis of data availability (field studies, known landslide occurrence times), representativeness (relevant soil types covered) and importance (damages, media coverage).

Table 4.4: Summary of Python scripts utilized in calibration. See Appendix A for full scripts.

Script Name	Function
Init_file_generator	Generates a number of initiation files by systematically varying parameters within given ranges.
TRIGRS_executor	Repeatedly executes TRIGRS from a list of initiation files and saves output at the desired location.
Analysis_multifiles	Calculates performance indexes for a set of output FS-files.
Analysis_magnitudes	Calculates the spatial distribution of predictions and error magnitudes on a single FS-file.

Non-calibrated model input was set according to Section 4.3. Then, for each sub-area, four sensitive parameters (cohesion, friction angle, diffusivity and saturated hydraulic conductivity) were systematically varied in simulations within ranges given in Table 4.5. Sensitive parameters were chosen according to relevant literature on sensitivity and back-calculation (e.g., Melchiorre and Frattini, 2012; Bordoni et al., 2015; Zieher et al., 2017), while ranges were chosen partly from literature (e.g., values reported by Melchiorre and Frattini, 2012) and partly by "try and fail" in order to cover both stability end-members.

Table 4.5: Parameters varied in saturated and unsaturated calibration. N refers to the number of simulations.

Saturated analysis, N=240			
Parameter	**Minimum**	**Maximum**	**# of values**
C [Pa]	2100	3600	4
Phi [deg.]	32	37	3
D0 [m^2/s]	0.00009	0.0009	4
K_S [m/s]	0.00001	0.0005	5

Unsaturated analysis, N=720			
Parameter	**Minimum**	**Maximum**	**# of values**
C [Pa]	600	999	5
Phi [deg.]	32	37	3
D0 [m^2/s]	0.00009	0.0009	4
K_S [m/s]	0.00001	0.0009	6
α_G [-]	7.88	15.077	2

Results were first evaluated in terms of correct and wrong predictions, with all landslides except Årnes assumed to occur at some point between 16:00 and 17:00. Parameter sets not adhering to basic requirements were neglected; such requirements were, following the approach of Zieher et al., (2017): 1) All cells are stable at 13:00 (before intense rainfall), and 2) most failure is predicted during and after the most intense rainfall (significant decrease in stability between 15:00 and 17:00).

Parameter combinations were further rank-ordered by evaluation criterion, and optimal parameters were identified. As argued in Section 6.3, different criteria measure different aspects of performance, and as the choice of one single criterion was challenging, two very different measures (D2PC and CSI) were applied. Parameters for Årnes and Slåtten were assumed representative for till and landslide material respectively, and exported to the entire study area where a limited number of simulations were performed due to computational limitations. Performance of the optimal, full-area simulation was finally subject to assessment through a novel approach; error magnitudes (Table 4.6), which considers non-binary performance, were spatially plotted in charts and maps along with contingencies

Table 4.6: Names, abbreviations and definitions of "error magnitudes".

Error term	**Abbreviation**	**Formula**
Confidence	C	1 - FS_{TP}
False confidence	FC	1 - FS_{FP}
Critical error magnitude	CEM	FS_{FN} - 1

Chapter 5

Results

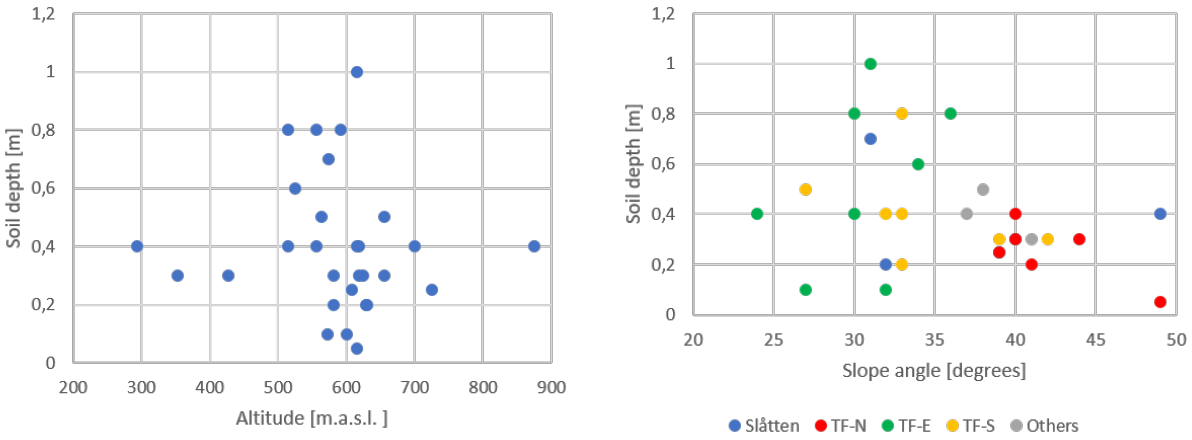
This chapter is organized in the following order: first, results from the statistical analysis of source areas are presented. Results from soil depth modelling come next next, forming an important basis for further simulations. Then, effects of evaluation criteria are investigated, with saturated and unsaturated results presented separately at the end.

5.1 Statistical analysis of landslide source areas

The majority of landslides initiated at some altitude, with only 11 source areas located below 400 m.a.s.l. (Fig. 5.1a). Typical slope angles are 25°– – 45°, although ranging from 18.9° to 53.5° (Fig. 5.1b). Though variance may slightly decrease, no significant relationship between slope angle and elevation is present (Fig. 5.1c). Areas classified as "no or limited vegetation" contain the majority of landslides (Fig. 5.1d); the majority occurred in deciduous and pine-dominated forests, while interestingly, only 3 of 91 landslides initiated in spruce-dominated areas. Landslides are mostly distributed between till, landslide deposits and "bare mountain" (Fig. 5.1e). Slope angles vary significantly within soil types, although landslide deposits and bare mountain generally facilitate failure at higher slope angles than till (Fig. 5.1f).

Figure 5.1: Results from the statistical analysis of source areas. Frequencies refer to the number of source area *cells*, not source areas per se. (a) Elevation distribution. (b) Slope angle distribution. (c) Relationship between slope angles and elevation. (d) Vegetation distribution. "Mixture" refers to varying quantities of pine, spruce and deciduous forest. (e) Soil type distribution. (f) Slope angles for source areas of different soil types.

5.2 Soil depth modelling



(a) (b)

Figure 5.2: Average source area soil depths from field studies. Observations are plotted vs. (a) altitudes and (b) average source area slope angles. Colors indicate sub-areas of Fig. 4.8 (TF = Tindefjellet, N, E and S = north, east and south respectively).

Soil depth appears uncorrelated with altitude, but slightly correlated with slope angle (Figs. 5.2a, b). Grouping observations in sub-areas indicates differences: large variations in soil depth (0.1m–0.9m for equal slope angles) occur in "Tindefjellet East", while landslides initiate at steep angles in "Tindefjellet North". A similar grouping on soil types was also tested, but no meaningful results were obtained.

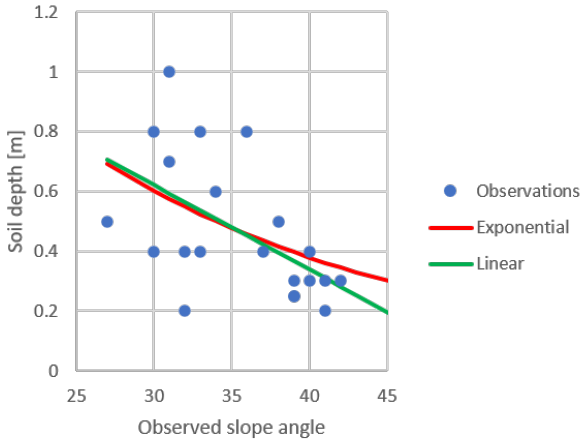


Figure 5.3: Regression model examples for correlating soil depth to slope angle.

Examples of regression models based on 21 relevant observations (low statistical robustness) are shown in Fig. 5.3, while corresponding equations and coefficients of determination are shown in Table 5.1. Regression was performed to maximize R^2 and minimize total absolute errors, but only linear and exponential regression was considered in light of conventional approaches reported in literature

Table 5.1: Coefficients of determination for soil depth regression models (calculated vs. observed thickness), digital elevation models (DEM slope angles of source area cells vs. observed slope angles) and combinations of both (predicted thickness from using DEM slope angles vs. observed thickness). θ is the slope angle.

	Regression model		DEM			Regression model + DEM	
Model	$2.41e^{(-0.046*\theta)}$	$1.47-0.028*\theta$	50m	10m	1m	Lin. + DEM 10m	Exp. + DEM 10m
R^2	30 %	28 %	15 %	58 %	68 %	14 %	17 %

Digital elevation models are increasingly accurate with resolution (Table 5.1), although errors of more than 10° are present even at 1*1m resolution. The combination of two error sources (soil depths and slope angles) results in decreased values of R^2 (14-17%) compared to regression models alone.

Finally, neither altitude nor slope angle shows correlation with well soil depths. But, "thin or discontinuous till" consistently yields shallower soil than "continuous till" (Fig. 5.4).

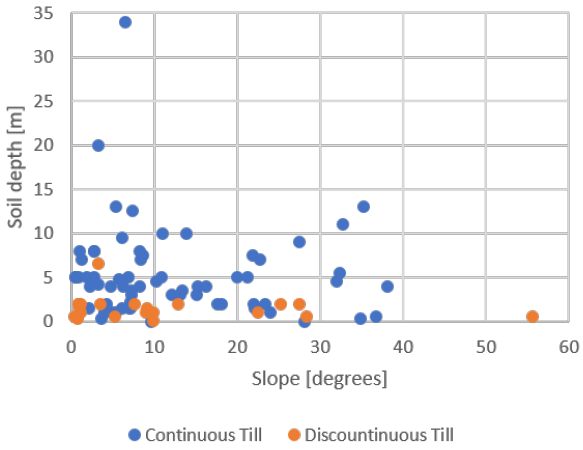


Figure 5.4: Soil depths from energy- and groundwater wells for two soil types vs. slope angles (1m DEM).

5.3 Evaluation criteria effects

The full area (50*50m DEM) contains 94 605 cells, of
which 156 cells represent landslide source areas. The ratio N/P (stable/negative cells divided by unstable/positive cells is thereby 605.4, while only 34 152 cells (36%) are identified as critical (Fig. 5.5).

Optimal parameter sets vary by evaluation criteria defined in Section 4.4.1 (Table 5.2). This is especially evident for relative rankings (among 240 simulations performed for criteria effect investigation), as e.g., set “cv6” ranks 64_{th} in terms of CSI. In fact, “Cv6” obtains the best D2PC, predicts the highest number of positives, but also has the highest FP/TP-ratio. On the other hand, “cv5” predicts less than half the true positives of cv6, performs worse at D2PC but maximizes CSI. D2PC* and MCC are optimized by equal parameter sets, while the difference between D2PC and D2PC* is modest. A visual representation of evaluation criteria effects on factor of safety is given in Fig. 5.6.

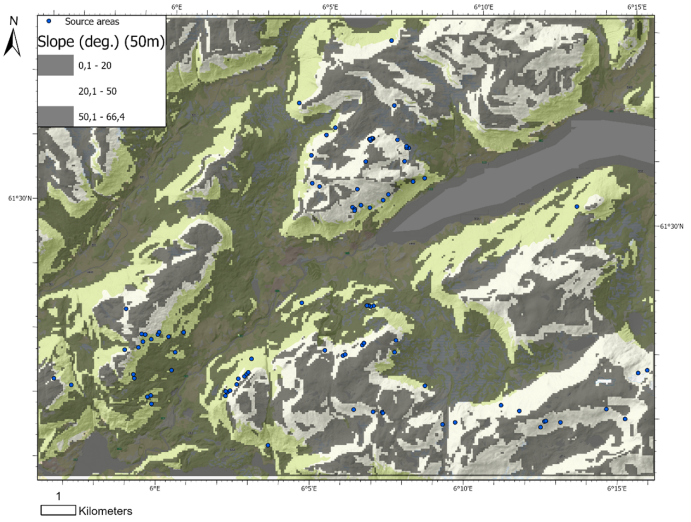


Figure 5.5: Critical cells (20°< slope < 50°, not shaded) for the 50*50m DEM.

Identified parameters vary; decreased K_s and/or D_0 apparently decreases stability, but no differences are found within geotechnical parameters.

Table 5.2: Key results from investigating evaluation criteria effects: performance, relative rankings and identified parameters for example parameter sets. "Criteria" refers to each set's optimized criteria. ϕ and C did not vary between presented sets. TN and FN are not included, but can be calculated knowing the area's total number of negative (94 605 - 156) and positive (156) cells.

Optimization		Rankings				Parameters	
Set	Criteria	D2PC	D2PC*	CSI	MCC	D0	Ks
cv2	D2PC*, MCC	4	1	21	1	0.0009	0.00005
cv5	CSI	19	17	1	12	0.0009	0.0005
cv6	D2PC	1	6	64	8	0.0006	0.00005

	Predictions		Performance				
Set	TP	FP	D2PC	D2PC*	CSI	MCC	FP/TP
cv2	67	8938	0.58	0.63	0.0074	0.046	133.40
cv5	32	3159	0.80	0.80	0.0097	0.039	98.72
cv6	71	11401	0.56	0.64	0.0061	0.042	160.58

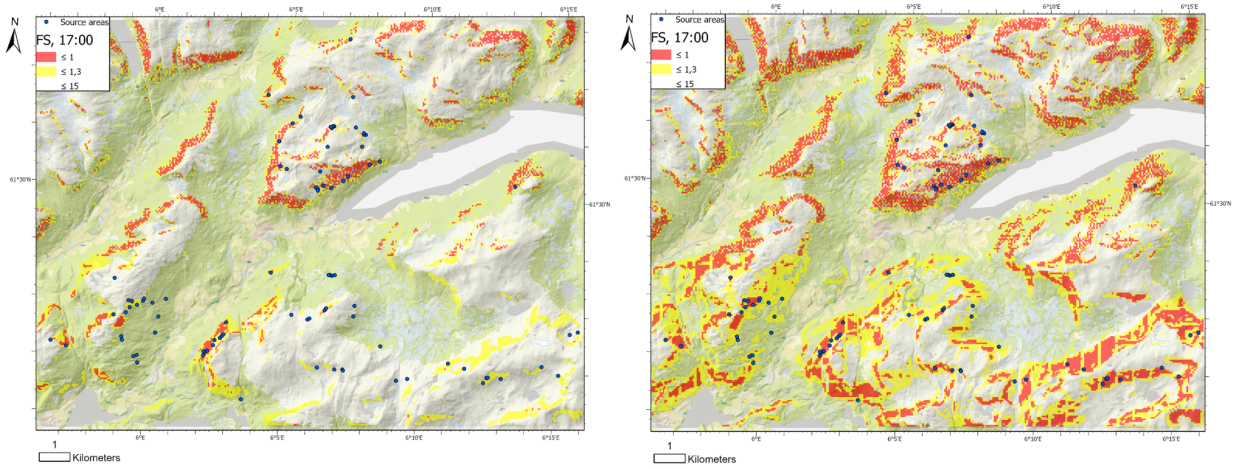


Figure 5.6: Comparison of factor of safety estimated by chosen parameter sets. (a) "cv5" and (b) "cv6", optimizing CSI and D2PC respectively.

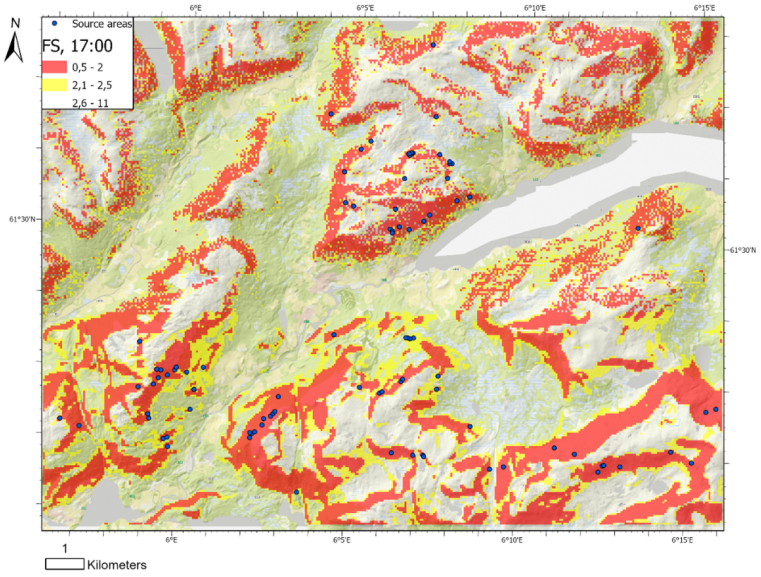
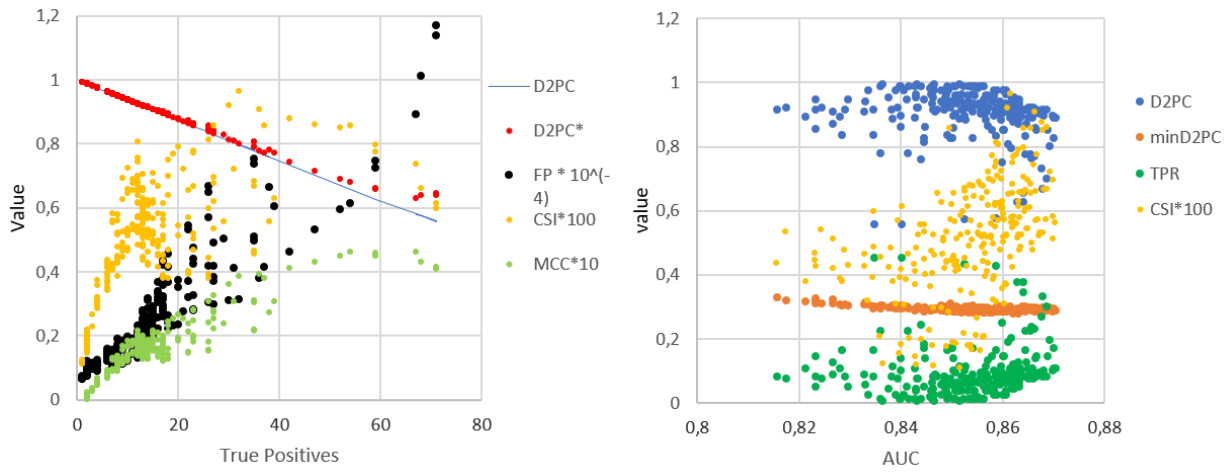


Figure 5.7: Unstable areas when applying an "optimal threshold" of $T_{FS} = 2$, which maximizes D2PC.

Adjusting T_{fs} makes for identification of an “optimal threshold” minimizing D2PC. As an example, optimal threshold for the parameter set with the lowest minimum D2PC is FS=2.1. When applied, this threshold results in 139 true positives, 24 319 false positives and D2PC=0.281 (Fig. 5.7).

Fig. 5.8a shows the relationship between true positives, false positives and selected evaluation criteria. Interestingly, D2PC is almost a linear function of true positives. D2PC and D2PC* differ only slightly, although the difference increases with numbers of true and false positives. Other measures show more variation by taking false positives into account; CSI maximizes at approximately the maximum TP/FP-ratio, while MCC is more biased towards true positives and maximizes at a less stable parameter set.

Fig. 5.8b shows the relationship between selected criteria and AUC. Increased AUC indicates decreased D2PC (and a slight decrease in minD2PC), but variation is considerable. AUC is also slightly related to true positive rate, while its strongest relationship is apparently that to CSI.



(a) (b)

Figure 5.8: Relationship between selected performance indicators. (a) Selected criteria vs. the number of true positives. (b) Selected criteria vs. AUC. CSI, MCC and the number of false positives are scaled for comparison.

