



University of Zagreb

Faculty of Mining, Geology and Petroleum Engineering

Marko Sinčić

**STATISTICALLY-BASED
METHODOLOGY FOR LARGE SCALE
LANDSLIDE SUSCEPTIBILITY
MODELLING OF SMALL AND SHALLOW
LANDSLIDES IN THE PANNONIAN
BASIN**

DOCTORAL DISSERTATION

Zagreb, 2025



University of Zagreb

Faculty of Mining, Geology and Petroleum Engineering

Marko Sinčić

**STATISTICALLY-BASED
METHODOLOGY FOR LARGE SCALE
LANDSLIDE SUSCEPTIBILITY
MODELLING OF SMALL AND SHALLOW
LANDSLIDES IN THE PANNONIAN
BASIN**

DOCTORAL DISSERTATION

Supervisors:

Professor Snježana Mihalić Arbanas, PhD
Senior Researcher Associate Mauro Rossi, PhD

Zagreb, 2025



Sveučilište u Zagrebu

Rudarsko-geološko-naftni fakultet

Marko Sinčić

**METODOLOGIJA MODELIRANJA
PODLOŽNOSTI NA KLIZANJE MALIH I
PLITKIH KLIZIŠTA U PANONSKOM
BAZENU U KRUPNOM MJERILU
PRIMJENOM STATISTIČKIH METODA**

DOKTORSKI RAD

Mentori:

prof. dr. sc. Snježana Mihalić Arbanas, redovita profesorica
dr. sc. Mauro Rossi, viši znanstveni suradnik

Zagreb, 2025

Information about the supervisor:

Professor Snježana Mihalić Arbanas, PhD

University of Zagreb

Faculty of Mining, Geology and Petroleum Engineering

Department of Geology and Geological Engineering

Information about the supervisor:

Senior Researcher Associate Mauro Rossi, PhD

National Research Council (CNR)

Research Institute for Geo-Hydrological Protection (IRPI)

Perugia office, Via Madonna Alta 126, Italy

ACKNOWLEDGEMENTS

This doctoral thesis is part of the *LandSlidePlan (Methodology development for landslide susceptibility assessment for land-use planning based on LiDAR technology)* and *Young researchers' career development* projects, financed by the Croatian Science Foundation (HRZZ) under the grant numbers IP-2019-04-9900 and DOK-2020-01-2432, respectively.

“The day science begins to study non-physical phenomena, it will make more progress in one decade than in all the previous centuries of its existence.”

- Nikola Tesla

ABSTRACT

Approximately 30% of Croatian territory is prone to landslides, which pose a significant threat to structure, infrastructure and population. A successful disaster risk reduction management implies mitigation of landslide impact by the development of landslide hazard zonation maps, which can express in probabilistic terms, where, when and how large landslides will occur. Susceptibility is key information for an effective hazard evaluation and quantifies the spatial probability of a landslide occurrence. In the framework of this doctoral dissertation, a methodology for assessing landslide susceptibility at a large scale is established. The methodology is developed for small and shallow landslides, which are widespread in the Pannonian Basin, and it is based on the use of statistically-based methods specifically applied for large scale landslide susceptibility modelling. The proposed methodology was developed on the following research framework: (i) analysing data quality; (ii) testing classification methods for continuous landslide conditioning factors; (iii) exploring sampling strategies for landslide data (i.e. stable and unstable areas); (iv) verifying the use of diversified statistically-based methods for susceptibility assessment; and (v) verifying the use of diversified evaluation metrics. The research framework was applied in three study areas characteristic and representative of the geo-environmental setting in the Pannonian Basin, i.e. part of Hrvatsko Zagorje (20 km²), part of Podsljeme area (21 km²) and whole Podsljeme area (130 km²). The results of data quality analysis showed that preparation of a representative landslide inventory and spatially accurate landslide conditioning factors are a key step, and it can be accomplished by utilizing mainly remote sensing data, i.e. high-density LiDAR (Light Detection and Ranging) point cloud and high-resolution digital orthophoto imagery. Machine learning methods, such as neural networks and random forests resulted in the highest performance metrics while being more sensitive to certain modelling parametrization, unlike logistic regression and information value methods, which demonstrated to be more stable, still being up to par with performance metrics. Furthermore, sampling stable areas as randomly generated points has to be preferred, whereas classifying continuous landslide conditioning factors shows optimal performance metrics when resulting in >10 classes, or ideally as much as possible (i.e. usage without classification). Practical application of preliminary large scale landslide susceptibility models over larger areas, was successfully tested in the Podsljeme area (130 km²) to increase cost-effectiveness of large-area susceptibility mapping. This required the selection of a smaller but geologically and geomorphologically representative area (Podsljeme area (21 km²)), a crucial step for a reliable susceptibility assessment. The doctoral research yielded new insights into the qualitative assessment of landslide susceptibility models, which proved to be highly significant alongside commonly used quantitative assessments. The methodological framework, based on relevant research findings, defines crucial modelling steps to obtain large scale landslide susceptibility maps. Alongside necessary input data, i.e. landslide inventories and landslide conditioning factors, these maps are applicable for landslide risk reduction and mitigation.

PROŠIRENI SAŽETAK

Modeliranje podločnosti na klizanje provodi se u okviru procjene podločnosti, a rezultat su karte zoniranja podločnosti koje imaju višestruku primjenu u upravljanju opasnostima od klizišta. Unatoč brojnim znanstvenim radovima tijekom prethodnih 40 godina o procjeni i kartama podločnosti na klizanje, ne postoji definirana metodologija za modeliranje podločnosti malih i plitkih klizišta u krupnom mjerilu. Glavna svrha doktorskog istraživanja bila je razviti metodologiju za procjenu podločnosti na klizanje u krupnom mjerilu primjenom statističkih metoda i testirati ju na pilot područjima u Panonskom bazenu. Metodologija je razvijena koristeći podatke daljinskih istraživanja visoke kvalitete dobivene LiDAR (Light Detection and Ranging) tehnologijom. Prva skupina istraživačkih tema obuhvaća istraživanja prostorne točnosti ulaznih podataka za modeliranje podločnosti na klizanje koje su odgovarajuće za krupno mjerilo. U drugoj skupini su istraživanja povezana sa sustavnim, detaljnim i istovremenim testiranjem parametara modeliranja podločnosti. Treća skupina istraživačkih tema usmjerena je na istraživanje mogućnosti za razvoj jedinstvenog pristupa procjeni podločnosti na klizanje u krupnom mjerilu, kao i za razvoj preporuka koje nedostaju u literaturi.

Ciljevi i hipoteze

Cilj doktorskog istraživanja vezan za prikupljanje ulaznih podataka uključuje razvijanje slojeva ulaznih podataka za izradu relevantnih podataka o klizištima te tematskih podataka visoke rezolucije i prostorne preciznosti za modele podločnosti u krupnom mjerilu. Drugi cilj su analize raspodjele klizišta i prostorne točnosti u različitim vrstama tematskih karata u svrhu odredbe optimalnih preuvjeta klizanja za modeliranje podločnosti na klizanje u krupnom mjerilu. Sljedeći cilj uključuje primjenu više metoda klasifikacije kontinuiranih tipova podataka preuvjeta klizanja u svrhu ispitivanja utjecaja klasifikacije na modeliranje podločnosti na klizanje. Cilj vezan za pripremu podataka o klizištima i preuvjetima klizanja za primjenu različitih statističkih metoda, uključuje također korelaciju po parovima (engl. *pairwise correlation*) da se procijene korelacije između klizišta i preuvjeta klizanja. Peti cilj je iz svega navedenog razviti niz modela podločnosti na klizanje uzimajući u obzir: klasifikaciju kontinuiranih tipova podataka preuvjeta klizanja, strategiju uzorkovanja stabilnih i nestabilnih područja te primjenu različitih statističkih metoda. Specifični cilj je razviti model podločnosti na klizanje u krupnom mjerilu za cijelo pilot područje koristeći model izrađen za manji, ali geomorfološki i geološki reprezentativni dio pilot područja. Posljednji cilj je definirati zaključke i preporuke za metodologiju modeliranja podločnosti na klizanje u krupnom mjerilu

za mala i plitka klizišta u Panonskom bazenu koji će proizaći iz svih provedenih istraživanja. Navedeni ciljevi temelje se na pet hipoteza. Prva hipoteza je da LiDAR oblak točaka velike gustoće i ortofoto snimci visoke rezolucije omogućuju pripremu potrebnih ulaznih podataka za razvoj preciznih modela podložnosti na klizanje u krupnom mjerilu. Prema drugoj hipotezi, reprezentativni inventar klizišta i prostorno točne karte preduvjeta klizanja omogućuju izradu modela podložnosti na klizanje visokog stupnja točnosti i predikcije, bez obzira na korištenu statističku metodu, strategiju uzorkovanja stabilnih i nestabilnih područja ili metodu klasifikacije kontinuiranih tipova podataka preduvjeta klizanja. Treća hipoteza pretpostavlja da optimizacija kontinuiranih tipova podataka preduvjeta klizanja i nepristrano uzorkovanje stabilnih i nestabilnih područja omogućuje povećanje stupnja točnosti i predikcije modela podložnosti na klizanje u krupnom mjerilu. Četvrta hipoteza je da često korišteni parametri poput AUC (engl. *Area Under the Curve*) i Cohen kappa koeficijenta (engl. *Cohen's kappa coefficient*) indeksa ne ocjenjuju dovoljno kvalitetu modela podložnosti na klizanje u krupnom mjerilu i mogu biti dodatno upotpunjeni s: (a) kvalitativnom usporedbom uvećanog prikaza karte; (b) kartama standardne devijacije vjerojatnosti podložnosti; i (c) kvantitativnom usporedbom karata zoniranja podložnosti klasificiranih na temelju prisutnosti pojave klizišta. Posljednja hipoteza je da model podložnosti na klizanje za manji, ali reprezentativni dio geomorfološki i geološki homogenog područja omogućuje izradu modela podložnosti na klizanje visokog stupnja predikcije za cijelo pilot područje.

Znanstveni doprinos

Znanstveni doprinosi doktorskog istraživanja su: (i) unaprjeđenje postupaka za dobivanje potpunog seta prostorno točnih ulaznih podataka (tj. tematskih podataka, podataka o klizištima i podataka o elementima pod rizikom) za modeliranje podložnosti na klizanje u krupnom mjerilu koristeći LiDAR oblak točaka velike rezolucije i ortofoto snimke; (ii) definiranje jedinstvenog pristupa za klasifikaciju kontinuiranih tipova podataka preduvjeta klizanja te strategiju uzorkovanja stabilnih i nestabilnih područja za različite statističke metode modeliranja podložnosti u svrhu povećanja predikcije modela za izradu karata krupnog mjerila; (iii) novi parametri potrebni za procjenu kvalitete modela podložnosti na klizanje u krupnom mjerilu; te (iv) inovativni postupak za smanjenje troškova izrade procjene podložnosti na klizanje u krupnom mjerilu u geomorfološki i geološki homogenim područjima. Objedinjeno sve navedeno predstavlja novu metodologiju za razvoj modela podložnosti na klizanje krupnog mjerila za mala i plitka klizišta u Panonskom bazenu i geomorfološki sličnim okolišima, što je glavni znanstveni doprinos doktorskog istraživanja (v).

Metode i postupci

U doktorskom istraživanju primijenjena je metodologija za modeliranje podložnosti na klizanje koja je općenito prihvaćena i koristi ju većina istraživača prilikom provođenja procjena podložnosti na klizanje. Sastoji od prikupljanja podataka, odabira jedinice kartiranja i statističke metode, te procjene stupnja točnosti i predikcije modela. Detaljnije, metodologiju ovog istraživanja možemo definirati kroz korake koji su sažeto opisani u nastavku. Prikupljanje relevantnih podataka o klizištima uključivalo je izradu reprezentativnog inventara klizišta vizualnim kartiranjem klizišta u krupnom mjerilu na temelju izvedenih karata iz digitalnog modela terena (DTM) visoke rezolucije (engl. *High resolution*) (HR-DTM), odnosno morfometrijskih karata, kako je opisano u Bernat Gazibara et al. (2019b). Optimalni HR-DTM odabran je na temelju kvalitativne procjene 24 različita HR-DTM-a izvedenih koristeći četiri metode interpolacije i šest rezolucija (Sinčić et al., 2022c). Reprezentativnost inventara klizišta određena je pomoću dijagrama učestalosti veličine površina klizišta (engl. *frequency-size distribution*), a dio kartiranih klizišta potvrđen je terenskom provjerom. Relevantni tematski podaci, potrebni za izradu karata preduvjeta klizanja, odabrani su na temelju okolišnih uvjeta na pilot područjima i uključivali su oko 20 slojeva koji pripadaju geomorfološkim, geološkim, hidrološkim ili antropogenim preduvjetima klizanja. Klasifikacije kontinuiranih tipova podataka preduvjeta klizanja napravljene su na sljedeće načine: nekoliko interpretacija krivulje studentiziranog kontrasta (engl. *studentized contrast*) (npr. Neuhäuser et al., 2012); granične vrijednosti definirane heuristički na temelju gustoće klizišta ili pomoću GIS alata odnosno prirodno grupiranje (engl. *Natural breaks*), kvantili (engl. *quantile*) i geometrijski interval (engl. *Geometrical interval*); i bez klasifikacije. Priprema, uređivanje i dovršavanje karata preduvjeta klizanja provedeni su na način da se zadovolji prostorna točnost modeliranja u krupnom mjerilu. Različite strategije za uzorkovanje nestabilnih područja uključivale su centroide, poligone, zaglađivanje terena (engl. *smooth*) na lokacijama klizišta i izrade *buffer* zona u okolici poligona klizišta. Izrada inventara stabilnih područja omogućila je uzorkovanje stabilnih područja kao centroida, poligona i *buffer* zona u okolici stabilnih poligona. Također, stabilna područja su uzorkovana kao nasumično generirane točke na području cijelog pilot područja. Kao jedinica za kartiranje koja je odgovarajuća, za modeliranje podložnosti na klizanje u krupnom mjerilu, korišteni su pikseli veličine 5 m budući da su alternativne jedinice nagiba terena dokazane kao neprihvatljive prema Bernat Gazibara et al. (2023b). Za modeliranje podložnosti na klizanje korišteno je šest statističkih metoda navedenih u Reichenbach et al. (2018), a koje pripadaju trima različitim pristupima unutar metoda temeljenih na podacima (engl. *data-driven*) u: (1) metoda informacijske vrijednosti (engl.

Information value) i metoda težinskih faktora (engl. *Weight of Evidence*) odgovaraju bivarijantnom pristupu primijenjenom u npr. van Westen et al. (1993); (2) logistička regresija (engl. *Logistic Regression*) odgovara multivarijantnom pristupu primijenjenom u npr. Rossi et al. (2021); i (3) metoda potpornih vektora (engl. *Support Vector Machine*), neuronske mreže (engl. *Neural Network*) i slučajne šume (engl. *Random Forests*) su pristupi strojnog učenja primijenjenog u npr. Merghadi et al. (2020). Procjena točnosti i predikcije (validacije) modela provedena je na više sljedećih načina primjenom: metoda krivulja operativnih karakteristika (engl. *Receiver Operating Characteristic*) (ROC), odnosno AUC-a za izražavanje stope uspješnosti (engl. *Success rate*) ili predikcije (engl. *Prediction rate*); AUC-a definiranog osjetljivošću (engl. *Hit Rate*) i specifičnošću (engl. *False Alarm Rate*); Cohen kappa koeficijenta (kako je primijenjeno u Rossi et al., 2021); i drugih pokazatelja kvalitete. Pouzdanost modela procijenjena je kartama standardne devijacije vjerojatnosti podložnosti kako bi se utvrdile zone niže pouzdanosti u prostoru u kojima modeli pokazuju razlike. Osim toga, provedena je kvalitativna i kvantitativna usporedba prikaza zona podložnosti u odnosu na stvarni prikaz terena. Također su kvantitativno uspoređene zone podložnosti i inventar klizišta na način da su provedene klasifikacije podložnosti na temelju zastupljenosti klizišta u pojedinim zonama (iskazano kao površina klizišta, za razliku od broja klizišta). Za istraživanja na području grada Zagreba na području istraživanja u Podsljemenskoj zoni, odabrano je manje pilot područje koje je geomorfološki i geološki reprezentativno. Na njemu je definiran niz preduvjeta podložnosti koristeći reprezentativni inventar klizišta, koji su zatim korišteni za razvoj modela podložnosti na klizanje za cijelu Podsljemensku zonu. Provedena je validacija modela podložnosti primijenjenog na većem pilot području kako bi se procijenile mogućnosti primjene ovog pristupa koji značajno smanjuje troškove izrade procjene podložnosti na klizanje.

Rezultati i zaključci

Glavni rezultat doktorske disertacije je metodologija za modeliranje podložnosti na klizanje u krupnom mjerilu za mala i plitka klizišta, testirana i validirana na pilot područjima u Panonskom bazenu. Pilot područje Podsljemenske zone (Zagreb) je urbani okoliš, a pilot područje u Hrvatskom Zagorju ruralni okoliš. Metodologija je razvijena na temelju 87 modela podložnosti na klizanje izrađenih od podataka o klizištima i preduvjetima klizanja visoke kvalitete koristeći tri grupe statističkih metoda (bivarijantne, multivarijantne i metode strojnog učenja). Svi rezultati doktorskog istraživanja su prikazana kroz četiri znanstvena rada. Metodološki doprinosi doktorske disertacije, zajedno sa zaključcima su dani kroz tri skupine

istraživačkih tema: (1) prostorna točnost u modeliranju u krupnom mjerilu; (2) sustavno testiranje parametara modeliranja; (3) jedinstveni pristupi i preporuke za krupno mjerilo.

Rezultati prve skupine istraživačkih tema relevantnih za prostornu točnost u modeliranju u krupnom mjerilu su vezani za:

(i) analize svih dostupnih ulaznih podataka za modeliranje podložnosti u krupnom mjerilu u Hrvatskom Zagorju,

(ii) analize mogućnosti izrade tematskih karata geoloških, geomorfoloških, hidroloških i antropogenih preduvjeta klizanja,

(iii) analize primjenjivosti karata preduvjeta klizanja u procjeni podložnosti na klizanje s obzirom na prostornu točnost,

(iv) donošenje zaključaka o postupcima i standardiziranim preporukama za izradu karata preduvjeta klizanja.

Rezultati druge skupine istraživačkih tema relevantni za sustavno testiranje parametara modeliranja dobiveni su temeljem:

(i) izrade 32 modela podložnosti na klizanje koristeći metode logističke regresije, slučajnih šuma, neuronskih mreža i metode potpornih vektora u kombinaciji s osam strategija uzorkovanja stabilnih i nestabilnih područja;

(ii) izrade 54 modela podložnosti na klizanje koristeći metode informacijske vrijednosti, logističke regresije, slučajnih šuma, neuronske mreže i metode potpornih vektora u kombinaciji s 11 scenarija za klasifikaciju kontinuiranih tipova podataka;

(iii) sustavne analize u svrhu razvoja jedinstvenog pristupa i preporuka za modeliranje podložnosti na klizanje u krupnom mjerilu.

Zaključci treće skupine istraživačkih tema relevantni za jedinstveni pristup i preporuke za krupno mjerilo su glavni zaključci doktorskog istraživanja koji proizlaze iz rezultata istraživanja iz prve dvije skupine. Glavni zaključci disertacije su sljedeći:

(i) LiDAR oblak točaka velike gustoće, ortofoto snimci visoke rezolucije i javno dostupni slojevi podataka predstavljaju nužne i dovoljne ulazne podatke za izradu reprezentativnih karata inventara klizišta i prostorno točnih preduvjeta klizanja.

(ii) LiDAR oblak točaka velike gustoće i ortofoto snimci visoke rezolucije omogućuju verifikaciju svih ulaznih podataka, odnosno provjeru prostorne točnosti zbog toga jer prikazuju stvarne uvjete na terenu.

(iii) Korištenje reprezentativnog inventara klizišta i prostorno točnih karata preduvjeta klizanja u modeliranju podložnosti koristeći statističke metode rezultira modelima podložnosti s visokim parametrima ocjene modela.

(iv) Korištenje odgovarajućih parametara modeliranja omogućuje metodama neuronske mreže i slučajne šume da rezultiraju najvišim parametrima ocjene modela, a logistička regresija i informacijska vrijednost metode pokazuju manju osjetljivost na parametre modeliranja i vjerojatnije rezultiraju sa zadovoljavajućim (i uprosječenim) rezultatima u odnosu na metode neuronske mreže i slučajne šume.

(v) Korištenje rastera kao neklasificiranih preduvjeta klizanja i linijskih vektora s velikim brojem *buffer* zona optimizira i pojednostavljuje primjenu kontinuiranih tipova podataka preduvjeta klizanja, dok je nizak broj klasa kod rastera ili linijskih vektora nepovoljna alternativa.

(vi) Uzorkovanje nestabilnih područja kao centroida i uzorkovanje stabilnih područja subjektivno (npr. inventar stabilnih područja) rezultira značajnim prostornim odstupanjem zona podložnosti. S druge strane, zaglađivanje terena (engl. *smooth*) na lokacijama klizišta može dovesti do povećanja vrijednosti ocjena parametara modeliranja dok nasumično generirana stabilna područja predstavljaju odgovarajuću opciju.

(vii) Parametri kao što su veličina i distribucija zona podložnosti na klizanje u odnosu na stvarno stanje terena, distribucija pojava klizišta u zonama, procjena pouzdanosti i/ili varijabilnosti zona podložnosti na više karata (npr. karte standardnih devijacija) te kvalitativna procjena u odnosu na stvarno stanje terena (npr. uvećani prikaz karte) omogućuju nove uvide u modele podložnosti na klizanje koji su nužni za procjenu primjenjivosti karata zoniranja podložnosti, a koje često korišteni parametri za procjenu stupnja točnosti i predikcije (npr. Cohen's kappa i AUC) ne omogućavaju.

(viii) Inventar klizišta izrađen za manji, ali reprezentativni dio geomorfološki i geološki homogenog područja omogućuje izradu modela podložnosti na klizanje visokog stupnja predikcije za cijelo pilot područje, značajno povećavajući isplativost procjene financijski i vremenski, smanjujući vrijeme izrade modela.

(ix) Bivarijantne metode (kao što su metoda težinskih faktora i informacijske vrijednosti korištene u ovom istraživanju), koje se mogu okarakterizirati kao zastarjele, pokazale su vrlo dobre rezultate u modeliranju podložnosti na klizanje, isto kao i algoritmi strojnog učenja kao što su slučajne šume, neuronske mreže ili potporni vektori.

(x) Kvaliteta ulaznih podataka ostaje najvažniji (vjerojatno i najzahtjevniji) faktor u modeliranju podložnosti na klizanju. Drugim riječima, koraci u pripremi reprezentativnog inventara klizišta i prostorno točnih preduvjeta klizanja trebaju biti u centru pozornosti istraživača.

KEYWORDS

landslide

susceptibility modelling

statistically-based methods

large scale

remote sensing

LiDAR

Pannonian Basin

KLJUČNE RIJEČI

klizište

modeliranje podložnosti

statističke metode

krupno mjerilo

daljinska istraživanja

LiDAR

Panonski bazen

TABLE OF CONTENTS

1. INTRODUCTION.....	1
1.1 Overview of previous research.....	1
1.2 Study area.....	2
1.3 Material	5
1.4 Methodology	9
1.5 Plan of research	13
1.6 Objectives and hypotheses	14
1.7 Scientific contribution	15
1.8 Doctoral dissertation structure.....	16
 2. ORIGINAL SCIENTIFIC PAPERS.....	 17
2.1 The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments.....	17
2.2 Comparison of conditioning factor classification criteria in large-scale statistically based landslide susceptibility models.....	55
2.3 A comprehensive comparison of stable and unstable area sampling strategies in large-scale landslide susceptibility models using machine learning methods.....	87
2.4 Landslide susceptibility assessment on a large scale in the Podsljeme area, City of Zagreb (Croatia)	122
 3. DISCUSSION	 136
3.1 Hypothesis (1)	136
3.2 Hypothesis (2)	137
3.3 Hypothesis (3)	139
3.4 Hypothesis (4)	140
3.5 Hypothesis (5)	141
3.6 Perspectives for future research	143
 4. CONCLUSION.....	 146
 5. LITERATURE	 149
 6. BIOGRAPHY OF THE AUTHOR	 168

LIST OF ABBREVIATIONS:

ALS	Airborne Laser Scanning
AUC	Area Under the Curve
BGM	Basic Geological Map
CTI	Compound Topographic Index
DEGM	Detailed Engineering Geological Map
DEM	Digital Elevation Model
DOF	Digital Orthophoto
DTM	Digital Terrain Model
EG	Engineering Geological
GIS	Geographic Information System
HR	High Resolution
IV	Information Value
LCF	Landslide Conditioning Factor
LiDAR	Light Detection and Ranging
LR	Logistic Regression
LSA	Landslide Susceptibility Assessment
LSM	Landslide Susceptibility Model
LSMg	Landslide Susceptibility Modelling
NN	Neural Networks
OSM	Open Street Map
PT	Proximity To
RF	Random Forests
ROC	Receiver Operating Characteristic
SD	Standard Deviation
SfM	Structure from Motion
SGA	State Geodetic Administration
SEI	Site Exposure Index
SVM	Support Vector Machine
WoE	Weight of Evidence
WMS	Web Map Service

LIST OF FIGURES

Figure 1-1 Geographic positions of study areas; in Europe (A), in Croatia (B), in NW Croatia (C). Detailed presentation of the study areas: whole Podsljeme Area in the City of Zagreb (D), selected representative Podsljeme Area (E) and selected representative part of Hrvatsko Zagorje (F)	4
Figure 1-2 Detailed workflow for LSAs, preliminary example for Chapter 2.2.....	10
Figure 1-3 Detailed workflow for LSAs, preliminary example for Chapter 2.3.....	11

LIST OF TABLES

Table 1-1 Overview of quantitative parameters of landslide inventories in Podsljeme Area (21 km ²) and Hrvatsko Zagorje (20 km ²).....	5
Table 1-2 Overview of source data used for LSMg in Podsljeme Area (21 km ²) (Chapter 2.2), Hrvatsko Zagorje (20 km ²) (Chapter 2.3) and Podsljeme Area (130 km ²) (Chapter 2.4).....	6
Table 1-3 Overview of source data for considered (and discarded) LCFs used in preliminary susceptibility analysis in Podsljeme Area (21 km ²) (Chapter 2.2).....	7
Table 1-4 Overview of source data for considered (and discarded) LCFs used in preliminary susceptibility analysis in Hrvatsko Zagorje (20 km ²) (Chapter 2.3)	8
Table 1-5 Relations of hypotheses and objectives of the doctoral research	15

1. INTRODUCTION

1.1 Overview of previous research

A landslide is the movement of a mass of rock, earth or debris down a slope (Cruden, 1991), whereas landslide zoning maps present territory divided into homogeneous areas considering the degree of susceptibility, hazard or risk (Fell, 2008a). Brabb (1984) defined landslide susceptibility as a ‘likelihood of a landslide occurring in a given area’, i.e. depicting only the spatial component of a landslide occurrence. Most significant recommendations, definitions, guidelines and overviews on landslide susceptibility assessment (LSA) were published in the last approx. 40 years in Varnes (1984), Soeters & van Westen (1996), Guzzetti et al. (1999), van Westen et al. (2003), van Westen et al. (2006), Guzzetti et al. (2006a,b), van Westen et al. (2008), Fell et al. (2008a,b), Corominas et al. (2014) and Reichenbach et al. (2018). Moreover, in recent years, some of the most important review papers dealing with specific topics of landslide susceptibility modelling (LSMg) and mapping are: GIS-based (Geographic Information System) modelling methodologies (Lee, 2019), hazard zonation techniques (Shano et al., 2020), machine learning algorithm comparison (Merghadi et al., 2020), a review of national case-studies (Dias et al., 2021) or review of studies from the certain geographical regions (Das et al., 2023). In a review paper, Reichenbach et al. (2018) emphasised the importance of statistically-based LSMg by analysing more than 550 peer-reviewed scientific articles published between 1983 and 2016. The paper considers the quality of LSAs through the evaluation criteria defined in Guzzetti et al. (2006a,b) and suggests steps for preparing LSAs. One of the paper’s main conclusions is that landslide investigators are more eager to experiment with different modelling techniques rather than acquiring good-quality input data.

Testing of modelling techniques according to particular steps of LSAs includes testing of different statistical methods (e.g. Wang et al., 2016; Chen et al., 2017), mapping units (e.g. Bornaetxea et al., 2018; Jacobs et al., 2020), inventory types (e.g. Guzzetti et al., 2012; Petschko et al., 2016), landslide conditioning factors (LCFs) (e.g. Donati & Turrini, 2002; Jebur et al., 2014; Gaidzik and Ramírez, 2021) and sampling strategies of stable (e.g. Hong et al., 2019; Fu et al., 2023) or unstable areas (e.g. Hussin et al., 2016; Dou et al., 2020). Literature related to software and hardware development and method comparisons frequently emphasize their advantages and disadvantages to reach satisfactory landslide susceptibility zonation (e.g. Tien Bui et al., 2016; Pourghasemi & Rahmadi, 2018; Merghadi et al., 2020; Youssef & Pourghasemi, 2021). Regardless of modelling techniques, remote sensing remains an essential

technology for input data collection as it enables obtaining both relevant landslide data as well as thematic data for LCFs (e.g. Chigira, 2004; Razak, 2011; Jaboyedoff et al., 2012; Razak et al., 2013; Scaioni et al., 2014; Tarolli, 2014; Ray et al., 2020). Despite the long tradition in the development of LSAs and known well-defined steps in LSAs, there are still open questions essential for optimal and unbiased LSMg related to: (i) measuring completeness and spatial accuracy of acquired data, (ii) usage of LCFs, (iii) unstable and stable area sampling, (iv) selection of statistical methods for LSMg, and (v) application of metrics for evaluation of LSA.

This research is focused on modelling landslide susceptibility of small and shallow landslides which occur commonly in the Pannonian Basin posing a substantial threat in hilly areas in Croatia (Vujnović, 2005; Mihalić et al., 2008; Bernat et al., 2014a,b; Bernat Gazibara et al., 2017a,b,c; Mihalić Arbanas et al., 2017; Bernat Gazibara 2019; Bernat Gazibara et al., 2019a,b; Sinčić et al., 2022a; Mihalić Arbanas et al., 2023a). The main purpose of the research is to develop a statistically-based methodology for large scale LSAs and to test and validate it in study areas representative of the Pannonian Basin geo-environmental setting. LSMg was performed using bivariate, multivariate and machine learning statistical methods based on high-quality data derived by remote sensing LiDAR (Light Detection and Ranging) technique. Special attention was given to obtaining good-quality input thematic and landslide data representing all characteristics of the targeted environments necessary for a large scale LSA.

Despite many papers that have explored the abovementioned open questions in LSMg, the research and development of a methodology for large scale opens additional questions specific to this scale of LSMg, and this research focuses on more of them. The first group of research topics deals with adequate spatial accuracy of input data for large scale modelling. A second group faces systematic testing and application of diversified susceptibility modelling settings to identify more effective methods. Finally, the third group aims to explore possibilities for standard large scale approaches and recommendations, which are missing in the literature.

1.2 Study area

The doctoral research is performed in the Croatian part of the Pannonian Basin, SE Europe. (Figure 1-1A). The Pannonian Basin is a megageomorfologic region (Bognar, 1999) which is in contact with the Dinarides in central Croatia and stretches further north and east into Hungary and, to a lesser extent, Serbia and Slovenia (Mihalić Arbanas et al., 2017).

Figure 1-1B shows City of Zagreb (County) and Varaždin County in NW part of Croatia where three study areas are located. Namely, whole large Podsljeme Area (130 km²) (Figure 1-

1D) and smaller part of Podsljeme Area (21 km²) (Figure 1-1E) represent two study areas in the City of Zagreb. The Podsljeme Area (21 km²) was chosen within larger Podsljeme Area (130 km²) as a study area to represent typical geomorphological settings of the entire Podsljeme Area (130 km²). Considering the local administrative division, the third study area is in Varaždin County, split between the City of Lepoglava and the Bednja Municipality (Figure 1-1C). Namely, Figure 1-1F shows the Hrvatsko Zagorje study area (20 km²) in more detail, typical for the rural part of NW Croatia in the Pannonian Basin.

Podsljeme Area and Hrvatsko Zagorje are hilly terrains with a significant amount of steep slopes formed in soils and soft rocks, making them prone to landslide processes. Climate in NW Croatia is continental, with a mild maritime influence and a mean annual rainfall of 887 mm (Podsljeme Area) and 874 mm (Hrvatsko Zagorje). In both study areas, rainfall distribution is also similar and can be described as unimodal with one dry season from December to March and one wet season from April to November. In NW Croatia, mean annual air temperature ranges from 6 to 11°C whereas in Podsljeme Area and Hrvatsko Zagorje the values are approximately 9 to 10°C (Zaninović et al., 2008). Long periods of rainfall and short periods of intensive rainfall with possible rapid snow melting, are the main triggers of landslides (Bernat et al., 2014a). Also, anthropogenic activity may significantly impact landslide occurrence (Bernat et al., 2014a,b). The majority of recently activated landslides can be classified as small (10-10³ m²) to medium-sized (<10⁵ m²) and superficial to moderately shallow (<20 m) soil slides (Mihalić Arbanas et al., 2017). A noteworthy exception is a large deep-seated Kostanjek landslide artificially caused by mining activities in the Podsljeme Area (Krkač et al., 2017, 2021).

The main difference between the Podsljeme Area and Hrvatsko Zagorje is the degree of urbanization. Namely, Podsljeme Area is highly urbanized, mostly residential area characterised by a dense road network, large number of buildings and high population density. On the other hand, Hrvatsko Zagorje has an extremely low population density, representing a typical Croatian rural environment in NW Croatia. Regardless of the population, in both areas, landslides caused significant material damage (mainly on roads) and related losses, especially because of the negative influence of climate changes and, more often, extreme precipitations in the last 30 years (Bernat et al., 2014a). In both areas, tools for effective hazard and risk management, such as reliable landslide maps (inventory and susceptibility maps), are missing, contributing to higher landslide risk caused by threats from small and shallow landslides.

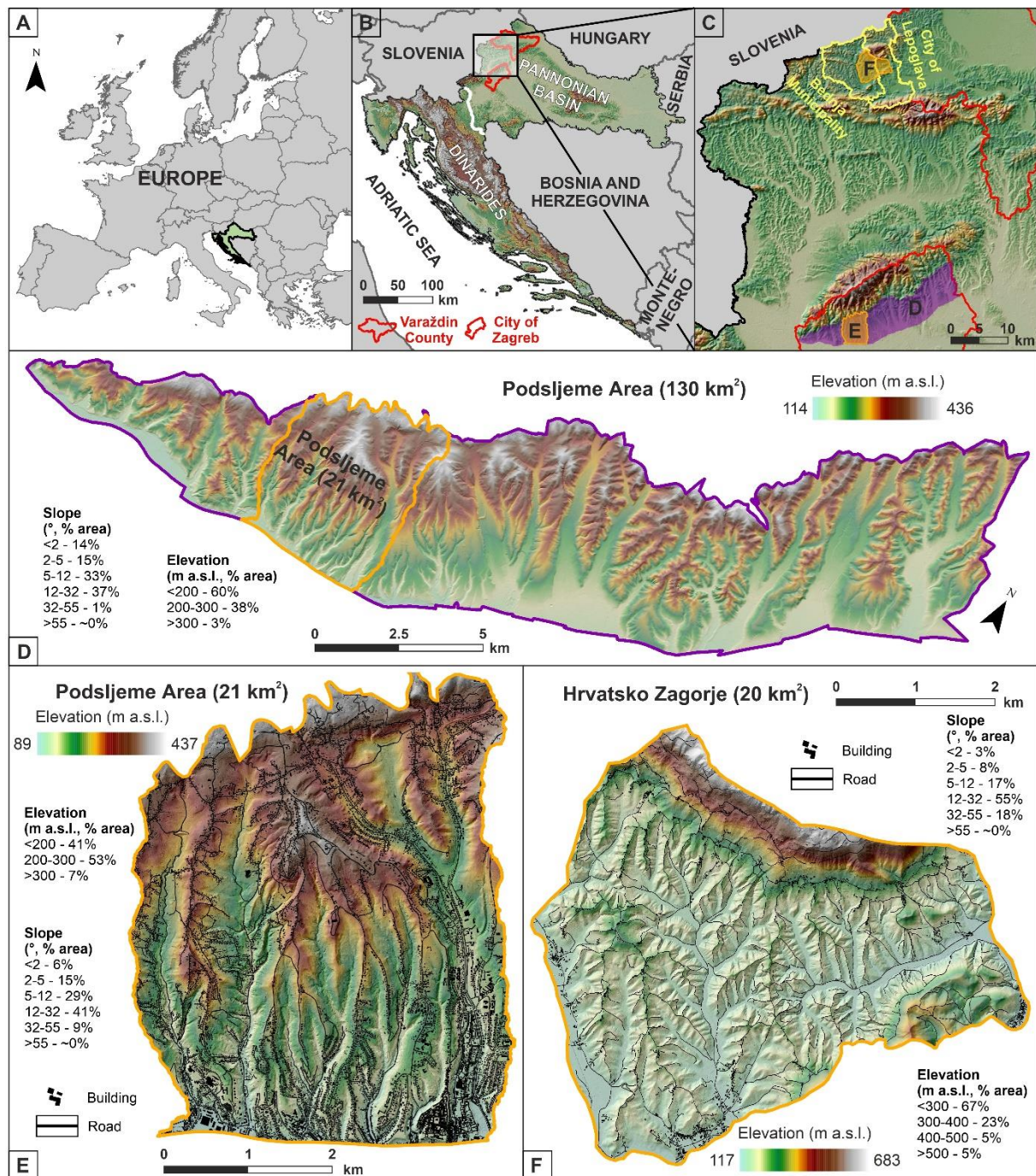


Figure 1-1 Geographic positions of study areas; in Europe (A), in Croatia (B), in NW Croatia (C). Detailed presentation of the study areas: whole Podsljeme Area in the City of Zagreb (D), selected representative Podsljeme Area (E) and selected representative part of Hrvatsko Zagorje (F)

The selection of study areas for the development of large scale LSA methodology was guided by the idea of satisfying the basic conditions for improvement of landslide risk management policies and practices, i.e. landslide susceptibility mapping and zonation. Reliable zonation maps in a large scale are the basic tool for mitigating the physical and economic impacts of landslides as well as their social consequences. Alongside a methodology to conduct landslide susceptibility models (LSMs) at a large scale, the doctoral research yields to the

realization of landslide susceptibility zonation maps emphasizing their applicability in the urban and spatial planning systems at local level that is crucial for decisions related to land use, construction and development.

Detailed information about Podsljeme Area (21 km²), Hrvatsko Zagorje (20 km²) and Podsljeme Area (130 km²) is found individually in Chapter 2.2 (Sinčić et al., 2025), Chapter 2.1 (Sinčić et al., 2022b) and Chapter 2.3 (Sinčić et al., 2024a), and Chapter 2.4 (Bernat Gazibara et al., 2023a), respectively.

1.3 Material

Airborne laser scanning (ALS) was performed in 2020 for the study areas Hrvatsko Zagorje (20 km²) and Podsljeme Area (21 km²) in the framework of the Croatian Science Foundation scientific project (Bernat Gazibara et al., 2022). As a result, a LiDAR derived high density bare earth point cloud is obtained, having ~15 points per 1 m² and ~0.3 m average point spacing. The LiDAR data were used to derive high resolution (HR) digital terrain models (DTMs) up to 15 cm resolution, i.e. source data for acquiring landslide data and thematic data. In addition, landslide data and other topographic data obtained in previous research of the bilateral Croatian-Japanese SATREPS FY2008 (Mihalić & Arbanas, 2012; Arbanas et al., 2017) and SATREPS FY2013 scientific projects were used, together with existing official and a few other public maps.

The landslide inventory for the Podsljeme Area (21 km²), used in this doctoral research, was originally developed by Bernat Gazibara et al. (2019a,b), but it was additionally supplemented and verified with the 2020 LiDAR data in this research (Table 1-1). Furthermore, a representative landslide inventory for Hrvatsko Zagorje study area was originally derived as a part of this research (Table 1-1).

Table 1-1 Overview of quantitative parameters of landslide inventories in Podsljeme Area (21 km²) and Hrvatsko Zagorje (20 km²)

	Polygons (N)	Density (N/km ²)	Min. (m ²)	Max. (m ²)	Mean (m ²)	Median (m ²)	St.dev. (m ²)	Total size (km ²)	Total size (km ²)
Podsljeme Area (21 km ²)	709	33	43	8064	726	425	941	0.5	2.4
Hrvatsko Zagorje (20 km ²)	912	45	3.3	13779	448	173	880	0.4	2.0

In the larger study area of the Podsljeme Area (130 km²), the Structure from Motion (SfM) Digital Elevation Model (DEM) from the SATREPS FY2008 project (Mihalić Arbanas et al., 2016; Arbanas et al., 2017) was used to derive thematic data for LCFs. Landslides were

mapped based on 1 m LiDAR DTM derivatives obtained in the framework of SATREPS FY2013 project. Besides that, additional cartographic data for the two study areas in the Podsljeme Area included only Basic Geological Maps (BGMs), Detailed Engineering Geological Map (DEGM), Topographic Map of Croatia and Land Use Map of the City of Zagreb. A summary of the source data used for LSMg in the doctoral research is given in Table 1-2.

Table 1-2 Overview of source data used for LSMg in Podsljeme Area (21 km²) (Chapter 2.2), Hrvatsko Zagorje (20 km²) (Chapter 2.3) and Podsljeme Area (130 km²) (Chapter 2.4)

Source data	Reference
Podsljeme Area (21 km²)	
LiDAR based DTM	Institutional data
BGM Sheet Zagreb	Šikić et al., 1978
DEGM	Miklin et al., 2007
Topographic Map of Croatia	URL-1
Land Use Map of City of Zagreb	City of Zagreb, 2011
Hrvatsko Zagorje (20 km²)	
LiDAR point cloud	Institutional data
LiDAR based DTM	Institutional data
LiDAR based HR-DTM derivatives	Institutional data
BGM Sheet Varaždin and BGM Sheet Rogatec	Šimunić et al., 1982, Aničić & Juriša, 1984
Topographic Map of Croatia	URL-1
DOF	URL-2
Open Street Map (OSM)	URL-3
Land Use Planning Maps	URL-4, URL-5
Podsljeme Area (130 km²)	
SfM DEM	Institutional data
LiDAR based DTM	Institutional data
BGM Sheet Zagreb and BGM Sheet Ivanić Grad	Šikić et al., 1978; Basch, 1983
Land Use Map of City of Zagreb	City of Zagreb, 2011

LiDAR point cloud, the HR-DTM derivatives, BGMs, Topographic Map of Croatia, Digital Orthophoto (DOF) Maps, OSM and Land Use Planning Maps were used in Hrvatsko Zagorje (Table 1-2). Due to limited data in this study area, a series of additional spatial analyses were performed to derive input data from other sources: EU-DEM (URL-6), Sentinel-2 DOF (URL-7), different State Geodetic Administration (SGA) Web Map Service (WMS) servers (e.g URL-8), SGA input data for DTM derivation (URL-9), and Corine Land Cover (URL-10). The complete source data overview and results of all analyses are described in a separate paper (Chapter 2.1). However, they are not listed in Table 1-2 because they were discarded from the susceptibility model in later phases of research.

The following tables give a summarized overview of all analysed LCFs in a particular study area, together with source data and notes if they were discarded in the final LCF selection. Table 1-3 for the Podsljeme Area (21 km²) and Table 1-4 for the Hrvatsko Zagorje (20 km²) show that approximately 50% more LCFs were used in preliminary analyses than in the final LSMg.

Table 1-3 Overview of source data for considered (and discarded) LCFs used in preliminary susceptibility analysis in Podsljeme Area (21 km²) (Chapter 2.2)

Thematic data group	LCF	Source data				
		LiDAR based DTM	BGM	DEGM	Topographic Map of Croatia	Land use map of City of Zagreb
Geomorphological	Elevation	X				
	Aspect	X				
	Slope	X				
	Dissection*	X				
	Curvature*	X				
	Roughness*	X				
	Contour density*	X				
Geological	BGM lithology		X			
	DEGM lithology*			X		
	PT ¹ BGM faults		X			
	PT ¹ DEGM faults*			X		
	PT ¹ BGM lithology contact		X			
	PT ¹ DEGM lithology contact*			X		
Hydrological	PT ¹ Drainage network	X				
	PT ¹ permanent streams*				X	
	PT ¹ all streams				X	
	Compound topographic index	X				
Anthropogenic	Land use					X
	PT ¹ Land use contact*					X
	PT ¹ buildings*					X
	PT ¹ roads*					X
	PT ¹ buildings and roads*					X

*discarded in final LSMg (Chapter 2.2)

¹ Proximity To

Table 1-4 Overview of source data for considered (and discarded) LCFs used in preliminary susceptibility analysis in Hrvatsko Zagorje (20 km²) (Chapter 2.3)

Thematic data group	LCF	Source data							
		LiDAR point cloud	LiDAR based DTM	LiDAR based HR-DTM derivatives	BGM	Topographic Map of Croatia	DOF	OSM	Land use planning maps
Geomorphological	Elevation		X						
	Aspect		X						
	Slope		X						
	Dissection*		X						
	Curvature*		X						
	Roughness*		X						
	Contour density*		X						
	Landform		X						
Geological	BGM lithology*				X				
	EG ¹ lithology			X	X				
	PT BGM faults*				X				
	PT BGM lithology contact*				X				
	PT EG ¹ lithology contact			X					
Hydrological	PT drainage network		X		X				
	PT permanent streams*					X			
	PT all streams*					X			
	PT springs*					X			
	CTI ^{2*}		X						
	SEI ³		X						
	IMI ⁴		X						
Anthropogenic	Land use	X		X		X	X	X	
	PT land use contact	X		X		X	X	X	
	PT buildings*	X							
	PT roads			X			X	X	X
	PT buildings and roads*	X		X			X	X	X

*discarded in final LSMg (Chapter 2.3)

¹ Engineering Geological

² Compound Topographic Index

³ Site Exposure Index

⁴ Integrated Moisture Index

1.4 Methodology

The methodology for LSMg used in doctoral research is commonly accepted and used by most researchers when conducting LSAs. Generally, it consists of input data acquisition (which includes relevant thematic data and landslide data), followed by a selection of mapping unit and statistical method, and lastly, evaluating model fitting and predictive performance followed by uncertainty assessment, as described in the interrelated steps (Reichenbach et al., 2018). Examples of detailed workflows used for analyses described in Chapter 2.2 and Chapter 2.3 are given in Figure 1-2 and Figure 1-3, respectively.

Obtaining relevant landslide data included compiling a representative landslide inventory by visual expert-based mapping using large scale HR-DTM derivatives (Razak et al., 2011), i.e. morphometric maps, as described in Bernat Gazibara (2019). Optimal HR-DTM was selected based on a qualitative assessment of 24 HR-DTMs derived by four interpolation methods and six resolutions (Sinčić et al., 2022c). A frequency-size distribution graph was used to evaluate the landslide inventory representativeness (Malamud et al., 2004), whereas the mapped landslides were verified by field survey. Relevant thematic data needed to derive LCF maps were selected based on the environmental settings of the study areas and in appropriate spatial accuracy (Krkač et al., 2023), including multiple data sources as described in Chapter 1.3. Considering the appropriate mapping unit, a 5 m pixel-based grid was used for large scale LSMg, since it has been proven by preliminary research, in the same study area (Bernat Gazibara et al., 2023b), that alternative slope units (Alvioli et al., 2016) are unacceptable. Six statistical methods grouped in three data-driven approaches were used in LSMg: (1) bivariate approach, i.e. information value (IV) (Yin & Yan, 1988) applied in Sarkar et al. (2013), Farooq & Akram (2021) and Krkač et al. (2023) and weights of evidence (WoE) (Bonham-Carter et al., 1989; Agterberg et al., 1990; Bonham-Carter, 1994) applied in van Westen (2002), Neuhäuser et al. (2012) and Sinčić et al. (2022a); (2) multivariate approach i.e. logistic regression (LR) (Cox, 1958) applied in Rossi et al. (2010), Hemasinghe et al. (2018) and Bornaetxea et al. (2018); and (3) machine learning approach i.e. support vector machine (SVM) (Cortes & Vapnik, 1995; Vapnik, 1995) applied in Yao et al. (2008), Pradhan (2013) and Kavzoglu et al. (2014), neural networks (NN) (Ripley, 1996; Venables & Ripley, 2002) applied in Lee (2007), Nefeslioglu et al. (2008), and Pascale et al. (2013) and random forests (RF) (Ho, 1995; Breiman, 2001) applied in Catani et al. (2013), Wang et al. (2021) and Sandić et al. (2023). Model fitting and predictive performance were evaluated by the Receiver Operating Characteristic (ROC) curve and related Area Under the Curve (AUC) and/or Cohen's kappa index based on hit rate and false alarm rate (Fawcett, 2006), i.e. derived from the confusion

matrix (Jolliffe & Stephenson, 2003). Moreover, AUC based on success and prediction rate curves (Chung & Fabri, 1999; Chung & Fabri, 2003) is also considered, observing cumulative landslide area against cumulative study area. Model uncertainty was estimated by standard deviation (SD) maps of probabilistic susceptibility values to estimate zones where models' predictions show low or high variability.

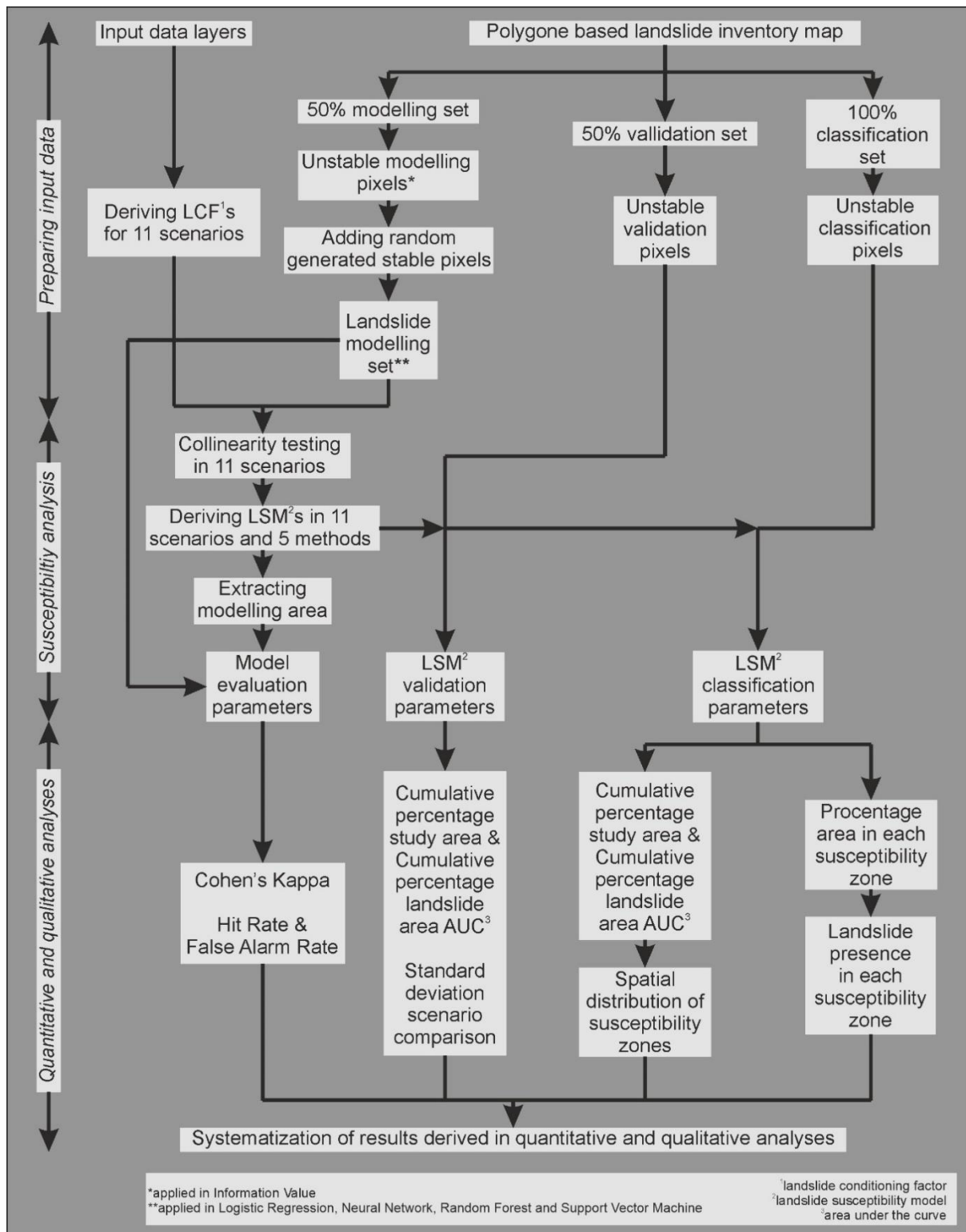


Figure 1-2 Detailed workflow for LSAs, preliminary example for Chapter 2.2

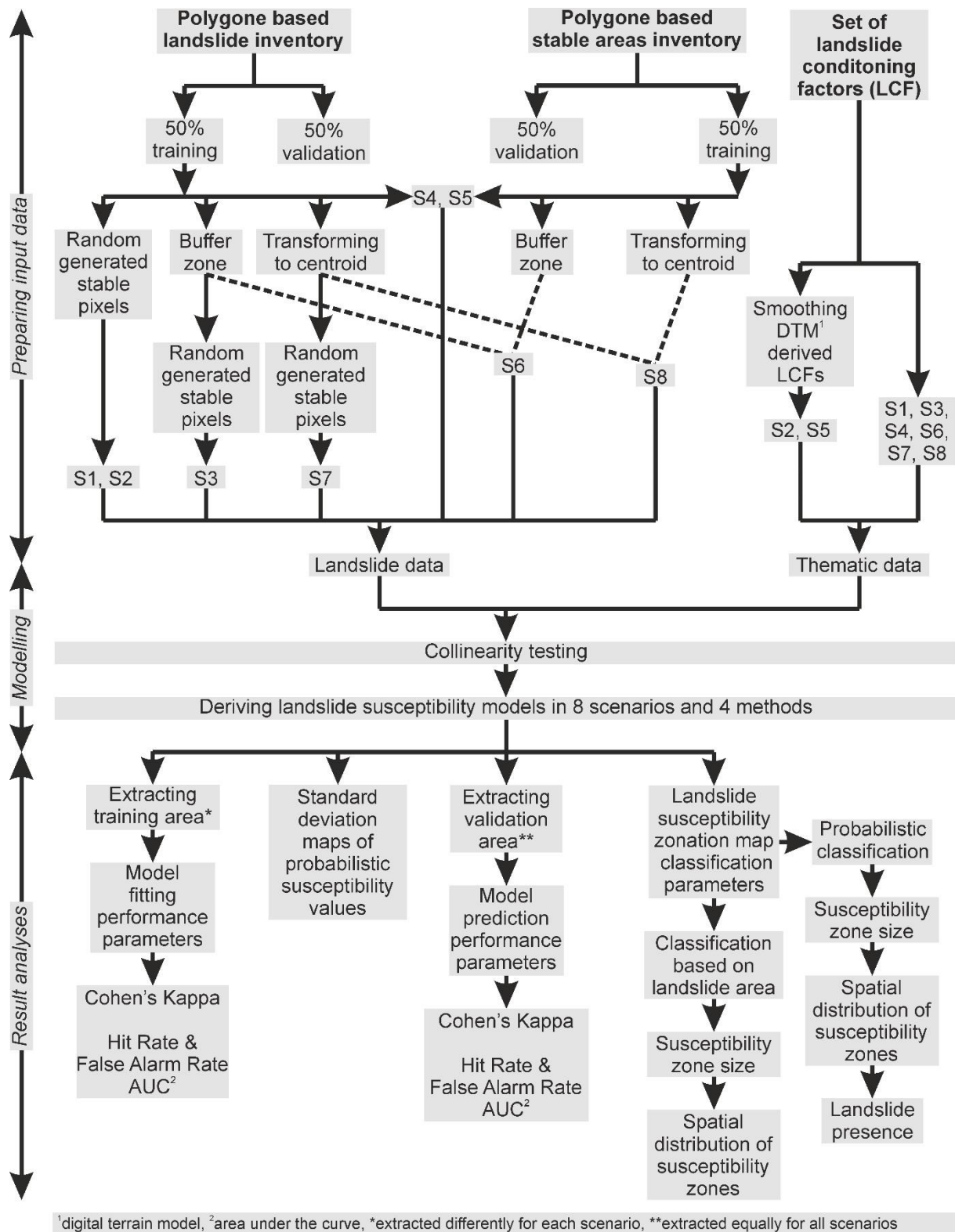


Figure 1-3 Detailed workflow for LSAs, preliminary example for Chapter 2.3

Moreover, a qualitative comparison of susceptibility maps using close-up-views and a quantitative comparison of susceptibility zonation maps classified based on landslide area presence (Mathew et al., 2007; Bernat Gazibara, 2019) or probabilistically (Bornaetxea et al., 2018) was also performed.

Specific methodologies, applied in Chapters 2.1, 2.2, 2.3 and 2.4., are adapted to the set objectives and therefore include a comprehensive comparison of different input data, defining scenarios for classification of continuous LCFs, scenarios for different data sampling and for setting up a workflow to develop a strategy that enables cost-effective development of large scale LSMs.

The input data analyses considered different data types, availability and scales, and provided a reference for their possible usage. From input data, including geomorphological, geological, hydrological and anthropogenic thematic maps, different LCF groups were defined, followed by proposing procedures for their derivation and correct use in the modelling. The procedures, performed based on quantitative and qualitative analysis of the landslide presence and spatial accuracy, gave recommendations for the preparation of input data for large scale LSAs (i.e., and essential information for a proper hazard assessment).

Scenarios for the classification of continuous LCFs were defined based on the classification criteria and insight found in the literature review (e.g. Jebur et al., 2014 and Yan et al., 2019). They vary from stretched values (Wang et al., 2021), partially stretched classes, heuristic approach (Sinčić et al., 2022a), classification based on studentized contrast (Neuhäuser et al., 2012) and landslide presence, and commonly used classification, such as Natural Neighbor (Huang et al., 2022), Quantiles (Yusof et al., 2015) and Geometrical intervals.

For sampling unstable areas, landslide centroid (Lee, 2007; Sarkar et al., 2013; Hemasinghe et al., 2018), landslide polygon (Pascale et al., 2013; Farooq & Akram, 2021; Sandić et al., 2023), landslide buffer zones to capture undisturbed morphological settings (Süzen & Doyuran, 2004) and a novel approach to smooth the DTM-based LCFs at landslide polygon location, were chosen. On the other hand, a novel inventory of stable areas was derived from which centroid, polygon and buffer stable areas are sampled. Lastly, random stable points are used as a reference point scenario, being the most used strategy for stable areas.

Geomorphologically and geologically representative small study area (about 16%) was selected from the whole Podsljeme Area (130 km²) to define a modelling set of landslide susceptibility conditions. Furthermore, the specific conditions in a smaller study area (21 km²) were also used to develop a LSM for the whole Podsljeme Area (130 km²). As a result,

validating the applied LSM to a large study area, enabled the evaluation of possibilities to reduce significantly the cost of LSAs by performing it at significantly smaller area.

LSMg was performed in ArcMap (ESRI, 2020a), MATLAB (The MathWorks Inc., 2021a,b), Microsoft Excel (Microsoft Corporation, 2018), and using the LAND-SUITE software by Rossi et al. (2022) (i.e., coded in R language, (R Core Team, 2023); run under RStudio, (Posit Software, PBC, 2021)). Additional software for specific procedures includes GlobalMapper (Blue Marble Geographics, 2020), CloudCompare (CloudCompare, 2022), ArcGIS Pro (ESRI, 2020b), QGIS (QGIS, 2020) and Surfer (Golden Software, 2020). CorelDRAW (Corel Corporation, 2020) and Grapher (Golden Software, 2018) were used for graphic editing, curve plotting and creating figures.

1.5 Plan of research

Based on the described methodology, available material and the defined study areas, a generalized plan of research is set in four main stages, as described below.

Stage 1 Landslide data

1a Acquiring landslide data for large scale LSMg: (i) deriving DTM derivatives for landslide mapping from LiDAR point cloud; (ii) expert-based visual landslide mapping on large scale; (iii) field verification of mapped landslides; (iv) finalising a representative landslide inventory map.

1b Processing and preparing landslide data for statistical analyses: (i) rasterization of a polygon based landslide inventory map; (ii) defining model evaluation and validation datasets.

Stage 2 Thematic data

2a Acquiring input data layers for LCF maps for large scale LSMs: (i) analysing data availability and determining the procedure for deriving LCFs; (ii) using LiDAR point cloud and large scale DOF imagery for deriving optimal LCFs; (iii) digitizing data; (iv) data editing and processing; (v) finalising input data layers for LCFs.

2b Processing and preparing LCFs for statistical analyses: (i) classifying continuous LCF data types using 11 classification methods, (ii) analysing the impact of derived LCFs, (iii) analysing landslide area distribution in LCFs derived from different input data layers.

Stage 3 LSMg

3a Deriving 87 large scale LSMs considering (i) six statistical methods, (ii) 11 classification scenarios for continuous LCF data types, and (iii) eight different strategies for sampling stable and unstable areas.

3b Evaluating LSM quality by determining the following quantitative and qualitative metrics: (i) AUC; (ii) Cohen's kappa; and (iii) susceptibility zonation maps based on probabilistic and landslide area presence approaches; (iv) distribution of landslides in susceptibility zones; (v) SD maps of probabilistic susceptibility values; (vi) visual grading of close-up map views.

Stage 4 Conclusions

Synthesize conclusions and recommendations for large scale LSMg of small and shallow landslides in soil and soft rock considering statistical methods, classification of continuous LCF data types and sampling strategies of stable and unstable areas.

1.6 Objectives and hypotheses

Objectives of the doctoral research are the following:

(1) Developing input data layers for deriving HR and spatially accurate relevant landslide and thematic data for susceptibility models in large scale, necessary for large scale LSMg.

(2) Analysing the distribution/repartition of landslide occurrence and spatial accuracy in different types of thematic data to determine optimal LCFs for large scale LSMg.

(3) Application of multiple classification methods of continuous LCF data types to determine classification impact on LSMg.

(4) Prepare landslide data and LCF data for application in different statistical methods, including a pairwise correlation analysis to assess correlations between landslides and LCFs.

(5) Developing a series of LSMs considering: (i) classification methods of continuous LCF data types, (ii) sampling strategy of stable and unstable areas and (iii) application of different statistical methods.

(6) Develop a large scale LSM for the entire pilot area by using a model developed for a smaller but geomorphologically and geologically representative part of the pilot area.

(7) Drawing out conclusions and recommendations for the methodology for large scale LSMg of small and shallow landslides.

The stated objectives are based on five hypotheses (Table 1-5):

(1) High density LiDAR point cloud and high resolution orthophoto imagery enable the preparation of necessary input data for developing accurate LSMs on a large scale.

(2) A representative landslide inventory map and spatially accurate LCF maps enables the derivation of LSMs of high fitting and predictive performance despite the used statistical

method, stable and unstable area sampling strategy or method of classification of continuous LCF data types.

(3) Optimisation of continuous LCF data type and unbiased sampling of stable and unstable areas enables an increase in the fitting and predictive performance of large scale LSMs.

(4) Commonly used metrics such as AUC and Cohen's kappa do not evaluate the quality of large scale LSMs sufficiently and could be additionally supplemented by the following metrics: (a) qualitative comparison of map close up views; (b) SD maps of probabilistic susceptibility values; and (c) quantitative comparison of susceptibility zonation maps classified based on landslide area presence.

(5) A landslide inventory map created for a smaller but representative part of a geomorphologically and geologically homogeneous area enables the derivation of LSMs of high predictive performance for the whole study area.

Table 1-5 Relations of hypotheses and objectives of the doctoral research

		Hypotheses, ordinal number*				
		(1)	(2)	(3)	(4)	(5)
Objectives, ordinal number*	(1)	X	X	X		X
	(2)	X				
	(3)			X		
	(4)		X	X		X
	(5)		X	X	X	
	(6)					X
	(7)	X	X	X	X	X

*hypotheses and objectives listed and numbered in the text above

1.7 Scientific contribution

Scientific contributions of the doctoral research are:

(1) Advancement of procedures to derive a complete set of spatially accurate input data for large scale LSMg (i.e. thematic data, landslide data, data about elements at risk) by using HR LiDAR point cloud and DOF imagery.

(2) Define a standard approach for classification of continuous LCF data types and sampling strategy of stable and unstable areas for different statistical susceptibility modelling methods for the purpose of increasing the model prediction performance in developing large scale maps.

(3) New metrics necessary for evaluating large scale LSM quality.

(4) Innovative procedure to reduce the cost of large scale susceptibility modelling in geomorphologically and geologically homogeneous areas.

(5) A new methodology for developing large scale LSMs for small and shallow landslides in the Pannonian Basin and geomorphologically similar environments.

1.8 Doctoral dissertation structure

The doctoral dissertation is organized in six Chapters. In Chapter 1, an overview of previous research is given, together with material and study area description, a summarized general methodology followed by research plan, established objectives, hypotheses and scientific contributions of the research. In Chapter 2, four original scientific papers are presented: Sinčić et al. (2022b), Sinčić et al. (2024a), Sinčić et al. (2025) and Bernat Gazibara et al. (2023a). Chapter 3 gives one discussion per hypothesis and states perspectives for future research. A research summary with highlighted main conclusions is given in Chapter 4. Finally, the doctoral thesis literature and bibliography of the author are presented in Chapter 5 and Chapter 6, respectively.

2. ORIGINAL SCIENTIFIC PAPERS

2.1 The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments

Authors:

Marko Sinčić, Sanja Bernat Gazibara, Martin Krkač, Hrvoje Lukačić, Snježana Mihalić Arbanas

Published in:

Land 2022, 11(8), 1360

Published on:

21 August 2022

DOI:

<https://doi.org/10.3390/land11081360>

Article

The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments

Marko Sinčić *, Sanja Bernat Gazibara, Martin Krkač, Hrvoje Lukačić and Snježana Mihalić Arbanas

Department of Geology and Geological Engineering, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Pierottijeva 6, HR-10000 Zagreb, Croatia; sanja.bernat@rgn.hr (S.B.G.); martin.krkač@rgn.hr (M.K.); hrvoje.lukacic@rgn.unizg.hr (H.L.); snjezana.mihalic@rgn.hr (S.M.A.)

* Correspondence: marko.sincic@rgn.unizg.hr; Tel.: +385-01-553-5776

Abstract: The objective of the study is to show that landslide conditioning factors derived from different source data give significantly different relative influences on the weight factors derived with statistical models for landslide susceptibility modelling and risk analysis. The analysis of the input data for large-scale landslide hazard assessment was performed on a study area (20.2 km²) in Hrvatsko Zagorje (Croatia, Europe), an area highly susceptible to sliding with limited geoinformation data, including landslide data. The main advantage of remote sensing technique (i.e., LiDAR, Light Detection and Ranging) data and orthophoto images is that they enable 3D surface models with high precision and spatial resolution that can be used for deriving all input data needed for landslide hazard assessment. The visual interpretation of LiDAR DTM (Digital Terrain Model) morphometric derivatives resulted in a detailed and complete landslide inventory map, which consists of 912 identified and mapped landslides, ranging in size from 3.3 to 13,779 m². This inventory was used for quantitative analysis of 16 input data layers from 11 different sources to analyse landslide presence in factor classes and thus comparing landslide conditioning factors from available small-scale data with high-resolution LiDAR data and orthophoto images, pointing out the negative influence of small-scale source data. Therefore, it can be concluded that small-scale landslide factor maps derived from publicly available sources should not be used for large-scale analyses because they will result in incorrect assumptions about conditioning factors compared with LiDAR DTM derivative factor maps. Furthermore, high-resolution LiDAR DTM and orthophoto images are optimal input data because they enable derivation of the most commonly used landslide conditioning factors for susceptibility modelling and detailed datasets about elements at risk (i.e., buildings and traffic infrastructure data layers).

Keywords: landslide; large-scale landslide hazard assessment; LiDAR; high-resolution orthophoto; landslide inventory; landslide conditioning factors; elements at risk

Citation: Sinčić, M.; Bernat Gazibara, S.; Krkač, M.; Lukačić, H.; Mihalić Arbanas, S. The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments. *Land* **2022**, *11*, 1360. <https://doi.org/10.3390/land11081360>

Academic Editor(s): Domenico Calcaterra, Quazi K. Hassan

Received: date: 15 June 2022

Accepted: date: 13 August 2022

Published: 21 August 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Sustainable strategic spatial planning and management requires knowing the spatial distribution of geohazards in populated areas and systematic mapping of landslide phenomena [1,2]. One of the priorities for an action plan in the Sendai Framework for Disaster Risk Reduction 2015–2030 [3] is understanding disaster risk in all its dimensions of vulnerability, capacity, exposure of persons and assets, hazard characteristics, and the environment. Based on this data, it is important to disseminate location-based disaster risk information to decision makers, the general public, and communities at risk. Therefore, one of the first prerequisites for risk reduction measures and mitigation of the landslide's consequences is creating prognostic landslide maps that should be implemented into the spatial planning system for restricting development in landslide-prone areas and defining

building conditions with respect to the landslide hazard. Reliable prognostic landslide maps on a large scale ($\leq 1:5000$) are the result of landslide susceptibility, hazard, and risk assessment, which requires good-quality input data, that is, detailed and complete landslide inventories [4,5] and appropriate resolution and spatial accuracy of geoenvironmental factors, triggering factors, and elements at risk [6,7]. Ref. [6] stated that the most crucial step in landslide susceptibility modelling is the identification and mapping of a suitable set of instability factors. Almost 20 years later, [8] conducted a critical review of statistical methods for landslide susceptibility modelling and pointed out that, nowadays, landslide investigators are more interested in experimenting with different modelling techniques than acquiring good-quality input data. Commonly used landslide conditioning factors for landslide susceptibility on a large scale are addressed in numerous recommendations, overviews, guidelines, and review papers, such as [5,8–12]. As stated in [5], the optimal selection of the scale and method of landslide susceptibility assessment is strongly dependent on the availability of spatial information, which is confirmed by several studies where the lack of data created serious limitations in research and, consequently, a decrease in the quality of landslide susceptibility models [13]. For large-scale landslide susceptibility assessments, researchers use a large spectre of different resolutions, from 1 m [14–16] to 5 m [17,18], 10 m [19,20], 12.5 m [21], 15 m [16], 20 m [22], and 25 m [23]. Geoenvironmental input data needed for landslide hazard assessment should be acquired using the same or similar scale [24], because inconsistency in geographic precision can lead to severe errors in susceptibility models [8].

Various remote sensing techniques and applications have been discussed in the last 15 years, considering their usefulness for landslide hazard mapping [25–29]. The main advantage of remote sensing techniques is that they enable 3D surface models with high precision and spatial resolution that can be used for an extensive area coverage analysis. The advantage of the LiDAR (Light Detection and Ranging) technique and data derived by ALS (Airborne Lidar Scanning) compared with other remote sensing techniques is ground detection in forested terrain and the possibility of creating high-resolution bare-earth digital elevation models (DEMs), which enables the identification and mapping of small and shallow landslides in densely vegetated areas [30–34]. Furthermore, the landslide maps obtained through the visual analysis of LiDAR-derived DTMs have better statistics for the landslide size (area) than the inventories obtained through field mapping or the interpretation of aerial photographs [31,32,35,36].

DEMs used for landslide susceptibility assessment are generated from different source data in various scales, that is, contours with (or without) points from topographic maps [17,18,20,22], aerial photographs based on traditional photogrammetry, LiDAR data [14–16], and satellite images in a resolution ranging from 25 cm [20] up to 15 or 30 m [23]. The accuracy of DEMs derived from space images mainly depends upon the image resolution, the height-to-base relation, and the image contrast. Geological and/or soil cartographic data are often available on smaller scales, such as 1:25,000 [17–19], 1:50,000 [19,20], or 1:100,000 [17]. Besides these, several other types of source data are used on scales smaller than suggested for large-scale landslide susceptibility assessments, such as 1:100,000 [23] or even 1:200,000 [22,23]. Consequently, a solution used in the literature is to apply the ‘average’ resolution, that is, 10 m resolution for landslide susceptibility assessment if the input data layers are stretched from 1:5000 to 1:50,000 [19], or from 0.25 m resolution to 1:50,000 [20], or from 1 m resolution to 1:50,000 [15], and from 1:5000 to 1:100,000 [17], respectively.

Usually, landslide conditioning factors are derived from different source data available for the subject study area, and are grouped in several categories [5,11] used by most researchers [8]. To overcome the problem of data unavailability or inadequate data resolution for the particular scale of landslide susceptibility assessment, researchers started implementing models without certain geoenvironmental factors [37].

The main objective of the paper is to show that landslide factor maps and elements at risk derived from different source data give significantly different relative influences

on the parameters for landslide susceptibility modelling and landslide risk analysis. To achieve this, a set of data layers was derived from high-resolution remote sensing data that served as an indicator of quality required for large-scale landslide hazard assessment: landslide inventory as well as geomorphological, geological, hydrological, and anthropogenic factors and elements at risk. LiDAR DTM-derived conditioning maps, as most precise, are compared with the factor maps of different resolutions and precisions by analysing landslide presence in factor classes. The analyses of available input data for the landslide hazard assessment were done for the study area in Croatia, located in the Pannonian Basin in Europe [38–40]. A study area of 20 km² in Hrvatsko Zagorje is highly susceptible to sliding and limited with geoinformation data, including landslide data. Therefore, the primary motivation for this study was to overcome the problem of having limited information for landslide hazard assessment often mentioned by researchers [13,17]. Furthermore, using high-resolution remote sensing data to the full extent and as a primary source should reduce the time and required resources for acquiring the necessary input data [29]. For this specific reason, airborne laser scanning (ALS) was undertaken during the leaf-off period in March 2020 for a study area of approximately 20 km² in the framework of the scientific research project methodology development for landslide susceptibility assessment for land use planning based on LiDAR technology (LandSlidePlan IP-2019-04-9900) funded by the Croatian Science Foundation [41].

The first step of the study was determining the quality of purchased LiDAR data for landslide inventory mapping of small and shallow landslides and the influence of resolution and interpolation methods on the accuracy of LiDAR DTM to enable further derivation of landslide conditioning factors. In addition to LiDAR data, other available DEMs and input data sets were used to derive several combinations of landslide conditioning factor maps to test different source data and various data processing settings. Additionally, it is important to emphasise that the study focuses on input data for large-scale landslide hazard assessment, meaning 1:5000 scales or larger, by analysing their applicability to the susceptibility modelling. In this study, we present a variety of landslide conditioning factors, despite what might be relevant for susceptibility modelling in the study area and their correctness. Furthermore, to address the geographic and thematic consistency of different geoenvironmental input data sets, which are usually not considered by landslide researchers [8], we present the adequate scale of input data to minimise deviations from actual environmental conditions. Moreover, using a complete LiDAR-based landslide inventory map, we compare the spatial distribution of the classes in landslide conditioning factors, pointing out the influence of small-scale input data sets often used by researchers on a large-scale landslide susceptibility assessment. The research hypothesis is that high-resolution remote sensing data (i.e., LiDAR DTM and orthophoto images) are optimal input data sets for large-scale landslide hazard assessments. The novelty of the study is that the analyses of landslide conditioning factors will prove that there is a difference between the relative influences of factor classes depending on the data source and that this difference can have a significant impact on the result in the form of weight factors for landslide susceptibility analysis. The contributions of the comprehensive analyses of 16 factors from 11 sources are to get concrete proof about what is usable for landslide hazard assessment on a large scale and opposite and which resources should not be used. Moreover, one of the innovations of the study is how to acquire good-quality input data by the improvement of the existing small-scale input data layers by LiDAR DTM-derived maps.

Additionally, presented landslide research deals with land use and land cover data. Derived input data layers are needed for large-scale landslide hazard assessments, enabling better input data for urban planning and development. Therefore, we emphasised elements at risk (buildings and roads) and anthropogenic input data layers (land-use) preparation, analysis, and usage in landslide research. Moreover, a finalised landslide in-

ventory map detected a significant amount of landslide occurrences resulting from extreme rainfall events in the last decade, that is, addressing the aim of the land–climate interaction.

2. Materials

2.1. Study Area

The study area comprises 20.2 km² of hilly terrain in the Hrvatsko Zagorje region in the northwestern part of Croatia (Figure 1). The study area belongs to the Varaždin County, that is, the Bednja Municipality (14.2 km²) and the City of Lepoglava (6.0 km²). According to the land-use planning maps from the Bednja Municipality [42] and Lepoglava City [43], the study encompasses 10.53 km² of forests (52%), 8.02 km² of agricultural areas (40%), and 1.63 km² of artificial surfaces (8%). The study area has an elevation range of 222–682 m a.s.l. with slope angles of 0°–85°. Furthermore, approximately 10% of the study area has slope angles <5°, 72% of the study area has 5°–32°, and 18% of the area has slope angles >32°, making most of the study area potentially prone to sliding. Most of the study area elevation ranges from 220 to 330 m a.s.l., with the exception of the northern part (southern slopes of Ravna Gora Hill) and, to a smaller extent, the southeast (Črešnjava Hill), where elevation values reach 682 and 430 m a.s.l., respectively. Despite this, all slope angles, including steep and very steep slopes, are equally distributed throughout the study area. The river Bednja and the Kamenica stream are the two main permanent water flows located at the western and eastern edge of the study area, respectively. Several temporary streams, mainly flowing south, and the Kamenica stream are tributaries to the Bednja river, with a general direction downstream to the east of the study area.

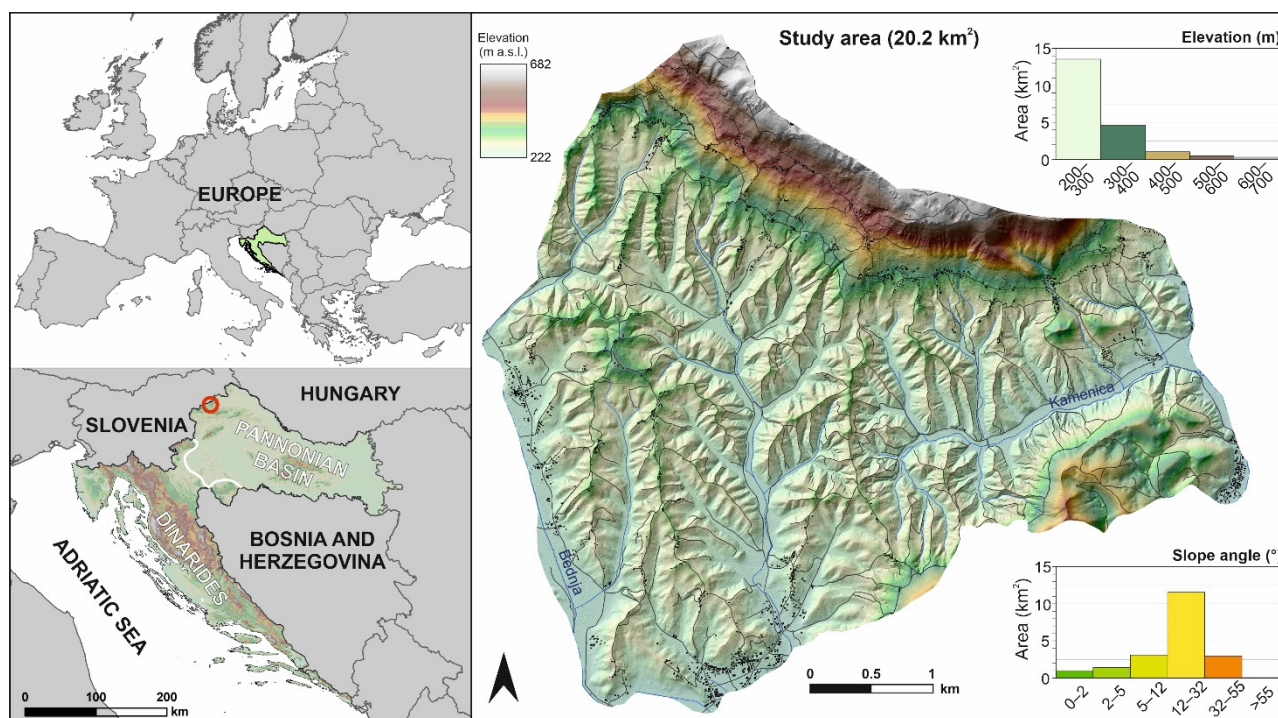


Figure 1. Study area and its location in Europe and Croatia.

The area is composed of Miocene (78%), Quaternary (14%), and Triassic sediments (7% of the study area) [44,45]. The Triassic sediments, located in the northeastern part of the study area, are composed of Lower Triassic sandstones, shales, dolomites, and limestones and Middle Triassic dolomites, limestones, and dolomitized breccias [46]. Dominantly, the Triassic sediments incline mainly towards the southwest with dip angles of

35°–60°. The Miocene sediments are sandstones; marls; sands; tuffs (Burdigalian); biogenic, sandy, and marly limestones; calcareous marls; and sandstones (Tortonian). The Miocene sediments are horizontal (in the western part of the study area) to steeply inclined (40°–55°) in the northeastern and eastern parts of the study area [46,47]. The Quaternary sediments, composed of sands, silts, and gravels, are located in the valleys.

The study area is characterized by a typical continental climate with a mild maritime influence. According to 59 years of meteorological data from the Varaždin station (approximately 30 km east of the study area), the mean annual precipitation is 861 mm. The majority of precipitations (approximately 70%) fall mainly from May to November [48]. Precipitations and human activity are the primary triggers of landslides in northwestern Croatia [49]. For example, in the winter of 2012/2013, in the same area, the prolonged heavy rainfall periods and the rapid melting of a thick snow cover caused abundant landslide events [50].

2.2. Data Availability

Landslide conditioning factors and a landslide inventory map present key data layers for landslide susceptibility assessments. According to [5], landslide conditioning factors used in landslide susceptibility analysis usually depend on resources and data availability. For that reason, several data sources were used in this study to ensure high-quality input data for future landslide hazard research and assessment. This chapter gives an overview of available data for deriving landslide conditioning factors and an inventory map as a preparatory phase for landslide hazard assessments on a large scale. The summary describes the data type, possibilities of data usage, availability, and scale (Table 1).

EU Digital Elevation Model (EU-DEM) [51] and Corine Land Cover (CLC) [52] are available data on a regional scale as a part of the Copernicus Land Monitoring Service (CLMS). The CLMS provides different types of geographical information without use restriction for a wide range of areas [53]. The EU-DEM was derived from the Shuttle Radar

Table 1. Overview of available data for the study area.

Data	Type of Data	Availability	Scale	Reference
EU-DEM	DTM	Free access	25 m resolution (small)	[51]
Elevation points, contours, and break lines	Input data for deriving DTMs	Free for scientific use	10 m resolution (medium)	[54]
Topographic Map (TK25) of Croatia	Topographic map	Free access *	1:25,000 (small)	SGA WMS server [55]
Croatian Basic Map (HOK)	Topographic map	Free access *	1:5000 (large)	SGA WMS server [56]
LiDAR point cloud	Input data for deriving HR-DTMs	Airborne laser scanning	From 0.15 to 5 m resolution (detailed)	Institutional data
Sentinel-2 orthophoto imagery	Orthophoto	Free access	10, 20, and 60 m resolution (small)	[57]
Digital Orthophoto Map (DOF)	Orthophoto	Free access *	1:5000 (large)	SGA WMS server [58]
Croatian Basic Geological Map	Geological map	Free for scientific use	1:100,000 (small)	[44,45]
CLC	Land use	Free access	Vector data 1:100,000 (small)	[51]
Open Street Map (OSM)	Traffic infrastructure, buildings	Free access	Vector data (large)	[59]
Land-use planning maps	Land use, traffic infrastructure	Free access	1:25,000 (small)	[42,43]

* View only for public use, for scientific use free upon request.

Topography Mission (SRTM), ASTER Global Digital Elevation Map (GDEM), and Russian topographic maps, providing elevation data in 25×25 m raster cells. With a vertical accuracy of ± 7 m root-mean-square deviation (RMSE), the EU-DEM is divided into 100×100 km tiles as GeoTIFF 32 bits [60].

The Geoportal of the State Geodetic Administration (SGA) represents the central access point for spatial data on a national level in Croatia. The SGA open-access data layers via Web Map Service (WMS) include the Basic Map of Croatia on a scale of 1:5000, topographic maps on scales of 1:25,000 and 1:100,000, a high-resolution digital orthophoto images of 0.5 m resolution for different periods, a digital cadastral plan, and the register of geographic names. In addition, named layers are available as WMS for either anonymous or registered users with restrictions about their usage (view only). Moreover, the SGA provides data (point elevation data, contour lines, and break lines) for DTM derivation, mainly obtained from digital photogrammetry, where image resolution must be at least 30 cm. After evaluating the source data, an SGA DTM in a 10×10 m resolution was derived using ArcGIS 10.8 Create TIN and TIN to Raster tools.

LiDAR data needed for deriving High-Resolution Digital Terrain Models (HR-DTMs) were acquired in the framework of the project 'Methodology development for landslide susceptibility assessment for land-use planning based on LiDAR technology (LandSlidePlan)' founded by the Croatian Science Foundation. Airborne laser scanning (ALS) was undertaken in March 2020, corresponding to Croatia's winter leaf-off period. The ALS was performed with a Eurocopter EC120B flying at 60 km/h at an altitude of 700 m a.s.l. The aircraft was equipped with a foto camera Hasselbad H60 and a Riegl LMS-Q780 long-range airborne laser scanner. The laser used a pulse rate frequency of 400 kHz and an average interval of 3.6 s between the shots. The number of satellites used for positioning ranged between 9 and 15, resulting in position accuracies of 0.03, 0.03, and 0.04 m for latitude, longitude, and elevation, respectively. The distance between the flight lines was approximately 400 m, followed by one cross-section flight line. The azimuth of the main direction was 115–295 and for the cross-section 44–224.

The Sentinel-2 mission supports land monitoring services, ensuring frequent and systematic coverage to support land cover mapping. Geoinformation data are available in 100×100 km tiles, in three resolutions, consisting of 13 spectral bands with a repeat cycle of 5 days [61].

The Basic Geological Map, Sheet Varaždin [45] and Sheet Rogatec [44], on a scale of 1:100,000 covers the study area and presents the only available source of geological information. Descriptions of geological settings for both sheets are given in geological notes [46,47]. Maps containing geological profiles and columns and geological notes provide spatial distribution and descriptive information about geological units, geological structures, and tectonics in the study area. However, considering the scale of the analyses, the data from the geological maps should be used with caution, especially regarding spatial precision, for example, positions of geological contact and faults are in the precision of ± 100 m.

Corine Land Cover (CLC) [52] for the 2018 period was derived by dominantly using satellite data, Sentinel-2, and Landsat-8 for gap filling. As a result, the geometric accuracy of CLC data is better than 100 m, and the thematic accuracy is greater than or equal to 85%. Furthermore, the CLC is available in vector type (shapefile) on three classification levels, ranging from a minimum of 5 robust classes in the first level to 44 detailed classes in the third level.

Open Street Map (OSM) is a community-driven initiative providing map data for various applications, including spatial files in polygon, line, or point type (shapefile) about buildings, roads, traffic infrastructure, transports, and several other spatial information. The data include classes, for example, subdivisions of each file, as presented in the [60].

Land-use planning maps for the City of Lepoglava [42] and the Bednja Municipality [43] contain information about the usage and purpose of the area inside the administrative

boundary. Land-use planning maps on a scale of 1:25,000 divide land use into several hierarchical levels. Moreover, city and municipality land-use planning maps provide up-to-date data about transportation routes, all types of roads, and railways, which could be useful for large-scale analyses.

3. Methodology

3.1. Processing LiDAR Data, DTM Derivation, and Verification

The LiDAR data were separated into sheets 500 × 500 m as LAS files for more convenient data management. Processing was performed by the vendor in the software MicroStation CONNECT (Exton, USA) in Terrasolid, followed by GPS information processing in Grafnav and IGI Aerooffice software (Kreuztal, Germany). Data processing included automatic classification and manual verification of the automated classification. The LiDAR point cloud was classified into four classes (Table 2).

Table 2. LiDAR point cloud data overview, classification, and quantitative distribution.

LiDAR Point Cloud			Default Class (Classified Points)	Ground Class (Bare-Earth Points)			Vegetation Class (Automatically Classified)	Noise Class
Amount (mil)	Point density (N/m ²)	Average point spacing (m)	Percentage (%)	Percentage (%)	Point density (N/m ²)	Average point spacing (m)	Percentage (%)	Percentage (%)
~623	30.84	0.18	24.20	52.19	16.09	0.28	23.34	0.26

LiDAR DTM for landslide inventory mapping was derived in 0.3 m resolution applying the Kriging interpolation method on point cloud of 16 points/m². Considering small and shallow landslides that dominantly occur in the study area, we performed detailed landslide mapping using the highest possible resolution of DTM. However, the derived 0.3 m resolution LiDAR DTM was not considered in further analyses because it is a too detailed topographic input data set for landslide susceptibility analysis on a large scale (i.e., 1:5000). Therefore, LiDAR DTMs needed for deriving landslide hazard assessment input data layers were derived in 1, 2, and 5 m resolutions using the IDW, Kriging, natural neighbor, ANUDEM, and local polynomial interpolation methods. DTM derivation was performed in the ArcGIS 10.8 software (Redlands, USA) using multipoint files for IDW, natural neighbor, Topo to Raster, and Kriging tools, and in the Surfer software (Golden, USA), using the source las file as input data for the local polynomial method.

For selecting the optimal DTM interpolation method, quantitative assessments were applied, combined with assessing technical limitations and difficulties encountered in used software. Fifteen different DTM combinations used for deriving landslide hazard assessment input data layers were assessed on the full extent of the study area. The quantitative assessment included a comparison of minimum, maximum, mean, and standard deviation values of HR-DTMs and available DTMs. Furthermore, by comparing source LiDAR data and interpolated DTM elevation values, mean absolute error (MAE) and root mean square error (RMSE) were calculated using Equations (1) [62] and (2) [62], respectively, as applied by [63]:

$$MAE = \frac{\sum_{i=1}^n |z_i - \hat{z}_i|}{n} \quad (1)$$

$$RMSE = \left[\frac{\sum_{i=1}^n (z_i - \hat{z}_i)^2}{n} \right]^{0.5} \quad (2)$$

where z_i is the elevation value of the source LiDAR point, $\hat{z}_i$ the interpolated elevation value of the HR-DTM or DTM raster, and n is the amount of sampled points. The amount

of sampled points (n parameter) was calculated for DTM quantitative assessment by using an expertly defined Equation (3):

$$n = \frac{A \times n_l}{A_l} \quad (3)$$

where A (m^2) is the study area surface, n_l (-) the number of mapped landslides, and A_l (m^2) the total area of mapped landslides in the finalized landslide inventory map.

Slope maps for each of the 15 derived DTMs were reclassified into 5-degree classes, and each class's percentage area was calculated. Histograms were defined to distinguish slope classes and landslide distribution within each slope class. Furthermore, the optimal LiDAR DTM was compared with EU-DEM and SGA DEM regarding the differences in slope histograms and landslide distributions.

3.2. Landslide Inventory Mapping

The topographic derivative datasets, derived from LiDAR DTM 0.3 m resolution, used to interpret the landslide morphology were: hillshade maps with different sun azimuth angles of 315° , 135° , and 45° and a constant altitude angle of 45° , slope maps, and 0.5 m contour lines. In addition, orthophoto images from 2014–2016 were used to check the morphological forms along roads and houses, such as artificial fills and cuts, similar to landslides on DTM derivatives. Landslide identification on the LiDAR DTM derivatives was based on visual interpretation of landslide features (e.g., concave main scarps, hummocky landslide bodies, and convex landslide toes). The mapping was performed at a large scale (1:100–1:500) to ensure the correct delineation of the landslide boundaries. As a result, each mapped landslide polygon was assigned with the confidence of landslide identification. The confidence in landslide identification was expressed as 'high confidence', where the landslide parts and morphology are clearly visible and easy to interpret, and 'low confidence', where the landslide parts are not clearly visible or missing and the landslide morphology is unclear [34,35]. Additionally, remediated landslides were not mapped because on the 0.3 m LiDAR DTM was impossible to identify the typical landslide morphology. Visually interpreted landslides on LiDAR DTM derivative maps were field-checked in December 2021 and January 2022 (Figure 2). However, systematic field mapping of all landslides from the inventory could not be carried out due to a relatively large area of the entire pilot area (20.2 km^2) and impassable parts of the terrain (neglected and overgrown land parcels).

3.3. Deriving Input Data Layers for Landslide Conditioning Factors and Elements at Risk

Ref. [5] states that landslide conditioning factors are not a prescribed uniform list, and their selection differs due to several factors [64]. For that reason, landslide susceptibility assessment input data layers were selected considering the scale, the purpose, suggested methods, and available data, and expertly based on the study area conditions. Moreover, according to [5], most of the conditioning factors presented in this study are highly applicable for large-scale landslide susceptibility assessments. Examining the study area as a part of landslide inventory verification and considering the application of the results, which are land-use spatial planning, special attention was given to anthropogenic conditioning factors. A total of 16 data layers for deriving landslide hazard assessment input data were selected and grouped into four categories: geomorphological, geological, hydrological, and anthropogenic (Table 3). Data preparation and digitalization for all four groups of input data layers were performed in a wider study area. As a result, a larger extent of vector data, such as roads, faults, and streams, was used to derive proximity factor maps. In addition, the spatial distribution of input data on a wider area enabled the proximity classes to be more accurate.

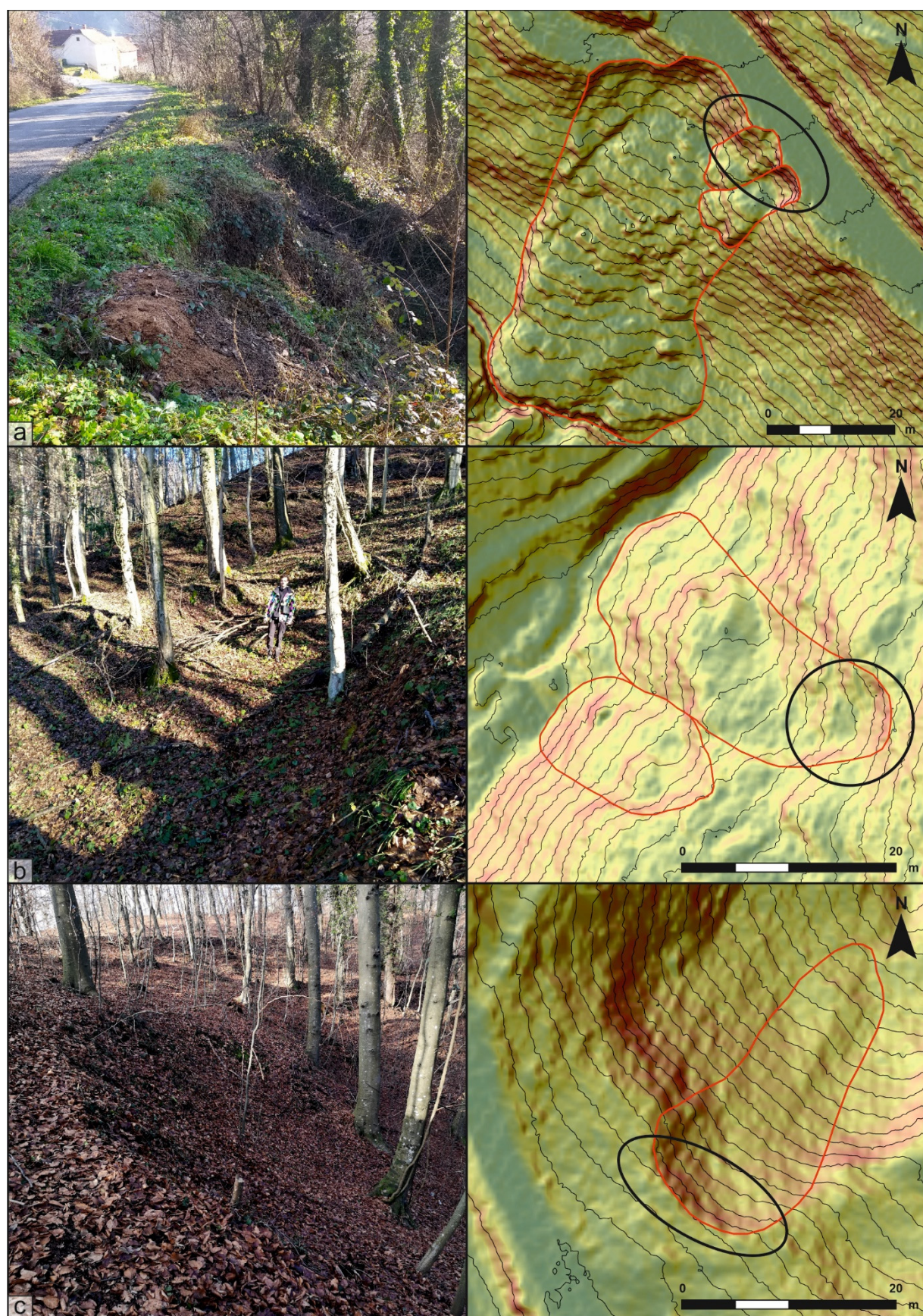


Figure 2. Examples of landslides identified on the LiDAR DTM derivatives and during field verification in the study area of Hrvatsko Zagorje, Croatia (17 December 2021): (a) close road proximity example, (b) forested area, close gully proximity example, (c) sparse forested area example.

Table 3. Source of input data layers for landslide hazard assessment in the study area.

Group	Source	Input Data Layer for Landslide Hazard Assessment
Geomorphological factors	25 m EU-DEM 10 m SGA DEM HR-LiDAR DTM	Elevation
		Slope gradient
		Slope orientation
		Terrain curvature
		Terrain roughness
		Terrain dissection
Geological factors	Basic Geological Maps (OGK) 1:100,000	Lithology (rock type)
	HR-LiDAR DTM	Geological contacts
	Basic Geological Maps (OGK) 1:100,000	Faults
Hydrological factors	25 m EU-DEM	Drainage network Topographic wetness
	10 m SGA DEM	
	HR-LiDAR DTM	
	Topographic Map of Croatia 1:25,000 (TK25)	Springs
	Croatian Basic Map 1:5000 (HOK)	Permanent and temporary streams
Anthropogenic factors	Sentinel-2 orthophoto imagery	Land-use Buildings Roads
	Digital Orthophoto Map (DOF) 1:5000	
	HR-LiDAR point cloud	
	HR-LiDAR DTM	
	Corine Land Cover	
	Open Street Map (OSM)	
	Land-use planning maps	

The geomorphological group of input data layers includes elevation, slope gradient, slope orientation, curvature, roughness, and terrain dissection. The elevation map is defined as raw DTM, used to prepare slope gradient, slope orientation, contour, and curvature map using Slope, Aspect, Contour, and Curvature tools in the 3D Analyst toolbox. Geomorphometry and Gradient Metrics Toolbox v2 [65] were used to apply the Surface Area Ratio and Dissection tool, which resulted in the roughness and terrain dissection conditioning factors, respectively.

The preliminary analysis showed deviations in geological contacts on the Basic Geological Map [44,45] in the geomorphological environment visible on the slope–hillshade map. Field verification of the Basic Geological Map confirmed that the deviations are significant and cannot be neglected. Moreover, in many landslide susceptibility assessments, stratigraphic units presented in geological maps are aggregated and used to describe rock characteristics [66], that is, forming engineering geological formations. Therefore, engineering geological formations were additionally mapped using LiDAR DTM derivatives to modify the geological contact, according to a suggestion by [67,68]. The suggested methodology demonstrated efficacy in forested areas and resulted in an engineering geological map. After the visual interpretation of LiDAR DTM derivatives, the engineering geological units and their contacts were chosen as the landslide input data. Faults were obtained by digitizing the Basic Geological Maps since their deviation is difficult to estimate by both field surveys and map comparison.

The hydrological group of input data layers consists of a drainage network, springs, permanent and temporary streams, and topographic wetness. The drainage network was derived using DTMs following the ArcGIS 10.8 tools procedure: (i) Fill, (ii) Flow Direction, (iii) Flow Accumulation, (iv) Greater Than, (v) Stream Link, (vi) Stream Order, and (vii) Stream To Feature. Different variations regarding constant values in step iv were tested, and validation was performed visually by overlaying the drainage network over a hill-

shade map. By observing the drainage network density, the average distance between different network orders, and the morphological fitting of the network on the convex and concave parts of the slope, we were able to determine the optimal solution subjectively. The topographic wetness was derived using DTMs as input for the Compound Topographic Index (CTI) tool available in the Geomorphometry and Gradient Metrics Toolbox v2 [65]. Springs and temporary and permanent streams were obtained by digitizing the available topographic map on a scale of 1:25,000. Two types of input data layers were selected: (i) permanent streams, which represent more particularly defined information, and (ii) permanent and temporary streams, which represent all streams, that is, water flows in the study area.

Anthropogenic input data layers were the most demanding to derive as they required four different input data sets and different processing methods, as shown in Figure 3. Starting from the LiDAR points, reaching robust buildings as a raster file is not an unambiguous process. Getting satisfactory results using verification to determine homogenization quality can differ depending on the LiDAR classification quality. Tools such as Focal statistics and Smooth in ArcGIS 10.8 and selecting vector files by their surface area allow a significant upgrade from the preliminary results after the reclassification stage. In the end, manually adding and removing buildings in the verification stage remained necessary to finalise input data. Considering the previous steps, this is achieved by overlaying the buildings on an up-to-date, high-resolution SGA orthophoto where buildings are well distinguished.

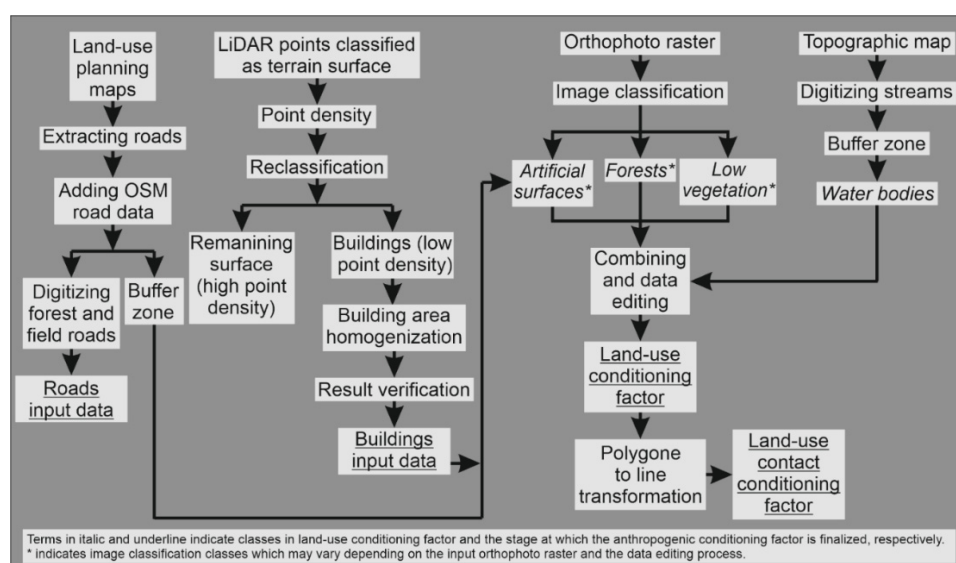


Figure 3. Flowchart of deriving anthropogenic input data layers.

Official city and municipality land-use planning maps present the most detailed source of information about roads and traffic infrastructure. However, since the availability depends on each administrative unit, the best alternative is OSM road data [60]. Therefore, after extracting the roads from the official land-use planning maps [42,43], OSM roads were added to supplement the data in a populated area. Furthermore, to complete the road network, the forest and field roads in unpopulated areas were digitized on a hillshade map and up-to-date orthophoto imagery WMS [59].

The land-use input data layer is the most complex to derive as it usually consists of artificial areas, agricultural areas, forest areas, and water bodies. Applying an Image classification tool in ArcGIS 10.8 on high-resolution orthophoto imagery enables the creation of a raster file consisting of several classes. Since the only water bodies in the study area are relatively small streams, they are indistinguishable on the SGA orthophoto and had

to be acquired using a different input map. After prospecting the study area using available data layers presenting water bodies, the topographic map on a scale of 1:25,000 [56] was used for digitizing streams that were buffered to an estimated width, presenting the water body class in the land-use data layer. Artificial surfaces consist of the building data layer and the buffer zone around road data layers, combined with the image classification results. The study area is low urbanized, so artificial surfaces are relatively small, and successful sampling for image classification was challenging. After successfully preparing individual classes of the land-use input data, the combining and data editing process consisted of homogenising the classes, that is, aggregation using the tool Focal statistics in the ArcGIS 10.8 software. Additionally, manual corrections were needed on the parts of the terrain where the previous methods showed poor results because of data preparation for landslide hazard assessments on a large scale.

The correlation between landslides and morphometric factors (elevation, curvature, roughness, dissection, and slope orientation), geological factors (lithological units based on the Basic Geological Map 1:100,000 and engineering geological units interpreted from LiDAR DTM), hydrological factors (drainage network and topographic wetness), and anthropogenic factors (land-use) derived from different input data sources is analysed and expressed through landslide percentage for each factor class. By comparing the area distribution of each factor map class and the landslide presence, that is, landslide number, it is possible to analyse the influence of the quality and resolution of input data on the accuracy and precision of landslide hazard assessment on a large scale.

4. Results

4.1. Landslide Inventory Map

In the study area in Hrvatsko Zagorje, initially, 904 landslides were identified and mapped on a scale of 1:100 to 1:500 on the LiDAR DTM derivatives. During the winter period in 2021/2022, 24% of the randomly selected landslides were checked on the field. From the 214 checked phenomena, 20 phenomena were rejected as landslides. Additionally, 28 phenomena were mapped in the field and added to the inventory, making the total number of landslides in the inventory 912 (Figure 4). Regarding the reliability of all mapped landslides, 525 landslides (58%) were identified as ‘high confidence’, and 387 landslides (42%) were identified as ‘low confidence’, that is, preassumed landslides.

During field verification, 4.9 km² or approximately 25% of the study area was checked. Of 912 interpreted landslides on LiDAR DTM, 24% or 222 randomly selected landslides were verified. A total of 163 verified landslides i.e. 73% were evaluated with ‘high confidence’ of visual identification on LiDAR DTM, and 59 or 27% landslides with ‘low certainty’ (Figure 5). Field checking resulted in 168 confirmed landslides (76%) and 35 presumed landslides (16%) based on hummocky morphology, while 19 landslides (8%) were not accessible due to neglect and overgrown land parcels. Furthermore, results from field checking were compared with the confidence of visual landslide identification on LiDAR DTM and 128 or 76% of confirmed landslides were evaluated with ‘high confidence’. In total, 35 landslides preassumed and not accessible during field checking were evaluated with ‘high confidence’, while 14 preassumed landslides and 5 inaccessible landslides were evaluated with ‘low confidence’ on LiDAR DTM. Most landslides can be dated as recently (re)activated due to the sharp appearance or a high degree of preservation of the landslide morphology. Therefore, the landslide inventory map of the study area represents a combination of seasonal inventory, landslides (re)activated in the winter of 2012/2013 and March 2018 [39], and historical landslide inventory.

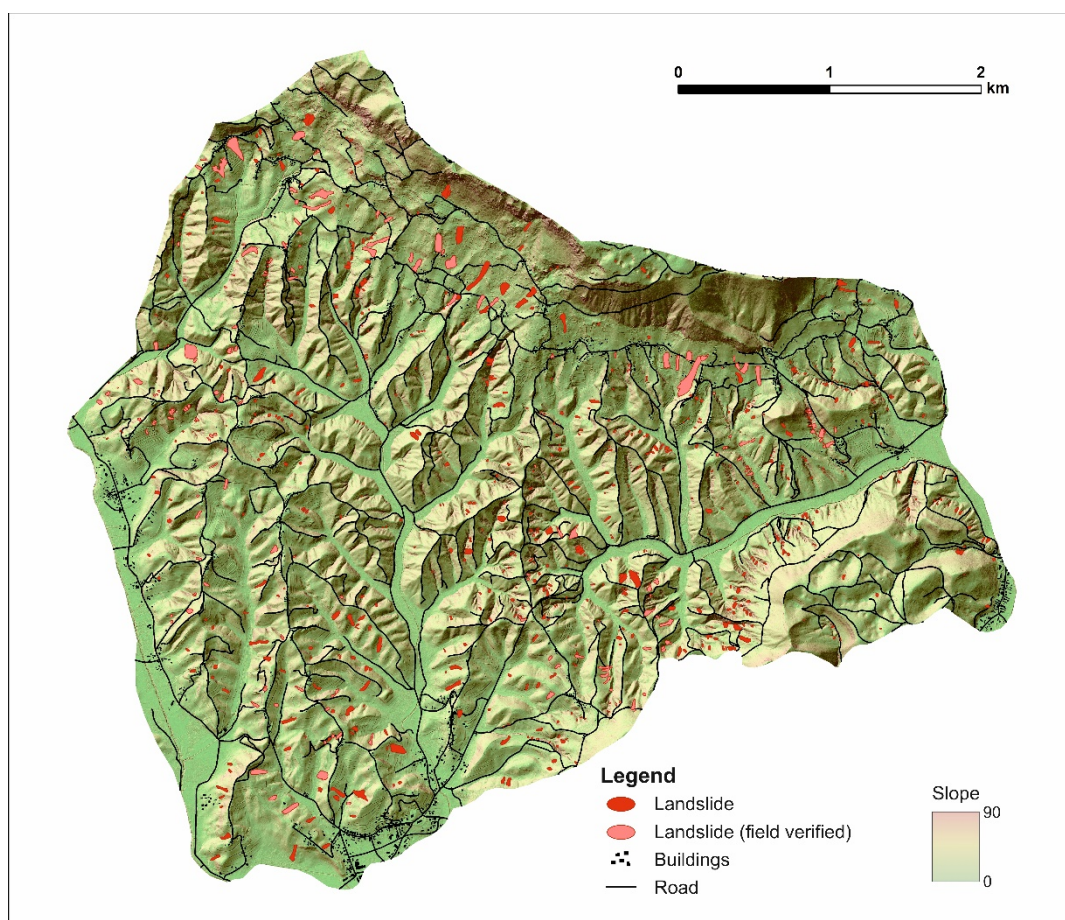


Figure 4. Spatial distribution of the landslides visually interpreted on the LiDAR DTM resolution 0.3 m.

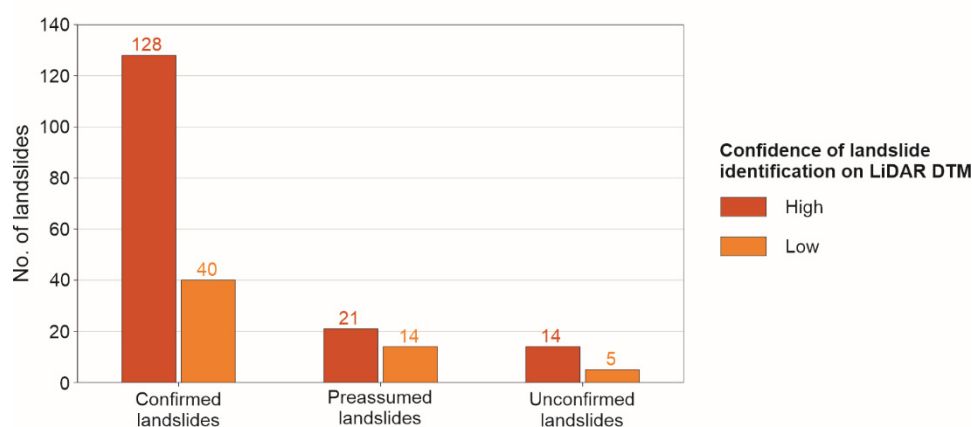


Figure 5. Classification of verified landslides concerning the confidence of visual identification on LiDAR DTM.

The total area of mapped landslides is 0.408 km² or 2.02% of the study area. The mean landslide density is 45.1 slope failures per square kilometre. The size of the recorded landslides ranges from a minimum value of 3.3 m² to a maximum of 13,779 m², whereas the average area is 448 m² (median = 173 m², std. dev. = 880 m²). The most frequent landslides in the inventory have an area of approximately 200 m², and almost 85% of the landslide bodies show a size between 40 and 2000 m². The small size of the landslides results from

geological conditions (mainly Miocene marls covered with residual soils) and geomorphological conditions, where the differences between the valley bottoms and the top of the hills are rarely higher than 100 m [69].

The frequency–size distribution of all mapped landslides in the pilot area (Figure 6) shows two scaling regimes: a positive power-law scaling for small landslides and a negative power-law scaling for medium and large landslides. The transition between the positive and the negative power-law relations can be used to distinguish between small and medium landslides. Based on the rollover at approximately 200 m², 48% of the mapped landslides are small (<200 m²), and 52% are medium and large (>200 m²) in size. The prevailing dominant type of landslides are shallow soil slides.

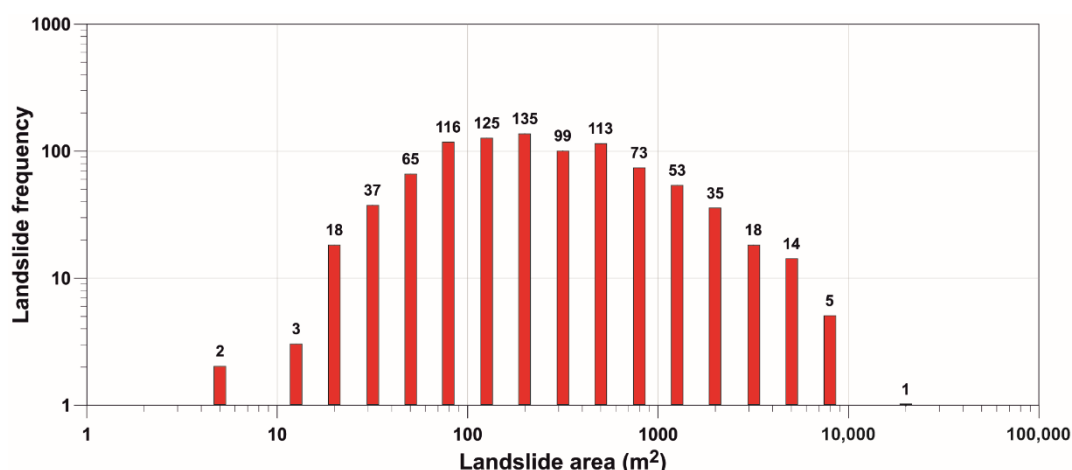


Figure 6. Frequency–size distribution of all mapped landslides in the study area (20 km²) of the Hrvatsko Zagorje region.

4.2. LiDAR DTMs

Results of a quantitative assessment for 15 LiDAR DTMs derived in three raster resolutions (1, 2, and 5 m), using five interpolation methods, and 0.3 Kriging LiDAR DTM used for landslide inventory mapping, 10 m SGA DEM and 25 m EU-DEM, are presented in Table 4. Interpolation of high point density las point cloud to lower resolution LiDAR DTMs induces smoothing of the extreme values, causing similar results and almost negligible differences in minimum, maximum, mean (μ), and standard deviation (σ) values. Apparent exceptions regarding the lowest minimum values are present for 1 m IDW and natural neighbor DTMs due to a few pixels with significantly lesser values than in other DTMs. Moreover, the standard deviation and mean values in the natural neighbor interpolation show slightly higher values ($\sigma = 78$ – 80.4 ; $\mu = 304.7$ – 305.8) compared with the other interpolation methods' stable values ($\sigma = 77.2$ – 77.4 ; $\mu = 304.4$ – 304.5). For all interpolation methods, minimum and maximum values reach their lowest, that is, highest values, in the 1 m resolution.

The n value (Equation (3)) for the MAE and RMSE analyses was 45,140. As expected, the MAE and RMSE values generally increase in lower resolutions. Local polynomial and Kriging interpolation methods show similar MAE and RMSE values, lower than the other interpolation methods. The highest difference between the local polynomial and Kriging interpolation was for the 2 m resolution, where MAE and RMSE of the local polynomial are lower and somewhat similar to ANUDEM values. Observing the sum of the MAE and RMSE values, the local polynomial interpolation method proved to be the most accurate, with the lowest values. IDW, natural neighbor, and ANUDEM showed similar results in the MAE values, somewhat higher than for a Kriging interpolation. Considering RMSE values, IDW and natural neighbor interpolation methods have the highest values compared to the local polynomial interpolation.

Table 4. Comparison of DTM statistical parameters for the study area.

Interpolation Method	Resolution	Min	Max	Mean	St.Dev.	MAE	RMSE
Inverse Distance Weighted	1	119.57	682.54	304.46	77.35	1.99	137.81
	2	222.39	681.64	304.45	77.35	3.22	176.55
	5	224.41	680.03	304.45	77.35	6.02	242.82
					SUM	11.40	555.19
Natural Neighbor	1	119.59	682.29	304.72	78.01	1.98	137.81
	2	222.40	681.70	304.98	78.64	3.21	176.55
	5	222.41	680.02	305.75	80.44	6.18	242.82
					SUM	11.38	557.19
ANUDEM	1	222.44	682.10	304.45	77.36	2.00	137.81
	2	222.60	680.94	304.41	77.29	2.44	146.40
	5	223.00	679.93	304.43	77.32	7.05	252.57
					SUM	11.50	536.78
Kriging	1	222.30	682.59	304.45	77.35	1.70	128.27
	2	222.39	682.58	304.45	77.34	3.16	175.38
	5	222.76	679.82	304.42	77.31	5.51	228.35
					SUM	10.37	532.00
Local Polynomial	1	222.32	682.73	304.45	77.34	1.70	128.27
	2	222.51	680.95	304.38	77.21	2.39	146.40
	5	222.50	679.84	304.42	77.31	5.51	228.35
					SUM	9.60	503.01
Kriging	0.3	116.77	682.77	304.45	77.35	0.27	50.11
SGA DEM	10	222.25	680.79	304.72	77.23	12.89	334.40
EU-DEM	25	226.72	679.71	311.48	78.18	38.56	291.50

The 0.3 m Kriging DTM resulted in minimal MAE and RMSE values due to its high resolution. As expected, 10 m SGA DEM has significantly higher MAE and RMSE values compared with the 5 m LiDAR DTMs. The RMSE analyses for the 25 m EU-DEM resulted in a lower value than for a 10 m SGA DTM. However, the MAE value for 25 m EU-DEM is the highest among all analysed DTMs.

The 15 LiDAR DTMs in resolutions 1, 2, and 5 m derived using five different interpolation methods, 0.3 m Kriging LiDAR DTM, 10 m SGA DEM, and 25 m EU-DEM (URL-3) were used to produce slope gradient maps. The slope maps were classified into 5° classes and overlain with the landslide inventory map of the study area (Figure 7). The five interpolation methods (Figure 7A–E) are characterised by: (i) similar trends regarding slope classes and landslide area distribution; (ii) slight differences between different resolutions; (iii) the highest area percentages in lower (<10°) and higher (>30°) slope classes for the 1 m resolutions and the highest area percentages in 10°–30° classes for 5 m resolutions; (iv) the most significant differences in slope classes between different resolutions present in 15°–20° and 20°–25° classes; (v) landslide area percentage showing minimum differences in the 0°–5° and 25°–30° slope classes; (vi) landslides being mostly present in the 15°–20° slope class; (vii) landslide presence in 5 m resolution slope map being predominant in the 10°–35° range, and least present in the <10° and >35° classes; (viii) the percentage of landslides and slope classes in the >60° zones being neglectable; (ix) landslide percentage steeply increasing until 15°–20° class, compared with a mild decrease trend following, reaching a minimum at 60°.

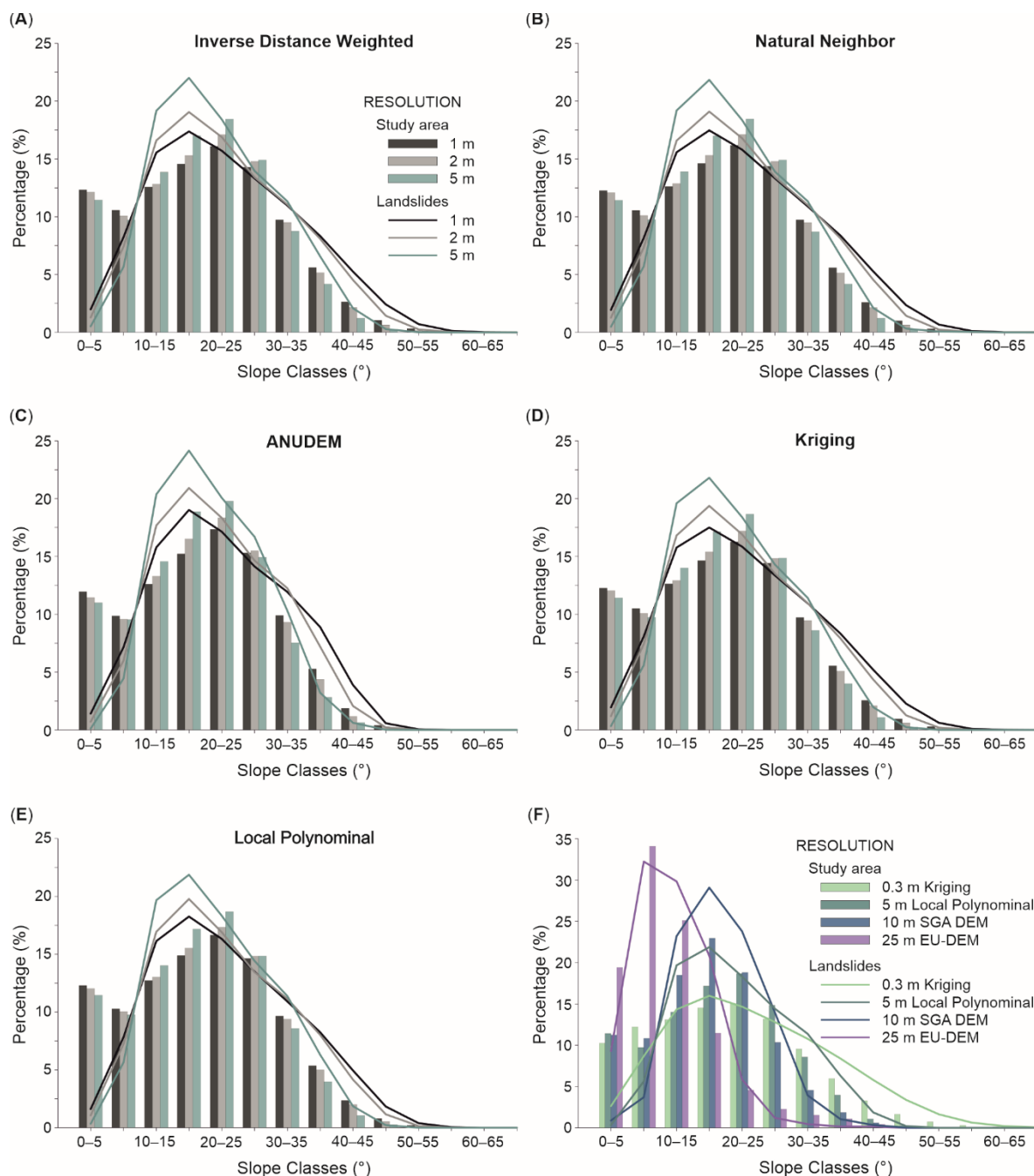


Figure 7. Comparison of slope class area distribution and landslide area distribution in the slope maps, derived from 1, 2, and 5 m LiDAR DTMs, 10 m SGA DEM, and 25 m EU-DEM.

The 10 m SGA DEM shows similar slope class areas and landslide area distributions compared with the 5 m LiDAR DTMs. On the other hand, 25 m EU-DEM depicts the most extreme values, with the most abundant slope class being 5°–10°, reaching nearly 35% in landslide area and slope area distributions. Moreover, the 0°–5° class of 25 m EU-DEM includes almost 10% of landslide areas, whereas, in all other scenarios, the 0°–5° class contains around 2.5% or less of the landslide areas. The 0.3 m Kriging LiDAR DEM has the most gentle decreasing trend in slope class distribution and the lowest maximum value of landslide presence in the 15°–20° slope class (15%).

Overall, the comparison of slope class and landslide area distributions for different resolutions and interpretation methods shows that the resolution of DEM is the most significant factor and, consequently, has an important impact on a landslide hazard assessment. Therefore, considering minimal differences in the slope class area and landslide

area distribution between interpolation methods, the relatively lowest values of MAE and RMSE parameters, the local polynomial interpolation, was chosen as the optimal method for deriving landslide conditioning factors in 1, 2, and 5 m resolutions.

4.3. Input Data Layers and Elements at Risk for Landslide Hazard Assessment

4.3.1. Geomorphological Input Data Layers

Six landslide geomorphological factors were derived using local polynomial LiDAR DTM in 1, 2, and 5 m resolutions; 10 m SGA DEM; and 25 m EU-DEM. DTMs and DEMs were used to derive slope gradient, slope orientation, terrain curvature, terrain dissection, and terrain roughness map. A comparison of minimum, maximum, mean, and standard deviation values for five continuous geomorphological factor maps is given in Table 5. Terrain elevation, slope gradient, and terrain dissection maps show minimum differences in the studied statistical parameters. In addition, terrain curvature parameters follow a gradually decreasing trend in values as raster resolution increases, whereas terrain roughness parameters gradually increase. The resulting geomorphological input data layers were classified as depicted in Figure 8A–E.

Table 5. Comparison of statistical parameters calculated from geomorphological landslide conditioning factors.

Input Data Layer	Statistical Parameter	Elevation	Slope Gradient	Terrain Curvature	Terrain Dissection	Terrain Roughness
1 m LiDAR DTM	min	222.32	0.00	−2860	0	1
	max	682.73	83.66	1699	1	9.05
	mean	304.45	19.71	-6×10^{-12}	0.50	1.09
	st.dev.	77.35	11.22	13.51	0.13	0.11
2 m LiDAR DTM	min	222.51	0.00	−1464	0	4
	max	680.95	79.42	562	1	21.79
	mean	304.38	19.50	2×10^{-16}	0.49	4.35
	st.dev.	77.21	10.90	9.67	0.14	0.39
5 m LiDAR DTM	min	222.51	0.00	−124	0	25
	max	679.84	71.46	142	1	78.63
	mean	304.42	18.95	7×10^{-17}	0.48	26.99
	st.dev.	77.31	10.21	4.97	0.18	2.04
10 m SGA DEM	min	222.25	0	−27.62	0	100
	max	680.79	58.23	30.63	1	182.93
	mean	304.72	17.01	0	0.46	106.18
	st.dev.	77.23	8.93	2.38	0.20	6.30
25 m EU-DEM	min	226	0	−2.08	0	625
	max	681	44.80	3.20	1	880.80
	mean	311.04	10.79	2.58	0.46	642.32
	st.dev.	78.25	7.24	0.37	0.22	27.90

In total, four classes were defined for the elevation map using an increment of 25 m a.s.l., and setting two edge classes at <250 and >300 m a.s.l., 1, 2, and 5 m LiDAR DTMs show (nearly) identical values, for both area distribution classes and landslide presence. All three DTMs reach maximum landslide presence in >300 m a.s.l. class, slightly more than in the 250–275 m a.s.l. class. Compared with the LiDAR DTM, the 10 m SGA DEM depicts lower differences in landslide presence and no differences in elevation classes area distribution, following the same trend. On the contrary, the 25 m EU-DEM has an increasing trend from first to last class, in both landslide presence and elevation classes distribution, reaching the maximum in the >300 m a.s.l. class.

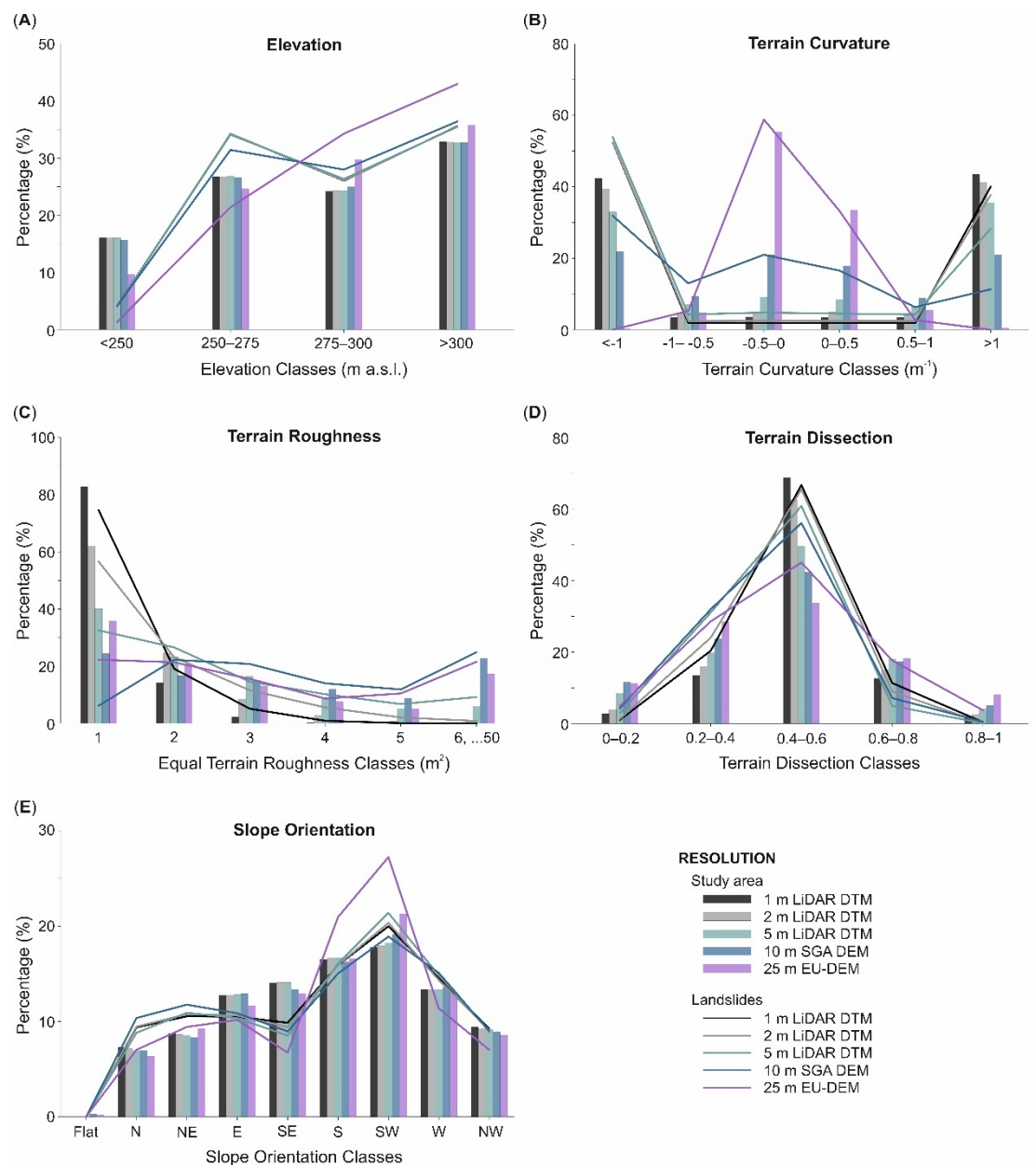


Figure 8. Comparison of geomorphological conditioning factors derived from LiDAR DTM with 10 m SGA DEM and 25 m EU-DEM: area distribution of elevation, terrain curvature, terrain roughness, terrain dissection, slope orientation classes, and landslide presence distribution.

The terrain curvature map was classified into six classes with a 0.5 increment. As a result, the class area distribution on LiDAR DTMs shows a high area percentage in the edge classes of terrain curvature and a lower area percentage in the middle classes (from -1 to 1). Additionally, the LiDAR DTMs show identical landslide presence distribution in all terrain curvature classes, reaching the maximum in the edge classes, that is, classes with values lower than -1 and higher than 1 . On the contrary, terrain curvature derived from 25 m EU-DEM shows opposite class area distribution and landslide presence than LiDAR DTM terrain curvature with maximum values of landslide presence and class area distribution in the middle -0.5 – 0 class. Therefore, the terrain curvature from 10 m SGA DEM shows a more uniform landslide presence and class area distribution than LiDAR DTMs and 25 m EU-DEM.

The terrain roughness map was normalized by equal classification in 50 classes. For practical comparison, 50 equal classes were resampled into six classes, where the 6th class represents the merged class from the 6th to the 50th classes. Significant differences are

present in the first roughness class, even between the LiDAR DTMs. A 1 m LiDAR DTM depicts maximum landslide presence and class area percentage in the 1st class, followed by 2 and 5 m LiDAR DTMs. On all three LiDAR DTMs, landslide presence and class area percentage values are gradually reduced, reaching minimum values in the 6th class. On the contrary, 10 m SGA DEM and 25 m EU-DEM show a more uniform distribution of landslide presence and class area percentage across all roughness classes. The 25 m EU-DEM has maximum values of landslide presence in the 1st and 6th classes, whereas the 10 m SGA DEM has minimum landslide presence in the 1st class and maximum in the 6th class.

The terrain dissection map was classified into five classes, with a 0.2 increment. All five studied DTMs and DEMs have similar distribution trends of landslide presence and class area percentage. For example, the 1 m LiDAR DTM shows minimum values of area percentage in all terrain dissection classes, except the middle class 0.4–0.6, where it is depicted with maximum value. Similarly, the 25 m EU-DEM shows the highest area percentage values in all terrain dissection classes except in the middle 0.4–0.6 class, where it depicts the lowest area percentage value compared with the other three LiDAR DTMs and SGA DEM. All derived terrain dissection maps show a gradual increase in landslide presence in the terrain dissection classes, starting from minimum values in the 0–0.2 class, reaching maximum in the middle 0.4–0.6 class, followed by a gradual decrease into minimum values in the 0.8–1 class.

The slope orientation map was classified into nine standard classes, that is, flat areas, north, northeast, east, southeast, south, southwest, west, and northwest. All five analysed slope orientation maps derived from DTMs and DEMs have similar distribution trends of landslide presence and class area percentage, with maximum values in the SW class and minimum values in the flat areas. An exception is the 25 m EU-DEM slope orientation map that shows lower landslide presence values in the SE and W classes and moderately higher values in the SW and S classes compared with the other LiDAR DTMs and 10 m SGA DEM.

Figure 9 presents derived geomorphological input data and close-up extents in the study area. Terrain curvature and terrain roughness maps seem less detailed than the rest of the landslide input data. However, all maps present detailed information about geomorphological settings by examining in a close-up view, that is, on a large scale. In general, maps present different relief settings, such as valleys, slopes, and hill ridges. Moreover, disturbed and nondisturbed terrain characteristics are visible, as well as sudden changes in the relief. Examining all maps, while giving extra attention to slope areas where landslides are likely to occur, it is visible that each map provides different information regarding the slope terrain, making the combination of selected geomorphological input data suitable for landslide hazard assessments.

Low slope gradient and terrain dissection values are highly represented in gullies and narrow valleys following main water flows. Moreover, low slope gradient values and continuous elevation values depict plains, larger valleys, and agricultural areas. Low terrain roughness values and close-to-zero values of terrain curvature maps are most widespread throughout the study area. However, examining the close-up extent of the terrain roughness map, the values drastically change on slopes, reaching maximum values near low slope locations. Likewise, terrain curvature values vary from low to high on slopes, depicting ridges and gullies. The morphological features of the terrain on the LiDAR DTM are sufficiently detailed to allow landslide hazard modelling on a large scale.

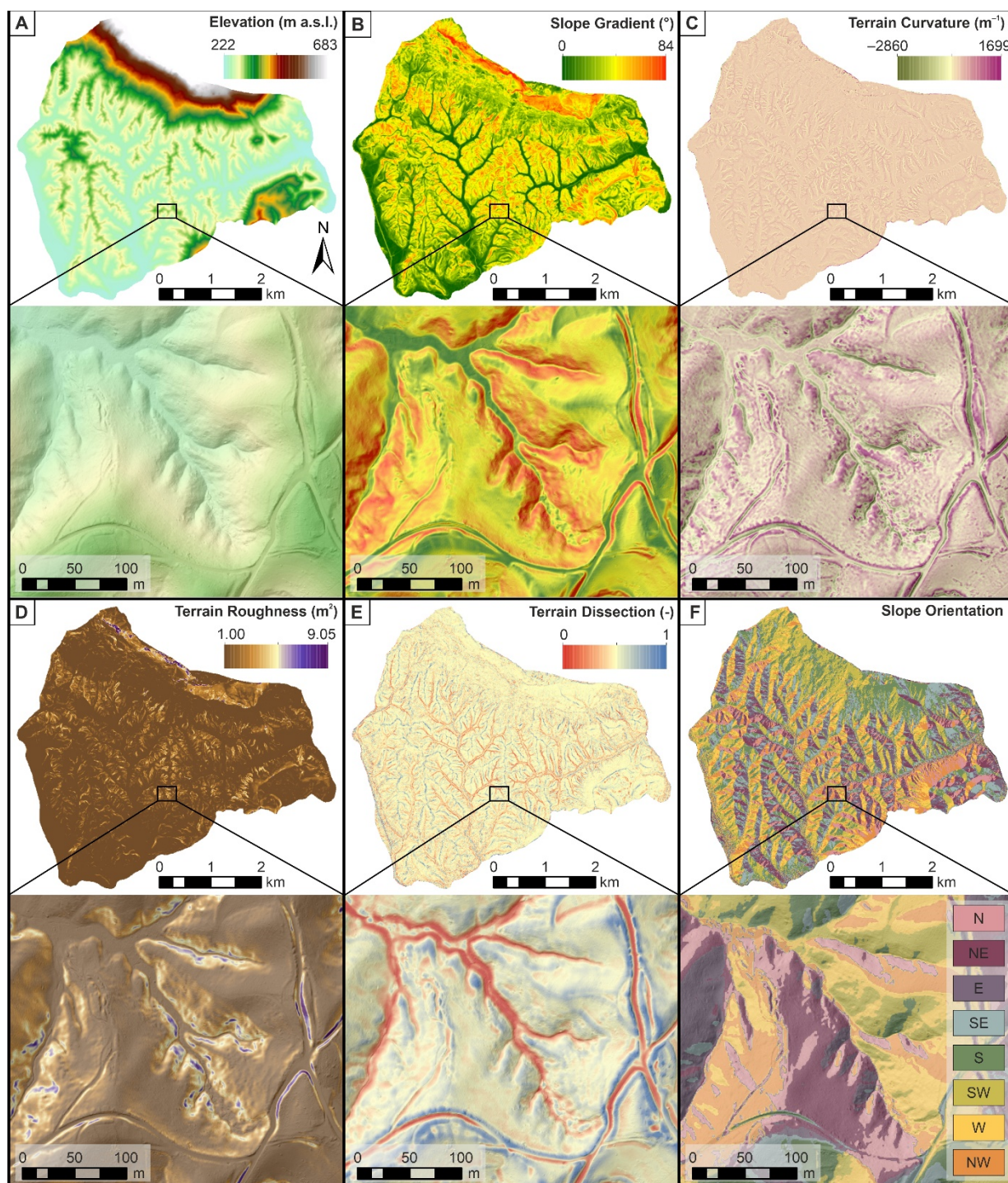


Figure 9. Geomorphological landslide conditioning factors derived from 1 m LiDAR DTM: (A) elevation, (B) slope gradient, (C) terrain curvature, (D) terrain roughness, (E) terrain dissection, (F) slope orientation.

4.3.2. Geological Input Data Layers

Lithological units and faults were digitized from the Basic Geological Maps sheets Varaždin and Rogatec [44,45]. Due to the inexact contact of the two sheets, subjective interpretation of a few faults and chronostratigraphic units was performed to unify data for the entire study area extent. In total, seven chronostratigraphic units providing soil and rock information and 48.62 km of faults were digitized (Figure 10). The small scale of the Basic Geological Maps (1:100,000) leads to an inaccurate geological contact deviating up to several hundred meters from actual environmental conditions.

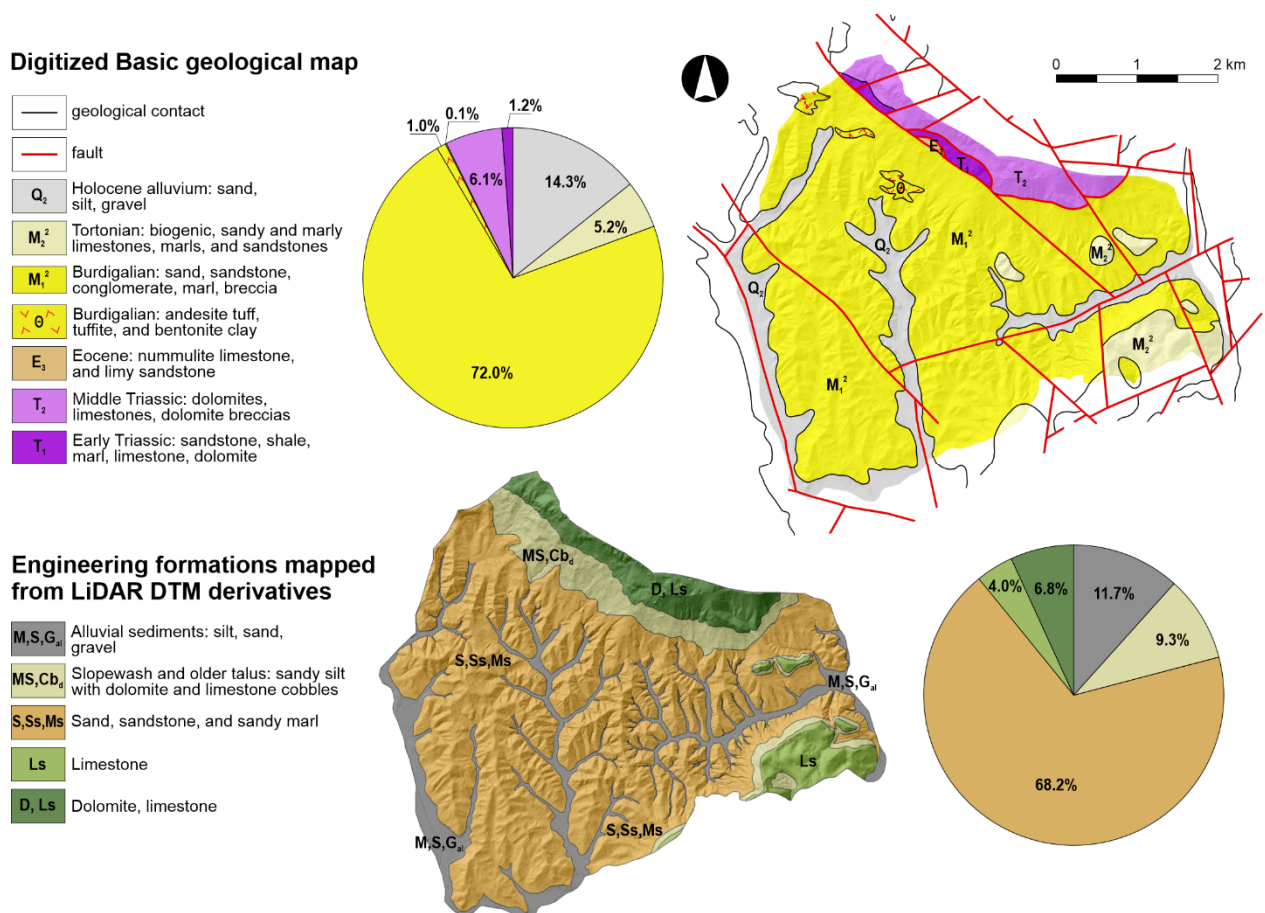


Figure 10. Geological input data: Basic Geological Map [44,45]; engineering geological unit map based on LiDAR DTM derivative interpretation.

Therefore, LiDAR DTM derivatives, such as hillshade, slope, curvature, and roughness map, were used to modify chronostratigraphic units and interpret engineering geological units. Based on visual interpretation of LiDAR DTM derivatives, five engineering geological units were identified in the pilot area, namely, (i) alluvium sediments; (ii) slopewash and older talus; (iii) sands, sandstones, and sandy marls; (iv) limestones; and (v) dolomites and limestones (Figure 10). Based on the homogeneity of identified lithologies forming the engineering geological units, they represent the engineering formations [70]. The interpreted engineering formations are briefly described in Table 6. Compared with chronostratigraphic units, the engineering geological units provide minimum deviations from the actual environmental conditions. Moreover, the spatial distribution of units' mapped LiDAR DTM derivatives is more accurate and detailed, especially in the alluvium sediments. The Holocene alluvium chronostratigraphic class is robustly represented in more extensive valleys. In contrast, the engineering formation of alluvium sediments from the engineering geological unit map is depictable in narrow valleys and plain areas, mapped based on the slope class 0° – 5° , providing the needed spatial distribution and accuracy for landslide hazard assessment on a large scale. Slopewash and older talus are an engineering geological unit that was not mapped on the Basic Geological Map and represent engineering soils on slopes transported by sheet erosion and gravitation. The presence of slopewash and older talus deposits was confirmed during the field surveys, and it was possible to determine the contact on the LiDAR DTM derivatives, such as the roughness and curvature maps. Limestone engineering geological formation was mapped by observing an area with sinkhole presence, mainly in the eastern part of the study area, in correspondence with the Tortonian chronostratigraphic class. Dolomite and limestone engineering formation depict Triassic deposits in the northern part of the study area,

mapped in more detail using LiDAR DTM derivatives (e.g., slope map), since the slopes built from Triassic dolomite and limestone are steeper than surrounding clastic engineering formations. Sand, sandstone, and sandy marl classes represent Burdigalian deposits mapped on the Basic Geological Map, following the previously mapped engineering formation units. Finalized geological input data layers consist of digitized chronostratigraphic units, geological contacts, and faults from the Basic Geological Maps; engineering geological units interpreted from LiDAR DTM derivatives; and the engineering geological unit contacts.

Table 6. Description of engineering formations interpreted from LiDAR DTM derivatives.

Engineering Formations		Description				
Name	Label	Geomorphological Setting and Origin	Engineering Material		Area km ²	%
Alluvial sediments	M, S, Gal	Soils in flood plains transported by fluvial processes	Fine and coarse-grained soils	Silt, sand, gravel	2.36	11.7
Slopewash and older talus	MS, Cb _d	Soils on slopes transported by sheet erosion and gravitation	Composite soil	Sandy silt with dolomite and limestone cobbles	1.88	9.3
Sands, sandstones, and sandy marls	S, Ss, SMs	Marine sediments and poorly cemented sedimentary rock	Coarse-grained soil, hard soil-soft rock	Sand, sandstone, sandy marl, clay	13.77	68.1
Limestones	Ls	Carbonaceous sedimentary rock; medium bedded to massive	Rock	Sandy limestone, marly limestone, calcareous marl	0.83	4.1
Dolomites and limestones	DI, Ls	Carbonaceous sedimentary rock; thinly to thickly bedded and massive	Rock	Dolomite, limestone, dolomitic breccia	1.38	6.8

The lithology input data layers, that is, the chronostratigraphic units from the digitized Basic Geological Map and the engineering geological units interpreted based on LiDAR DTM, minimally differ considering area class distribution (Figure 11). Exceptions are slopewash and older talus sediments, which are not mapped on the Basic Geological Map. Regardless of the similar area class distribution, the difference in spatial distribution, that is, geographic accuracy, is significant, as shown in Figure 10. Both lithological input data layer maps show high landslide presence in the sand, sandstone, and marl class and none or minimal landslide presence in the alluvial sediments, limestone, and dolomite and limestone classes. The Basic Geological Map has more landslides than the engineering formations map in the sand, sandstone, and marl class. Moreover, the slopewash and other talus class have significant landslide presence considering the low area percentage of the class.