Criteria effects vary with study area (10*10m DEM). Årnes includes 2 positive and 7126 negative cells (N/P = 3563), while 75% of cells are critical. Slåtten includes 66 positive and 25 218 negative cells (N/P = 382) of which 54% of cells are critical, with higher complexity in source area properties. Statistics for both sub-areas are shown in Table 5.3. One single parameter set optimizes all criteria for Årnes, while for Slåtten, two sets are identified. Here, D2PC, D2PC* and MCC are maximized by the same set, while again, CSI is optimized by less predicted landslides and higher ratio of TP/FP.

Table 5.3: Evaluation criteria effects in sub-areas. Relative rankings and performance for selected sets in Årnes ("ar") and Slåtten ("sl").

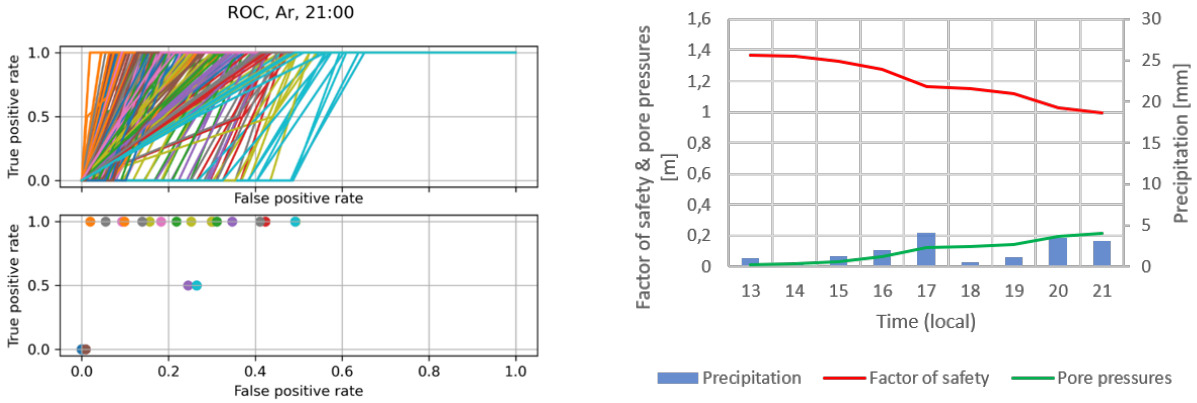
	Rankings				Predictions		Performance				
Parameter set	**D2PC**	**D2PC***	**CSI**	**MCC**	**#TP**	**#FP**	**D2PC**	**D2PC***	**CSI**	**MCC**	**FP/TP**
ar29	1	1	1	1	2	138	0.019	0.026	0.014	0.12	69
sl8	1	1	6	1	48	2660	0.29	0.34	0.018	0.10	55.42
sl9	12	12	1	3	43	2165	0.36	0.38	0.019	0.10	50.35

5.4 Parameter back-calculation

5.4.1 Saturated analysis

Årnes

The Årnes sub-area includes two failure cells whose predicted soil depths (36cm) are lower than observed (50cm). Considerable differences in performance can be observed between parameter sets; e.g., false positive rates required for identifying both failure cells range from 2% to 65% (Fig. 5.9a). AUC ranges from 87.6% to 93.6%. Of 240 simulations (at 21:00), 20 identify both failure cells, 4 identify 1 cell and 216 identify no cells. Of the above-mentioned 24 sets identifying failure, 23 do so prematurely (at 20:00) yielding “ar29" as optimal (results by various criteria in Table 5.3). Corresponding estimated parameters are shown in Table 5.4, which also includes identified optimal parameters when soil depths at failure cells are adjusted according to field studies.



(a) (b)

Figure 5.9: (a) ROC-curves and points for sub-area Årnes at 21:00 local time. Curves indicate overall performance, while points represent performance at $T_{FS} = 1$ (b) Precipitation, factor of safety and pore pressures for Årnes failure cell (47,35). Note that the left vertical axis denotes both factor of safety and pore pressures.

Table 5.4 also shows selected statistics in time for "ar29". All cells are initially stable, until 16 false positives occur at 20:00 after modest, continuous rainfall (Fig. 5.9b). Factor of safety declines as pore pressures increase, dropping below 1 at 20:30 when pore pressures reach 0.2m. The correct prediction of both failure cells induces 138 false positives, all located within similar slope ranges. AUC increases with time indicating infiltration, groundwater and runoff computations improve the model's ability to rank-order cells.

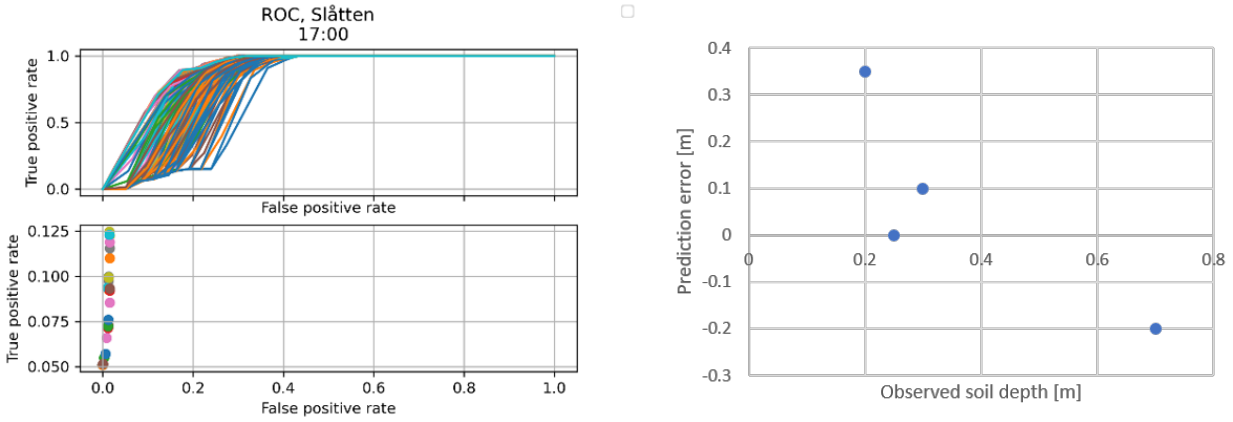
Table 5.4: Estimated parameters and statistics in time for sets "Ar29" (predicted soil depths) and "Ar_{obs}" (observed soil depths). Units: C = Pa, Phi (ϕ) = degrees, D_0 = m/s^2 and K_S = m/s.

Set	C	Phi	D0	Ks
Ar29	2100	35	6e-4	2e-4
Ar_obs	2100	35	3e-4	8e-5

Time	minFS	#TP	#FP	AUC
13:00	1,24	0	0	90.2%
17:00	1,03	0	0	91.9%
20:00	0.99	0	16	92.7%
21:00	0.98	2	38	93.6%

Slåtten

The Slåtten sub-area includes 10 landslides represented by 66 failure cells. These are mainly located in landslide material close to a steep wall, with observed and predicted soil depths as depicted in Fig. 5.10b.



(a) (b)

Figure 5.10: (a) ROC-curves and points for the Slåtten sub-area at 17:00 local time. (b) Observed soil depths and corresponding errors in soil depth predictions for Slåtten source areas subject to field studies.

Compared to Årnes, less variation between parameter sets is observed (Fig. 5.10a). AUC at 17:00 ranges from 71.6% to 91.6%, and both increases and decreases for different parameter sets throughout the modeled period (average change in AUC = 0.14%). General performance is poor; FPR required for TPR=1 ranges from 30% to >40% (Fig. 5.10a). Of 240 parameter sets, no sets obtain TPR=1. Specifically, 5 sets identify 48 true positives (TPR=72.7%), 5 sets identify 47 true positives while a total of 33 sets identify a minimum of one landslide cell. All sets are stable at 13:00; however, many identify instability prematurely (at 16:00) but are not neglected from evaluation due to heavy rainfall occurring in that period. Performance in terms of criteria generally decreases from 16:00 to 17:00, due to a faster increase in #FP compared to #TP.

Two optimal parameter sets are, depending on evaluation criterion, identified for Slåtten at 17:00 (Table 5.5). These are similar in terms of C and ϕ, while the higher stability of SL5 is attributed to increased D_0 and K_S. Interestingly, parameters identified for for Årnes ("ar29", Table 5.4) provide more stability; when these are applied for the Slåtten area, 28 true positives and 1637 false positives are identified yielding a slightly higher TP/FP-ratio than "SL12" and "SL5".

Table 5.5: Identified parameters and performance at various times for sets SL12 and SL5. Landslides are assumed to occur between 16:00 and 17:00. Units: C = Pa, Phi (ϕ) = degrees, D_0 = m/s^2 and K_S = m/s.

Set	C	Phi	D0	Ks
SL12	2100	32	3e-4	5e-5
SL5	2100	32	9e-4	5e-4

Set	Time	minFS	#TP	#FP	AUC	D2PC	CSI
SL12	13:00	1.30	0	0	91.9%	1.000	0.000
	16:00	0.92	45	1898	92.8%	0.33	0.023
	17:00	0.92	48	3092	90.5%	0.30	0.015
SL5	13:00	1.30	0	0	91.9%	1.000	0.000
	16:00	0.95	16	785	91.1%	0.76	0.019
	17:00	0.92	38	2352	90.2%	0.43	0.016

Table 5.5 also shows key statistics over time for both parameter sets. Both estimate complete stability until 16:00, when a large drop in FS is observed in response to intense precipitation of 24.2mm and 20.5mm (Table 4.1). Minimum factor of safety is equal for both sets, indicating that parameter differences affect FS not in the most unstable cells. Maximum AUC is reached at 16:00 for "sl12", but constantly decreases for "sl5".

Study area

The full study area, which contains 91 landslides, 605 failure cells and 933425 cells in total, is modeled with parameter sets "f-d" (ar29+SL12) and "f-c" (ar29+SL5), optimizing D2PC and CSI in sub-areas respectively. Due to the source area soil type distribution of Fig. 5.1e, "bare rock" was, as for Årnes, treated as till throughout the study area.. Key statistics over time are shown in Table 5.6. Again, no failure is identified at 13:00, but for both sets, a number of false positives are identified at 16:00 followed by a further decline in stability towards 17:00. Instability develops gradually towards 21:00 and the occurrence of the Årnes landslide. AUC increases with time, but is relatively high already at 13:00.

Table 5.6: Performance in time for full study area parameter sets "f-d" and "f-c".

Set	Time	#TP	#FP	AUC	D2PC	CSI
f-d	13:00	0	0	85.0%	1	0
	17:00	106	14977	86.3%	0.83	0.0068
	20:00	161	47681	87.7%	0.73	0.0033
	21:00	197	84789	87.8%	0.68	0.0023
f-c	13:00	0	0	85.0%	1	0
	17:00	91	6501	86.4%	0.85	0.0128
	20:00	154	35648	87.8%	0.75	0.0042
	21:00	194	70532	88.0%	0.68	0.0027

Parameter set “f-d”, optimizing D2PC in Slåtten, is most unstable throughout the modeled time period. At 20:00, “f-d” and “f-c” identifies 161 and 154 true positives (of 605 in total), and 47681 and 35648 false positives respectively. Thus, for obtaining 7 additional true positives, 12033 false positives are required, but “f-d” still outperforms "f-c" in terms of D2PC. However, “f-c” performs better on other measures of Table 5.6. Differences between parameters, which is mainly that of more false positives for "f-d", are visualized in Figs. 5.11 & 5.12. It is also worth noting that the application of "ar29"-parameters for the entire study area yields 133 true positives and 32 019 false positives at 17:00, and 169 true positives and 65 028 false positives at 21:00.

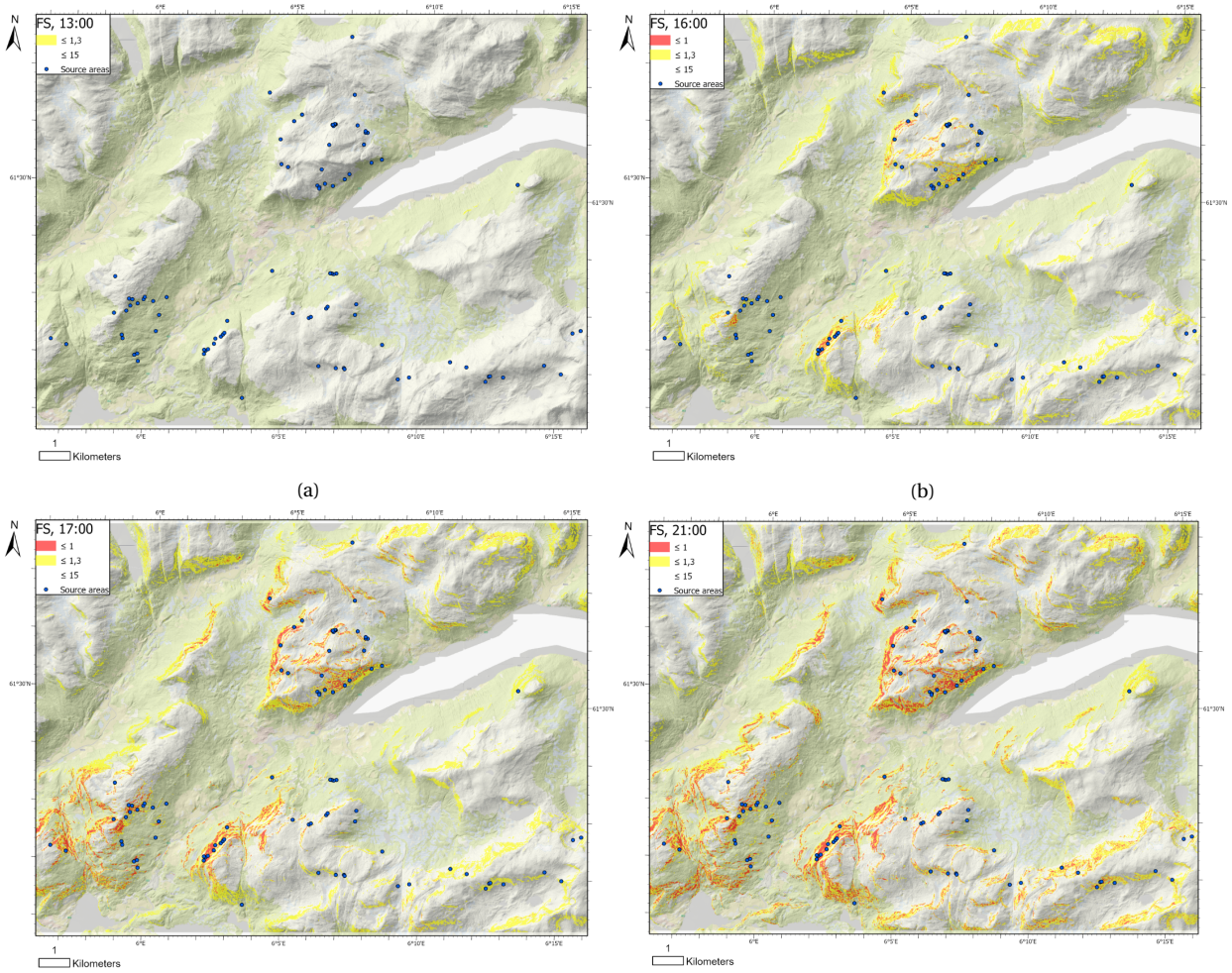
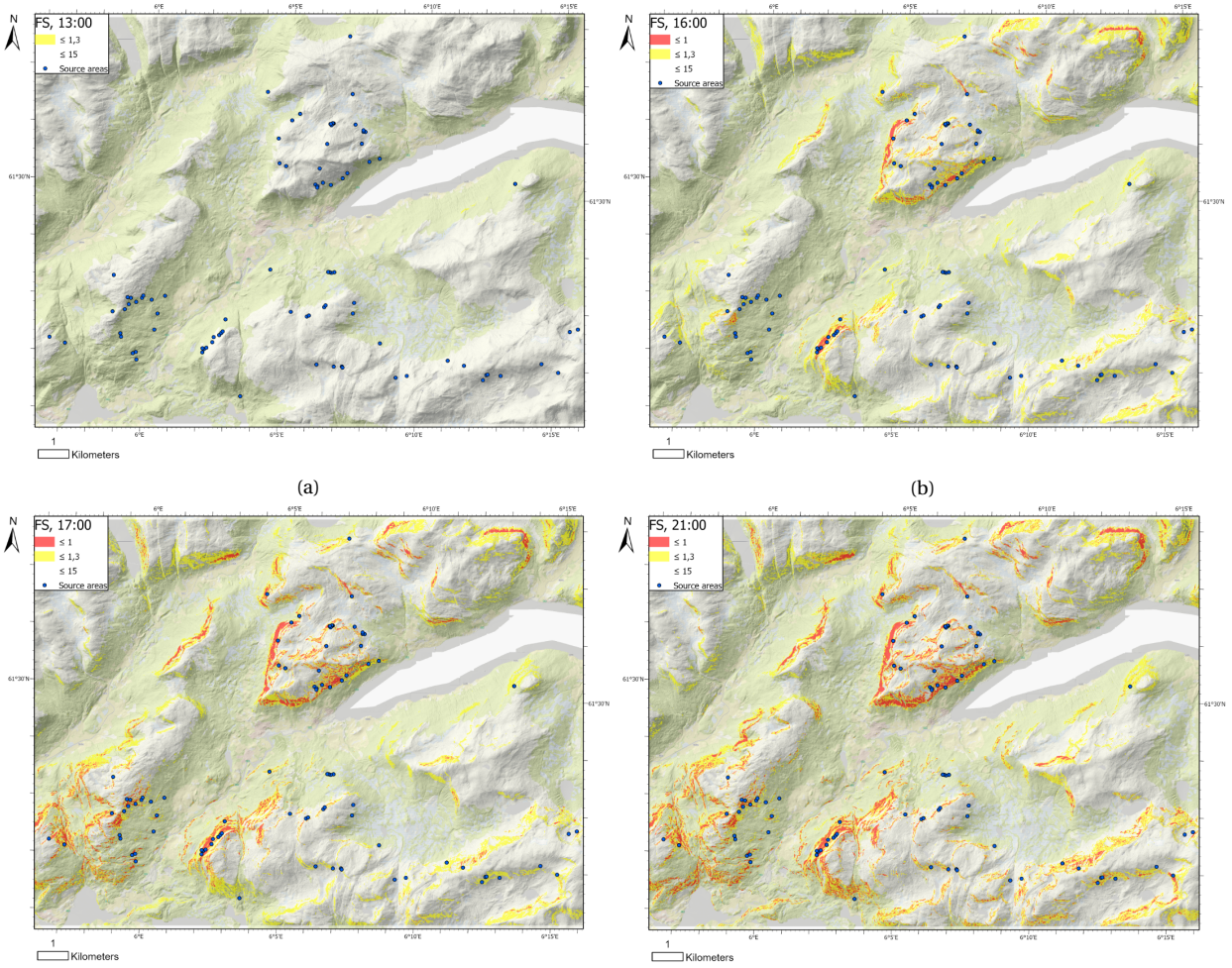


Figure 5.11: Factor of safety development in time for parameter set "f-c"



(a) (b)

(c) (d)

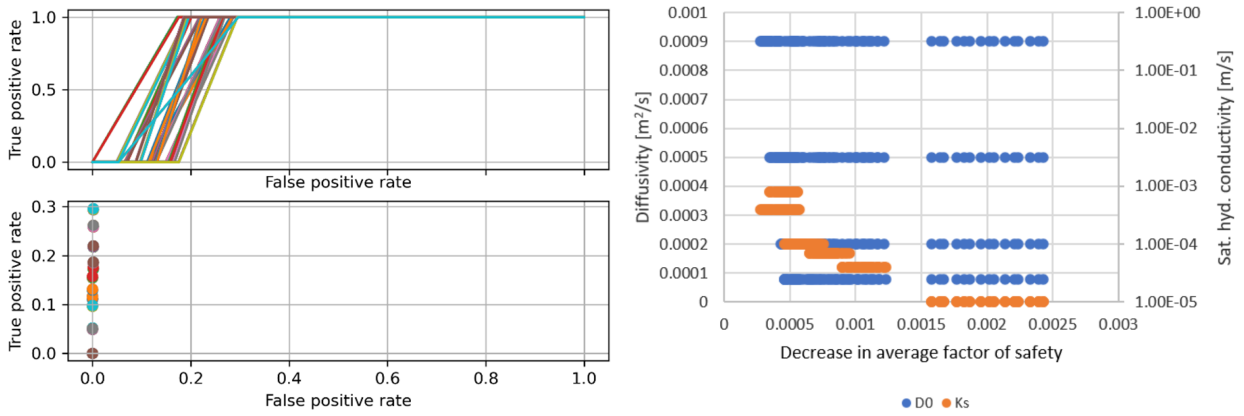
Figure 5.12: Factor of safety development in time for parameter set "f-d".