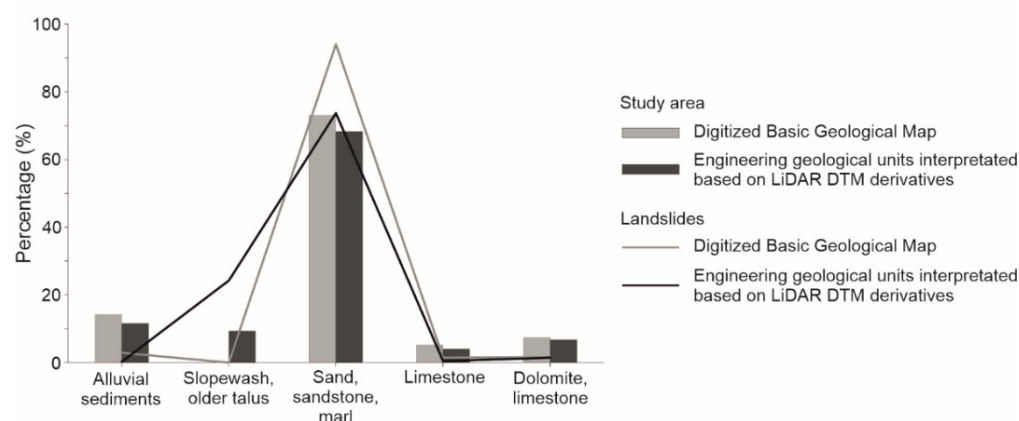


Figure 11. Comparison of engineering geological units derived from LiDAR DTM with the digitized Basic Geological Map: area distribution of lithology and landslide presence distribution.

4.3.3. Hydrological Input Data Layers

Based on the topographic map on a scale of 1:25,000, 43 springs and 55.10 km of streams (Figure 12A) were digitized. All digitized streams were classified as permanent (19.57 km) and temporary (35.53 km) streams. After digitizing, the Smooth tool in ArcGIS 10.8 was applied to stream a polyline data layer. Smoothing decreased the rough edges and sudden direction changes caused by digitizing, leading to more representative environment conditions. As a result of digitizing, it was possible to derive two types of landslide conditioning factors: proximity to all streams and proximity to permanent streams.

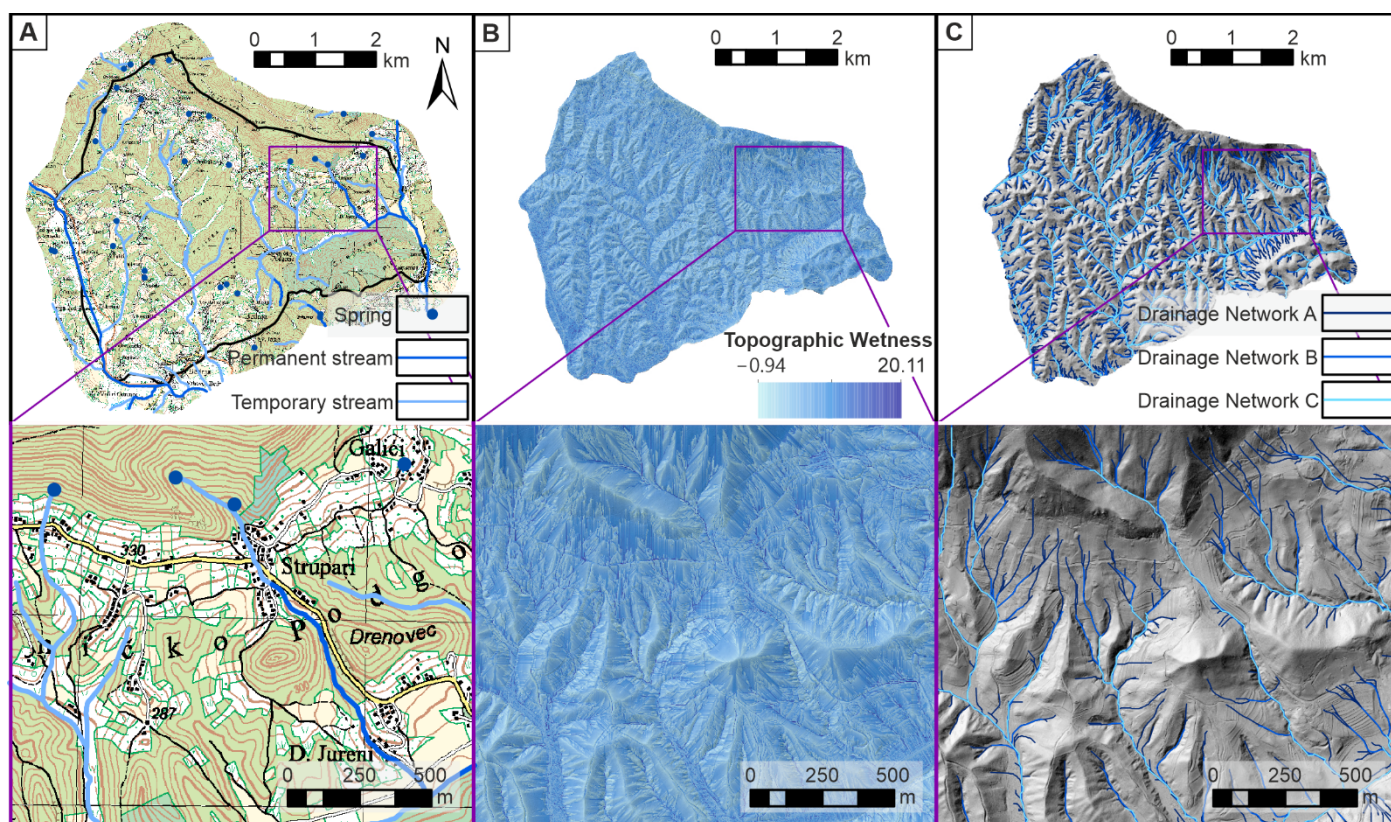


Figure 12. Stages of hydrological input data preparation for the study area and close-up extent. (A) Topographic map and digitized springs, permanent and temporal streams; (B) derived topographic wetness map; (C) hillshade map and derived drainage networks.

Figure 12B depicts topographic wetness derived from a 5 m LiDAR DTM of the study area, followed by a close-up extent. Hilltops and ridges are visible as areas with low topographic wetness values. The values increase on downhill slopes, reaching the maximum in gullies and valleys, whereas medium values are present on undisturbed slopes.

A topographic wetness input data layer derived in 1, 2, and 5 m resolutions resulted in uncategorical rasters with generally low differences in statistical information, except for minimum values (Table 7). 10 m SGA DEM and 25 m EU-DEM topographic wetness maps show moderate differences compared with the LiDAR DTMs, having relatively higher minimum and mean values and moderately lower maximum values. The most significant changes in statistical parameters are depicted when comparing the 1 m LiDAR DTM topographic wetness to the 25 m EU-DEM. An exception is the standard deviation value, where the most significant difference is between 5 m LiDAR DTM and 25 m EU-DEM.

Table 7. Comparison of topographic wetness statistical parameters derived from 1, 2, and 5 m LiDAR DTM, 10 m SGA DEM, and 25 m EU-DEM.

Input Data Layers	Topographic Wetness			
	Min	Max	Mean	St.Dev.
1 m LiDAR DTM	−0.94	20.12	4.93	1.76
2 m LiDAR DTM	0.21	19.90	5.3	1.80
5 m LiDAR DTM	1.70	19.42	5.87	1.88
10 m SGA DEM	3.16	18.45	6.48	1.74
25 m EU-DEM	4.55	17.01	7.85	1.54

The drainage networks (Figure 12C) were derived from 1, 2, and 5 m LiDAR DTMs. Visually examining the results and comparing the drainage networks, priority was given to the resolution that reflects actual environmental conditions in the best way. Finally, the 5 m resolution LiDAR DTM was used as the input for further deriving drainage network variations, depending on the methodology step iv Greater Than tool in ArcGIS 10.8 (Section 3.3), using three values for deriving different drainage network branchings. That is, the three selected options included constant values of 100, 500, and 1500, resulting in Drainage Network A, Drainage Network B, and Drainage Network C classes (Figure 12C). The necessity to iteratively test the options when deriving the drainage network is easily depictable in the close-up extent (Figure 12C), where Drainage Network A describes the gullies in distributed slopes, followed by losing much of the branching details in Drainage Network B, and finally describing only main water flows in valleys, that is, ignoring disturbed terrain in the slopes, in Drainage Network C. Drainage Network A contain 265 km of polyline data and a density of 13.1 km/km². In contrast, Drainage Network B includes 118 km of polyline data with a density of 5.8 km/km², and Drainage Network C has 72 km of polylines with a density of 3.6 km/km².

For further analysis, buffer zones of 50 m were derived around all three drainage network data layers. As a result, proximity to drainage network maps with four classes were defined and compared regarding class area distribution and landslide presence. Distribution trends (Figure 13) indicate that class area distribution follows the landslide presence in all four classes for all three proximity-to-drainage-network maps. In other words, high-class area percentage values result in high landslide presence and vice versa. Drainage Networks A and B depict steep and moderate decreases in the class area from the first 0–50 m to the last class of more than 150 m. On the contrary, Drainage Network C has a gentle decrease, that is, nearly equal class area distribution. All three proximity-to-drainage-network maps have the highest landslide presence in the class 0–50 m. Proximity-to-drainage-network map A has nearly 100% of landslides in the 0–50 m class, that is, reaching the minimum value in the 50–100 m class and having the classes 100–150 m and more

than 150 m without landslides. Proximity-to-drainage-network maps B and C reach minimum landslide presence in the class of more than 150 m.

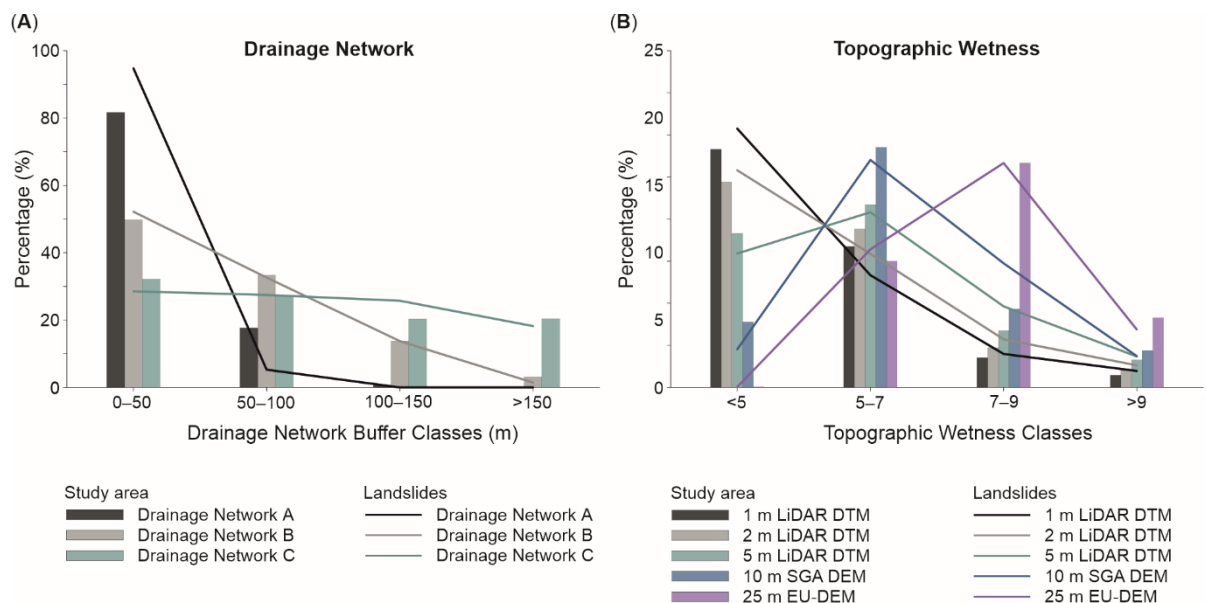


Figure 13. Comparison of hydrological conditioning factors derived from LiDAR DTM with 10 m SGA DEM and 25 m EU-DEM: area distribution of drainage network and topographic wetness classes and landslide presence distribution.

4.3.4. Anthropogenic Input Data Layers and Elements at Risk

Available information for deriving anthropogenic input data were: (i) 10 m Sentinel-2 orthophoto, (ii) 0.5 m SGA orthophoto, (iii) LiDAR point cloud and LiDAR DTM, (iv) Open Street Map (OSM), (v) land-use planning maps, and (vi) Corine Land Cover. A comparison of the 10 m Sentinel-2 orthophoto imagery and 0.5 m SGA orthophoto imagery on a close-up extent (Figure 14) demonstrated the importance of the orthophoto resolution for preparing data layers for landslide hazard assessments on a large scale. For deriving input data for landslide input data maps and elements at risk, Sentinel-2 orthophoto with low resolution makes buildings, roads, and other relevant information indistinguishable. On the contrary, the high-resolution SGA orthophoto provides reliable information about environmental conditions and the degree of urbanisation. In this study, SGA orthophoto in resolution 0.5 m enables verification of several derived input data and automated mapping of anthropogenic data, that is, building, roads and infrastructure, and land use.

Using LiDAR points classified as terrain, a point density raster (Figure 15A) was derived using the Point Density tool in ArcGIS 10.8. LiDAR point cloud with high average point distance exposed nonterrain surface. Using near-minimum point density value as a cut-off reclassification value resulted in a two-class raster. One class contained mainly buildings and a significant amount of small-area artefacts. The artefacts mainly presented areas of high-density vegetation where ALS captured a few or no terrain surface points, resulting in minimum point density values. Since most of these areas were smaller than the average building size, they were successfully removed by selecting polygons with a predefined minimum surface size. Results were verified with 0.5 m SGA orthophoto, and minimum changes were made to finalize the building data layer, such as removing excess areas and manually adding buildings. The used methodology provided automatic detection of most buildings with high precision on a large scale. As a result, 2017 individual building polygons were defined, ranging from 2 to 2620 m² and with a mean value of 63.12 m². Buildings were generally adequately located, and the process reduced manual digitizing. However, building boundaries are often irregular and show deviations from the actual building appearance on the SGA orthophoto. However, for creating input data

maps for landslide hazard assessment, that is, deriving buffer zones for proximity to buildings, these deviations are neglectable. Furthermore, smoothing building polygons would solve the problem of boundary irregularity if the resulting data layer is used to define elements at risk in the presented study area.

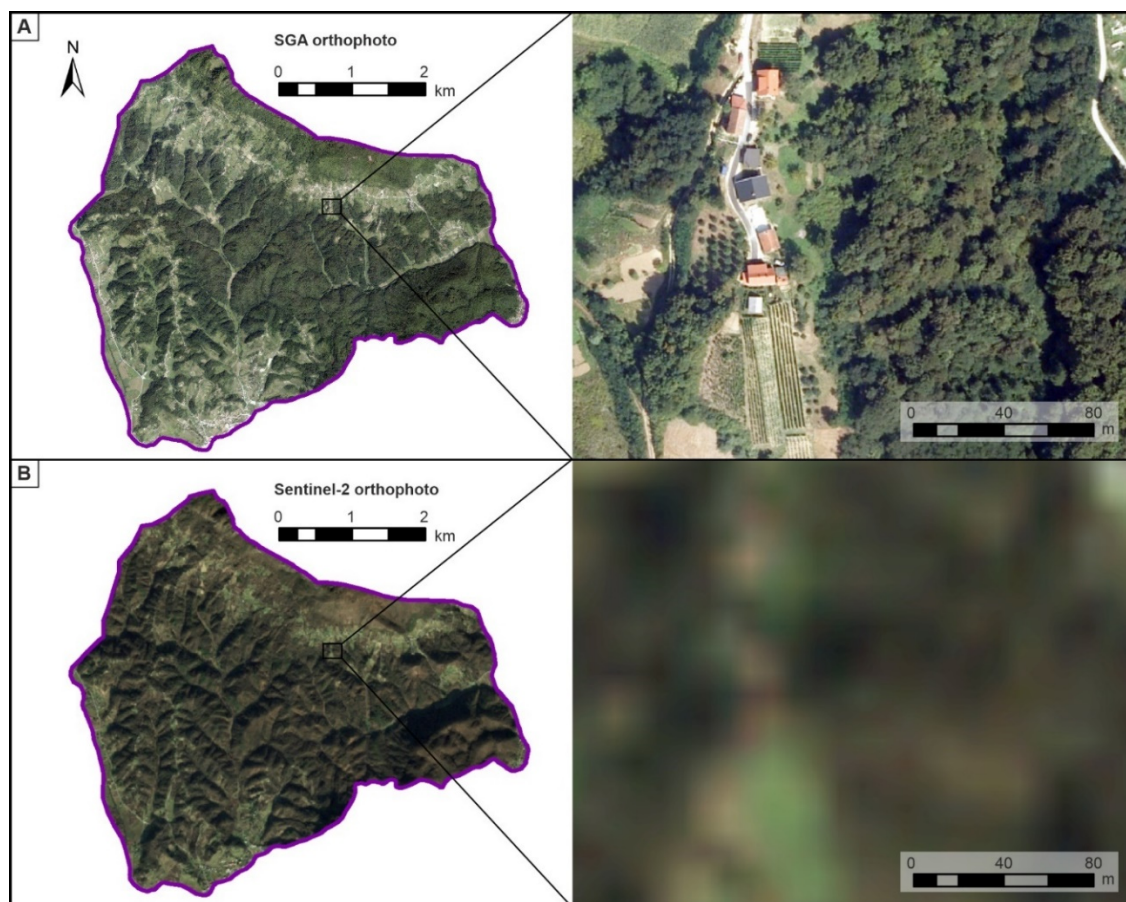


Figure 14. The study area and close-up extent. (A) SGA orthophoto (resolution 0.5 m); (B) Sentinel-2 orthophoto (resolution 10 m).

Since the land-use planning maps of the City of Lepoglava do not contain roads, OSM roads were used. In combination with the road data layer from the land-use planning map of the Bednja Municipality, a draft version of the road data layer for the study area extent was derived. Prospecting the derived road data layer on SGA orthophoto showed low spatial deviation, that is, a satisfactory level of precision for large-scale landslide hazard assessment. However, further digitizing on orthophoto maps (Figure 15B) and hillshade (Figure 15C) was needed to complete the data since the prospecting revealed unmapped roads in forested and low-populated areas. These mainly included forest pathways, unregistered roads, field roads, and similar, clearly visible on the hillshade map derived from 1 m LiDAR DTM and orthophoto imagery. Furthermore, applying the smoothing tool on the digitized road polylines increased the road curvature to actual conditions in the terrain. Therefore, the prepared road data layer for the subject study area can be used for further derivation of landslide conditioning factors and elements at risk. In total, 218.43 km of roads were identified, out of which 38.06 km are from the OSM (City of Lepoglava), 53.09 km of roads are from the land-use planning map of Bednja Municipality, and 127.28 km were digitized from high-resolution remote sensing data (Figure 15D).

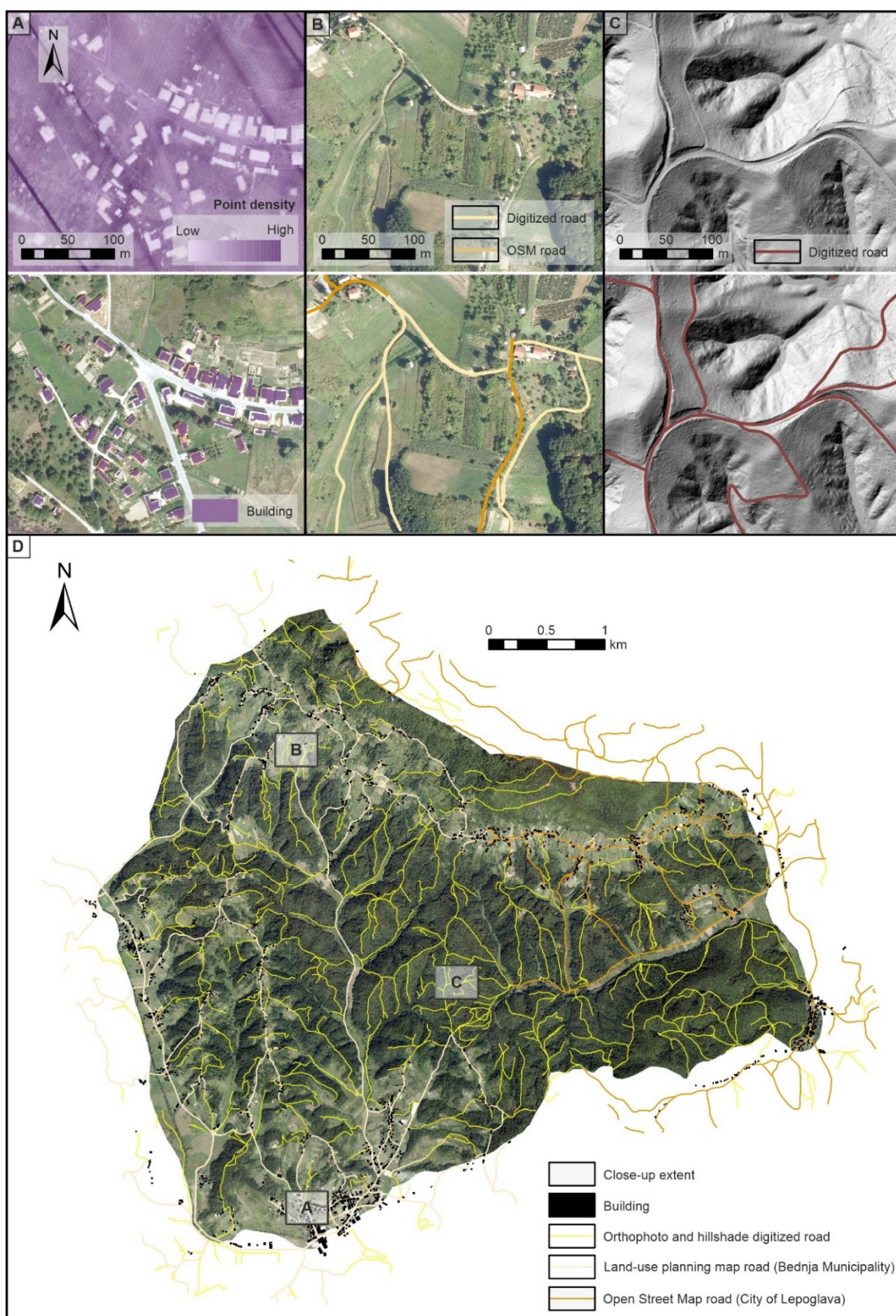


Figure 15. Stages of preparing anthropogenic input data layers. (A) The close-up extent of point density map and derived buildings, (B) the close-up extent of orthophoto raster and digitized roads, (C) the close-up extent of hillshade map and digitized roads, (D) final building and road data layer.

SGA orthophoto with a resolution of 0.5 m was prospected to determine which land use classes are possible to capture using the Image classification tool in ArcGIS 10.8 to

derive land-use map A (Figure 16A) using the suggested methodology described in Section 3.3 (Figure 3). Moreover, the prospection provided the information on which colour combinations, that is, different samples, represent land-use classes. Considering the high resolution and quality of the SGA orthophoto, three land-use categories were defined, that is, agricultural areas and low vegetation, forest areas and high vegetation, and artificial areas. After verification of the resulting land-use map, it was concluded that two classes, agricultural areas and low vegetation and forest areas and high vegetation, show relatively high accuracy and precision, while artificial areas were mapped with reduced success. To minimise the manual editing and digitizing process, the sample-taking process for each land-use class was performed iteratively, visually prospecting the results after each iteration by comparing them to the SGA orthophoto imagery. Despite the optimization iterations, the process resulted in a significant amount of small polygons improperly classified. To address the issue and aggregate the final land-use classes of agricultural areas and low vegetation, and forest areas and high vegetation, Focal statistics and Selecting by area tools in ArcGIS 10.8 were used. The artificial area class was finalized by merging the building data and a buffer zone of the road data layer. Furthermore, the water bodies class was defined using the buffer zone around digitized permanent streams from hydrological input data layers. After finalizing all four land-use classes, data merging included verification of land-use boundary zones, that is, checking overlaps and gaps.

Land-use map B was derived using the most recent land-use spatial planning maps (Figure 16B) for the City of Lepoglava and the Bednja Municipality. The spatial planning maps consist of a large number of classes, which were generalized into four land-use classes to ease usage for landslide hazard assessment. After combining original spatial planning classes, the three final classes were artificial areas, agricultural areas and pastures, and forest areas (Figure 16B). The main problem on the derived land-use map is the artificial area, which is largely oversized, often including agricultural areas in the buildings' immediate proximity. Moreover, the contact between the agricultural area and pasture class and forest area class deviates from the environmental conditions on high-resolution SGA orthophoto imagery. This resulted in agricultural area and pasture class overspread in the forest area class, which drastically influenced the spatial accuracy of the derived land-use map B in the study area.

Corine Land Cover (CLC) presents land use map C (Figure 16C). Considering large-scale landslide hazard assessment, all levels of classification showed little difference regarding land-use classes in the study area. Therefore, the first hierarchy level was selected for a representative CLC land-use map (Figure 16C). Regardless of the information details, the distribution of the land-use classes was unprecise, with extensive deviations from the actual environmental conditions. Unlike land-use map B, derived from current spatial planning maps, land-use C resulted in a relatively similar spatial distribution to our proposed methodology (land-use A). However, the deviations in land-use contact on CLC are even more severe than the spatial planning maps (land-use B map), which is expected considering the scale and the source of data used to prepare the CLC data layer. The CLC land-use map does not consist of roads or buildings in the artificial surfaces class, as it only captured discontinuous urban fabric at the very south of the study area. Moreover, the agricultural area class on CLC does not contain low vegetation pastures and valleys, which are well captured in land-use maps A and B. Land-use maps B and C provide too generalized and homogenous land-use class distribution compared with high-resolution SGA orthophoto imagery. Moreover, land-use boundary lengths between classes derived from land-use maps A, B, and C are 357.48, 233.64, and 55.73 km. Having the most extended contact length, land-use map A represents the optimal input data layer for preparing the anthropogenic landslide conditioning factors.

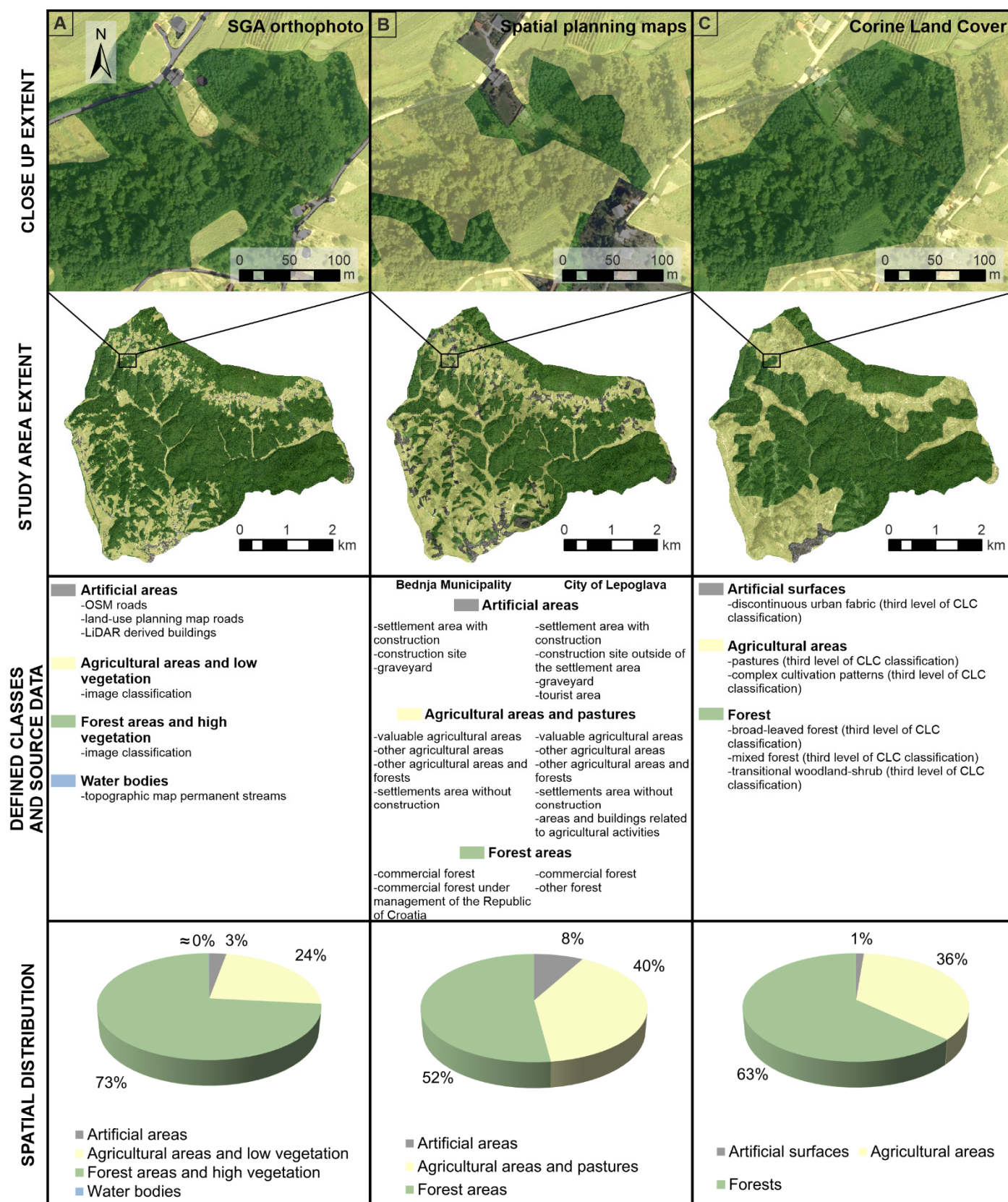


Figure 16. Land-use data layers in the close-up extent and study area, source data and defined classes for each land-use map, and spatial area distribution of land-use classes on: (A) land-use map A—SGA orthophoto; (B) land-use map B—spatial planning maps of the City of Lepoglava and the Bednja Municipality; (C) land-use map C—Corine Land Cover (CLC).

Land-use maps B and C show similar class area distribution and minimum landslide presence difference (Figure 17). On the contrary, land-use map A has moderate differences in class area distribution in the agricultural area and low vegetation class and forest area and high vegetation class compared with land-use maps B and C, and significant differences in landslide presence. Based on the conducted analysis, all three land-use data layers have the highest landslide presence in the forest area and high vegetation class. In the artificial area class, land-use maps A and C depict no landslide presence, whereas land-use map B shows a small landslide presence.

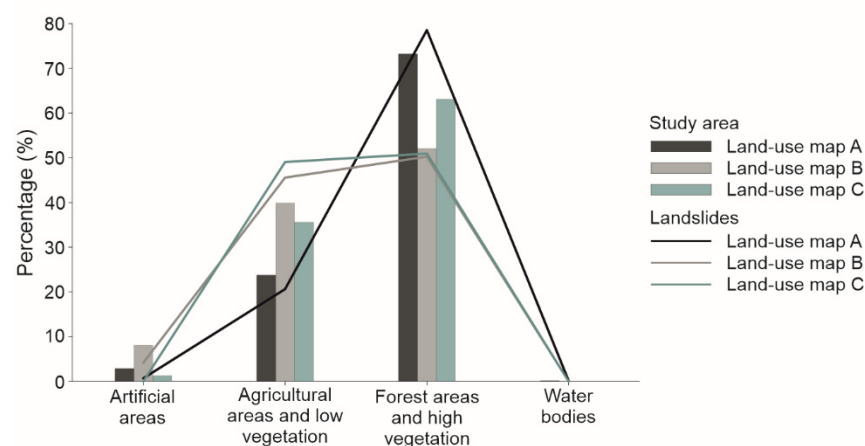


Figure 17. Comparison of the land-use area distribution and landslide presence distribution on land-use map A—SGA orthophoto; land-use map B—spatial planning maps of the City of Lepoglava and the Bednja Municipality; and land-use map C—Corine Land Cover (CLC).

5. Discussion

5.1. Landslide Inventory Map Interpreted Based on LiDAR DTM Derivatives

Landslide inventory mapping for this study was performed on the LiDAR DTM derivatives for the 20.2 km² area in the Hrvatsko Zagorje region. Geomorphological mapping of the landslide inventory was inadequate because most landslides are located in forested and seminatural areas. The ability to follow a landslide boundary accurately in the forest is limited by the reduced visibility of the slope, a failure due to dense vegetation cover, the local perspective, the size of the landslide, and the fact that the landslide boundary is often indistinct [71]. Automated landslide mapping using vegetation cover algorithms or interpretation of landslides using SAR (Synthetic Aperture Radar) images is not possible due to dense vegetation cover and specific landslide morphology, that is, small and shallow landslides characterized by small displacements. LiDAR DTM enables a distant view of the slope and the landslide phenomena, which is preferable during landslide inventory mapping to reach better results, and a more accurate and complete landslide inventory map [25]. However, the accuracy of landslide inventories made by visual interpretation of LiDAR DTM largely depends on the researcher's experience in landslide identification and mapping [72]. That is, certain practice and expertise in landslide identification on morphometric maps are needed to interpret the characteristic landslide morphology on maps derived from LiDAR DTM.

During the landslide inventory mapping, the landslide identification procedure was carried out several times for the entire study area. Each time, 'new' landslides were identified, which were not observed in the previous process of interpretation of a morphometric map derived from LiDAR DTM. Results from field checking were compared with the confidence of visual landslide identification on LiDAR DTM, and 76% of confirmed landslides were evaluated with 'high confidence'. Based on the landslide identification confidence, percentage of confirmed landslides during field checking, and the landslide frequency–normal size distribution, it can be concluded that LiDAR DTM is a very useful

tool for landslide mapping in the Neogen sediments of NW Croatia [35], despite the landslide characteristics (small and shallow landslide bodies) and the dense vegetation cover (forested areas). Furthermore, the landslide inventory derived by visual interpretation of LiDAR DTM contains a representative number of all landslide sizes in the pilot area, which is necessary for a reliable landslide hazard assessment on a large scale.

5.2. LiDAR DTM for Landslide Susceptibility Assessment

Statistical parameters combined with the nearly identical slope histograms and landslide distribution indicate minimal differences between analysed LiDAR DTMs. However, in all interpolation methods, the 1 m resolution shows the highest value for the maximum and the lowest for the minimum parameter. The 0.3 m resolution Kriging DTM histogram confirms that higher resolutions enable reaching more extreme values and capture relief details in the study area more evenly. However, the quantity of landslide presence is different as the extreme values are lost, that is, losing the relief details due to lower resolution. Finally, the 25 m EU-DEM significantly deviates from the actual environment conditions compared with the LiDAR DTMs and 10 m SGA DEM.

The RMSE and MAE parameters were applied to determine minimum deviations of the interpolation methods. The 0.3 m Kriging LiDAR DTM resulted in the lowest RMSE and MAE values, utilizing the high-density points. However, LiDAR DTMs with a very high resolution will result in too fragmented landslide susceptibility zones in large-scale modelling. For that reason, 1 to 5 m resolution LiDAR DTMs should be used in a large-scale landslide hazard assessment depending on the average landslide size in the research area. Furthermore, considering the tested variations of the LiDAR data, we concluded that if the LiDAR data density matches the landslide hazard assessment scale, the selection of interpolation method for deriving LiDAR DTM will not have an influence. However, this would likely not be the case with limited and insufficient point density LiDAR data. Therefore, determining the minimum point density of LiDAR data for specific scales of landslide hazard assessment and deriving input data layers remain an open question.

5.3. Input Data Layers for Landslide Conditioning Factors and Elements at Risk

The geomorphological (elevation, slope orientation, terrain curvature, terrain dissection, terrain roughness) and hydrological (topographic wetness) input data layers derived from LiDAR DTM have the trend of minimum differences in mean and standard deviation values between resolutions. Based on the comparison of class area distribution for all geomorphological and hydrological input data layers, it can be concluded that there are significant differences between LiDAR DTMs and available 10 m and 25 DEMs because actual environment conditions are represented with significant limitations. Different resolutions and sources, that is, spatial accuracy of the input data, will affect the landslide susceptibility modelling because most statistical methods are based on landslide density in the conditioning factor class. In other words, different conditioning factor class areas and landslide densities in one of the landslide conditioning factor maps directly affect the estimated landslide susceptibility, which is especially pronounced in large-scale assessments. The influence of the input data spatial accuracy on the final susceptibility map should be the subject of quantitative analysis and further research.

Subjective interpretation of the only available geological data for the study area, the Basic Geological Map, leads to significant data loss and results in deviations from actual environmental conditions that should be avoided when conducting large-scale landslide susceptibility assessments. Using geological data from small-scale geological maps and information from the geological data sheets enables the researchers to get acquainted with the terrain and the geological settings. LiDAR data, that is, high-resolution DTM derivatives, enable distinguishing the geological settings in more detail, improving geological contacts, and aggregating chronostratigraphic units into engineering formation units, which are more suitable for landslide susceptibility assessments. The limitations of this approach could be the presence of different geological settings across the study area, in

which engineering units cannot be distinguished from the LiDAR DTM derivatives. Moreover, field mapping as an alternative to small-scale robust geological data would be too time-consuming, and using only LiDAR DTM derivative maps alone cannot provide sufficient data. Without LiDAR DTM derivatives, the geological data on a scale of 1:100,000 could be used only for small-scale landslide susceptibility assessments. However, the process described in our study showed that combining small-scale geological maps and LiDAR DTM derivatives resulted in significantly reduced deviations from the actual environmental conditions, except for the geological structures, that is, fault data set. In our study area, in Hrvatsko Zagorje, it was impossible to map structural elements such as faults on LiDAR DTM derivatives due to the geological setting in which they are not clearly visible, that is, 93% of the study area is covered with Miocene and Quaternary sediments.

Streams displayed on the topographic maps on a scale of 1:25,000 are visible in the LiDAR DTM derivatives, such as hillshade maps. Moreover, streams on the LiDAR DTM derivatives sometimes can be similar to forest paths, and field checking is necessary. Additionally, the topographic map legend enables more information about streams, enabling their classification into permanent and temporary, which is impossible from the digitisation of LiDAR DTM derivatives. On the other hand, digitized streams from the topographic maps present higher-order flow in the drainage network, providing roughly five times fewer streams in length than the drainage network. For that reason, determining the robustness of input data is crucial to making it fit for the selected scale of the analysis. For example, comparing the three different drainage networks, we concluded that the scenario that manages to cover gullies in steep slopes without having the drainage network edges too close to each other should be considered optimal. However, testing the values of the buffer zones around the drainage network was not in the scope of this study, except for comparative purposes, and they should be further analysed because they could significantly impact the landslide susceptibility model. Therefore, combining small-scale topographic data and LiDAR DTM derivatives can provide optimal hydrological input data layers for landslide susceptibility assessment on a large scale. An exception is spring data that are impossible to map by the interpretation of LiDAR DTM derivatives and should be digitized from topographic maps.

Almost all roads (except forest roads and paths) in the study area are well distinguishable on the 50 cm SGA orthophoto imagery. However, available city and municipality land-use planning maps and OSM road data layers were used to reduce the digitising process. Furthermore, OSM data are satisfactorily precise but generally incomplete, missing buildings and roads in rural study areas. In that case, LiDAR point cloud and a hillshade map can be used to complement building and road data layers. The object-based classification of land-use classes depends on the quality of capturing the most common land-use classes, such as artificial surfaces and agricultural and forest areas. Moreover, the accuracy of the derived land-use data highly depends on the researchers' ability to select necessary samples of each class on the orthophoto. We suggest prospecting the study area to determine the distribution of each class and in what manner each class is visible on the high-resolution orthophoto. The results are often heterogeneous and misleading, especially for agricultural and forest classes. We found that homogenizing the preliminary results to remove misleading polygons, to the extent that the data are still as heterogeneous as possible, provides the most detailed information for large-scale landslide hazard assessments. That is, artificial areas, agricultural areas and low vegetation, forest areas and high vegetation, and water bodies are the four defined classes, giving descriptions accordingly to the areas they cover, compared with the actual environment conditions presented in the orthophoto imagery. Streams and similar water bodies on LiDAR DTM derivatives often look similar to other natural shapes, such as gullies or forest paths, so we suggest prospecting the source files to mitigate the misinterpretation. In any case, spatial distribution can be misleading, as two different land-use maps could have a

similar class area distribution, whereas their spatial distribution in the study area can deviate significantly. Applying the proposed method to derive building and road data layers and land-use maps from LiDAR data and high-resolution orthophoto images reduces the number of manual corrections needed in larger study areas. Moreover, the proposed methodology results in optimal input data considering the amount of detail and minimum deviations to the actual environmental conditions.

6. Conclusions

The preparation of input data and cartographic thematic layers is the most important and demanding step in the framework of landslide hazard assessment because all applied methodologies and derived results strongly depend on the quality of input data. Therefore, the input data for landslide hazard assessment on a large scale were prepared and analysed for the study area (20.2 km²) in Hrvatsko Zagorje (Croatia), where only limited medium-scale topographic and small-scale geological data are available.

The result of the visual interpretation of LiDAR DTM morphometric derivatives is the landslide inventory map, which consists of 912 identified and mapped landslides, ranging in size from 3.3 to 13,779 m² and with an average landslide density 45.1 landslides/km². Based on the normal distribution of mapped landslides and the assessment that 65% of landslides are reliably identified, it can be concluded that the inventory of visually mapped landslides for the study area of the 20.2 km² is detailed and complete. The prevailing landslide types are shallow and small to medium-sized soil slides that cause minor property and infrastructure damages. However, due to high landslide density, small and shallow landslides can cause significant damage on local roads in Hrvatsko Zagorje, and should not be neglected during landslide susceptibility analysis. Considering the preservation of landslide morphology throughout the study area, it can be concluded that most of the landslides in the inventory are recent and (re)activated by extreme hydrometeorological events in the winter of 2012/2013 and February 2018. The result is a combination of seasonal and historical landslide inventory, which are high-quality and representative input data for further landslide hazard assessment on a large scale. The quality of the inventory was also proved by field-checking, which confirmed the methodology of landslide identification and mapping as well as resulting high-quality inventory and its representativeness as input data for further testing of landslide conditioning factors and verification of the relative importance of factor classes.

Prior to deriving all the landslide conditioning factors, the analysis of how the interpolation method and resolution affect the quality of the LiDAR DTMs was performed. The LiDAR DTMs' statistical parameters, the area distribution of slope map classes derived from LiDAR DTMs, and the landslide presence in slope classes showed that high-density LiDAR point cloud enables the derivation of high-quality DTMs at an appropriate resolution, regardless of the interpolation method. Furthermore, 15 LiDAR DTMs were compared with available DEMs, that is, 10 m SGA DEM and 25 m EU-DEM. As a result, 10 m and 25 m available DEMs significantly differ from LiDAR DTMs, considering landslide presence in slope class area distribution, indicating that they should not be used as input data for landslide hazard assessment on a large scale.

The high-resolution remote sensing data, LiDAR DTM, and orthophoto images are optimal input data sets for large-scale landslide hazard assessments because they enable the following: (i) derivation of geomorphological, geological, hydrological, and anthropogenic input data layers on a large scale, except geological structures and spring data; (ii) verification of input data layers; (iii) derivation of input data layers with the best fit to the actual environmental conditions (spatial accuracy testing); and (iv) efficient derivation of input data layers considering both price and time. Having representative spatial data depicting actual environmental conditions in the needed scale enables unambiguous selection of the optimal input data layers for landslide mapping, landslide susceptibility, and hazard modelling, as well as for the derivation of accurate data on elements at risk for landslide risk assessment. The resolution and spatial accuracy of input data layers for

landslide hazard assessment, class area distribution of landslide conditioning factors, and landslide number in a single landslide factor class directly affect the susceptibility assessment in statistics-based models. Furthermore, for iterative processes, such as drainage network derivation or object-based image classification of land-use data, high-resolution input data used for verification enable optimization of the settings needed for the scale of the analysis. Moreover, when preparing geological data or mapping engineering geological units, the LiDAR DTM derivatives allow significant corrections of imperfections in the spatial accuracy of the robust small-scale geological information.

The presented research study in a relatively small study area (20.2 km²) in Hrvatsko Zagorje showed that the high-resolution remote sensing data provide a quick and affordable tool that enables derivation of the most commonly used landslide conditioning factors and elements at risk for landslide risk analysis on a large scale, regardless of geomorphological and geological settings and the degree of urbanization. Furthermore, high-resolution LiDAR data are the only possible solution for landslide inventory mapping in highly vegetated and unreachable areas, which can be applied in NW Croatia (approximately 7000 km²) based on similar geomorphological and geological characteristics and the landslide type. The LiDAR-based landslide inventory and all other landslide maps produced from landslide conditioning factors and elements at risk, derived from high-resolution remote sensing data, enable a high confidence level during the spatial planning system implementation.

Author Contributions: Conceptualization, M.S. and S.B.G.; methodology, M.S. and S.B.G.; software, M.S. and H.L.; validation, M.S., S.B.G., and M.K.; formal analysis, M.S. and H.L.; investigation, M.S.; resources, M.S., S.B.G., M.K., and S.M.A.; data curation, M.S., S.B.G., M.K., H.L., and S.M.A.; writing—original draft preparation, M.S. and S.B.G.; writing—review and editing, S.B.G., M.K., and S.M.A.; visualization, M.S., M.K., and H.L.; supervision, S.B.G., M.K., and S.M.A.; project administration, S.M.A. and S.B.G.; funding acquisition, S.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been fully supported by the Croatian Science Foundation under the project methodology development for landslide susceptibility assessment for land use planning based on LiDAR technology, LandSlidePlan (HRZZ IP-2019-04-9900, HRZZ DOK-2020-01-2432).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The generated and analyzed datasets shown in the study are available from the corresponding author on reasonable request.

Acknowledgments: We would like to thank anonymous reviewers for their suggestions and comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Bell, R.; Glade, T.; Granica, K.; Heiss, G.; Leopold, P.; Petschko, H.; Pomaroli, G.; Proske, H.; Schweigl, J. Landslide Susceptibility Maps for Spatial Planning in Lower Austria. In *Landslide Science and Practice*; Margottini, C., Canuti, P., Sassa, K., Eds.; Springer: Berlin, Heidelberg, Germany, 2013; pp. 467–472, ISBN 978-3-642-31324-0.
2. Clague, J.J.; Roberts, N.J. Landslide Hazard and Risk. In *Landslides*; Clague, J.J., Stead, D., Eds.; Cambridge University Press: Cambridge, UK, 2012; pp. 1–9, ISBN 978-0-511-74036-7.
3. UN. *Sendai Framework for Disaster Risk Reduction 2015–2030*; The United Nations Office for Disaster Risk Reduction, Ed.; UN: Geneva, Switzerland, 2015.
4. Glade, T. Landslide Hazard Assessment and Historical Landslide Data—An Inseparable Couple? In *The Use of Historical Data in Natural Hazard Assessments*; Glade, T., Albini, P., Francés, F., Eds.; Springer: Dordrecht, The Netherlands, 2001; Volume 17, pp. 153–168, ISBN 978-90-481-5762-4.
5. van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial Data for Landslide Susceptibility, Hazard, and Vulnerability Assessment: An Overview. *Eng. Geol.* **2008**, *102*, 112–131. <https://doi.org/10.1016/j.enggeo.2008.03.010>.

6. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-Scale Study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [https://doi.org/10.1016/s0169-555x\(99\)00078-1](https://doi.org/10.1016/s0169-555x(99)00078-1).
7. van Westen, C.J.; van Asch, T.W.J.; Soeters, R. Landslide Hazard and Risk Zonation—Why Is It Still so Difficult? *Bull. Eng. Geol. Environ.* **2006**, *65*, 167–184. <https://doi.org/10.1007/s10064-005-0023-0>.
8. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A Review of Statistically-Based Landslide Susceptibility Models. *Earth-Sci. Rev.* **2018**, *180*, 60–91. <https://doi.org/10.1016/j.earscirev.2018.03.001>.
9. Soeters, R.; Van Westen, C.J. Slope Instability Recognition Analysis and Zonation. In *Landslides: Investigation and mitigation*; Turner, A.K., Schuster, R.L., Eds.; Special report/Transportation Research Board, National Research Council; National Academy Press: Washington, DC, USA, 1996; pp. 129–177, ISBN 978-0-309-06208-4.
10. Aleotti, P.; Chowdhury, R. Landslide Hazard Assessment: Summary Review and New Perspectives. *Bull. Eng. Geol. Environ.* **1999**, *58*, 21–44. <https://doi.org/10.1007/s100640050066>.
11. Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.-P.; Fotopoulou, S.; Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F.; et al. Recommendations for the Quantitative Analysis of Landslide Risk. *Bull. Eng. Geol. Environ.* **2013**, *73*, 209–263. <https://doi.org/10.1007/s10064-013-0538-8>.
12. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for Landslide Susceptibility, Hazard and Risk Zoning for Land Use Planning. *Eng. Geol.* **2008**, *102*, 85–98. <https://doi.org/10.1016/j.enggeo.2008.03.022>.
13. Sinčić, M.; Bernat Gazibara, S.; Krkač, M.; Mihalić Arbanas, S. Landslide Susceptibility Assessment of the City of Karlovac Using the Bivariate Statistical Analysis. *MGPB* **2022**, *38*, 149–170. <https://doi.org/10.17794/rgn.2022.2.13>.
14. Yusof, N.M.; Pradhan, B.; Shafri, H.Z.M.; Jebur, M.N.; Yusoff, Z. Spatial Landslide Hazard Assessment along the Jelapang Corridor of the North-South Expressway in Malaysia Using High Resolution Airborne LiDAR Data. *Arab. J. Geosci.* **2015**, *8*, 9789–9800. <https://doi.org/10.1007/s12517-015-1937-x>.
15. Petschko, H.; Brenning, A.; Bell, R.; Goetz, J.; Glade, T. Assessing the Quality of Landslide Susceptibility Maps—Case Study Lower Austria. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 95–118. <https://doi.org/10.5194/nhess-14-95-2014>.
16. Gaidzik, K.; Ramírez-Herrera, M.T. The Importance of Input Data on Landslide Susceptibility Mapping. *Sci. Rep.* **2021**, *11*, 19334. <https://doi.org/10.1038/s41598-021-98830-y>.
17. Pellicani, R.; Van Westen, C.J.; Spilotro, G. Assessing Landslide Exposure in Areas with Limited Landslide Information. *Landslides* **2014**, *11*, 463–480. <https://doi.org/10.1007/s10346-013-0386-4>.
18. Zêzere, J.L.; Pereira, S.; Melo, R.; Oliveira, S.C.; Garcia, R.A.C. Mapping Landslide Susceptibility Using Data-Driven Methods. *Sci. Total Environ.* **2017**, *589*, 250–267. <https://doi.org/10.1016/j.scitotenv.2017.02.188>.
19. Lee, S.; Min, K. Statistical Analysis of Landslide Susceptibility at Yongin, Korea. *Environ. Geol.* **2001**, *40*, 1095–1113. <https://doi.org/10.1007/s002540100310>.
20. Vojteková, J.; Vojtek, M. Assessment of Landslide Susceptibility at a Local Spatial Scale Applying the Multi-Criteria Analysis and GIS: A Case Study from Slovakia. *Geomat. Nat. Hazards Risk* **2020**, *11*, 131–148. <https://doi.org/10.1080/19475705.2020.1713233>.
21. Nath, S.K.; Sengupta, A.; Srivastava, A. Remote Sensing GIS-Based Landslide Susceptibility & Risk Modeling in Darjeeling–Sikkim Himalaya Together with FEM-Based Slope Stability Analysis of the Terrain. *Nat. Hazards* **2021**, *108*, 3271–3304. <https://doi.org/10.1007/s11069-021-04823-5>.
22. Roșca, S.; Bilașco, Ș.; Petrea, D.; Fodorean, I.; Vescan, I.; Filip, S.; Măguț, F.-L. Large Scale Landslide Susceptibility Assessment Using the Statistical Methods of Logistic Regression and BSA—Study Case: The Sub-Basin of the Small Niraj (Transylvania Depression, Romania); Landslides and Debris Flows Hazards. *Nat. Hazards Earth Syst. Sci. Discuss.* **2015**, *3*, 7171–7201.
23. Xing, Y.; Yue, J.; Guo, Z.; Chen, Y.; Hu, J.; Travé, A. Large-Scale Landslide Susceptibility Mapping Using an Integrated Machine Learning Model: A Case Study in the Lvliang Mountains of China. *Front. Earth Sci.* **2021**, *9*, 722491. <https://doi.org/10.3389/feart.2021.722491>.
24. Santangelo, M.; Marchesini, I.; Bucci, F.; Cardinali, M.; Fiorucci, F.; Guzzetti, F. An Approach to Reduce Mapping Errors in the Production of Landslide Inventory Maps. *Nat. Hazards Earth Syst. Sci.* **2015**, *15*, 2111–2126. <https://doi.org/10.5194/nhess-15-2111-2015>.
25. Guzzetti, F.; Mondini, A.C.; Cardinali, M.; Fiorucci, F.; Santangelo, M.; Chang, K.T. Landslide Inventory Maps: New Tools Old Problem. *Earth-Sci. Rev.* **2012**, *112*, 42–66. <https://doi.org/10.1016/j.earscirev.2012.02.001>.
26. Jaboyedoff, M.; Oppikofer, T.; Abellán, A.; Derron, M.-H.; Loye, A.; Metzger, R.; Pedrazzini, A. Use of LIDAR in Landslide Investigations: A Review. *Nat. Hazards* **2012**, *61*, 5–28. <https://doi.org/10.1007/s11069-010-9634-2>.
27. Scaioni, M.; Longoni, L.; Melillo, V.; Papini, M. Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. *Remote Sens.* **2014**, *6*, 9600–9652. <https://doi.org/10.3390/rs6109600>.
28. Tarolli, P. High-Resolution Topography for Understanding Earth Surface Processes: Opportunities and Challenges. *Geomorphology* **2014**, *216*, 295–312. <https://doi.org/10.1016/j.geomorph.2014.03.008>.
29. Ray, R.L.; Lazzari, M.; Olutimehin, T. Remote Sensing Approaches and Related Techniques to Map and Study Landslides. In *Landslides—Investigation and Monitoring*; Ray, R., Lazzari, M., Eds.; IntechOpen: London, UK, 2020; ISBN 978-1-78985-823-5.
30. Chigira, M.; Duan, F.; Yagi, H.; Furuya, T. Using an Airborne Laser Scanner for the Identification of Shallow Landslides and Susceptibility Assessment in an Area of Ignimbrite Overlain by Permeable Pyroclastics. *Landslides* **2004**, *1*, 203–209. <https://doi.org/10.1007/s10346-004-0029-x>.

31. Eeckhaut, M.V.D.; Poesen, J.; Verstraeten, G.; Vanacker, V.; Nyssen, J.; Moeyersons, J.; van Beek, L.P.H.; Vandekerckhove, L. Use of LIDAR-Derived Images for Mapping Old Landslides under Forest. *Earth Surf. Process. Landf.* **2007**, *32*, 754–769. <https://doi.org/10.1002/esp.1417>.
32. Razak, K.A.; Straatsma, M.W.; van Westen, C.J.; Malet, J.-P.; de Jong, S.M. Airborne Laser Scanning of Forested Landslides Characterization: Terrain Model Quality and Visualization. *Geomorphology* **2011**, *126*, 186–200. <https://doi.org/10.1016/j.geomorph.2010.11.003>.
33. Domlija, P. Identification and Classification of Landslides and Erosion Phenomena Using the Visual Interpretation of the Vinodol Valley Digital Elevation Model. Ph.D. Thesis, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb: Zagreb, Croatia, 2018.
34. Bernat Gazibara, S. Methodology for Landslide Mapping Using High Resolution Digital Elevation Model in the Podsljeme Area (City of Zagreb). Ph.D. Thesis, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb: Zagreb, Croatia, 2019.
35. Bernat Gazibara, S.; Krkač, M.; Mihalić Arbanas, S. Landslide Inventory Mapping Using LiDAR Data in the City of Zagreb (Croatia). *J. Maps* **2019**, *15*, 773–779. <https://doi.org/10.1080/17445647.2019.1671906>.
36. Bernat Gazibara, S.; Krkač, M.; Mihalić Arbanas, S. Verification of Historical Landslide Inventory Maps for the Podsljeme Area in the City of Zagreb Using LiDAR-Based Landslide Inventory. *MGPB* **2019**, *34*, 45–58. <https://doi.org/10.17794/rgn.2019.1.5>.
37. Jebur, M.N.; Pradhan, B.; Tehrany, M.S. Optimization of Landslide Conditioning Factors Using Very High-Resolution Airborne Laser Scanning (LiDAR) Data at Catchment Scale. *Remote Sens. Environ.* **2014**, *152*, 150–165. <https://doi.org/10.1016/j.rse.2014.05.013>.
38. Mihalić Arbanas, S.; Bernat Gazibara, S.; Sećanj, M.; Damjanović, V.; Oršanić, D.; Penović, S.; Krkač, M.; Cindrić Kalin, K.; Domlija, P.; Jagodnik, V.; et al. Landslide Risk Management in Croatia: Current State. In Proceedings of the 4th Regional Symposium on Landslides in the Adriatic—Balkan Region, Sarajevo, Bosnia and Herzegovina, 23–25 October 2019; Društvo za geotehniku u Bosni i Hercegovini: Tuzla, Bosnia and Herzegovina, 2019; pp. 243–250.
39. Bernat Gazibara, S.; Cindrić, K.; Erak, M.; Krkač, M.; Sećanj, M.; Domlija, P.; Arbanas, Ž.; Mihalić Arbanas, S. Landslide Hazard Analysis in National-Scale for Landslide Risk Assessment in Croatia. In Proceedings of the 4th Regional Symposium on Landslides in the Adriatic—Balkan Region, Sarajevo, Bosnia and Herzegovina, 23–25 October 2019; Društvo za geotehniku u Bosni i Hercegovini: Prijedor, Bosnia and Herzegovina, 2019; pp. 175–182.
40. Mihalić Arbanas, S.; Sećanj, M.; Bernat Gazibara, S.; Krkač, M.; Begić, H.; Džindo, A.; Zekan, S.; Arbanas, Ž. Landslides in the Dinarides and Pannonian Basin—from the Largest Historical and Recent Landslides in Croatia to Catastrophic Landslides Caused by Cyclone Tamara (2014) in Bosnia and Herzegovina. *Landslides* **2017**, *14*, 1861–1876. <https://doi.org/10.1007/s10346-017-0880-1>.
41. Bernat Gazibara, S.; Mihalić Arbanas, S.; Sinčić, M.; Krkač, M.; Lukačić, H.; Jagodnik, P.; Arbanas, Ž. LandSlidePlan-Scientific Research Project on Landslide Susceptibility Assessment in Large Scale. In Proceedings of the 5th Regional Symposium on Landslides in Adriatic-Balkan Region, Rijeka, Croatia, 23–26 March 2022; Faculty of Civil Engineering, University of Rijeka and Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb: Rijeka, Croatia, 2022.
42. URL-1. Available online: http://Arhiva.Vzz.Hr/Images/Stories/Prostorni-Plan/LEPOGLAVA_ID/LEPOGLAVA_ID2/K1_KORISTENJE_I_NAMJENA_POVRSINA_HTRS.Pdf (accessed on 15 June 2022).
43. URL-2. Available online: http://Arhiva.Vzz.Hr/Images/Stories/Prostorni-Plan/BEDNJA/2-ID-PPUO-Bednja-05-2017/K1_KORISTENJE_I_NAMJENA_POVRSINA_ID2.Pdf (accessed on 15 June 2022).
44. Aničić, B.; Juriša, M. Basic geological map, scale 1:100,000, Rogatec, Sheet 33–68. 1984.
45. Šimunić, A.; Pikija, M.; Hečimović, I. Basic geological map, scale 1:100,000, Varaždin, Sheet 33–69. 1982.
46. Šimunić, A.; Pikija, M.; Hečimović, I.; Šimunić, A. Geological notes for Basic geological map, scale 1:100,000, Varaždin, Sheet 33–69. 1982.
47. Aničić, B.; Juriša, M. Geological notes for Basic geological map, scale 1:100,000, Rogatec, Sheet 33–68. 1983.
48. Zaninović, K.; Gajić-Čapka, M.; Perčec Tadić, M.; Vučetić, M.; Milković, J.; Bajić, A.; Cindrić, K.; Cvitan, L.; Katušin, Z.; Kaučić, D.; et al. Climate atlas of Croatia 1961–1990, 1971–2000. 2008.
49. Bernat, S.; Mihalić Arbanas, S.; Krkač, M. Inventory of Precipitation Triggered Landslides in the Winter of 2013 in Zagreb (Croatia, Europe). In *Landslide Science for a Safer Geoenvironment*; Sassa, K., Canuti, P., Yin, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2014; pp. 829–835, ISBN 978-3-319-05049-2.
50. Bernat, S.; Mihalić Arbanas, S.; Krkač, M. Landslides Triggered in the Continental Part of Croatia by Extreme Precipitation in 2013. In *Engineering Geology for Society and Territory—Volume 2*; Lollino, G., Giordan, D., Crosta, G.B., Corominas, J., Azzam, R., Wasowski, J., Sciarra, N., Eds.; Springer International Publishing: Cham, Switzerland, 2015; pp. 1599–1603, ISBN 978-3-319-09056-6.
51. URL-3. Available online: <https://Land.Copernicus.Eu/Imagery-in-Situ/Eu-Dem/Eu-Dem-v1.1?Tab=download> (Eu Dem_v11_E40N20.Tif (accessed on 15 June 2022)).
52. URL-4. Available online: <https://Land.Copernicus.Eu/Pan-European/Corine-Land-Cover/Clc2018?Tab=download> (CLC2018_CLC2012_V2018_20b2.Gdb) (accessed on 15 June 2022).
53. URL-5. Available online: <https://Land.Copernicus.Eu/about> (accessed on 15 June 2022).

54. URL-7. Available online: <https://Dgu.Gov.Hr/UserDocsImages/Dokumenti/Pristup%20informacijama/Zakoni%20i%20ostali%20propisi/Specifikacije/SP%20za%20izradu%20digitalnih%20ortofotokarata.ZIP> (accessed on 15 June 2022).
55. URL-8. Available online: <http://Geoportal.Dgu.Hr/Wms?Layers=TK25> (accessed on 15 June 2022).
56. URL-9. Available online: <http://Geoportal.Dgu.Hr/Wms?Layers=HOK> (accessed on 15 June 2022).
57. URL-10. Available online: <https://SciHub.Copernicus.Eu/Dhus/#/Home> (accessed on 15 June 2022).
58. URL-11. Available online: <http://Geoportal.Dgu.Hr/Wms?Layers=DOF> (accessed on 15 June 2022).
59. URL-12. Available online: <http://Download.Geofabrik.de/Europe.Html> (accessed on 15 June 2022).
60. URL-6. Available online: https://Land.Copernicus.Eu/User-Corner/Technical-Library/Eu-Dem-v1-1-User-Guide/At_download/File (accessed on 15 June 2022).
61. URL-13. Available online: https://Sentinels.Copernicus.Eu/Documents/247904/685211/Sentinel-2_User_Handbook.Pdf/8869acdf-Fd84-43ec-Ae8c-3e80a436a16c?T=1438278087000 (accessed on 15 June 2022).
62. Mayer, D.G.; Butler, D.G. Statistical Validation. *Ecol. Model.* **1993**, *68*, 21–32. [https://doi.org/10.1016/0304-3800\(93\)90105-2](https://doi.org/10.1016/0304-3800(93)90105-2).
63. Razak, K.A.; Santangelo, M.; Van Westen, C.J.; Straatsma, M.W.; de Jong, S.M. Generating an Optimal DTM from Airborne Laser Scanning Data for Landslide Mapping in a Tropical Forest Environment. *Geomorphology* **2013**, *190*, 112–125. <https://doi.org/10.1016/j.geomorph.2013.02.021>.
64. Glade, T.; Crozier, M.J. A Review of Scale Dependency in Landslide Hazard and Risk Analysis. In *Landslide Hazard and Risk*; Glade, T., Anderson, M., Crozier, M.J., Eds.; Wiley: Hoboken, NJ, USA, 2005; pp. 75–138, ISBN 978-0-471-48663-3.
65. Evans, J.S.; Oakleaf, J.; Cushman, S.A.; Theobald, D. An ArcGIS Toolbox for Surface Gradient and Geomorphometric Modeling, Version 2.0-0. 2014. Available online: <http://evansmurphy.wix.com/evansspatial> (accessed on 2 December 2015).
66. Carrara, A.; Guzzetti, F.; Cardinali, M.; Reichenbach, P. Use of GIS Technology in the Prediction and Monitoring of Landslide Hazard. *Nat. Hazards* **1999**, *20*, 117–135. <https://doi.org/10.1023/A:1008097111310>.
67. Jagodnik, P.; Bernat Gazibara, S.; Arbanas, Ž.; Mihalić Arbanas, S. Engineering Geological Mapping Using Airborne LiDAR Datasets—An Example from the Vinodol Valley, Croatia. *J. Maps* **2020**, *16*, 855–866. <https://doi.org/10.1080/17445647.2020.1831980>.
68. Jagodnik, P.; Bernat Gazibara, S.; Jagodnik, V. Types and Distribution of Quarternary Deposits Originating from Carbonate Rock Slopes in the Vinodol Valley, Croatia—New Insight Using Airborne LiDAR Data. *MGPB* **2020**, *35*, 57–77. <https://doi.org/10.17794/rgn.2020.4.6>.
69. Krkač, M.; Bernat Gazibara, S.; Sinčić, M.; Lukačić, H.; Mihalić Arbanas, S. Landslide Inventory Mapping Based on LiDAR Data: A Case Study from Hrvatsko Zagorje (Croatia). In Proceedings of the 5th ReSyLAB “Landslide Modelling & Applications”, Rijeka, 23–26 March 2022; Peranić, J., Vivoda Prodan, M., Bernat Gazibara, S., Krkač, M., Snježana Mihalić, A., Arbanas, Ž., Eds.; Faculty of Civil Engineering, University of Rijeka, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb: Rijeka, Zagreb, 2022; pp. 81–86.
70. Dearman, W.R. *Engineering Geological Mapping*; Elsevier: Amsterdam, The Netherlands, 1991; ISBN 978-0-7506-1010-0.
71. Santangelo, M.; Cardinali, M.; Rossi, M.; Mondini, A.C.; Guzzetti, F. Remote Landslide Mapping Using a Laser Rangefinder Binocular and GPS. *Nat. Hazards Earth Syst. Sci.* **2010**, *10*, 2539–2546. <https://doi.org/10.5194/nhess-10-2539-2010>.
72. Van Den Eeckhaut, M.; Vanwalleghe, T.; Poesen, J.; Govers, G.; Verstraeten, G.; Vandekerckhove, L. Prediction of Landslide Susceptibility Using Rare Events Logistic Regression: A Case-Study in the Flemish Ardennes (Belgium). *Geomorphology* **2006**, *76*, 392–410. <https://doi.org/10.1016/j.geomorph.2005.12.003>.

2.2 Comparison of conditioning factor classification criteria in large-scale statistically based landslide susceptibility models

Authors:

Marko Sinčić, Sanja Bernat Gazibara, Mauro Rossi, Snježana Mihalić Arbanas

Published in:

Natural Hazards and Earth System Sciences, 25 (2025), 183-206

Published on:

7 January 2025

DOI:

<https://doi.org/10.5194/nhess-25-183-2025>



Comparison of conditioning factor classification criteria in large-scale statistically based landslide susceptibility models

Marko Sinčić¹, Sanja Bernat Gazibara¹, Mauro Rossi², and Snježana Mihalić Arbanas¹

¹Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Pierottijeva 6, 10000 Zagreb, Croatia

²Consiglio Nazionale delle Ricerche, Istituto di Ricerca per la Protezione Idrogeologica, via Madonna Alta 126, 06128 Perugia, Italy

Correspondence: Marko Sinčić (msincic@rgn.hr)

Received: 14 February 2024 – Discussion started: 26 February 2024

Revised: 10 September 2024 – Accepted: 19 October 2024 – Published: 7 January 2025

Abstract. The large-scale landslide susceptibility assessment (LSA) is an important tool for reducing landslide risk through the application of resulting maps in spatial and urban planning. The existing literature more often deals with LSA modelling techniques, and the scientific research very rarely focuses on acquiring relevant thematic and landslide data, necessary to achieve reliable results. Therefore, the paper focuses on the crucial step of classifying continuous landslide conditioning factors for susceptibility modelling by presenting an innovative comprehensive analysis that resulted in 54 landslide susceptibility models to test 11 classification criteria (scenarios which vary from stretched values, partially stretched classes, heuristic approach, classification based on studentized contrast and landslide presence, and commonly used classification criteria, such as natural neighbour, quantiles and geometrical intervals) in combination with 5 statistical methods. The large-scale landslide susceptibility models were derived for small and shallow landslides in the pilot area (21 km²) located in the City of Zagreb (Croatia), which occur mainly in soils and soft rocks. Some of the novelties in LSA are the following: scenarios using stretched landslide conditioning factor values or classification with more than 10 classes prove more reliable; certain statistical methods are more sensitive to the landslide conditioning factor classification criteria than others; all the tested machine learning methods give the best landslide susceptibility model performance using continuous stretched landslide conditioning factors derived from high-resolution input data. The research highlights the importance of qualitative assessments, alongside commonly used quantitative metrics, to verify spatial ac-

curacy and to test the applicability of derived landslide susceptibility maps for spatial planning purposes.

1 Introduction

Addressing landslide hazard is commonly done by zoning, i.e. deriving landslide susceptibility, hazard and risk zoning maps (Corominas et al., 2014). Brabb (1984) defined landslide susceptibility as a “likelihood of a landslide occurring in a given area”, indicating the spatial component exclusively. Soeters and van Westen (1996), Guzzetti et al. (1999), van Westen et al. (2008), Fell et al. (2008a, b), Corominas et al. (2014), and Reichenbach et al. (2018) represent some of the most significant progress done in the field of research considering landslide susceptibility. One of the most popular approaches to derive landslide susceptibility models (LSMs) is using statistical methods, where the most recent and detailed review of statistically based landslide susceptibility models is given in Reichenbach et al. (2018), whereas Merghadi et al. (2020) emphasize only machine learning methods by reviewing algorithm performance. The literature (e.g. van Westen et al., 2008; Corominas et al., 2014) has shown two main groups of data needed for applying statistical methods in landslide susceptibility assessments (LSAs), i.e. landslide inventory maps and landslide conditioning factors (LCFs). Moreover, Reichenbach et al. (2018) identified acquiring relevant landslide and thematic information (i.e. LCF) as the first two steps for preparing a LSA. Many papers discuss different methods (e.g. Wang et al., 2016a; Chen et al., 2017; Merghadi et al., 2020), mapping units (e.g. Bor-

naetxea et al., 2018; Jacobs et al., 2020), inventory types (e.g. Guzzetti et al., 2012; Petschko et al., 2016), and the importance and/or selection of LCFs (Donati and Turrini, 2002; Jebur et al., 2014; Gaidzik and Ramírez, 2021), which are all necessary steps in a LSA according to Reichenbach et al. (2018).

This research is situated in the European Pannonian Basin, i.e. in the City of Zagreb in the hilly region of the southern foothills of Medvednica, which are susceptible to sliding. Despite long-term investigation on landslide phenomena, Jurak et al. (1996) and Mihalić (1998) identified the lack of sustainable landslide inventories and landslide hazard maps as the main issue in the landslide risk management in the Republic of Croatia. As a result, a landslide inventory map (Bernat Gazibara et al., 2019a) and a landslide susceptibility map (Bernat Gazibara et al., 2023) were derived, followed by emphasizing the importance and necessity of large-scale landslide susceptibility maps in the system of spatial and urban planning (Mihalić Arbanas et al., 2023). Regardless of the degree of urbanization, landslide occurrence is commonly related to geomorphological, geological and climate settings, as well as anthropogenic factors. Furthermore, Bernat Gazibara et al. (2017) indicate that the main landslide triggers in the area are precipitation and snow melting, i.e. long continuous precipitation periods or short precipitation periods of high intensity. The input data for LSA in this study were successfully acquired during previous research investigations in the study area (e.g. Bernat Gazibara et al., 2017, 2019a, 2023) and past large-scale assessment in Croatia (Sinčić et al., 2022a), but their optimal application in landslide susceptibility modelling has remained an open question. The relevance and new insight into acquiring input data for preparing LCFs on a large scale (i.e. 1 : 5000) were provided by Sinčić et al. (2022a) and proved by analysing the predictive performances of large-scale LSAs by Krkač et al. (2023).

This relevant aspect of LSA was addressed by Jebur et al. (2014), which used weight of evidence (WoE), logistic regression (LR) and support vector machine (SVM) methods in a large-scale case study to compare a lidar LCF set with a set containing additional LCFs such as land use and geology. Area under the curve (AUC) comparison between the two scenarios favoured lidar-only-derived LCFs, addressing the number and type of LCFs used in the research but not their classification criteria. Similarly, Dou et al. (2015) applied LR and statistical index methods to demonstrate that six LCFs with high correlation to landslide occurrence result in better success and prediction rate than a complete set of 15 LCFs. Donati and Turrini (2002) show that when using categorical LCFs, only a few classes can significantly influence the LSA, confirming the importance of how LCFs are classified. These studies showed the need for relevant LCF acquisition and selection criteria, whereas the leave-one-out test available in LAND-SUITE software (Rossi et al., 2022), certainty factor models (Dou et al., 2015) and variable im-

portance (Shirvani, 2020) are some methods used for LCF selection.

LCF classification issues were discussed by Yan et al. (2019), who defined two classification criteria for LCFs selected by reviewing classification criteria in different studies available in the literature. Specifically, the 2 classification criteria for each of the 5 LCFs result with fewer or more classes which were used to develop 32 LSM scenarios to test all the combinations. From the 32 derived maps, a low difference of 0.03 AUC between minimum and maximum AUC was identified. Namely, the best result was obtained when all LCFs were used in a scenario that included more classes in most LCFs. Xiao et al. (2021) implemented a different strategy, aimed to reduce local correlations among LCFs by reclassifying them to increase LSA accuracy. As stated by Bonham-Carter et al. (1990) and Neuhäuser and Terhorst (2007), contrast or studentized contrast (C_{st}) metrics can be used to determine classification cut-off values in continuous LCFs when using a bivariate approach such as WoE or information value (IV). Moreover, generalizing continuous LCFs enables maximized spatial relations and statistical robustness (Neuhäuser et al., 2012). Concretely, Mathew et al. (2007) used the C_{st} curve maximum to split the distance to roads, drainage and lineament LCFs into two classes. Furthermore, Neuhäuser et al. (2012) applied the C_{st} curve in the WoE method to convert continuous LCFs to categorical by observing maximum and local maximum C_{st} values. Jebur et al. (2014) used mainly the quantile criteria to derive 10 classes for stretched rasters and a heuristic approach for buffer zones from vector lines (i.e. proximities or distances). On the other hand, Yusof et al. (2015) applied only the quantile criteria to classify LCFs into 10 classes, based on the approach by Tehrany et al. (2013, 2014) in flood susceptibility mapping. Huang et al. (2020a, b) and Huang et al. (2022) applied the natural break classification criteria and defined eight classes in each LCF. Wang et al. (2016b, 2020), Cui et al. (2017), and Zhao and Chen (2020) used what seems to be equal intervals defined heuristically without specifications, whereas Wang et al. (2021) did not classify continuous LCFs and applied them as stretched rasters in the analysis. Several papers dealing with large-scale LSAs (e.g. Vojteková and Vojtek, 2020; Xing et al., 2021) also do not specify the classification criteria but present the LCFs with popular equal interval classes likely defined heuristically. Similarly, Wang et al. (2016b) used equal intervals to classify slope and buffer zones for LCFs, whereas a heuristic approach was selected for stretched rasters. In most of the mentioned studies, which applied a heuristic approach, a relatively low number of classes in LCFs are often defined, i.e. from four to six classes. A detailed comparison of all these criteria and techniques is still missing, especially on large-scale assessments and with adequate metrics. Based on the latter and as confirmed by Huang et al. (2020b) and Xing et al. (2021), there is no uniform approach to classifying continuous LCFs. It can be concluded that no systematic analysis to compare the

criteria was done, as researchers are instead applying what was already confirmed to be successful in various individual case studies.

Few papers deal with LCF processing once they are selected, i.e. discussing the process of classifying continuous LCFs (e.g. Yan et al., 2019; Xiao et al., 2021), which is the scope of this study. Therefore, this paper compares 11 criteria used to define the processing of continuous LCFs, i.e. transform relevant input data layers into LCFs suitable for application in landslide susceptibility modelling. Unlike Yan et al. (2019), who are more oriented to case study combinations of LCFs with different derived LCF classes, we aim to develop a uniform approach for classifying continuous LCFs, leading to defining a first step for large-scale LSA methodology. Besides testing the 11 scenarios where continuous LCFs are categorized, each scenario is applied in 5 statistical landslide susceptibility methods, including IV, LR, neural network (NN), random forest (RF) and SVM. Such methods were considered to analyse the influence of LCF classification criteria on LSA accuracy using different statistical modelling approaches, and they were selected because they are commonly used in LSAs (Reichenbach et al., 2018).

Remote sensing has proven helpful for LSAs, leading to better landslide hazard mitigation strategies. Concretely, different remote sensing techniques are widely applied for data acquisition when studying landslide hazard, as presented by Scaioni et al. (2014) and Ray et al. (2020), whereas in this paper, geomorphological and hydrological LCFs are based on lidar (light detection and ranging) point cloud data acquired by airborne laser scanning (ALS). Furthermore, the landslide inventory map was derived by mapping on morphometric maps derived from a high-resolution DTM (digital terrain model), which shows a significant advantage compared to other methods for mapping small landslides under vegetation (Razak et al., 2013; Bernat Gazibara et al., 2019a). Jebur et al. (2014) argue that lidar-derived LCFs could be sufficient for LSAs where geological LCFs (e.g. soil and geology) are not available, whereas Sinčić et al. (2022a) were able to derive a complete set of LCFs together with a set of elements at risk of large-scale LSAs by exclusively using high-resolution lidar and orthophoto data in combination with publicly available small-scale geological data.

Besides a study case presented by Yusof et al. (2015) using LR and an evidential belief function, comparison of LSA methods on a large scale is not common in the literature. Hence, one of the novelties of this paper is presenting a detailed comparison, including the quantitative and qualitative perspective of five popular LSA methods applied on a large scale with relevant input data of high spatial accuracy. Zêzere et al. (2017) point out that the best AUC metric values do not necessarily define the best LSM. Similarly, Vakhshoori and Zare (2018) emphasize that AUC values might be deceiving or ineffective in detecting uncertainties in spatial prediction and can only indicate general reliability. To address the issue, Vakhshoori and Zare (2018) suggest studying additional

metrics, such as Cohen's κ , to acquire additional information about the LSM. Considering the susceptibility quality level (SQL) introduced by Guzzetti et al. (2006a, b), Reichenbach et al. (2018) identified a low number of papers published with a high SQL rank. Addressing model fitting and predictive performance and measuring uncertainties in the derived models, 54 LSMs based on combinations of 5 methods (IV, LR, RF, NN, SVM) and 11 LCF classification criteria derived in this paper are considered high SQL, ensuring systematic and unbiased comparison. Moreover, researching the application of LSMs in spatial planning (Mihalić Arbanas et al., 2023), we qualitatively present derived LSMs in close-up views and high resolution to measure the applicability of the used zonation method, i.e. determine visually susceptibility class area distribution. However, choosing the optimal zonation method is out of the scope of this study, and the one presented in this paper serves exclusively for uniform comparison of derived LSMs. Moreover, it should be stated that optimization of used statistical methods is a topic beyond the objective of this research. We argue that large-scale LSAs using high-quality input data should also be measured qualitatively as expert judgement by observing real environmental conditions on high-resolution DTM derivatives. The observation provides necessary insight into map quality and applicability, which is not detected by commonly used quantitative approaches (e.g. AUC).

2 Materials

2.1 Study area

The study area is located in the City of Zagreb, the capital of Croatia, in the northwest part of the country. It encompasses 21 km² of the southern slopes of Medvednica, i.e. the western part of the Podsljeme area. The City of Zagreb belongs to the European Pannonian Basin, whereas the Podsljeme area encompassing the study area is depicted with hilly relief, with 90 % of the study area steeper than 5° (Bernat Gazibara et al., 2019b). Basic geological settings can be described as Upper Miocene and Quaternary sediments making up 92 % of the study area, described in detail in the geological map 1 : 100 000 and supplementary geological notes developed by Šikić et al. (1972, 1979), respectively. With a dense population and a high degree of urbanization, which is increasing, the Podsljeme area has been under research regarding landslide phenomena in the last 50 years, starting with the first landslide inventory map by Šikić (1967) and following geomorphological landslide maps from 1979 (Polak et al., 1979) and a geomorphological inventory from 2006 (Miklin et al., 2007). More recently, Bernat Gazibara et al. (2014a, b) described landslide events triggered by intensive rainfall. Moreover, Mihalić Arbanas et al. (2016) identified the lack of a suitable landslide inventory as one of the critical issues in landslide risk management in the City of

Zagreb. As a result, a lidar-based landslide inventory map for the 21 km² in this study area was derived (Bernat Gazibara et al., 2019a, b), followed by a susceptibility assessment (Bernat Gazibara et al., 2023). Satisfactory results were obtained, confirming the proposed methodology for large-scale LSAs. Considering the highly urbanized environment, population density and a continuous increase of human-induced landslides in the Podsljeme area (Jurak et al., 2008; Mihalić Arbanas et al., 2014), a large-scale landslide susceptibility assessment is necessary for adequate landslide management in the study area.

2.2 Input data

The first landslide inventory map developed from lidar data for the study area was completed by Bernat Gazibara et al. (2019a, b) based on ALS, which was performed during the leaf-off period in Croatia in 2013. Furthermore, another lidar ALS was performed in 2020 (Bernat Gazibara et al., 2022, 2023) to verify the existing landslide inventory map, providing a multi-temporal insight into landslide occurrence. Namely, the landslide inventory map consists of 702 mapped polygons, with the most frequent landslide area being 400 m² and a density of 33 landslides per square kilometre. The input data layers needed to derive LCFs are prepared from source data and are classified according to their origin into continuous and categorical, as presented in Table 1. Furthermore, continuous input data layers can be subdivided into vector lines (e.g. geological contacts, faults, drainage network, all streams) and stretched rasters (e.g. elevation, slope, terrain wetness). On the other hand, lithology (rock type) and land use made up the polygon group of categorical input data layers, unlike aspect being the only categorical raster type. Continuous stretched rasters are presented by edge values (Fig. 1a, b, e), whereas categorical input data layers are depicted by their classes (Fig. 1c, d, g). Furthermore, the continuous line input data layers' spatial presence is illustrated in Fig. 1d–f.

We note that the used geological input data are on a small scale, but considering preliminary analysis in the LCF selection process, they resulted in being more relevant to landslide occurrences in terms of the leave-one-out test in comparison to the alternative geological map on a 1 : 5000 scale. Furthermore, the mentioned small-scale geological input data were also applied in Bernat Gazibara et al. (2023), where they yielded excellent results.

3 Methodology

3.1 Preparing input data

Categorical LCFs (aspect, lithology, land use) are equivalent to input data layers presented in Table 1 and were used equally in all 11 scenarios, an exception being aspect in scenario S11, as explained further in the section. Classes in cat-

egorical LCFs were ordered according to frequency ratio values defined by the presence of an unstable training landslide set. Continuous LCFs were derived differently from input data layers (Table 1, Fig. 1) for 11 scenarios and are briefly summarized in Table 2.

Continuous stretched rasters (elevation, slope, terrain wetness) were reclassified into 100 equal classes for scenario S1. Regarding continuous line vector LCFs (geological contact, faults, drainage network, all streams), 100 equal multiple buffer ring zones were derived to finalize scenario S1 LCFs (Fig. 2a). Similarly to scenario S1, scenario S2 is defined by 50 equal classes, as illustrated in Fig. 2b. The aim of defining scenarios S1 and S2 is to simulate the continuous input data layer to a high and low detailed extent, respectively. In scenario S3, LCFs were defined by heuristically classifying and defining buffer zones for continuous rasters and line input data layers, respectively. The subjective approach was led by researchers' experiences (e.g. Sinčić et al., 2022b; Krkač et al., 2023) and previous work in the study area (e.g. Bernat Gazibara et al., 2023).

For each of the 100 scenario S1 LCF classes, C_{st} values were calculated based on the bivariate approach, i.e. observing class pixel size and landslide pixels presence in the class. Namely, C_{st} is defined as a ratio of contrast (C) to its standard deviation $s(C)$ (Bonham-Carter, 1994):

$$C_{st} = \frac{C}{s(C)}, \quad (1)$$

where

$$C = W^+ - W^-, \quad (2)$$

where W^+ and W^- indicate positive and negative weight factors in the WoE method, respectively. Standard deviation $s(C)$ is defined as

$$s(C) = \sqrt{S(W^+)^2 + S(W^-)^2}, \quad (3)$$

where $S(W^+)^2$ and $S(W^-)^2$ are variances of weights defined by Bishop et al. (1975). The C_{st} curve was defined for each LCF, followed by observing positive and negative C_{st} trends throughout the 100 classes. Cut-off lines for scenario S4 were determined by dividing the C_{st} curve into positive and negative segments, as depicted in Fig. 2c. To define scenario S5, each scenario S4 positive segment was split into two by defining an additional cut-off line at the highest peak point (Fig. 2d).

Similarly, scenario S6 was defined by an additional cut-off line from scenario S5 at the lowest peak point of negative segments (Fig. 2e). For the theoretical example in Fig. 2c–e, the LCFs for scenarios S4, S5 and S6 would result in 3, 4 and 6 classes, respectively.

For scenario S7, the number of expected and mapped landslide pixels was determined for each of the 100 classes previously defined in scenario S1. For stretched rasters (e.g. elevation and slope), expected landslides are defined by the

Table 1. Overview of source data used for derivation of input data layers.

Source data	Scale	Input data layer	Obtained by	Categorization
Lidar point cloud	5 m resolution	Elevation	Interpolation	Continuous stretched raster
		Slope	ArcGIS 10.8 slope tool	Continuous stretched raster
		Aspect	ArcGIS 10.8 aspect tool	Categorical (classified) raster
		Terrain wetness	Evans et al. (2014)	Continuous stretched raster
		Drainage network	ArcGIS 10.8 Spatial Analyst Toolbox	Continuous line (vector)
Croatian basic Geological map	1 : 100 000	Lithology (rock type)	Digitization	Categorical polygon (vector)
		Geological contact	Digitization	Continuous line (vector)
		Faults	Digitization	Continuous line (vector)
Topographic map of Croatia	1 : 25 000	Streams	Digitization (State Geodetic Administration, 2024)	Continuous line (vector)
Land use planning maps	Large scale	Land use	City of Zagreb (2011)	Categorical polygon (vector)

Table 2. Summary of 11 defined scenarios considering different categorization criteria.

Scenario	Description of categorization criteria	
	Stretched rasters	Line vector
S1	100 equal classes	100 equal buffer zones
S2	50 equal classes	50 equal buffer zones
S3	Heuristic classification based on researcher experience	
S4	Studentized contrast defined on S1 classes, positive and negative segments	
S5	Studentized contrast defined on S1 classes, positive segments split into two, negative segments	
S6	Studentized contrast defined on S1 classes, positive segments split into two, negative segments split into two	
S7	Classification based on landslide density in S1 classes, i.e. expected vs. mapped landslide area	Classification based on landslide density in S1 classes, Eq. (4)
S8	10 classes defined by natural break (NB) classification method applied on input data layers (stretched raster)	10 classes defined by natural break (NB) classification method applied on scenario S1 classes
S9	10 classes defined by quantile (Q) classification method applied on input data layers (stretched raster)	10 classes defined by quantile (Q) classification method applied on scenario S1 classes
S10	10 classes defined by geometrical interval (GI) classification method applied on input data layers (stretched raster)	10 classes defined by geometrical interval (GI) classification method applied on scenario S1 classes
S11	Input data layers (stretched raster)	Scenario S1 classes

hypothesis stating that in each of the 100 LCF classes the landslide density should be equal to the total study area landslide density, whereas the mapped landslide area was acquired by simple observation. After calculating the mapped and expected landslide area for each class, landslide density curves considering each class are constructed (Fig. 2f). Furthermore, the cut-off lines were determined by trend changes of landslide area presence, i.e. at the points where expected

or mapped landslides change to being higher or lower than the other.

For an example given in Fig. 2f, the trends change three times resulting in three cut-off lines, i.e. four classes. On the other hand, the constant expected number of landslides A_{exp} for line vector LCFs used in scenario S7 was defined by

$$A_{\text{exp}} = \frac{i}{i_{\text{max}}} \times A_{\text{tot}}, \tag{4}$$

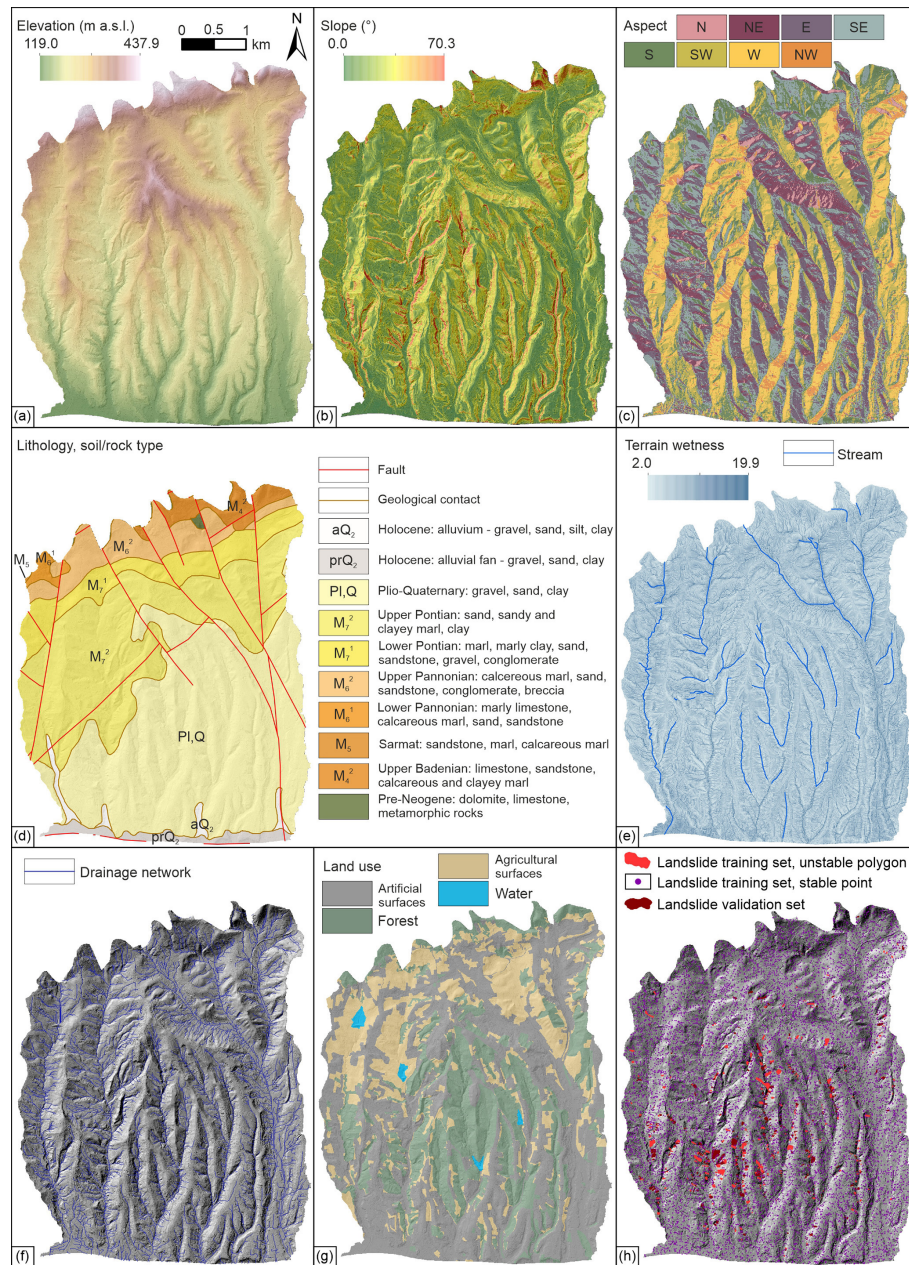


Figure 1. Input data layers for landslide susceptibility analyses: (a) elevation; (b) slope; (c) aspect; (d) lithology (soil/rock type), geological contact, and faults; (e) terrain wetness and streams; (f) drainage network; (g) land use; and (h) landslide dataset.

where i is the buffer interval size, $i_{\max}$ the maximum reached buffer distance and A_{tot} total landslide area. The equation aims to define an equal expected landslide area in each buffer ring. Then, the area of mapped landslides in each buffer ring is compared to the constant A_{exp} . Similarly, as with stretched rasters, trends in relations between mapped landslides being lower or higher than the constant A_{exp} define the cut-off lines. It should be noted that a training landslide dataset was used for calculations needed to define scenarios S4–S7.

Continuous stretched rasters and scenario S1 vector lines were reclassified using natural break (NB), quantile (Q) and geometrical interval (GI) classification criteria. As a result, scenarios S8, S9 and S10 are developed, with each having 10 classes, defined by NB, Q and GI reclassification criteria, respectively. Lastly, scenario S11 was defined by using continuous stretched rasters without classifying them, i.e. as input data layers containing edge values rather than classes. Moreover, the aspect input data layer was applied with its original stretched values (i.e. 0–360°) representing a continuous

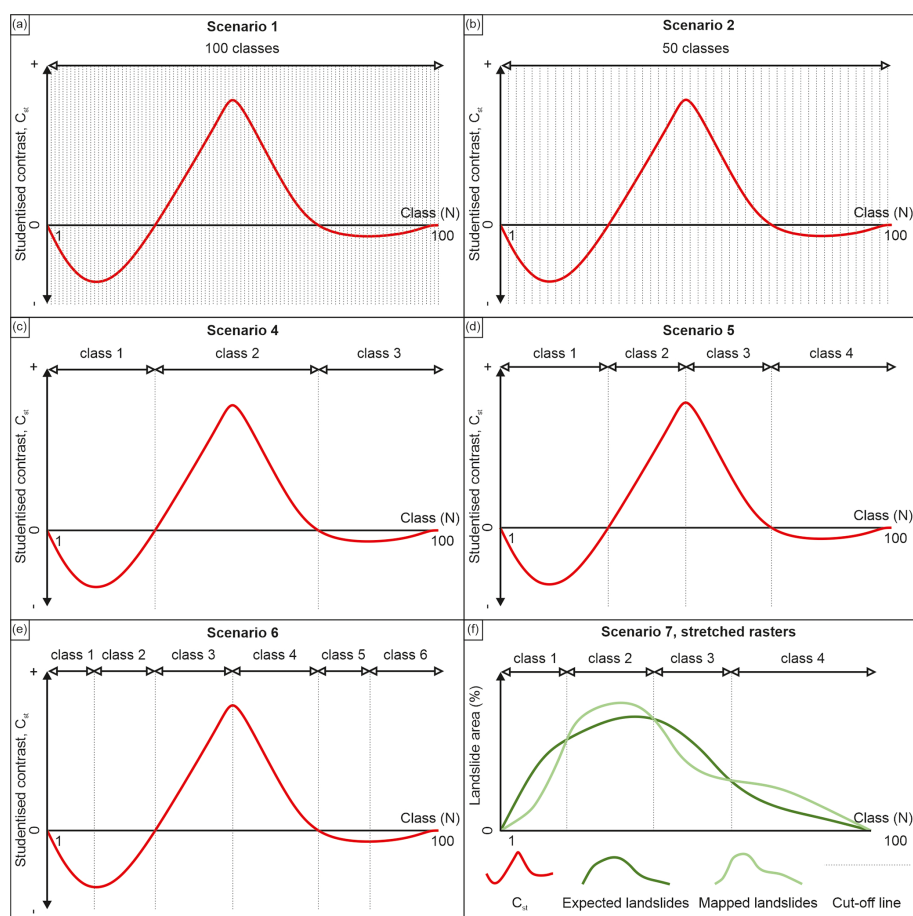


Figure 2. Theoretical example of the studentized contrast and landslide density curves for the methodology applied to define cut-off values in scenarios S1 (a), S2 (b), S4 (c), S5 (d), S6 (e) and S7 (f), respectively.

LCF, unlike being categorical in the remaining 10 scenarios. Lastly, scenario S1 line continuous LCFs with 100 equal classes were used in scenario S11 to simulate an input data layer as closely as possible. The presented classification criteria for 11 scenarios were applied uniformly, meaning each stretched raster and/or line vector was processed equally inside each scenario yet methodologically different from other scenarios. Scenarios for proposed LCF classification criteria vary from stretched (i.e. S11); partially stretched (i.e. S1 and S2); heuristic (i.e. S3); classified based on studentized contrast curve (i.e. S4, S5 and S6); classified based on expected and mapped landslide presence (i.e. S7); and lastly the commonly used classification criteria such as NB, Q and GI (i.e. S8, S9 and S10). The main aim is to test the stated classification criteria relevance and determine its influence and necessity while using different statistical methods in a large-scale case study.

3.2 Susceptibility analyses

Figure 3 synthesizes the methodology applied in this study that can be split into “preparing input data”, “modelling”,

and “quantitative and qualitative analyses” steps. Namely, relevant landslide and thematic (input data layers) information was obtained as described in Sect. 2. Preparing input data involves deriving continuous LCF sets applicable in 11 classification scenarios and defining a fixed landslide dataset. Susceptibility analysis includes deriving 54 LSMs using the selected mapping unit with 5 statistical methods applied with the 11 classified factor scenarios. The quantitative and qualitative analyses considered three directions: (i) model evaluation parameters, (ii) LSM validation parameters and (iii) LSM classification parameters. Commonly used quantitative parameters are examined in all three directions, whereas the qualitative approach is made only for the LSM classification, focusing on the LSM applicability, i.e. observing spatial distribution of susceptibility classes and variability of susceptibility values. The described workflow ensures the development of landslide susceptibility maps of high-quality rank according to the SQL (Guzzetti et al., 2006a, b). Lastly, the quantitative and qualitative analysis results were summarized, enabling drawing the conclusions.

Preparing input data begins with splitting the polygon-based landslide inventory map randomly into two sets, each containing the same number of polygons. Unstable training polygons (first set) were transformed into pixels, i.e. 5 m raster chosen as appropriate for a LSA on a large scale done in this study. Unstable training pixels were subtracted from the study area extent, and then an equal number of stable pixels were randomly selected in the remaining territory. This ensured an unbiased landslide training set with an equal number of stable and unstable pixels for deriving LSMs, which also defines the model training dataset for model evaluation parameters. An unstable set of polygons used for validation (second set) and all unstable polygons used for classification were transformed into 5 m rasters and used for determining LSM validation and classification parameters, respectively.

As a first step in susceptibility analyses (modelling), LCF collinearity testing was performed individually for 11 scenarios in LAND-SUITE (Rossi et al., 2022). Correlations were examined by detecting LCF collinearity regarding Pearson's r absolute value of 0.5 as the cut-off threshold. Namely, values higher than 0.5 indicate collinearity between two examined LCFs, suggesting excluding one from the further susceptibility analyses. Selected LCFs showed no collinearity in all 11 scenarios. Furthermore, 54 LSMs were derived using the prepared landslide dataset and 11 LCF sets in the 5 selected methods, i.e. information value (IV), logistic regression (LR), neural network (NN), random forest (RF) and support vector machine (SVM). It should be noted that the IV method was not applied in scenario S11, as described further below. The methods' optimization is out of the scope of this study, as they are already widely known, and their usage is discussed by researchers in LSA studies. For a detailed theoretical background about the methods applied in this paper, the readers should consult van Westen et al. (1993) and Merghadi et al. (2020).

IV is a simple bivariate statistical method developed by Yin and Yan (1988), based on landslide density, using the following equation:

$$I_i = \log \frac{S_i N}{N_i S}, \quad (5)$$

where S_i is the number of landslide pixels in the observed class, N_i the number of pixels in the observed class, S the number of landslide pixels used for model training, N number of pixels in the study area and I_i the information value of the observed variable. Positive and negative I_i values indicate instability and stability, respectively, whereas higher values indicate a stronger relationship. Detailed methodology for using the bivariate statistical approach in landslide susceptibility analyses is given in van Westen (2002), whereas IV is applied by Sarkar et al. (2013), Farooq and Akram (2021), and Krkač et al. (2023) in different LSAs.

To achieve posterior probability, i.e. probabilistic [0,1] susceptibility values, the numerical I_i is converted using the

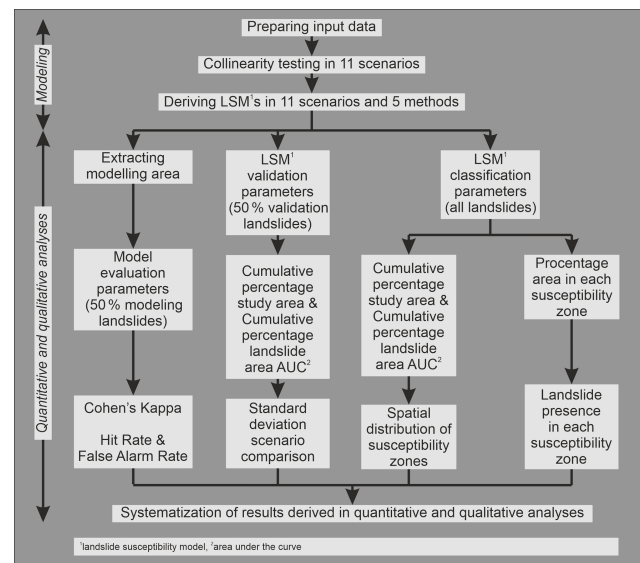


Figure 3. Landslide susceptibility modelling workflow.

following equation:

$$y = \frac{e^{f(x)}}{1 + e^{f(x)}}, \quad (6)$$

where $f(x)$ is the numerical susceptibility value (input) and y the probabilistic susceptibility value (output) as defined by Bonham-Carter (1994). The described equation has a 0.5 cut-off value, defining < 0.5 and > 0.5 values as stable and unstable, respectively.

IV is the only method in this study using only unstable modelling pixels for training, whereas other methods also require the randomly generated stable pixels. Consequently, IV LSM is not derived for scenario S11 as IV only applies to LCFs with classes, i.e. not compatible with continuous stretched input data layers that define scenario S11. On the other hand, scenario S1 with 100 equal classes approximates scenario S11 and can be considered the closest alternative.

Introduced in the early work of Cox (1958), LR today corresponds to the most common statistical classification method used in LSAs (Reichenbach et al., 2018), as seen in the study cases from Rossi et al. (2010), Hemasinghe et al. (2018) and Bornaetxea et al. (2018), with practical code in a software application available in LAND-SUITE (Rossi et al., 2022). A linear fitting function (Z) between landslides for n number of conditioning variables is defined by the following equation:

$$Z = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_n X_n, \quad (7)$$

where b_0 is the intercept of the model, b_n the partial regression coefficients and X_n the conditioning variable. Lastly, a common application of the LR method includes converting resulting Z values to probabilistic output by applying Eq. (6)

as mentioned in Bornaetxea et al. (2018) and Merghadi et al. (2020).

NN is a two-stage regression or classification model that can handle multiple quantitative responses (Hastie et al., 2009). Among artificial NN methods, in this study feedforwarding was applied, indicating the flow of information exclusively in one direction. NN models are generally composed of simple circuits of nodes connected to each other (Merghadi et al., 2020), defined by three layers, i.e. input layer, hidden layer and output layer. The structure applied in this paper is as follows: input layer, first fully connected layer, rectified linear unit activation function, final fully connected layer, softmax function and output. The softmax function which was applied to the final fully connected layer is defined as

$$f(x_i) = \frac{e^{(x_i)}}{\sum_{j=1}^K e^{(x_j)}}, \quad (8)$$

where K is the number of classes of response variables in the final fully connected layer and x_i each input, resulting in probabilistic $[0,1]$ values, i.e. $f(x_i)$. Merghadi et al. (2020) summarize NN as a “black box” method, unable to interpret relations between input and out variables. However, the method is popular in LSAs (Reichenbach et al., 2018), with Habumugisha et al. (2022) even testing different settings to develop a method comparison, such as convolutional, deep and recurrent NN models. Furthermore, the successful applicability of different NN variations is found in the work of Lee (2007), Nefeslioglu et al. (2008) and Pascale et al. (2013), as well as in the LAND-SUITE software (Rossi et al., 2022).

Introducing the concept of bagging and random feature selection by Ho (1995) and Breiman (2001), RF is based on decision trees and provides an improvement over bagged trees (James et al., 2013). Bagging is a procedure introduced to reduce the variance of statistical learning methods, particularly useful for decision trees (James et al., 2013). Defining p as the number of total predictors and m as the number of predictors taken at each split, in bagging, decision trees are built using the following expression:

$$m = p, \quad (9)$$

compared to RF, where the relation is defined as

$$m = \sqrt{p}. \quad (10)$$

Namely, the algorithm does not consider most predictors in each decision tree in the RF method. As a result, RF tends to produce precise results and has an increasing popularity in LSAs from 2010 onward (Merghadi et al., 2020), as depicted in several papers such as Catani et al. (2013), Wang et al. (2021) and Sandić et al. (2023). Moreover, ensemble methods, including RF, are remarked to have excellent performance by Merghadi et al. (2020).

SVM is introduced by Cortes and Vapnik (1995) and Vapnik (1995), whereas James et al. (2013) describe SVM as an extension of a support vector classifier that can convert a linear classifier into one that automatically produces non-linear boundaries. Moreover, SVM can deal with linearly separable data and linearly non-separable data, whereas points that constrain the width of the margin, i.e. are closest to the optimal hyperplane, are called support vectors. To better classify most training observations, the support vector classifier allows for a certain number of observations on the wrong side of the hyperplane or the margin (James et al., 2013). The SVM uses kernel functions to transform originally non-separable data from two-dimensional into three-dimensional feature space (Ballabio and Sterlacchini, 2012), where the optimal hyperplane is constructed and later used to classify new data. Binary classification applied in this study (i.e. 0 being stable and 1 unstable pixels) results in better prediction efficiency than a one-class SVM application (Yao et al., 2008). Also, Eq. (6) is applied to convert numerical susceptibility values to probabilistic ranging $[0,1]$. Merghadi et al. (2020) identified SVM starting a continuous increase in usage in landslide susceptibility studies since 2010 (e.g. Yao et al., 2008; Pradhan, 2013; Kavzoglu et al., 2014) and characterized SVM as having above-average performance with moderately easy implementation.

3.3 Quantitative and qualitative analysis

A landslide training set containing unstable and stable modelling pixels was examined to determine model evaluation parameters, i.e. fitting performance by determining Cohen's κ and AUC for false alarm rate and hit rate metrics. Cohen's κ and AUC are often used statistical parameters in landslide susceptibility analyses to evaluate the model, e.g. Pourghasemi et al. (2021) and Tyagi et al. (2023). Concretely, Cohen's κ value interpretation was defined in Landis and Koch (1977) as (i) almost perfect (0.8–1), (ii) substantial (0.61–0.8), (iii) moderate (0.41–0.60), (iv) fair (0.21–0.40), (v) slight (0.00–0.20) or (vi) poor (< 0.00), and the equation to determine the metric is

$$\text{Cohen's } \kappa = \frac{2 \times (TP \times TN - (FN \times FP))}{(TP + FP) \times (FP + TN) + (TP + FN) \times (FN + TN)}, \quad (11)$$

where TP denotes true positives (i.e. landslide pixels classified as unstable), FN denotes false negatives (i.e. landslide pixels classified as stable), TN denotes true negatives (i.e. non-landslide pixels classified as stable) and FP denotes false positives (i.e. non-landslide pixels classified as unstable) (Gorsevski et al., 2006).

In this study, the susceptibility values were split into 100 classes with 0.01 susceptibility intervals, i.e. defining the classification thresholds used for AUC calculation. TP, FN, TN and FP pixels were determined using 0.5 as a probabilistic susceptibility cut-off value for each interval. Moreover,

when the receiver operating characteristic (ROC) curve was defined, the AUC was calculated to further estimate model fitting performance as in Rossi et al. (2010), Bornaetxea et al. (2018) and Wang et al. (2021), whereas the hit rate and false alarm rate values were defined by Fawcett (2006) as

$$\text{hit rate} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (12)$$

and

$$\text{false alarm rate} = \frac{\text{FP}}{\text{FP} + \text{TN}}. \quad (13)$$

AUC values closer to 1 indicate a perfect prediction, contrary to a 0.5 value corresponding to random prediction. An example of describing performance given by the AUC values is as follows: (i) < 0.7 for poor, (ii) 0.7–0.8 for fair, (iii) 0.8–0.9 for good and (iv) > 0.9 for excellent (Fressard et al., 2014).

Unstable validation pixels are used to derive a cumulative percentage study area and cumulative percentage landslide area, defining prediction performance as introduced by Chung and Fabri (1999, 2003). The metric is often used in bivariate LSA approach (e.g. van Westen et al., 2003; Sinčić et al., 2022b; Bernat Gazibara et al., 2023) and is defined as the success rate and the prediction rate for fitting and predictive performance, respectively.

The standard deviation (SD) maps for susceptibility values are implemented to measure each scenario's stability and method's stability, resulting in qualitatively detecting uncertainty zones. Namely, 11 SD maps are developed for each of the 11 scenarios by observing the 5 applied methods in each scenario, i.e. to measure deviations among the methods. Furthermore, 5 SD maps describing criteria are defined by observing 11 scenarios applied for each method, i.e. to measure deviations among scenarios. Considering the probabilistic result of each LSM from 0.0 to 1.0 SD, classes are determined with 0.1 threshold intervals to observe their presence and spatial distribution in the study area. Qualitatively, all classified SD maps are presented to illustrate spatial distribution, which is also measured quantitatively by determining the percentage area of each SD class in every derived map.

By observing all unstable pixels, AUC values were determined for the cumulative percentage study area and cumulative percentage landslide area curve. The latter presents AUC for LSM classification, whereas all unstable pixels are the sum of unstable pixels used for fitting performance (model evaluation) and predictive performance (validation).

Observing both stable and unstable pixels for model evaluation by Cohen's κ , false alarm rate and hit rate AUC metrics were chosen due to their equal presence in the model training. On the other hand, measuring only unstable pixels for validation and classification by developing a cumulative percentage study area and cumulative percentage landslide area curve was selected to emphasize landslide occurrence exclusively for validation and classification without considering stable pixels. It should be noted that all AUC values

presented in this paper, i.e. fitting, predictive and classification performance, are expressed in rates, meaning 100 and 50 correlate to 1.0 and 0.5, respectively.

For uniform comparison, LSMs were classified in this paper according to probabilistic susceptibility values by using cut-off values from Bornaetxea et al. (2018), resulting in five susceptibility zones: (i) very low (0.0–0.2), (ii) low (0.2–0.45), (iii) medium (0.45–0.55), (iv) high (0.55–0.8) and (v) very high (0.8–1.0). Quantitatively, for every 54 LSMs, the class area size is observed, including landslide presence in each class. Furthermore, classified LSMs are qualitatively illustrated with training and validation landslides to observe spatial distribution, including close-up views by overlapping them on high-resolution hillshade maps to investigate spatial accuracy, robustness and pixelization degree.

4 Results

4.1 Landslide conditioning factors

In this section, differences between the 11 scenarios described for classifying continuous LCFs are explained on elevation LCF maps, quantitatively (Table 4, Fig. 4) and qualitatively (Fig. 5). Moreover, a general overview of all LCFs with the number of classes through 11 scenarios is presented in Table 3, as well as class area distribution in Fig. 7. Following the described methodology, all LCFs in S1 and S2 scenarios keep a rather stable number of classes at a 90–101 and 46–51 range, respectively. The shortage of classes in slope and terrain wetness LCFs is due to missing certain values in the original stretched raster. As intended, with a fixed number of 10 classes, S8 to S10 scenarios have identical numbers of classes for all LCFs. Regarding the heuristic S3 scenario, the number of classes varies from 7 classes at elevation to 13 classes in proximity to faults. A relatively low number of classes are depicted in S4 to S6 scenarios where continuous rasters have 2 to 6 classes, whereas vector line (buffer) LCFs show a larger span from 2 to 25 classes. Namely, proximity to geological contact has significantly more classes than other LCFs, followed by proximity to streams and faults. An exception is proximity to drainage network LCF, having two to four classes in S4–S6 scenarios. As intended, the number of classes in each LCF increases from scenario S4 to scenario S6. The S7 scenario has the lowest number of classes in all LCFs, the highest number of classes being four at proximity to geological contact and the lowest two classes in proximity to drainage network, streams and faults LCF. Edge raster values and edge buffer values defined for the S11 scenario representing input data layer files are depicted in Table 3.

To obtain 100 classes for the S1 scenario and 50 classes for the S2 scenario, equal interval values in the elevation LCF are 3.2 and 6.4 m a.s.l., respectively. Furthermore, the heuristic interval for the scenario S3 is chosen at 50 m a.s.l. Scenarios S4 to S6 are defined by C_{st} curves based on scenario S1

Table 3. Distribution of the number of classes (or values) for 7 continuous LCFs in 11 scenarios.

Number of classes in a continuous LCF (<i>N</i>)								
	Elevation	Slope	Terrain wetness	Prox. to streams	Prox. to drainage network	Prox. to Geological contact	Prox. to fault	
S1	100	91	90	101	100	99	99	
S2	50	46	46	51	50	50	50	
S3	7	10	12	12	8	12	13	
S4	3	2	3	8	2	13	6	
S5	4	3	3	11	3	20	9	
S6	6	4	4	14	4	25	12	
S7	3	3	3	2	2	4	2	
S8	10	10	10	10	10	10	10	
S9	10	10	10	10	10	10	10	
S10	10	10	10	10	10	10	10	
Edge raster values				Edge buffer values (m)				
S11	From	122.15	0.71	2.14	12	1.7	12.5	13
	To	439.94	80.26	19.90	1212	170	1237.5	1287

Table 4. Elevation reclassification cut-off value distribution in 10 scenarios.

Class (<i>N</i>)	Elevation (119.0–436.9 m a.s.l.) reclassification cut-off values									
	S1	S2	S3	S4	S5	S6	S7	S8	S9	S10
1	122.2	125.3	150.0	153.9	153.9	138.1	157.1	152.5	150.0	148.9
2	125.3	131.7	200.0	249.3	179.4	153.9	246.2	178.6	167.4	171.1
3	128.5	138.1	250.0	436.9	249.3	192.1	436.9	202.2	183.6	187.6
4	131.7	144.4	300.0	–	436.9	249.3	–	224.5	198.5	199.7
5	134.9	150.8	350.0		–	274.8		245.7	213.4	216.2
6	138.1	157.1	400.0			436.9		266.8	229.5	238.4
7	141.2	163.5	> 400			–		290.4	145.7	268.3
8	144.4	169.8	–					316.5	261.8	308.8
9	147.6	176.2						351.2	287.9	363.3
10	150.8	182.6						436.9	436.9	436.9
11	153.9	188.9						–	–	–
...	...	...								
49	274.8	430.6								
50	278.0	436.9								
51	281.1	–								
...	...									
99	433.8									
100	436.9									

classes, as illustrated in Fig. 4. For elevation LCF, the curve is defined by two negative segments in which C_{st} values reach roughly -9 and -7 minimum values, respectively. The positive segment has a steep increase at class 12, leading to a C_{st} value of 16 at class 24, followed by a decrease in value reaching 0 in class 41, followed by below 0 values. Lastly, class 63 and higher classes have a C_{st} value of 0, indicating no landslide presence. Figure 4 illustrates that when defining scenario S7, the relation in expected and mapped landslides changes twice, defining three classes. Namely, there

are more expected landslides in classes 1 to 13 and 41 to 63, whereas from 13 to 41 there are more mapped than expected landslides. Regardless of having an equal number of classes, the reclassification cut-off values significantly differ for scenarios S8, S9 and S10, as depicted in the example on elevation (Table 4). On the other hand, despite having a different methodology, scenarios S4 and S7 show minimal differences in the cut-off values for the elevation LCF.

Considering class area distribution illustrated in Fig. 5 for the elevation LCF map, scenarios S1 and S11 (Fig. 5a and

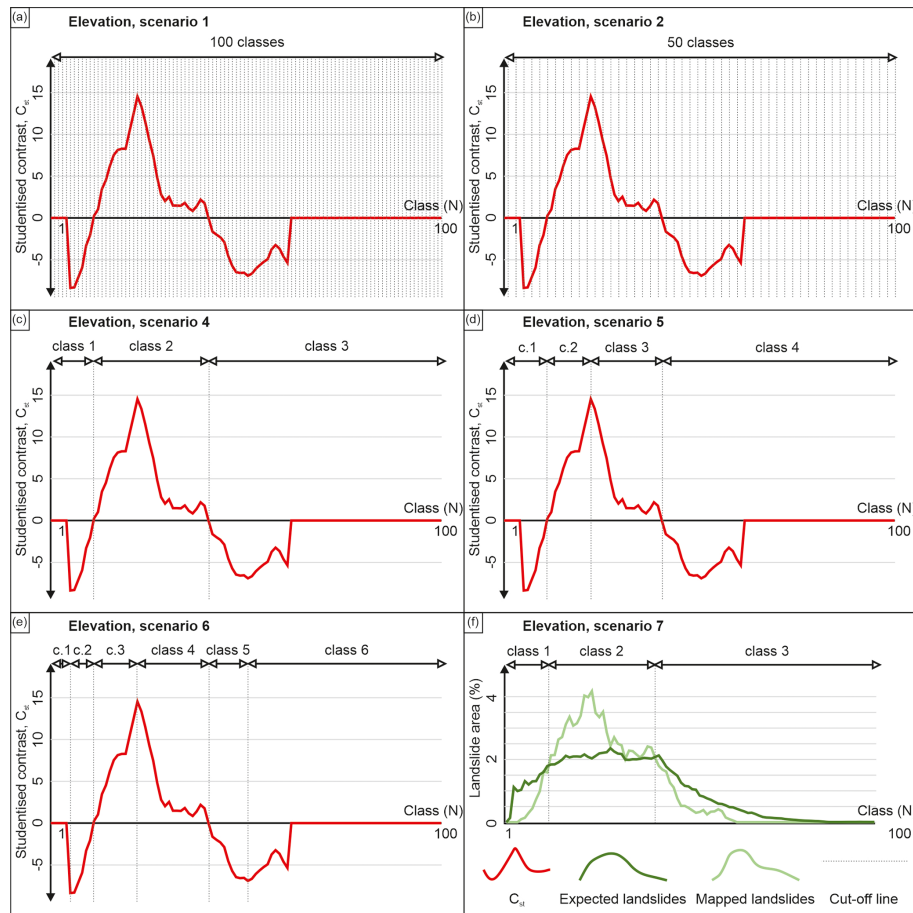


Figure 4. Elevation landslide conditioning factor class cut-off values determined for scenarios S1 (a), S2 (b), S4 (c), S5 (d) and S6 (e) based on studentized curve and for scenario S7 (f) based on landslide density curve.

k, respectively) visually show little difference as intended, S1 being a simulation, i.e. approximation of S11, depicting a continuous change from low to high altitudes. With 50 instead of 100 classes, the S2 scenario depicts a somewhat rougher transition through altitude classes (Fig. 5b). Starting at three classes in scenario S4 (Fig. 5d), the second class defined by a positive C_{st} curve trend splits into two, making up four classes in the S5 scenario (Fig. 5e).

Finally, two classes defined by negative C_{st} curve trends split individually into two additional classes, making up a final count of six classes for the S6 scenario (Fig. 5f). The S7 scenario is visually identical to S4 due to almost identical cut-off values depicted in Table 3. Scenarios S8, S9, and S10 (Fig. 5h, i, j) visually represent moderate differences, mainly visible in the areas of highest altitudes, i.e. the northernmost and central parts of the study area defined by classes 9 and 10. The latter classes are predominantly expressed and prevail in scenario S9 compared to scenarios S8 and S10.

Due to significant class number differences, scenarios S1 and S2 have drastically less class area size than other scenarios (Fig. 6) in all applied LCFs. Concretely, in these two sce-

narios, the maximum class area is found to be around 10 % in proximity to geological contact, proximity to drainage network and terrain wetness LCFs. On the contrary, in the elevation LCF, the class area size does not exceed more than 5 % in any class. After roughly the 30th class in scenario S2, the class area size does not exceed more than 1 %. Similarly, in scenario S1, the classes in the interval from 60 to 100 usually have nearly 0 % class area size. In scenarios S1 and S2, elevation, slope, proximity to all streams and terrain wetness LCFs have an increasing trend in class area size, reaching a maximum followed by a decreased trend. Proximity to geological contact, proximity to fault and proximity to drainage network show only a decreasing trend in scenarios S1 and S2, an exception being proximity to drainage network in scenario S1 with a short exchange of increasing and decreasing trends. For scenarios S3 to S10, all LCFs have around 10 to 14 classes, except proximity to geological contact, which has 24 classes. In any case, for all LCFs in scenarios S3 to S7, most area is contained in the first few classes, e.g. up to class 3 or 4. In elevation, slope and proximity to geological contact LCFs, the maximum class area size is roughly 60 %,

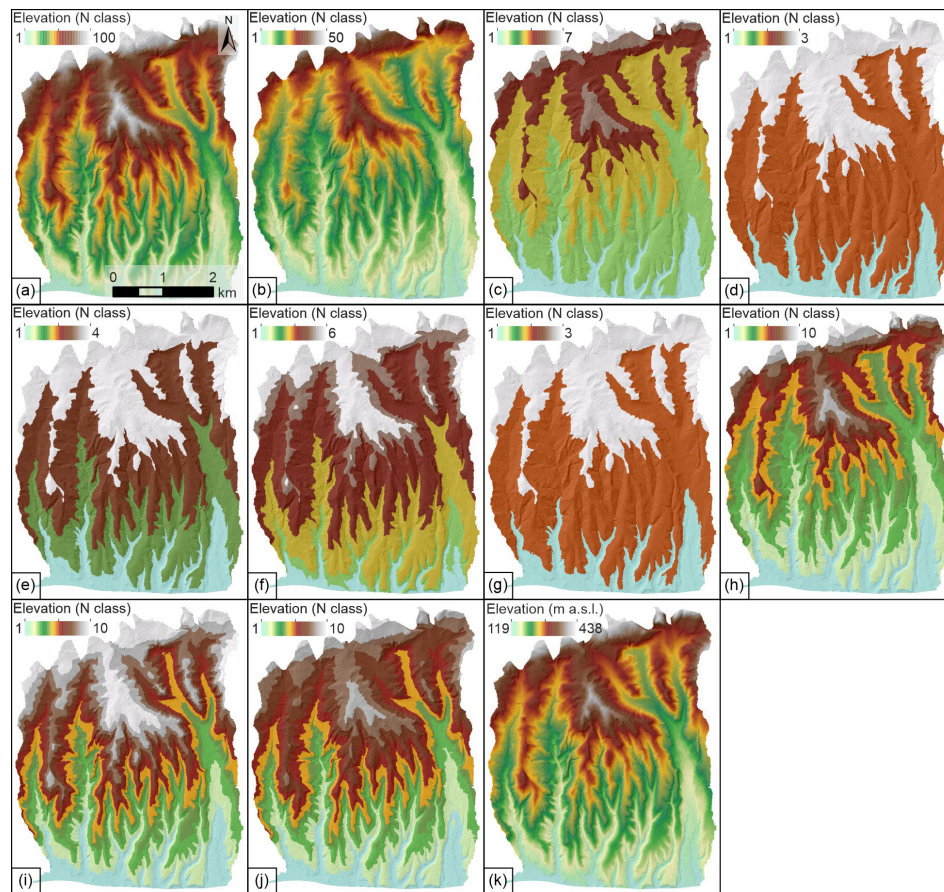


Figure 5. Maps of the spatial distribution of elevation landslide conditioning factor classes for 11 scenarios.

compared to roughly 80 % class area size in terrain wetness, proximity to faults, drainage network and all streams' LCFs. Generally, in scenarios S3 to S7 the class area size tends to drastically change from class to class, unlike scenarios S8 to S10, which depict slight changes. As methodologically defined, scenario S9 contains the 10 % class area trend through all scenarios, whereas scenario S9 has significantly less class area size in classes 7 to 10, compared to classes 1 to 6.

4.2 Landslide susceptibility model evaluation

Model evaluation parameters consider both stable and unstable pixels used for training the model and expose their fitting performance as illustrated in Fig. 7. Namely, Cohen's κ index has substantial agreement in all 11 scenarios for IV, LR, NN and SVM methods, ranging roughly from 0.6 to 0.7. The latter methods follow the same trend in all scenarios, with NN showing the best performance (0.7), followed by LR (0.65) and, finally, SVM and IV (0.6). RF method showed perfect agreement with 1.0 or near 1.0 values in S1, S2, S3, S6, S8, S9, S10 and S11 scenarios and drastically lower (i.e. 0.7) in the S7 scenario. Scenarios S4 and S5 in RF performed slightly worse but still with almost perfect agreement, having

Cohen's κ values of 0.88 and 0.96, respectively. Regarding false alarm rate and hit rate AUC values, the RF method follows the behaviour of Cohen's κ values, having excellent performance and nearly identical trends as Cohen's κ through 11 scenarios. IV, LR and SVM methods depict nearly identical AUC values in all 11 scenarios, ranging from roughly 85.5 to 88.5. In scenarios S4, S5, S6, and S7, the latter methods show values closer to 85.5, whereas the IV method has somewhat poorer results in scenarios S8, S9 and S10. Having the most stable AUC values in all 11 scenarios, the NN method performed excellent with AUC values ranging from 90.5 to 92, which is moderately better than IV, LR and SVM yet significantly worse than perfect agreements in the RF method.

4.3 Landslide susceptibility model validation

Unlike results in model evaluation, where studied metrics showed significant differences between five methods, by validating the LSMs, we notice fewer differences with clustered AUC values between methods (Fig. 8). Measuring the predictive performance of LSMs on an independent landslide dataset results in scenarios S1, S2, S3, S8, S9, S10 and S11 defined by 10 or more classes in LCFs as better solutions. In

the latter scenarios, the AUC values range from roughly 86.5 to 88.5 for all methods, compared to a decrease in scenarios S4, S5, S6 and S7, where AUC values reach approximately 84 to 87.5. Generally, methods are similar in all scenarios and deviate minimally in scenarios S1, S2, S3, S8, S9, S10 and S11 (approx. 1.5) and moderately in scenarios S4 to S7. Concretely, S11 has proven as a scenario with minimum AUC deviations (< 1) between methods, compared to the highest deviations in scenarios S4, S5, S6 and S7 (approx. 3). LR has the best prediction performance in all scenarios with the highest values reaching up to 88.5. Furthermore, LR AUC values are higher by up to 3 points compared to RF, which has the least predictive performance, except in scenarios S1 and S2. Having lower AUC values in most scenarios than LR, NN, and SVM, the IV value method nearly has the best predictive performance in the S7 scenario. On the contrary, LR, NN and SVM outperform IV in scenarios S8, S9 and S10. Lastly, IV, NN and SVM have similar predictive performances in scenarios S4, S5 and S6, outperforming RF significantly.

4.4 Landslide susceptibility model classification

AUC classification values, measuring all unstable pixels as a combination of fitting and predictive performance, show very high values (Fig. 9). IV, LR, NN and SVM methods show low deviations from scenario to scenario, mainly clustered with AUC values in the range of 84.5 to 89. RF method outperforms the other methods by having AUC values ranging lowest 89 in S7 up to > 93 in S11 and S1. After RF, NN proved to be an alternative, having slightly higher AUC values than IV, LR and SVM, which differentiate minimally in each observed scenario. All methods show lower AUC values in scenarios S4, S5 and S6, with a minimum in scenario S7.

In scenarios S1, S2, S3, S8, S9, S10 and S11, all methods show AUC values higher than 86. Lastly, the RF method is the only method significantly influenced by the number of LCF classes, with significantly lower AUC values in scenarios S4, S5, S6 and S7.

To identify classification parameters, derived LSMs are firstly classified in zones according to probabilistic susceptibility values as follows: (i) very low (0.0–0.2), (ii) low (0.2–0.45), (iii) medium (0.45–0.55), (iv) high (0.55–0.8) and (v) very high (0.8–1.0). Generally observing the susceptibility zones, noticeable differences are distinguishable among methods rather than scenarios (Fig. 10). All methods show $< 10\%$ class area in a very high zone, with RF, SVM and IV often having $< 5\%$. All 54 LSMs have a medium-susceptibility zone smaller than 10% , with RF having a minimum as small as 2% . Moreover, the RF method has the smallest area in zones of high and very high susceptibility. Class area percentage changes are most significant in very low and low classes, whereas very high, high and medium classes show low differences in all methods and most scenarios. IV and SVM methods tend to have fewer differences

between low and very low classes, i.e. having almost equal areas in both zones. RF has the most percentage area in the very low class, followed by NN and LR. However, RF, NN and LR have nearly similar class area sizes in the low class. Observing very low and low zones as one, IV, LR, NN and SVM show a lot of similarities, unlike RF, which has exceptionally high values, reaching $> 85\%$ area for the cumulative area of the two classes in most scenarios. Similarly, very high and high zones contain less than 10% of the map area in the RF method, whereas the values reach around 20% for the other methods. In other words, the most significant differences are depicted individually in the relation between very low and low zone and very high and high zone. For the latter, the RF method favours having an extremely large very low zone and an extremely small very high zone. On the contrary, IV and SVM tend to have rather large low and high zones. NN has around a 10% smaller very low zone compared to RF, followed by LR 10% smaller than NN. Lastly, IV and SVM have around 40% of the map area in the very low zone.

The sums of landslide presence in zones of very high and high susceptibility are extremely similar in all 54 LSMs (Fig. 11). However, observing the two zones individually, noticeable differences are depicted. Namely, LR, NN and RF methods have higher landslide presence percentages in very high zones resulting in lower presence in high zones, whereas on the contrary, IV and SVM methods have less landslide presence in very high zones and higher landslide presence in the high zones. Concretely, IV and SVM have around or less than 35% landslide presence in very high zones, compared to other methods, which often have $> 50\%$. The minimum is in the IV method scenario S7 and SVM method scenarios S4 and S5 at around 20% or less landslide presence in very high susceptibility zones. RF has the maximum landslide presence in very high zones in each scenario, valuing on average 62% , followed by NN (approx. 56%) and LR (approx. 51%). Observing landslide presence in very low, low and medium zones, minimal differences are noticeable, an exception being IV and SVM methods, which stand out with extremely low landslide presence in very low susceptibility zones. Moreover, the RF method has the highest number of landslides present in very high susceptibility zones in each scenario and the highest landslide presence in very low zones, reaching up to 5% . Interestingly, the average landslide presence area in the medium-susceptibility zone is 5.5% , i.e. rather low.

5 Discussion

5.1 Discussion regarding LCF classification criteria

Considering the quantitative results in Sect. 4.2, 4.3 and 4.4 from a perspective of 11 scenarios for continuous LCF classification, moderate differences are found in fitting perfor-

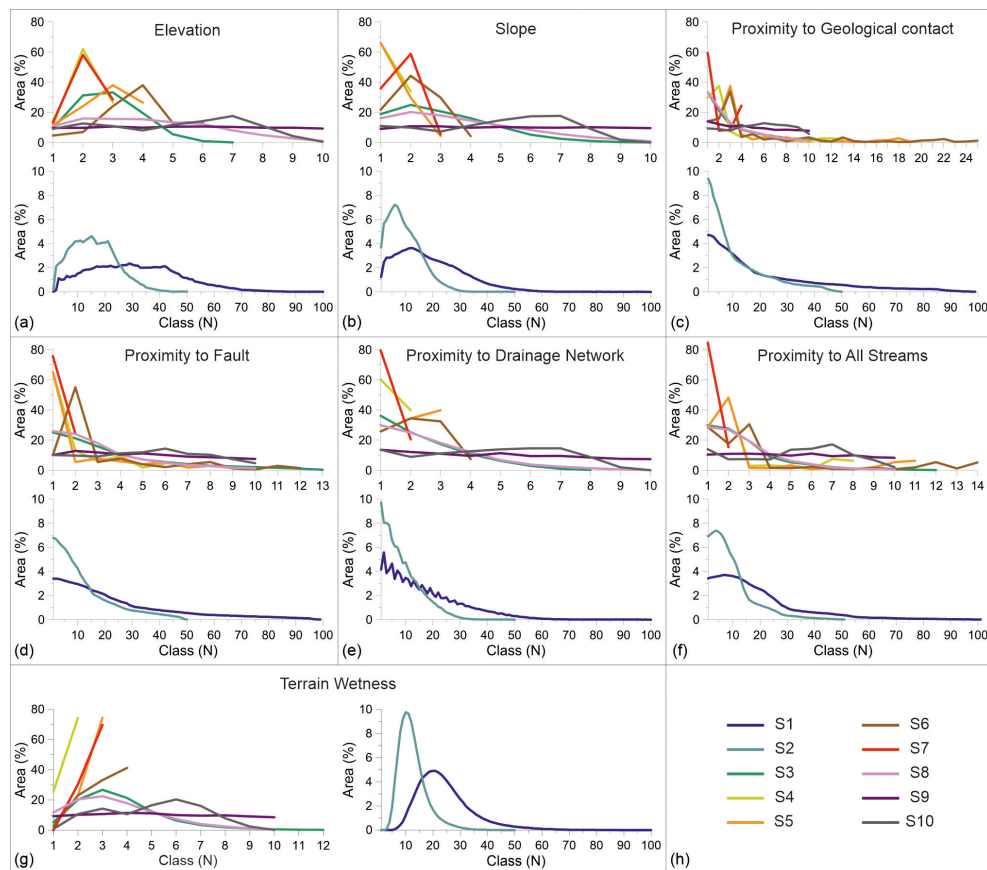


Figure 6. Graphs of class area distribution for continuous landslide conditioning factors in scenarios S1 to S10.

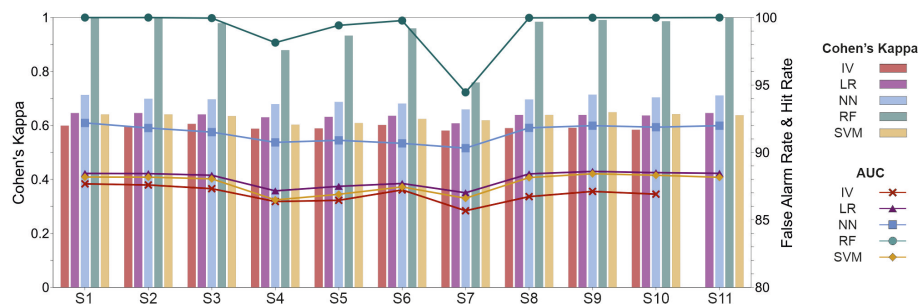


Figure 7. AUC and Cohen's κ quantitative parameters describing fitting performance of the 54 derived landslide susceptibility models.

mance and low differences in predictive and classification performance. However, the noted differences occur repeatedly in scenarios S4 to S7 and are represented by poorer performance in all studied metrics. A likely reason behind the results is too few classes in the LCFs. Concretely, often showing extraordinary results, scenario S7 has four or fewer classes in the continuous LCFs. On the contrary, opposite classification criteria with a large number of classes, such as scenarios S1 to S3 and S8 to S10, result in satisfactory results in all scenarios and are highly similar. That implies that the

closer the classification is to the continuous behaviour, the higher the LSM fitting and predictive performance.

The 11 classified SD maps are illustrated in Fig. S1 in the Supplement for each LCF classification scenario, representing differences in susceptibility values between the 5 applied methods. Namely, classes of higher SD values indicate higher uncertainties in the LSMs. The SD maps are visually similar, indicating low differences in susceptibility values from scenario to scenario. The 0.0–0.1 SD class is most represented in all scenarios, with area presence ranging from minimum approx. 50 % in scenario S5 to > 65 % in

scenario S7. The 0.1–0.2 SD class and 0.2–0.3 SD class value are equally represented in all scenarios, their area percentage being approx. 35 % and 5 %–10 %, respectively. An exception is S7, where the 0.2–0.3 SD class has less than 5 %, and in S11, where the 0.1–0.2 SD class has < 30 % area. Qualitatively observing spatial distribution, SD class distribution does not differ significantly from scenario to scenario. An exception is scenario S11, where the 0.0–0.1 SD zone has clustered areas, unlike dispersed and likely pixelized, i.e. distributed into smaller zones as visible in other scenarios. Surprisingly, scenarios S7 and S11, which are methodologically opposite considering the number of classes, have a high > 60 % map area in the 0.0–0.1 SD class. Moreover, S7 stands out by having the least area in the 0.2–0.5 SD range, i.e. having > 95 % of the area in SD values < 0.2. If they are present, 0.3–0.4 and 0.4–0.5 SD classes appear in the same locations in all scenarios. Scenario S11 stands out with the most area in the 0.3–0.4 SD class but still < 1 % of the study area. Generally, the 0.3–0.5 SD range is least present and neglectable in all scenarios when compared to other classes. By observing the 11 close-up views (Fig. S2), the differences can be observed by comparing two groups, defined by the number of classes in LCFs used to derive a LSM, and scenario S7 as a standalone scenario. Namely, the first group consists of scenarios with fewer classes in LCFs, i.e. S4, S5 and S6. Scenarios with mainly 10 or more classes comprise the second group, i.e. scenarios S1, S2, S3, S8, S9, S10 and S11. The first group is less pixelized, with a low number of standalone pixels and highly expressed susceptibility zones, unlike the second group, which has heterogeneous and pixelized small zones.

Considering the above, noticeable differences between the 11 scenarios can be traced to the number of classes present in LCFs, not to the method applied to create the classes. Consequently, scenarios S4, S5, S6 and S7 can be outlined as the less favourable ones. The latter scenarios show poorer fitting, predictive and classification performance results. Moreover, qualitatively, they have robust susceptibility zones, i.e. losing susceptibility information on a large scale, which is essential for high spatial accuracy and applicability in spatial planning. On the other hand, robustness can ease the classification process to determine susceptibility zones if the values related to robustness are spatially accurate. Also, compared to other scenarios, the classification process is relatively time-consuming for scenarios S4–S7. Commonly used and straightforward heuristic, NB, Q and GI classification criteria applied in scenarios S3, S8, S9 and S10 represent satisfactory options, considering the number of classes defined. With most researchers using around 10 classes in different methods, considering results from scenarios S1 and S2, a significantly larger number of classes can also be applied to reach reliable results.

Lastly, scenario S11 can be suggested as a uniform approach to performing landslide susceptibility modelling. Scenario S11 shares excellent performance with several other

scenarios. However, it stands out as the most consistent considering predictive performance with low differences between the five used methods. Similarly, it is the least pixelized in the SD map and has > 60 % of the map area in the 0.0–0.1 SD zone, proving low uncertainties between methods. Moreover, directly applying stretched rasters removes a step in landslide susceptibility modelling often done by researchers, enabling more technical simplicity and reducing the time needed due to avoiding the classification of the input data layers. Considering the bivariate approach, where applying stretched rasters such as in scenario S11 without classes is not possible, we suggest scenario S1 as the closest alternative, i.e. the optimal solution.

It should be stated that the set of prepared LCFs in this study is appropriate for achieving excellent performance for deriving LSMs. Moreover, the study aimed to test the 11 criteria on such a set; however, the relevance and/or importance of each LCF in the LSA was not individually determined. Hence, some of the LCFs may be of poor relation to LSA quality, and classifying them differently is irrelevant. In other words, the 11 classification scenarios would likely present more differences on a limited number of highly important LCFs. The latter is addressed by applying the classification criteria uniformly for all continuous LCFs and proposing a uniform solution for a relevant set of LCFs, not limited only to the most significant LCFs. Similarly, applying a non-representative landslide inventory map would likely cause substantial deviations to the results presented in this study. Future work considering this topic could lead to the idea presented by Yan et al. (2019), where LCFs in all scenarios are mutually combined and tested to develop an optimal solution, which would lead to a tremendous number of LSMs in this paper and was therefore discarded, rather emphasizing the investigation into developing a uniform approach.

5.2 Discussion regarding different statistical methods

Unlike low differences between the 11 scenarios, the 5 applied methods showed substantial deviations from each other in different quantitative and qualitative performances as seen in Sect. 4.2, 4.3, 4.4. and 4.6. Namely, RF outperforms other methods significantly in model evaluation, followed by the NN method, whereas IV, LR and SVM differ minimally and show the lowest Cohen's κ and AUC values. Furthermore, predictive performance favours LR, whereas RF performed poorly, leaving IV, NN and SVM in between. Lastly, classification metrics favour RF due to perfect results in model evaluation. Moreover, unlike IV and SVM, RF has the most landslide presence in a very high susceptibility zone. On the other hand, RF also excels in the area presence of a very low susceptibility zone, having it by far the largest compared to other methods while keeping a low percentage of landslide presence.

Examining SD maps of susceptibility values for each method individually (Fig. S3), uncertainty through 11 sce-

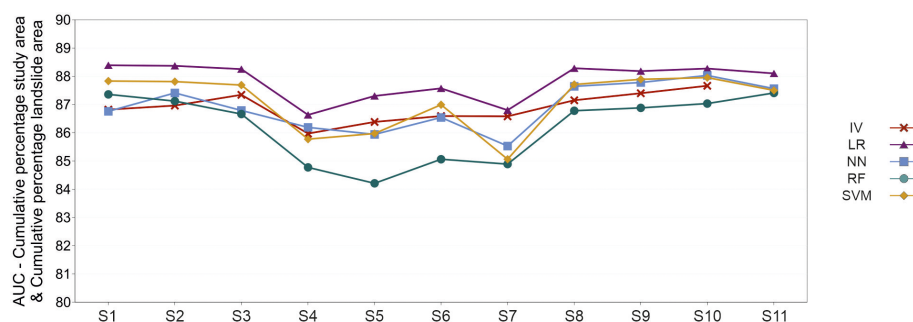


Figure 8. AUC quantitative parameter describing predictive performance of 54 derived landslide susceptibility models.

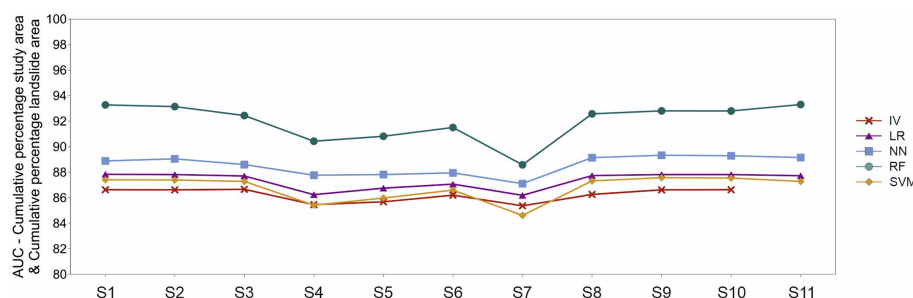


Figure 9. AUC quantitative parameter describing classification performance of 54 derived landslide susceptibility models.

narios is measured for each method. Surprisingly, the IV, LR and SVM methods show minimal SD values for most of the study area. Namely, the SVM method has 99 % of the study area in the 0.0–0.1 SD zone, whereas IV and LR methods have < 5 % of the study area in the 0.1–0.2 SD zone. In addition, SVM and IV methods display noticeable uncertainty in the south and the southeastern part of the study area, whereas the LR method has them distributed throughout the study area. Conversely, RF and NN methods have significantly less area in the lowest SD class.

The quantitative and qualitative analysis applied to compare 54 derived LSMs showed the advantages and disadvantages of the 5 used methods, justifying the broader approach in this study. In model evaluation, RF often classified the training pixels perfectly, whereas the predictive performance was significantly lower, leaving an open question to investigate further. Nonetheless, a rigorous approach with an unbiased training dataset and 50 % to 50 % landslide inventory splitting still enabled high predictive performance. Moreover, RF clusters the values close to 0 and 1 susceptibility, leaving low- and high-susceptibility zones relatively small. However, considering landslide presence in the RF susceptibility zones, extreme landslide density in the very high zone can be observed. Interestingly, the IV method can be regarded as equally usable despite the modelling using only unstable pixels. Moreover, IV was poorly influenced by LCFs having few classes, similar to LR and SVM. All three methods showed extremely low SD values, indicating that all 11 scenarios are equally suitable for their application con-

sidering the resulting susceptibility values. On the contrary, NN and RF are affected by the lack of more classes in LCFs (i.e. should not be used recklessly). Considering the classification parameters, low landslide presence was detected in very high susceptibility zones for both IV and SVM. Despite using 100 classes for LCFs in scenario S1, SVM was the only method that had well-characterized susceptibility zones in close-up views, unlike other methods, which were rather pixellized (Fig. S4). NN and even to a greater extent RF seem to outperform IV, LR and SVM considering quantitative metrics. However, the qualitative approach depicted certain drawbacks which should be noted. On the other hand, certain advantages were noted in the IV and SVM methods, despite their poorer performance. Consequently, we suggest using LR as a starting point, being the most stable with the least extraordinary results, whereas for IV, NN, RF and SVM methods, the classification criteria can drastically influence the LSM quality; i.e. optimization should be applied as well as considering the LSA's purpose.

Considering that in many LSA comparison papers the mentioned model evaluation and/or validation metrics reach satisfactory results in the tested scenarios, in this study we emphasized the classification parameters for the following reasons:

- i. Satisfactory quantitative metrics results are expected in model evaluation (fitting performance) and LSM validation (predictive performance) due to complete input data of high spatial accuracy.

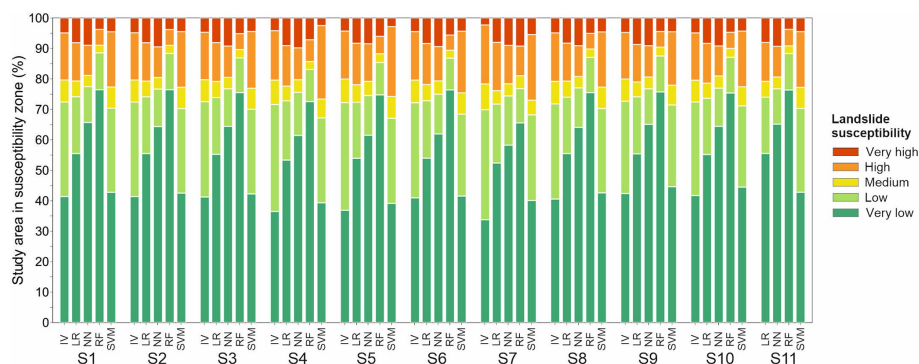


Figure 10. Landslide susceptibility zone area distribution graph for 54 derived landslide susceptibility models.

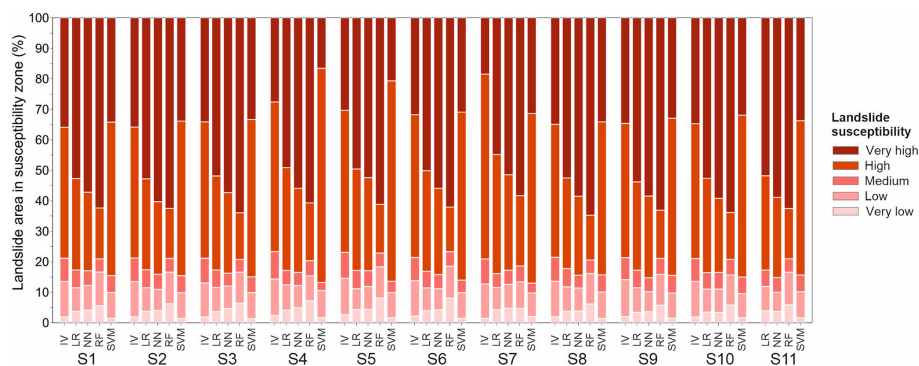


Figure 11. Landslide area distribution in a landslide susceptibility zone graph for 54 derived landslide susceptibility models.

- ii. Classification parameters take into consideration all identified unstable pixels from the complete landslide inventory.
- iii. Spatial distribution of classified LSMs allows insight regarding application in the spatial planning system, combined with susceptibility zone relation to elements at risk (e.g. buildings, roads).
- iv. Classification parameters define relations of all identified landslide presence in each defined susceptibility class, essential for application and for defining landslide protocols for LSA.

Putting the differences aside, all methods resulted in objectively excellent LSMs and confirmed the previous LSA done by Bernat Gazibara et al. (2023) in the study area. However, modelling on a large scale and aiming towards application in spatial planning, using a qualitative approach has proven to be of great significance by providing new insights into landslide susceptibility modelling. Without the latter, a complete picture of the used methods could not be acquired. LSA is often carried out to optimize certain parameters (e.g. method, mapping unit, inventory type) by developing dozens of LSMs which are further compared. When landslide susceptibility modelling reaches a point where most LSMs have a high fitting and predictive performance, we argue that ad-

ditional metrics (e.g. qualitative approach and classification parameters) are needed, depending on the scope of the LSA.

6 Conclusions

The presented work is based on geomorphologically significant and spatially accurate thematic data for LSMs, making it one of the rarest pieces of scientific research to implement the recommendations for LSA by Reichenbach et al. (2018). More precisely, Reichenbach et al. (2018) identified that LSA researchers are more eager to experiment with modelling techniques than to focus on acquiring relevant thematic and landslide data. Moreover, existing literature deals with preparing thematic input data in a general way and lacks a uniform approach for continuous LCF classification. This work represents a comprehensive analysis of different classification criteria for continuous LCFs and applies the most commonly used statistical methods evaluated by several quantitative and qualitative metrics to obtain representative conclusions and novelties regarding large-scale LSMs.