5.4.2 Unsaturated analysis

Årnes

No parameter set from the saturated analysis identifies instability when applied in unsaturated analysis, but adjusted ranges given in Table 4.5 result in ROC-curves of Fig. 5.13a. Identification of both failure cells, which is achieved by 33% of sets for both values of α_G, requires large false positive rates compared to the saturated analysis.

However, no set fulfills the requirement of stability before rainfall. In fact, pore pressures and FS remain nearly constant throughout the event for all sets (Fig. 5.13b). Decrease in average factor of safety is in the range of 0.005 – 0.0025; decreased K_S induces more change, while variations in D_0 remain negligible. Cohesion in the range of 600–800 Pa is required for instability at 21:00, resulting in unconditional instability at most slope angles > 41° throughout the modeled period.

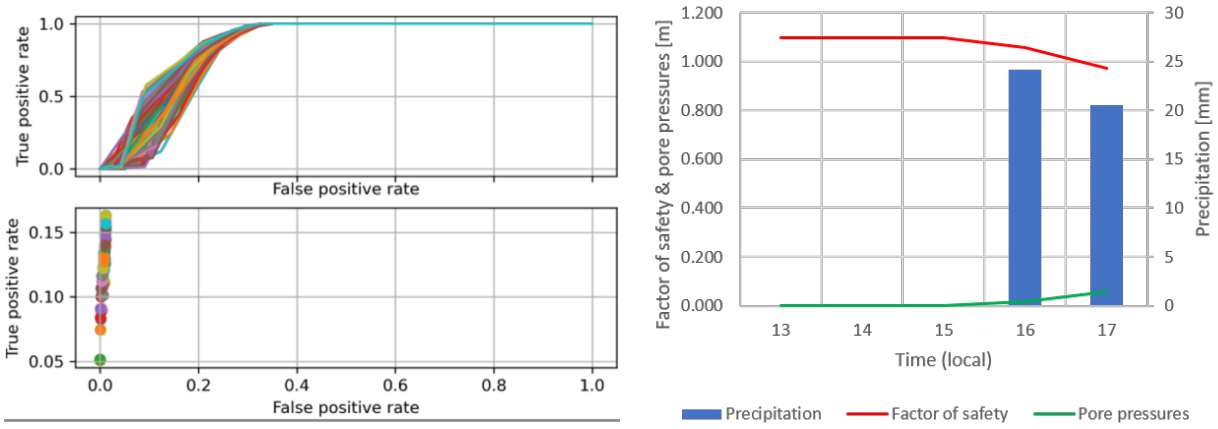


(a) (b)

Figure 5.13: (a) ROC-curves and points for unsaturated analysis, Årnes, 21:00 local time. (b) Applied diffusivity and cohesion vs. decrease in factor of safety between 13:00 and 21:00.

Slåtten

High false false positive rates are required also for Slåtten (Fig. 5.14a). 324 sets identify failure at 17:00, of which 36 are stable at 13:00. All sets are unstable at 16:00, but still considered for evaluation. The decrease in FS over time is larger compared to Årnes, but small compared to the saturated analysis. True positives range from 0—49, while false positives range from 1310—4140. AUC-values – still decreasing as rainfall proceeds – range from 0.81-–0.88.



(a) (b)

Figure 5.14: (a) ROC-curves and points for unsaturated analysis, Slåtten, 17:00 local time. (b) Precipitation, factor of safety and pore pressures for correctly predicted cell (81,85) of the Slåtten sub-area.

Among the sets characterized as stable before the beginning of rainfall, "sd332" obtains 20 true positives, 2877 false positives and optimal scores in terms of D2PC and CSI at 17:00 (Table 5.7). Minimum FS is well below one, identifying 1484 false positives prematurely at 16:00. Lower soil strength parameters and different hydraulic parameters are estimated compared to saturated analysis, while failure occurs at considerably lower pore pressures (Fig. 5.14b).

Table 5.7: Estimated parameters and performance in time for Slåtten parameter set "sd332". Units: C = Pa, Phi (ϕ) = degrees, $D_0 = m/s^2$ and K_S = m/s.

Set	C	Phi	D0	Ks
Sd332	999	35	9e-4	4e-5

Set	Time	minFS	#TP	#FP	AUC	D2PC	CSI
sd332	13:00	1.02	0	0	90%	1	0
	16:00	0.88	2	1484	86%	0.97	0.0013
	17:00	0.75	20	2877	85%	0.71	0.0068

Saturated vs. unsaturated analysis

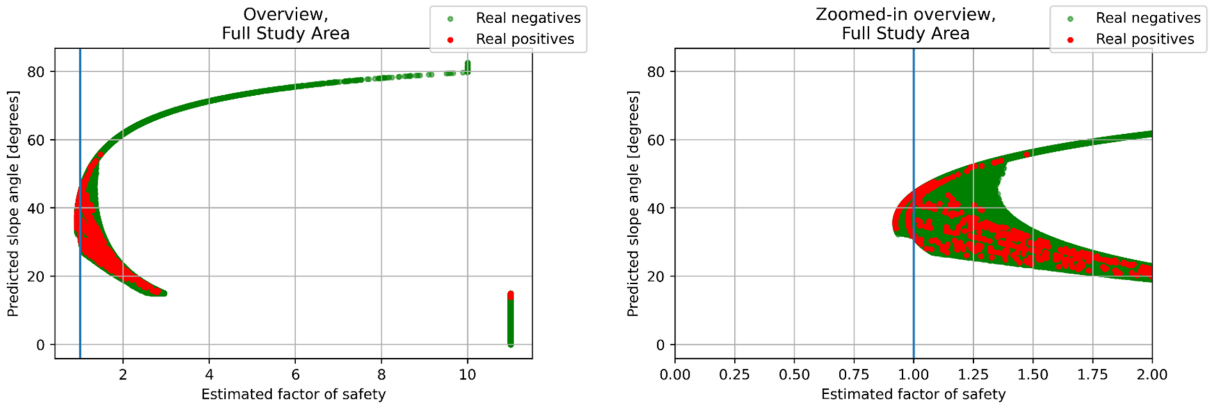
Model performance across all applied measures is superior for saturated analysis, especially for the Årnes sub-area. Generally, little change in factor of safety is observed in unsaturated analysis; this requires low estimated soil strength and identifies few true positives and numerous false positives if basic requirements are satisfied. AUC decreases over time for unsaturated analysis, compared to a consistent increase in saturated analysis. As a consequence, the unsaturated analysis is not applied in the full study area.

.

5.4.3 Spatial and non-binary assessment of model performance

In the following section, the best-performing (subjectively, by D2PC and CSI) parameter combination "f-c" is further evaluated. The section is primarily an example of how results may be plotted and analyzed..

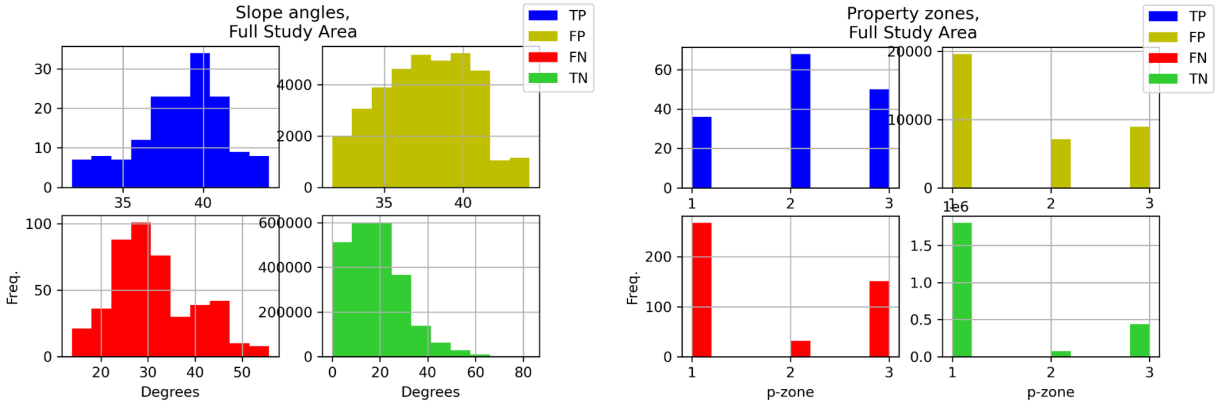
Examining the distribution of all modeled cells, a strong relationship between slope angle and estimated FS is evident (Fig. 5.15a). Variation in slope angle is largest at ca. 1 > FS > 1.37 (Fig. 5.15b), while both increasing and decreasing slope angles outside od this range increases FS. Instability is predicted between ca. 30° and 45°. False negatives locate at a range of slope angles, with specifically large critical error magnitudes (Table 4.6) at low angles.



(a) (b)

Figure 5.15: An overview of all modeled cells with the parameter combination "f-c". In the case of perfect prediction, real positives (landslide cells) and real negatives would locate on each side of the line representing FS=1. Cells of slope angle below 15° are given FS=1 due to chosen settings in TRIGRS. (a) Full overview and (b) zoomed-in overview exploring behaviour close to T_{FS}.

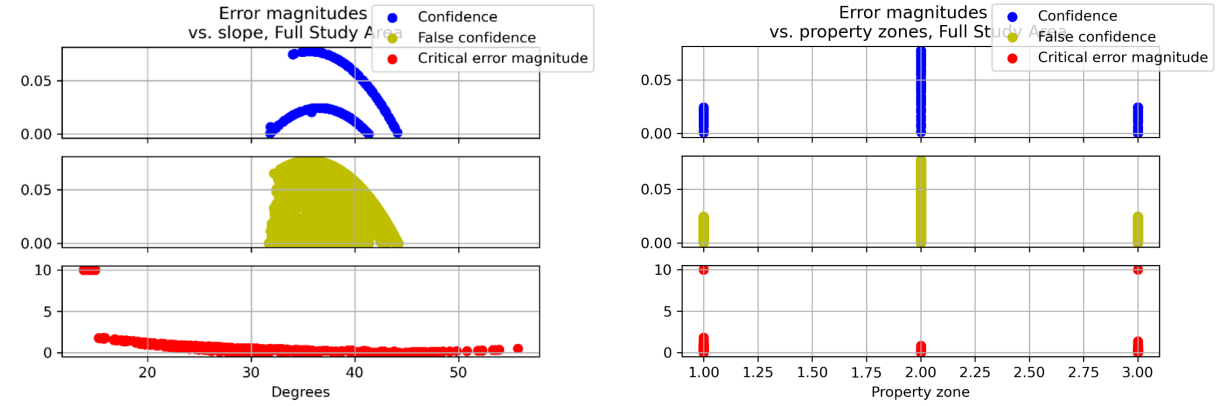
Both true and false negatives are generally predicted on low slope angles (Fig. 5.16a). The majority of true positives occur in landslide material, while most false positives and false negatives locate in till (Fig. 5.16b). Few true negatives occur in landslide material.



(a) (b)

Figure 5.16: Predictions by parameter combination "f-c" (TP=true positives, FP=false positives etc.) vs. (a) slope angles and (b) property zones (1=till, 2=landslide material, 3=bare rock *properties as till*).

Fig 5.17 shows error magnitudes in relation to spatial data. Confidence and false confidence quickly reduces away from its maximum at slope angles of ca. 35°, while critical error magnitude increases with decreased slope angle. The former two are considerably higher in landslide material than till.



(a) (b)

Figure 5.17: Scatter plots of error magnitudes obtained by parameter combination "f-c" vs. (a) slope angles and (b) property zones (1=till, 2=landslide material, 3=bare rock with properties as till)

Error magnitudes may also be visualized on maps. An example is shown Fig. 5.18, providing the distribution of "false confidence" for the parameter combination "f-c". Areas of under-prediction of stability are identifiable – specifically in relation to landslide material below steep cliffs.

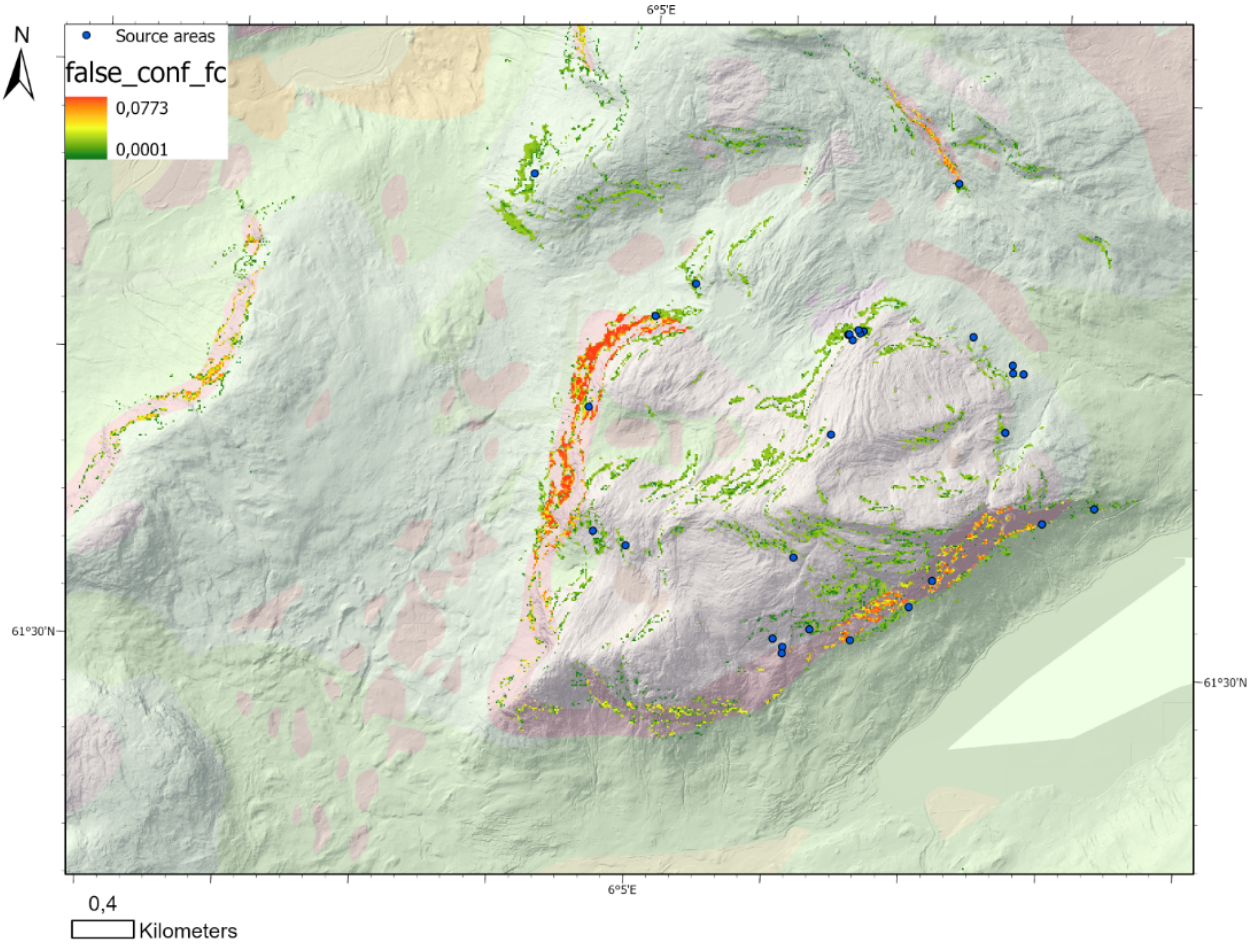


Figure 5.18: False confidence ($1 - FS_{FP}$) in the Tindefjellet area for the parameter combination "f-c".

Chapter 6

Discussion

6.1 Where, how and to what degree has TRIGRS been applied?

The literature review provided interesting results; some are discussed in this section, while others are referred to throughout other parts of Chapter 6 when appropriate.

An impressive number of papers were identified, especially during the last 3-5 years. TRIGRS' simplicity, availability and recognition, makes it among the most commonly used physically based model, and following the trend, numerous applications in new areas are expected in the near future. Practical usage appears challenging, limited to one paper.

Different methods of reporting geological and climatic conditions are represented. Standardization would be appreciated, but would likely obscure each area's specific features too severely. Applications also show notable variation in sources, reliability and diligence of input data. The estimation of soil properties ranges from crude assumptions to detailed field studies. Many applications devote surprisingly little effort to the matter, and a detailed review of e.g., test densities or qualities would be interesting. Soil depth is often assumed uniform, while approaches to distributed soil modelling appear rather arbitrary; the number of measurements is typically limited, and relationships to explanatory variables are sometimes assumed because other's have before. Nonetheless, it might be argued that even uncertain distributed soil modelling is superior to uniform depth in most cases. Stabilizing effects of vegetation is largely neglected, while surprising is its lack of discussion in most papers. Although Godt et al. (2008) argued that applying e.g., uniform root cohesion is physically incorrect, avoidance of the issue is strange compared to the rigorousness in other areas. A number of papers leave crucial and less crucial inputs unmentioned.

The second research questions considered TRIGRS' performance, and the underlying question "under which conditions does TRIGRS perform well?". Unfortunately, any concise answer was unattainable due to a remarkable variation in evaluation practices. Comparing results from different evaluation criteria and correlating these to e.g., the reliability of soil properties would be a major task, while as argued in Section 6.3, comparing results even from

similar criteria may be unachievable without specific considerations. Still, some standardization (e.g., as recently for landslide early warning systems, see Piciullo et al., 2020) may increase model understanding and help guide best practices.

6.2 Back-calculation of soil parameters for the Jølster case study

Soil parameter back-calculation may be separated in two: *objective* back-calculation identifies actual physical properties (e.g., angle of internal friction) while *relative* back-calculation identifies model-specific properties not necessarily with any physical meaning. As little is known about the spatial distribution of soil properties in Jølster and more generally, Norway, both were of interest for this thesis.

6.2.1 What limits soil parameter back-calculation?

However, several issues limit both objective and relative back-calculation. Assumptions that must be satisfied include:

- TRIGRS is sufficiently physically correct and consistent.
- Parameter calibration is well executed.
- Un-calibrated input data (e.g., topography, soil depth) is accurate.
- Landslides are represented with good spatial and temporal accuracy, and initiated through modellable mechanisms.

Limitations of the TRIGRS model

TRIGRS is a *model* – not an accurate representation of reality. Simplifications in procedures for computing runoff, infiltration, pore pressures and slope stability are discussed elsewhere (Baum et al., 2008; Iverson, 2000), but selected issues encountered in application to the Jølster case study are also discussed here.

For infiltration, assumptions regarding landslide thickness being less than the square root of the upslope contributing area and rainfall duration being much shorter than the steady state groundwater response time (Iverson, 2000) are considered appropriate for the Jølster case study. By contrast, no lateral groundwater flow is unrealistic, but according to Baum et al. (2008), this produces only minor adverse effects for intense, short-duration storms. Runoff computations do not include rate of overland flow, preventing modelling of landslide initiation through surface erosion.

The infinite slope stability model (Equation 2.5) includes, as the name suggests, assumptions on the length of slopes. Milledge et al. (2012) found that low length/depth (L/D)–ratios for source areas would usually yield too high estimate factor of safety, and that ratios >25 were appropriate. L/D ratios for Jølster landslides are typically 15–50, but certain outliers are present. The infinite slope stability model also neglects cell-to-cell forces, contributing to conservativeness as resisting forces are usually larger than driving forces (Stancanelli et al., 2017). On the other hand, less conservativeness would be obtained in the unlikely case of Jølster landslides being non-planar failures (Baum et al., 2008).

Overall effects from model simplifications on parameter back-calculations within the model are difficult to quantify, but doubtless, accurate input data does not guarantee accurate results and vice versa.

Limitations of the calibration procedure

Only four parameters, assumed the most sensitive based on analyses in literature (e.g., Melchiorre and Frattini, 2012, Zieher et al., 2017) were varied during saturated calibration in this thesis. The choice of keeping all other parameters constant was made mainly due to computational requirements, and their inclusion may have altered both estimated parameters and model performance. The potential for equifinality (see e.g., Foulon and Rousseau, 2018), in which e.g., high C may compensate for low ϕ or vice versa, is present both among calibrated parameters and due to ill-assumed, un-calibrated parameters.

Parameter ranges were loosely set based on values reported by Melchiorre and Frattini (2012) and Opsal (2017), although the use of any reference table would in hindsight be sufficient. The appropriateness of ranges is uncertain, as is whether certain combinations (e.g., high K_S, low D_0) are physically possible. Ranges did not cover the desired stability conditions, with e.g., D2PC being far from maximized for parameter combination "f-d" (Table 5.6).

Sampling intervals within parameter ranges, chosen partly to abide certain script limitations, were larger than desired. Optimal parameter combinations were therefore likely missed, and smaller intervals would have improved "fine-tuning". Spatial intervals of model output (1h) were also large (landslides may have occurred at e.g., 16:30), but considered acceptable due to unknown landslide initiation times.

The choice of which evaluation criteria would govern back-calculation is an important issue. As discussed in Section 6.3, different criteria consider different aspects of performance, while in theory, back-calculation of one landslide requires only one criterion (does factor of safety drop below 1 at the time of initiation?). For Årnes, the latter was approximately fulfilled (Fig. 5.9b), but for Slåtten, the attempted calibration of the entire sub-area simultaneously means estimated parameters represent unaccounted trade-offs between true and false positives decided by the criteria applied. The association between landslide material and soil parameters of Fig. 5.5 is therefore questionable. Optimally, each source area would be back-calculated separately, but largely due to the variation in slope angles discussed in Section 6.4.1, this would have resulted in vastly different estimated parameter combinations of which usefulness is dubious.

Limitations of the digital topography

Slope angles are crucial in slope stability modelling; these are relevant in hydrological and stability calculations as well as soil depth predictions, and their uncertainties may affect factor of safety comparably to those of pore water pressure and friction angle (Haneberg, 2006).

The accuracy of DEMs typically increases with resolution (Sarma et al., 2020; Fig. 5.1), while a wide range of resolutions are applied in literature (Fig. 3.3d). Therefore, maximum resolution (1*1m) – especially in areas adjacent to steep slopes such as Slåtten – would have been preferential for parameter back-calculation. This was impractical due to computational requirements, and the less accurate 10*10m DEM was considered an adequate substitute. In fact, for the ability to predict landslides per se, even 50*50m may have been satisfactory; too "localized" topography does not always best represents landslide-initiating factors (Catani et al., 2010), and instances are reported where lowering resolution to e.g., 30*30m improves modelling performance (e.g., Cama et al., 2016)). Evidently, no "per-

fect" resolution exists (Claessens et al., 2005), but certainly, the relatively crude resolution applied in this thesis' back-calculation induces errors due to its inability to capture local, important features.

Soil depth predictions by a simple approach

Soil depth strongly influences the computation of factor of safety even at small ranges of variation (Melchiorre & Frattini, 2012). Generally, increased soil depth yields decreased factor of safety, but the opposite may also occur depending on interactions with other parameters (Melchiorre & Frattini, 2012).

Negative exponential or linear relationships between slope angle and soil depth lack considerations of specific processes within geomorphological units (Catani et al., 2010), but are common in literature (Fig. 5.2). For the regression analysis in Section 5.2, R^2 = 30% indicates a weak, negative relationship is present also for Jølster source areas. The strength of correlation is unimpressive, but its comparison to others is uncertain; relationships are often assumed without basis in field tests, while those that are, are seldom reported in terms of R^2. Significantly improved performance may have been obtained using e.g., the "Geomorphologically Indexed Soil Thickness" (GIST) model (Catani et al., 2010) or a more deliberate empirical approach (e.g., Tesfa et al., 2009), although this would likely require more data than currently available. In any case, the identified empirical relationship predicts unrealistically low soil depths at low slope angles (largely irrelevant for modelling), but succeeds in predicting little or no soil on steep slopes not prone to failure. Errors are inevitable, and e.g., too high predicted soil depth likely overestimates soil strength during back-calculation. The relationship is statistically weak due to an insufficient sample size, as well as measurements being source area *averages*. Also, measurements may represent unfortunate combinations of soil depth, slope angle and other factors leading to failure, rather than the actual distribution of soil depth. The spatial distribution of soil depth in Norway is a topic for further research, and may include field tests, various (combinations of) explanatory variables and perhaps a process-based approach considering e.g., glacial propagation and different sedimentation regimes.

What about vegetation?

Vegetation generally increases slope stability, but as sown in the literature review, only a few authors mention this issue in relation to TRIGRS (Fig. 5.1d). Therefore, although detailed spatial data was available (Section 4.2.5), the lack of examples discouraged considerations of vegetation in this thesis as well. However, Jølster's slopes below 600–700 m.a.s.l. are largely vegetated, and Fig. 5.1d indicates vegetation *may* have contributed to stability in the 2019 event; few landslides occurred in e.g., spruce, but as this may be a result of the spatial distribution of spruce in relation to e.g., soil types and slope angles, conclusions cannot be drawn. Also telling is the fact that a majority of landslides occurred in "no tress or limited vegetation (Fig. 5.1d).

Jølster slopes are, regardless, treated as completely unvegetated by TRIGRS; applying e.g., root cohesion at specific areas may, although discouraged by Godt et al. (2008), have reduced the number of false positives (e.g., Park et al., 2013)) and prevented the likely estimation of too high soil strength. As a side-note, quantitative information on the stabilizing effects of typical Norwegian vegetation appears scarce, and may be a topic for further research.

Saturated vs. unsaturated analysis and assumptions on initial conditions

The choice of analysis method is based on assumed initial conditions, and may alter estimated parameters and TRIGRS' ability to identify stable and unstable areas. The unsaturated version performed much poorer in sub-areas compared to the saturated version (Table 5.7), and was therefore not pursued for modelling the full study area. Especially for Årnes (Section 5.4.2), the combination of stabilizing suction and low effective K_S in unsaturated soil resulted in almost constant FS with time (Fig. 5.13b). As a consequence, low cohesion (600– 800 Pa) was required for identifying any landslides, causing failure at nearly all steep slopes even before rainfall. The unsaturated version performed better for Slåtten, but still worse than the saturated version across all measures. A testimony to poor performance was the generally decreasing AUC in time for both study areas; transient rainfall infiltration *reduced* TRIGRS' ability to rank-order cells and although neglecting timing, better performance may have been obtained with nothing but adjustment of T_{FS}. Acceptable unsaturated performance has been obtained by others (e.g, Baum et al., 2008; Schilirò et al., 2015), but no instances comparing both versions have been identified.

Why did the saturated analysis yield superior results? Unsaturated analysis is not unconditionally better (Luca Schilirò, personal communication, March 10, 2021), requiring suitable input data and case studies. As emphasized by Baum et al. (2008):

> "Successful use of the unsaturated model requires initial testing on single-cell grids to learn model behaviour. Results of these tests will guide selection of program control parameters, such as nmax, nzs, and tx, as well as ranges of physical properties to achieve accurate results with efficient computation." (Rephrased).

Extensive testing was beyond the scope of this thesis, and applied unsaturated parameters (Table 4.2) are of uncertain validity due to unrepresentability of the till sample and uncertainties in Rosetta Lite. In fact, the complexity increase caused by addition of unsaturated parameters may have been undesirable considering the scarce knowledge of soil properties.

TRIGRS is sensitive to initial conditions (Baum et al., 2008), and the dry initial conditions assumed by unsaturated analysis appear valid according to Sections 4.3.3 and 4.3.4. However, completely unsaturated conditions are rarely achieved in Norway, and partial saturation in combination with modest rainfall may explain the lack of change in pore pressures and FS observed for Årnes. Locating the groundwater level at the lower soil boundary may also be too un-conservative. Conversely, the relatively high applied background infiltration rate is conservative, but its effect on model behavior remains uncertain.

In any case, TRIGRS' unsaturated version appears – despite unusually dry conditions – unsuited for this thesis' specific study area, precipitation regime and data availability. whether applicability is better for other Norwegian events is a topic for further research that should include single cell testing, in situ soil data and a small, well-understood study area.

Other sources of error

Unmodellable landslide initiation mechanisms may be an important factor limiting the accuracy of property back-calculation (see discussion in Section 6.4.2). Moreover, the applied radar-estimated precipitation (discussed by Sandvoll, 2020) may contain inaccuracies due to low spatial (250*250m) and temporal (1h) resolution.

6.2.2 Can objective soil parameters be estimated?

For reasons discussed in the previous section, estimated parameters from saturated analysis are not given much physical meaning. Although estimated parameters (Tables 5.4, 5.5) are within expected ranges for the assumed material types (e.g., M. van Genuchten et al., 1991), this may be a result of luck. Comparisons should accordingly be made to calibrated (e.g., Melchiorre and Frattini, 2012) rather than reported (e.g., Opsal, 2017) soil properties. In essence, accurate modelling results do not guarantee accurate inputs, which conversely questions meticulous efforts in obtaining accurate input data.

6.2.3 Can relative soil parameters be estimated?

Although objective parameters appear inestimable, TRIGRS may still be useful in identifying relative differences in properties between soil types. These differences may refer to specific parameters, or a general "soil stability" (the combined interaction of soil parameters, soil depth etc.). In the case of their existence, associating parameters to soil types (Årnes = till, Slåtten = landslide material) and transferring these to the full study area was expected to improve modelling compared to single sub-area of full study-area calibration.

Different soil properties were estimated for sub-areas (Tables 5.4 & 5.5). This *may* indicate soil type differences, but may also be caused by a sub-optimal calibration procedure for Slåtten's multiple landslides (Section 6.2.1). In fact, parameter variability within landslide material may be larger than between landslide material and till, and the definition of one set of parameters for each soil type may be too crude of an arrangement.

Årnes-parameters applied for the full study area gave a model more unstable at 17:00, more stable at 21:00 and with generally lower TP/FP-ratios compared to "f-d" and "f-c"; i.e., depending on evaluation criteria, using only the Årnes area for calibration may have been both "better" and "worse" compared to the proposed two-area approach. Hypothetical results of calibration in other single sub-areas (e.g., Slåtten or Tindefjellet) are unknown, and would again depend on evaluation criteria. Calibration using the full study area was unachievable due to computational requirements, and whether this would improve performance and optimize TRIGRS for e.g., warning purposes is uncertain.

Despite being unidentifiable by calibration, indications of parameter differences between soil types may still be present. High slope angles for landslide material source areas indicate higher "stability" than till, adhering to the assumption of coarse-grained material originating from rockfall. High false confidence in distinct zones of landslide material below cliffs (Fig. 5.18) and "half-circles" of Fig. 5.17a representing different soil types may indicate likewise, but may as well be a function of "f-c"-parameters providing too little stability.

6.3 On the choice of evaluation criteria

Assessing "quality" and performance of physically based models is important (Corominas & Mavrouli, 2011). The choice of one or more criteria decides which model feature receives attention and which parameters are back-calculated; e.g., should one maximize the number of landslides identified, or minimize the number of false positives? As this thesis' literature review identified a number of differing practices (Fig. 3.5a), a discussion on the matter was considered necessary before and during application of TRIGRS. The following section is not at all conclusive, but based on relevant theory and results obtained in Section 5.3. It considers performance measurement of PBMs in general, selected criteria tested or applied in this thesis, and argues in favor of a more careful selection of relevant measures.

6.3.1 Performance in general

What, in a general sense, does one wish to measure in physically based landslide modelling? An important distinction lies between what was named *overall* and *threshold-specific* performance (Section 4.4.1). "Objective" back-calculation of soil properties (discussion in Section 6.2.2) requires threshold-specific ($T_{FS} = 1$) evaluation, while its "relative" counterpart (discussion in Section 6.2.3) may do with any T_{FS}. Likewise, in calibration for e.g., warning purposes, T_{FS} may simply be adjusted in order to suit specific needs. Although the overall performance of a predictor may be important in some circumstances and may indicate good threshold-specific performance, the latter is not necessarily true as clearly shown in Section 6.3.2. Similarly, the distinction between *objective* performance (absolute results) and *relative* performance (to other results) is important. The former is the sole part that should be reported and e.g., compared between study areas, while the latter is suitable only for performance of models or parameter sets within similar study areas, input data qualities etc. However, as discussed below, "objective" evaluation in physically based modelling is no easy task.

6.3.2 On ROC and related performance measures

ROC, with related measures AUC and D2PC (Section 2.4), is the commonest evaluation approach identified in literature (Fig. 3.5a). Nevertheless, one of its fundamental properties deserves extra consideration:

- False positive rate (FPR) is dependent on the number of true negatives. This number is again dependent on the area under consideration, so that inclusion of valley floors or altitude plains in analysis – which correctly predicted stability is trivial – reduces FPR and increases apparent performance.

In physically based landslide modelling, the number of true negatives is typically large, with limited numbers of landslides and susceptible slopes (e.g. Schilirò et al., 2021). Consequently, as for this thesis (e.g., Table 5.5), impressive but disproportionate values of AUC are reported (Table 3.5b). Although an ROC-graph considers fractions and is insensitive to changes in the proportion of positives to negatives (P/N) (Fawcett, 2006), it is to changes in proportions FP/N. With reference described in Section 6.3.1, AUC is not "objective"; it should not be reported or compared across study areas unless care is taken to explain the basis upon which it is used.

Also, AUC considers overall, not threshold-specific performance. As noted by Fawcett (2006), visually analyzing ROC-curves is important as performance may vary across the threshold range. For this case study, threshold-specific measures CSI and D2PC were found to generally improve with AUC (Fig. 5.8b), but considerable variation was present (e.g., AUC = 0.85 yields CSI between 0.0011 and 0.0086). Accordingly, although commonly reported *as if*, AUC should not be treated as indicative of threshold-specific performance and was, for the above reasons, not applied as an evaluation criterion for calibration in this thesis (Section 4.4.1).

D2PC is derived from the ROC-curve and is similarly not an objective measure. The high number of true negatives induces a bias; true positive are over-valued in performance estimation, and the strength of bias depends on the factor N/P. High N/P (which is typically the case) makes false positives largely irrelevant, while a lower N/P increasingly relates good performance to good TP/FP-ratio. In fact, D2PC increases as long as less than N/P (equal to 605 for the full study area of Section 5.3) false positives are required for each additional true positive. Implications were evident in Fig. 5.8a; D2PC effectively measured the number of true positives, unaffected by the number of false positives. Correspondingly, maximizing D2PC by up-adjusting T_{FS} correctly predicts the majority of landslides but also a remarkable number of false positives (Fig. 5.7). I.e., the relationship between marginal costs and benefits of false and true positives respectively is case-dependent, and this should be clearly accounted for when the measure is applied.

Despite shortcomings, D2PC was chosen as one of two evaluation criteria for this thesis' back-calculation. As an attempt to reduce the effect of true negatives, the term critical cells (N*) was defined. The difference between D2PC* and D2PC depends on N*/N, and as accordingly, slightly worse performance was indicated using the former (Fig. 5.8a, Table 5.2). The magnitude of reduction in estimated performance increased as TPR approached its maximum, as certain source areas were outliers requiring excessive false positives in order to be identified. The removal of unconditionally stable cells – which definition should be performed with greater care than here – may be considered an improvement, but only through increasing the desired/optimal TP/FP-ratios and without solving the underlying problem.

6.3.3 On other, selected criteria

CSI, which contrary to ROC is not influenced by true negatives, was chosen as the other, threshold-specific evaluation criterion. Within landslide modelling, where false positives by far outnumber true positives and false negatives, it practically measures the TP/FP-ratio (using CSI in favor of TP/FP was, in hindsight, unnecessary). CSI is a counterweight to D2PC, evident in terms of identified parameter sets and their relative rankings (Table 5.2).

MCC was also applied as an alternative measure, and may be considered a moderate tradeoff between true and false positives (Fig. 5.8a). It will not be further discussed, but may have interesting potential in this context and readers are referred to Matthews (1975) and e.g., Chicco and Jurman (2020).

6.4 Is the Jølster case study suitable for the application of TRIGRS?

TRIGRS' ability to separate stable and unstable areas, at correct times, governs the accuracy of (relative) parameter back-calculation and the model's potential for practical utility in e.g., susceptibility assessments and warning. On a large scale, parameter combinations presented in this study perform reasonably well in identifying unstable areas (Figs. 5.11 & 5.12). However, on a cell-by-cell basis, performance is limited by two, interrelated challenges:

- High false positive rates are inevitable when aiming for an acceptable number of correctly predicted landslides
- Numerous false negatives, of high "critical error magnitude", occur unless extreme conservativeness is applied

A high false positive rate (FPR = FP/P) is, in other words, the price paid to minimize false negative rate (FNR = FN/P). This is commonly observed in literature: as an example, Zieher et al., (2017) predicted 73% and 95% of landslides in two separate events, accompanied by FPRs of ca. 30–40%. For comparison, this study's parameter combination "f-c" identified 32% of landslides (which may or may not be acceptable depending on goal), requiring an FPR of 7.6%. Importantly, this was the combination yielding optimal performance in terms of TP/FP - most modeled parameter sets performed much worse. The phenomena is mainly explained by two features:

- Nearly uniform soil properties and the prediction of soil depth makes slope angle a too strong predictor of factor of safety.
- Jølster landslides do not show a similar relationship to slope angle as encoded in TRIGRS, likely due to initiation mechanisms not accounted for.

6.4.1 Nearly uniform soil properties and dependency on slope angles

NGU soil maps were the only basis for spatially varying soil properties in the Jølster case study. Soil maps are commonly utilized in literature (Fig. 3.4b), and their application is considered an improvement compared to uniform properties. Also, the applied soil maps are largely intended for land use planning and show dominating types in upper layers only (NGU, 2017), making reliability uncertain. In fact, Rouault (2020) showed how source areas often locate within the "bare rock" zone due to the existence of soil pockets above bedrock; ths was also the case for Jølster, leading to all mountainous areas being treated as till and effectively reducing the number of relevant soil types to only two. Assumed homogeneity within soil types, the neglection of vegetation and the same assumed relationship between soil depth and slope angle for the entire study meant *nearly* uniform soil properties were applied. However, these are inherently variable even over small areas, and as typically in physically based landslide modelling (Stancanelli et al., 2017), limited knowledge of their distribution impair modelling performance also here.

Nearly uniform soil properties eliminate much variance in estimated stability between slopes. Fig. 5.15a clearly demonstrates how factor of safety is almost entirely dependent on slope angle. This is not surprising, and consistent with findings in parameter sensitivity analyses (e.g., Melchiorre and Frattini, 2012; Zieher et al., 2017). In practice, although spatially distributed precipitation and topography/drainage effects provide some variation, only a fraction of slopes (ca. $30° - -45°$) have the *potential* to be modeled as unstable. Source areas that fall outside this "instability range", of which shape, size and location is altered by assumptions on soil properties, soil depth and the chosen T_{FS}, are accordingly left unidentifiable. Implications were evident in Årnes' unsaturated analysis (Section 5.4.2), where in order to correctly predict two source area cells, cohesion (and the instability range's lower slope angle limit) was brought down until most slopes angles higher than the source areas' were given FS<1.

6.4.2 Are Jølster landslides modellable?

If slope angles of source areas were within the aforementioned "instability range", fewer incorrect predictions would occur. But, this is not the case; they vary widely in (although generally within ranges reported in literature for shallow landslides) (Section 5.1). Are mechanisms not accounted for in TRIGRS responsible for such variation?