We defined 11 classification criteria for the 7 continuous LCFs applied in the landslide susceptibility modelling: elevation, slope, terrain wetness, proximity to geological contact, faults, streams and drainage network. Also, two criteria were applied to the categorical aspect LCF, whereas the re-

maintaining categorical LCFs, i.e. lithology (rock/type) and land use, remained constant in all scenarios. Scenarios were defined based on the classification criteria applied in the literature review and/or modifications of the criteria and vary by stretched values, partially stretched classes, the heuristic approach, classification based on studentized contrast and landslide presence, and commonly used classification (such as natural neighbour, quantiles and geometrical intervals). By applying 5 statistical methods (i.e. IV, LR, NN, RF and SVM), 54 LSMs were derived, providing comprehensive insight into the research issue. Also, a strict landslide sampling approach was used, with 50 % of landslide polygons used for training and the other 50 % for validation. Quantitative metrics to measure each derived LSM include false alarm rate and hit rate AUC, Cohen's κ model evaluation), AUC prediction rate values, SD class percentage area (model validation), AUC, susceptibility zone percentage area, and landslide area presence in susceptibility zones (model classification). The complete set of studied metrics proved to be necessary, pointing out the advantages and disadvantages of used methods and classification scenarios. Considering general performance, all the derived LSMs proved to be reliable considering usually studied metrics, i.e. AUC for fitting and predictive performance.

The analysis presented in the paper resulted in a set of large-scale LSMs created from representative and spatially accurate input data, including detailed lidar-based landslide inventory and high-resolution thematic data. Based on 54 reliable LSMs, specific conclusions with practical application could be made. The landslide susceptibility modelling was done on 21 km² in the Podsljeme area, using a 5 m pixel as a mapping unit suitable for large-scale LSA. The study area is characterized by small and shallow landslides and can be representative of similar environments with high-quality input data for landslide susceptibility modelling. The main conclusions and novelties derived from the presented comprehensive large-scale landslide susceptibility analysis are following:

- (i.) Due to using relevant input data with sufficient spatial accuracy, landslide susceptibility modelling performed by any statistical method or any LCF classification scenario in this paper resulted in a highly reliable LSM.
- (ii.) Any of the suggested scenarios to classify continuous LCFs is appropriate if it resulted in roughly 10 or unlimitedly more classes in the LCF, suggesting the higher importance of the number of classes in LCFs than the method of how the classes were created. In other words, a low number of classes in LCFs, such as < 5, are likely to perform poorly and should be avoided.
- (iii.) Applying input data layers as stretched rasters (scenario S11) and line vectors as buffers with > 10 buffer zones simplifies the susceptibility modelling process and provides a uniform solution to preparing LCFs.

- (iv.) Quantitative classification parameters and uncertainty metrics, as well as qualitative comparison (e.g. close-up views to verify spatial accuracy) applied in this study, are necessary metrics to evaluate optimal settings for large-scale landslide susceptibility modelling as they depict LSM characteristics unidentified by standard quantitative fitting and/or prediction metrics.
- (v.) Optimal method selection remains an open question and generally should be considered regarding the final applicability of the LSA, whereas in this study the LR method presents the most stable and representative option, and the RF method offers optimal performance when appropriately applied, achieving far better performance.
- (vi.) NN and RF methods are more sensitive to the LCF classification criteria than IV, LR and SVM.

However, using the same strategy on a different scale or with an incomplete dataset (either irrelevant LCFs or a non-representative landslide inventory) remains an open question which requires additional research. Moreover, comprehensive comparison using a variety of parameters provides new insight into derived LSMs. For LSM validation, quantitative metrics are not representative due to the lack of a spatial accuracy component, which is especially important in large-scale LSA and was proven in this paper. The studied qualitative metrics are maps of SD classes for model validation, whereas the spatial distribution of susceptibility zones and landslides in full extent and additionally with elements at risk in close-up views are used to describe classified LSMs. Presented results and performed validation of LSMs contribute to the preparation of recommendations, evaluation, and use of LSMs and associated terrain zonation proposed by Reichenbach et al. (2018). The recommended qualitative metrics enable the verification of the LSMs for practical application in spatial planning at the local level because they enable comparison with real environmental conditions and elements at risk that are visible on close-up views and high-resolution hillshade maps.

Data availability. Data supporting the research are obtainable from the corresponding author upon reasonable request.

Supplement. The supplement related to this article is available online at: <https://doi.org/10.5194/nhess-25-183-2025-supplement>.

Author contributions. MS and SBG conceptualized the draft study design, followed by finalization of the research concept with MR. MS carried out the analyses and wrote the manuscript under SBG's supervision. All authors commented on and analysed the results and were included in the discussion, whereas SBG, MR and SMA fully revised the manuscript.

Competing interests. The contact author has declared that none of the authors has any competing interests.

Disclaimer. Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

Acknowledgements. The authors would like to thank Paola Reichenbach for discussions and useful suggestions during the conceptualization phase of the research. We would also like to thank the reviewers for their suggestions and comments.

Financial support. This research has been fully supported by the Croatian Science Foundation under the project "Methodology development for landslide susceptibility assessment for land-use planning based on LiDAR technology", LandSlidePlan (grant nos. HRZZ IP-2019-04-9900, HRZZ DOK-2020-01-2432).

Review statement. This paper was edited by Mario Parise and reviewed by two anonymous referees.

References

- Ballabio, C. and Sterlacchini, S.: Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy, *Math. Geosci.*, 44, 47–70, <https://doi.org/10.1007/s11004-011-9379-9>, 2012.
- Bernat, S., Mihalić Arbanas, S., and Krkač, M.: Landslides triggered in the continental part of Croatia by extreme precipitation in 2013, in: *Engineering geology for society and territory, Landslide Process.*, 2, 1599–1603, 2014a.
- Bernat, S., Mihalić Arbanas, S., and Krkač, M.: Inventory of precipitation triggered landslides in the winter of 2013 in Zagreb (Croatia, Europe), in: *Proceedings of the 3rd World Landslide Forum, Landslide Science for a Safer Geoenvironment: Volume 2: Methods of Landslide Studies*, 3rd World Landslide Forum, Beijing, China, 2–6 June 2014, 829–836, 2014b.
- Bernat Gazibara, S., Krkač, M., Sećanj, M., and Mihalić Arbanas, S.: Identification and Mapping of Shallow Landslides in the City of Zagreb (Croatia) Using the LiDAR-Based Terrain Model, in: *Advancing Culture of Living with Landslides*, Springer International Publishing, Cham, 1093–1100, https://doi.org/10.1007/978-3-319-53498-5_124, 2017.
- Bernat Gazibara, S., Krkač, M., and Mihalić Arbanas, S.: Verification of historical landslide inventory maps for the Podsljeme area in the City of Zagreb using LiDAR-based landslide inventory, *The Mining-Geology-Petroleum Engineering Bulletin*, 34, 45–58, <https://doi.org/10.17794/rgn.2019.1.5>, 2019a.
- Bernat Gazibara, S., Krkač, M., and Mihalić Arbanas, S.: Landslide inventory mapping using LiDAR data in the City of Zagreb (Croatia), *J. Maps*, 15, 773–779, <https://doi.org/10.1080/17445647.2019.1671906>, 2019b.
- Bernat Gazibara, S., Mihalić Arbanas, S., Sinčić, M., Krkač, M., Lukačić, H., Jagodnik, P., and Arbanas, Ž.: LandSlidePlan - Scientific research project on landslide susceptibility assessment in large scale, in: *Proceedings of the 5th regional symposium on landslides in Adriatic – Balkan Region, 5th regional symposium on landslides in Adriatic – Balkan Region*, Rijeka, Croatia, 23–26 March, 99–106, 2022.
- Bernat Gazibara, S., Sinčić, M., Krkač, M., Lukačić, H., and Mihalić Arbanas, S.: Landslide susceptibility assessment on a large scale in the Podsljeme area, City of Zagreb (Croatia), *J. Maps*, 19, 1–11, <https://doi.org/10.1080/17445647.2022.2163197>, 2023.
- Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W.: *Discrete Multivariate Analysis: Theory and Practice*, 557 pp., <https://doi.org/10.1177/014662167700100218>, 1975.
- Bonham-Carter, G. F., Agterberg, F. P., and Wright, D. F.: Weights of evidence modelling: a new approach to mapping mineral potential, *Geological Survey of Canada*, 89-9, <https://doi.org/10.4095/128059>, 1990.
- Bonham-Carter, G. F.: *Geographic Information Systems for Geoscientists: Modelling with GIS*, Pergamon, Ottawa, ISBN 0080424201, 1994.
- Bornaetxea, T., Rossi, M., Marchesini, I., and Alvioli, M.: Effective surveyed area and its role in statistical landslide susceptibility assessments, *Nat. Hazards Earth Syst. Sci.*, 18, 2455–2469, <https://doi.org/10.5194/nhess-18-2455-2018>, 2018.
- Brabb, E. E.: Innovative Approaches to Landslide Hazard Mapping, in: *Proceedings of 4th International Symposium on Landslides*, 307–324, 1984.
- Breiman, L.: Random Forests, *Mach. Learn.*, 45, 5–32, <https://doi.org/10.1023/A:1010933404324>, 2001.
- Catani, F., Lagomarsino, D., Segoni, S., and Tofani, V.: Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues, *Nat. Hazards Earth Syst. Sci.*, 13, 2815–2831, <https://doi.org/10.5194/nhess-13-2815-2013>, 2013.
- Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D. T., Duan, Z., and Ma, J.: A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility, *Catena (Amst)*, 151, 147–160, <https://doi.org/10.1016/j.catena.2016.11.032>, 2017.
- Chung, C. J. F. and Fabbri, A. G.: Probabilistic Prediction Models for Landslide Hazard Mapping, *Photogramm. Eng. Remote Sens.*, 65, 1389–1399, 1999.
- Chung, C.-J. F. and Fabbri, A. G.: Validation of Spatial Prediction Models for Landslide Hazard Mapping, *Nat. Hazards*, 30, 451–472, <https://doi.org/10.1023/B:NHAZ.0000007172.62651.2b>, 2003.
- City of Zagreb: Land use map of City of Zagreb, <https://registri.nipp.hr/izvori/view.php?id=68> (last access: 10 December 2024), 2011.
- 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., and Smith, J. T.: Recommendations for the quantitative analysis of landslide risk, *B. Eng. Geol. Environ.*, 73, 209–263, <https://doi.org/10.1007/s10064-013-0538-8>, 2014.

- Cortes, C. and Vapnik, V.: Support-vector networks, *Mach. Learn.*, 20, 273–297, <https://doi.org/10.1007/BF00994018>, 1995.
- Cox, D. R.: The Regression Analysis of Binary Sequences, *J. Roy. Stat. Soc.*, 20, 215–242, 1958.
- Cui, K., Lu, D., and Li, W.: Comparison of landslide susceptibility mapping based on statistical index, certainty factors, weights of evidence and evidential belief function models, *Geocarto Int.*, 32, 935–955, <https://doi.org/10.1080/10106049.2016.1195886>, 2017.
- Donati, L. and Turrini, M. C.: An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy), *Eng. Geol.*, 63, 277–289, [https://doi.org/10.1016/S0013-7952\(01\)00087-4](https://doi.org/10.1016/S0013-7952(01)00087-4), 2002.
- Dou, J., Tien Bui, D., P. Yunus, A., Jia, K., Song, X., Revhaug, I., Xia, H., and Zhu, Z.: Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan, *PLoS One*, 10, e0133262, <https://doi.org/10.1371/journal.pone.0133262>, 2015.
- Evans, J. S., Oakleaf, J., Cushman, S. A., and Theobald, D.: An arc gis toolbox for surface gradient and geo-morphometric modeling, version 2.0-0., <https://evansmurphy.wixsite.com/evansspatial/arcgis-gradient-metrics-toolbox> (last access: 6 February 2024), 2014.
- Farooq, S. and Akram, M. S.: Landslide susceptibility mapping using information value method in Jhelum Valley of the Himalayas, *Arab. J. Geosci.*, 14, 824, <https://doi.org/10.1007/s12517-021-07147-7>, 2021.
- Fawcett, T.: An introduction to ROC analysis, *Pattern Recognit. Lett.*, 27, 861–874, <https://doi.org/10.1016/j.patrec.2005.10.010>, 2006.
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., and Savage, W. Z.: Guidelines for landslide susceptibility, hazard and risk zoning for land use planning, *Eng. Geol.*, 102, 85–98, <https://doi.org/10.1016/j.enggeo.2008.03.022>, 2008a.
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., and Savage, W. Z.: Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning, *Eng. Geol.*, 102, 99–111, <https://doi.org/10.1016/j.enggeo.2008.03.014>, 2008b.
- Fressard, M., Thiery, Y., and Maquaire, O.: Which data for quantitative landslide susceptibility mapping at operational scale? Case study of the Pays d’Auge plateau hillslopes (Normandy, France), *Nat. Hazards Earth Syst. Sci.*, 14, 569–588, <https://doi.org/10.5194/nhess-14-569-2014>, 2014.
- Gaidzik, K. and Ramírez-Herrera, M. T.: The importance of input data on landslide susceptibility mapping, *Sci. Rep.*, 11, 19334, <https://doi.org/10.1038/s41598-021-98830-y>, 2021.
- Gorsevski, P. V., Gessler, P. E., Foltz, R. B., and Elliot, W. J.: Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis, *T. GIS*, 10, 395–415, <https://doi.org/10.1111/j.1467-9671.2006.01004.x>, 2006.
- Guzzetti, F., Carrara, A., Cardinali, M., and Reichenbach, P.: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy, *Geomorphology*, 31, 181–216, [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1), 1999.
- Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., and Galli, M.: Estimating the quality of landslide susceptibility models, *Geomorphology*, 81, 166–184, <https://doi.org/10.1016/j.geomorph.2006.04.007>, 2006a.
- Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F., and Cardinali, M.: Landslide hazard assessment in the Collazzone area, Umbria, Central Italy, *Nat. Hazards Earth Syst. Sci.*, 6, 115–131, <https://doi.org/10.5194/nhess-6-115-2006>, 2006b.
- Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., and Chang, K.-T.: Landslide inventory maps: New tools for an old problem, *Earth Sci. Rev.*, 112, 42–66, <https://doi.org/10.1016/j.earscirev.2012.02.001>, 2012.
- Habumugisha, J. M., Chen, N., Rahman, M., Islam, M. M., Ahmad, H., Elbeltagi, A., Sharma, G., Liza, S. N., and Dewan, A.: Landslide Susceptibility Mapping with Deep Learning Algorithms, *Sustainability*, 14, 1734, <https://doi.org/10.3390/su14031734>, 2022.
- Hastie, T., Tibshirani, R., and Friedman, J.: The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Springer, New York, 745, <https://doi.org/10.1007/b94608>, 2009.
- Hemasinghe, H., Rangali, R. S. S., Deshapriya, N. L., and Samarakoon, L.: Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka), *Procedia Eng.*, 212, 1046–1053, <https://doi.org/10.1016/j.proeng.2018.01.135>, 2018.
- Ho, T. K.: Random decision forests, in: Proceedings of 3rd International Conference on Document Analysis and Recognition, 278–282, <https://doi.org/10.1109/ICDAR.1995.598994>, 1995.
- Huang, F., Zhang, J., Zhou, C., Wang, Y., Huang, J., and Zhu, L.: A deep learning algorithm using a fully connected sparse autoencoder neural network for landslide susceptibility prediction, *Landslides*, 17, 217–229, <https://doi.org/10.1007/s10346-019-01274-9>, 2020a.
- Huang, F., Cao, Z., Guo, J., Jiang, S.-H., Li, S., and Guo, Z.: Comparisons of heuristic, general statistical and machine learning models for landslide susceptibility prediction and mapping, *Catena (Amst)*, 191, 104580, <https://doi.org/10.1016/j.catena.2020.104580>, 2020b.
- Huang, F., Tao, S., Li, D., Lian, Z., Catani, F., Huang, J., Li, K., and Zhang, C.: Landslide Susceptibility Prediction Considering Neighborhood Characteristics of Landslide Spatial Datasets and Hydrological Slope Units Using Remote Sensing and GIS Technologies, *Remote Sens. (Basel)*, 14, 4436, <https://doi.org/10.3390/rs14184436>, 2022.
- Jacobs, L., Kervyn, M., Reichenbach, P., Rossi, M., Marchesini, I., Alvioli, M., and Dewitte, O.: Regional susceptibility assessments with heterogeneous landslide information: Slope unit- vs. pixel-based approach, *Geomorphology*, 356, 107084, <https://doi.org/10.1016/j.geomorph.2020.107084>, 2020.
- James, G., Witten, D., Hastie, T., and Tibshirani, R.: An Introduction to Statistical Learning, Springer, New York, 606 pp., <https://doi.org/10.1007/978-1-4614-7138-7>, 2013.
- Jebur, M. N., Pradhan, B., and Tehrany, M. S.: Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale, *Remote Sens. Environ.*, 152, 150–165, <https://doi.org/10.1016/j.rse.2014.05.013>, 2014.
- Jurak, V., Ivan Matković, Miklin, Ž., and Snježana Mihalić: Data analysis of the landslides in the Republic of Croatia: Present state and perspectives, in: Landslides, 1923–1929, 1996.

- Jurak, V., Ortolan, Ž., Ivšić, T., Herak, M., Šumanovac, F., Vukelić, I., Jukić, M., and Šurina, Z.: Geotehničko i seizmičko mikrozoniranje grada Zagreba – pokušaji i ostvarenje (Geotechnical and seismological microzonation of Zagreb City - attempts and accomplishments), in: *Zbornik radova s konferencije razvitak Zagreba, Razvitak Zagreba*, Zagreb, Croatia, 1–2 February 2008, 99–108, 2008 (in Croatian).
- Kavzoglu, T., Sahin, E. K., and Colkesen, I.: Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression, *Landslides*, 11, 425–439, <https://doi.org/10.1007/s10346-013-0391-7>, 2014.
- Krkač, M., Bernat Gazibara, S., Sinčić, M., Lukačić, H., Šarić, G., and Snježana, M. A.: Impact of input data on the quality of the landslide susceptibility large-scale maps: A case study from NW Croatia, in: *Progress in Landslide Research and Technology*, edited by: Alcántara-Ayala, I., Arbanas, Ž., Cuomo, S., Huntley D., Konagai, K., Mihalić Arbanas, S., Mikoš, M., Sassa, K., Tang, H., and Tiwari, B., https://doi.org/10.1007/978-94-007-2162-3_36, 2023.
- Landis, J. R. and Koch, G. G.: The measurement of observer agreement for categorical data, *Biometrics*, 33, 159–174, 1977.
- Lee, S.: Landslide susceptibility mapping using an artificial neural network in the Gangneung area, Korea, *Int. J. Remote Sens.*, 28, 4763–4783, <https://doi.org/10.1080/01431160701264227>, 2007.
- Mathew, J., Jha, V. K., and Rawat, G. S.: Weights of evidence modelling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand, *Curr. Sci.*, 95, 628–638, 2007.
- Merghadi, A., Yunus, A. P., Dou, J., Whiteley, J., ThaiPham, B., Bui, D. T., Avtar, R., and Abderrahmane, B.: Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance, *Earth Sci. Rev.*, 207, 103225, <https://doi.org/10.1016/j.earscirev.2020.103225>, 2020.
- Mihalić, S.: Recommendations for Landslide Hazard and Risk Mapping in Croatia, *Geol. Croatica*, 51, 195–204, <https://doi.org/10.4154/GC.1998.15>, 1988.
- Mihalić Arbanas, S., Bernat, S., Fabijanović, S., and Arbanas, Ž.: The analysis of historical landslide information from the area of the City of Zagreb and Primorsko-Goranska County, in: *Proceedings of the 1st Regional Symposium on Landslides in the Adriatic-Balkan Region – Landslide and Flood Hazard Assessment*, 91–96, https://www.researchgate.net/publication/271908916_Analysis_of_Historical_Landslide_Information_from_the_Area_of_the_City_of_Zagreb_and_Primorsko-Goranska_County (last access: 10 December 2024), 2014.
- Mihalić Arbanas, S., Krkač, M., Bernat Gazibara, S.: Application of advanced technologies in landslide research in the area of the City of Zagreb (Croatia, Europe), *Geol. Croatica*, 69, 179–192, <https://doi.org/10.4154/gc.2016.18>, 2016.
- Mihalić Arbanas, S., Bernat Gazibara, S., Krkač, M., Sinčić, M., Lukačić, H., Jagodnik, P., and Arbanas, Ž.: Landslide Detection and Spatial Prediction: Application of Data and Information from Landslide Maps, in: *Progress in Landslide Research and Technology*, Volume 1, edited by: Alcántara-Ayala, I., Arbanas, Ž., Huntley, D., Konagai, K., Mikoš, M., Sassa, K., Sassa, S., Tang, H., and Tiwari, B., Fullerton, CA, USA, 195–212, https://doi.org/10.1007/978-3-031-18471-0_16, 2023.
- Miklin, Ž., Mlinar, Ž., Brkić, Ž., Hečimović, I., and Dolić, M.: Detailed engineering geological map of Podsljeme urbanised zone at a scale of 1 : 5000 (DIGK-Phase I), 2007.
- Nefeslioglu, H. A., Gokceoglu, C., and Sonmez, H.: An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps, *Eng. Geol.*, 97, 171–191, <https://doi.org/10.1016/j.enggeo.2008.01.004>, 2008.
- Neuhäuser, B. and Terhorst, B.: Landslide susceptibility assessment using “weights-of-evidence” applied to a study area at the Jurassic escarpment (SW-Germany), *Geomorphology*, 86, 12–24, <https://doi.org/10.1016/j.geomorph.2006.08.002>, 2007.
- Neuhäuser, B., Damm, B., and Terhorst, B.: GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model, *Landslides*, 9, 511–528, <https://doi.org/10.1007/s10346-011-0305-5>, 2012.
- Pascale, S., Parisi, S., Mancini, A., Schiattarella, M., Conforti, M., Sole, A., Murgante, B., and Sdao, F.: Landslide Susceptibility Mapping Using Artificial Neural Network in the Urban Area of Senise and San Costantino Albanese (Basilicata, Southern Italy), 473–488, https://doi.org/10.1007/978-3-642-39649-6_34, 2013.
- Polak, K., Klemar, M., Nejkova, M., Radošević, N., Stepan, Z., Miroslav, M., and Križanić, Z.: Lithological analysis and categorisation of terrain according to soil stability of Medvednica slopes in the area of the City of Zagreb, *Geoteknika-Geoexpert, Zagreb*, 102 pp., 1979 (in Croatian).
- Petschko, H., Bell, R., and Glade, T.: Effectiveness of visually analyzing LiDAR DTM derivatives for earth and debris slide inventory mapping for statistical susceptibility modeling, *Landslides*, 13, 857–872, <https://doi.org/10.1007/s10346-015-0622-1>, 2016.
- Pourghasemi, H. R., Sadhasivam, N., Amiri, M., Eskandari, S., and Santosh, M.: Landslide susceptibility assessment and mapping using state-of-the art machine learning techniques, *Nat. Hazards*, 108, 1291–1316, <https://doi.org/10.1007/s11069-021-04732-7>, 2021.
- Pradhan, B.: A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS, *Comput. Geosci.*, 51, 350–365, <https://doi.org/10.1016/j.cageo.2012.08.023>, 2013.
- Ray, R.L., Lazzari, M., and Olutimehin, T.: Remote Sensing Approaches and Related Techniques to Map and Study Landslides, in: *Landslides – Investigation and Monitoring*, IntechOpen, <https://doi.org/10.5772/intechopen.93681>, 2020.
- Razak, K. A., Santangelo, M., Van Westen, C. J., Straatsma, M. W., and de Jong, S. M.: Generating an optimal DTM from airborne laser scanning data for landslide mapping in a tropical forest environment, *Geomorphology*, 190, 112–125, <https://doi.org/10.1016/j.geomorph.2013.02.021>, 2013.
- Reichenbach, P., Rossi, M., Malamud, B. D., Mihir, M., and Guzzetti, F.: A review of statistically-based landslide susceptibility models, *Earth Sci. Rev.*, 180, 60–91, <https://doi.org/10.1016/j.earscirev.2018.03.001>, 2018.
- Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C., and Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, *Geomorphology*, 114, 129–142, <https://doi.org/10.1016/j.geomorph.2009.06.020>, 2010.
- Rossi, M., Bornaetxea, T., and Reichenbach, P.: LAND-SUITE V1.0: a suite of tools for statistically based landslide sus-

- ceptibility zonation, *Geosci. Model Dev.*, 15, 5651–5666, <https://doi.org/10.5194/gmd-15-5651-2022>, 2022.
- Sandić, C., Marjanović, M., Abolmasov, B., and Tošić, R.: Integrating landslide magnitude in the susceptibility assessment of the City of Doboj, using machine learning and heuristic approach, *J. Maps*, 19, <https://doi.org/10.1080/17445647.2022.2163199>, 2023.
- Sarkar, S., Roy, A. K., and Martha, T. R.: Landslide susceptibility assessment using Information Value Method in parts of the Darjeeling Himalayas, *J. Geol. Soc. India*, 82, 351–362, <https://doi.org/10.1007/s12594-013-0162-z>, 2013.
- Scaioni, M., Longoni, L., Melillo, V., and Papini, M.: Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives, *Remote Sens. (Basel)*, 6, 9600–9652, <https://doi.org/10.3390/rs6109600>, 2014.
- Shirvani, Z.: A Holistic Analysis for Landslide Susceptibility Mapping Applying Geographic Object-Based Random Forest: A Comparison between Protected and Non-Protected Forests, *Remote Sens. (Basel)*, 12, 434, <https://doi.org/10.3390/rs12030434>, 2020.
- Sinčić, M., Bernat Gazibara, S., Krkač, M., Lukačić, H., and Mihalić Arbanas, S.: The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments, *Land (Basel)*, 11, 1360, <https://doi.org/10.3390/land11081360>, 2022a.
- Sinčić, M., Bernat Gazibara, S., Krkač, M., and Mihalić Arbanas, S.: Landslide susceptibility assessment of the City of Karlovac using the bivariate statistical analysis, *Rudarsko-geološko-naftni zbornik*, 37, 149–170, <https://doi.org/10.17794/rgn.2022.2.13>, 2022b.
- Soeters, R. and van Westen, C. J.: Slope Instability Recognition Analysis and Zonation, in: *Landslides: Investigation and Mitigation*, edited by: Turner, K. T. and Schuster, R. L., Washington DC, 129–177, https://www.researchgate.net/publication/209803184_Slope_instability_Recognition_analysis_and_zonation (last access: 10 December 2024), 1996.
- State Geodetic Administration: WMS server, topographic map: <https://geoportal.dgu.hr/services/tk/wms> (last access: 6 February 2024), 2024.
- Šikić, V.: Engineering geology of Zagreb – north and south, 1967.
- Šikić, K., Basch, O., and Šimunić, A.: Basic Geological Map, scale 1:100 000, Zagreb sheet, 1972.
- Šikić, K., Basch, O., and Šimunić, A.: Geological notes for Basic geological map, scale 1:100 000, Zagreb sheet, 75 pp., 1979.
- Tehrany, M. S., Pradhan, B., and Jebur, M. N.: Spatial prediction of flood susceptible areas using rule based decision tree (DT) and a novel ensemble bivariate and multivariate statistical models in GIS, *J. Hydrol. (Amst)*, 504, 69–79, <https://doi.org/10.1016/j.jhydrol.2013.09.034>, 2013.
- Tehrany, M. S., Pradhan, B., and Jebur, M. N.: Flood susceptibility mapping using a novel ensemble weights-of-evidence and support vector machine models in GIS, *J. Hydrol. (Amst)*, 512, 332–343, <https://doi.org/10.1016/j.jhydrol.2014.03.008>, 2014.
- Tyagi, A., Tiwari, R. K., and James, N.: Mapping the landslide susceptibility considering future land-use land-cover scenario, *Landslides*, 20, 65–76, <https://doi.org/10.1007/s10346-022-01968-7>, 2023.
- Vakhshoori, V. and Zare, M.: Is the ROC curve a reliable tool to compare the validity of landslide susceptibility maps?, *Geomatics, Nat. Hazard. Risk*, 9, 249–266, <https://doi.org/10.1080/19475705.2018.1424043>, 2018.
- van Westen, C. J.: Use of weights of evidence modeling for landslide susceptibility mapping, 23 pp., https://filetransfer.itc.nl/pub/westen/GISSIZ/Chapter6-Exercises/GISexercises/Datadrivenmodels/Exercisedata_drivenmodels2002.pdf (last access: 10 December 2024), 2002.
- van Westen, C. J., van Duren, I., Kruse, H. M. G., and Terlien, M. T. J.: Training Package for Geographic Information Systems in Slope Stability Zonation, Part 1: Theory, ITC Publication, Eeschede, 245 pp., 1993.
- van Westen, C. J., Rengers, N., and Soeters, R.: Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment, *Nat. Hazards*, 30, 399–419, <https://doi.org/10.1023/B:NHAZ.0000007097.42735.9e>, 2003.
- van Westen, C. J., Castellanos, E., and Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview, *Eng. Geol.*, 102, 112–131, <https://doi.org/10.1016/j.enggeo.2008.03.010>, 2008.
- Vapnik, V. N.: The Nature of Statistical Learning Theory, Springer New York, New York, NY, <https://doi.org/10.1007/978-1-4757-2440-0>, 1995.
- Vojteková, J. and Vojtek, M.: Assessment of landslide susceptibility at a local spatial scale applying the multi-criteria analysis and GIS: a case study from Slovakia, *Geomatics, Nat. Hazard. Risk*, 11, 131–148, <https://doi.org/10.1080/19475705.2020.1713233>, 2020.
- Wang, L.-J., Guo, M., Sawada, K., Lin, J., and Zhang, J.: A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network, *Geosci. J.*, 20, 117–136, <https://doi.org/10.1007/s12303-015-0026-1>, 2016a.
- Wang, Q., Li, W., Wi, Y., Pei, Y., Xing, M., and Yang, D.: A comparative study on the landslide susceptibility mapping using evidential belief function and weights of evidence models, *J. Earth Syst. Sci.*, 125, 645–662, <https://doi.org/10.1007/s12040-016-0686-x>, 2016b.
- Wang, G., Chen, X., and Chen, W.: Spatial Prediction of Landslide Susceptibility Based on GIS and Discriminant Functions, *ISPRS Int. J. Geoinf.*, 9, 144, <https://doi.org/10.3390/ijgi9030144>, 2020.
- Wang, H., Zhang, L., Luo, H., He, J., and Cheung, R. W. M.: AI-powered landslide susceptibility assessment in Hong Kong, *Eng. Geol.*, 288, 106103, <https://doi.org/10.1016/j.enggeo.2021.106103>, 2021.
- Xiao, T., Yu, L., Tian, W., Zhou, C., and Wang, L.: Reducing Local Correlations Among Causal Factor Classifications as a Strategy to Improve Landslide Susceptibility Mapping, *Front. Earth Sci. (Lausanne)*, 9, 781674, <https://doi.org/10.3389/feart.2021.781674>, 2021.
- Xing, Y., Yue, J., Guo, Z., Chen, Y., Hu, J., and Travé, A.: Large-Scale Landslide Susceptibility Mapping Using an Integrated Machine Learning Model: A Case Study in the Lvliang Mountains of China, *Front. Earth Sci. (Lausanne)*, 9, 722491, <https://doi.org/10.3389/feart.2021.722491>, 2021.
- Yan, G., Liang, S., Gui, X., Xie, Y., and Zhao, H.: Optimizing landslide susceptibility mapping in the Kongtong District, NW China: comparing the subdivi-

- sion criteria of factors, *Geocarto. Int.*, 34, 1408–1426, <https://doi.org/10.1080/10106049.2018.1499816>, 2019.
- Yao, X., Tham, L. G., and Dai, F. C.: Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China, *Geomorphology*, 101, 572–582, <https://doi.org/10.1016/j.geomorph.2008.02.011>, 2008.
- Yin, K. L. and Yan, T. Z.: Statistical prediction model for slope instability of metamorphosed rocks, in: *Proceedings of Fifth International Symposium on Landslides*, 1269–1272, 1988.
- Yusof, N. M., Pradhan, B., Shafri, H. Z. M., Jebur, M. N., and Yusoff, Z.: Spatial landslide hazard assessment along the Jelapang Corridor of the North-South Expressway in Malaysia using high resolution airborne LiDAR data, *Arab. J. Geosci.*, 8, 9789–9800, <https://doi.org/10.1007/s12517-015-1937-x>, 2015.
- Zêzere, J. L., Pereira, S., Melo, R., Oliveira, S. C., and Garcia, R. A. C.: Mapping landslide susceptibility using data-driven methods, *Sci. Total Environ.*, 589, 250–267, <https://doi.org/10.1016/j.scitotenv.2017.02.188>, 2017.
- Zhao, X. and Chen, W.: Optimization of Computational Intelligence Models for Landslide Susceptibility Evaluation, *Remote Sens. (Basel)*, 12, 2180, <https://doi.org/10.3390/rs12142180>, 2020.



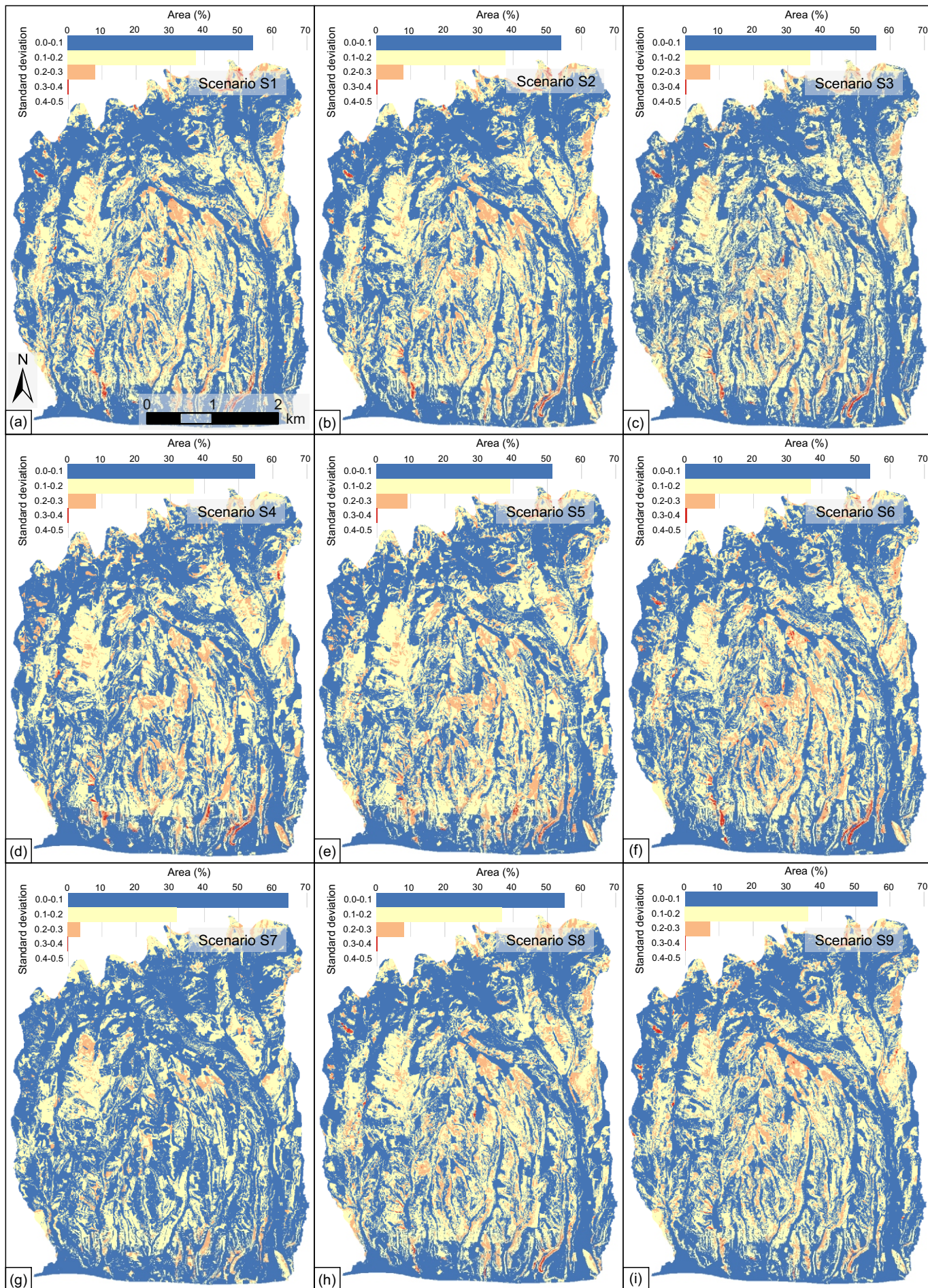
Supplement of

Comparison of conditioning factor classification criteria in large-scale statistically based landslide susceptibility models

Marko Sinčić et al.

Correspondence to: Marko Sinčić (msincic@rgn.hr)

The copyright of individual parts of the supplement might differ from the article licence.



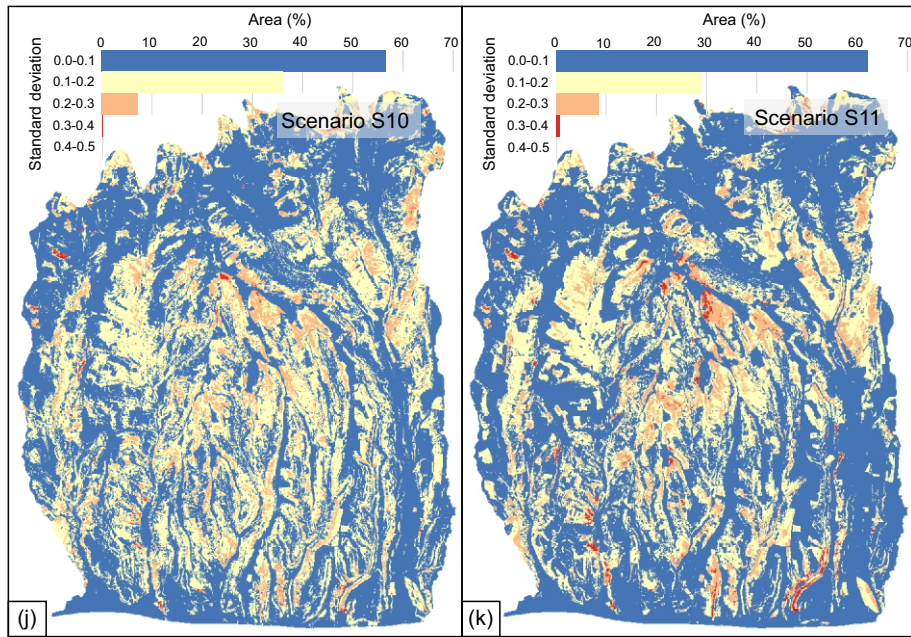
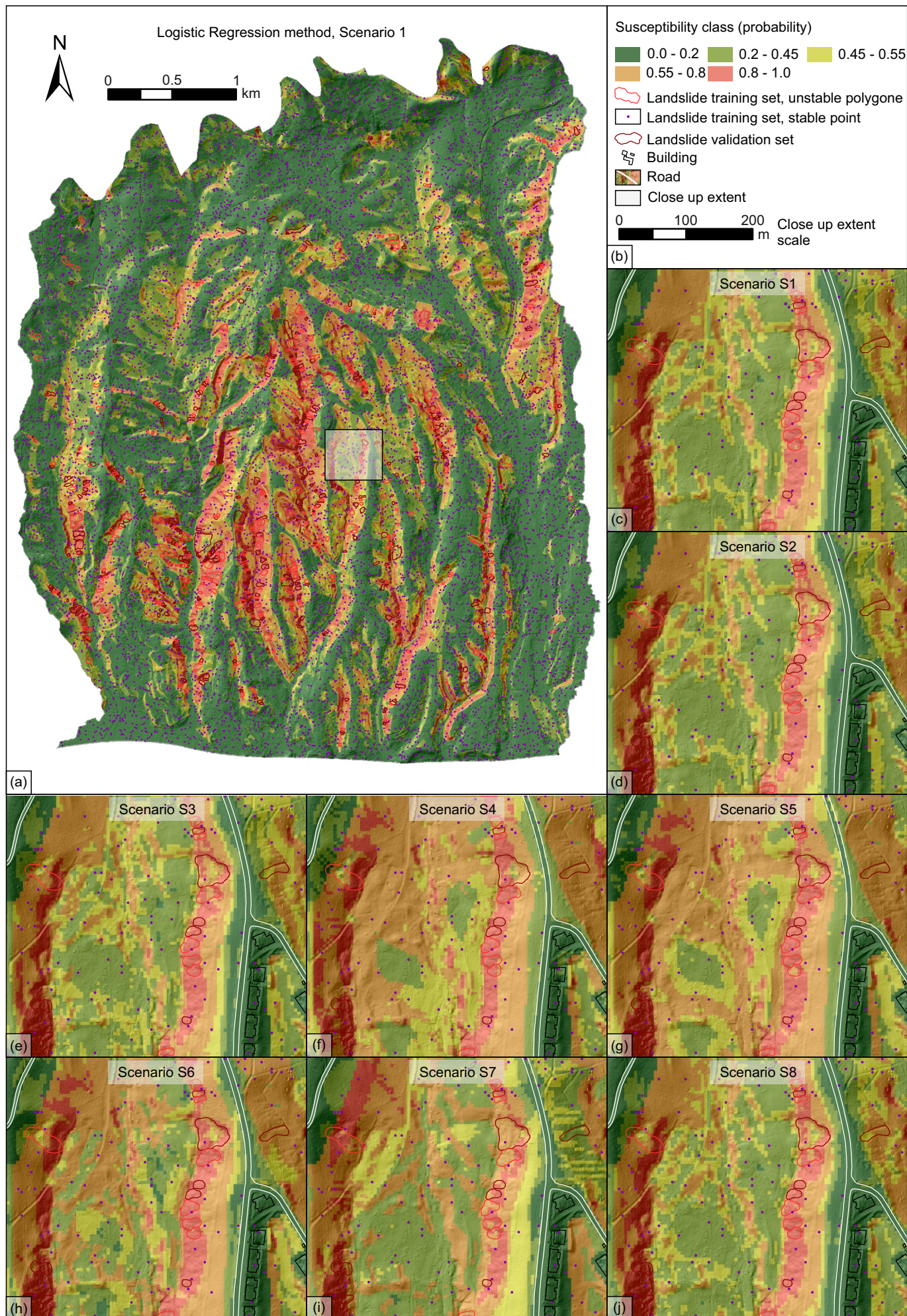


Figure S1 Standard deviation maps and class area distribution defined through five statistically based methods for each scenario (i.e. scenarios S1 to S11)



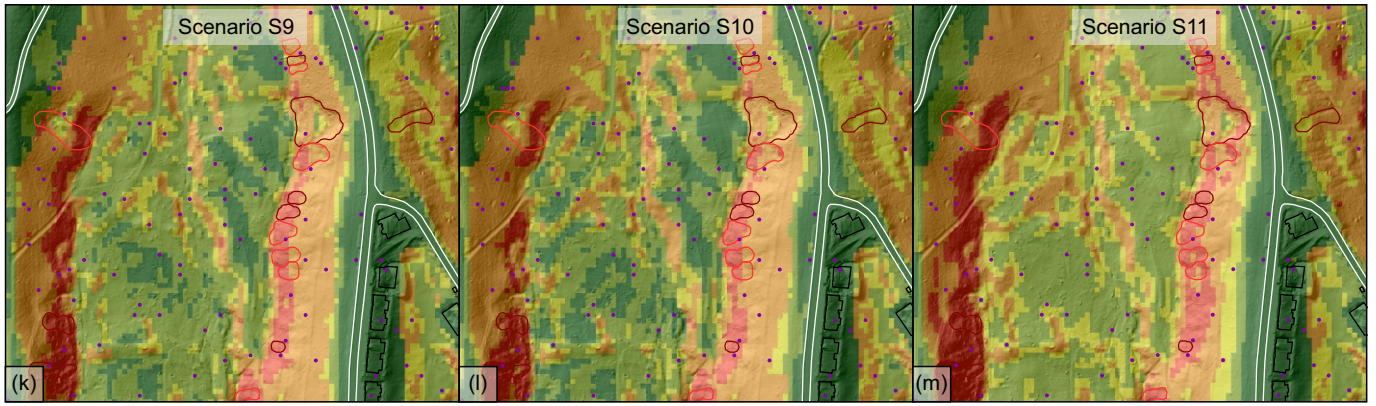


Figure S2 Landslide susceptibility maps derived using logistic regression method: (a) scenario 11 study area view; (b) legend; (c) to (m) close up view for S1 to S11 scenarios

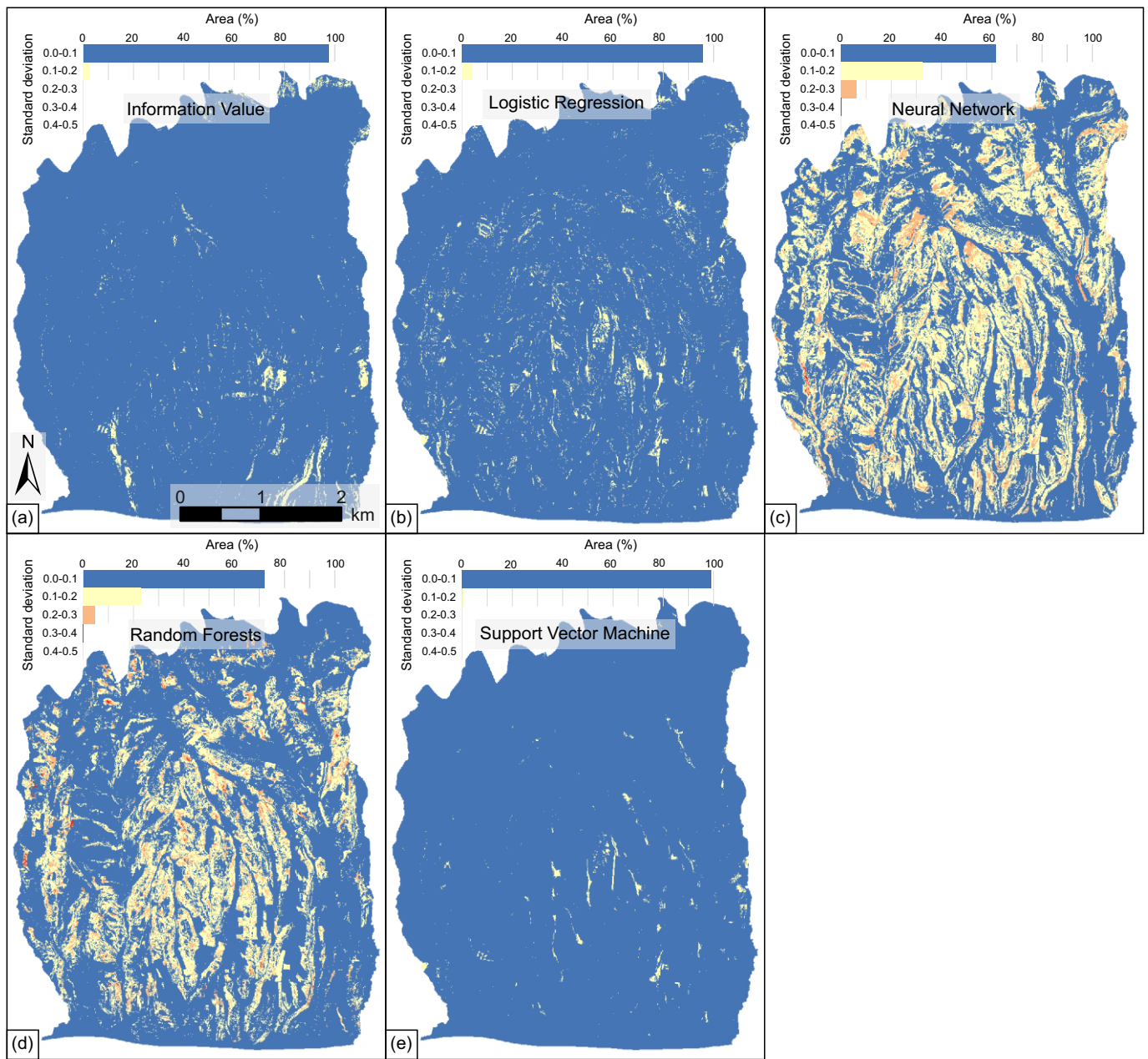


Figure S3. Standard deviation maps and class area distribution defined through 11 scenarios for each method, i.e. IV (a), LR (b), NN (c), RF (d), SVM (e)

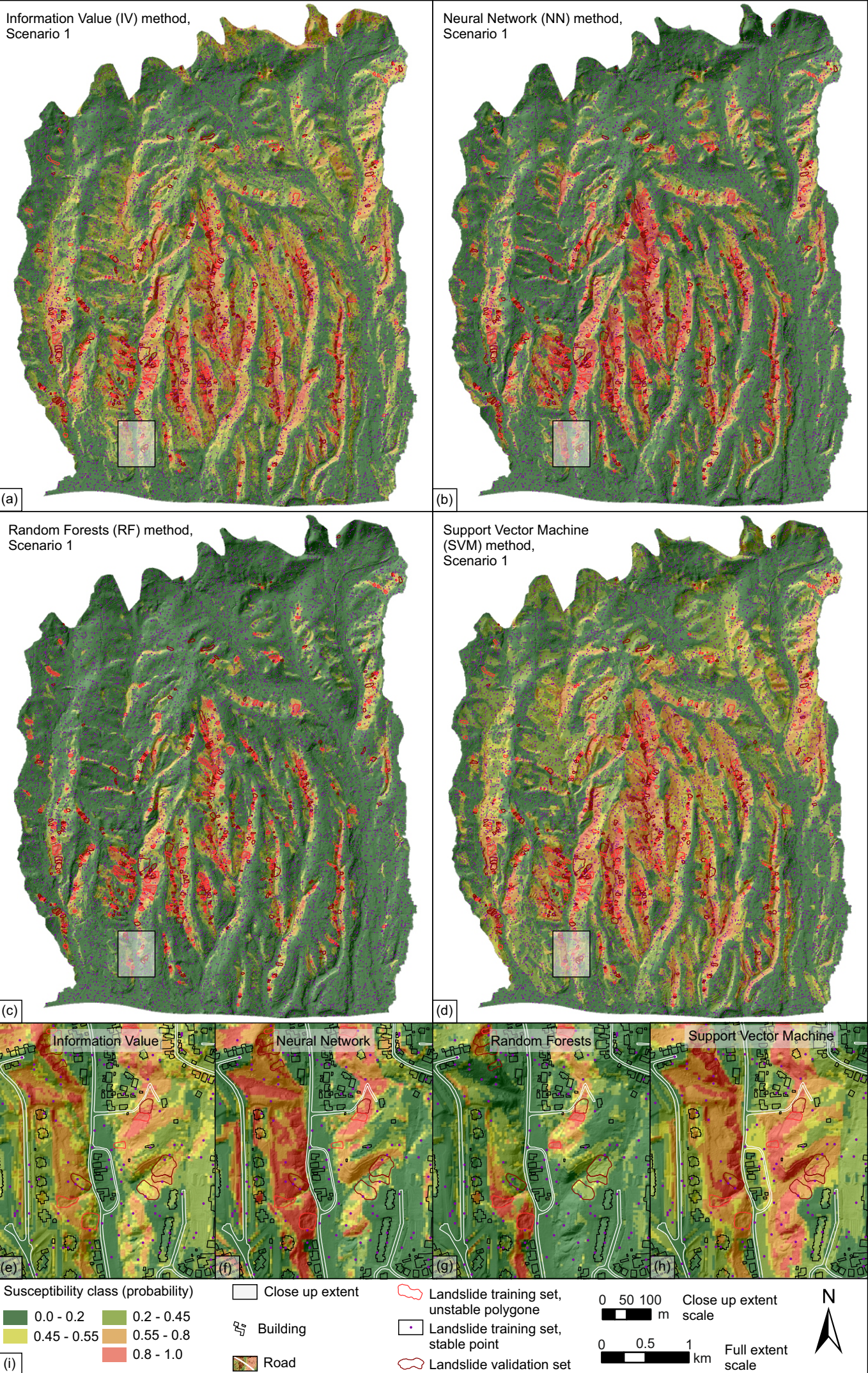


Figure S4 Landslide susceptibility maps derived in scenario S1: (a) Information Value method, study area view and (e) close up view; (b) Neural Network method, study area view and (f) close up view; (c) Random Forests method, study area view and (g) close up view; (d) Support Vector Machine method, study area view and (h) close up view; (i) legend

2.3 A comprehensive comparison of stable and unstable area sampling strategies in large-scale landslide susceptibility models using machine learning methods

Authors:

Marko Sinčić, Sanja Bernat Gazibara, Mauro Rossi, Martin Krkač, Snježana Mihalić Arbanas

Published in:

Remote Sensing 2024, 16(16), 2923

Published on:

9 August 2024

DOI:

<https://doi.org/10.3390/rs16162923>



Article

A Comprehensive Comparison of Stable and Unstable Area Sampling Strategies in Large-Scale Landslide Susceptibility Models Using Machine Learning Methods

Marko Sinčić ^{1,*} , Sanja Bernat Gazibara ¹ , Mauro Rossi ² , Martin Krkač ¹ and Snježana Mihalić Arbanas ¹

¹ Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, 10000 Zagreb, Croatia; sanja.bernat@rgn.unizg.hr (S.B.G.); martin.krkač@rgn.unizg.hr (M.K.); snjezana.mihalic@rgn.unizg.hr (S.M.A.)

² Istituto di Ricerca per la Protezione Idrogeologica, Consiglio Nazionale delle Ricerche, 06128 Perugia, Italy; mauro-rossi@cnr.it

* Correspondence: marko.sincic@rgn.unizg.hr

Abstract: This paper focuses on large-scale landslide susceptibility modelling in NW Croatia. The objective of this research was to provide new insight into stable and unstable area sampling strategies on a representative inventory of small and shallow landslides mainly occurring in soil and soft rock. Four strategies were tested for stable area sampling (random points, stable area polygon, stable polygon buffering and stable area centroid) in combination with four strategies for unstable area sampling (landslide polygon, smoothing digital terrain model derived landslide conditioning factors, polygon buffering and landslide centroid), resulting in eight sampling scenarios. Using Logistic Regression, Neural Network, Random Forest and Support Vector Machine algorithm, 32 models were derived and analysed. The main conclusions reveal that polygon sampling of unstable areas is an imperative in large-scale modelling, as well as that subjective and/or biased stable area sampling leads to misleading models. Moreover, Random Forest and Neural Network proved to be more favourable methods (0.804 and 0.805 AUC, respectively), but also showed extreme sensitivity to the tested sampling strategies. In the comprehensive comparison, the advantages and disadvantages of 32 derived models were analysed through quantitative and qualitative parameters to highlight their application to large-scale landslide zonation. The results yielded by this research are beneficial to the susceptibility modelling step in large-scale landslide susceptibility assessments as they enable the derivation of more reliable zonation maps applicable to spatial and urban planning systems.

Keywords: landslide susceptibility; sampling strategy; large scale; machine learning; remote sensing; Croatia



Citation: Sinčić, M.; Bernat Gazibara, S.; Rossi, M.; Krkač, M.; Mihalić Arbanas, S. A Comprehensive Comparison of Stable and Unstable Area Sampling Strategies in Large-Scale Landslide Susceptibility Models Using Machine Learning Methods. *Remote Sens.* **2024**, *16*, 2923. <https://doi.org/10.3390/rs16162923>

Academic Editor: Domenico Calcaterra

Received: 12 June 2024

Revised: 29 July 2024

Accepted: 31 July 2024

Published: 9 August 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Landslide zoning maps are useful tools in managing landslide hazards as they present territory divided into homogeneous areas, considering the degree of susceptibility, hazard or risk [1], whereas a landslide susceptibility map depicts the spatial probability of landslide occurrence [2]. Intensive research on landslide susceptibility has been carried out in the last 40 years, starting with the key report by Varnes and International Association for Engineering Geology Commission on Landslides and other Mass-Movements [3], followed by the most significant publications [1,4–9] in the period from 1996 to 2013. The increasing trend in landslide susceptibility studies has resulted in more recent review papers on specific topics, such as a review of national studies [10], certain geographical regions [11] or hazard zonation techniques [12]. Furthermore, reviews on various aspects of landslide susceptibility modelling are given in [13] (from 1983 to 2016), [14] (from 2005 to 2016) and [15] (from 1999 to 2018).

In Croatia (Europe), refs. [16,17] identified the necessity of applying large-scale landslide susceptibility maps in spatial and urban planning. The main purpose of landslide

susceptibility maps is to provide information about the spatial probability of small and shallow landslides occurrences that limit potential land-use. This study focused on a hilly area (approx. 20 km²) located in the Pannonian Basin (NW Croatia) with prevailing precipitation-triggered landslides [18] that caused damage to buildings and infrastructure. In previous research, great attention was also given to developing relevant landslide and thematic input data [19,20]. This paper deals with an open question of sampling strategies of stable and unstable areas in the given environmental settings, focusing on the determination of its influence in statistically based landslide susceptibility modelling for application in large-scale spatial and urban planning. The choice of sampling strategy for stable/unstable mapping unit selection is a crucial issue in landslide susceptibility assessments, as the type and parametrization of the sampling approach significantly change landslide susceptibility models (LSMs).

The random selection of stable points usually serves as a reference for comparison with many other creative and novel approaches that have been introduced into the literature. For example, [21] and/or [22] analysed and compared the following most significant sampling methods of stable points: Bioclim [23], domain methods [24], one-class support vector algorithm [25], Improved Target Space Exteriorization Sampling [26], buffer control sampling, information value, mini-batch K-Medoids [27] and integrative sampling combination [22]. Furthermore, ref. [28] introduced similarity-based sampling, ref. [29] used Newmark-based sampling, while [30] presented an objective Mahalanobis distances-based method. More simple approaches deal with restricting the location of the sampling area to low-susceptibility zones based on preliminary LSMs [31], either by means of different buffer zones around landslides [32] or by means of a rectangle zone in lowlands with (and without) buffer zones [32]. Ref. [33] favoured sampling stable areas from randomly generated circles with a diameter equal to the mean landslide width, whereas [34] proved that landslide susceptibility is sensitive to random sampling of stable data in lithologically heterogeneous areas.

Considering unstable area sampling, ref. [35] compared landslide scarp centroid, scarp points on a 50 m grid and the entire scarp polygon in two pilot areas, and generally found low differences in success and prediction rates. Research conducted by [36] shows better results with landslide scarp and body polygons over scarp and body centroids. Similarly, ref. [37] characterize polygon sampling with better accuracy and lower uncertainty, whereas [38] found low differences between polygon scarp area and polygon landslide area. A case study by [39] highlighted that choosing a single point at a 50 m distance from the scarp area results in higher uncertainties, similar to the conclusion of [40], claiming that the point can be misleading and results in lower accuracies. For the sampling of landslides as centroids, ref. [41] suggest that landslide run-out should be included separately in the landslide inventory to derive more reliable LSMs. Ref. [42] introduced seed cells, i.e., zones that represent undisturbed morphological settings before the occurrence of a landslide, followed by [43], who found little differences between seed cells and landslide scarp, while also indicating the lowest results in point sampling. Generally, in landslide susceptibility studies that do not investigate sampling strategies, the most commonly used are landslide points (e.g., centroid) [44–49] and polygons [50–54].

In the past few years, the notable increase in research studies considering the stable area sampling underlines that this topic is of high interest. Besides that, studies commonly investigate traditional and/or experimental strategies, dealing with sampling either stable areas or unstable areas, whereas the investigation of both simultaneously is rarely considered. For instance, ref. [55] defined 14 scenarios based on two strategies to sample unstable areas (i.e., landslide core and landslide extension) and seven strategies for stable areas sampling (random sampling, three buffer options and three areas defined as very low-susceptibility zones). Based on this research, it arises that a random selection of stable areas may enlarge the uncertainty of modelling, whereas the landslide core is suggested as a more suitable option for unstable area sampling. Besides the sampling strategies relating to stable or unstable areas, their size ratios are also discussed by means of oversampling

or undersampling [21,56]. The most common size ratio sampling encompasses an equal amount of stable and unstable areas [22,29,32,55,57,58]. Furthermore, ref. [57] suggested a minimum of 400 stable and unstable points to be used for a case study, whereas [35] and [59] managed to achieve excellent predictive performance by training a model with as low as 10% of the available unstable samples. All of the abovementioned shows that the results and conclusions vary from study to study, and currently, no uniform proposal exists. Moreover, a lack of papers dealing with sampling both stable and unstable areas is evident, as well as the necessity to test the sampling strategies by means of their influence on the particular applicability of a LSM, especially in large-scale analysis.

The main objective of this paper is to determine a more objective and effective sampling strategy for stable and unstable areas for large-scale landslide susceptibility modelling. Specifically, this is carried out in hilly areas with prevailing small and shallow landslides to propose the type and parametrisation of sampling approach that can be used for the development of landslide zonation maps. In order to achieve a comprehensive comparison of sampling strategies, we tested four strategies for stable area sampling (random points, stable area polygon, stable polygon buffering and stable area centroid) in combination with four strategies for unstable area sampling (landslide polygon, smoothing digital terrain model (DTM)-derived landslide conditioning factors (LCFs), polygon buffering and landslide centroid). Based on combining the defined sampling strategies, eight modelling scenarios are prepared to apply to four commonly used machine learning methods (e.g., [14,60–62]), namely, Logistic Regression (LR), Neural Network (NN), Random Forests (RFs) and Support Vector Machine (SVM), resulting in 32 LSMs.

Considering the environmental setting of hilly areas with small and shallow landslides and the scope of the modelling, that is, large-scale susceptibility zonation for land-use planning, this research is based on very high-resolution (HR) remote sensing data (LiDAR, i.e., Light Detection and Ranging DTM-derived maps and orthophoto images of HR) necessary for 5 m pixel-based susceptibility analysis. The novelties of this research are twofold: (i) it highlights the possible pros and cons of landslide sampling in probabilistic zonation based on combinations of multiple sampling strategies of stable and unstable areas; and (ii) it introduces novel sampling strategies to test their applicability to large-scale zonation maps for spatial and urban planning at the local level, which requires accurate susceptibility information derived from representative landslide and thematic mapping.

2. Material

2.1. Study Area

The study area is in NW Croatia. It spans 20.2 square kilometers across the Bednja Municipality and Lepoglava City (Figure 1). According to [63], landslides are abundant in the area, with more than 1500 in NW Croatia occurring from 2006 to 2014, mainly triggered by precipitation events and intense snow melt [18]. The study area is predominantly hilly, with 90% of the area having slope angles of $> 5^\circ$ [19], indicating a high risk of landslides. Approximately 78% of the area is composed of Miocene sediments, i.e., sandstones, marls, sands, tuffs (Burdigalian) and biogenic, sandy, and marly limestones, calcareous marls and sandstone (Tortonian). Sands, silts and gravels located in the valleys are Quaternary lithologies that comprise 14% of the study area, whereas dolomites, limestones, sandstones, dolomitized breccias and shales represent Triassic deposits (7%) [64–67]. Bednja Municipality and Lepoglava City are predominantly rural areas, i.e., poorly urbanized areas with low population densities. Ref. [19] described land-use as follows: artificial areas (3% of the study area), agricultural areas and low vegetation (24% of the study area), forest areas and high vegetation (73% of the study area) and water bodies (~0% of the study area). The two main water flows in the study area are the river Bednja and the Kamenica stream. Perennial and temporary water flows are 5.7 and 24.5 km in length, respectively. The climate of the area can be described as continental, with a mild maritime influence, with most rainfall occurring from May to November [68]. The closest meteorological station to the study area is 30 km west, which measured a mean annual precipitation of approximately 874 mm from

1949 to 2020 [69]. As identified by [20], more than 50% of landslides are located 25 to 75 m from roads, whereas <5% of landslides are found 0 to 25 m distance from buildings. Consequently, the necessity for landslide management and developing landslide susceptibility maps in the study area is of the utmost importance [16].

2.2. Landslide Data

As a first step to modelling the landslide susceptibility of the study area, a geomorphological landslide inventory map was interpreted (Figure 1) based on HR-DTM [19], created from a high-density LiDAR point cloud with 30 cm bare-earth point spacing [20]. The resulting 0.3m HR-DTM was suitable for mapping small and shallow landslides under vegetation [70–72] with high precision and accuracy [20,72]. The finalized geomorphological landslide inventory map, which included field verification, contained 912 phenomena with a density of 45 landslides per square kilometer and an average size of 448 m² [19,20] (Table 1). Generally, landslides can be categorized as small and shallow, mainly occurring in soil and soft rock.

Table 1. Comparison of statistical parameters for landslide and stable area inventories.

	Amount of Polygons (N)	Density (N/km ²)	Min. (m ²)	Max. (m ²)	Average (m ²)	Median (m ²)	Standard Deviation (m ²)	Total Size (km ²)
Landslide inventory	912	45	3.3	13,779	448	173	880	0.41
Stable area inventory	912	45	4.4	6986	421	194	662	0.38

Stable areas (Figure 1) were mapped using HR morphometric maps derived from HR DTM, depicting accurate environmental/morphometrical conditions. The inventory of the stable areas was mapped by the same group of landslide mappers as the landslide inventory, based on researchers' experience, and the following set of rules was considered: (i) define stable areas in all geological environments, (ii) uniformly distribute the stable areas throughout the entire study area and (iii) the finalized inventory should have statistical parameters similar to the landslide inventory map. Table 1 shows a comparison of the statistical parameters for stable and unstable inventories and Figure 2 illustrates the distributions of landslide centroids, centroids from mapped stable polygons, and randomly generated stable points in different slope classes, which are all needed for the definition of susceptibility analysis across eight scenarios. It can be concluded that the landslide inventory and the inventory of stable areas have the opposite trend considering the distribution of mapped features in slope classes, while randomly generated stable points follow the distribution of slope classes for the entire study area.

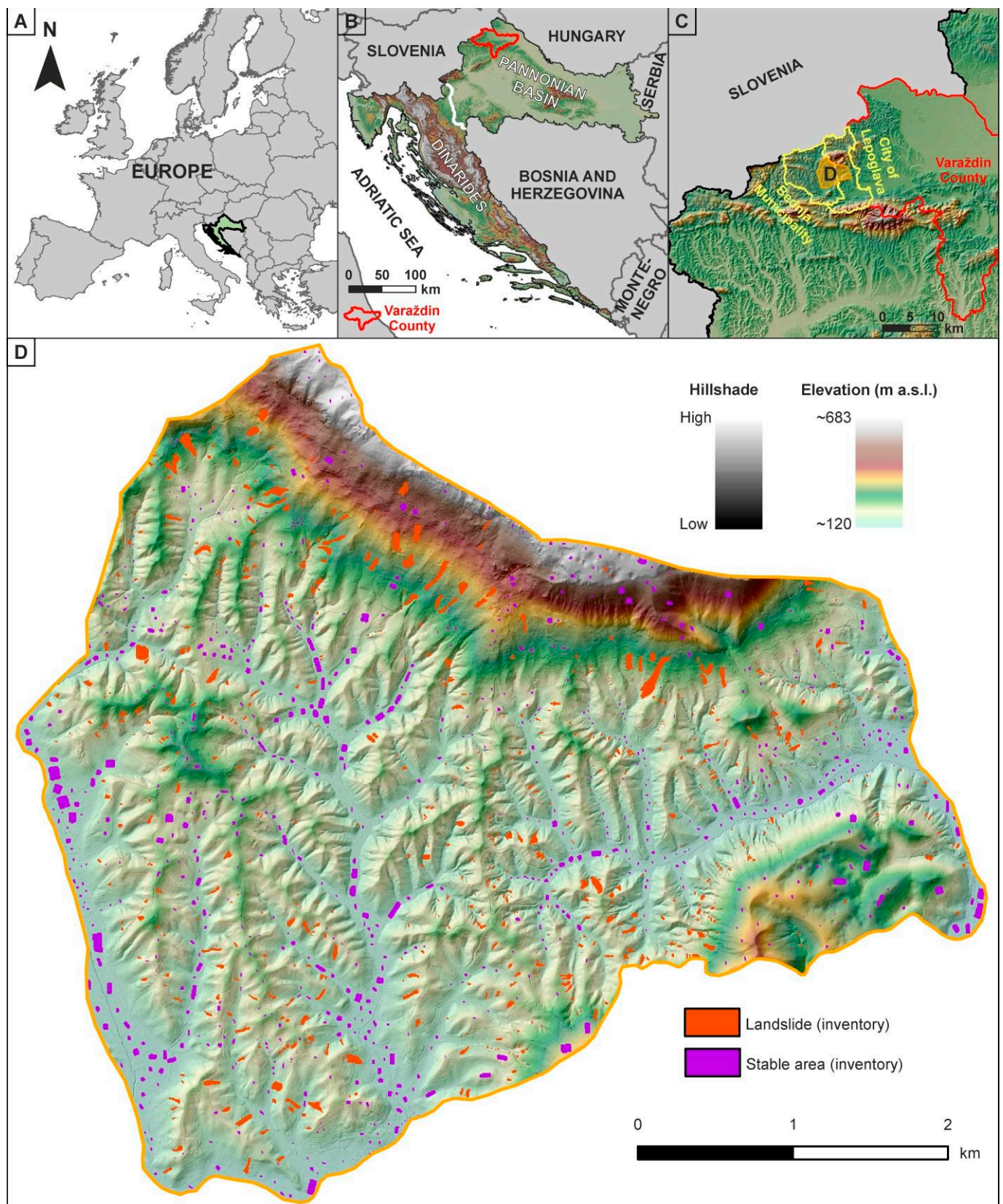


Figure 1. Study area location in Europe (A), Croatia (B), NW Croatia (C) and mapped landslide and stable area inventories in the study area (D).

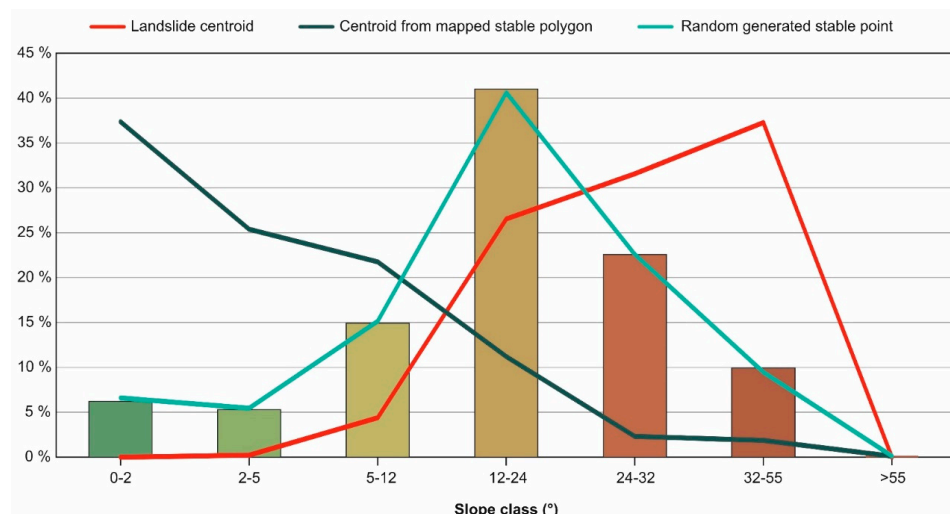


Figure 2. Distributions of landslide centroids, centroids from mapped stable polygons and randomly generated stable point in different slope classes.

2.3. Thematic Data

Following recommendations by [13], the geo-environmental information used for landslide susceptibility modelling in this study was selected considering the scale and the scope of the analysis and the data's quality, accuracy and relevance. Ref. [19] provided a detailed analysis of data availability for landslide hazard assessments in the study area, reaching the conclusion that HR LiDAR DTM and orthophoto images represent optimal input data to derive LCFs. Furthermore, in a preliminary susceptibility analysis [73], the application of spatially inaccurate LCFs resulted in a LSM with roughly 10% lower success and prediction Area Under the Curve (AUC) values than the model that used LCFs derived from HR remote sensing data. Considering the above, a 5 m-resolution LiDAR-based DTM was used to derive the geomorphological and hydrological LCFs used in this study. Furthermore, 0.3 m-resolution LiDAR-based HR DTM derivatives enabled the modification of the small-scale geological maps [64,67], as described in [19], following the methodology used by [74] and [75]. To achieve large-scale spatial accuracy, 0.5 m-resolution orthophoto images [76] served as a basis to derive anthropogenic LCFs, which were supplemented with a variety of different source files (see Figure 3 in [19]).

3. Methodology

3.1. General Workflow

The susceptibility modelling workflow in this study generally followed the interrelated steps described in [13]. As depicted in Figure 3, the workflow can be grouped into modelling and result analysis. Preparing input data consists of developing training and validation datasets needed for eight scenarios defined by different sampling strategies used for stable and unstable areas. It should be stated that in all 32 LSMs, a rigorous 50% ratio of training and validation data was applied, which is uncommon, as researchers more often choose a significantly larger training dataset (e.g., [32,37,55,58]). The reason for this rigorous approach is to test the model under the most unfavourable conditions, which we find necessary for a large-scale landslide susceptibility assessment. Moreover, the complete and representative landslide inventory and spatially accurate LCFs enable high model performance. The stable and unstable datasets required for the susceptibility modelling were derived from polygon-based landslide and stable area inventory maps. Similarly, a set of LCFs was prepared using either regular or smoothing sampling for application to eight scenarios. After finalizing the datasets, collinearity testing between LCF maps was performed, followed by the development of 32 LSMs using four machine learning methods in eight scenarios. Finally, the derived LSMs were analysed using appropriate metrics for fitting, predictive and verification performance, as well as measuring uncertainty [13].

and deriving landslide susceptibility zonation maps. As even high evaluation metrics can favour a scenario that has evident model misclassifications (i.e., opposed to reality) [32], we opted for the mentioned parameters, as well as qualitative assessments.