TRIGRS considers initiation by ideal infiltration and pore pressure increase only. Although it is assumed that no extreme landslide event in Norway has been spatially mapped in similar detail (Section 4.2.1), little is known about Jølster landslides' initiation mechanisms. They have been classified by others (Lindsay et al., in prep.; Rüther et al., in prep.), but as both the Norwegian and international classification are based on motion rather than initiation, little relevant information can be inferred. In fact, assessing initiation mechanisms may be impossible even through field studies, and the type of landslide may change as propagation proceeds; e.g., debris flows may initiate by either slope-toe erosion, gradual material entrainment (Welcker, 2011) or debris slides (Hungr et al., 2014). Further, terms "jordskred" and "flomskred" overlap (Jensen et al., 2015), and in practice this thesis' case study includes unknown proportions of debris slides, debris avalanches, debris flows and "debris floods" (Hungr et al., 2001). Accordingly, Sandvoll (2020) and Rüther et al. (in prep.) found mechanisms such as surface erosion, cracking and locally high pore pressures likely initiators for a proportion of landslides, and although neither the statistical analysis of source areas (Section 5.1) nor results from modelling (Section 5.4) can validate this, they indicate the presence of *some* alternative mechanisms. These may explain impressive error magnitudes (Fig. 5.15, several source areas are given FS>2) and accordingly, the attempted back-calculation of such outliers results in soil properties of little relation to reality along with high false positive rates.

Outliers exist despite that to limit adverse effects of variable initiation, "unsuitable" source areas were manually removed (Section 4.2.1). Some were with confidence (exclusion of "river erosion" and "washed-away grass mats"), while others were not (locations within existing channels do not guarantee initiation by erosion). The neglection of relevant source areas and vice versa is therefore likely, but the approach is still considered an improvement compared to modelling all available landslides.

6.4.3 TRIGRS practical utility in Jølster and similar events

Issues discussed in previous sections limit TRIGRS' potential for practical utility. Identified model parameters should not be applied elsewhere without extensive testing. Calibration of a physically based model is highly case specific; geotechnical and hydraulic parameters have uncertain physical validity, and the relationship between slope angle and soil depth is not tested elsewhere in Norway.

For assessing susceptibility in the Jølster study area, the estimated parameter combinations identify generally exposed slopes - although this is mostly a range of slope angles. However, numerous landslides are unidentified (only identified by extreme conservativeness) and there is little or no relationship between a given factor of safety and probability of failure. Interesting insights may have been obtained through the application of rainfall events of varying statistically estimated return periods (e.g., Peres & Cancelliere, 2016), but whether usefulness of results would be more than those of simpler statistical methods (data-driven methods, Section 2.3.3), is uncertain and a topic for further research.

Detailed warning is not desirable due to similar issues. More potential may be found in large-scale, area-wise warning where performance in terms of typical evaluation criteria comes second to the prediction of *any* occurrence of landslides. It may be speculated that an extensively tested and calibrated TRIGRS model – which e.g., indicates landslide activity when a certain number of cells are given FS < 1 – could be a better alternative than today's warning systems (typically combinations of antecedent and acute precipitation), but such an approach being worth the added effort is unlikely.

Chapter 7

Conclusions

A comprehensive literature review for the period 2008—2020 was conducted, identifying 42 papers applying TRIGRS in 12 countries. The frequency of applications has increased in recent years, and the model is applied on a range of study area extents and qualities of input data. A concise assessment of TRIGRS' performance is at the moment unachievable; results are case-specific, while effort in and approaches to evaluation vary considerably.

Soil parameter back-calculation through calibration was performed in two separate sub-areas assumed representative of specific soil types. Back-calculation of "objective" soil parameters is unachievable due to limitations in TRIGRS, the chosen calibration procedure, inaccuracies in input data, the neglection of unsaturated soil mechanics and uncertain landslide characteristics. Estimated parameters are therefore not given much physical meaning.

Back-calculation of "relative" soil parameters is challenging due to similar limitations. Especially, the attempted back-calculation of Slåtten's numerous landslides and consequent definition of a single, representative parameter combination – adhering to a specific evaluation criterion– was in hindsight not ideal. Estimated parameter differences between soil types are therefore attributed to issues in data and calibration more than actual physical conditions – although, higher slope angles of landslide material source areas do indicate increased "strength" compared to till. Even though calibration in two, soil type-representative sub-areas is subject to limitations, and did not improve full study area modeling performance compared to e.g., calibration in one sub-area only, the approach may be viable in cases of more reliable data.

Soil depth is crucial in physically based modeling, but often assumed uniform across large areas. For the Jølster case study, a weak correlation between soil depths and slope angles of source areas was indicated. Identified negative linear and exponential relationships (R^2 = ca. 30%) are consistent with findings in literature, but only 21 measurements limit statistical power. Using distributed soil depth modeling in the application of TRIGRS likely improves results, especially on steep slopes where soil depth is limited. Soil depth prediction by various approaches and explanatory variables is a topic of further research.

TRIGRS' unsaturated version appears, despite unusually dry initial conditions, unsuitable for the Jølster case study. Considerably poorer performance was obtained compared to the saturated analysis, with little change in factor of safety, a gradual decrease in AUC and under-estimation of stability on steep slopes. This is attributed

to an inappropriate precipitation regime (for one sub-area), a lack of testing for optimal combinations of model parameters, wrongly assumed initial conditions and knowledge of soil properties not fit for the added complexity involved in unsaturated analysis.

Two final parameter combinations for the full study area were presented, optimizing evaluation criteria “distance to perfect classification” (D2PC) and “critical success index” (CSI) respectively. Predictions from one combination were then subject to assessment through a – to my knowledge – novel approaches of spatial distributions and calculation of error magnitudes. Both combinations perform reasonably well on a large scale, identifying general areas where landslides of the case event occurred. Accordingly, TRIGRS may be calibrated to indicate landsliding at certain levels of predicted instability, but if such approach would trump statistical methods for susceptibility assessment and rainfall thresholds for warning is unknown. On a smaller scale, assumed homogeneity of soil type properties, groundwater table locations and soil depth distributions leaves slope angle a strong influence on estimated factor of safety. As slope angle varies considerably for Jølster source areas, a tradeoff is inevitable between a) identification of landslide cells, along with high false positive rates or b) low false positive rates, at the cost of false negatives of high “critical error magnitudes”. Slope angle variations of source areas and research by others indicate a proportion of Jølster landslides were initiated by e.g., small-scale surface erosion, local cracking and preferential flow – mechanisms not accounted for by TRIGRS.

Numerous approaches to performance evaluation were identified in literature, with the commonest being "receiver operating characteristics" (ROC) and related measures. To explore how the choice of evaluation criteria affects the notion of "good" performance, parameter calibration using "area under the curve" (AUC), D2PC, CSI and “Matthews’ correlation coefficient” (MCC) was conducted on the Jølster case study area with the following findings:

- Overall (through a range of cut-off values) and threshold-specific (e.g., at factor of safety = 1) performance must be distinguished; although AUC indicates threshold-specific performance in terms of e.g., D2PC, CSI and MCC, it cannot be considered a valid proxy as considerable variation occurs.

- Measured performance through evaluation criteria is not objective, but dependent on characteristics of the study area, landslides and input data. They should accordingly not be reported or compared unless care is taken to account for specific circumstances. AUC and D2PC involve a bias depending on the ratio of negative to positive cells (N/P); this may effectively result in it measuring true positive rate only and induce artificially high scores. The experimental definition of “critical cells” and the altered D2PC* reduces this bias but does not solve the underlying problem.

- Evaluation criteria such as AUC, D2PC, CSI and MCC measure different aspects of performance, and may be applied in conjunction (with other criteria as well) for a broad assessment of results.

In general, the importance of specific research goals and the deliberate choice of suitable evaluation criteria is stressed. Improved practices – perhaps through standardization – would likely be a helpful step in further advancing the field of physically based landslide modelling.

References

Agersten, S., Andersen, A. S. H., Berger, A., Dyrrdal, A. V., Køltzow, M., & Tunheim, K. (2019). *Intense byger med storekonsekvenser i Sogn og Fjordane 30. juli 2019* (tech. rep.). MET. Bergen.

Al-Umar, M., Fall, M., & Daneshfar, B. (2020). GIS-based modeling of snowmelt-induced landslide susceptibility of sensitive marine clays. *Geoenvironmental Disasters*, *7*(1). https://doi.org/10.1186/s40677-020-0142-8

Alvioli, M., & Baum, R. L. (2016). Parallelization of the TRIGRS model for rainfall-induced landslides using the message passing interface. *Environmental Modelling and Software*, *81*, 122–135. https://doi.org/10.1016/j.envsoft.2016.04.002

Alvioli, M., Guzzetti, F., & Rossi, M. (2014). Scaling properties of rainfall induced landslides predicted by a physically based model. *Geomorphology*, *213*, 38–47. https://doi.org/10.1016/j.geomorph.2013.12.039

Alvioli, M., Melillo, M., Guzzetti, F., Rossi, M., Palazzi, E., von Hardenberg, J., Brunetti, M. T., & Peruccacci, S. (2018). Implications of climate change on landslide hazard in Central Italy. *Science of the Total Environment*, *630*, 1528–1543. https://doi.org/10.1016/j.scitotenv.2018.02.315

An, H., Viet, T. T., Lee, G., Kim, Y., Kim, M., Noh, S., & Noh, J. (2016). Development of time-variant landslide-prediction software considering three-dimensional subsurface unsaturated flow. *Environmental Modelling and Software*, *85*, 172–183. https://doi.org/10.1016/j.envsoft.2016.08.009

Andersson, A., Lambrechts, L., & Nord, B. A. (2018). Ber Søviknes om 800 mill. ekstra til skred- og flomsikring. https://www.bt.no/nyheter/innenriks/i/dd8dpB/ber-soeviknes-om-800-mill-ekstra-til-skred-og-flomsikring

Bargel, T. H., Blikra, L. H., Høst, J., Sletten, K., & Stalsberg, K. (2008). Graasteinen12_06. *Gråasteinen - geologi for samfunnet i 150 år - arven etter kjerulf*. NGU.

Baum, R. L., Savage, W. Z., & Godt, J. W. (2002). *TRIGRS-A Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis* (tech. rep.). USGS. Golden, Colorado.

Baum, R. L., Savage, W. Z., & Godt, J. W. (2008). *TRIGRS — A Fortran Program for Transient Rainfall Infiltration and Grid-Based Regional Slope-Stability Analysis, Version 2.0* (tech. rep. No. 2008-1159). USGS. Reston, Virginia. https://doi.org/Open-FileReport2008âĂŞ1159

Bjordal, H., & Helle, T. E. (2011). *Skred og flom på veg: Statistiske betraktninger* (tech. rep.). Statens Vegvesen. http://www.vegvesen.no/klimaogtransport

Bordoni, M., Meisina, C., Valentino, R., Bittelli, M., & Chersich, S. (2015). Site-specific to local-scale shallow landslides triggering zones assessment using TRIGRS. *Natural Hazards and Earth System Sciences*, *15*(5), 1025–1050. https://doi.org/10.5194/nhess-15-1025-2015

Bryhni, I., & Askheim, S. (2020). Loenulykkene. https://snl.no/Loenulykkene

Cama, M., Conoscenti, C., Lombardo, L., & Rotigliano, E. (2016). Exploring relationships between grid cell size and accuracy for debris-flow susceptibility models: a test in the Giampilieri catchment (Sicily, Italy). *Environmental Earth Sciences*, *75*(3), 1–21. https://doi.org/10.1007/s12665-015-5047-6

Carrara, A., Crosta, G., & Frattini, P. (2008). Comparing models of debris-flow susceptibility in the alpine environment. *Geomorphology*, *94*(3-4), 353–378. https://doi.org/10.1016/j.geomorph.2006.10.033

Catani, F., Segoni, S., & Falorni, G. (2010). An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. *Water Resources Research*, *46*(5), 1–15. https://doi.org/10.1029/2008WR007450

Chen, Y., Zhao, L., Wang, Y., Jiang, Q., & Qi, D. (2019). Precipitation data and their uncertainty as input for rainfall-induced shallow landslide models. *Frontiers of Earth Science*, *13*(4), 695–704. https://doi.org/10.1007/s11707-019-0791-7

Cheung, P., Wong, M., & Yeung, H. (2006). Application of Rainstorm Nowcast to Real-time Warning of Landslide Hazards in Hong Kong WMO PWS Workshop on Warnings of Real-Time Hazards. *WMO PWS Workshop on Warnings of Real-Time Hazards by Using Nowcasting Technology*.

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics*, *21*(1). https://doi.org/10.1186/s12864-019-6413-7

Chien, L. K., Hsu, C. F., & Yin, L. C. (2015). Warning model for shallow landslides induced by extreme rainfall. *Water (Switzerland)*, *7*(8), 4362–4384. https://doi.org/10.3390/w7084362

Ciurleo, M., Mandaglio, M. C., & Moraci, N. (2019). Landslide susceptibility assessment by TRIGRS in a frequently affected shallow instability area. *Landslides*, *6*, 175–188. https://doi.org/10.1007/s10346-018-1072-3

Claessens, L., Heuvelink, G. B., Schoorl, J. M., & Veldkamp, A. (2005). DEM resolution effects on shallow landslide hazard and soil redistribution modelling. *Earth Surface Processes and Landforms*, *30*(4), 461–477. https://doi.org/10.1002/esp.1155

Corominas, J., van Westen, C., Frattini, P., Cascini, L., Malet, J. P., Fotopoulou, S., Catani, F., Van Den Eeckhaut, M., Mavrouli, O., Agliardi, F., Pitilakis, K., Winter, M. G., Pastor, M., Ferlisi, S., Tofani, V., Hervás, J., & Smith, J. T. (2014). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, *73*(2), 209–263. https://doi.org/10.1007/s10064-013-0538-8

Corominas, J., & Mavrouli, O. (2011). *Recommended Procedures for Validating Landslide Hazard and Risk Models and Maps* (tech. rep.). Safeland.

Crosta, G. B., & Frattini, P. (2003). Distributed modelling of shallow landslides triggered by intense rainfall. *Natural Hazards and Earth System Science*, *3*(1), 81–93. https://doi.org/https://doi.org/10.5194/nhess-3-81-2003

Cruden, D. M. (2003). The first classification of landslides? *Environmental and Engineering Geoscience*, *9*(3), 197–200. https://doi.org/10.2113/9.3.197

Cruden, D. M., Thomson, S., Bornhold, B. D., Chagnon, J., Locat, J., Evans, S., Heginbottom, J., Moran, K., Piper, D., Powell, R., Prior, D., & Quigley, R. (1989). Landslides: Extent and economic significance in Canada. In Brabb & Harrod (Eds.), *Landslides: Extent and economic significance* (pp. 1–23). Balkema. https://www.researchgate.net/publication/328174712

Delmonaco, G., Leoni, G., Margottini, C., Puglisi, C., & Spizzichino, D. (2003). Large scale debris-flow hazard assessment: a geotechnical approach and GIS modelling. *Natural Hazards and Earth System Sciences*, *3*, 443–455. https://doi.org/https://doi.org/10.5194/nhess-3-443-2003

Det Kongelige Finansdepartement. (2014). Rundskriv R-109/14: Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv.

Dikshit, A., Satyam, N., & Pradhan, B. (2019). Estimation of Rainfall-Induced Landslides Using the TRIGRS Model. *Earth Systems and Environment*, *3*(3), 575–584. https://doi.org/10.1007/s41748-019-00125-w

Dingman, S. L. (2015). *Physical Hydrology: Third Edition*. Waveland Press.

Dowling, C. A., & Santi, P. M. (2014). Debris flows and their toll on human life: A global analysis of debris-flow fatalities from 1950 to 2011. *Natural Hazards*, *71*(1), 203–227. https://doi.org/10.1007/s11069-013-0907-4

Erickson, D. J., Oglesby, R. J., Elliott, S., Steffen, W., & Brasseur, G. (2008). Challenges in Earth System Modelling: Approaches and Applications. *Developments in integrated environmental assessment* (pp. 297–306). Elsevier. https://doi.org/10.1016/S1574-101X(08)00617-0

Esri Inc. (2021). ArcGIS Pro (Version 2.7.2). https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview.

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

Fleming, R. W., & Taylor, F. A. (1980). Estimating the Costs of Landslide Damage in the United States. *Geological Survey Circular*, *832*. https://doi.org/https://doi.org/10.3133/cir832

Flentje, P. N., Miner, A., Whitt, G., & Fell, R. (2007). *Guidelines for landslide susceptibility, hazard and risk zoning for land use Guidelines for landslide susceptibility, hazard and risk zoning for land use planning planning* (tech. rep.). Australian Geomechanics Society.

Foulon, É., & Rousseau, A. N. (2018). Equifinality and automatic calibration: What is the impact of hypothesizing an optimal parameter set on modelled hydrological processes? *Canadian Water Resources Journal*, *43*(1), 47–67. https://doi.org/10.1080/07011784.2018.1430620

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

Furseth, A. (2006). *Skredulykker i Norge*. Tun Forlag.

Gariano, S. L., & Guzzetti, F. (2016). Landslides in a changing climate. *Earth-Science Reviews*, *162*, 227–252. https://doi.org/10.1016/j.earscirev.2016.08.011

Geertsema, M. (2013). Quick Clay. In P. T. Bobrowsky (Ed.), *Encyclopedia of naturalhazards. encyclopedia of earth sciences series*. Springer. https://doi.org/https://doi.org/10.1007/978-1-4020-4399-4{_}282

Ghezzehei, T. A., Kneafsey, T. J., & Su, G. W. (2007). Correspondence of the Gardner and van Genuchten-Mualem relative permeability function parameters. *Water Resources Research, 43*(10). https://doi.org/10.1029/2006WR005339

Gioia, E., Speranza, G., Ferretti, M., Godt, J. W., Baum, R. L., & Marincioni, F. (2016). Application of a process-based shallow landslide hazard model over a broad area in Central Italy. *Landslides, 13*(5), 1197–1214. https://doi.org/10.1007/s10346-015-0670-6

Gjertsen, A. K., & Nilsen, J.-E. (2012). *SAT-SKOG: Et skogkart basert på tolking av satellittbilder* (tech. rep.). NIBIO.

Godt, J. W., Baum, R. L., Savage, W. Z., Salciarini, D., Schulz, W. H., & Harp, E. L. (2008). Transient deterministic shallow landslide modeling: Requirements for susceptibility and hazard assessments in a GIS framework. *Engineering Geology, 102*(3-4), 214–226. https://doi.org/10.1016/j.enggeo.2008.03.019

Grelle, G., Soriano, M., Revellino, P., Guerriero, L., Anderson, M. G., Diambra, A., Fiorillo, F., Esposito, L., Diodato, N., & Guadagno, F. M. (2014). Space-time prediction of rainfall-induced shallow landslides through a combined probabilistic/deterministic approach, optimized for initial water table conditions. *Bulletin of Engineering Geology and the Environment, 73*(3), 877–890. https://doi.org/10.1007/s10064-013-0546-8

Guerrero-Bote, V. P., & Moya-Anegón, F. (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of Informetrics, 6*(4), 674–688. https://doi.org/10.1016/j.joi.2012.07.001

Guha-Sapir, D. (2018). *2018 Review of Disaster Events, supplementary information* (tech. rep.). CRED. Brussels.

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*. https://doi.org/10.1016/j.earscirev.2019.102973

Haldorsen, S., & Jenssen, P. D. (1983). Some hydraulic properties of Sandy-silty Norweigan Tills. *Acta Geologica Hispanica*. https://www.researchgate.net/publication/39118230

Hammond, C., Hall, D., Miller, S., & Swetik, P. (1992). *Level I Stability Analysis (LISA) Documentation for Version 2.0* (tech. rep.). United States Department of Agriculture.

Haneberg, W. C. (2006). Effects of digital elevation model errors on spatially distributed seismic slope stability calculations: An example from Seattle, Washington. *Environmental and Engineering Geoscience, 12*(3), 247–260. https://doi.org/10.2113/gseegeosci.12.3.247

Hanssen-Bauer, I., & Hisdal, H. (2017). *Climate in Norway 2100* (tech. rep.). Norwegian Centre for Climate Services (NCCS). https://www.researchgate.net/publication/316922280

Haque, U., da Silva, P. F., Devoli, G., Pilz, J., Zhao, B., Khaloua, A., Wilopo, W., Andersen, P., Lu, P., Lee, J., Yamamoto, T., Keellings, D., Jian-Hong, W., & Glass, G. E. (2019). The human cost of global warming: Deadly landslides and their triggers (1995–2014). *Science of the Total Environment, 682*, 673–684. https://doi.org/10.1016/j.scitotenv.2019.03.415

Harbitz, C. B., Glimsdal, S., Løvholt, F., Kveldsvik, V., Pedersen, G. K., & Jensen, A. (2014). Rockslide tsunamis in complex fjords: From an unstable rock slope at Åkerneset to tsunami risk in western Norway. *Coastal Engineering*, *88*, 101–122. https://doi.org/10.1016/j.coastaleng.2014.02.003

Hefre, H., Gauer, P., Høydal, Ø., Sandersen, F., Kristensen, K., Breien, H., Ekseth, K., Sverdrup Thygeson, K., & Gisnås og Katrine Mo, K. (2019). *Faresonekartlegging i Jølster kommune* (tech. rep.). Norges vassdrags- og energidirektorat. Oslo.

Hidayat, R., Sutanto, S. J., Hidayah, A., Ridwan, B., & Mulyana, A. (2019). Development of a landslide early warning system in Indonesia. *Geosciences (Switzerland)*, *9*(10). https://doi.org/10.3390/geosciences9100451

Hsu, Y. C., & Liu, K. F. (2019). Combining TRIGRS and DEBRIS-2D models for the simulation of a rainfall infiltration induced shallow landslide and subsequent debris flow. *Water (Switzerland)*, *11*(5). https://doi.org/10.3390/w11050890

Hungr, O., Evans, S. G., Bovis, M. J., & Hutchinson, J. N. (2001). A review of the classification of landslides of the flow type. *Environmental and Engineering Geoscience*, *7*(3), 221–238. https://doi.org/10.2113/gseegeosci.7.3.221

Hungr, O., Leroueil, S., & Picarelli, L. (2014). The Varnes classification of landslide types, an update. *Landslides*, *11*(2), 167–194. https://doi.org/10.1007/s10346-013-0436-y

Iverson, R. M. (2000). Landslide triggering by rain infiltration. *Water Resources Research*, *36*(7), 1897–1910. https://doi.org/10.1029/2000WR900090

Jensen, O., Devoli, G., Rustad, K., Verhage, A., Viklund, M., Larsen, J., & Kristensen, L. (2015). *Terminologi for naturfare* (tech. rep.). Norges vassdrags- og energidirektorat.

Kalsnes, B., & Capobianco, V. (2019). *Use of vegetation for landslide risk mitigation* (tech. rep.).

Kanungo, D. P., Arora, M. K., Sarkar, S., & Gupta, R. P. (2006). A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. *Engineering Geology*, *85*(3-4), 347–366. https://doi.org/10.1016/j.enggeo.2006.03.004

Kim, D., Im, S., Lee, C., & Woo, C. (2013). Modeling the contribution of trees to shallow landslide development in a steep, forested watershed. *Ecological Engineering*, *61*, 658–668. https://doi.org/10.1016/j.ecoleng.2013.05.003

Kim, D., Im, S., Lee, S. H., Hong, Y., & Cha, K. S. (2010). Predicting the rainfall-triggered landslides in a forested mountain region using TRIGRS model. *Journal of Mountain Science*, *7*(1), 83–91. https://doi.org/10.1007/s11629-010-1072-9

Kjekstad, O., & Highland, L. (2009). Economic and social impacts of landslides. In K. Sassa & P. Canuti (Eds.), *Landslides - disaster risk reduction* (pp. 573–587). Springer. https://doi.org/https://doi.org/10.1007/978-3-540-69970-5{_}30

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

Kumar, P., Pathania, A., Mali, N., Chaturvedi, P., Sihag, P., Agarwal, S., Singh, R., Uday, K. V., & Dutt, V. (2019). *Landslide Debris-Flow Prediction using Ensemble and Non-Ensemble Machine-Learning Methods Cognitive Modelling View project Landslide risk assessment and management View project Landslide Debris-Flow Prediction using Ensemble and Non-Ensemble Machine-Learning Methods* (tech. rep.). https://www.researchgate.net/publication/336241696

Lee, J. H., & Park, H. J. (2016). Assessment of shallow landslide susceptibility using the transient infiltration flow model and GIS-based probabilistic approach. *Landslides*, *13*(5), 885–903. https://doi.org/10.1007/s10346-015-0646-6

Lehmann, P., von Ruette, J., & Or, D. (2019). Deforestation Effects on Rainfall-Induced Shallow Landslides: Remote Sensing and Physically-Based Modelling. *Water Resources Research*, *55*(11), 9962–9976. https://doi.org/10.1029/2019WR025233

Liao, Z., Hong, Y., Kirschbaum, D., Adler, R. F., Gourley, J. J., & Wooten, R. (2011). Evaluation of TRIGRS (transient rainfall infiltration and grid-based regional slope-stability analysis)'s predictive skill for hurricane-triggered landslides: A case study in Macon County, North Carolina. *Natural Hazards*, *58*(1), 325–339. https://doi.org/10.1007/s11069-010-9670-y

Lindsay, E., Frauenfelder, R., Rubensdotter, L., & Rüther, D. (n.d.). (In prep.) Change detection with Sentinel-1 and Sentinel-2 imagery for landslide event inventory.

Liu, C. N., & Wu, C. C. (2008). Mapping susceptibility of rainfall-triggered shallow landslides using a probabilistic approach. *Environmental Geology*, *55*(4), 907–915. https://doi.org/10.1007/s00254-007-1042-x

Marin, R. J., & Mattos, Á. J. (2020). Physically-based landslide susceptibility analysis using Monte Carlo simulation in a tropical mountain basin. *Georisk*, *14*(3), 192–205. https://doi.org/10.1080/17499518.2019.1633582

Marin, R. J., & Velásquez, M. F. (2020). Influence of hydraulic properties on physically modelling slope stability and the definition of rainfall thresholds for shallow landslides. *Geomorphology*, *351*, 106976. https://doi.org/10.1016/j.geomorph.2019.106976

Matthews, B. W. (1975). Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochimica et Biophysica Acta*, 442–451. https://doi.org/https://doi.org/10.1016/0005-2795(75)90109-9

Melchiorre, C., & Frattini, P. (2012). Modelling probability of rainfall-induced shallow landslides in a changing climate, Otta, Central Norway. *Climatic Change*, *113*(2), 413–436. https://doi.org/10.1007/s10584-011-0325-0

Mergili, M., Marchesini, I., Rossi, M., Guzzetti, F., & Fellin, W. (2014). Spatially distributed three-dimensional slope stability modelling in a raster GIS. *Geomorphology*, *206*, 178–195. https://doi.org/10.1016/j.geomorph.2013.10.008

MET. (2021). Seklima: Observasjoner og værstatistikk. https://seklima.met.no/observations/

Meyer, N. K., Dyrrdal, A. V., Frauenfelder, R., Etzelmüller, B., & Nadim, F. (2012). Hydrometeorological threshold conditions for debris flow initiation in Norway. *Natural Hazards and Earth System Science*, *12*(10), 3059–3073. https://doi.org/10.5194/nhess-12-3059-2012

Milledge, D. G., Griffiths, D. V., Lane, S. N., & Warburton, J. (2012). Limits on the validity of infinite length assumptions for modelling shallow landslides. *Earth Surface Processes and Landforms*, *37*(11), 1158–1166. https://doi.org/10.1002/esp.3235

Miralles, D. G., Gash, J. H., Holmes, T. R., De Jeu, R. A., & Dolman, A. J. (2010). Global canopy interception from satellite observations. *Journal of Geophysical Research Atmospheres*, *115*(16). https://doi.org/10.1029/2009JD013530

Mitchell, J. K., & Soga, K. (2005). *Fundamentals of soil behavior* (3rd ed.). John Wiley & Sons, Inc.

Montgomery, D. R., & Dietrich, W. E. (1994). A physically based model for the topographic control on shallow landsliding. *WATER RESOURCES RESEARCH*, *30*(4). https://doi.org/https://doi.org/10.1029/93WR02979

Montrasio, L., Valentino, R., & Losi, G. L. (2011). Towards a real-time susceptibility assessment of rainfall-induced shallow landslides on a regional scale. *Natural Hazards and Earth System Science*, *11*(7), 1927–1947. https://doi.org/10.5194/nhess-11-1927-2011

Mualem, Y. (1976). A New Model for Predicting the Hydraulic Conductivity of Unsaturated Porous Media. *Water Resources Research*, *12*(3). https://doi.org/https://doi.org/10.1029/WR012i003p00513

NGU. (2017). PRODUKTARK: LØSMASSER N50/N250. http://geo.ngu.no/mapserver/LosmasserWMS?VERS

NIBIO. (2016). *Produktspesifikasjon: AR50 (Arealressurskart i målestokk 1:50.000)* (tech. rep.). NIBIO.

NNPP. (2020). Naturskader for over en halv milliard i 2019. https://www.naturskade.no/aktuelt/2020/naturskader-for-over-en-halv-milliard/

NVE. (2011). *Plan for skredfarekartlegging - Status og prioriteringer innen oversiktkartlegging og detaljert skredfarekartlegging i NVEs regi* (tech. rep.). Norges vassdrags- og energidirektorat (NVE). Oslo.

Olsen, L., Fredin, O., & Olesen, O. (2013). Quaternary geology of Norway. *Geological Survey of Norway Special Publication 13*.

Opsal, Ø. L. (2017). Shear strength of dry tills from the southern half of Norway in relation to bedrock geology. *Norwegian Journal of Geology*, *97*. https://doi.org/10.17850/njg97-2-04

Opsal, Ø. L., & Langeland, J. M. (2019). Geological parameters in relation to bedrock geology and shear strength of dry tills: samples from the southern half of Norway. *Bulletin of Engineering Geology and the Environment*, *78*(3), 1359–1370. https://doi.org/10.1007/s10064-018-1236-3

Pack, R. T., & Goodwin, C. N. (2001). Assessing Terrain Stability in a GIS using SINMAP. *15th annual GIS conference*.

Paniagua, P. (2020). Unsaturated soil, rainfall induced landslides and slope stability (lecture, TBA5155, NTNU).

Park, D. W., Nikhil, N. V., & Lee, S. R. (2013). Landslide and debris flow susceptibility zonation using TRIGRS for the 2011 Seoul landslide event. *Natural Hazards and Earth System Sciences*, *13*(11), 2833–2849. https://doi.org/10.5194/nhess-13-2833-2013

Perera, E. N., Jayawardana, D. T., Jayasinghe, P., Bandara, R. M., & Alahakoon, N. (2018). Direct impacts of landslides on socio-economic systems: a case study from Aranayake, Sri Lanka. *Geoenvironmental Disasters*, *5*(1). https://doi.org/10.1186/s40677-018-0104-6

Peres, D. J., & Cancelliere, A. (2016). Estimating return period of landslide triggering by Monte Carlo simulation. *Journal of Hydrology*, *541*, 256–271. https://doi.org/10.1016/j.jhydrol.2016.03.036

Piciullo, L., Tiranti, D., Pecoraro, G., Cepeda, J. M., & Calvello, M. (2020). Standards for the performance assessment of territorial landslide early warning systems. *Landslides*, *17*(11), 2533–2546. https://doi.org/10.1007/s10346-020-01486-4

Raia, S., Alvioli, M., Rossi, M., Baum, R. L., Godt, J. W., & Guzzetti, F. (2014). Improving predictive power of physically based rainfall-induced shallow landslide models: A probabilistic approach. *Geoscientific Model Development*, *7*(2), 495–514. https://doi.org/10.5194/gmd-7-495-2014

Rankka, K., Andersson-Sköld, Y., Hulten, C., Larsson, R., Leroux, V., & Dahlin, T. (2004). *Quick clay in Sweden* (tech. rep.). Swedish Geotechnical Institute. Linköping. https://www.researchgate.net/publication/265232950

Redshaw, P., & Bottomley, J. (2020). *The Global Landslide Hazard Map Final Project Report* (tech. rep.). World Bank.

Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., & Guzzetti, F. (2018). A review of statistically-based landslide susceptibility models. *Earth-Science Reviews*, *180*, 60–91. https://doi.org/10.1016/j.earscirev.2018.03.001

Richards, L. A. (1931). Capillary conduction of liquids through porous mediums. *Journal of Applied Physics*, *1*(5), 318–333. https://doi.org/10.1063/1.1745010

Ritchie, H., & Roser, M. (2014). Natural Disasters. https://ourworldindata.org/natural-disasters

Rokoengen, K., Jespersen, M. N., Kleiv, R. A., & Sæterbø, E. (2001). The 1345 slide and flood disaster in the Gauldalen valley, Mid-Norway: A new interpretation. *Norsk Geografisk Tidsskrift*, *55*(2), 57–70. https://doi.org/10.1080/00291950121138

Rouault, C. (2020). Extreme Multiple Landslide Events in Norway: An Investigation of Rainfall and Snowmelt Induced Soil Landslide Detection and Forecasting.

Rüther, D., Hefre, H., & Rubensdotter, L. (n.d.). (In prep.) Extreme precipitation induced landslide event on June 30, 2019 in Jølster, western Norway.

Saadatkhah, N., Kassim, A., & Lee, L. M. (2015). Hulu Kelang, Malaysia regional mapping of rainfall-induced landslides using TRIGRS model. *Arabian Journal of Geosciences*, *8*(5), 3183–3194. https://doi.org/10.1007/s12517-014-1410-2

Saadatkhah, N., Mansor, S., Kassim, A., Lee, L. M., Saadatkhah, R., & Sobhanmanesh, A. (2016). Regional modeling of rainfall-induced landslides using TRIGRS model by incorporating plant cover effects: case study in Hulu Kelang, Malaysia. *Environmental Earth Sciences*, *75*(5). https://doi.org/10.1007/s12665-016-5326-x

Sandersen, F., & Høydal, Ø. (2019). *Faresonekartlegging i Jølster kommune* (tech. rep. No. 57). Norges vassdrags- og energidirektorat. Oslo.

Sandvoll, S. (2020). Spatial and temporal distribution of the extreme precipitation and landslide events in Jølster 30th of July 2019.

Sarma, C. P., Dey, A., & Krishna, A. M. (2020). Influence of digital elevation models on the simulation of rainfall-induced landslides in the hillslopes of Guwahati, India. *Engineering Geology*, *268*, 105523. https://doi.org/10.1016/j.enggeo.2020.105523

Saulnier, G. M., Beven, K., & Obled, C. (1997). Including spatially variable effective soil depths in TOPMODEL. *Journal of Hydrology*, *202*(1-4), 158–172. https://doi.org/10.1016/S0022-1694(97)00059-0

Schaap, M. G., Leij, F. J., Van Genuchten, M. T., & Brown, G. E. (2001). ROSETTA: a computer program for estimating soil hydraulic parameters with hierarchical pedotransfer functions. *Journal of Hydrology*, *251*. http://www.ussl.ars.usda.gov/.

Schilirò, L., Esposito, C., & Mugnozza, G. S. (2015). Evaluation of shallow landslide-triggering scenarios through a physically based approach: An example of application in the southern Messina area (northeastern Sicily, Italy). *Natural Hazards and Earth System Sciences*, *15*(9), 2091–2109. https://doi.org/10.5194/nhess-15-2091-2015

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

Schilirò, L., Cevasco, A., Esposito, C., & Mugnozza, G. S. (2018). Shallow landslide initiation on terraced slopes: Inferences from a physically based approach. *Geomatics, Natural Hazards and Risk*, *9*(1), 295–324. https://doi.org/10.1080/19475705.2018.1430066

Schuster, R. L. (1978). Chapter 1: Introduction. *Landslides: Analysis and control*. Transportation Research Board.

SCOPUS. (2019). Scopus fact sheet. https://www.elsevier.com/__data/assets/pdf_file/0017/114533/Scopus_GlobalResearch_Factsheet2019_FINAL_WEB.pdf

Seefelder, C., Koide, S., & Mergili, M. (2017). Does parameterization influence the performance of slope stability model results? A case study in Rio de Janeiro, Brazil. *Landslides*, *14*(4), 1389–1401. https://doi.org/10.1007/s10346-016-0783-6

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

Simoni, S., Zanotti, F., Bertoldi, G., & Rigon, R. (2008). Modelling the probability of occurrence of shallow landslides and channelized debris flows using GEOtop-FS. *Hydrological Processes*, *22*(4), 532–545. https://doi.org/10.1002/hyp.6886

Simunek, J., Saito, H., Sakai, M., & Van Genuchten, M. (2008). *The HYDRUS-1D Software Package for Simulating the One-Dimensional Movement of Water, Heat, and Multiple Solutes in Variably-Saturated Media Dissolution of NTO, DNAN, and Insensitive Munitions Formulations and Their Fates in Soils View project* (tech. rep.). Department of Environmental Sciences, University of California Riverside. Riverside, California. https://www.researchgate.net/publication/271515313

Sorbino, G., Sica, C., & Cascini, L. (2010). Susceptibility analysis of shallow landslides source areas using physically based models. *Natural Hazards*, *53*(2), 313–332. https://doi.org/10.1007/s11069-009-9431-y

SSB. (2016). Befolkning - årleg, per 1. januar - Tabellar - SSB. https://www.ssb.no/befolkning/statistikker/folkemengde/aar-per-1-januar/2016-02-19?fane=tabell&sort=nummer&tabell=256001

Stancanelli, L. M., Peres, D. J., Cancelliere, A., & Foti, E. (2017). A combined triggering-propagation modeling approach for the assessment of rainfall induced debris flow susceptibility. *Journal of Hydrology*, *550*, 130–143. https://doi.org/10.1016/j.jhydrol.2017.04.038

Tang, Y., Yin, K. l., Liu, L., Zhang, L., & Fu, X. l. (2017). Dynamic assessment of rainfall-induced shallow landslide hazard. *Journal of Mountain Science*, *14*(7), 1292–1302. https://doi.org/10.1007/s11629-016-4353-0

Taylor, D. (1948). *Fundamentals of Soil Mechanics*. John Wiley; Sons, Inc. https://doi.org/https://doi.org/10.1097/00010694-194808000-00008

Terratec. (2016). *NDH Jølster 2pkt 2016_Prosjektrapport* (tech. rep.).

Tesfa, T. K., Tarboton, D. G., Chandler, D. G., & McNamara, J. P. (2009). Modeling soil depth from topographic and land cover attributes. *Water Resources Research*, *45*(10). https://doi.org/10.1029/2008WR007474

Thorsnæs, G., & Askheim, S. (2020). Jølster – tidligere kommune – Store norske leksikon. https://snl.no/J%C3%B8lster_-_tidligere_kommune

Tianchi, L. (1994). Landslide Disasters and Human Responses in China. *Mountain Research and Development*, *14*(4), 341–346. https://doi.org/https://doi.org/10.2307/3673730

Tran, T. V., Alvioli, M., Lee, G., & An, H. U. (2018). Three-dimensional, time-dependent modeling of rainfall-induced landslides over a digital landscape: a case study. *Landslides*, *15*(6), 1071–1084. https://doi.org/10.1007/s10346-017-0931-7

Tran, T. V., Lee, G., An, H., & Kim, M. (2017). Comparing the performance of TRIGRS and TiVaSS in spatial and temporal prediction of rainfall-induced shallow landslides. *Environmental Earth Sciences*, *76*(8), 1–16. https://doi.org/10.1007/s12665-017-6635-4

UNISDR. (2009). *UNISDR Terminology on Disaster Risk Reduction* (tech. rep.). www.preventionweb.net

Vanapalli, S. K., & Fredlund, D. G. (2000). Comparison of Different Procedures to Predict Unsaturated Soil Shear Strength. *Geo-Denver 2000*, *287*. https://doi.org/https://doi.org/10.1061/40510(287)13

van Genuchten, M., Leij, F. J., & Yates, S. R. (1991). *The RETC Code for Quantifying the Hydraulic Functions of Unsaturated Soils* (tech. rep.). U. S. Environmental Protection Agency.

van Genuchten, M. T. (1980). A Closed-form Equation for Predicting the Hydraulic Conductivity of Unsaturated Soils. *Soil Science Society of America Journal*, *44*(5), 892–898. https://doi.org/10.2136/sssaj1980.03615995004400050002x

Varnes, D. J. (1978). Slope Movement Types and Processes. In R. L. Schuster & R. J. Krizek (Eds.), *Landslides, analysis and control, transportation research board, special report no. 176*.