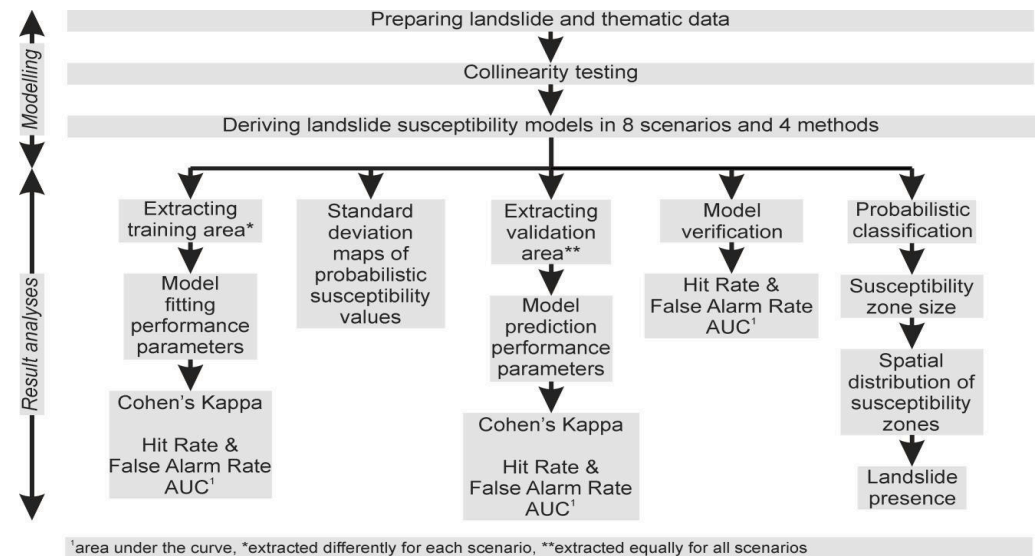


Figure 3. Workflow for applied landslide susceptibility modelling.

3.2. Preparing Stable and Unstable Datasets

Eight different modelling scenarios were defined and named based on a combination of sampling strategies for stable and unstable pixels or considering different LCF preparation procedures (Table 2). Namely, polygon-based landslide and stable area inventories were split into two equally large sets, defining the training and validation polygons. The eight training datasets were rasterized and used in the susceptibility analyses. Scenarios S1-PR_r (Figure 4A), S2-PR_s (Figure 4A), S4-PM_r (Figure 4C) and S5-PM_s (Figure 4C) sample unstable areas at the mapped landslide location, as illustrated in Figure 5A. Random stable pixels were generated in an equal number as unstable pixels to define scenarios S1-PR_r and S2-PR_s, opposite to scenarios S4-PM_r and S5-PM_s, which use the mapped stable pixels from the stable area inventory. To our knowledge, an experience-based stable area inventory has not been developed or utilized in landslide susceptibility modelling, even though a few researchers have attempted “non-susceptibility” zonation (e.g., [77]). Furthermore, scenarios S2-PR_s and S5-PM_s use smoothened DTM-derived LCFs (e.g., Figure 5C), i.e., elevation, slope, landform curvature, aspect, site exposure index and integrated moisture index. DTM was smoothened in ArcMap 10.8 by using the mean statistics type “Raster calculator” tool exclusively on the entire unstable pixel area defined by the landslide inventory map. It should be stated that the smoothing process can lead to different terrain changes due to settings in the “Raster calculator” and that several tests were conducted to achieve the desired effect in our study. Namely, the scale of the analysis and, consequently, pixel size and DTM quality, represent the parameters of interest. The smoothened DTM was further used to derive the previously stated LCFs to simulate a proxy of the terrain conditions prior to the occurrence of a landslide (i.e., undisturbed morphological settings [42]). Scenario S3-BR_r (Figure 4B) uses buffered unstable areas (Figure 5B) with randomly generated stable pixels. Similarly, scenario S6-BbM_r (Figure 4D) is defined by buffered unstable areas and buffered stable areas. For scenarios S3-BR_r and S6-BbM_r, the aim of sampling buffered areas is to train the model on undisturbed terrain, i.e., avoiding the direct influence of modified terrain conditions. The buffering distance on the polygon is set to 6.25 m, i.e., 1.25 times the pixel resolution (5 m). Scenarios S7-CR_r (Figure 4E) and S8-CcM_r (Figure 4F) use landslide centroids as unstable areas, whereas the stable pixels in scenario S7-CR_r are defined randomly. Stable area polygons are

transformed into centroids and used as stable pixels in scenario S8-CcM_r. It should be stated that the randomly generated stable pixels needed for scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r were generated throughout the entire study area, excluding only the area already selected to have unstable training pixels. As a result, an unbiased approach was ensured by allowing random stable pixels to be generated even in unstable validation pixels. The extent of the described model training for eight sampling scenarios is summarized with abbreviations and quantified by the number of pixels for unstable and stable areas in Table 2. From 456 pixels in scenarios S7-CR_r and S8-CcM_r to 10,099 pixels in scenario S3-BR_r, all eight scenarios have a similar or equal amount of unstable and stable pixels.

Table 2. Defined model training extents for eight scenarios.

Model Training Extents								
Scenario	Unstable Area			Stable Area			DTM-Derived LCFs	Abbr. *
	Sampling Type	Abbr. *	Pixel (N)	Sampling Type	Abbr. *	Pixel (N)		
S1	Landslide polygon	P	7793	Randomly generated stable point	R	7793	Regular	_r
	A reference point scenario, using commonly used landslide polygon sampling and random stable points							
S2	Landslide polygon	P	7793	Randomly generated stable point	R	7793	Smooth	_s
	A novel scenario testing unstable area sampling by smoothing DTM derived LCFs to simulate terrain conditions prior to landslide occurrence, i.e., capturing undisturbed terrain conditions at landslide locations							
S3	Buffer zone on landslide polygon	B	10,099	Randomly generated point	R	10,067	Regular	_r
	An alternative to scenario S2, using previously researched buffer zones to capture terrain conditions not influenced by landslide presence							
S4	Landslide polygon	P	7793	Mapped stable polygon	M	7543	Regular	_r
	A novel scenario testing stable area sampling by sampling mapped stable polygons and using unstable landslide polygons as a reference point							
S5	Landslide polygon	P	7793	Mapped stable polygon	M	7543	Smooth	_s
	Testing unstable areas as in scenario S2 with an addition of novel mapped stable polygons							
S6	Buffer zone on landslide polygon	B	10,099	Buffer zone on mapped stable polygon	bM	9573	Regular	_r
	Testing unstable areas as in scenario S3 with an addition of novel mapped stable polygons							
S7	Landslide centroid	C	456	Randomly generated stable point	R	457	Regular	_r
	A reference point scenario, using commonly used landslide centroid sampling and random stable points							
S8	Landslide centroid	C	456	Centroid from mapped stable polygon	cM	456	Regular	_r
	A novel scenario testing stable area sampling by sampling centroids from mapped stable polygons and using unstable landslide centroid as a reference point							

* Abbreviation.

The extent of validation is equal for all eight scenarios, and it consists of the second half of the landslide and stable areas polygons, which were excluded from all model training scenarios (Figure 4G). In other words, the extent of the validation dataset was unseen when training the model in all eight scenarios and is based on mapped locations for both stable and unstable areas, thus providing a uniform measurement of predictive performance for a valid comparison of results. It contains 8543 and 7773 pixels created by the rasterization of 456 mapped landslide and stable area polygons, respectively. Having nearly identical amounts of stable and unstable pixels distributed throughout the entire study area, the extent of validation provides satisfactory information about the predictive performance of the landslide susceptibility model as described by Cohen's Kappa index and AUC. Moreover, the mapped polygons that defined the extent of validation represent accurate yet heuristically defined environment conditions with high spatial accuracy in

relation to stable and unstable areas. As a result, an unbiased approach is ensured, meaning that the comparison is objective due to the uniform approach used for all eight scenarios.

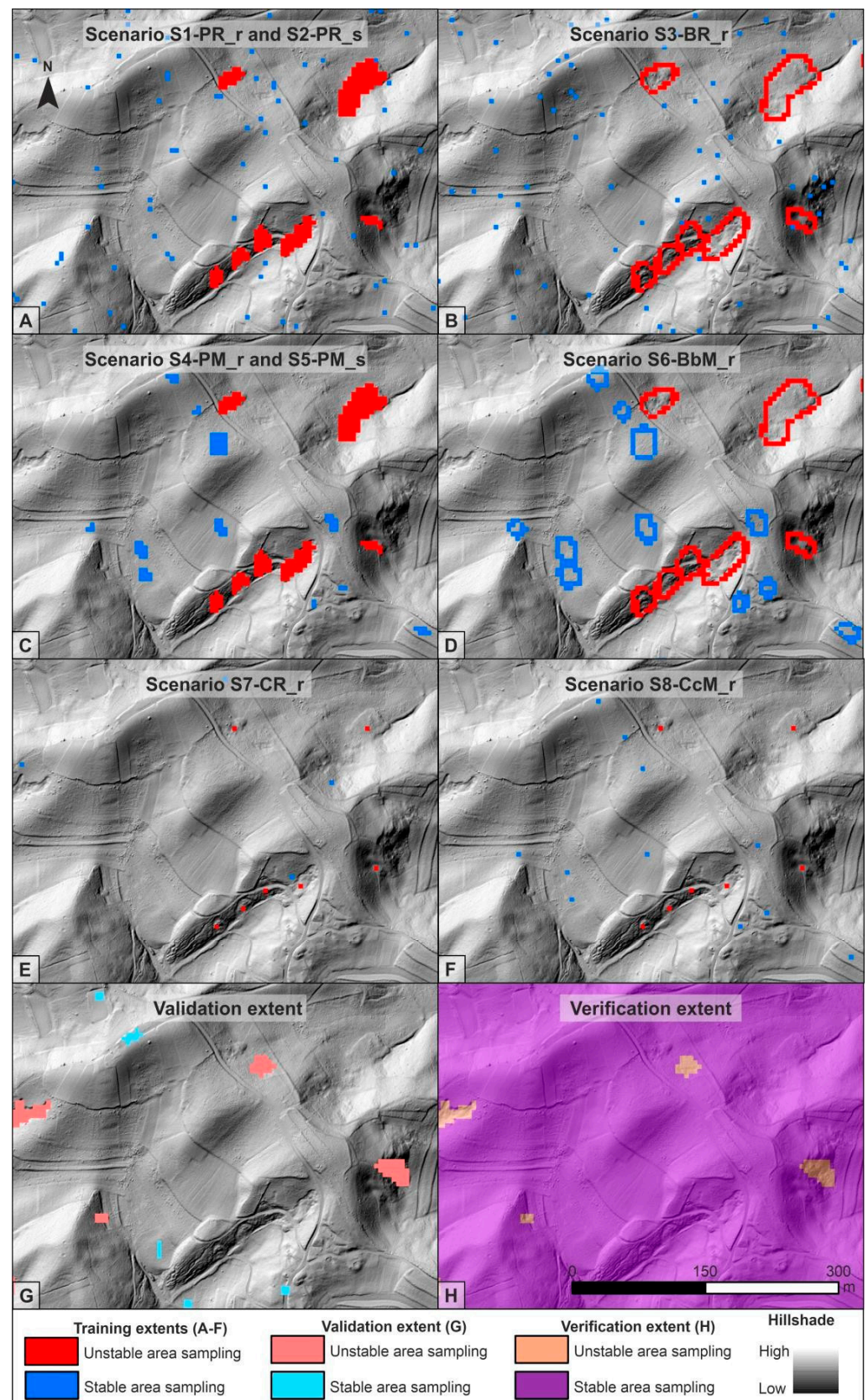


Figure 4. Close up examples for stable and unstable pixel sampling in training, validation and verification.

Model verification was also calculated to equal extents in all models by using AUC to examine the validation landslide polygons as unstable pixels (8543 pixels) and the remaining study area as stable pixels (800,267) (Figure 4H). The purpose of this is to quantify the performance of model agreement on unseen landslide data, but doing so on the entire study area, i.e., 808,810 pixels, unlike the predictive performance, which examines the extent of validation on 16,316 pixels. This metric provides necessary insight into the agreement performance of the entire study area, which is necessary for a large-scale landslide susceptibility assessment.

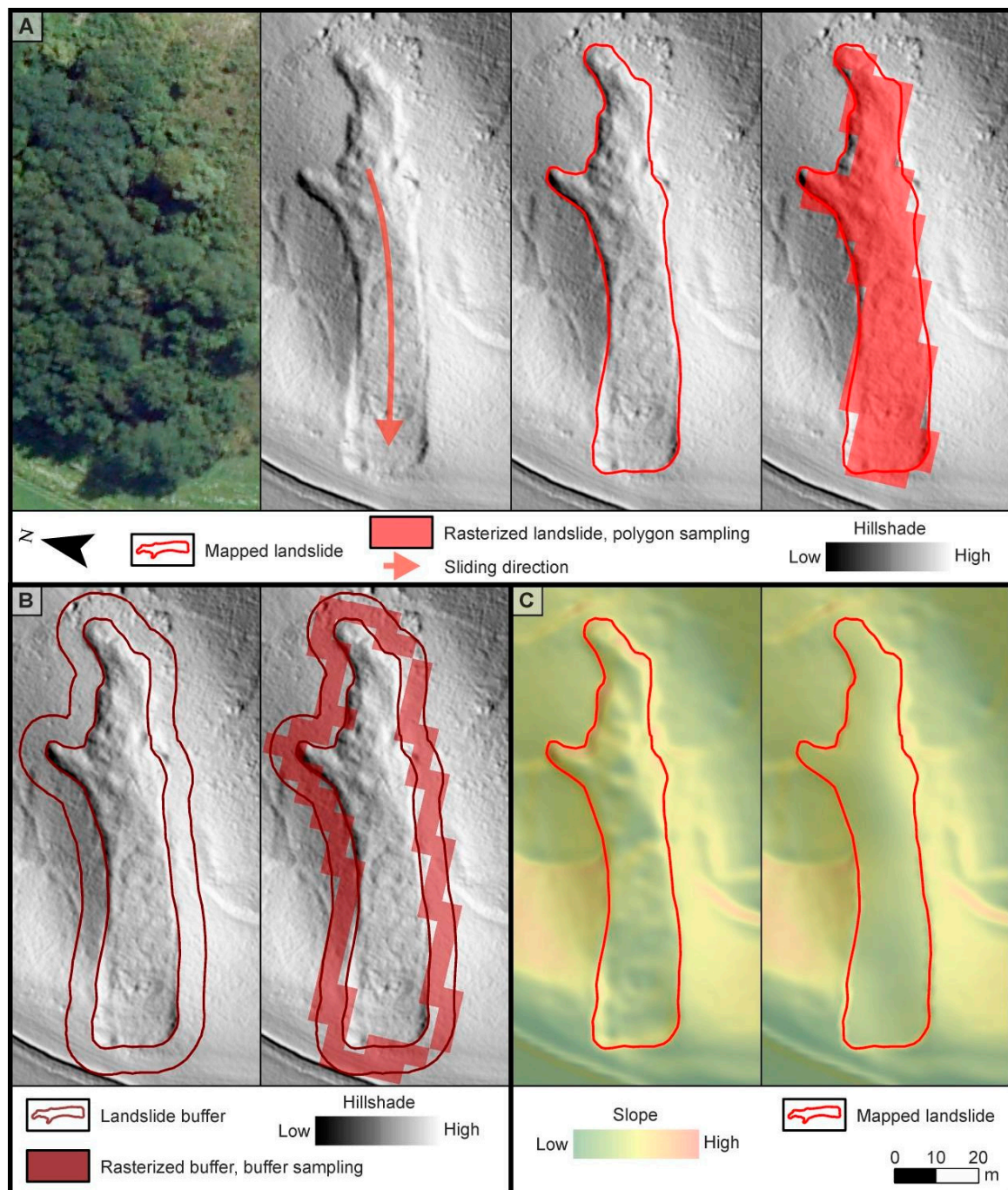


Figure 5. Examples of sampling strategies: (A) polygon sampling, (B) buffer sampling and (C) smoothed LCF sampling.

3.3. Preparing Thematic Data

Considering the thematic data, i.e., LCFs, a set of 12 maps grouped as geomorphological, geological, hydrological and anthropogenic is prepared from the source data by using the methods or tools described in Table 3. The LiDAR point cloud stands out as the most significant source data as it enables the derivation of a DTM (elevation LCF), which was further used as source data for all geomorphological and hydrological LCFs. Moreover, the LiDAR point cloud was used in anthropogenic LCFs and to derive a HR-DTM, which was necessary for the modification of geological LCFs. It should be stated that stretched rasters, i.e., elevation, slope, landform curvature, aspect, site exposure index and integrated moisture index were used in susceptibility analysis without a classification method, i.e., as stretched rasters, whereas the 5 m buffer zone intervals derived from line vectors resulted in a sufficient number of classes and maximally simulate continuous behaviour [78]. The spatial distribution of the stretched raster LCF values, LCFs defined by buffer zones (proximity to engineering formations, drainage network, traffic infrastructure and land-use contact) and classes of categorical LCFs (engineering formations and land-use) are illustrated in Figure 6. All LCFs were processed to an equal number of pixels (i.e., rows and columns) to a 5 m resolution.

Table 3. Overview of source data and methodology to derive landslide conditioning factors.

Landslide Conditioning Factor		Source Data	Obtained by
Geomorphological	Elevation	LiDAR point cloud (class 2, bare earth)	Interpolation
	Landform curvature		ArcGIS 10.8 Landform curvature tool [79]
	Aspect		ArcGIS 10.8 Aspect tool
Geological	Engineering formations	Croatian Basic Geological Maps [64,67], HR-LiDAR DTM	Digitization, visual interpretation of HR-LiDAR DTM derivatives
	Proximity to engineering formations	Engineering formations	ArcGIS 10.8 Multiple Ring Buffer tool
Hydrological	Proximity to drainage network	Elevation	ArcGIS 10.8 Spatial Analyst Toolbox
	Site exposure index		ArcGIS 10.8 Site Exposure Index tool [79]
	Integrated moisture index		ArcGIS 10.8 Integrated Moisture Index tool [79]
Anthropo-genic	Land-use	Digital orthophoto, LiDAR point cloud, HR-LiDAR DTM, Open Street Map, Land-use planning maps	see [19] Figure 3
	Proximity to traffic infrastructure	Roads input data (see [19] Figure 3)	ArcGIS 10.8 Multiple Ring Buffer tool
	Proximity to land-use contact	Land-use	ArcGIS 10.8 Multiple Ring Buffer tool

Table 4 depicts an overview of continuous (i.e., stretched rasters and line vectors) and categorical LCFs, including a comparison of regular and smooth sampled LCFs. Namely, the difference in minimum, maximum, mean and standard deviation statistical parameters between regular and smoothed LCFs is either not present or insignificant. This is not surprising because <2% of the study area was smoothed. It should be noted that the geological and anthropogenic LCFs do not consist of DTM-derived LCFs that were smoothed. Deriving 5 m buffer zones for the LCFs, which are defined by their proximity to line vectors, resulted in a maximum of 260 and 585 m zones for proximity to traffic infrastructures and land-use contact anthropogenic LCFs, respectively. Proximity to the drainage network is defined with 32 classes, i.e., reaching a maximum distance of 160 m, whereas the proximity to engineering formation LCF is defined with 104 classes and a maximum distance of 520 m.

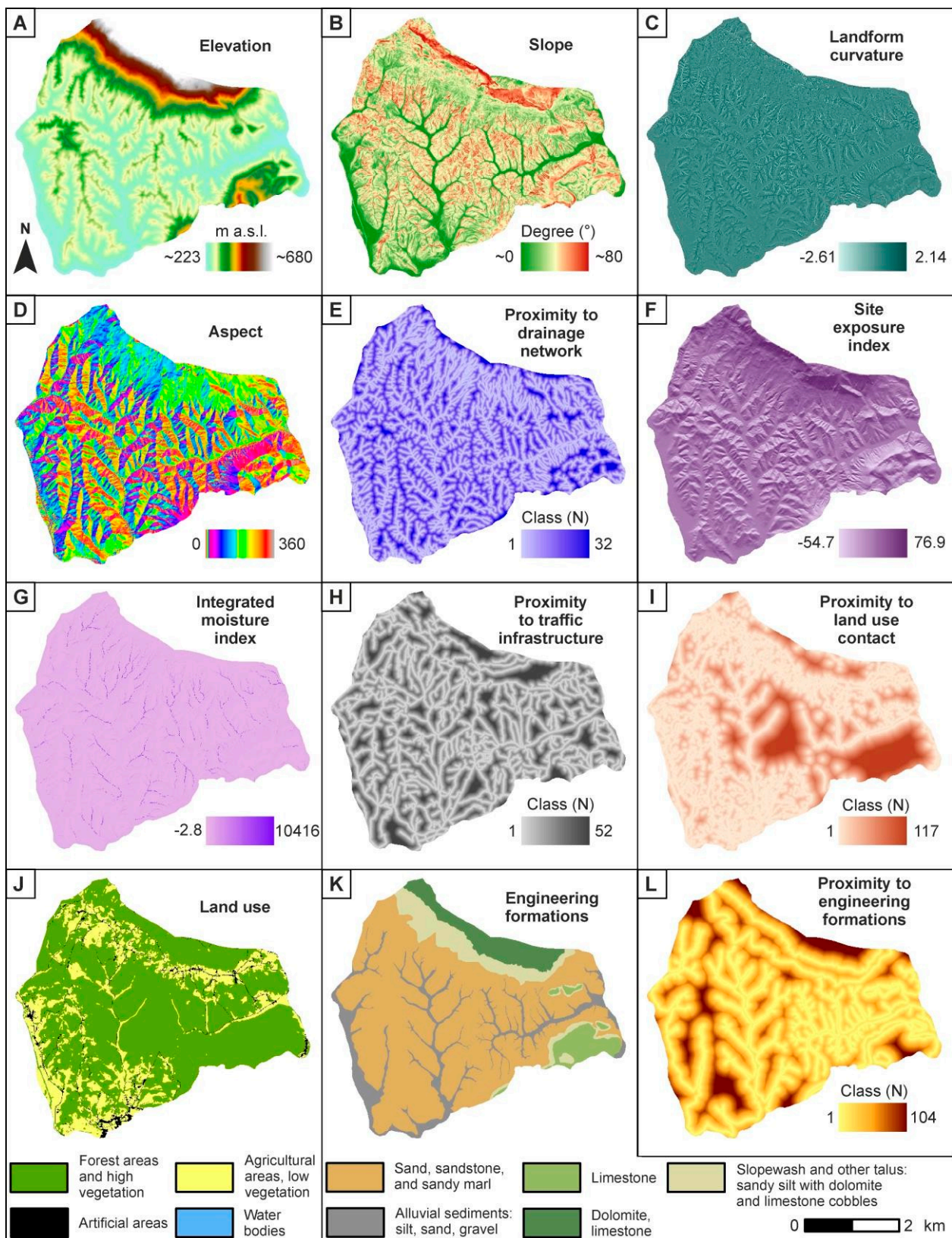


Figure 6. Landslide conditioning factors used in the susceptibility modelling: (A) elevation, (B) slope, (C) landform curvature, (D) aspect, (E) proximity to drainage network, (F) site exposure index, (G) integrated moisture index, (H) proximity to traffic infrastructure, (I) proximity to land use contact, (J) land use, (K) engineering formations and (L) proximity to engineering formations.

Table 4. Categorization and statistical parameters for landslide conditioning factors.

Landslide Conditioning Factors (LCFs)								
	Name	Continuous LCFs						Categorical LCFs
		Stretched Raster				Line Vector		Classes (N)
		Min.	Max.	Mean	St. Dev.	Interval (m)	Classes (N)	
Geomorphological	Elevation ^R	222.5	679.8	304.4	77.3			
	Elevation ^S	222.5	679.8	304.4	77.3			
	Slope ^R	0	80.3	18.9	10.2			
	Slope ^S	0	80.3	18.9	10.2			
	Landform curvature ^R	−2.6	2.1	2.1	0.1			
	Landform curvature ^S	−2.6	2.1	2.0	0.1			
	Aspect ^R	0	360	183.8	89.6			
	Aspect ^S	0	360	183.8	89.6			
Geological	Engineering formations							5
	Proximity to engineering formations					5	104	
Hydrological	Proximity to drainage network					5	32	
	Site exposure index ^R	−55	77	2.9	14.8			
	Site exposure index ^S	−55	77	2.9	14.8			
	Integrated moisture index ^R	−2.8	10,416	115.2	248.0			
	Integrated moisture index ^S	−2.8	10,416	114.3	242.0			
	Land-use							4
Anthropogenic	Proximity to traffic infrastructure					5	52	
	Proximity to land-use contact					5	117	

^R indicates regular stretched raster used in scenarios S1-PR_r, S3-BR_r, S4-PM_r, S6-BBM_r, S7-CR_r and S8-CcM_r;

^S indicates smoothened stretched raster used in scenarios S2-PR_s and S5-PM_s.

3.4. Modelling

After the development of relevant stable and unstable datasets and the thematic information according to the inter-related steps proposed by [13], an appropriate mapping unit was selected based on the preliminary susceptibility modelling conducted in [80]. In a preliminary study, 5 m-resolution pixels were selected over the slope units to adequately represent spatial conditions in large-scale susceptibility modelling. After the selection of the appropriate mapping unit, the prepared LCFs were tested for collinearity considering Pearson's *R* absolute value. An absolute value of 0.5 for Pearson's coefficient was selected as a cut-off value to exclude LCFs showing collinearity. Out of 12 previously described LCF maps, none showed collinearity and all of them were considered to be a complete set of LCFs that could be further used in the analysis. A further step was the statistical modelling of landslide susceptibility. Multiple studies (e.g., [61,62,81,82]) prove that statistical methods differ in terms of certain advantages and disadvantages, resulting in different LSMs, as shown by the comparison metrics. For that reason, and to provide new insight into large-scale susceptibility modelling with different methods, LR, NN, RF

and SVM methods were used for landslide susceptibility modelling. The modelling was performed in MATLAB version 9.10.0.1602886 [83] software. Specifically, the “Statistics and Machine Learning Toolbox” was used [84], where “fitlinear”, “fitcnet”, “fitcensemble” and “fitsvm” functions relate to LR, NN, RF and SVM classifiers, respectively [85]. The readers are referred to [78] for details about method application, whereas [61] explain the theoretical background of machine learning algorithms. The four statistical methods in combination with eight sampling scenarios resulted with 32 different LSMs.

3.5. Comparison Metrics for Results Analysis

The previously explained sampling strategies define the training extent on which the fitting performance is measured in each scenario. Namely, the training extents differ due to the use of different sampling strategies but nonetheless keep an (nearly) equal amount of stable and unstable pixels, enabling the evaluation of the classification performance in the confusion matrix. For a uniform comparison of the 32 landslide susceptibility models, the validation and verification extents are constant. On the defined training, validation and verification extents, probabilistic susceptibility values are observed considering the 0.5 threshold to define a confusion matrix. As a result, the stable sampling pixels classified as unstable or stable, and the unstable sampling pixels classified as stable or unstable [86] are identified in each observed extent. The four variables are further used to calculate Cohen’s Kappa index [87] and to construct a Receiver Operating Characteristic (ROC) curve to determine the AUC values [88]. Both metrics are commonly used in landslide susceptibility modelling to describe fitting and predictive performance (e.g., [58,62,81,89–91]). Namely, Cohen’s Kappa values of < 0.2 can be considered to be of slight and poor agreement, unlike > 0.8 values, which indicate an almost perfect agreement [87], whereas 0.5 AUC values indicate random prediction, contrary to 1.0 values that suggest a perfect fit [92]. LCF importance was tested using a leave-one-out test and quantified via AUC and Cohen’s Kappa fitting performance values.

Moreover, to test modelling variability, 12 standard deviation (SD) maps of probabilistic susceptibility values were derived considering the eight sampling scenarios and four applied statistical methods. Uncertainty metrics can be used to identify changes in probabilistic susceptibility values for each method throughout the eight scenarios and for each scenario considering the four statistical methods. Namely, four SD maps describe the uncertainty for each method based on probabilistic values in each of the eight sampling scenarios, whereas probabilistic values in the four applied methods define eight SD maps for each of the eight sampling scenarios. As a result, the uncertainty metric provides insight into method and scenario stability and applicability, i.e., sensitivity to specific modelling settings. SD maps were classified in 0.1 intervals, i.e., 0.0–0.1, 0.1–0.2, 0.2–0.3, 0.3–0.4 and 0.4–0.5 classes. Low SD values, i.e., the 0.0–0.1 class, indicate minimum uncertainty, opposite to the high uncertainty, defined by the 0.4–0.5 SD class.

The zonation of probabilistic susceptibility models ranging from 0.0 to 1.0 was performed by using 0.0–0.2 (very low), 0.2–0.45 (low), 0.45–0.55 (medium), 0.55–0.8 (high) and 0.8–1.0 (very high) intervals [58]. Furthermore, landslide presence and zone sizes were examined and quantified for the derived zones.

4. Results

4.1. Model Fitting Performance

Cohen’s Kappa and AUC metrics, which describe the fitting performance of the 32 derived LSMs by applying LR, NN, RF and SVM methods in eight sampling strategies, are shown in Figure 7. Models derived using the RF method show perfect agreement in all eight scenarios by having 1.0 Cohen’s Kappa and AUC values. In other words, all of studied stable and unstable training pixels studied had probabilistic susceptibility values of < 0.5 and > 0.5 , respectively. Generally, across all eight scenarios, LR and SVM methods differed minimally. Those two methods showed lower fitting performance compared to the NN method by having approx. 0.1 to 0.15 lower Cohen’s Kappa and AUC values

in scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r. In scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r the difference is extremely low in terms of AUC (i.e., <0.025) and low in terms of Cohen's Kappa (i.e., 0.05 to 0.15). The models were shown to be more sensitive to the sampling strategy of stable areas, as suggested in the differences between scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r, with one group where the stable areas are randomly generated and the second group defined with scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r, where stable areas are mapped. The mapped stable area group resulted in a Cohen's Kappa of approx. 0.4 and AUC values approx. 0.25 higher in scenarios S4-PM_r, S5-PM_s and S6-BbM_r compared to scenarios S1-PR_r, S2-PR_s and S3-BR_r. Centroid sampling resulting from scenarios S7-CR_r and S8-CcM_r showed lower differences, but they were still significant, following the same trend with a higher Cohen's Kappa and AUC values for the mapped stable area scenario. Differences between polygon, smoothing and buffering sampling strategies proved to be of low relevance to the model fitting performance due to similar performance between scenarios S1-PR_r, S2-PR_s and S3-BR_r considering unstable areas, and S5-PM_s, S6-BbM_r and S7-CR_r considering stable areas. The NN method is the only exception in scenario S3-BR_r, defined by the buffering sampling strategy as having somewhat poorer performance. Furthermore, scenario S7-CR_r, defined by centroid sampling, showed better fitting performance than scenarios S1-PR_r, S2-PR_s, and S3-BR_r, which were defined by the polygon sampling of unstable areas. LR, NN and SVM methods show better agreement in scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r whose common characteristic is using mapped stable areas. Concretely, LR and SVM resulted in Cohen's Kappa values of approx. 0.8 and AUC values of 0.95–0.975, whereas the NN method performed even slightly better. Scenario S6-BbM_r showed minimally lower results than scenarios S4-PM_r and S5-PM_s, unlike S8-CcM_r, which was defined by centroid sampling and performed better.

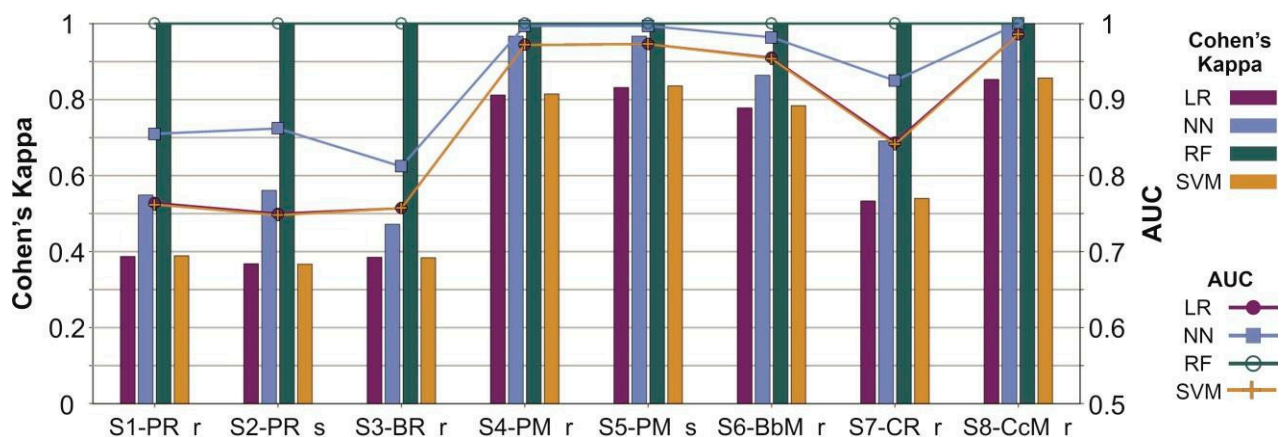


Figure 7. Cohen's Kappa and AUC fitting metrics for 32 derived landslide susceptibility models.

To estimate the influence of the LCF in the susceptibility analysis, a leave-one-out test was performed for the SVM method in scenarios S1-PR_r and S4-PM_r, representing the two opposite groups of scenarios. For both models, Cohen's Kappa and AUC were determined to measure the changes in fitting performance (Table 5). By comparing the leave-one-out test results with the values in Figure 7, we note that only a few LCFs have a significant influence. Namely, the absence of the engineering formation LCF in scenario S1-PR_r causes a drastic decline in fitting performance, followed by moderate decline when leaving out the proximity to the drainage network LCF. Slope and engineering formations are the most significant LCFs in scenario S4-PM_r, whereas the absence of either still enables the model to achieve high Cohen's Kappa and AUC values. Generally, the absence of most of the LCFs minimally changes model fitting performance.

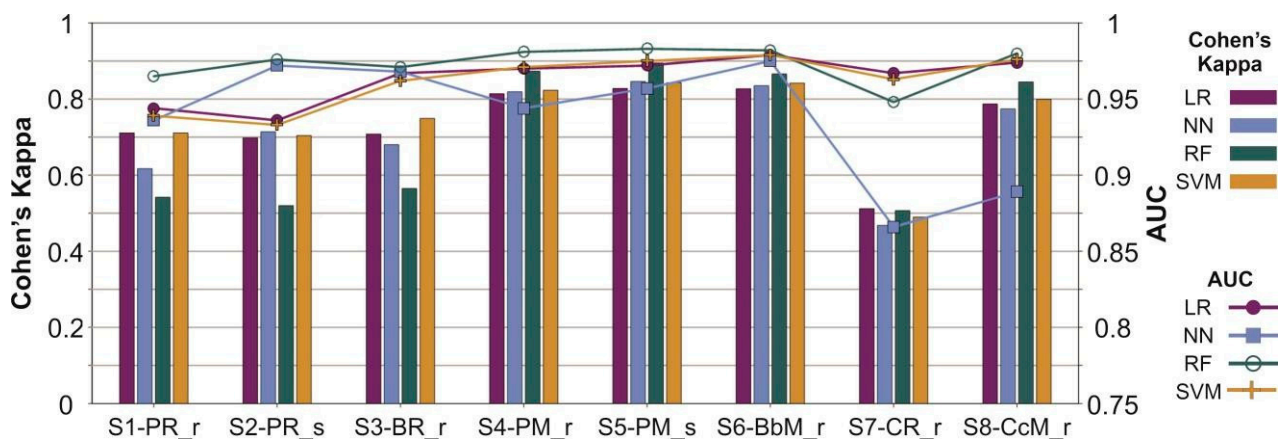
Table 5. Leave-one-out test results conducted for S1-PR_r and S4-PM_r scenarios using SVM method.

Excluded LCF	S1-PR_r (SVM Method)		S4-PM_r (SVM Method)	
	Cohen's Kappa	AUC	Cohen's Kappa	AUC
Elevation	0.392	0.761	0.813	0.972
Slope **	0.376	0.756	0.725	0.939
Landform curvature	0.365	0.752	0.812	0.971
Aspect	0.388	0.762	0.816	0.971
Engineering formations *	0.342	0.715	0.807	0.950
Proximity to engineering formations	0.381	0.755	0.815	0.972
Proximity to drainage network	0.356	0.739	0.817	0.959
Site exposure index	0.383	0.760	0.815	0.969
Integrated moisture index	0.389	0.762	0.815	0.972
Land-use	0.377	0.756	0.805	0.967
Proximity to traffic infrastructure	0.386	0.762	0.814	0.972
Proximity to land-use contact	0.384	0.758	0.814	0.967

* and ** indicate the most significant LCFs in scenario S1-PR_r and S4-PM_r, respectively.

4.2. Model Predictive Performance

When examining model predictive performance (Figure 8), extremely high AUC values in the range of approx. 0.925 to 0.975 are present in all methods and scenarios, except for the models derived from the NN method in scenarios S7-CR_r and S8-CcM_r, which were defined by centroid sampling. On the other hand, Cohen's Kappa metric indicates very high agreement in scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r, with roughly >0.8 values, whereas scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r show values of around 0.7 or lower. The lowest results when considering Cohen's Kappa are identified in scenario S7-CR_r at around 0.5 in all four methods, followed by 0.55 values in the RF method in scenarios S1-PR_r, S2-PR_s and S3-BR_r. LR, NN and SVM models in scenarios S1-PR_r, S2-PR_s and S3-BR_r have approx. 0.7 Cohen's Kappa values, except in a model derived from the NN method in the S1-PR_r scenario, with a Cohen's Kappa value of 0.6. Generally, the 32 derived LSMs showed excellent predictive performance. Due to the extent of validation, similar to fitting performance, high agreement is found in scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r, whose stable area sampling is based on the stable inventory. Lastly, the extreme differences in the centroid sampling strategy in scenario S7-CR_r should be noted, as well as the importance of studying two quantitative metrics (i.e., Cohen's Kappa and AUC).

**Figure 8.** Cohen's Kappa and AUC predictive metrics for 32 derived landslide susceptibility models.

4.3. Probabilistic-Based Zonation

By observing zone area sizes (Figure 9) based on a probabilistic approach and landslide presence in each zone (Figure 10), two groups of similar results across the scenarios are

identified as follows: Scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r make up one group of similar results, where very low and low zones are most present in the study area (>55%), followed by smaller area of medium and high zones and, lastly, an extremely small area of very high zones. A nearly equal distribution with approx. 70% landslide presence in very high and high zones is evident, except in scenario S7-CR_r, where the amount is roughly 50%. Consequently, scenario S7-CR_r has a significant number of landslides in the low-susceptibility zone (e.g., 30%). Scenario S4-PM_r has an extremely expressed very high zone (approximately 50%), followed by S5-PM_s, S8-CcM_r and S6-BbM_r, which comprise the second group. Very low zones are similar in size to group one, unlike low, moderate and high zones, which shrunk significantly. Landslides in the very high zone are present up to 90%, generally leaving very low and low zones with an insignificant number of landslides. This is highly expressed in scenarios S4-PM_r and S5-PM_s, and somewhat less in scenarios S6-BbM_r and S8-CcM_r.

Considering differences between LR, NN, RF and SVM methods, LR and SVM can be described as similar by showing less polarized susceptibility zones and landslide area presence. Concretely, LR and SVM models have equally expressed very low and low zones, unlike NN and RF, which have more area presence in the very low zone.

All methods depict a moderate zone 5–10% in size; however, exceptions are visible in scenarios S4-PM_r, S5-PM_s and S6-BbM_r, where the zone makes up <5% of the study area. The NN method has the largest very high zone in each scenario, followed by RF and LR. Moreover, the NN method demonstrated extreme polarization in scenario S8-CcM_r by depicting only very low and very high susceptibility zones. RF and NN methods result in nearly identical landslide presence in scenarios S4-PM_r, S5-PM_s and S6-BbM_r, whereas RF results in significantly more landslides in scenarios S1-PR_r and S2-PR_s. On the other hand, SVM and LR have only around 10% landslide presence in very high zones in scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r, indicating way higher landslide presence in high zones. Furthermore, the differences among methods in each scenario are the highest in terms of landslide presence between very high and high susceptibility zones, leaving very low, low and moderate zone areas relatively stable in each scenario. Similarly, the cumulative area of very high and high zone sizes is nearly equal in each scenario, i.e., among all methods, with the exception for RF in scenarios S1-PR_r and S2-PR_s. The LSMs classified for eight scenarios and four methods are represented in the Supplementary Material in Figures S1–S8.

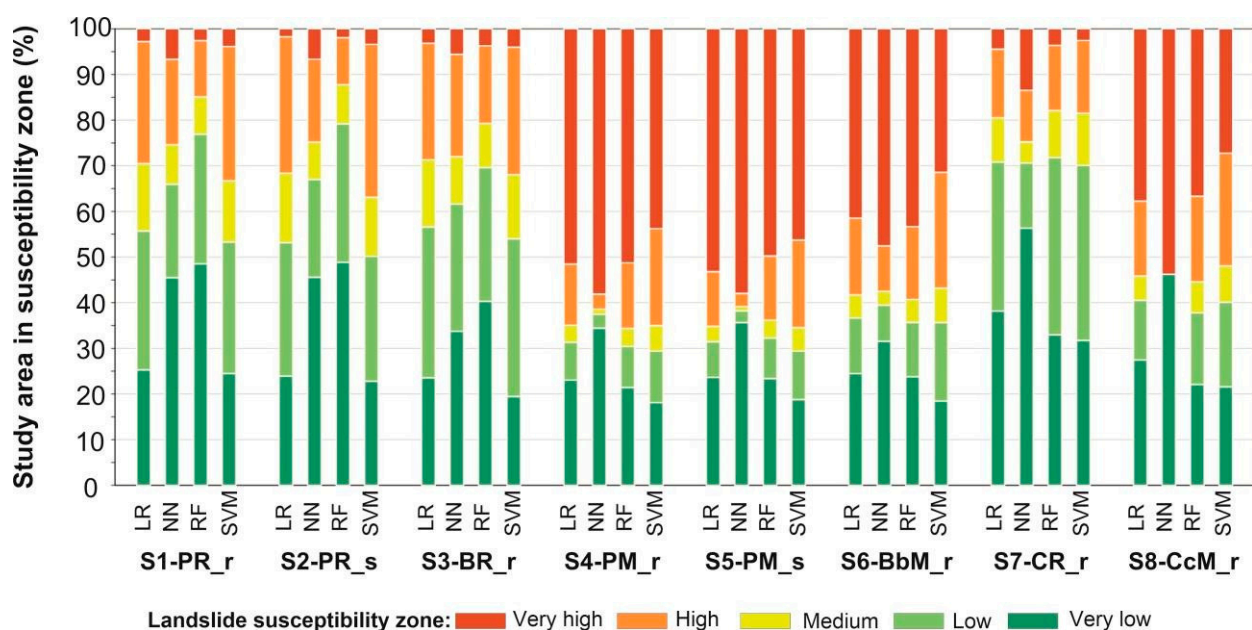


Figure 9. Landslide susceptibility zone area size based on probabilistic zonation.

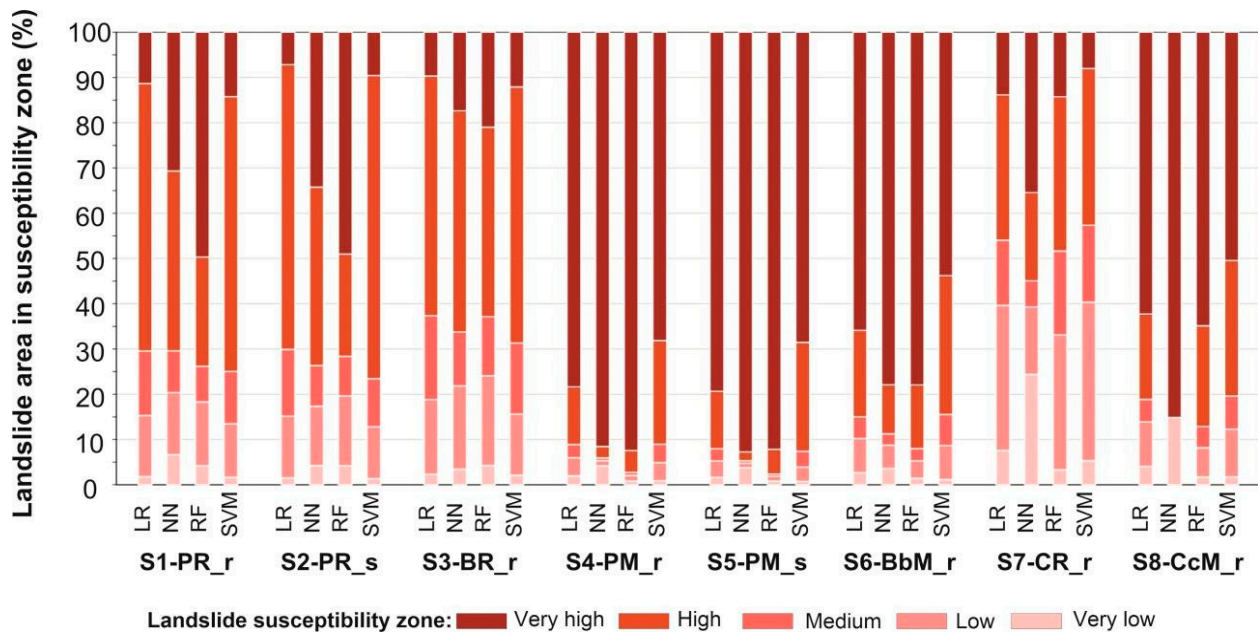


Figure 10. Landslide area presence in susceptibility zones based on probabilistic zonation.

5. Discussion

5.1. Model Uncertainty Considering Applied Methods and Sampling Scenarios

Considering the variability in the probabilistic susceptibility values from the derived LSMs, Figure 11 illustrates the spatial distribution and area size of the SD classes for eight sampling scenarios (Figure 11A–H) and four susceptibility methods (Figure 11I–L). Scenarios with random stable pixels showed lower deviations, i.e., S1-PR_r, S2-PR_s, S3-BR_r, and S7-CR_r, which generally account for >90% of the study area in 0.0–0.2 standard deviation classes, unlike scenarios S4-PM_r, S5-PM_s and S8-CcM_r, which account for approx. 20 or more percent of the study areas in >0.2 standard deviation classes, indicating extreme variability. It should be noted that scenario S6-BbM_r, where mapped stable pixels are used, showed greater similarity to scenarios S1-PR_r, S2-PR_s, S3-BR_r and S7-CR_r, indicating the positive influence of using unstable buffer zone pixels. Similarly, scenario S3-BR_r, which also uses buffer areas for unstable pixels, has significantly more study areas defined with the lowest SD values compared to scenarios S1-PR_r and S2-PR_s. Moreover, the difference between sampling unstable areas on regular (S1-PR_r and S4-PM_r) or smoothed (S2-PR_s and S5-PM_s) rasters resulted in negligible differences. The point sampling of unstable pixels did not significantly influence the variability, i.e., scenarios S1-PR_r and S7-CR_r showed minimal differences. On the other hand, the point sampling of unstable pixels (S8-CcM_r) increases variability, unlike the polygon sampling of unstable pixels (S4-PM_r), which showed less variability. Considering the spatial distribution, lowlands have the lowest SD values, whereas the northern part of the study area, narrowly stretching from west to east, is defined as having the highest standard deviation values, which is likely due to slope wash and other talus classes in the engineering formation LCF.

The SVM method showed the least sensitivity to sampling strategies by relating to >80% of the study area in the <0.2 standard deviation classes (Figure 11L), followed by LR, covering approx. 10% more of the study area in the 0.2–0.3 SD class (Figure 11I). SVM and LR have negligible areas present, with SD values of >0.3, unlike the RF method, which covers approx. 18% of the study area in the 0.3–0.4 SD class, mainly present in the wider-center area (Figure 11K). With >60% study area covered by >0.2 SD values, the NN method shows extreme variability in terms of sampling strategies (Figure 11J). Namely, only alluvial sediment areas and the southeastern area depicted by the limestone geological engineering unit represent low SD values when using the NN method, whereas all four methods show minimal SD values in terms of the engineering geological unit of the alluvial sediments.

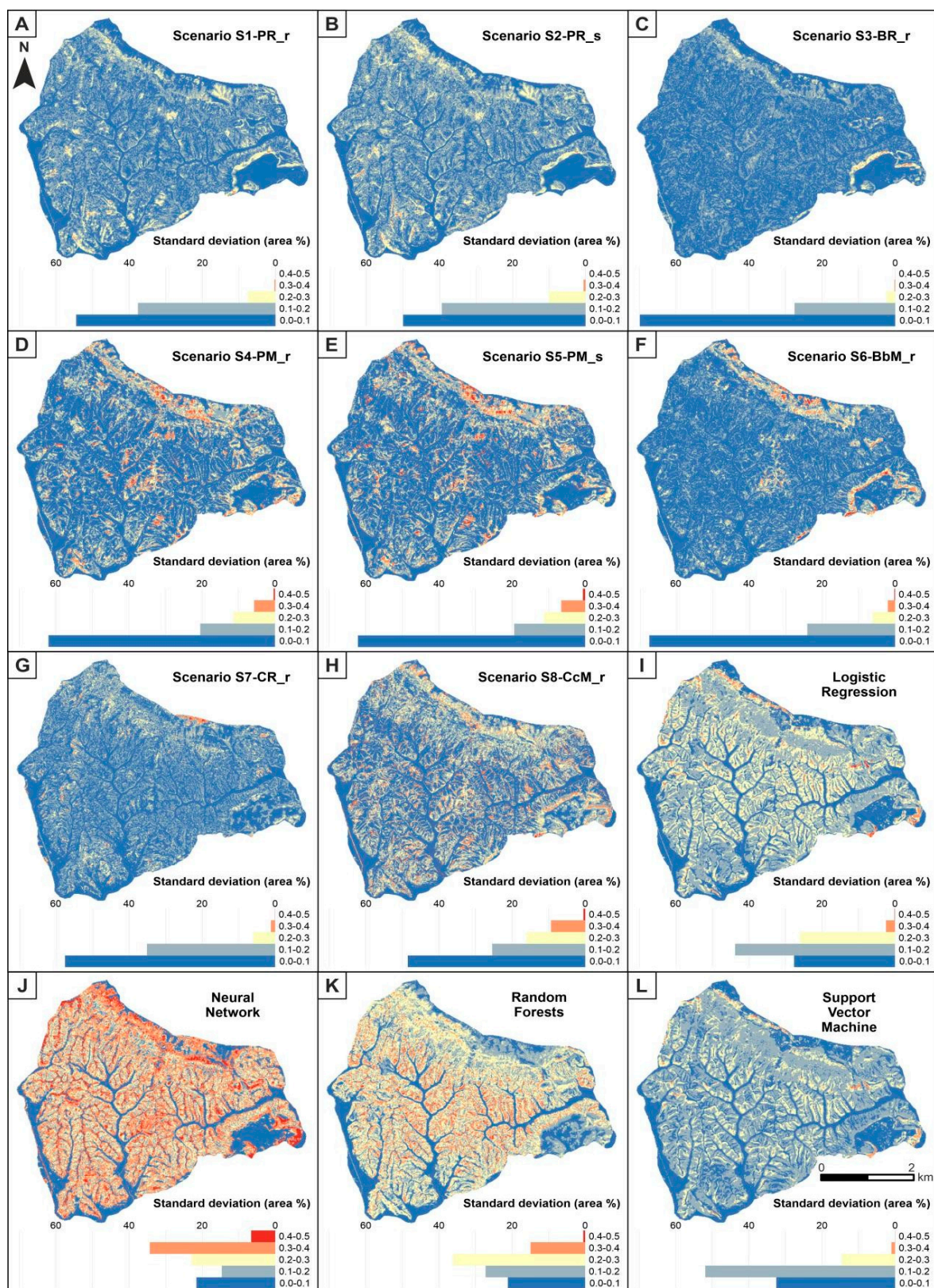


Figure 11. Standard deviation maps of probabilistic susceptibility values for eight scenarios: (A) scenario S1-PR_r, (B) scenario S2-PR_s, (C) scenario S3-BR_r, (D) scenario S4-PM_r, (E) scenario S5-PM_s, (F) scenario S6-BbM_r, (G) scenario S7-CR_r, (H) scenario S8-CcM_r and four methods: (I) Logistic Regression, (J) Neural Network, (K) Random Forests, (L) Support Vector Machine.

5.2. Analysing Model Fitting, Predictive and Classification Performance

Considering the variability described above and the fitting, predictive and classification parameters from Section 4 (i.e., Figures 7–10), interesting findings can be identified, as we will discuss below. The fitting performance for scenarios S1-PR_r, S2-PR_s and S3-BR_r is significantly lower compared to other scenarios, likely related to the significant number of randomly generated stable points in areas that are otherwise highly susceptible. Nonetheless, this was necessary as it ensured an unbiased approach, i.e., the model was unaware of the existence of the validation data. This is not the case for RF, which performed perfectly in all eight scenarios. On the other hand, the predictive parameters are incredibly high, meaning that all 32 models successfully classified the locations of the unseen 50% of stable and unstable polygons from the inventories. Anomalies are visible in the RF method, showing somewhat lower results compared to other methods in scenarios S1-PR_r, S2-PR_s and S3-BR_r. Perfect agreement in RF fitting performance and the anomalies in predictive performance point to overfitting behaviour. Moreover, the NN method yielded significantly lower AUC values in scenarios S7-CR_r and S8-CcM_r, indicating sensitivity to the number (i.e., points) of training pixels applied.

By comparing scenario S7-CR_r to scenarios S1-PR_r, S2-PR_s and S3-BR_r, a dispersion of landslides in all zones is evident. Namely, by changing to point sampling of unstable areas, landslides are no longer clustered in very high and high susceptibility zones but are more equally distributed, losing the purpose of the zones. Similarly, probabilistic classification can yield extremely large very high susceptibility zones (scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r), resulting in low landslide density, which is not expected for a very high susceptibility zone.

In terms of measuring the influence of smoothing LCFs, in scenarios S2-PR_s and S5-PM_s, negligible results are present compared to scenarios S1-PR_r and S4-PM_r, respectively. This could be because the model is less sensitive to the DTM-derived variables (elevation, slope, landform curvature, aspect, site exposure index and integrated moisture index). The latter was not addressed within the scope of this study as we opted to test the sampling strategies on a complete LCF set with no collinearity prepared to yield satisfactory AUC results. On the other hand, by having smaller low and very low zones in scenario S3-BR_r when compared to scenarios S1-PR_r and S2-PR_s, the NN and RF methods show sensitivity to the buffering of unstable areas (i.e., scenarios S3-BR_r and S6-BbM_r).

5.3. Model Verification

The predictive performance in terms of AUC and Cohen's Kappa (Figure 8) identifies scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r as more favourable, and an additional AUC was calculated to investigate model applicability. Table 6 depicts the model verification AUC values defined by examining the unstable validation pixels as landslide occurrences and the remaining study area as stable occurrences when considering all pixels in the entire study area. Unrealistically large very high susceptibility zones defined by a probabilistic zonation presented in Figure 9 showed scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r to be less favourable, which is also evident when looking to the AUC values shown in Table 6. Namely, the values correlate to poor agreement (approximately 0.7 or lower), with an exception being RF in scenarios S4-PM_r and S5-PM_s (0.726 and 0.727, respectively).

Table 6. Model verification AUC values in 32 derived LSMs.

Method	S1-PR_r	S2-PR_s	S3-BR_r	S4-PM_r	S5-PM_s	S6-BbM_r	S7-CR_r	S8-CcM_r
LR	0.790	0.772	0.754	0.683	0.662	0.690	0.737	0.696
NN	0.764	0.805	0.760	0.670	0.679	0.690	0.710	0.661
RF	0.796	0.804	0.753	0.726	0.727	0.698	0.757	0.705
SVM	0.787	0.769	0.760	0.680	0.659	0.697	0.729	0.705

Furthermore, scenario S7-CR_r performs significantly poorer compared to scenarios S1-PR_r, S2-PR_s and S3-BR_r, whereas the NN and RF methods show the best results in scenario S2-PR_s (0.805 and 0.804, respectively). The RF method clearly stands out as having the best performance in most scenarios, followed by NN. On the other hand, NN is extremely sensitive to the sampling strategies used for stable areas (e.g., 0.661 in S8-CcM_r), as also indicated by the SD variability metric (Figure 11J). The results of scenarios S1-PR_r and S2-PR_s are relatively similar to those obtained in previous studies, i.e., [73] and [80]. Lastly, high fitting and predictive metrics can favour a sampling strategy that is opposed to reality, e.g., there are evident misclassifications by the model. It should be noted that the latter was only detectable after adding additional metrics (e.g., model verification). Scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r, which rely on a stable area inventory, obviously fit this misclassification category, whereas a similar effect was reported in [32], with a rectangular area limiting the stable area sampling zone to a valley. However, [32] could not quantitatively identify the misclassifying model, whereas the AUC metric presented in Table 6 proves to be a successful solution. Furthermore, as a result of experimenting with stable area sampling, overestimated high susceptibility zone issues (i.e., scenarios S4-PM_r, S5-PM_s, S6-BbM_r and S8-CcM_r in our study) were also reported in [28] and [21].

5.4. Highlighting the Unstable Centroid Sampling Disadvantages

Unstable polygon versus unstable centroid sampling proved to be a point of interest, so to further qualitatively investigate the differences, the SVM models, which demonstrated the most minor variability, were selected to compare scenarios S1-PR_r and S7-CR_r in detail (Figure 12). Figure 12A,B confirm the spatial differences in probabilistic zone sizes and landslide presence distribution for scenarios S1-PR_r and S7-CR_r, which were previously also noted (Figures 9 and 10). A close up point of interest is illustrated in Figure 12C,D, based on which we conclude that centroid sampling to train the model cannot yield the probabilistic zonation which represents accurate environmental conditions, e.g., for most of the mapped landslide polygons to be present in the high and very high probabilistic zone. The following can be traced to model verification AUC (Table 6) and the landslide area presence (Figure 10), and not to the AUC parameters used for predictive performance of LSMs which are in this case misleading (generally >0.9).

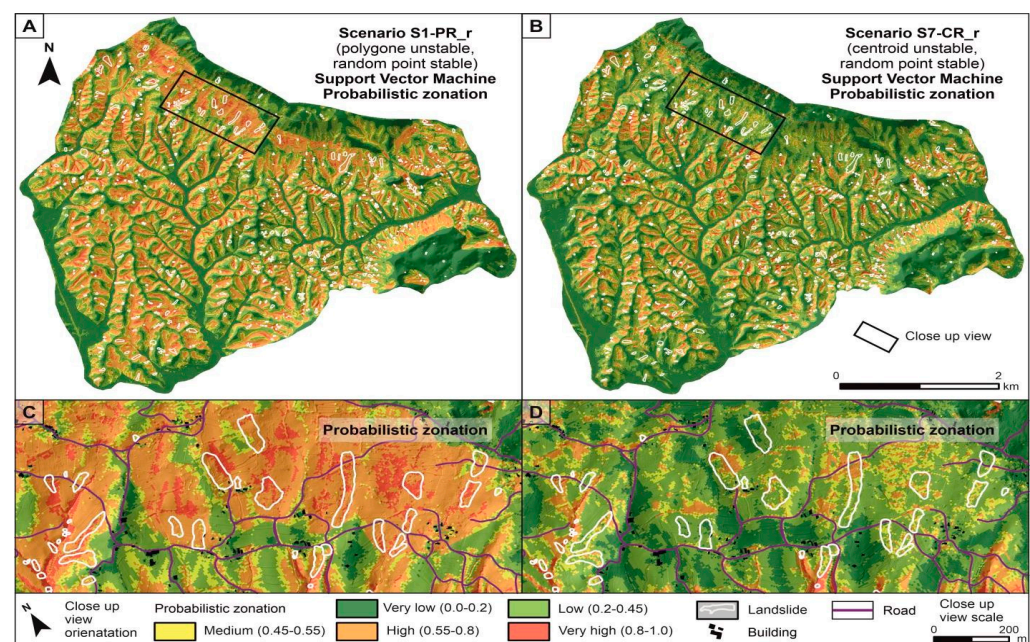


Figure 12. Zonation susceptibility map derived using the Support Vector Machine method in Scenarios S1-PR_r and S7-CR_r. Namely: (A,B): full extent probabilistic zonation; (C,D): close up view of probabilistic zonation.

6. Conclusions

This research investigates the influence of stable and unstable sampling strategies in landslide susceptibility modelling for their applicability to large-scale spatial planning systems. The modelling considered the use of eight different sampling strategies and four machine learning methods and a probabilistic zonation approach was applied. The relevant input data, i.e., spatially accurate thematic and landslide data, were carefully selected according to suggestions in the literature and previous research relating to the study area to describe local geomorphological, geological, hydrological and anthropogenic conditions in the Pannonian basin.

A novelty of this research is in the derivation of a stable areas inventory that can be used to investigate the sampling of stable pixels and smoothing of DTM-derived LCFs to capture terrain conditions before the occurrence of a landslide. The eight sampling strategies included four options for unstable pixels and stable pixels each. Namely, landslide polygon, smoothing DTM derived LCFs, polygon buffering, and landslide centroid represent four sampling strategies used for unstable pixels, whereas random points, stable area polygon, stable polygon buffering and stable area centroid were used to define the stable pixels. The processed input data were further applied in LR, NN, RF and SVM machine learning algorithms, representing commonly used methods in state-of-the-art landslide susceptibility modelling. In total, 32 LSMs were derived and analysed, considering fitting, predictive and model verification performance, as well as the SD variability for eight sampling strategies and four methods used. Furthermore, 32 zonation maps based on probabilistic zonation were described by analysing zone area, spatial distribution and landslide presence within the zones. The most significant conclusions yielded by the comprehensive comparison are as follows:

- (i) Smoothing DTM-derived LCFs at landslide locations slightly reduces model variability, whereas RF and NN models in scenario S2-PR_s using unstable polygons and randomly generated stable points represent the best LSMs
- (ii) Buffering only stable or stable and unstable polygons results in the least model variability, with a significant decline in zonation performance;
- (iii) Using the proposed stable area inventory to define stable modelling pixels by using any tested strategy drastically lowers zonation quality, which was undetectable by commonly used fitting and predictive performance metrics;
- (iv) The NN method, followed by RF, is extremely sensitive to the tested sampling strategies, whereas SVM showed the least variability, followed by similar results for LR. Additionally, with proper settings, the RF method yields LSMs that are generally better than LR, NN and SVM models;
- (v) Sampling landslides as centroids, which is extremely common in landslide susceptibility assessments, should be avoided when developing LSMs for large-scale spatial planning purposes due to the severe inability of the model to depict spatially accurate zones, which were quantitatively and qualitatively evaluated;
- (vi) A qualitative assessment of classified LSMs, quantifying zonation area size and landslide presence, represents a crucial parameter to estimate LSMs for application to large-scale spatial planning systems.

Besides the specific conclusions noted, a novelty to emphasize in terms of large-scale landslide susceptibility modelling for applicability to spatial planning systems is that commonly used quantitative metrics, i.e., Cohen's Kappa and AUC for fitting and predictive performance, may yield misleading LSM evaluations and result in inappropriate zonation maps, e.g., poor spatial distribution when considering landslide presence. In this paper, calculating additional AUCs, i.e., model verification and qualitative assessment, enabled us to distinguish more favourable LSMs. Among the different modelling strategies, an open question remains in terms of defining a systematic approach in selecting an optimal LSM, which would consider not only the studied metrics but also take into consideration the spatial accuracy and/or susceptibility zones highlighted in this paper. Furthermore, it should be noted that large-scale landslide susceptibility models should be based on

the most detailed, representative and reliable polygon-based landslide inventory possible because that will enable accurate and representative landslide susceptibility assessments and responsible and sustainable spatial planning and risk management at a local level.

Supplementary Materials: The following supporting information can be downloaded at <https://www.mdpi.com/article/10.3390/rs16162923/s1>, Figure S1: Scenario S1-PR_r; Figure S2: Scenario S2-PR_s; Figure S3: Scenario S3-BR_r; Figure S4: Scenario S4-PM_r; Figure S5: Scenario S5-PM_s; Figure S6: Scenario S6-BbM_r; Figure S7: Scenario S7-CR_r; Figure S8: Scenario S8-CcM_r.

Author Contributions: Conceptualization, M.R., S.B.G. and M.S.; methodology, M.S., S.B.G. and M.R.; software, M.S.; validation, S.B.G., M.R. and S.M.A.; formal analysis, M.S.; investigation, M.S.; resources, M.S., S.B.G., M.R., M.K. and S.M.A.; data curation, M.S. and S.B.G.; writing—original draft preparation, M.S.; writing—review and editing, S.B.G., M.R., M.K. and S.M.A.; visualization, M.S. and M.K.; supervision, S.B.G. and S.M.A.; project administration, S.M.A., S.B.G. and M.K.; funding acquisition, M.S., M.R., S.B.G., M.K. and S.M.A. All authors have read and agreed to the published version of the manuscript.