Verdal Historielag, . (2020). Verdalsraset. https://verdal-historielag.no/verdal/verdalsraset/

Vieira, B. C., Fernandes, N. F., Augusto Filho, O., Martins, T. D., & Montgomery, D. R. (2018). Assessing shallow landslide hazards using the TRIGRS and SHALSTAB models, Serra do Mar, Brazil. *Environmental Earth Sciences*, *77*(6), 1–15. https://doi.org/10.1007/s12665-018-7436-0

Viet, T. T., Lee, G., Thu, T. M., & An, H. U. (2017). Effect of Digital Elevation Model Resolution on Shallow Landslide Modeling Using TRIGRS. *Natural Hazards Review*, *18*(2), 1–12. https://doi.org/10.1061/(ASCE)NH.1527-6996.0000233

Vihinen, M. (2012). How to evaluate performance of prediction methods? Measures and their interpretation in variation effect analysis. *BMC genomics*, *13*(Suppl 4). https://doi.org/10.1186/1471-2164-13-S4-S2

Weidner, L., Oommen, T., Escobar-Wolf, R., Sajinkumar, K. S., & Samuel, R. A. (2018). Regional-scale back-analysis using TRIGRS: an approach to advance landslide hazard modeling and prediction in sparse data regions. *Landslides*, *15*(12), 2343–2356. https://doi.org/10.1007/s10346-018-1044-7

Welcker, C. (2011). *Bulking Debris Flow Initiation and Impacts* (Doctoral dissertation). University of Idaho. https://doi.org/10.13140/RG.2.1.1840.5207

WFPA. (2020). Deep-seated and shallow rapid landslides: know the difference. https://www.wfpa.org/news-resources/blog/deep-seated-landslides-shallow-landslides-washington/

Winter, M. G., Shearer, B., Palmer, D., Peeling, D., Harmer, C., & Sharpe, J. (2016). The Economic Impact of Landslides and Floods on the Road Network. *Procedia Engineering*, *143*, 1425–1434. https://doi.org/10.1016/j.proeng.2016.06.168

Wu, S. J., Hsiao, Y. H., Yeh, K. C., & Yang, S. H. (2017). A probabilistic model for evaluating the reliability of rainfall thresholds for shallow landslides based on uncertainties in rainfall characteristics and soil properties. *Natural Hazards*, *87*(1), 469–513. https://doi.org/10.1007/s11069-017-2773-y

Yu, M.-h. (2002). Advances in strength theories for materials under complex stress state in the 20th century. *Applied Mechanics Reviews*, *55*(3), 169–218. https://doi.org/10.1115/1.1472455

Zêzere, J. L., Trigo, R. M., & Trigo, I. F. (2005). Shallow and deep landslides induced by rainfall in the Lisbon region (Portugal): assessment of relationships with the North Atlantic Oscillation. *Natural Hazards and Earth System Sciences*, *5*, 331–344.

Zhang, H. Q., Tannant, D. D., Jing, H. W., Nunoo, S., Niu, S. J., & Wang, S. Y. (2015). Evolution of cohesion and friction angle during microfracture accumulation in rock. *Natural Hazards*, *77*(1), 497–510. https://doi.org/10.1007/s11069-015-1592-2

Zhuang, J., Peng, J., Wang, G., Iqbal, J., Wang, Y., Li, W., Xu, Q., & Zhu, X. (2017). Prediction of rainfall-induced shallow landslides in the Loess Plateau, Yan'an, China, using the TRIGRS model. *Earth Surface Processes and Landforms*, *42*(6), 915–927. https://doi.org/10.1002/esp.4050

Zieher, T., Rutzinger, M., Schneider-Muntau, B., Perzl, F., Leidinger, D., Formayer, H., & Geitner, C. (2017). Sensitivity analysis and calibration of a dynamic physically based slope stability model. *Natural Hazards and Earth System Sciences*, *17*(6), 971–992. https://doi.org/10.5194/nhess-17-971-2017

Appendices

Appendix A: Python Scripts

Initialization-file-creator:

```
"""
Created on January 7, 2021
@author: Harald B. Larsen

This script:
-creates a modified copy of an initial initiation file
-saves its parameter values and output nametag in an excel sheet
-repeat
"""
import numpy as np
import shutil
import pandas as pd
import os

#Input: Import file
file_r = open('faobs.txt', 'r') #Reads the initial initiation file
lines = file_r.readlines() #Creates a list of each line as strings

#Input: Rows and columns of values to change:
line1 = 10 #First property zones
line2 = 13 #Second property zone
ci = 0 #Starting index of cohesion
pi = 6 #Friction angle (phi)
di = 17 #Diffusivity (D0)
ki = 23 #K_sat
tag_line = 48 #Row of the nametag
tag_ind = 3 #Index of nametag
tag = 0 #Initial tag, makes first nametag = 1

#Input: Parameter range of values: (NB: CANNOT BE EQUAL at any point with current code(!)
c_range = [2100, 2500, 3000, 3600]
p_range = [32, 35, 37]
d_range = [0.0009, 0.0006, 0.0003, 0.00009]
k_range = [0.00001, 0.00005, 0.00008, 0.0002, 0.0005]
simulations = len(c_range)*len(p_range)*len(d_range)*len(k_range) #Number of simulations

#Empty parameter lists for saving in excel:
c_list = list()
p_list = list()
d_list = list()
k_list = list()
tag_list = list()

```

```
#Iterate through all possible combinations of parameters:
#Replace 4-digit cohesion values [int]
for a in range(0, len(c_range)):
    c = c_range[a]
    lines[line1] = lines[line1].replace(lines[line1][ci:ci+4], str(c))
    lines[line2] = lines[line2].replace(lines[line2][ci:ci+4], str(c))

    #Replace 2-digit phi value [int]
    for b in range(0, len(p_range)):
        p = p_range[b]
        lines[line1] = lines[line1].replace(lines[line1][pi:pi+2], str(p))
        lines[line2] = lines[line2].replace(lines[line2][pi:pi+2], str(p))

        #Replace "4-digit" D0-value (scientific number)
        for e in range(0, len(d_range)):
            d = d_range[e]
            d_sci_str = np.format_float_scientific(d).replace('.', '')
            lines[line1] = lines[line1].replace(lines[line1][di:di+4], d_sci_str.replace('0'
    , ''))
            lines[line2] = lines[line2].replace(lines[line2][di:di+4], d_sci_str.replace('0'
    , ''))

            #Replace "4-digit" k_sat value (scientific number)
            for f in range(0, len(k_range)):
                k = k_range[f]
                k_sci_str = np.format_float_scientific(k).replace('.', '')
                lines[line1] = lines[line1].replace(lines[line1][ki:ki+4], k_sci_str.replace
    ('0', ''))
                lines[line2] = lines[line2].replace(lines[line2][ki:ki+4], k_sci_str.replace
    ('0', ''))


                #For each resulting initiation file: resulting file:

                #Save new parameter values to lists:
                c_list.append(c), p_list.append(p), d_list.append(d), k_list.append(k)

                #Change nametag (nb, max nr = 999)
                tag = str(int(tag)+1) #Adds 1 to the nametag string
                lines[tag_line] = lines[tag_line].replace(lines[tag_line][tag_ind:tag_ind
    +4], tag+'\n') #Replace nametag
                tag_list.append(tag) #List of tags later stored in excel

                #Write modified lines to initial init_file:
                file_w = open('sl0.txt', 'w') #Opens the file in writing mode
                for eachline in lines:
```

```
85                      file_w.write(eachline) #Writes each line (including modified lines) to a
      new version of the file
86                  file_w.close() #Close the file
87
88                  #Create a copy of the init_file with new, recognizable name:
89                  shutil.copyfile('sl0.txt', 'sl' + tag + '.txt')
90
91 #Prepare list and save in excel:
92 init_data = np.transpose(np.vstack((c_list, p_list, d_list, k_list, tag_list))) #Stack and
     transpose
93 colu = ['cohesion', 'phi', 'D0', 'k_sat', 'nametag'] #Column headings
94 init_data_df = pd.DataFrame(init_data, columns=colu) #DataFrames are easy to work with
95 init_data_df.to_excel('initfiles_sl_run1.xlsx') #Save documentation in excel
```

TRIGRS-executor:

```
"""
@author: Harald B. Larsen

This script:
1) Reads a list of initiation files, copied as paths
2  Changes the name of an initiation file for recognition by TRIGRS
3) Executes TRIGRS on the initiation file
4) Changes the init_file back to its original name
REPEAT
"""

import numpy as np
import subprocess
import os

#Provide list of initialization file names:
initfiles = np.loadtxt('init_files_list.txt', dtype=str)

#Iterate through inititiation files and execute TRIGRS on each:
for x in range(0, len(initfiles)):
    initfile = initfiles[x][1:-1] #Pick one init_file, strip unneccessary quotes

    os.rename(initfile, 'tr_in.txt') #Rename initialization file for recognition by TRIGRS

    subprocess.run('TRIGRS.exe')  #Execute TRIGRS:

    os.rename('tr_in.txt', initfile) #Back to original initiaion file name:

    #Input: Save log file:
    os.rename('trigrslog.txt', 'log_'+ initfile[-9:]) #Check starting index of nametag (-9
    here)
```

Analysis-multifiles:

```
"""
@author: Harald B. Larsen

This script:
-takes a list of FoS-files, along with a file representing real failuiire
-runs external tools for ROC-analysis and summary statistic
-plots all ROC-curves / points in one figure, and stores all results in an excel document.
"""

import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
from analysis_multiplefiles_funcs import ROC
from analysis_multiplefiles_funcs import stat
plt.rcParams['axes.grid'] = True #Grids for plots

#Input: Import files and specify figure titles:
real = np.loadtxt('failure_r10.txt', skiprows=6) #Binary file representing failure
filenames = np.loadtxt('fs_file_list.txt', dtype=str) #List of FoS-files to evaluate, copied
    as paths

roc_title = 'ROC on one test file, 17:00\n -J lster' #Choose a title for ROC-plots
roc_figname = 'roc_analysis_test.png' #Choose a name/path for saving ROC-fiure
excel_name = 'analysis_Jolster_test.xls' #Spceify name/path of resulting excel data file

#Empty lists for storing results from multiple files:
minFS_L = list()
maxFS_L = list()
avgFS_L = list()

P_L = list() #Real positives
TP_L = list() #True positives
FP_L = list() #False positives
D2PC_1_L = list() #Distance to perfect classification
TPR_1_L = list() #True positive rate
FPR_1_L = list() #False positive rate

TPr_list_L = list() #List of lists
FPr_list_L = list() #List of lists
AUC_L = list() #Area under the curve
minD2PC_L = list() #Minimum distance to perfect classification
o_thresh_L = list() #Optmal threshold to minimize D2PC

#Create figure for ROC-plots
```

```
fig, ax = plt.subplots(2,1, sharex=True)
fig.autolayout : True

for x in range(0, len(filenames)):#Iterate through filenames
    fname = filenames[x][1:-1] #Pick one model-FoS filename
    model = np.loadtxt(fname, skiprows=6) # Load to a numpy array and skip information rows

    #Call summary statistics function
    minFS, maxFS, avgFS = stat(model)

    #Call ROC-curve tool
    Ps, TPs, FPs, TPr_list, FPr_list, AUC, minD2PC, opt_threshold, D2PC_1, TPr_1, FPr_1 =
    ROC(real, model)

    #Append results to lists
    minFS_L.append(minFS), maxFS_L.append(maxFS), avgFS_L.append(avgFS)
    P_L.append(Ps)
    TP_L.append(TPs),FP_L.append(FPs)
    D2PC_1_L.append(D2PC_1)
    TPR_1_L.append(TPr_1)
    FPR_1_L.append(FPr_1)
    TPr_list_L.append(TPr_list)
    FPr_list_L.append(FPr_list)
    AUC_L.append(AUC)
    minD2PC_L.append(minD2PC)
    o_thresh_L.append(opt_threshold)#Threshold with lowest D2PC

    #Plot each ROC-curve and point
    ax[0].plot(FPr_list, TPr_list)

    #Plot ROC-points:
    ax[1].scatter(FPr_1, TPr_1, label=fname)

#Fix ROC-figure
fig.suptitle(roc_title)
ax[0].set_xlabel('False positive rate')
ax[0].set_ylabel('True positive rate')
ax[1].set_xlabel('False positive rate')
ax[1].set_ylabel('True positive rate')
fig.legend()
fig.savefig(roc_figname, dpi=300)

#Combine lists for storing data in excel:
all_data1 = np.vstack((filenames, minFS_L, maxFS_L, avgFS_L, FP_L, TP_L))
all_data2 = np.vstack((P_L,D2PC_1_L, TPR_1_L, FPR_1_L, AUC_L, minD2PC_L, o_thresh_L))
all_data_all = np.vstack((all_data1, all_data2))
```

```
all_data = np.transpose(all_data_all)

col = ['filename','minFS','maxFS','avgFS','false positives', 'true positives', 'positives','
    D2PC_1','TP4_1','FPr_1','AUC','minD2PC','opt_thresh']
data_df = pd.DataFrame(all_data, columns=col)

#Specify name of resulting excel file:
data_df.to_excel(excel_name)
```

Analysis-multifiles-functions:

```
"""
@author: Harald B. Larsen

Summary Statistics
"""
def stat(model): #Takes a modeled FoS-file
    import numpy as np


    #Basic Statistics:
    minFS = np.amin(model)
    maxFS = np.amax(model)
    avgFS = np.average(model)

    return minFS, maxFS, avgFS

"""
ROC-Analysis
This script
-Takes a model-FoS file and a real failure file, and, for each threshold-increment:
-Creates a contingency matrix (through addition) and counts TPs, FPs etc.
-Calculates various performance indexes
-Finds approximate AUC and related terms by calculating performance at each threshold
    increments
-Saves results in an excel document
"""
def ROC(real, model): #Real failure file and model FoS-file

 import numpy as np

 #Create empty lists for data storage
 TP_list = list()
 FP_list = list()
 TPr_list = list()
 FPr_list = list()
 D2PC_list = list()
 threshold_list = list()

 #Iterate through a range of thresholds:
 for x in range(3,120): #Threshold range: 0.3-12 (more than enough)

     #Create contingency matrix (TP = 1, FP = 2, FN = 3, TN = 4):
     p = model <= 0.1*x #Booelan model positives
     model_newvals = np.select([p==1, p==0], [3,1]) #Replace values before addition on next
```

```
    line
     cwv = model_newvals + real #Combine modeled and real positives, "contingency wrong
    values"
     contingency = np.select([cwv==4, cwv==3, cwv==2, cwv==1], [1,2,3,4]) #Replace with
    nicer values

     #Count results
     TP = np.count_nonzero(contingency == 1) #True positives
     FP = np.count_nonzero(contingency == 2) #False positives
     FN = np.count_nonzero(contingency == 3) #False negatives
     TN = np.count_nonzero(contingency == 4) #True negatives
     Ps = TP + FN #Total number of model positives
     Ns = FP + TN #Total number of model negatives

     #Calculate performance indexes (might add others):
     TPr = TP/Ps #True positive rate
     FPr = FP/Ns #False positive rate
     FNr = 1-TPr #False negative rate
     D2PC = np.sqrt(FPr**2 + FNr**2) #Distance to perfect classification

     #Append results to lists:
     TP_list.append(TP)
     FP_list.append(FP)
     TPr_list.append(TPr)
     FPr_list.append(FPr)
     D2PC_list.append(D2PC)
     threshold_list.append(0.1*x)

#Calculate AUC;
AUC = 0 #Initial
for x in range (0,len(TPr_list)-1):
        AUC = AUC + ((TPr_list[x+1]+TPr_list[x])/2)*(FPr_list[x+1]-FPr_list[x]) #Approximate
    integration

#Find minimum D2PC and corresponding threshold value:
minD2PC = min(D2PC_list) #Minimum value
minD2PCindex= D2PC_list.index(minD2PC) #Index of minimum value
opt_threshold = threshold_list[minD2PCindex] #Threshold value yielding minimum D2PC

#Results at threshold = 1 (Default):
index_1 = threshold_list.index(1) #The index (in all lists) at which threshold=1
TPs = TP_list[index_1] #Number of true positives
FPs = FP_list[index_1] #Number of false positives
TPr_1 = TPr_list[index_1] #True positive rate
FPr_1 = FPr_list[index_1] #False positive rate
D2PC_1 = D2PC_list[index_1] #Distance to perfect classification
```

```
86
87 #Returning variables:
88 return Ps, TPs, FPs, TPr_list, FPr_list, AUC, minD2PC, opt_threshold, D2PC_1, TPr_1, FPr_1
```

Analysis-magnitudes:

```
"""
@author: Harald B. Larsen

This script:
-Takes ONE FoS-file, along with a real binary failure file and various other spatial data
-Calls external evaluation functions
-Plots data in numerous figures and saves these figures in the current working directory (or as specified)
"""
#Script needs work if there are noe false negatives!
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

from evaluation_funcs_HBL import contingency
from evaluation_funcs_HBL import spat_bin
from evaluation_funcs_HBL import non_spat_bin

plt.rcParams['axes.grid'] = True #Auto-set grids for figures

"""Preferences"""
#Input: Landslides
failure = np.loadtxt('failure_r10.txt', skiprows=6) #Binary real failure grid
fos_file_f = np.loadtxt("OutputTRfs_min_fa0_5.txt", skiprows=6) #FoS file at the time of landslide occurence
#NB: If analyzing results BEFORE failure (FPs), the above files should be change appropriately

#Input: Spatial data
slope = np.loadtxt("spat_data\slope_10.txt", skiprows=6) #Slope grid
alt = np.loadtxt("spat_data\dem_10.txt", skiprows=6)  #Altitude grid
p_zones = np.loadtxt("spat_data\propzones_3class_10m_sa.txt", skiprows=6) #Property zone grid
v_p_zones = np.loadtxt("spat_data\propzones_veg_2class.txt", skiprows=6) #Evt kun veg, men g r for veg_propzones for  yeblikket !

#Input: Plots
eval_tag = 'test' #The nametag to be added to plot titles

#Input: Storage options - specify names/paths
cont_mat_store = 'cont_mat_test' #Contingency matrix

#Input: Percentile of positive cells at which to calculate UIL:
UIL_perc = 90
```

```
40
41
42 """Binary, Spatial Evaluation"""
43 #Create contingency matrix:
44 cont_mat = contingency(failure, fos_file_f, cont_mat_store) #Also saves cont_mat as txt_file
       !
45
46 #Call binary spatial analysis function:
47 binspat_df = spat_bin(cont_mat, slope, alt, p_zones, v_p_zones)
48
49 #Plot Binary Spatial Data:
50 #Create empty figures:
51 s_fig, s_ax = plt.subplots(2,2) #Slope distribution figure, 4 subplots
52 a_fig, a_ax = plt.subplots(2,2) #Altitude distribution figure
53 p_fig, p_ax = plt.subplots(2,2) #Property zone distribution figure
54 v_fig, v_ax = plt.subplots(2,2) #Vegetation zone distribution figure
55
56 #Plot histograms of contingency-matrix-divided spatial data in each subplot:
57 s_ax[0][0].hist(binspat_df['s_TP'], label='TP', color='b') #Slopes of True positives
58 s_ax[0][1].hist(binspat_df['s_FP'], label='FP', color='y') #Slopes of false positives
59 s_ax[1][0].hist(binspat_df['s_FN'], label='FN', color='r')
60 s_ax[1][1].hist(binspat_df['s_TN'], label='TN', color='limegreen')
61 a_ax[0][0].hist(binspat_df['a_TP'], label='TP', color='b') #Altitudes
62 a_ax[0][1].hist(binspat_df['a_FP'], label='FP', color='y')
63 a_ax[1][0].hist(binspat_df['a_FN'], label='FN', color='r')
64 a_ax[1][1].hist(binspat_df['a_TN'], label='TN', color='limegreen')
65 p_ax[0][0].hist(binspat_df['p_TP'], label='TP', color='b') #Property Zones
66 p_ax[0][1].hist(binspat_df['p_FP'], label='FP', color='y')
67 p_ax[1][0].hist(binspat_df['p_FN'], label='FN', color='r')
68 p_ax[1][1].hist(binspat_df['p_TN'], label='TN', color='limegreen')
69 v_ax[0][0].hist(binspat_df['v_TP'], label='TP', color='b', bins=np.linspace(0,12,13)) #
       Vegetation Property Zones
70 v_ax[0][1].hist(binspat_df['v_FP'], label='FP', color='y', bins=np.linspace(0,12,13))
71 v_ax[1][0].hist(binspat_df['v_FN'], label='FN', color='r', bins=np.linspace(0,12,13))
72 v_ax[1][1].hist(binspat_df['v_TN'], label='TN', color='limegreen', bins=np.linspace(0,12,13)
       )
73 #Set axis labels:
74 s_ax[1][0].set_xlabel('Degrees')
75 s_ax[1][1].set_xlabel('Degrees')
76 s_ax[1][0].set_ylabel('Freq.')
77 a_ax[1][0].set_xlabel('m.a.s.l.')
78 a_ax[1][1].set_xlabel('m.a.s.l.')
79 a_ax[1][0].set_ylabel('Freq.')
80 p_ax[1][0].set_xlabel('p-zone')
81 p_ax[1][1].set_xlabel('p-zone')
82 p_ax[1][0].set_ylabel('Freq.')
```

```
v_ax[1][0].set_xlabel('v_p-zone')
v_ax[1][1].set_xlabel('v_p-zone')
v_ax[1][0].set_ylabel('Freq.')
#Set titles and legends
s_fig.legend()
s_fig.suptitle('Slope Distributions\n' + eval_tag)
a_fig.legend()
a_fig.suptitle('Altitude Distributions\n' + eval_tag)
p_fig.legend()
p_fig.suptitle('Property Zone Dist.\n' + eval_tag)
v_fig.legend()
v_fig.suptitle('Vegetation property zone Dist.\n' + eval_tag)

s_fig.savefig('s_fig_' + eval_tag + '.png', dpi=300)
a_fig.savefig('a_fig' + eval_tag + '.png', dpi=300)
p_fig.savefig('p_fig' + eval_tag + '.png', dpi=300)
v_fig.savefig('v_fig' + eval_tag + '.png', dpi=300)


"""Non-binary Spatial Evaluation"""
#Call non-binary spatial evaluation function:
CR, FCR, CEMR, contspat_df, positives, negatives, pos_slopes, neg_slopes, UIL, CTN,
    crit_cells = non_spat_bin(fos_file_f, cont_mat, slope, alt, p_zones, v_p_zones, eval_tag
    , UIL_perc)
#note that this function saves visualization txts of NoData = 0, which can be prepared and
    visualized!

#Plot non-binary data:

#Major scatter plot representing all cells:
overview, ax = plt.subplots(1,1)
ax.scatter(negatives, neg_slopes, c='g', label='Real negatives', alpha=0.5, s=10)
ax.scatter(positives, pos_slopes, c='r', label='Real positives', s=10)
ax.axvline(x=1)
ax.set_xlabel('Model Factor of Safety')
ax.set_ylabel('Predicted slope (degrees)')
overview.legend()
overview.suptitle('Overview of modeling results\n' + eval_tag)
overview.savefig('overview_'+eval_tag +'.png', dpi=400)
ax.set_xlim(0,2)
overview.suptitle('Focused overview\n' +eval_tag)
overview.savefig('overview_zoom_'+eval_tag+'.png', dpi=400)

#Scatter plots (pairwise spatial value and corresponding error):
s_figm, s_axm = plt.subplots(3,1, sharex=True) #Slope figure
a_figm, a_axm = plt.subplots(3,1, sharex=True) #Altitude figure
```

```
126 p_figm, p_axm = plt.subplots(3,1, sharex=True) #Property zone figure
127 v_figm, v_axm = plt.subplots(3,1, sharex=True) #Vegetation zone figure
128
129 s_axm[0].scatter(contspat_df['slopes_C'], contspat_df['C_slopes'], label='Confidence', color
    ='b')
130 a_axm[0].scatter(contspat_df['alts_C'], contspat_df['C_alts'], label='Confidence', color='b'
    )
131 p_axm[0].scatter(contspat_df['p_zones_C'], contspat_df['C_p_zones'], label='Confidence',
    color='b')
132 v_axm[0].scatter(contspat_df['v_p_zones_C'], contspat_df['C_v_p_zones'], label='Confidence',
     color='b')
133
134 s_axm[1].scatter(contspat_df['slopes_FC'], contspat_df['FC_slopes'], label='False confidence
    ', color='y')
135 a_axm[1].scatter(contspat_df['alts_FC'], contspat_df['FC_alts'], label='False confidence',
    color='y')
136 p_axm[1].scatter(contspat_df['p_zones_FC'], contspat_df['FC_p_zones'], label='False
    confidence', color='y')
137 v_axm[1].scatter(contspat_df['v_p_zones_FC'], contspat_df['FC_v_p_zones'], label='False
    confidence', color='y')
138
139 s_axm[2].scatter(contspat_df['slopes_CEM'], contspat_df['CEM_slopes'], label='Critical error
     magnitude', color='r')
140 a_axm[2].scatter(contspat_df['alts_CEM'], contspat_df['CEM_alts'], label='Critical error
    magnitude', color='r')
141 p_axm[2].scatter(contspat_df['p_zones_CEM'], contspat_df['CEM_p_zones'], label='Critical
    error magnitude', color='r')
142 v_axm[2].scatter(contspat_df['v_p_zones_CEM'], contspat_df['CEM_v_p_zones'], label='Critical
     error magnitude', color='r')
143
144 s_axm[2].set_xlabel('degrees')
145 a_axm[2].set_xlabel('m.a.s.l.')
146 p_axm[2].set_xlabel('Property Zone')
147 v_axm[2].set_xlabel('Vegetation Property Zone')
148
149 s_figm.legend()
150 s_figm.suptitle('Error magnitudes\n vs. slope, '+eval_tag)
151 a_figm.legend()
152 a_figm.suptitle('Error magnitudes\n vs. altitude, '+eval_tag)
153 p_figm.legend()
154 p_figm.suptitle('Error magnitudes\n  vs. property zones, '+eval_tag)
155 v_figm.legend()
156 v_figm.suptitle('Error magnitudes\n  vs. veg. property zones, '+eval_tag)
157
158 s_figm.savefig('s_figm_'+eval_tag +'.png', dpi=300)
159 a_figm.savefig('a_figm_'+eval_tag+'.png', dpi=300)
```

```
160 p_figm.savefig('p_figm_'+eval_tag+'.png', dpi=300)
161 v_figm.savefig('v_figm_'+eval_tag+'.png', dpi=300)
162
163 #Some other calculations:
164 crit_cells_perc = (crit_cells/np.count_nonzero(fos_file_f))*100
165
166 CTNr = CTN / np.count_nonzero(cont_mat==4) #Critical negative ratio (CTN/TN)
167
168 #Print results:
169 print('Ci: ', CR)
170 print('FCi: ', FCR)
171 print('CEMi: ', CEMR)
172 print('For true negatives, a total of ', CTN, ' lies below ', UIL, ' (UIL), yielding a CTNr
        of ', CTNr, '.')
173 print('A total of ', crit_cells, ' cells (CS), yielding ', crit_cells_perc, '% of total
        cells (CSr), lie below UIL')
```

Analysis-magnitudes-functions:

```
"""
Functions called by "Analysis_magnitudes"
"""

def contingency(failure, fos_file_f, cont_mat_store):

    import numpy as np

    #Contingency Matrix (1=TP, 2=FP, 3=FN, 4=TN)
    bol_pos = fos_file_f <=1 #Booelan model positives (b=model)
    bol_pos_2 = np.select([bol_pos==1, bol_pos==0], [3,1])  #Replacing values before
    addition
    cont_1 = bol_pos_2 + failure #adding positives to real file
    cont_mat = np.select([cont_1==4, cont_1==3, cont_1==2, cont_1==1], [1,2,3,4]) #Nicer
    values
    np.savetxt(cont_mat_store, cont_mat)
    return cont_mat



"""
Binary Spatial Analysis
This script:
-takes in a contingency matrix and various spatial data
-investigates WHERE contingengencies occur in relation to slope etc. by:
-saving e.g. slope values of TPs, FPs etc in separate lists in a DataFrame
"""

def spat_bin(cont_mat, slope, alt, p_zones, v_p_zones):
    import pandas as pd

    #Create 4 lists (one for each contingency) per spatial grid  and total lists
    slopes_TP = list()
    slopes_FP = list()
    slopes_FN = list()
    slopes_TN = list()
    alts_TP = list()
    alts_FP = list()
    alts_FN = list()
    alts_TN = list()
    p_zones_TP = list()
    p_zones_FP = list()
    p_zones_FN = list()
    p_zones_TN = list()
```

```
    v_p_zones_TP = list()
    v_p_zones_FP = list()
    v_p_zones_FN = list()
    v_p_zones_TN = list()

    slopes_total = list()
    alts_total = list()
    p_zones_total = list()
    v_p_zones_total = list()

    #Iterate through each cell in cont_mat, and store corresponding spatial data:
    rows, cols = cont_mat.shape
    for x in range(0, rows):
        for y in range(0,cols):
            if cont_mat[x][y] == 1: #True positives
                slopes_TP.append(slope[x][y])
                alts_TP.append(alt[x][y])
                p_zones_TP.append(p_zones[x][y])
                v_p_zones_TP.append(v_p_zones[x][y])
            elif cont_mat[x][y] == 2: #False positives
                slopes_FP.append(slope[x][y])
                alts_FP.append(alt[x][y])
                p_zones_FP.append(p_zones[x][y])
                v_p_zones_FP.append(v_p_zones[x][y])
            elif cont_mat[x][y] == 3: #False negatives
                slopes_FN.append(slope[x][y])
                alts_FN.append(alt[x][y])
                p_zones_FN.append(p_zones[x][y])
                v_p_zones_FN.append(v_p_zones[x][y])
            elif cont_mat[x][y] == 4: #True negatives
                slopes_TN.append(slope[x][y])
                alts_TN.append(alt[x][y])
                p_zones_TN.append(p_zones[x][y])
                v_p_zones_TN.append(v_p_zones[x][y])

    #Concatenate all data:
    slopes_total.extend((slopes_TP, slopes_FP, slopes_FN, slopes_TN)) #All slope values,
    each list on one row
    slopes_df = pd.DataFrame(slopes_total).transpose() #Transpose to one list per column,
    values vertically
    slopes_df.columns= ['s_TP', 's_FP', 's_FN', 's_TN']

    alts_total.extend((alts_TP, alts_FP, alts_FN, alts_TN))
    alts_df = pd.DataFrame(alts_total).transpose()
    alts_df.columns= ['a_TP', 'a_FP', 'a_FN', 'a_TN']

```

```
    p_zones_total.extend((p_zones_TP, p_zones_FP, p_zones_FN, p_zones_TN))
    p_zones_df = pd.DataFrame(p_zones_total).transpose()
    p_zones_df.columns= ['p_TP', 'p_FP', 'p_FN', 'p_TN']

    v_p_zones_total.extend((v_p_zones_TP, v_p_zones_FP, v_p_zones_FN, v_p_zones_TN))
    v_p_zones_df = pd.DataFrame(v_p_zones_total).transpose()
    v_p_zones_df.columns= ['v_TP', 'v_FP', 'v_FN', 'v_TN']

    binspat_df = pd.concat([slopes_df, alts_df, p_zones_df, v_p_zones_df], axis=1) #
    Concatenenate horizontally

    return binspat_df



def non_spat_bin(fos_file_f, cont_mat, slope, alt, p_zones, v_p_zones, eval_tag, UIL_perc):
#NB: Script does not work if there are e.g. zero false negatives!
    import numpy as np
    import pandas as pd

    #Create emtpy lists for data storage:
    confidence_matrix = np.zeros(cont_mat.shape) #Confidence
    false_conf_matrix = np.zeros(cont_mat.shape) #False confidence
    CEM_matrix = np.zeros(cont_mat.shape) #Critical error magnitude

    slopes_C = list() #Slope values of true positives
    C_slopes = list() #Confidence values correspodning to slope values of true positives
    slopes_FC = list() #Slope values for false negatives etc..
    FC_slopes = list()
    slopes_CEM = list()
    CEM_slopes = list()
    alts_C = list()
    C_alts = list()
    alts_FC  = list()
    FC_alts = list()
    alts_CEM  = list()
    CEM_alts = list()
    p_zones_C = list()
    C_p_zones = list()
    p_zones_FC = list()
    FC_p_zones = list()
    p_zones_CEM = list()
    CEM_p_zones = list()
    v_p_zones_C = list()
    C_v_p_zones = list()
    v_p_zones_FC = list()
```

```
    FC_v_p_zones = list()
    v_p_zones_CEM = list()
    CEM_v_p_zones = list()

    TPs = list() #List of all TP-values etc.
    FPs = list()
    FNs = list()
    TNs = list()
    slopes_TP = list() #List of all slope values for TPs etc.
    slopes_FP = list()
    slopes_FN = list()
    slopes_TN = list()

    #Iterate through the contingency matrix and fill lists:
    rows, cols = cont_mat.shape
    for a in range(0,rows):
        for b in range(0,cols):
            if cont_mat[a][b] == 1: #A true positive
                confidence_matrix[a][b] = 1-fos_file_f[a][b] #Calcualte confidence and fill
                slopes_C.append(slope[a][b]) #Slope values for TPs
                C_slopes.append(confidence_matrix[a][b]) #Corresponding C-values
                alts_C.append(alt[a][b])
                C_alts.append(confidence_matrix[a][b])
                p_zones_C.append(p_zones[a][b])
                C_p_zones.append(confidence_matrix[a][b])
                v_p_zones_C.append(v_p_zones[a][b])
                C_v_p_zones.append(confidence_matrix[a][b])
                TPs.append(fos_file_f[a][b]) #List of all TP-values for the large scatter
    plot
                slopes_TP.append(slope[a][b]) #List of all slopes for TPs for the large
    scatter plot
                #The above can obviously be improved as e.g. all C-lists are effectively the
     same! May consider improving later.
            elif cont_mat[a][b] == 2: #FP
                false_conf_matrix[a][b] = 1-fos_file_f[a][b]
                slopes_FC.append(slope[a][b])
                FC_slopes.append(false_conf_matrix[a][b])
                alts_FC.append(alt[a][b])
                FC_alts.append(false_conf_matrix[a][b])
                p_zones_FC.append(p_zones[a][b])
                FC_p_zones.append(false_conf_matrix[a][b])
                v_p_zones_FC.append(v_p_zones[a][b])
                FC_v_p_zones.append(false_conf_matrix[a][b])
                FPs.append(fos_file_f[a][b])
                slopes_FP.append(slope[a][b])
            elif cont_mat[a][b] == 3: #FN
```

```
                CEM_matrix[a][b] = fos_file_f[a][b]-1
                slopes_CEM.append(slope[a][b])
                CEM_slopes.append(CEM_matrix[a][b])
                alts_CEM.append(alt[a][b])
                CEM_alts.append(CEM_matrix[a][b])
                p_zones_CEM.append(p_zones[a][b])
                CEM_p_zones.append(CEM_matrix[a][b])
                v_p_zones_CEM.append(v_p_zones[a][b])
                CEM_v_p_zones.append(CEM_matrix[a][b])
                FNs.append(fos_file_f[a][b])
                slopes_FN.append(slope[a][b])
            elif cont_mat[a][b] == 4: #TN
                TNs.append(fos_file_f[a][b])
                slopes_TN.append(slope[a][b])

    #Save files for visualization, zero as nodata!
    np.savetxt('confidence_matrix_' + eval_tag +'.txt', confidence_matrix)
    np.savetxt('false_conf_matrix_' + eval_tag +'.txt', false_conf_matrix)
    np.savetxt('CEM_matrix_' + eval_tag + '.txt', CEM_matrix)

    #Calculate Ratios
    CR = confidence_matrix.sum()/np.count_nonzero(confidence_matrix) #Confidence Ratio
    FCR = false_conf_matrix.sum()/np.count_nonzero(false_conf_matrix) #False confidence
    ratio
    CEMR = CEM_matrix.sum()/np.count_nonzero(CEM_matrix) #Critical error magnitude ratio

    #Put together lists for large scatter plot:
    positives = TPs + FNs
    negatives = FPs + TNs
    pos_slopes = slopes_TP + slopes_FN
    neg_slopes = slopes_FP + slopes_TN

    #Upper Instability Limit - Needs some more thinking, I like the idea.
    pos_sort = positives
    pos_sort.sort()
    UIL = np.percentile(pos_sort, UIL_perc) #Upper Instability Limit (stable above)
    CTN = np.count_nonzero(TNs<= UIL) #Critical True Negatives
    Critical_Cells = np.count_nonzero(fos_file_f <= UIL) #Cells below UIL

    #Save all spatial data in one DataFrame for plotting by main:
    #Empty total lists:
    slopes_tot = list()
    alts_tot = list()
    p_zones_tot = list()
    v_p_zones_tot = list()
    C_tot = list()
```

```
    FC_tot = list()
    CEM_tot = list()

    slopes_tot.extend((slopes_C, slopes_FC, slopes_CEM))
    alts_tot.extend((alts_C, alts_FC, alts_CEM))
    p_zones_tot.extend((p_zones_C, p_zones_FC, p_zones_CEM))
    v_p_zones_tot.extend((v_p_zones_C, v_p_zones_FC, v_p_zones_CEM))
    C_tot.extend((C_slopes, C_alts, C_p_zones, C_v_p_zones))
    FC_tot.extend((FC_slopes, FC_alts, FC_p_zones, FC_v_p_zones))
    CEM_tot.extend((CEM_slopes, CEM_alts, CEM_p_zones, CEM_v_p_zones))

    slopes_df = pd.DataFrame(slopes_tot).transpose()
    alts_df = pd.DataFrame(alts_tot).transpose()
    p_zones_df = pd.DataFrame(p_zones_tot).transpose()
    v_p_zones_df = pd.DataFrame(v_p_zones_tot).transpose()
    C_df = pd.DataFrame(C_tot).transpose()
    FC_df = pd.DataFrame(FC_tot).transpose()
    CEM_df = pd.DataFrame(CEM_tot).transpose()

    slopes_df.columns = ['slopes_C', 'slopes_FC', 'slopes_CEM']
    alts_df.columns = ['alts_C', 'alts_FC', 'alts_CEM']
    p_zones_df.columns = ['p_zones_C', 'p_zones_FC', 'p_zones_CEM']
    v_p_zones_df.columns = ['v_p_zones_C', 'v_p_zones_FC', 'v_p_zones_CEM']
    C_df.columns = ['C_slopes', 'C_alts', 'C_p_zones', 'C_v_p_zones']
    FC_df.columns = ['FC_slopes', 'FC_alts', 'FC_p_zones', 'FC_v_p_zones']
    CEM_df.columns = ['CEM_slopes', 'CEM_alts', 'CEM_p_zones', 'CEM_v_p_zones']

    cont_spat_df = pd.concat([slopes_df, alts_df, p_zones_df, v_p_zones_df, C_df, FC_df,
    CEM_df], axis=1)
    #Spatial data to excel: Unneccessary, and supersedes maximum rows and columns when using
     10x10m DEM.

    return CR, FCR, CEMR, cont_spat_df, positives, negatives, pos_slopes, neg_slopes, UIL,
    CTN, Critical_Cells
```

Harald Birkeland Larsen

Physically Based Modelling of Shallow Landslides with "TRIGRS"

NTNU
Norwegian University of
Science and Technology