Funding: This research has been fully supported by the Croatian Science Foundation under the project ‘Methodology development for landslide susceptibility assessment for land-use planning based on LiDAR technology’, LandSlidePlan (HRZZ IP-2019-04-9900, HRZZ DOK-2020-01-2432) and by Institutional project 311980036 ModKLIZ, co-funded by the Faculty of Mining, Geology and Petroleum Engineering. The work was co-funded by the Geomorphology group of the Istituto di Ricerca per la Protezione Idrogeologica, Consiglio Nazionale delle Ricerche (Perugia, Italy).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Acknowledgments: We would like to thank anonymous reviewers for their suggestions and comments.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for Landslide Susceptibility, Hazard and Risk Zoning for Land Use Planning. *Eng. Geol.* **2008**, *102*, 85–98. [\[CrossRef\]](#)
2. Guzzetti, F.; Reichenbach, P.; Cardinali, M.; Galli, M.; Ardizzone, F. Probabilistic Landslide Hazard Assessment at the Basin Scale. *Geomorphology* **2005**, *72*, 272–299. [\[CrossRef\]](#)
3. Varnes, D.J. *Landslide Hazard Zonation: A Review of Principles and Practice*; Education, Scientific and Cultural Organization: Paris, France, 1984.
4. Soeters, R.; van Westen, C.J. Slope Instability Recognition Analysis and Zonation. In *Landslides: Investigation and Mitigation*; Turner, K.T., Schuster, R.L., Eds.; National Academy Press: Washington, DC, USA, 1996; pp. 129–177.
5. Guzzetti, F.; Carrara, A.; Cardinali, M.; Reichenbach, P. Landslide Hazard Evaluation: A Review of Current Techniques and Their Application in a Multi-Scale Study, Central Italy. *Geomorphology* **1999**, *31*, 181–216. [\[CrossRef\]](#)
6. van Westen, C.J.; Rengers, N.; Soeters, R. Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment. *Nat. Hazards* **2003**, *30*, 399–419. [\[CrossRef\]](#)
7. van Westen, C.J.; Castellanos, E.; Kuriakose, S.L. Spatial Data for Landslide Susceptibility, Hazard, and Vulnerability Assessment: An Overview. *Eng. Geol.* **2008**, *102*, 112–131. [\[CrossRef\]](#)
8. Fell, R.; Corominas, J.; Bonnard, C.; Cascini, L.; Leroi, E.; Savage, W.Z. Guidelines for Landslide Susceptibility, Hazard and Risk Zoning for Land-Use Planning. *Eng. Geol.* **2008**, *102*, 99–111. [\[CrossRef\]](#)
9. Corominas, J.; van Westen, C.; Frattini, P.; Cascini, L.; Malet, J.P.; Fotopoulou, S.; Catani, F.; Van Den Eeckhaut, M.; Mavrouli, O.; Agliardi, F.; et al. Recommendations for the Quantitative Analysis of Landslide Risk. *Bull. Eng. Geol. Environ.* **2014**, *73*, 209–263. [\[CrossRef\]](#)
10. Dias, H.C.; Hölbling, D.; Grohmann, C.H. Landslide Susceptibility Mapping in Brazil: A Review. *Geosciences* **2021**, *11*, 425. [\[CrossRef\]](#)
11. Das, S.; Sarkar, S.; Kanungo, D.P. A Critical Review on Landslide Susceptibility Zonation: Recent Trends, Techniques, and Practices in Indian Himalaya. *Nat. Hazards* **2023**, *115*, 23–72. [\[CrossRef\]](#)
12. Shano, L.; Raghuvanshi, T.K.; Meten, M. Landslide Susceptibility Evaluation and Hazard Zonation Techniques—A Review. *Geoenviron. Disasters* **2020**, *7*, 18. [\[CrossRef\]](#)
13. Reichenbach, P.; Rossi, M.; Malamud, B.D.; Mihir, M.; Guzzetti, F. A Review of Statistically-Based Landslide Susceptibility Models. *Earth Sci. Rev.* **2018**, *180*, 60–91. [\[CrossRef\]](#)
14. Pourghasemi, H.R.; Rahmati, O. Prediction of the Landslide Susceptibility: Which Algorithm, Which Precision? *Catena* **2018**, *162*, 177–192. [\[CrossRef\]](#)

15. Lee, S. Current and Future Status of GIS-Based Landslide Susceptibility Mapping: A Literature Review. *Korean J. Remote Sens.* **2019**, *35*, 179–193.
16. Mihalić Arbanas, S.; Bernat Gazibara, S.; Krkač, M.; Sinčić, M.; Lukačić, H.; Jagodnik, P.; Arbanas, Ž. Landslide Detection and Spatial Prediction: Application of Data and Information from Landslide Maps. In *Progress in Landslide Research and Technology, Volume 1 Issue 2*, 2022; Alcantara-Ayala, I., Arbanas, Ž., Huntley, D., Konagai, K., Mikoš, M., Sassa, K., Sassa, S., Tang, H., Tiwari, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 195–212. [\[CrossRef\]](#)
17. Bernat Gazibara, S.; Mihalić Arbanas, S.; Sinčić, M.; Krkač, M.; Lukačić, H.; Jagodnik, P.; Arbanas, Ž. LandSlidePlan -Scientific Research Project on Landslide Susceptibility Assessment in Large Scale. In Proceedings of the Proceedings of the 5th Regional Symposium on Landslides in Adriatic—Balkan Region, Rijeka, Croatia, 23–26 March 2022; pp. 99–106.
18. Bernat, S.; Mihalić Arbanas, S.; Krkač, M. Inventory of Precipitation Triggered Landslides in the Winter of 2013 in Zagreb (Croatia, Europe). In *Landslide Science for a Safer Geoenvironment*; Springer International Publishing: Cham, Switzerland, 2014; pp. 829–835.
19. Sinčić, M.; Bernat Gazibara, S.; Krkač, M.; Lukačić, H.; Mihalić Arbanas, S. The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments. *Land* **2022**, *11*, 1360. [\[CrossRef\]](#)
20. Krkač, M.; Bernat Gazibara, S.; Sinčić, M.; Lukačić, H.; Mihalić Arbanas, S. Landslide Inventory Mapping Based on LiDAR Data: A Case Study from Hrvatsko Zagorje (Croatia). In Proceedings of the 5th ReSyLAB, Rijeka, Croatia, 23–26 March 2022; pp. 81–86.
21. Hong, H.; Miao, Y.; Liu, J.; Zhu, A.-X. Exploring the Effects of the Design and Quantity of Absence Data on the Performance of Random Forest-Based Landslide Susceptibility Mapping. *Catena* **2019**, *176*, 45–64. [\[CrossRef\]](#)
22. Fu, Z.; Wang, F.; Dou, J.; Nam, K.; Ma, H. Enhanced Absence Sampling Technique for Data-Driven Landslide Susceptibility Mapping: A Case Study in Songyang County, China. *Remote Sens.* **2023**, *15*, 3345. [\[CrossRef\]](#)
23. Busby, J.R. Bioclim—A Bioclimatic Analysis and Prediction System. In *Nature Conservation*; Margules, C.R., Augstin, M.P., Eds.; CSIRO: Canberra, Australia, 1993; pp. 64–68.
24. Carpenter, G.; Gillison, A.N.; Winter, J. DOMAIN: A Flexible Modelling Procedure for Mapping Potential Distributions of Plants and Animals. *Biodivers. Conserv.* **1993**, *2*, 667–680. [\[CrossRef\]](#)
25. Scholkopf, G. Support Vector Method for Novelty Detection. *Adv. Neural Inf. Process Syst.* **2008**, *12*, 582–588.
26. Xiao, C.; Tian, Y.; Shi, W.; Guo, Q.; Wu, L. A New Method of Pseudo Absence Data Generation in Landslide Susceptibility Mapping with a Case Study of Shenzhen. *Sci. China Technol. Sci.* **2010**, *53*, 75–84. [\[CrossRef\]](#)
27. Hu, J.; Xu, K.; Wang, G.; Liu, Y.; Khan, M.A.; Mao, Y.; Zhang, M. A Novel Landslide Susceptibility Mapping Portrayed by OA-HD and K-Medoids Clustering Algorithms. *Bull. Eng. Geol. Environ.* **2021**, *80*, 765–779. [\[CrossRef\]](#)
28. Zhu, A.-X.; Miao, Y.; Liu, J.; Bai, S.; Zeng, C.; Ma, T.; Hong, H. A Similarity-Based Approach to Sampling Absence Data for Landslide Susceptibility Mapping Using Data-Driven Methods. *Catena* **2019**, *183*, 104188. [\[CrossRef\]](#)
29. Xi, C.; Han, M.; Hu, X.; Liu, B.; He, K.; Luo, G.; Cao, X. Effectiveness of Newmark-Based Sampling Strategy for Coseismic Landslide Susceptibility Mapping Using Deep Learning, Support Vector Machine, and Logistic Regression. *Bull. Eng. Geol. Environ.* **2022**, *81*, 174. [\[CrossRef\]](#)
30. Rabby, Y.W.; Li, Y.; Hilafu, H. An Objective Absence Data Sampling Method for Landslide Susceptibility Mapping. *Sci. Rep.* **2023**, *13*, 1740. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Zhou, C.; Wang, Y.; Cao, Y.; Singhc, R.P.; Ahmed, B.; Motagh, M.; Wang, Y.; Chen, L. Non-Landslide Sampling and Ensemble Learning Techniques to Improve Landslide Susceptibility Mapping. *Nat. Hazards Earth Syst. Sci.* **2023**. [\[CrossRef\]](#)
32. Lucchese, L.V.; de Oliveira, G.G.; Pedrollo, O.C. Investigation of the Influence of Nonoccurrence Sampling on Landslide Susceptibility Assessment Using Artificial Neural Networks. *Catena* **2021**, *198*, 105067. [\[CrossRef\]](#)
33. Conoscenti, C.; Rotigliano, E.; Cama, M.; Caraballo-Arias, N.A.; Lombardo, L.; Agnesi, V. Exploring the Effect of Absence Selection on Landslide Susceptibility Models: A Case Study in Sicily, Italy. *Geomorphology* **2016**, *261*, 222–235. [\[CrossRef\]](#)
34. Dornik, A.; Drăguț, L.; Oguchi, T.; Hayakawa, Y.; Micu, M. Influence of Sampling Design on Landslide Susceptibility Modeling in Lithologically Heterogeneous Areas. *Sci. Rep.* **2022**, *12*, 2106. [\[CrossRef\]](#)
35. Hussin, H.Y.; Zumpano, V.; Reichenbach, P.; Sterlacchini, S.; Micu, M.; van Westen, C.; Bălteanu, D. Different Landslide Sampling Strategies in a Grid-Based Bi-Variate Statistical Susceptibility Model. *Geomorphology* **2016**, *253*, 508–523. [\[CrossRef\]](#)
36. Dou, J.; Yunus, A.P.; Merghadi, A.; Shirzadi, A.; Nguyen, H.; Hussain, Y.; Avtar, R.; Chen, Y.; Pham, B.T.; Yamagishi, H. Different Sampling Strategies for Predicting Landslide Susceptibilities Are Deemed Less Consequential with Deep Learning. *Sci. Total Environ.* **2020**, *720*, 137320. [\[CrossRef\]](#)
37. Huang, F.; Yan, J.; Fan, X.; Yao, C.; Huang, J.; Chen, W.; Hong, H. Uncertainty Pattern in Landslide Susceptibility Prediction Modelling: Effects of Different Landslide Boundaries and Spatial Shape Expressions. *Geosci. Front.* **2022**, *13*, 101317. [\[CrossRef\]](#)
38. Petschko, H.; Bell, R.; Leopold, P.; Heiss, G.; Glade, T. Landslide Inventories for Reliable Susceptibility Maps in Lower Austria. In *Landslide Science and Practice*; Margottini, C., Canuti, P., Sassa, K., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 281–286.
39. Poli, S.; Sterlacchini, S. Landslide Representation Strategies in Susceptibility Studies Using Weights-of-Evidence Modeling Technique. *Nat. Resour. Res.* **2007**, *16*, 121–134. [\[CrossRef\]](#)
40. Simon, N.; Crozier, M.; de Roiste, M.; Rafek, A.G. Point Based Assessment: Selecting TheBest Way to Represent Landslide Polygon as Point Frequency in Landslide Investigation. *Electron. J. Geotech. Eng.* **2013**, *18*, 775–784.
41. Lai, J.-S.; Chiang, S.-H.; Tsai, F. Exploring Influence of Sampling Strategies on Event-Based Landslide Susceptibility Modeling. *ISPRS Int. J. Geoinf.* **2019**, *8*, 397. [\[CrossRef\]](#)

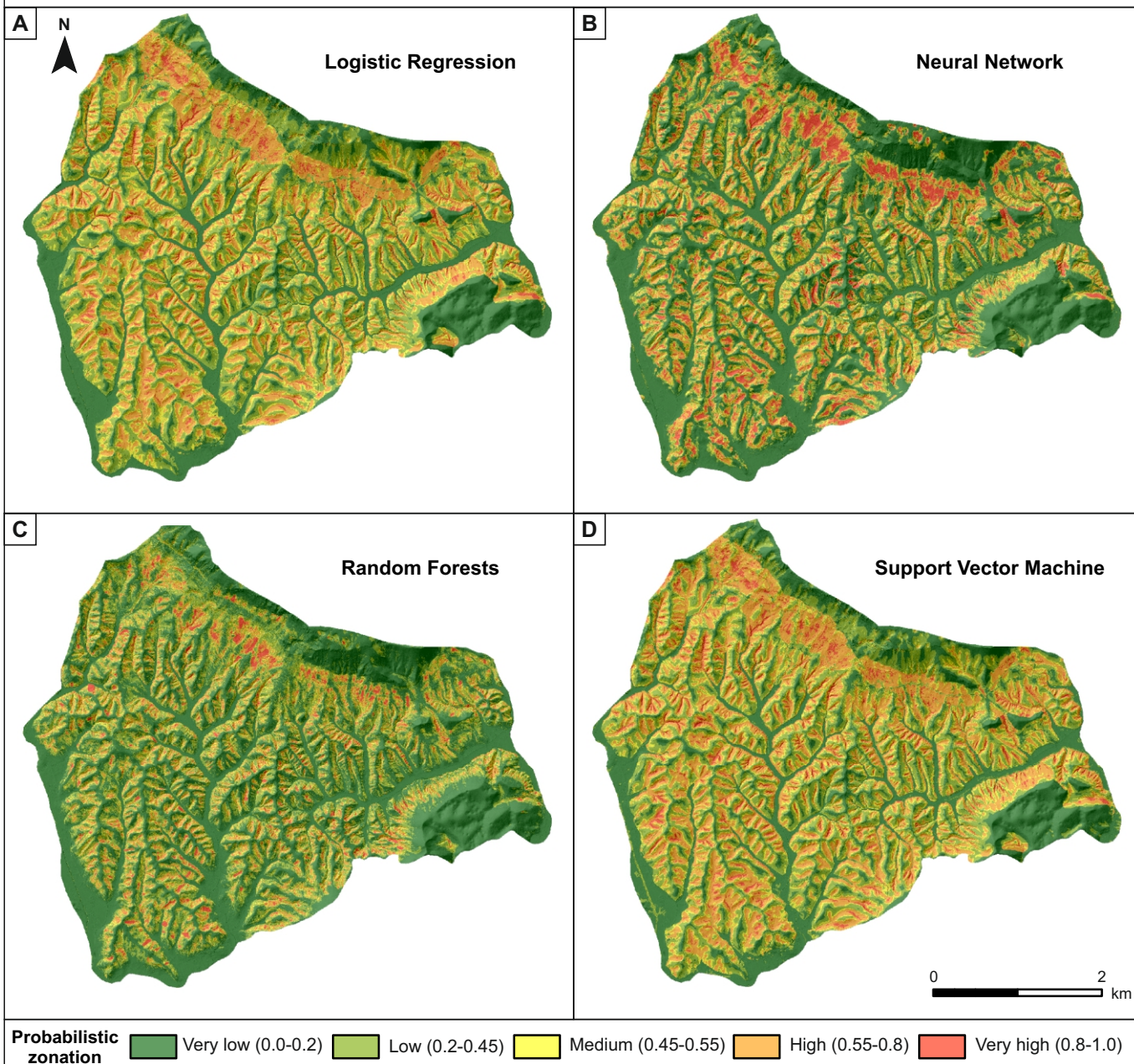
42. Süzen, M.L.; Doyuran, V. Data Driven Bivariate Landslide Susceptibility Assessment Using Geographical Information Systems: A Method and Application to Asarsuyu Catchment, Turkey. *Eng. Geol.* **2004**, *71*, 303–321. [[CrossRef](#)]
43. Yilmaz, I. The Effect of the Sampling Strategies on the Landslide Susceptibility Mapping by Conditional Probability and Artificial Neural Networks. *Environ. Earth Sci.* **2010**, *60*, 505–519. [[CrossRef](#)]
44. Lee, S. Landslide Susceptibility Mapping Using an Artificial Neural Network in the Gangneung Area, Korea. *Int. J. Remote Sens.* **2007**, *28*, 4763–4783. [[CrossRef](#)]
45. Yao, X.; Tham, L.G.; Dai, F.C. Landslide Susceptibility Mapping Based on Support Vector Machine: A Case Study on Natural Slopes of Hong Kong, China. *Geomorphology* **2008**, *101*, 572–582. [[CrossRef](#)]
46. Sarkar, S.; Roy, A.K.; Martha, T.R. Landslide Susceptibility Assessment Using Information Value Method in Parts of the Darjeeling Himalayas. *J. Geol. Soc. India* **2013**, *82*, 351–362. [[CrossRef](#)]
47. Hemasinghe, H.; Rangali, R.S.S.; Deshapriya, N.L.; Samarakoon, L. Landslide Susceptibility Mapping Using Logistic Regression Model (a Case Study in Badulla District, Sri Lanka). *Procedia Eng.* **2018**, *212*, 1046–1053. [[CrossRef](#)]
48. Wang, H.; Zhang, L.; Luo, H.; He, J.; Cheung, R.W.M. AI-Powered Landslide Susceptibility Assessment in Hong Kong. *Eng. Geol.* **2021**, *288*, 106103. [[CrossRef](#)]
49. Sinčić, M.; Bernat Gazibara, S.; Krkač, M.; Mihalić Arbanas, S. Landslide Susceptibility Assessment of the City of Karlovac Using Bivariate Statistical Analysis. *Rudarsko-Geološko-Naftni Zb.* **2022**, *37*, 149–170. [[CrossRef](#)]
50. Pascale, S.; Parisi, S.; Mancini, A.; Schiattarella, M.; Conforti, M.; Sole, A.; Murgante, B.; Sdao, F. Landslide Susceptibility Mapping Using Artificial Neural Network in the Urban Area of Senise and San Costantino Albanese (Basilicata, Southern Italy). *Lect. Notes Comput. Sci.* **2013**, *7974*, 473–488. [[CrossRef](#)]
51. Catani, F.; Lagomarsino, D.; Segoni, S.; Tofani, V. Landslide Susceptibility Estimation by Random Forests Technique: Sensitivity and Scaling Issues. *Nat. Hazards Earth Syst. Sci.* **2013**, *13*, 2815–2831. [[CrossRef](#)]
52. Pradhan, B. A Comparative Study on the Predictive Ability of the Decision Tree, Support Vector Machine and Neuro-Fuzzy Models in Landslide Susceptibility Mapping Using GIS. *Comput. Geosci.* **2013**, *51*, 350–365. [[CrossRef](#)]
53. Farooq, S.; Akram, M.S. Landslide Susceptibility Mapping Using Information Value Method in Jhelum Valley of the Himalayas. *Arab. J. Geosci.* **2021**, *14*, 824. [[CrossRef](#)]
54. Sandić, C.; Marjanović, M.; Abolmasov, B.; Tošić, R. Integrating Landslide Magnitude in the Susceptibility Assessment of the City of Doboj, Using Machine Learning and Heuristic Approach. *J. Maps* **2023**, *19*. [[CrossRef](#)]
55. Guo, Z.; Tian, B.; Zhu, Y.; He, J.; Zhang, T. How Do the Landslide and Non-Landslide Sampling Strategies Impact Landslide Susceptibility Assessment?—A Catchment-Scale Case Study from China. *J. Rock Mech. Geotech. Eng.* **2024**, *16*, 877–894. [[CrossRef](#)]
56. Song, Y.; Yang, D.; Wu, W.; Zhang, X.; Zhou, J.; Tian, Z.; Wang, C.; Song, Y. Evaluating Landslide Susceptibility Using Sampling Methodology and Multiple Machine Learning Models. *ISPRS Int. J. Geoinf.* **2023**, *12*, 197. [[CrossRef](#)]
57. Petschko, H.; Brenning, A.; Bell, R.; Goetz, J.; Glade, T. Assessing the Quality of Landslide Susceptibility Maps—Case Study Lower Austria. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 95–118. [[CrossRef](#)]
58. Bornaetxea, T.; Rossi, M.; Marchesini, I.; Alvioli, M. Effective Surveyed Area and Its Role in Statistical Landslide Susceptibility Assessments. *Nat. Hazards Earth Syst. Sci.* **2018**, *18*, 2455–2469. [[CrossRef](#)]
59. Bernat Gazibara, S.; Sinčić, M.; Krkač, M.; Jagodnik, P.; Lukačić, H.; Mihalić Arbanas, S. Influence of the Landslide Inventory Completeness on the Accuracy of the Landslide Susceptibility Modelling: A Case Study from the City of Zagreb (Croatia). In Proceedings of the 6th World Landslide Forum, Florence, Italy, 14–17 November 2023; p. 644.
60. Tien Bui, D.; Tuan, T.A.; Klempe, H.; Pradhan, B.; Revhaug, I. Spatial Prediction Models for Shallow Landslide Hazards: A Comparative Assessment of the Efficacy of Support Vector Machines, Artificial Neural Networks, Kernel Logistic Regression, and Logistic Model Tree. *Landslides* **2016**, *13*, 361–378. [[CrossRef](#)]
61. Merghadi, A.; Yunus, A.P.; Dou, J.; Whiteley, J.; ThaiPham, B.; Bui, D.T.; Avtar, R.; Abderrahmane, B. Machine Learning Methods for Landslide Susceptibility Studies: A Comparative Overview of Algorithm Performance. *Earth Sci. Rev.* **2020**, *207*, 103225. [[CrossRef](#)]
62. Youssef, A.M.; Pourghasemi, H.R. Landslide Susceptibility Mapping Using Machine Learning Algorithms and Comparison of Their Performance at Abha Basin, Asir Region, Saudi Arabia. *Geosci. Front.* **2021**, *12*, 639–655. [[CrossRef](#)]
63. Bernat, S.; Mihalić Arbanas, S.; Krkač, M.; Sećanj, M. Catalog of Precipitation Events That Triggered Landslides in Northwestern Croatia. In Proceedings of the 2nd Regional Symposium on Landslides in the Adriatic-Balkan Region, Belgrade, Serbia, 14–15 May 2015; pp. 103–107.
64. Aničić, B.; Juriša, M. *Basic Geological Map, Scale 1:100,000, Rogatec, Sheet 33–68*; Geological Department: Belgrade, Serbia, 1984.
65. Aničić, B.; Juriša, M. *Geological Notes for Basic Geological Map, Scale 1:100,000, Rogatec, Sheet 33–68*; Geological Department: Belgrade, Serbia, 1983.
66. Šimunić, A.; Pikija, M.; Hečimović, I.; Šimunić, A. *Geological Notes for Basic Geological Map, Scale 1:100,000, Varaždin, Sheet 33–69*; Geological Department: Belgrade, Serbia, 1982.
67. Šimunić, A.; Pikija, M.; Hečimović, I. *Basic Geological Map, Scale 1:100,000, Varaždin, Sheet 33–69*; Geological Department: Belgrade, Serbia, 1982.
68. Zaninović, K.; Gajić-Čapka, K.; Perčec Tadić, M.; Vučetić, M.; Milković, J.; Bajić, A.; Cindrić, K.; Cvitan, L.; Katušin, Z.; Kaučić, D. *Climate Atlas of Croatia 1961–1990, 1971–2000*; Croatian Meteorological and Hydrological Service: Zagreb, Croatia, 2008.
69. URL-1. Available online: https://meteo.hr/klima.php?section=klima_podaci¶m=k1&Grad=varazdin (accessed on 6 June 2024).

70. Razak, K.A.; Straatsma, M.W.; van Westen, C.J.; Malet, J.P.; de Jong, S.M. Airborne Laser Scanning of Forested Landslides Characterization: Terrain Model Quality and Visualization. *Geomorphology* **2011**, *126*, 186–200. [\[CrossRef\]](#)
71. Bernat Gazibara, S.; Krkač, M.; Mihalić Arbanas, S. Verification of Historical Landslide Inventory Maps for the Podsljeme Area in the City of Zagreb Using LiDAR Based LiDAR Landslide Inventory. *Rudarsko-Geološko-Naftni Zb.* **2019**, *34*. [\[CrossRef\]](#)
72. Bernat Gazibara, S.; Krkač, M.; Mihalić Arbanas, S. Landslide Inventory Mapping Using LiDAR Data in the City of Zagreb (Croatia). *J. Maps* **2019**, *15*, 773–779. [\[CrossRef\]](#)
73. Krkač, M.; Bernat Gazibara, S.; Sinčić, M.; Lukačić, H.; Šarić, G.; Snježana, M.A. Impact of Input Data on the Quality of the Landslide Susceptibility Large-Scale Maps: A Case Study from NW Croatia. In *Progress in Landslide Research and Technology*; Alcántara-ra-Ayala, I., Arbanas, Ž., Cuomo, S., Huntley, D., Konagai, K., Mihalić Arbanas, S., Mikoš, M., Sassa, K., Tang, H., Tiwari, B., Eds.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 135–146. [\[CrossRef\]](#)
74. Jagodnik, P.; Bernat Gazibara, S.; Jagodnik, V. Typed and Distribution of Quaternary Deposits Originating from Carbonate Rock Slopes in the Vinodol Valley, Croatia—New Insight Using Airborne LiDAR Data. *Rudarsko-Geološko-Naftni Zb.* **2020**, *35*, 57–77. [\[CrossRef\]](#)
75. Jagodnik, P.; Bernat Gazibara, S.; Arbanas, Ž.; Mihalić Arbanas, S. Engineering Geological Mapping Using Airborne LiDAR Datasets—An Example from the Vinodol Valley, Croatia. *J. Maps* **2020**, *16*, 855–866. [\[CrossRef\]](#)
76. URL-2. Available online: <http://Geoportal.Dgu.Hr/Wms?Layers=DOF> (accessed on 6 June 2024).
77. Marchesini, I.; Ardizzone, F.; Alvioli, M.; Rossi, M.; Guzzetti, F. Non-Susceptible Landslide Areas in Italy and in the Mediterranean Region. *Nat. Hazards Earth Syst. Sci.* **2014**, *14*, 2215–2231. [\[CrossRef\]](#)
78. Sinčić, M.; Bernat Gazibara, S.; Rossi, M.; Mihalić Arbanas, S. Comparison of Conditioning Factors Classification Criteria in Large Scale Statistically Based Landslide Susceptibility Models. *Nat. Hazards Earth Syst. Sci.* **2024**. [\[CrossRef\]](#)
79. Evans, J.S.; Oakleaf, J.; Cushman, S.A.; Theobald, D. *An Arc Gis Toolbox for Surface Gradient and Geo-Morphometric Modeling*, Version 2.0-0; 2014. Available online: <https://evansmurphy.wixsite.com/evansspatial/arcgis-gradient-metrics-toolbox> (accessed on 1 June 2024).
80. Bernat Gazibara, S.; Sinčić, M.; Rossi, M.; Reichenbach, P.; Krkač, M.; Lukačić, H.; Jagodnik, P.; Šarić, G.; Mihalić Arbanas, S. Application of LAND-SUITE for Landslide Susceptibility Modelling Using Different Mapping Units: A Case Study in Croatia. In *Progress in Landslide Research and Technology, Volume 2 Issue 2*, 2023; Alcántara-Ayala, I., Arbanas, Ž., Huntley, D., Konagai, K., Mihalić Arbanas, S., Mikoš, M., Ramesh, M.V., Sassa, K., Sassa, S., Tang, H., et al., Eds.; Springer: Berlin/Heidelberg, Germany, 2023; pp. 343–354. [\[CrossRef\]](#)
81. Rossi, M.; Guzzetti, F.; Reichenbach, P.; Mondini, A.C.; Peruccacci, S. Optimal Landslide Susceptibility Zonation Based on Multiple Forecasts. *Geomorphology* **2010**, *114*, 129–142. [\[CrossRef\]](#)
82. Wang, L.-J.; Guo, M.; Sawada, K.; Lin, J.; Zhang, J. A Comparative Study of Landslide Susceptibility Maps Using Logistic Regression, Frequency Ratio, Decision Tree, Weights of Evidence and Artificial Neural Network. *Geosci. J.* **2016**, *20*, 117–136. [\[CrossRef\]](#)
83. The MathWorks Inc. *MATLAB, Version: 9.10.0.1602886 (R2021b)*; The MathWorks Inc.: Natick, MA, USA, 2021. Available online: <https://www.mathworks.com> (accessed on 6 June 2024).
84. The MathWorks Inc. *Statistics and Machine Learning, Toolbox: 12.1 (R2021)*; The MathWorks Inc.: Natick, MA, USA, 2021. Available online: <https://www.mathworks.com/products/statistics.html> (accessed on 6 June 2024).
85. URL-3. Available online: <https://www.mathworks.com/help/stats/choose-a-classifier.html> (accessed on 6 June 2024).
86. Gorsevski, P.V.; Gessler, P.E.; Foltz, R.B.; Elliot, W.J. Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis. *Trans. GIS* **2006**, *10*, 395–415. [\[CrossRef\]](#)
87. Landis, J.R.; Koch, G.G. The Measurement of Observer Agreement for Categorical Data. *Biometrics* **1977**, 159–174. [\[CrossRef\]](#)
88. Fawcett, T. An Introduction to ROC Analysis. *Pattern Recognit. Lett.* **2006**, *27*, 861–874. [\[CrossRef\]](#)
89. Rossi, M.; Reichenbach, P. LAND-SE: A Software for Statistically Based Landslide Susceptibility Zonation, Version 1.0. *Geosci. Model Dev.* **2016**, *9*, 3533–3543. [\[CrossRef\]](#)
90. Rossi, M.; Bornaetxea, T.; Reichenbach, P. LAND-SUITE V1.0: A Suite of Tools for Statistically Based Landslide Susceptibility Zonation. *Geosci. Model Dev.* **2022**, *15*, 5651–5666. [\[CrossRef\]](#)
91. Tyagi, A.; Tiwari, R.K.; James, N. Mapping the Landslide Susceptibility Considering Future Land-Use Land-Cover Scenario. *Landslides* **2023**, *20*, 65–76. [\[CrossRef\]](#)
92. Sweets, J.A. Measuring the Accuracy of Diagnostic Systems. *Science* **1988**, *240*, 1285–1293. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.

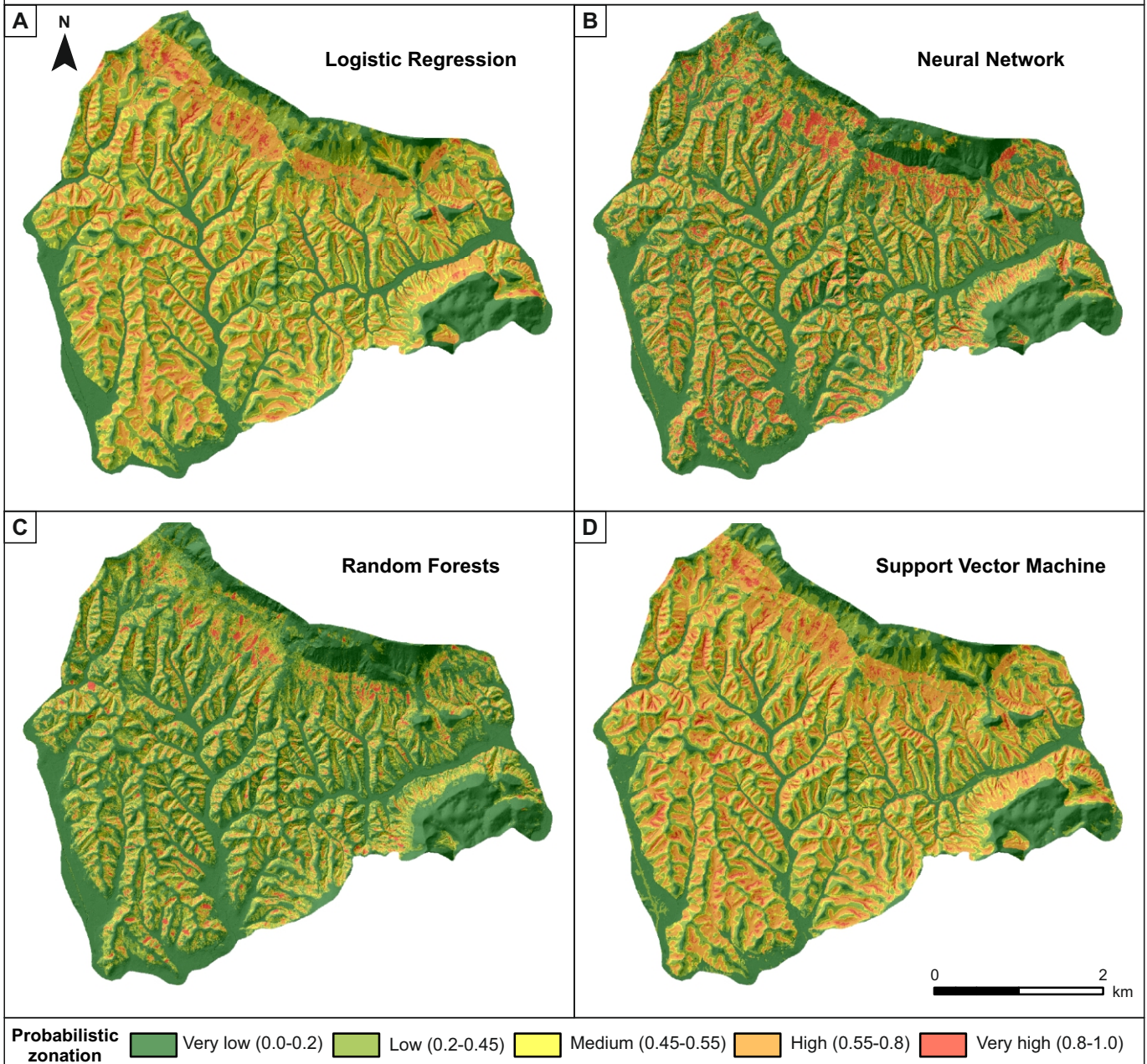
Supplementary material S1

Scenario S1-PR_r (polygon unstable, random point stable, regular LCF)



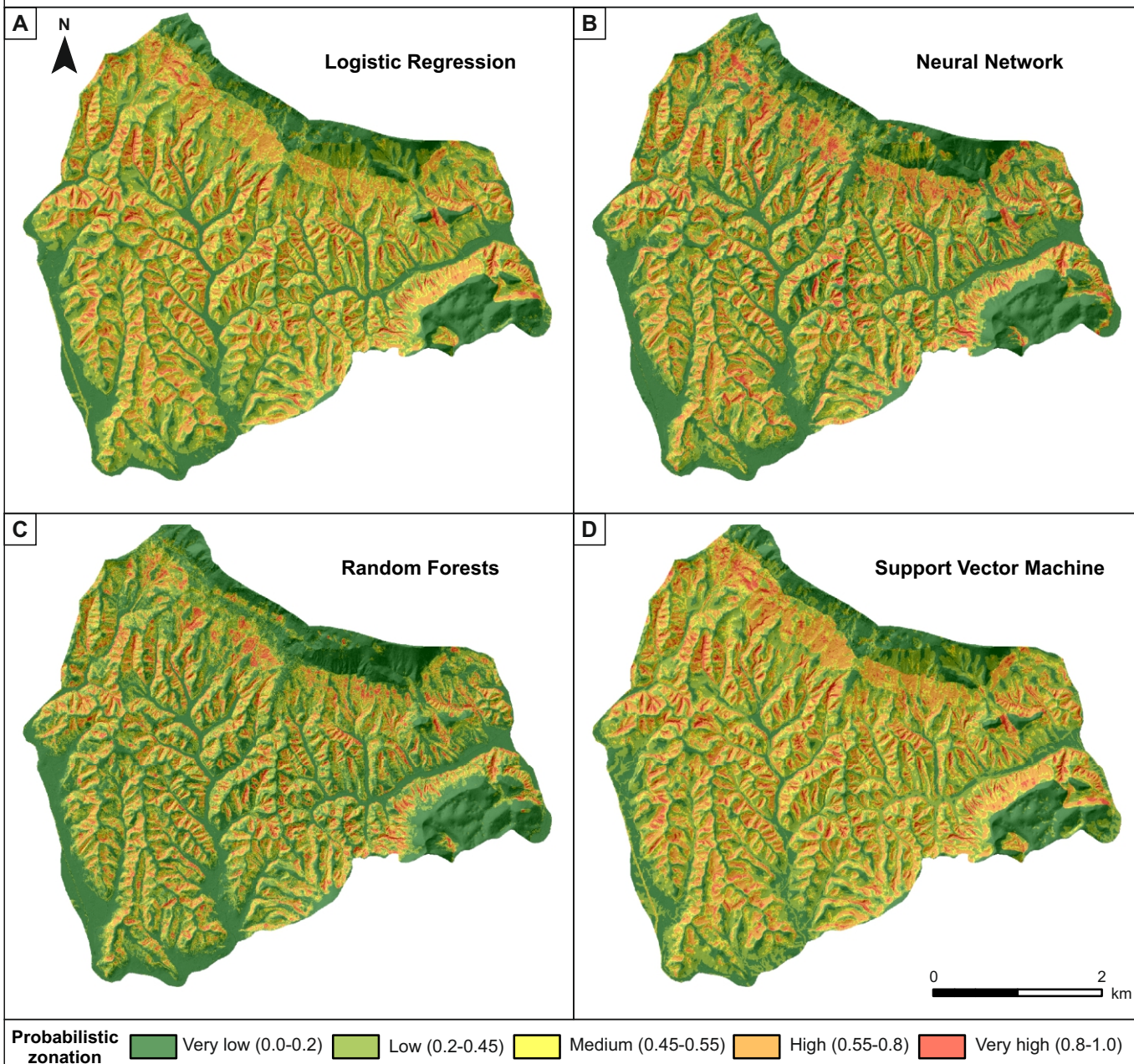
Supplementary material S2

Scenario S2-PR_s (polygon unstable, random point stable, smoothed LCF)



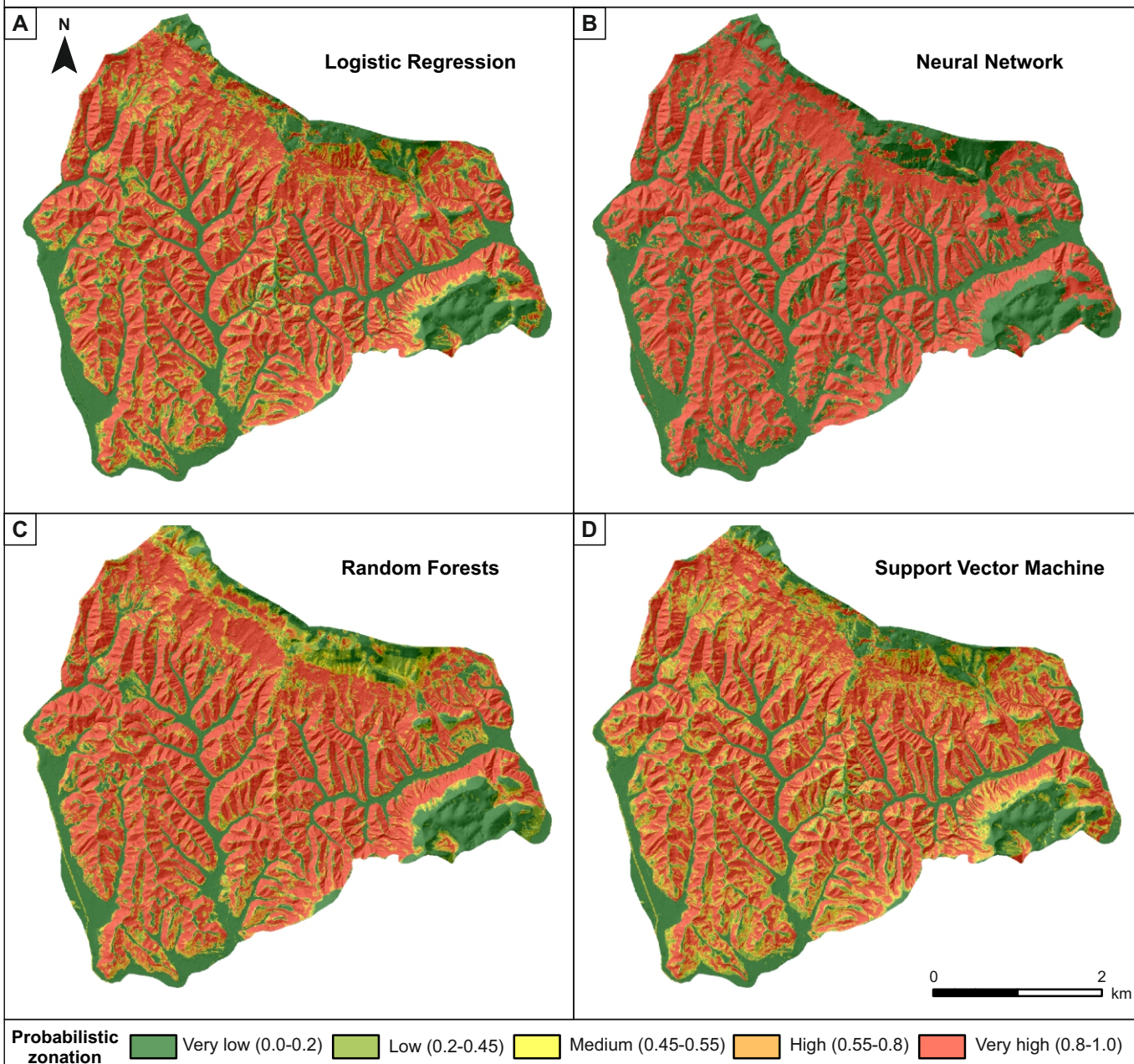
Supplementary material S3

Scenario S3-BR_r (buffer polygon unstable, random point stable, regular LCF)



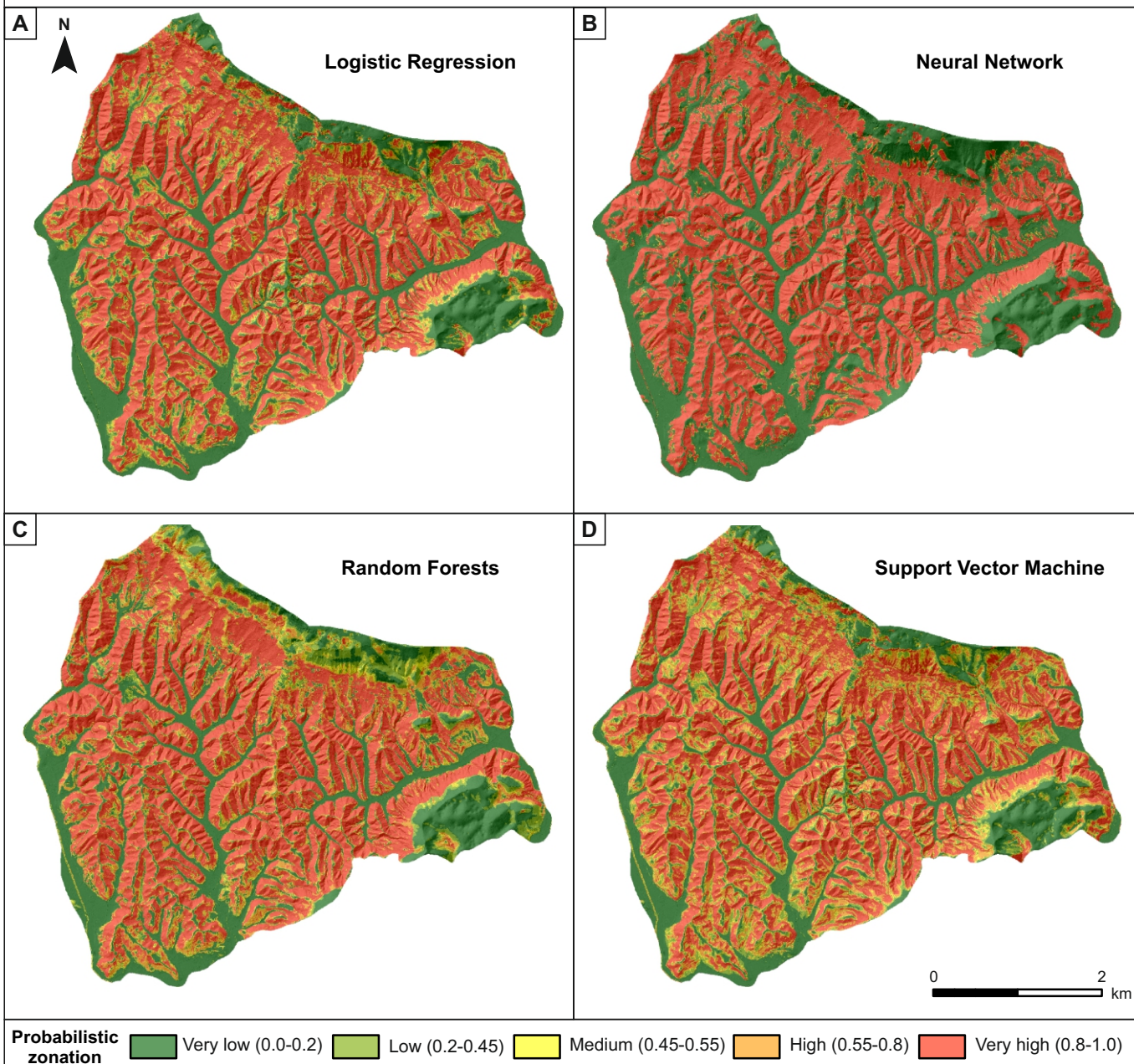
Supplementary material S4

Scenario S4-PM_r (polygon unstable, mapped stable polygon, regular LCF)



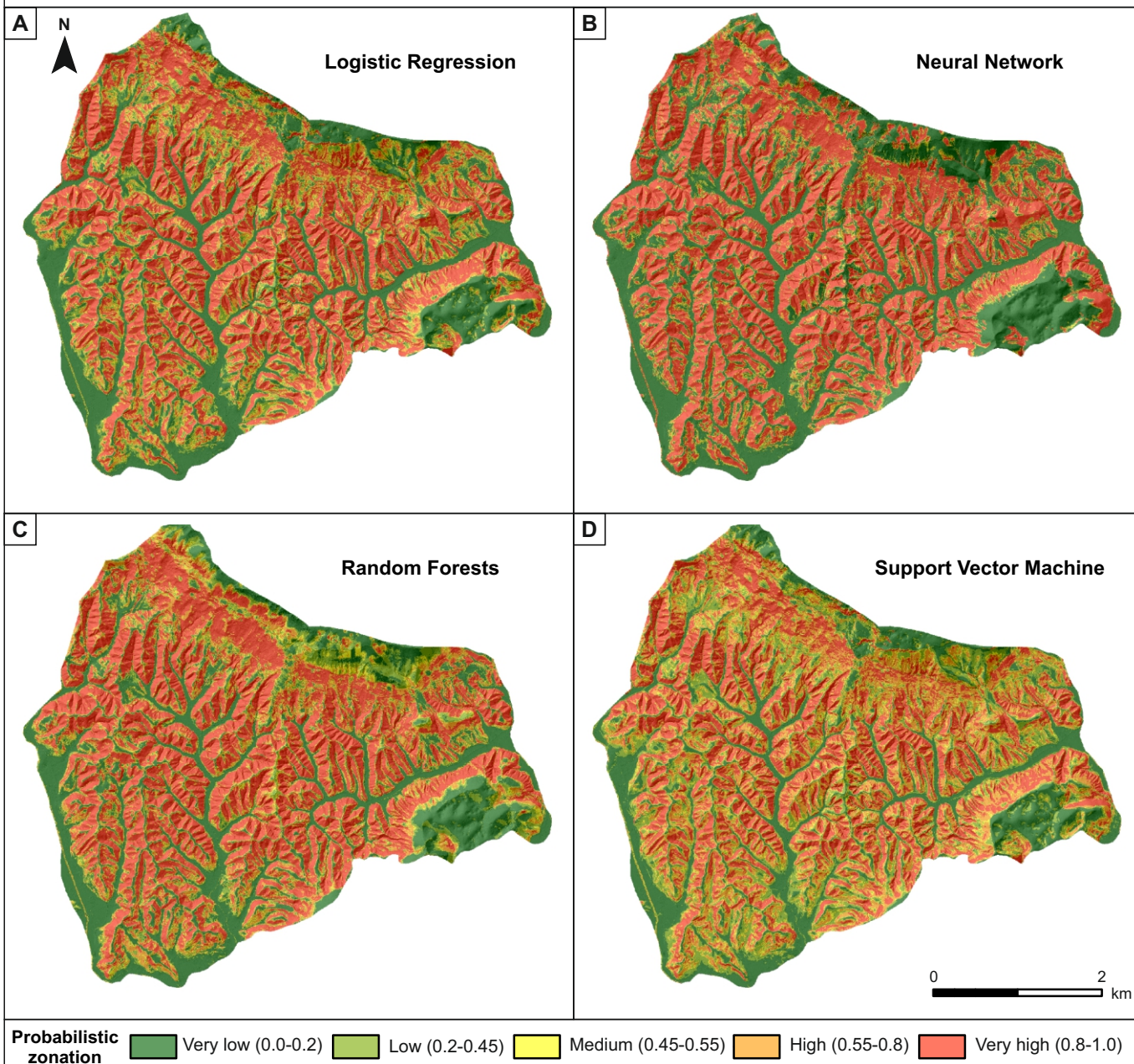
Supplementary material S5

Scenario S5-PM_s (polygon unstable, mapped stable polygon, smoothed LCF)



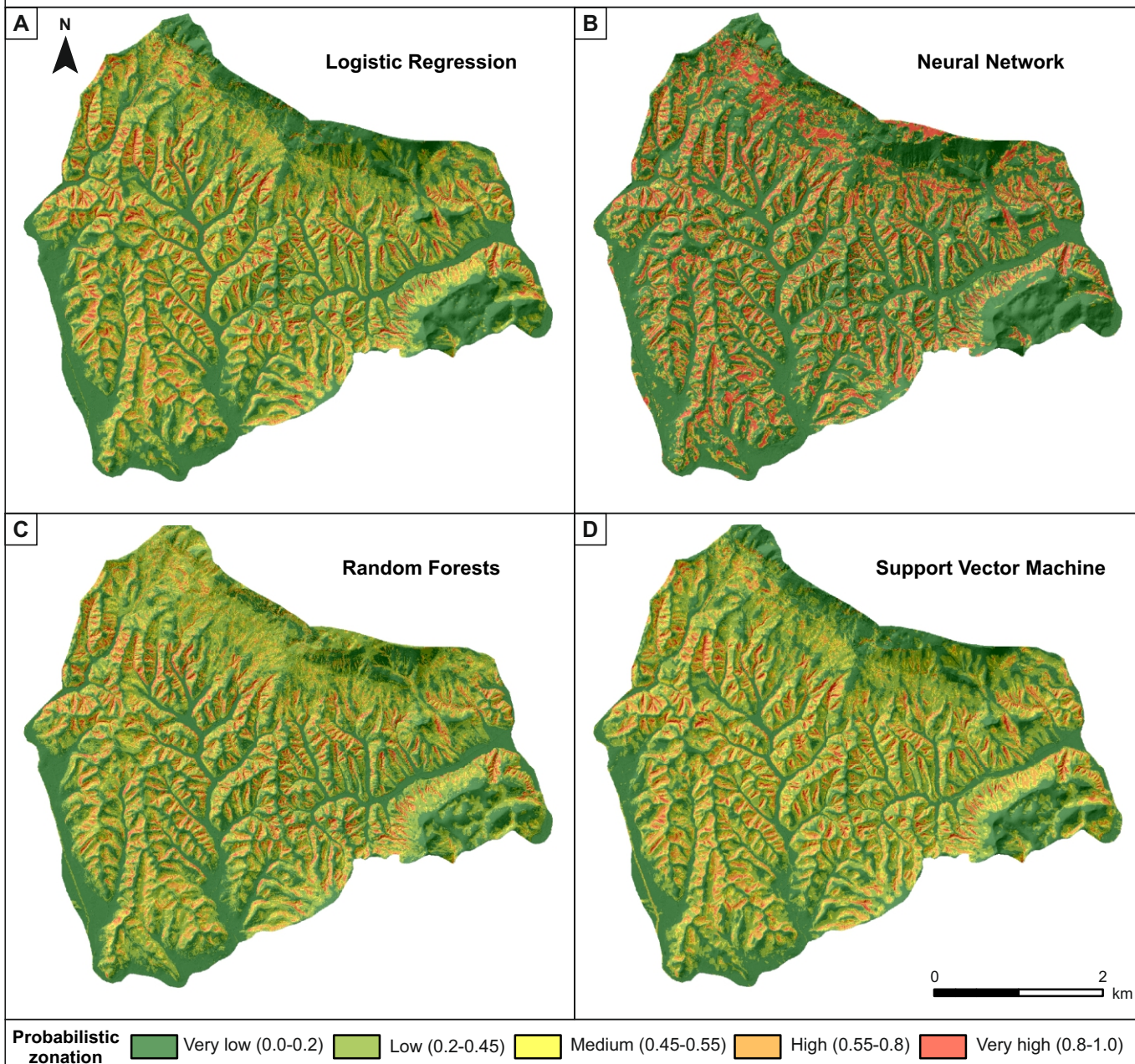
Supplementary material S6

Scenario S6-BbM_r (buffer polygon unstable, buffer mapped stable polygon, regular LCF)



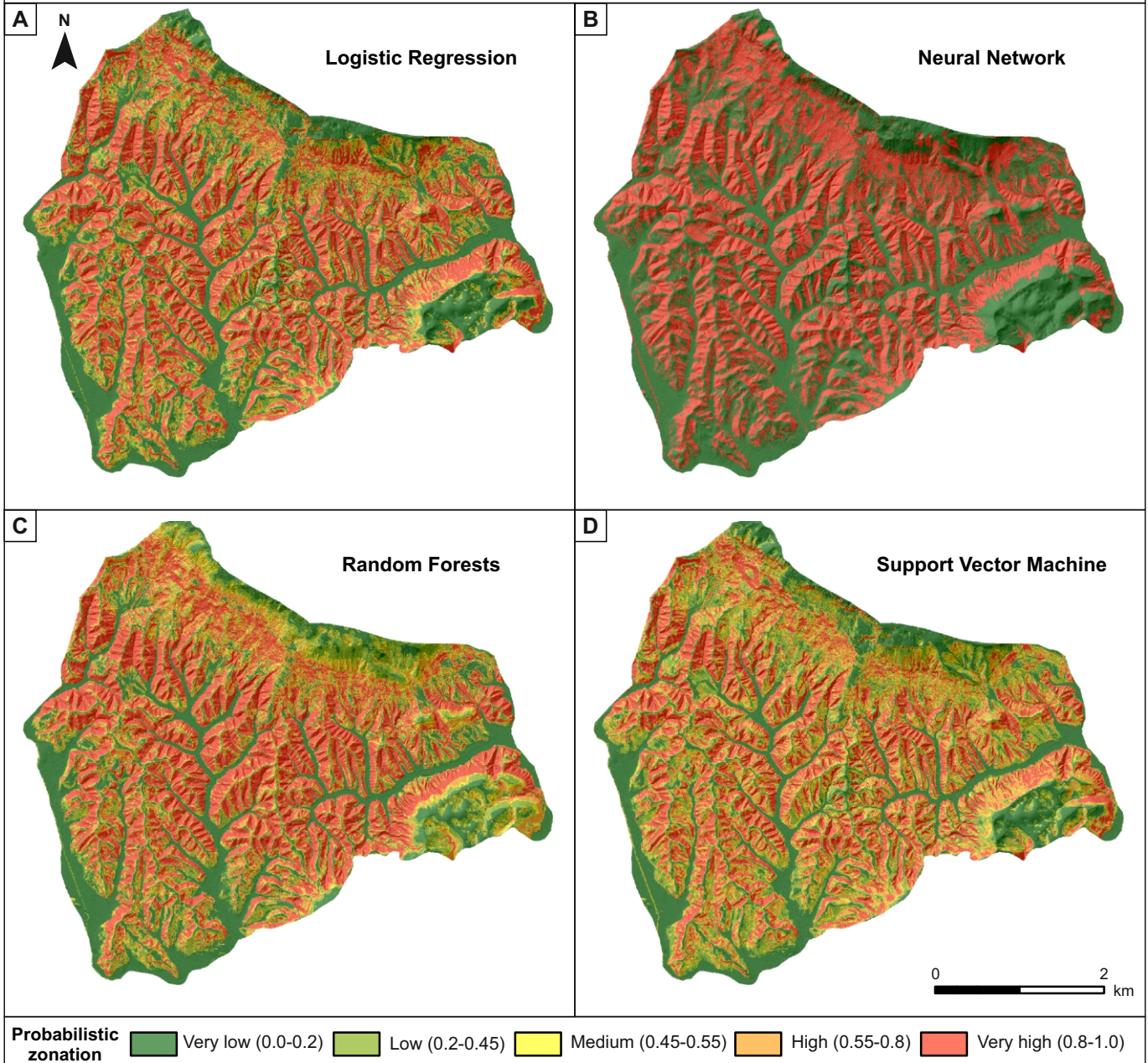
Supplementary material S7

Scenario S7-CR_r (landslide centroid unstable, random point stable, regular LCF)



Supplementary material S8

Scenario S8-CcM_r (landslide centroid unstable, centroid from mapped stable polygon, regular LCF)



2.4 Landslide susceptibility assessment on a large scale in the Podsljeme area, City of Zagreb (Croatia)

Authors:

Sanja Bernat Gazibara, Marko Sinčić, Martin Krkač, Hrvoje Lukačić, Snježana Mihalić Arbanas

Published in:

Journal of Maps 19(1)

Published on:

17 January 2023

DOI:

<https://doi.org/10.1080/17445647.2022.2163197>



Landslide susceptibility assessment on a large scale in the Podsljeme area, City of Zagreb (Croatia)

Sanja Bernat Gazibara , Marko Sinčić, Martin Krkač , Hrvoje Lukačić and Snježana Mihalić Arbanas

Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Zagreb, Croatia

ABSTRACT

The study presents a landslide susceptibility assessment on a large scale in the City of Zagreb (Croatia). The susceptibility analysis was performed using the Weight of Evidence model in the pilot area (21 km²) and applying the obtained weight values for each class of conditioning factors in the study area (130 km²). The input data were LiDAR-based landslide inventory and six conditioning factors derived from 5 m LiDAR DTM, 5 m SfM DEM, and geological and land-use maps. The validation of the susceptibility assessment for the study area was evaluated with a ROC curve, which showed a high prediction rate (AUC = 84.4%), similar to the prediction rate for the pilot area (AUC = 86.9%). Based on the results, it can be concluded that the proposed method for large-scale landslide susceptibility assessment, where susceptibility conditions are defined in smaller pilot areas, can be applied to larger research areas with similar geomorphological and geological conditions.

ARTICLE HISTORY

Received 1 July 2022
Revised 29 November 2022
Accepted 21 December 2022

KEYWORDS

Landslide susceptibility modelling; large scale; bivariate statistical method; LiDAR; Croatia

1. Introduction

A landslide susceptibility map of a certain area is a valuable tool in landslide hazard management as it identifies the degree of the spatial probability of an area to landslide occurrence (Guzzetti et al., 1999). Statistically-based methods for landslide susceptibility modelling are most often used for areas of 10 to 10,000 km², i.e. for landslide maps on a large scale (1: 5,000 to 1: 25,000), and the result is a quantitative landslide susceptibility assessment. The main assumption in quantitative statistical methods is that landslides occur throughout a study area within the same landslide conditioning factor combinations (Corominas et al., 2013). Furthermore, a lack of knowledge about different landslide types and geomorphological processes in the research area affects the selection of relevant landslide susceptibility factors and the final landslide susceptibility assessment. Therefore, the quality of the derived landslide susceptibility map mostly depends on the quality of the input data, i.e. the spatial resolution and accuracy of the landslide conditioning factor maps and the completeness and accuracy of the landslide inventory map (Ardizzone et al., 2002). Of all the additional data that a landslide inventory map can contain, spatial accuracy is the minimal and the most important requirement during landslide mapping (Aleotti & Chowdhury, 1999; Fell et al., 2008a, 2008b; Malamud et al., 2004). For large-scale inventory mapping, landslides are most often mapped as polygons, i.e. the researcher in the

field or on a high-resolution remote sensing image identifies landslide features and interprets the landslide boundary. However, in the case of landslide inventory mapping for a susceptibility assessment of larger areas, landslides are often mapped as a point, e.g. a point in the landslide centre or on the main scarp (Atkinson et al., 1998; Beguería, 2006; Eeckhaut et al., 2007; Felicísimo et al., 2013). The advantage of point-only landslide mapping is the higher mapping speed, i.e. it is possible to map much larger areas in a relatively short period, and the disadvantage is that the landslide runout is not considered in landslide susceptibility modelling. In the last fifteen years, various remote sensing techniques and applications have been used for landslide hazard mapping (Guzzetti et al., 2012; Jaboyedoff et al., 2012; Scaioni et al., 2014; Tarroli, 2014; Ray et al., 2020). The advantage of the LiDAR (Light Detection and Ranging) technique and data derived by ALS (Airborne Lidar Scanning) compared to other remote sensing techniques is ground detection in forested terrain and the possibility of creation of high-resolution bare-earth digital elevation models (DEMs), which enables the identification and mapping of small and shallow landslides in densely vegetated areas (Bernat Gazibara, 2019; Chigira et al., 2004; Domlija, 2018; Eeckhaut et al., 2007; Razak et al., 2011). Furthermore, the landslide maps obtained through the visual analysis of LiDAR-derived DTMs had better statistics for the landslide size (area) than the inventories obtained through

CONTACT Sanja Bernat Gazibara sbernat@rgn.hr Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Pierotijeva 6, Zagreb 10000, Croatia

Supplemental map for this article is available online at <https://doi.org/10.1080/17445647.2022.2163197>

© 2023 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group on behalf of Journal of Maps

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

field mapping or the interpretation of aerial photographs (Bernat Gazibara et al., 2019a, 2019b; Eeckhaut et al., 2007; Razak et al., 2011).

Hilly area in the Podsljeme area, City of Zagreb (Croatia), is prone to small ($10\text{--}10^3\text{ m}^2$) to medium-sized ($<10^5\text{ m}^2$) and superficial to moderately shallow ($<20\text{ m}$) soil slides (Mihalić Arbanas et al., 2017). An exception is the Kostanjek landslide (Krkač et al., 2017, 2020) in the western part of the Podsljeme area, a large deep-seated landslide ($32 \times 10^6\text{ m}^3$) caused by mining activities in 1965. Despite the relatively small volume of most landslides, they cause significant damage to buildings and infrastructure because of the high landslide density. The main landslide problem in the Podsljeme area is a lack of prognostic maps and adequate slope stability assessments. For the pilot area (21 km^2) in the Podsljeme area, City of Zagreb (Croatia), the detailed and complete LiDAR-based landslide inventory map was prepared by (Bernat Gazibara et al., 2019b). LiDAR-based landslide inventory (21 km^2) will be used for landslide susceptibility modelling on a large scale by applying the statistical method of the Weight of Evidence (WoE) model. The main objective of the study is to derive a landslide susceptibility map for the study area (130 km^2) based on susceptibility analysis of the input data layers for the pilot area (21 km^2) in the Podsljeme area. Furthermore, the study will analyse the possibility of defining susceptibility conditions in smaller pilot areas on detailed LiDAR DTM input data sets and applying them for large-scale landslide susceptibility assessment of large research areas in northwestern Croatia due to similar geomorphological and geological conditions. Therefore, landslide susceptibility assessment in the Podsljeme area can provide necessary information for disaster loss reduction and serve as an example and guideline for sustainable land use planning at a local level in the City of Zagreb and northwestern Croatia.

2. Study area

The City of Zagreb is located in the western part of NW Croatia, and it is part of the European Pannonian Basin. Within the city limits, the relief changes in the north to south direction from the ridge of Medvednica Mt., its southeastern slopes, through the alluvial floodplain of the Sava River to and to the Vukomeričke Gorice hills (Mihalić Arbanas et al., 2012). The southeast foothill step of the Medvednica Mt., i.e. Podsljeme area, represents a denudation-accumulation structure in which slope movements and fluvial-denudation geomorphological processes predominate (Bognar, 1992). The pilot area for which a detailed and complete LiDAR-based landslide inventory (Bernat Gazibara et al., 2019b) has been prepared is located in the western part of the Podsljeme area and has a

total area of 21 km^2 . Due to similar geomorphological, geological and hydrological conditions (Figure 1) and degree of construction, it is possible to apply the results of landslide susceptibility modelling from the pilot area (21 km^2) to the broader area of the Podsljeme area, i.e. study area (130 km^2).

According to the official land use of the City of Zagreb, there are approx. 70,400 buildings in the study area, with approx. 273,000 inhabitants (CBS, 2011), i.e. the average population density is 2090 inhabitants/ km^2 . The current land use includes 52.1 km^2 of agricultural areas, 40.7 km^2 of artificial surfaces, 37.3 km^2 of forests, and 0.5 km^2 of water surfaces. The elevation ranges from 114 to 437 m a.s.l., and the prevailing slope angles (59% of the study area) range from 5° to 20° , while 71% of slopes have slope angles $>5^\circ$. The study area is mostly comprised of Miocene and Quaternary sediments (Šikić, 1995), while the oldest rocks are from the late Paleozoic era. The top parts of Miocene deposits are fine-grained soils, mostly silts, while marls and limestones predominantly characterise the middle Miocene deposits. The Quaternary deposits are heterogeneous mixtures of mostly impermeable clayey soils. The geological contact between the Miocene sandy silts and the Quaternary clayey soils is highly susceptible to sliding (Jurak et al., 2008).

The climate of the City of Zagreb is continental with a mild maritime influence and mean annual precipitation (MAP) of 887 mm (Zaninović et al., 2008). In the study area, landslides are mainly triggered by rainfall, e.g. more than 50 landslides were (re)activated in the winter period of 2012/2013 (Bernat et al., 2014a, 2014b). In the Podsljeme area, data for landslide inventories were systematically collected five times in the last 50 years (Miklin et al., 2007; Podolszki, 2014; Polak et al., 1979; Šikić, 1967). In addition, the latest landslide inventory map was made for the pilot area (21 km^2) in the Podsljeme area based on a visual interpretation of LiDAR DTM derivatives (Bernat Gazibara, 2019; Bernat Gazibara et al., 2019b). Furthermore, the only available landslide susceptibility map for the Podsljeme area in the City of Zagreb was made based on field geomorphological mapping (Polak et al., 1979). A landslide susceptibility map from 1979 depicts four zones of relative landslide susceptibility, with 10.5% of the area classified as stable terrain, 78% as conditionally stable slopes, 11% as conditionally unstable slopes and 0.5% as unstable slopes.

3. Data and methods

Airborne laser scanning (ALS) of the 21 km^2 pilot area was undertaken in the frame of scientific research project “Methodology development for landslide susceptibility assessment for land-use planning based on LiDAR technology” (LandSlidePlan, HRZZ IP-2019-

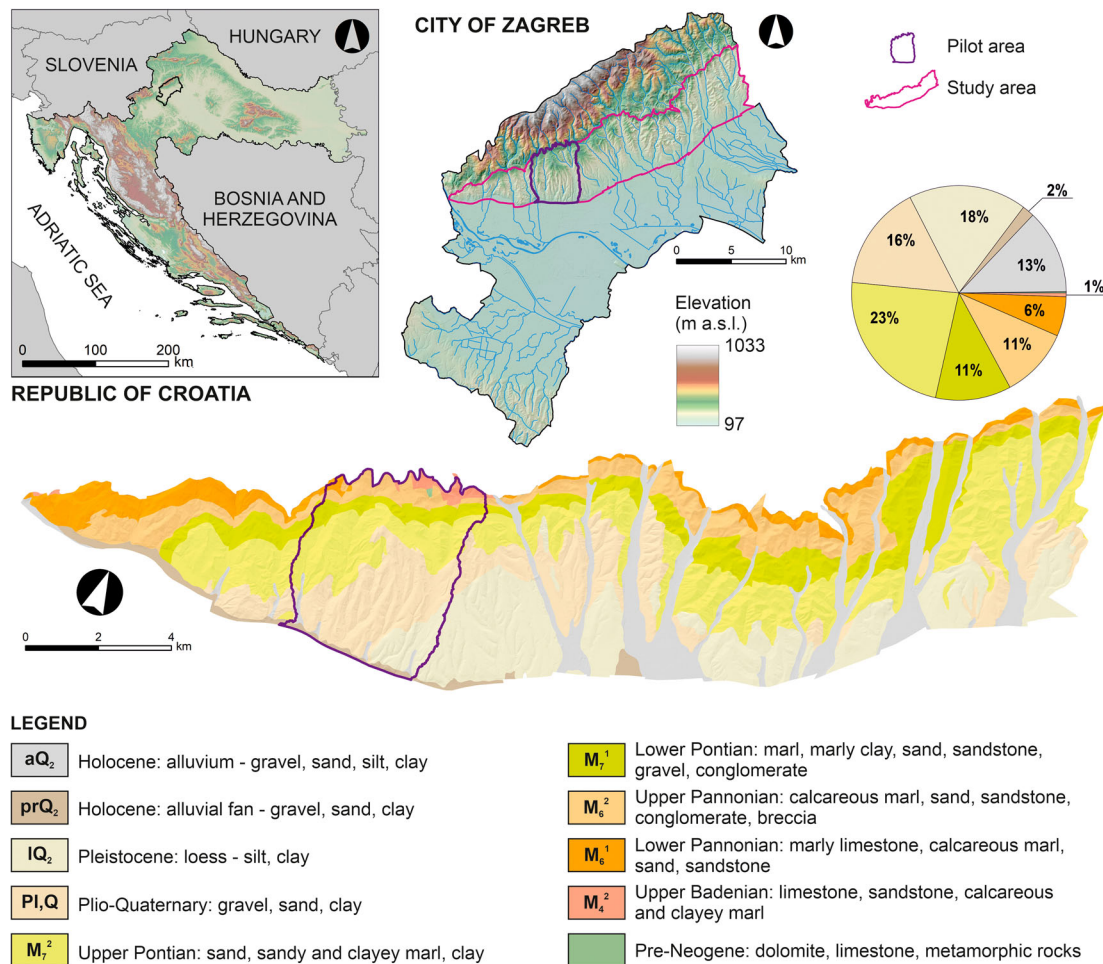


Figure 1. Location of the pilot area (21 km²) and the study area (130 km²) within the City of Zagreb and its lithostratigraphic setting.

04-9900), funded by the Croatian Science Foundation (Bernat Gazibara et al., 2022) in March 2020, corresponding to Croatia's winter leaf-off period. The ALS was performed with a Eurocopter EC 120B flying at 60 km/h at an altitude of 700 m a.s.l. The aircraft was equipped with a Reigl LMS-Q780 long-range airborne laser scanner. The obtained point cloud has 49% of points classified as terrain, while 33% were unclassified, 18% were classified as vegetation, and <0.5% of points were classified as noise. The density of the bare earth points is 13.41 points/m², i.e. the average point spacing is 0.2 m. LiDAR DTM for the visual interpretation of landslides was derived in 0.3 m resolution, using Kriging interpolation. LiDAR DTM for preparing geomorphological and hydrological conditioning factors for the pilot area (21 km²) was derived in 5 m resolution using the Local polynomial interpolation method.

The LiDAR DTM derivatives used to interpret the landslide morphology were hillshade, slope angle and 0.5 m contour lines. In addition, orthophoto images from 2011 to 2020 were used to check the morphological forms along roads and houses, such as artificial fills and cuts, similar to landslides on DTM

derivatives. Landslide identification on the LiDAR DTM derivatives was based on visual interpretation of landslide features by experienced landslide expert (e.g. concave main scarps, hummocky landslide bodies and convex landslide toes). As detailed described in Bernat Gazibara et al. (2019b), landslide mapping for the pilot area (21 km²) in the Podsljeme area was performed at a large-scale (1:100–1:500) to ensure the correct delineation of the boundaries based on landslide features. As a result, the total number of landslides in the inventory is 702, while the total area of mapped landslides is 0.5 km² or 2.43% of the pilot area. The mean landslide density is 33.3 slope failures per square kilometre. The size of the mapped landslides ranges from a minimum value of 43 m² to a maximum of 8,064 m², while the most frequent landslides in the inventory have an area of 400 m². The detailed and complete landslide inventory map of the pilot area represents a combination of seasonal inventory, landslides (re)activated in the winter of 2012/2013 and March 2018, and historical landslide inventory (Bernat Gazibara et al., 2019b). The landslide inventory for the pilot area (21 km²) was used for landslide susceptibility modelling and defining

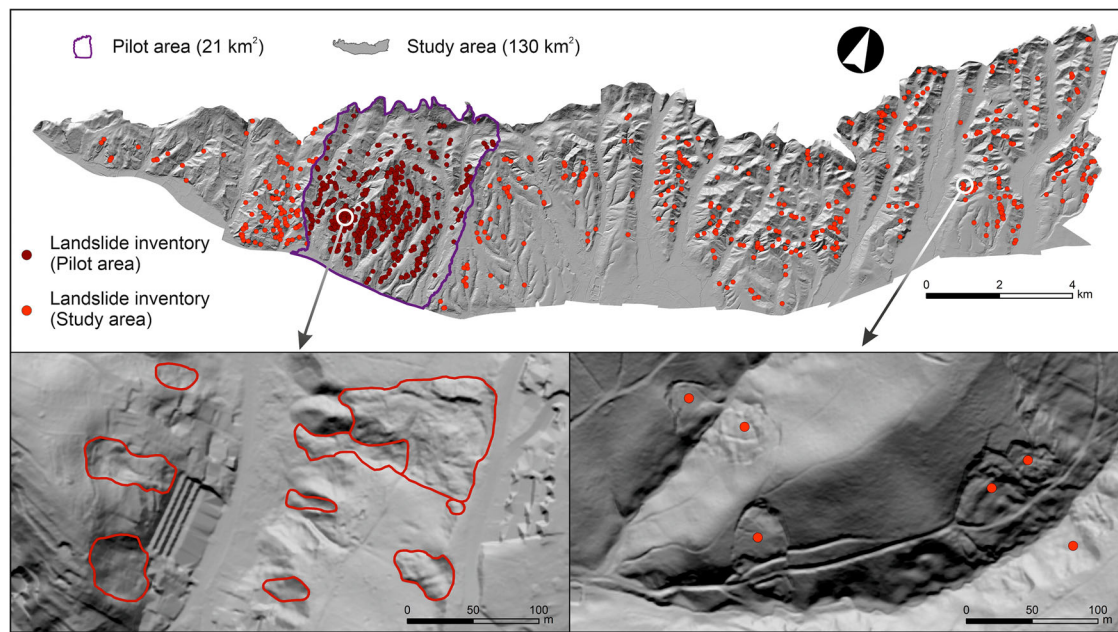


Figure 2. Landslide inventories for the pilot area (21 km²) and the study area (130 km²) on the hillshade derived from LiDAR DTM.

weight values for each factor map class that will be further used for landslide susceptibility assessment of the study area (130 km²). The landslide inventory for the susceptibility verification of the study area consists of 507 landslides depicted with the point in the centre of the identified landslide feature (Figure 2). The landslide inventory for the study area (130 km²) is not sustainably complete but can be considered representative for landslide susceptibility verification.

Input data layers for landslide susceptibility assessment were selected expertly based on the study area conditions, the purpose and scale, and data availability (Bernat Gazibara, 2019). As a result, a total of 6 landslide conditioning factors (Table 1) were selected and prepared for the pilot area (21 km²) and study area (130 km²): (i) slope gradient map; (ii) terrain dissection map; (iii) lithology map; (iv) proximity to geological contact; (v) proximity to drainage network; and (vi) land-use map. In addition, data preparation and digitalisation of input data for proximity to geological contact and drainage network were done for a broader study area. As a result, a more considerable extent of vector data was used to derive proximity factor maps. The geomorphological and hydrological

conditioning factors for the pilot area (21 km²) were prepared from 5 m LiDAR DTM and for the study area (130 km²) from 5 m SfM DEM, which in Croatia represents more accessible data than LiDAR data. Considering the LiDAR point cloud as a reference point, the root mean square error (RMSE) for 5 m LiDAR DTM is 0.559, and for 5 m sFM DEM is 1.15. The mean absolute error (MAE) for 5 m LiDAR DTM is 0.328, and for 5 m sFM DEM is 0.859.

Furthermore, the pairwise conditional independence test between all landslide conditioning factor classes was analysed using LAND-SUITE V1.0 (Rossi et al., 2021; Rossi & Reichenbach, 2016), i.e. LAND-SVA: LANDslide – Susceptibility Variable Analysis which includes Pairwise correlation analysis and Multicollinearity test. Detailed descriptions of the applied LAND-SUITE tests are described in the paper by Rossi et al. (2021).

In this paper, a bivariate statistical method, i.e. Weights-of Evidence (WoE) model (Agterberg et al., 1990; Bonham-Carter, 1994; Bonham-Carter et al., 1990; Sinčić et al., 2022) was applied in a GIS environment to derive quantitative spatial information on the predisposition to landslides. Using the WoE method,

Table 1. List of landslide conditioning factors used in the susceptibility analyses in the Podsljeme area.

Landslide conditioning group	Landslide conditioning factor, alias	Source	Method and/or derivation tool
Geomorphological group	Slope gradient	LiDAR DTM 5 m	ArcGIS 10.8 Slope tool
	Terrain dissection, TD	SfM DEM 5 m	Evans et al., (2014)
Geological group	Lithology (rock type)	Basic Geological Map 1:100 000 sheet Zagreb, (Šikić 1967)	Digitisation
	Proximity to geological contact	and sheet Ivanić Grad (Basch, 1983)	Digitisation, ArcGIS 10.8 Multiple Ring Buffer tool
Hydrological group	Proximity to drainage network	LiDAR DTM 5 m	ArcGIS 10.8
		SfM DEM 5 m	Spatial Analyst toolbox
Anthropogenic group	Land-use	(City of Zagreb 2011)	–

each conditioning factor map overlaps with the landslide inventory map to obtain the frequency of landslides in each class of all landslide factor maps. Then, the relative influence of the observed landslide factor is determined by comparing the landslide density in the factor classes with the landslide density in the entire pilot area. After calculating densities, different methods can determine weighting factors (Coe et al., 2004). In the case of the WoE method, the final weight of evidence value (W_{map}), used for landslide susceptibility modelling, is acquired using Equation (3) (van Westen et al., 2008):

$$W_i^+ = \log_e \frac{Npix_1/(Npix_1 + Npix_2)}{Npix_3(Npix_3 + Npix_4)} \quad (1)$$

$$W_i^- = \log_e \frac{Npix_2(Npix_1 + Npix_2)}{Npix_4(Npix_3 + Npix_4)} \quad (2)$$

$$W_{\text{map}} = W_i^+ - W_i^- + \sum W_i^- \quad (3)$$

Where: $Npix_1$ – number of pixels for present landslides and present potential landslide causal factor; $Npix_2$ – number of pixels for present landslides and absent potential landslide causal factor; $Npix_3$ – number of pixels for absent landslides and present potential landslide causal factor; $Npix_4$ – number of pixels for absent landslides and absent potential landslide causal factor; W_i^+ – positive weights (indicating the importance of the presence of the factor); W_i^- – negative weights (indicating the importance of the absence of the factor); W_{map} – weight of evidence value (weight factor).

The accuracy of derived landslide susceptibility maps was performed using ROC curves (Lobo et al., 2008), i.e. a graphical representation of sensitivity to 1-specificity (false positive rate). If the ROC curve is derived from the set of landslides used for susceptibility modelling, it represents the success rate of the landslide susceptibility map. Likewise, the ROC curve derived with landslide inventory for model validation shows the prediction rate of the landslide susceptibility map. Out of 702 landslides in the LiDAR-based inventory for the pilot area (21 km²), 50% of the landslides were selected for landslide susceptibility analyses, while the remaining 50% of the landslides were used to verify the results in the pilot area. Additional, 507 point-mapped landslides were used to verify the landslide susceptibility map for the whole extent of the study area (130 km²).

Furthermore, landslide susceptibility zonation was done using the ROC curve cut-off values defined by Bernat Gazibara (2019): (i) low landslide susceptibility class was defined as the area where only 1% of the cumulative landslide area is located; (ii) medium landslide susceptibility class is the area where maximum 5% of the cumulative landslide area is located; (iii) high landslide susceptibility class is the area where maximum 15% of the cumulative landslide area is located; (iv) very high landslide susceptibility class is the area where 85% of the cumulative area landslide landslides is located.

4. Results

Before performing landslide susceptibility analysis in the Podsljeme area, the pairwise conditional independence test of the input landslide conditioning factors was done using Pairwise correlation analysis in the LAND-SUITE V1.0 (Rossi et al., 2021). Inspection of the correlation results (Table 2) confirms that there are no correlations within six landslide conditioning factors selected for landslide susceptibility assessment, assuming a Pearson's R absolute value of 0.5 as the threshold for detecting correlations. The most significant pairwise correlation was calculated between slope gradient and land use, where the value of the Pearson correlation coefficient is -0.44 . Since the value is lower than the threshold, both maps were included in further analyses because they represent important conditioning factors for the landslide susceptibility assessment in the study area.

According to the analyses of landslide conditioning maps, the slope gradient classes generally have the highest impact on the landslide susceptibility. The W_{map} values show that the most susceptible classes are those with gradients between 25 and 45°. Terrain dissection classes between 0.2 and 0.6 influence negatively on slope stability. The only lithology units negatively influencing landslide susceptibility are Plio-Quaternary and Lower Pontian, with relatively low W_{map} values. In the proximity to the geological contact map, classes further from the contact generally negatively influence landslide susceptibility, which is not expected. The only classes in the proximity to the drainage network factor map and in the land-use factor map, which negatively influence landslide susceptibility, are class within 50 m of the drainage network and the forests area class.

Table 2. Results of the Pairwise correlation analysis in the LAND-SUITE V1.0 (Rossi et al., 2021).

	Slope gradient	Terrain dissection	Lithology (rock type)	Proximity to geological contact	Proximity to drainage network
Terrain dissection	-0.06				
Lithology (rock type)	-0.26	0.05			
Proximity to geological contact	0.12	-0.02	-0.34		
Proximity to drainage network	-0.10	0.24	0.08	-0.11	
Land-use	-0.44	0.12	0.27	-0.14	0.15

The landslide susceptibility analysis requires the classification of each conditioning factor into a number of classes. The slope gradient and terrain dissection were divided into ten classes, i.e. the class range of slope degree was 5°, and the class range of terrain dissection was 0.1. The lithology was divided into ten classes based on the Basic Geological Map 1:100

000 (Basch, 1983; Šikić et al., 1978). The proximity to geological contact was divided into nine classes with a class range of 100 m. The proximity to drainage was divided into five classes with a class range of 50 m. Finally, land use was classified into four classes, namely artificial areas, agricultural areas, forests and water areas. For each class of the six landslide

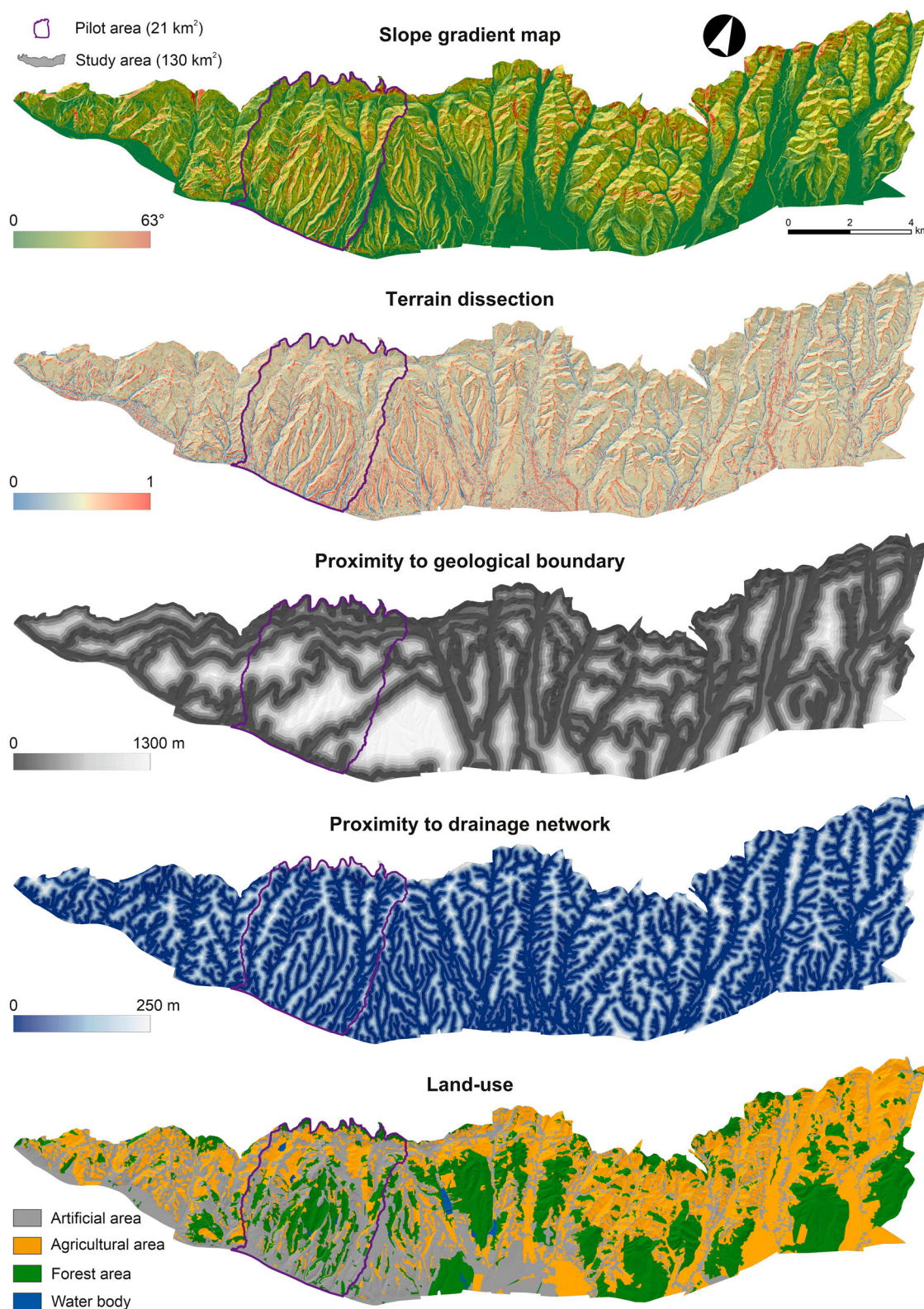


Figure 3. Landslide conditioning factors used in landslide susceptibility modelling of the study area in the Podsljeme area, City of Zagreb.

conditioning maps (Figures 1 and 3) in the pilot area (21 km²), a weight value was defined by applying the WoE model (Table 3).

Landslide susceptibility maps for the pilot area (21 km²) and the study area (130 km²) were derived by summing (overlapping) the factor maps according to the assigned weight values. As a result, the W_{map} value for the WoE model in the pilot area (Figure 4a) ranges from -12.14 to 6.03 , and for the study area, from -12.45 to 6.03 . The performance of the WoE model was evaluated using the ROC curve, i.e. the area under the curve (AUC). The success rate of the susceptibility model in the pilot area (21 km²) calculated based on the training set is $\text{AUC} = 85.3\%$, and the predictive rate is $\text{AUC} = 86.9\%$. In addition,

validation of the landslide susceptibility map for the study area (130 km²) was performed with the point-based landslide inventory, consisting of 507 landslides, which showed the high predictive rate of the susceptibility model, $\text{AUC} = 84.4\%$. Moreover, success and prediction rates for the pilot area (21 km²) and prediction rate for the study area (130 km²) differ by less than 3%, indicating the high reliability and quality of the model. Furthermore, the pilot area (21 km²) resulted in two landslide susceptibility maps. The first landslide susceptibility map was based on LiDAR data, i.e. geomorphological and hydrological conditioning factors were derived from 5 m LiDAR DTM, and the second landslide susceptibility map was the product of landslide conditioning factors

Table 3. Calculated weights for landslide conditioning factors based on the WoE model.

LCF group	LCF map	Class description	Class area (km ²)	Landslide area (km ²)	W_{map}
Geomorphological	Slope gradient (°)	0–5	3.95	0.0025	−3.2747
		5–10	5.23	0.0220	−1.3743
		10–15	4.35	0.0376	−0.5299
		15–20	3.39	0.0588	0.3234
		20–25	2.19	0.0624	0.9172
		25–30	1.10	0.0508	1.4263
		30–35	0.52	0.0334	1.7246
		35–40	0.22	0.0135	1.6144
		40–45	0.05	0.0025	1.3306
		>45	0.01	0.0003	0.5685
		SUM	21.01	0.2838	
	Terrain dissection	0.0–0.1	0.36	0.0012	−1.4586
		0.1–0.2	0.79	0.0075	−0.4022
		0.2–0.3	1.40	0.0226	0.1568
		0.3–0.4	2.71	0.0529	0.4065
		0.4–0.5	4.89	0.0899	0.3924
		0.5–0.6	5.05	0.0722	0.0364
		0.6–0.7	3.08	0.0291	−0.4487
		0.7–0.8	1.61	0.0069	−1.2496
		0.8–0.9	0.82	0.0012	−2.3483
		0.9–1.0	0.31	0.0003	−2.6102
Geological	Lithology (rock type)	SUM	21.01	0.2838	
		Holocene: alluvium	0.33	0.0047	−0.1408
		Holocene: alluvial fan	0.47	0.0000	0
		Pleistocene	1.88	0.0063	−1.6752
		Plio-Quaternary	7.77	0.1620	0.6337
		Upper Pontian	6.14	0.0977	0.0464
		Lower Pontian	1.81	0.0077	−1.4218
		Upper Pannonian	1.75	0.0053	−1.7728
		Lower Pannonian	0.16	0.0000	0
		Upper Badenian	0.67	0.0000	0
	Proximity to geological contact (m)	Pre-Neogene	0.03	0.0000	0
		SUM	21.01	0.2838	
		0–100	7.63	0.1114	0.1251
		100–200	4.89	0.0477	−0.4118
		200–300	2.88	0.0283	−0.3636
		300–400	2.07	0.0330	0.1896
		400–500	1.35	0.0176	−0.0434
		500–600	0.93	0.0143	0.1384
		600–700	0.66	0.0154	0.5722
		700–800	0.41	0.0094	0.5546
Hydrological	Proximity to drainage network (m)	>800	0.20	0.0069	1.0017
		SUM	21.01	0.2838	
		0–50	8.65	0.1523	0.4159
		50–100	6.35	0.0817	−0.1647
		100–150	3.99	0.0366	−0.5589
		150–200	1.63	0.0113	−0.8139
		>200	0.39	0.0019	−1.1243
		SUM	21.01	0.2838	
Anthro-pogenic	Land-use	Artificial areas	10.44	0.0191	−3.2600
		Forests	5.89	0.2316	1.8521
		Agricultural areas	4.56	0.0327	−1.3837
		Water areas	0.11	0.0005	−1.8470
		SUM	21.01	0.2838	

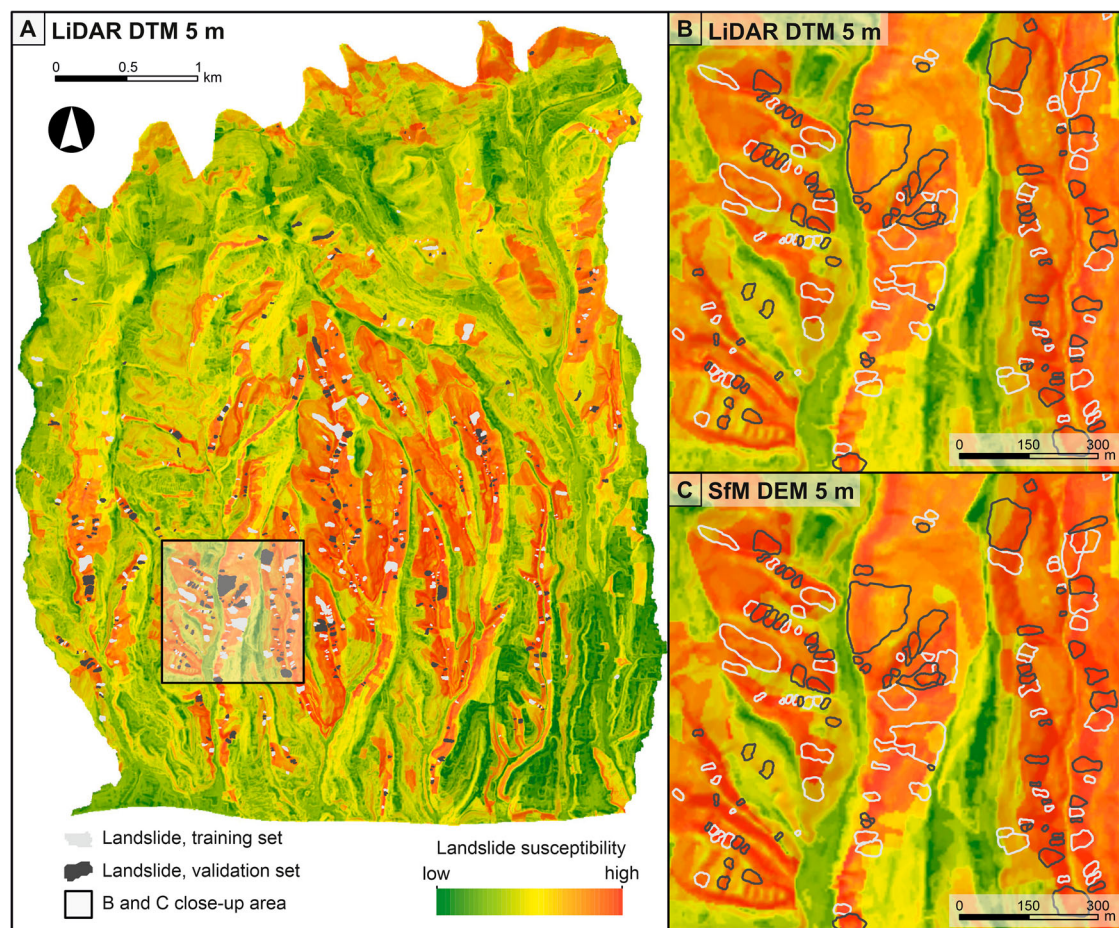


Figure 4. Result of the landslide susceptibility analysis in the pilot area (21 km²): (a) unclassified landslide susceptibility map; (b) the close-up extent of the landslide susceptibility map based on 5 m LiDAR DTM input data; (c) the close-up extent of the landslide susceptibility map based on 5 m SfM DEM input data.

derived from 5 m SfM DEM. Comparing the landslide susceptibility values in two locations of the pilot area (Figure 4b and c) shows minimal differences in derived landslide susceptibility maps, i.e. using high-quality 5 m LiDAR DTM does not result in higher spatial accuracy in regards to 5 m SfM DEM.

The landslide susceptibility zonation was done using the ROC curve cut-off values defined by Bernat Gazibara (2019), i.e. 1%, 5% and 85% of the cumulative landslide densities defined four susceptibility zones. As a result, the landslide susceptibility map of the pilot area (21 km²) was reclassified into the zones of low (8.3 km² or 39%), medium (4.4 km² or 21%), high (1.9 km² or 9%) and very high (6.4 km² or 31%) susceptibility. Furthermore, the landslide susceptibility map of the study area (130 km²), reclassified based on cut-off values defined on the ROC curve for the pilot area (21 km²), includes 51.1 km² or 39% of the low, 29.7 km² or 23% of the moderate, 13.9 km² or 11% of the high and 35.6 km² or 27% of the very high susceptibility zones. Therefore, there are minimal differences in landslide zone distribution in the pilot area (21 km²) and study area (130 km²), leading to the conclusion that the pilot area was representatively selected for landslide susceptibility analysis.

5. Discussion and conclusions

For the first time, the landslide susceptibility map for the part of the Podsljeme area (130 km²) in the City of Zagreb, Croatia, was prepared based on available input data, i.e. LiDAR-based landslide inventory (Bernat Gazibara et al., 2019b), 5 m LiDAR DTM, 5 SfM DEM, Basic Geological Map 1:100 000 (Basch, 1983; Šikić et al., 1978) and official land use of the City of Zagreb in large scale. Landslide susceptibility analysis was done for the pilot area (21 km²) using the WoE model and six conditioning factors, namely slope gradient, terrain dissection, lithology (rock type), proximity to a geological contact, proximity to drainage network, and land use. As a result, the success rate of the derived landslide susceptibility map (21 km²) is AUC = 85.3%, while the prediction rate is AUC = 86.9%. Based on the high accuracy of the landslide susceptibility model for the pilot area, the weight values of each class of six landslide conditioning factors were applied for the same factor maps in the study area (130 km²). Validation of the resulting susceptibility map for the study area shows a high prediction rate, i.e. the AUC is 84.4%. Therefore, it can be concluded that the proposed method for large-scale

landslide susceptibility assessment, where susceptibility conditions are defined in smaller pilot areas, can be applied to larger research areas where similar geomorphological and geological conditions prevail.

Three conditioning factors, namely slope gradient, terrain dissection, and proximity to drainage network, were derived from high-resolution LiDAR DTM and SfM DEM, and the land use conditioning map was derived from large-scale vector data available for the entire area of the City of Zagreb. The remaining two conditioning factor maps, lithology (rock type) and proximity to a geological contact were derived from the only available small-scale Basic Geological Map (M 1:100.000). Using small-scale data when conducting large-scale landslide susceptibility assessments leads to significant data loss and results in deviations from actual environmental conditions that should be avoided. On the other hand, high-resolution LiDAR data enables distinguishing the geological settings in more detail, improving geological contacts, and aggregating chronostratigraphic units into engineering formation units (Sinčić et al., 2022), which are more suitable for large-scale modelling and at the same time enables the improvement of the derived landslide susceptibility assessment. Based on the high prediction rate of the presented large-scale landslide susceptibility model, we can conclude that six selected conditioning factor maps represent a good correlation between the landslide occurrence and environmental conditions in the study area, and the large-scale geological data will increase the spatial accuracy of the final susceptibility maps.

Landslide susceptibility analysis was based on small to medium landslides, therefore the area of the large, deep-seated Kostanjek landslide (1 km²) was not recognised as an area of high landslide susceptibility. Furthermore, the landslide susceptibility maps for the pilot area based on 5 m LiDAR DTM and 5 m SfM DEM resulted in minimal differences in landslide susceptibility distribution. It follows that it is possible to conduct landslide susceptibility analysis in smaller pilot areas (10–30 km²) with available high-resolution data sets and apply them to larger regions with input data available in relatively lower resolution. The landslide susceptibility zonation for the study area (130 km²) in the Podsljeme area was done based on ROC curve cut-off values defined by Bernat Gazibara (2019), i.e. 1, 5, and 85% of the cumulative landslide densities defined four susceptibility zones. As a result, the landslide susceptibility map of the study area was classified into the low (51.1 km² or 39%), medium (29.7 km² or 23%), high (13.9 km² or 11%) and very high (35.6 km² or 27%) susceptibility zone. In the end, it can be concluded that different input data quality (LiDAR DTM vs SfM DEM) can provide similar information regarding large-scale landslide susceptibility assessment. Also, the presented large-scale

landslide susceptibility assessment would show significantly different results if we did not use sustainable complete, LiDAR-based landslide inventory for the modelling. Therefore, the derived landslide susceptibility map can be used in large-scale spatial planning and civil protection system, but also for more detailed preliminary analysis that will define areas with higher landslide susceptibility where detailed landslide mapping using high-resolution data is necessary to conduct.

Software

ArcMap 10.8 was used for the data analysis and construction of the main map and figures. The map and figures were imported into CorelDRAW 2020 for graphic adjustments and the final layout.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Croatian Science Foundation [grants number HRZZ IP-2019-04-9900 and HRZZ DOK-2020-01-2432], and by the SATREPS FY2008 under Grant for Croatian-Japanese joint research project.

ORCID

Sanja Bernat Gazibara  <http://orcid.org/0000-0002-5530-4381>

Martin Krkač  <http://orcid.org/0000-0002-0232-3394>

Snježana Mihalić Arbanas  <http://orcid.org/0000-0003-4036-3722>

Data availability statement

The data that support the findings of this study are available from the corresponding author, S.B.G, upon reasonable request.

References

- Agterberg, F. P., Bonham-Carter, G. F., & Wright, D. F. (1990). Statistical pattern integration for mineral exploration. In G. Gaál, & D. F. Merriam (Eds.), *Computer applications in resource estimation* (pp. 1–21). Elsevier.
- Aleotti, P., & Chowdhury, R. (1999). Landslide hazard assessment: Summary review and new perspectives. *Bulletin of Engineering Geology and the Environment*, 58 (1), 21–44. <https://doi.org/10.1007/s100640050066>
- Ardizzone, F., Cardinali, M., Carrara, A., Guzzetti, F., & Reichenbach, P. (2002). Impact of mapping errors on the reliability of landslide hazard maps. *Natural Hazards and Earth System Sciences*, 2(1/2), 3–14. <https://doi.org/10.5194/nhess-2-3-2002>
- Atkinson, P., Jiskoot, H., Massari, R., & Murray, T. (1998). Generalized linear modelling in geomorphology. *Earth Surface Processes and Landforms*, 23(13), 1185–1195.

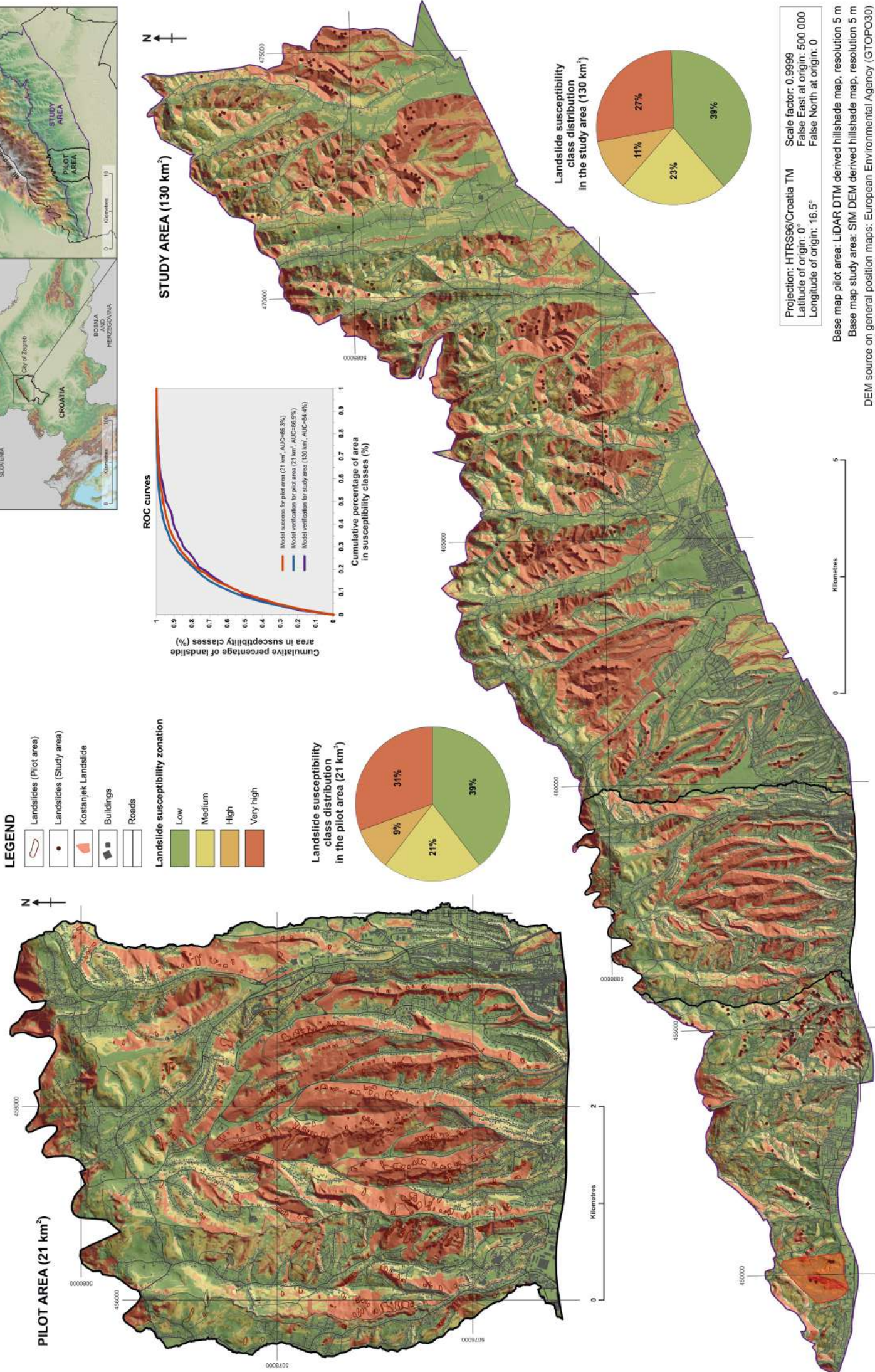
- [https://doi.org/10.1002/\(SICI\)1096-9837\(199812\)23:13<1185::AID-ESP928>3.0.CO;2-W](https://doi.org/10.1002/(SICI)1096-9837(199812)23:13<1185::AID-ESP928>3.0.CO;2-W)
- Basch, O. (1983). Basic geological map, scale 1:100,000, Ivanić – Grad, Sheet L33-81.
- Begueria, S. (2006). Validation and evaluation of predictive models in hazard assessment and risk management. *Natural Hazards*, 37(3), 315–329. <https://doi.org/10.1007/s11069-005-5182-6>
- Bernat, S., Mihalić Arbanas, S., & Krkač, M. (2014a). Landslides triggered in the continental part of Croatia by extreme precipitation in 2013. In G. Lollino, D. Giordan, G. B. Crosta, J. Corominas, R. Azzam, J. Wasowski, & N. Sciarra (Eds.), *Engineering geology for society and territory – volume 2* (pp. 1599–1603). Springer International Publishing.
- Bernat, S., Mihalić Arbanas, S., & Krkač, M. (2014b). Inventory of precipitation triggered landslides in the winter of 2013 in Zagreb (Croatia, Europe). In K. Sassa, P. Canuti, & Y. Yin (Eds.), *Landslide science for a safer geoenvironment* (pp. 829–835). Springer International Publishing.
- Bernat Gazibara, S. (2019). *Methodology for landslide mapping using high resolution digital elevation model in the Podsljeme area (City of Zagreb)* [Doctoral thesis]. Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb.
- Bernat Gazibara, S., Krkač, M., & Mihalić Arbanas, S. (2019a). Verification of historical landslide inventory maps for the Podsljeme area in the City of Zagreb using LiDAR-based landslide inventory. *Rudarsko-geološko-naftni Zbornik*, 34(1), 45–58. <https://doi.org/10.17794/rgn.2019.1.5>
- Bernat Gazibara, S., Krkač, M., & Mihalić Arbanas, S. (2019b). Landslide inventory mapping using LiDAR data in the City of Zagreb (Croatia). *Journal of Maps*, 15(2), 773–779. <https://doi.org/10.1080/17445647.2019.1671906>
- Bernat Gazibara, S., Mihalić Arbanas, S., Sinčić, M., Krkač, M., Lukačić, H., Jagodnik, P., & Arbanas, Ž. (2022). LandSlidePlan -Scientific research project on landslide susceptibility assessment in large scale. In J. Peranić, M. Vivoda Prodan, S. Bernat Gazibara, M. Krkač, S. Mihalić Arbanas, & Ž. Arbanas (Eds.), *Proceedings of the 5th regional symposium on landslides in Adriatic-Balkan Region* (pp. 99–106). Faculty of Civil Engineering, University of Rijeka and Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb.
- Bognar, A. (1992). Engineering-geomorphologic mapping – possibilities and development. *Acta Geographica Croatica*, 27, 173–185.
- Bonham-Carter, G. (1994). *Geographic information systems for geoscientists: Modelling with GIS* (1st ed.). Pergamon.
- Bonham-Carter, G. F., Agterberg, F. P., & Wright, D. F. (1990). Weights of evidence modelling: A new approach to mapping mineral potential.
- CBS. (2011). *Statistical reports – 2011 census*. Croatian Bureau of Statistics.
- Chigira, M., Duan, F., Yagi, H., & Furuya, T. (2004). Using an airborne laser scanner for the identification of shallow landslides and susceptibility assessment in an area of ignimbrite overlain by permeable pyroclastics. *Landslides*, 1(3), 203–209. <https://doi.org/10.1007/s10346-004-0029-x>
- City of Zagreb. (2011). Land use map of City of Zagreb.
- Coe, J. A., Godt, J. W., Baum, R. L., et al. (2004). Landslide susceptibility from topography in Guatemala. In W. Lacerda, M. Ehrlich, S. Fontoura, & A. Sayão (Eds.), *Landslides: Evaluation and stabilisation/glisement de Terrain: Evaluation et Stabilisation, set of 2 volumes: Proceedings of the Ninth International Symposium on Landslides, June 28–July 2, 2004 Rio de Janeiro, Brazil* (pp. 69–78). CRC Press.
- 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. (2013). Recommendations for the quantitative analysis of landslide risk. *Bulletin of Engineering Geology and the Environment*, 73, 209–263. <https://doi.org/10.1007/s10064-013-0538-8>
- Domlija, P. (2018). *Identification and classification of landslides and erosion phenomena using the visual interpretation of the Vinodol Valley digital elevation model* [Doctoral thesis]. Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb.
- Eeckhaut, M. V. D., Poesen, J., Verstraeten, G., Nyssen, J., Moeyersons, J., van Beek, L. P. H., & Vandekerckhove, L. (2007). Use of LIDAR-derived images for mapping old landslides under forest. *Earth Surface Processes and Landforms*, 32(5), 754–769. <https://doi.org/10.1002/esp.1417>
- Evans, J. S., Oakleaf, J., Cushman, S. A., & Theobald, D. (2014). An arcgis toolbox for surface gradient and geomorphometric modeling, version 2.0-0.
- Feliciísimo, ÁM, Cuartero, A., Remondo, J., & Quirós, E. (2013). Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: A comparative study. *Landslides*, 10(2), 175–189. <https://doi.org/10.1007/s10346-012-0320-1>
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W. Z., & on behalf of the JTC-1 Joint Technical Committee on Landslides and Engineered Slopes. (2008a). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102(3-4), 85–98. <https://doi.org/10.1016/j.enggeo.2008.03.022>
- Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W. Z., & on behalf of the JTC-1 Joint Technical Committee on Landslides and Engineered Slopes. (2008b). Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*, 102(3-4), 99–111. <https://doi.org/10.1016/j.enggeo.2008.03.014>
- Guzzetti, F., Carrara, A., Cardinali, M., & Reichenbach, P. (1999). Landslide hazard evaluation: A review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 31(1-4), 181–216. [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1)
- Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T. (2012). Landslide inventory maps: new tools old problem. *Earth-Science Reviews*, 112(1-2), 42–66. <https://doi.org/10.1016/j.earscirev.2012.02.001>
- Jurak, V., Ortolan, Ž., Ivšić, T., Herak, M., Šumanovac, F., Vukelić, I., Jukić, M., & Šurina, Z. (2008). Geotechnical and seismic microzonation of Zagreb – attempts and realisation. In J. Radić (Ed.), *Proceedings of the conference development of Zagreb* (pp. 99–108). SECON HDGK d.o.o.
- Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R., & Pedrazzini, A. (2012). Use of LIDAR in landslide investigations: A review. *Natural*

- Hazards*, 61, 5–28. <https://doi.org/10.1007/s11069-010-9634-2>
- Krkač, M., Bernat Gazibara, S., Arbanas, Ž., Sečanj, M., & Mihalić Arbanas, S. (2020). A comparative study of random forests and multiple linear regression in the prediction of landslide velocity. *Landslides*, 17(11), 2515–2531. <https://doi.org/10.1007/s10346-020-01476-6>
- Krkač, M., Špoljarić, D., Bernat, S., & Arbanas, S. M. (2017). Method for prediction of landslide movements based on random forests. *Landslides*, 14(3), 947–960. <https://doi.org/10.1007/s10346-016-0761-z>
- Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: A misleading measure of the performance of predictive distribution models. *Global Ecology and Biogeography*, 17(2), 145–151. <https://doi.org/10.1111/j.1466-8238.2007.00358.x>
- Malamud, B. D., Turcotte, D. L., Guzzetti, F., & Reichenbach, P. (2004). Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, 29(6), 687–711. <https://doi.org/10.1002/esp.1064>
- Mihalić Arbanas, S., Krkač, M., Bernat, S., & Arbanas, Ž. (2012). Use of existing data in the City of Zagreb (Croatia, Europe) for the purpose of geo-planning. In Z. Yingxin, C. Jungang, & R. Sterling (Eds.), *Advances in underground space development* (pp. 465–477). Research Publishing.
- Mihalić Arbanas, S., Sečanj, M., Bernat Gazibara, S., Krkač, M., Begić, H., Džindo, A., Zekan, S., & Arbanas, Ž. (2017). Landslides in the Dinarides and Pannonian Basin—from the largest historical and recent landslides in Croatia to catastrophic landslides caused by Cyclone Tamara (2014) in Bosnia and Herzegovina. *Landslides*, 14(6), 1861–1876. <https://doi.org/10.1007/s10346-017-0880-1>
- Miklin, Ž., Mlinar, Ž., Brkić, Ž., Hećimović, I., & Dolić, M. (2007). Detailed engineering geological map of Podsljeme urbanised zone at a scale of 1: 5,000 (DIGK-Phase I). Croatian Geological Survey.
- Podolszki, L. (2014). *Stereoscopic analysis of landslides and landslide susceptibility on the southern slopes of the Medvednica Mt.* [Doctoral thesis]. Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb.
- Ray, R. L., Lazzari, M., & Olutimehin, T. (2020). Remote sensing approaches and related techniques to map and study landslides. In R. Ray, & M. Lazzari (Eds.), *Landslides - investigation and monitoring*. IntechOpen. <https://doi.org/10.5772/intechopen.93681>
- Polak, K., Klemar, M., Nejškova, M., Radošević, N., Stepan Z., Miroslav M., & Križanić Z. (1979). Lithological analysis and categorisation of terrain according to soil stability of Medvednica slopes in the area of the City of Zagreb.
- Razak, K. A., Straatsma, M. W., van Westen, C. J., Malet, J.-P., & de Jong, S. M. (2011). Airborne laser scanning of forested landslides characterization: Terrain model quality and visualization. *Geomorphology*, 126(1-2), 186–200. <https://doi.org/10.1016/j.geomorph.2010.11.003>
- Rossi, M., Bornaetxea, T., & Reichenbach, P. (2021). LAND-SUITE V1.0: a suite of tools for statistically-based landslide susceptibility zonation. *Earth and space science informatics*.
- Rossi, M., & Reichenbach, P. (2016). LAND-SE: A software for statistically based landslide susceptibility zonation, version 1.0. *Geoscientific Model Development*, 9(10), 3533–3543. <https://doi.org/10.5194/gmd-9-3533-2016>
- Scaioni, M., Longoni, L., Melillo, V., & Papini, M. (2014). Remote sensing for landslide investigations: An overview of recent achievements and perspectives. *Remote Sensing*, 6(10), 9600–9652. <https://doi.org/10.3390/rs6109600>
- Šikić, K. (1967). Engineering geology of Zagreb – north and south.
- Šikić, K. (1995). *Geological guide of Mt. Medvednica*. Croatian Geological Institute-Croatian Geological Society, INA–Oil Industry.
- Šikić, K., Basch, O., & Šimunić, A. (1978). Basic geological map, scale 1:100,000, Zagreb, Sheet L33-80.
- Sinčić, M., Bernat Gazibara, S., Krkač, M., & Mihalić Arbanas, S. (2022). Landslide susceptibility assessment of the City of Karlovac using the bivariate statistical analysis. *Rudarsko-geološko-naftni Zbornik*, 37(2), 149–170. <https://doi.org/10.17794/rgn.2022.2.13>
- Tarolli, P. (2014). High-Resolution Topography for Understanding Earth Surface Processes: Opportunities and Challenges. *Geomorphology*, 216, 295–312. <https://doi.org/10.1016/j.geomorph.2014.03.008>
- van Westen, C. J., Castellanos, E., & Kuriakose, S. L. (2008). Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*, 102(3-4), 112–131. <https://doi.org/10.1016/j.enggeo.2008.03.010>
- Zaninović, K., Gajić-Čapka, M., Perčec Tadić, M., et al. (2008). Climate atlas of Croatia 1961 – 1990, 1971 – 2000.

LANDSLIDE SUSCEPTIBILITY ASSESSMENT ON A LARGE SCALE IN THE PODSLJEME AREA, CITY OF ZAGREB (CROATIA)

Sanja Bernat Gazibara¹, Martin Krkač¹, Marko Sinčić¹, Hrvoje Lukačić¹, Snježana Mihalčić Arbanas¹

¹ University of Zagreb, Faculty of Mining, Geology and Petroleum Engineering, Pierottijeva 6, 10000 Zagreb, Croatia



Map Design

The landslide susceptibility map for the pilot area (21 km²) in the City of Zagreb was derived based on LiDAR-based landslide inventory, 5 m LiDAR DTM derivatives (slope, terrain dissection and drainage network), geological data and land use using Weight of Evidence (WoE) model. The landslide susceptibility map for the study area (130 km²) was derived based on pilot area weight values for each class of conditioning factors in the study area (130 km²). All input data sets, preparation of landslide condition factors and landslide susceptibility analysis were done in ArcGIS 10.8. Landslide susceptibility maps for the pilot area (21 km²) and study area (130 km²) were derived as raster files and converted to vectors. After raster to polygon conversion, the Smooth tool in ArcGIS 10.8 was applied. The hillshade maps with a spatial resolution of 5m were exported from ArcMap 10.8 as a TIFF file. Landslide susceptibility maps, roads, buildings, and landslides were exported from ArcMap 10.8 as an EMF file. The final map with map title and authorship, map projection and datum information and general position map were constructed in Corel Draw 2020. The hillshade map is underneath the landslide susceptibility zonation layer (50% transparency), giving a vivid impression of the terrain and landslide susceptibility in the pilot and study area. The final map shows the position of mapped landslides and landslide susceptibility with respect to the buildings and roads in the pilot and study area.

3. DISCUSSION

3.1 Hypothesis (1)

High density LiDAR point cloud and high resolution orthophoto imagery enable the preparation of necessary input data for developing accurate landslide susceptibility models on a large scale.

The hypothesis is **confirmed**, as explained in detail in Chapter 2.1 (Sinčić et al., 2022b) and proven further in LSMg performed in Chapter 2.2. (Sinčić et al., 2025), Chapter 2.3 (Sinčić et al., 2024a) and Chapter 2.4 (Bernat Gazibara et al., 2023a). Considering source data for compilation of a landslide inventory, small and shallow landslides covered by vegetation cannot be delineated from widely used data such as DEM derivatives, DOF or Synthetic Aperture Radar leaving HR-DTM derivatives based on a high density LiDAR point cloud as the only alternative. Unlike other remote sensing techniques, LiDAR point cloud enables classification of points, thus removing the vegetation and revealing bare-earth relief, i.e. a DTM. Furthermore, based on the statistical analysis (frequency-size distribution graphs) performed with landslide data prepared for the study areas Hrvatsko Zagorje and Podsljeme Area (21 km²), large scale landslide inventory maps are considered as representative inventories ('universal distribution' according to Malamud et al. (2004)), ensuring the required accuracy for the large scale LSMg. Considering source data for the derivation of a series of LCF maps, more extensive and complex conclusions emerged from this research based on a systematic overview and mutual comparison of source data and analyses about input data creation. Geomorphological and most hydrological LCFs rely on DEMs or DTMs as source data, and the comparison showed significant differences (Sinčić et al., 2022b) due to variations in resolution or the origin of the source data. Despite the determined differences, doctoral research also proved that using a SfM DEM to derive LCFs resulted in a successful LSA for the whole Podsljeme Area (130 km²), equally as using a LiDAR based DTM did for the smaller part of the same Podsljeme Area (21 km²) (Bernat Gazibara et al., 2023a; Sinčić et al., 2025) or study area Hrvatsko Zagorje (Sinčić et al., 2024a). An important benefit of using high density LiDAR point cloud are the HR-DTM derivatives which enabled significant correction of spatially inaccurate geological LCFs, (small scale cartographic unit of soil/rock type and the geological contacts) presented in Sinčić et al. (2024a). The small scale geological source data available for the entire country, provides basic information for generating large scale geological maps by modifying spatial irregularities and inhomogeneities of geological units by using HR-DTM derivatives based on

a high density LiDAR point cloud. Anthropogenic LCFs are the most demanding to develop because they require following specific procedures and combined use of several source data layers. However, high density LiDAR point cloud and HR-DOF imagery are only necessary to obtain high spatial accuracy anthropogenic LCFs in a large scale LSA, whereas the remaining source data serve to speed up the finalizing process. It should be noted that small scale source data that are often publicly available (e.g. OSM), can serve as a good basic information, and can reduce time spent in collecting data, as in example with digitizing roads (or traffic infrastructure in general) (Sinčić et al., 2022b). For instance, out of approximately 214 km of traffic infrastructure, more than 40% was obtained through small scale source data, and the remaining was supplemented by digitizing on high density LiDAR point cloud derivatives (e.g. HR hillshade map) and HR-DOF imagery (Sinčić et al., 2022b). Detailed comparison conducted by Sinčić et al. (2022b) emphasized the high quality of LCFs containing land use information derived by the abovementioned procedure (using suggested workflow), compared to LCFs derived from spatial planning maps and, in particular, Corine Land Cover that depict environmental conditions with severe spatial deviations. High density LiDAR point cloud (and derivatives) and HR-DOF enables the derivation of all LCFs needed for a large scale LSA, except springs, faults and partially lithology. It should be noted that limitations for identifying faults and some lithological contacts are caused by geological origin and characteristics of geological settings in soft rocks, i.e. the limitations might not be present in a different geological environment.

A layer depicting real environmental conditions is necessary to evaluate spatial accuracy of any input data (either landslide inventory or LCFs necessary for a LSA). Due to the conclusions listed above, high density LiDAR point cloud and HR-DOF imagery represent the most optimal input data because they enable an extensive real environmental representation of the terrain and vegetation with required spatial accuracy.

3.2 Hypothesis (2)

A representative landslide inventory map and spatially accurate landslide conditioning factor maps enables the derivation of landslide susceptibility models of high fitting and predictive performance despite the used statistical method, stable and unstable area sampling strategy or method of classification continuous landslide conditioning factor data types.

By testing different statistical methods (Sinčić et al., 2024a and Sinčić et al., 2025), stable and unstable area sampling strategies in Hrvatsko Zagorje study area (Sinčić et al., 2024a) and methods of continuous LCFs classification in smaller Podsljeme study area (Sinčić et al., 2025), the stated hypothesis can be considered as **partially confirmed**. Among five statistical methods, eight sampling strategies and 11 methods for classifying continuous LCFs tested in 86 LSMs, only sampling strategies based on a stable areas inventory leads to LSMs which can be described with low predictive performance.

Testing this hypothesis resulted in determination of certain advantages and disadvantages for LSM development considering the three tested groups of modelling settings. Namely, despite having high validation performance, the verification identified that sampling stable areas from a derived stable area inventory results in LSMs with poor predictive performance in four tested statistical methods (LR, RF, NN, SVM) (Sinčić et al., 2024a). Verification AUC for LSMs based on a mapped stable inventory range from 0.659 to 0.727, i.e. significantly lower compared to a maximum of 0.805 achieved with randomly generated stable areas using NN method (Sinčić et al., 2024a). Smoothing DTM-based LCFs and using landslide polygons proved as an optimal solution for sampling unstable areas, followed by regular sampling of landslide polygons with similar results and sampling as buffer zones around landslide polygon with moderately worse results. Testing classification methods for continuous LCFs showed minimally 0.84 AUC values for both fitting and predictive performance in all five tested methods, indicating very good results. Moreover, when using >10 classes in LCFs, the AUC values reach even higher values in both fitting (e.g. 1.0 for RF and 0.92 for NN) and predictive performance (e.g. >0.88 for LR).

Considering the statistical methods, NN and particularly RF can result in highest performance metrics while being sensitive to certain modelling settings such as sampling strategies or classification method of continuous LCFs noted above. On the other hand, LR followed by IV depict lower sensitivity to tested modelling settings and represent an excellent starting point in LSMg with very good evaluation metric results. A significant finding is that some modelling parameters, such as sampling strategies or classification method of continuous LCFs have higher influence on the model performance than the used statistical method.

It should be stated that high fitting and predictive performance do not necessarily imply a landslide susceptibility zonation map of satisfying quality, which has practical applicability to be used as a thematic map for spatial and urban planning purposes with reliable information, e.g. for land use management decisions. For instance, there are several LSMs derived in Sinčić et al. (2024a) which have high fitting and predictive performance, but the LSM flaws were only

revealed with additional verification metrics and qualitative assessment. That means the fitting and predictive performance (AUC and Cohen's kappa) should be additionally supplemented with other metrics. Moreover, more complex evaluation using metrics discussed in Chapter 3.4. will lead to a more comprehensive decision about an optimal LSM.

3.3 Hypothesis (3)

Optimisation of continuous landslide conditioning factor data type and unbiased sampling of stable and unstable areas enables an increase in the fitting and predictive performance of large scale landslide susceptibility models.

The stated hypothesis was tested in Sinčić et al. (2024a) and Sinčić et al. (2025) and can be considered as **confirmed**. From 11 defined classification scenarios for continuous LCFs, the scenarios which results in a low amount of classes were clearly identified with lower fitting and predictive performance (Sinčić et al., 2025). Namely, classification based on the studentised contrast curve (scenarios S4, S5, S6) and in particular on density (scenario S7) were characterized as unfavourable, whereas scenario S11 (defined to simulate stretched behavior maximally for line vectors and using stretched rasters without modification) showed low variability between methods, having simple technical preparation and excellent fitting and predictive results, representing the optimal scenario. Furthermore, the number of classes proved to be of higher importance than the method which defined the classes, making any scenario which results in >10 classes acceptable. Sampling strategies analyzed in Sinčić et al. (2024a) resulted in poor predictive performance (verification) for scenarios using stable areas based on the stable area inventory (scenarios S4-PM_r, S5-PM_s, S6-BbM_r, S8-CcM_r). On the contrary, the most commonly used randomly generated stable points proved as an appropriate option (scenarios S1-PR_r, S2-PR_s, S3-BR_r, S7-CR_r). This is due to high model sensitivity to stable area sampling. Namely, stable areas must be spread throughout the entire study area, without the researcher influence on the stable area location (e.g. mapping a stable areas inventory). After a thorough comparison, sampling unstable areas as centroids (scenario S7-CR_r) or buffer zones around polygon (S3-BR_r), showed as less favorable scenarios compared to regular polygon sampling (scenario S1-PR_r). On the other hand, smoothing the landslide area locations (scenario S2-PR_s) enables an increase in verification AUC values for NN and RF methods, slightly optimising the parameter. To summarize, despite the confirmation of the hypothesis, one should also take caution not to decrease the LSM fitting and predictive

performance by experimenting with the optimisation. Moreover, as noted in Sinčić et al. (2024a) and Sinčić et al. (2025), more parameters, besides only fitting and predictive performance, should be taken into account when evaluating large scale LSMs (Chapter 3.4).

3.4 Hypothesis (4)

Commonly used metrics such as AUC and Cohen's kappa do not evaluate the quality of large scale landslide susceptibility models sufficiently and could be additionally supplemented by the following metrics: (a) qualitative comparison of map close up views; (b) standard deviation maps of probabilistic susceptibility values; and (c) quantitative comparison of susceptibility zonation maps classified based on landslide area presence.

Based on analyses of 87 LSMs described in Chapters 2.2 (Sinčić et al., 2025), 2.3 (Sinčić et al., 2024a) and 2.4 (Bernat Gazibara et al., 2023a), the stated hypothesis is **confirmed**. Sinčić et al. (2024a) demonstrate that in some cases, AUC and Cohen's kappa were severely misleading in LSAs due to the sampling extents for the confusion matrix. Similar experience was noted by other researchers (e.g. Lucchese et al., 2021) and it was identified that experimenting with uncommon sampling strategies caused the misleading conclusions. To resolve the issue, additional verification AUC was implemented in analyses described in Sinčić et al. (2024a), whereas in Sinčić et al. (2024a) and Sinčić et al. (2025) the three metrics mentioned in the hypothesis provided additional new important understanding of the LSMs quality. *The qualitative comparison of map close up views* enables necessary insight into susceptibility zone distribution in relation to the location of buildings, roads and other elements at risk clearly visible on large scale maps. Moreover, the close up views are necessary for more detailed comparison and accurate distinction of terrain parts of different landslide susceptibility, e.g. identification of certain slope areas which are prone to sliding. When developing a large number of LSMs (such as in analyses described in Sinčić et al. (2024a) and Sinčić et al. (2025), the *standard deviation maps of probabilistic susceptibility values* proved as an excellent metric for measuring uncertainty, as suggested in Reichenbach et al. (2018). The SD maps enabled an effective comparison of the used statistical methods (IV, LR, RF, NN, SVM in Sinčić et al. (2025) and LR, RF, NN, SVM in Sinčić et al. (2024a)) as well as comparison of the tested parameters (classification criteria of continuous LCFs from Sinčić et al., 2025, and sampling strategies in Sinčić et al., 2024a). The SD map comparison was done simultaneously, visually and by quantification of SD values and its zone areas. Zones with high SD values indicate areas

of uncertainty caused by a tested statistical method or parameter. Hence, SD maps provide two critical information about the intensity of uncertainty and its spatial distribution. Insight into LSMs given by the SD maps could not be substituted by any other metric, nor should the new information given by the metric be neglected. Finally, the *quantitative comparison of susceptibility zonation maps classified based on landslide area presence* is an indispensable indicator of LSM quality when the map is applied in spatial or urban planning (Bernat Gazibara et al., 2023a) and it can be considered as a necessity as the zonation directly influences the decision-making processes by users who need more specific information about landslides. The susceptibility zonation based on landslide area presence brings simultaneously two new groups of information that simplify the readability and interpretation of maps: (1) landslide area in each susceptibility zone and (2) proportion of landslide area in landslide susceptibility zone area. By such comparisons, differences between LSMs are easily identified and this metric proved practical and of high importance for practical application of LSA.

3.5 Hypothesis (5)

A landslide inventory map created for a smaller but representative part of a geomorphologically and geologically homogeneous area enables the derivation of landslide susceptibility models of high predictive performance for the whole study area.

The hypothesis is exclusively addressed in Chapter 2.4 (Bernat Gazibara et al., 2023a), and can be considered as **confirmed**. Namely, the smaller Podsljeme study area (21 km²) geomorphologically and geologically represents the whole Podsljeme Area (130 km²). This was ensured by having the same LCF classes present in both the smaller and whole study areas. In total, six LCFs are used, at least one representative from geomorphological, geological, hydrological and anthropogenic groups. The landslide inventory map for the smaller area is certainly representative because it was mapped using HR-DTM derivatives, and it was used for LSM development. However, in the whole Podsljeme Area, landslide data were collected to be representative to measure predictive performance in a way that they are mapped homogeneously through the entire 130 km². Moreover, landslides in the smaller area are mapped as polygons and in the whole area as centroids. The first part of the experiment included testing the fitting and predictive performance of LSM developed only for the smaller area by splitting landslide inventory into two parts with equal amounts of landslide polygons. Fitting performance measured 0.853 AUC whereas predictive performance measured 0.869 AUC. Both metrics are

considered extremely high, especially considering the unbiased approach and a rigorous 50% to 50% approach between training and validation landslides. The second part included testing the predictive performance in the whole Podsljeme study area by applying weight factors, which the model learned in the first part of the experiment, i.e. on the 50% of training landslides of the smaller study area. An AUC value of 0.844 proved excellent model predictive performance thus confirming the hypothesis. The high resolution (5×5 pixels) of the landslide susceptibility map and its high spatial accuracy enable the practical application of the resulting map on a large scale for spatial planning purposes. It is particularly important that the proposed procedure significantly reduces the time needed to create landslide inventory maps as it reduces the extent of the area for mapping landslides and leads to high efficiency of LSAs. Regardless of high predictive performance, the LSM of a large Podsljeme study area should only be considered as preliminary. Its primary application should be identifying highly susceptible areas where detailed landslide mapping (resulting in representative polygon-based landslide inventory) is needed as input to develop a final large scale LSA. A crucial part of the methodology is selecting the size of a small representative area, as proven by Sinčić et al. (2024b) with the conclusion that the size of 20 km^2 is minimal because the predictive performance drastically decreases when using smaller areas in LSMg. On the other hand, analyses of randomly selecting as low as 10% of landslides in the smaller study area also enabled high predictive performance for the remaining 90% of landslides (Sinčić et al., 2024c). Whereas 20 km^2 is found sufficient for the Podsljeme area (130 km^2), the representative size area for any other region of interest is likely to oscillate significantly depending on the geological and geomorphological settings of the entire region, as well as its spatial distribution. Despite the high predictive performances of the models, qualitative assessment of the susceptibility zonation maps should remain imperative because described methodological changes of input landslide data are likely to influence high spatial variability in the zonation maps (Sinčić et al., 2024c). The experiment that confirmed the hypothesis was performed using the WoE method (Bernat Gazibara et al., 2023a), and it should be generally stated that the suggested procedure was not tested by machine learning algorithms such as LR, RF, NN or SVM. An important limitation of the proposed methodology is related to the type of landslides that are subject to susceptibility assessment. Namely, small and shallow landslides were mapped in the smaller Podsljeme study area which resulted in high predictive performance for small and shallow landslides which were mapped in the whole Podsljeme Area (130 km^2) (Bernat Gazibara et al., 2023a). An issue occurred with a large (approx. 1 km^2) deep-seated landslide Kostanjek whose area was poorly classified in the LSM.

3.6 Perspectives for future research

Considering the LSA research gaps, the overview of previous international research related to LSMg and landslide management status in Croatia, this doctoral research opens novel perspectives on the following topics:

(i) Extensive work done in large scale landslide inventory mapping (Bernat Gazibara 2019a; Bernat Gazibara, 2019b; Jagodnik et al., 2020; Lukačić et al., 2022; Bernat Gazibara et al., 2023c) and input data research (Sinčić et al., 2022b; Krkač et al., 2023), followed by susceptibility modelling (Sinčić et al., 2022a; Bernat Gazibara et al., 2023a; Bernat Gazibara et al., 2023b; Sinčić et al., 2024a; Sinčić et al., 2025) sets the ground for continuing the research with landslide hazard and risk modelling assessment to produce other types of landslide zoning maps (Fell et al., 2008a) necessary for landslide management. Landslide hazard and risk maps would represent a significant upgrade integrating the temporal component of threatening landslide events and consequences as a risk component, respectively.

(ii) As this doctoral research yielded a significant amount of LSMs, and conclusions that enable the selection of adequate methodology for LSAs, the next step in the derivation of large scale zonation maps is a research of zonation methods (e.g. Cantarino et al., 2019) to optimise and verify necessary parameters. The zonation part of the susceptibility mapping procedure should be largely influenced by stakeholders because the zonation in the “street-light colour” scheme should also be supplemented by certain actions of responsible services, professionals or decision-makers, leading to landslide protocols (Guzzetti et al., 2000; van Den Eeckhaut et al., 2009). The zonation and the landslide protocol are a difficult challenge because they should include relevant stakeholders (Reichenbach et al., 2018) and practical components specific to a particular system of development or management. Pioneering but extensive research of different landslide zonation maps and determining preliminary landslide protocols in the Croatian system of spatial planning and civil protection was undertaken in the framework of the PRI-MJER scientific project (URL-11). Based on numerous discussions with the relevant stakeholders, “Guidelines for Application of Landslide Maps in Croatia” (Mihalić Arbanas et al., 2023b) were published, which serve as a starting point for more detailed elaboration of landslide protocols for specific types of landslide zonation maps for particular stakeholder. The practical application of maps further requires the establishment of map quality standards. The second pioneering publication in Croatia on the topic is “Guidelines for Development of Landslide Maps in Croatia” (Bernat Gazibara et al., 2023d) published in the framework of LandSlidePlan scientific projects (Bernat Gazibara et al., 2022; URL-12). The outcomes of this

doctoral research can serve to further develop landslide mapping protocols and standards in Croatia and abroad.

(iii) Driven by experience from this doctoral research, a similar methodology could be conducted for the development of LSMs applicable on a medium scale which is used for regional application in Croatia (e.g. 1:25.000). Namely, with the exemption of the abovementioned preliminary statistically-based LSA described in Bernat Gazibara et al. (2023a), all published medium-scale landslide susceptibility maps in Croatia (Mihalić Arbanas et al., 2023b) were derived using a heuristic approach. A systematic approach to develop more reliable and statistically-based medium and/or small scale LSMs for the derivation of zonation maps for regional-level applications is necessary and would be beneficial to the policymakers and all other stakeholders.

(iv) As presented in this doctoral research, there is a lack of medium and large scale geological and lithological maps in Croatia (Sinčić et al., 2022b), which can be overcome by using small scale maps without modification (Bernat Gazibara et al., 2023a; Sinčić et al., 2025) or with modification (Sinčić et al., 2024a) using LIDAR data. It should be stated that the LSA carried out by Sinčić et al. (2024a) found EG units as either the most important LCF or in fewer cases of medium importance (equivalent to many others). The possibility to develop a LSA without geological LCFs requires further research of their influence in LSMg. This could be important to reduce some steps in LSMg, optimise the LSA procedure and possibly reduce production cost and time consumption.

(v) According to the example of a successful LSA of a larger homogenous study area based on training the LSM on a representative smaller study area (Bernat Gazibara et al., 2023a), a similar research could be conducted by using training areas which are geographically apart. For instance, it could be possible to develop a LSM in Hrvatsko Zagorje study area by training the model in Podsljeme Area or vice versa. Or alternatively, the successful LSA could be expected by combining LSMs derived from more than one training study areas. A significant issue to this application is that it is limited to similar geological and geomorphological settings because of the same landslide type occurrences and similar influence of LCFs. Satisfactory LSM development in such an experiment would lead to a more effective LSA, i.e. significant reduction of time and cost for preparing the input data. Specifically, if a representative landslide inventory map and a complete set of LCFs from Hrvatsko Zagorje or Podsljeme Area can yield a high quality LSM in another study area in Pannonian Basin, this would eliminate the need for developing a representative landslide inventory map in the new study area, as well as determining the LCFs. A second important limitation is that, regardless of the high quality

LSM, the LSA would be considered only as a preliminary step. Its main application is to provide information about areas which are more prone to sliding, i.e. narrowing down the study area size where detailed large scale investigation is truly needed. Having a reduced study area size after the preliminary LSA step enables more efficient development of needed representative polygon based landslide inventory and spatially accurate LCFs for a final large scale LSA.

(vi) As qualitative parameters proved as an essential asset in the LSM evaluation, further investigations could lead to novel and better parameters. For instance, systematically categorizing certain qualitative LSM characteristics could lead to objective quantification. Quantifying the qualitative LSM characteristics should lead to parameters describing the spatial accuracy of LSMs, currently missing in the literature due to commonly used numerical parameters (e.g. AUC). Furthermore, a strong motivation towards such research should be the fact that different conclusions about LSM quality are found from qualitative metrics used in this doctoral thesis and the commonly used numerical parameters.

4. CONCLUSION

The main result of the doctoral dissertation is a methodology for assessing landslide susceptibility at a large scale. The methodology is developed for small and shallow landslides, tested and validated in the Pannonian Basin, i.e., in study areas of the Podsljeme Area (Zagreb) and Hrvatsko Zagorje, representing urban and rural environments, respectively. The methodology was developed based on 87 landslide susceptibility models (LSMs) derived from high quality landslide and thematic input data by using three groups of statistically-based methods (i.e. bivariate, multivariate and machine learning). All results of the doctoral research are presented in four scientific papers. Methodological contributions of the dissertation, together with the conclusions, are given by three research topics: (1) spatial accuracy for large scale model; (2) systematic testing of modelling settings; (3) standard large scale approaches and recommendations.

Research topic 1: Spatial accuracy for large scale modelling

The first paper (Sinčić et al., 2022b) is exclusively dedicated to analyses of input data for large scale landslide susceptibility modelling (LSMg) by testing all available source data for Hrvatsko Zagorje study area. The paper analyses and elaborates the possibilities to derive geological, geomorphological, hydrological and anthropogenic thematic input data ending with systematization and comparison of its applicability for landslide susceptibility assessments (LSAs). The research of a complete set of landslide conditioning factors (LCFs) and the most important elements at risk (buildings and roads), yielded the conclusion that most of them can be derived using high resolution (HR) remote sensing data, i.e. high density LiDAR (Light Detection and Ranging) point cloud and HR digital orthophoto (DOF) imagery. Moreover, an alternative solution is given for small scale geological maps because the remote sensing data enables increasing spatial accuracy of geological contacts and accompanying geological units. Finally, the procedures for deriving anthropogenic LCFs are defined, together with a conclusion that high density LiDAR data and orthophoto imagery represent necessary and irreplaceable source data. All the conclusions about spatial accuracy for large scale modelling were additionally confirmed in LSMg by Sinčić et al. (2024a, 2025). In the paper by Bernat Gazibara et al. (2023a), it is approved that high quality input data collected on smaller area (specifically on 16% of the total area) are representative for development of LSM for a significantly larger area. This conclusion was derived using Weight of Evidence method applied for training the model on a smaller part of the Podsljeme Area (21 km²) followed by applying the weighted factors on the wider Podsljeme Area (130 km²).

Research topic 2: Systematic testing of modelling settings

Systematic testing of different modelling settings, developed on 86 LSMs in the doctoral research was organized as follows. The importance of evaluation parameters was highlighted in Sinčić et al. (2024a), where logistic regression (LR), random forests (RF), neural networks (NN) and support vector machine (SVM) methods were used to develop a total of 32 LSMs. In the research, stable and unstable area sampling strategies were tested in eight defined scenarios, introducing novel approaches, such as stable area inventory and smoothing landslide areas. The remaining strategies were centroids, polygons, buffering polygons for unstable areas and randomly generated points for stable areas. The final step in the analyses included a comprehensive comparison which pointed out the advantages and disadvantages of each method and sampling strategy for the case study in Hrvatsko Zagorje. The research by Sinčić et al. (2025) was done in the Podsljeme Area (20 km²) and encompassed 54 LSMs developed by either information value (IV), LR, RF, NN or SVM method, providing new insight into continuous LCF classification criteria and emphasizing the influence of using a variety of evaluation parameters. Namely, to test the classification criteria parameter for continuous LCFs, 11 scenarios were defined based on the literature review and/or modifications of the criteria, and vary from stretched values, partially stretched classes, heuristic approach, classification based on studentized contrast and landslide presence, and commonly used classification, such as Natural Neighbor, Quantiles and Geometrical intervals.

Research topic 3: Standard large scale approaches and recommendation

By emphasizing the crucial input data (Sinčić et al., 2022b), testing multiple modelling settings by deriving 86 LSMs (Sinčić et al., 2024a, Sinčić et al., 2025) and investigating cost-effectiveness practical application (Bernat Gazibara et al., 2023a), the most important findings defined by the methodology can be summarised as standard approaches and recommendations as follows:

- (i) High density LiDAR point cloud, HR-DOF imagery and publicly available data layers represent necessary and sufficient input data for the derivation of a representative landslide inventory map and spatially accurate LCFs for large scale LSMg.
- (ii) High density LiDAR point cloud and HR-DOF imagery enable verification of all input data, i.e. testing spatial accuracy by depicting real environmental conditions.
- (iii) Using a representative landslide inventory map and spatially accurate LCFs in statistically-based LSMg results in LSMs characterized by high evaluation metrics.

(iv) With adequate modelling parameter settings, NN and particularly RF yield the highest evaluation metrics, whereas LR followed by IV show lower sensitivity to parameter settings and more likely result in satisfactory (and average) results.

(v) Using stretched rasters as unclassified LCFs and line vectors with a high amount of buffer zones optimises and simplifies the continuous LCF application, whereas using a low amount of classes in either stretched rasters or line vectors is unfavourable.

(vi) Sampling unstable areas as centroids and sampling stable areas subjectively (e.g. stable areas inventory) results in severe spatial deviations of susceptibility zones. On the other hand, smoothing landslide areas can result in an increase of model evaluation metrics and randomly generated stable areas represent the good solution.

(vii) Size and distribution of susceptibility zones considering the real environmental conditions, distribution of landslides in the zones, estimation of uncertainty and/or variability on multiple susceptibility maps (e.g. standard deviation maps) and qualitative assessment considering the real environmental conditions (e.g. close up views) are metrics which enable new insights into LSMs. These new insights are often not identified by commonly used fitting and predictive metrics (e.g. Cohen's kappa and Area Under the Curve) and they are necessary for evaluating applicability of zonation maps.

(viii) A landslide inventory map created for a smaller but representative part of a geomorphologically and geologically homogeneous area, enables the derivation of preliminary LSMs of high predictive performance for the whole study area, leading to cost-effectiveness and reducing development time.

(ix) Bivariate methods (such as WoE and IV used in this doctoral dissertation) which might be characterized as outdated nowadays showed extremely well results in LSMg, and up to par with machine learning algorithms such as RF, NN or SVM.

(x) The quality of input data remains the most important (and likely most challenging) factor in LSMg, i.e. the steps in preparation of a relevant landslide inventory and spatially accurate LCFs should take most of the researcher attention.

5. LITERATURE

Agterberg, F.P., Bonham–Carter, G.F. & Wright, D.F. (1990) Statistical Pattern Integration for Mineral Exploration. In: Gaal, G. & Merriam, D.F. (eds) *Computer Applications in Resource Estimation: Prediction and Assessment for Metals and Petroleum*. Pergamon, Oxford, 1-21.

Alvioli, M., Marchesini, M., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F. & Guzzetti, F. (2016) Automatic delineation of geomorphological slope units with r.slopeunits v1.0 and their optimization for landslide susceptibility modelling. *Geoscientific Model Development*, 9, 3975–3991, <https://doi.org/10.5194/gmd-9-3975-2016>.

Aničić, B. & Juriša, M. (1984) *Basic geological map, Sheet 33–68 Rogatec, scale 1:100.000*. Belgrade, Geological Survey Zagreb.

Arbanas, Ž., Mihalić Arbanas, S., Sassa, K., Marui, H., Fukuoka, H., Krkač, M., Vivoda Prodan, M., Bernat Gazibara, S. & Domlija, P. (2017) The Croatian-Japanese SATREPS Joint Research Project on Landslides (IPL-161). In: Sassa, K., Mikoš, M. & Yin, Y. (eds) *Advancing Culture of Living with Landslides, Volume 1*. Springer, Cham, 395-409, https://doi.org/10.1007/978-3-319-59469-9_35.

Basch, O. (1983) *Basic geological map, Sheet 33-81 Ivanić – Grad, scale 1:100.000*. Belgrade, Geological Survey Zagreb.

Bernat, S., Mihalić Arbanas, S. & Krkač, M. (2014a) Landslides triggered in the continental part of Croatia by extreme precipitation in 2013. In: Lollino G., Giordan D., Crosta G., Corominas J., Azzam R., Wasowski J. & Sciarra N. (eds) *Engineering Geology for Society and Territory, Volume 2: Landslide Processes*. Springer, Heidelberg, 1599-1603.

Bernat, S., Mihalić Arbanas, S. & Krkač, M. (2014b) Inventory of precipitation triggered landslides in the winter of 2013 in Zagreb (Croatia, Europe). In: Sassa K., Canuti P. & Yin Y. (eds.) *Landslide Science for a Safer Geoenvironment, Volume 2: Methods of Landslide Studies*. Springer, Cham, 829-836.

Bernat Gazibara, S., Krkač, M., Sečanj, M., Begić, H., & Mihalić Arbanas, S. (2017a) Extreme rainfall events and landslide activation in Croatia and Bosnia and Herzegovina. In: Jemec Auflič, M., Mikoš, M. & Verbovšek, T. (eds) *Proceedings of the 3rd Regional Symposium on Landslides in the Adriatic - Balkan Region*. Geological Survey of Slovenia, Ljubljana, Slovenia.

Bernat Gazibara, S., Krkač, M., Sečanj, M. & Mihalić Arbanas, S. (2017b) Identification and mapping of shallow landslides in the City of Zagreb (Croatia) using the LiDAR-based terrain model. In: Mikoš, M., Tiwari, B., Yin, Y. & Sassa K. (eds) *Advancing Culture of Living with Landslides, Volume 2: Advances in Landslide Science*. Springer, Cham, 1093-1100, https://doi.org/10.1007/978-3-319-53498-5_124.

Bernat Gazibara, S., Mihalić Arbanas, S., Krkač, M. & Sečanj, M. (2017c) Catalog of precipitation event that triggered landslides in northwestern Croatia. In: Abolmasov, B., Marjanović, M. & Đurić, U. (eds) *Proceedings of the 2nd Regional Symposium on Landslides in the Adriatic - Balkan Region*. Belgrade, Faculty of Mining and Geology, University of Belgrade, 103-107.

Bernat Gazibara, S. (2019) Methodology for landslide mapping using high resolution digital elevation model in the Podsljeme area (City of Zagreb), Doctoral thesis, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb.

Bernat Gazibara, S., Krkač, M. & Mihalić Arbanas, S. (2019a) Verification of historical landslide inventory maps for the Podsljeme area in the City of Zagreb using LiDAR-based landslide inventory. *The Mining-Geological-Petroleum Engineering Bulletin*, 34, 1, 45–58, <https://doi.org/10.17794/rgn.2019.1.5>.

Bernat Gazibara, S., Krkač, M. & Mihalić Arbanas, S. (2019b) Landslide inventory mapping using LiDAR data in the City of Zagreb (Croatia). *Journal of Maps*, 15, 773–779, <https://doi.org/10.1080/17445647.2019.1671906>.

Bernat Gazibara, S., Mihalić Arbanas, S., Sinčić, M., Krkač, M., Lukačić, H., Jagodnik, P. & Arbanas, Ž. (2022) LandSlidePlan - Scientific research project on landslide susceptibility assessment in large scale. In: Peranić, J., Vivoda Prodan, M., Bernat Gazibara, S., Krkač, M., Mihalić Arbanas, S. & Arbanas, Ž. (eds) *Landslide Modelling & Applications: Proceedings of*

the 5th Regional Symposium on Landslides in the Adriatic-Balkan Region. Rijeka, Faculty of Civil Engineering, University of Rijeka, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, 99-106.

Bernat Gazibara, S., Sinčić, M., Krkač, M., Lukačić, H. & Mihalić Arbanas, S. (2023a) Landslide susceptibility assessment on a large scale in the Podsljeme area, City of Zagreb (Croatia). *Journal of Maps*, 19, 1, <https://doi.org/10.1080/17445647.2022.2163197>.

Bernat Gazibara, S., Sinčić, M., Rossi, M., Reichenbach, P., Krkač, M., Lukačić, H., Jagodnik, P., Šarić, G. & Mihalić Arbanas, S. (2023b) Application of LAND-SUITE for landslide susceptibility modelling using different mapping units. In: Alcántara-Ayala, I., Arbanas, Ž., Cuomo, S., Huntley, D., Konagai, K., Mihalić Arbanas, S., Mikoš, M., Sassa, K., Sassa, S., Tang, H. & Tiwari, B. (eds) *Progress in Landslide Research and Technology, Volume 2 Issue 2*. Springer, Cham, 343-354. https://doi.org/10.1007/978-3-031-44296-4_19.

Bernat Gazibara, S.; Jagodnik, P., Lukačić, H., Sinčić, M., Krkač, M., Šarić, G., Arbanas, Ž. & Mihalić Arbanas, S. (2023c) Landslide and Soil Erosion Inventory Mapping Based on High-Resolution Remote Sensing Data: A Case Study from Istria (Croatia). In: Alcántara-Ayala, I., Arbanas, Ž., Cuomo, S., Huntley, D., Konagai, K., Mihalić Arbanas, S., Mikoš, M., Sassa, K., Sassa, S., Tang, H. & Tiwari, B. (eds) *Progress in Landslide Research and Technology Volume 2 Issue 1*. Springer, Cham, 363-375. https://doi.org/10.1007/978-3-031-39012-8_18.

Bernat Gazibara, S., Sinčić, M., Krkač, M., Jagodnik, P. & Mihalić Arbanas, S. (2023d) *Guidelines for Development of Landslide Maps in Croatia*. Zagreb, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Faculty of Civil Engineering, University of Rijeka. Available at: <https://landslideplan.eu/wp-content/uploads/2023/12/RGF-SMJERNICE-WEB-BIG.pdf>. Accessed: 15.1.2025.

Blue Marble Geographics (2020) *Global Mapper v22.0.0*. Available at: <https://www.bluemarblegeo.com/global-mapper/>. Accessed: 15.1.2025.

Bognar, A. (1999) Geomorphological regionalisation of Croatia. *Acta Geographica Croatia*, 34,1, 7-26. Available at: <https://hrcak.srce.hr/84572>. Accessed: 15.1.2025.

Bonham-Carter, G.F. (1994) *Geographic Information Systems for Geoscientists: Modelling with GIS*. Ottawa, Pergamon.

Bonham-Carter, G.F., Agterberg, F.P. & Wright, D.F. (1989) Weights of evidence modeling: a new approach to mapping mineral potential. In: Agterberg, F.P. & Bonham-Carter, G.F. (eds) *Statistical Applications in the Earth Sciences*. Ottawa, Geological Survey of Canada, 89-9, 171-183.

Bornaetxea, T., Rossi, M., Marchesini, I. & Alvioli, M. (2018) Effective surveyed area and its role in statistical landslide susceptibility assessments. *Natural Hazards and Earth System Sciences*, 18, 2455-2469, <https://doi.org/10.5194/nhess-18-2455-2018>.

Brabb, E.E. (1984) Innovative approaches to landslide hazard mapping. In: International Symposium on Landslides, International Society of Soil Mechanics and Foundation Engineering Sub-Committee on Landslides, International Association of Dredging Companies, National Research Council Canada & Canadian Geotechnical Society (eds) *Proceedings of 4th International Symposium on Landslides*. Toronto, Canadian Geotechnical Society, 307-324.

Breiman, L. (2001) Random Forests. *Machine Learning*, 45, 5-32, <https://doi.org/10.1023/A:1010933404324>.

Cantarino, I., Carrion, M.A., Goerlich, F. & Martínez-Ibáñez, V. (2019) A ROC analysis-based classification method for landslide susceptibility maps. *Landslides*, 16, 265-282, <https://doi.org/10.1007/s10346-018-1063-4>.

Catani, F., Lagomarsino, D., Segoni, S. & Tofani, V. (2013) Landslide susceptibility estimation by random forests technique: sensitivity and scaling issues. *Natural Hazards and Earth System Sciences*, 13, 2815-2831, <https://doi.org/10.5194/nhess-13-2815-2013>.

Chen, W., Xie, X., Wang, J., Pradhan, B., Hong, H., Bui, D.T., Duan, Z. & Ma, J. (2017) A comparative study of logistic model tree, random forest, and classification and regression tree models for spatial prediction of landslide susceptibility. *Catena*, 151, 147-160, <https://doi.org/10.1016/j.catena.2016.11.032>.

Chigira, M., Duan, F., Yagi, H. & Furuya, T. (2004) Using an Airborne Laser Scanner for the Identification of Shallow Landslides and Susceptibility Assessment in an Area of Ignimbrite Overlain by Permeable Pyroclastics. *Landslides*, 1, 203–209, <https://doi.org/10.1007/s10346-004-0029-x>.

Chung, C.-J.F. & Fabbri, A.G. (1999) Probabilistic Prediction Models for Landslide Hazard Mapping. *Photogrammetric Engineering Remote Sensing*, 65, 1389–1399. Available at: https://www.researchgate.net/publication/209802926_Probabilistic_prediction_models_for_landslide_hazard_mapping. Accessed: 15.1.2025.

Chung, C.-J.F. & Fabbri, A.G. (2003) Validation of Spatial Prediction Models for Landslide Hazard Mapping. *Natural Hazards*, 30, 451–472, <https://doi.org/10.1023/B:NHAZ.00000007172.62651.2b>.

City of Zagreb (2011) *Land use map of City of Zagreb*. Available at: <https://registri.nipp.hr/izvori/view.php?id=68>. Accessed: 15.1.2025.

CloudCompare (2022) *CloudCompare version 2.12.alpha [GPL software]*. Available at: <http://www.cloudcompare.org/>. Accessed: 15.1.2025.

Corel Corporation (2020) *CorelDRAW® Graphics Suite 2020 – version 22*. Available at: <https://www.coreldraw.com/en/>. Accessed: 15.1.2025.

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, 209–263, <https://doi.org/10.1007/s10064-013-0538-8>.

Cortes, C. & Vapnik, V. (1995) Support-vector networks. *Machine Learning*, 20, 273–297, <https://doi.org/10.1007/BF00994018>.

Cox, D.R. (1958) The Regression Analysis of Binary Sequences. *Journal of the Royal Statistical Society*, 20, 215–242, <https://doi.org/10.1111/j.2517-6161.1958.tb00292.x>.

Cruden, D.M. (1991) A simple definition of a landslide. *Bulletin of the International Association of Engineering Geology*, 43, 27-29, <https://doi.org/10.1007/BF02590167>.

Das, S., Sarkar, D. & Kanungo, D.P. (2023) A critical review on landslide susceptibility zonation: recent trends, techniques, and practices in Indian Himalaya. *Natural Hazards*, 115, 23-72, <https://doi.org/10.1007/s11069-022-05554-x>.

Dias, H.C., Hölbling, D. & Grohmann, C.H. (2021) Landslide Susceptibility Mapping in Brazil: A Review. *Geosciences*, 11, 425, <https://doi.org/10.3390/geosciences11100425>.

Donati, L. & Turrini, M.C. (2002) An objective method to rank the importance of the factors predisposing to landslides with the GIS methodology: application to an area of the Apennines (Valnerina; Perugia, Italy). *Engineering Geology*, 63, 277–289, [https://doi.org/10.1016/S0013-7952\(01\)00087-4](https://doi.org/10.1016/S0013-7952(01)00087-4).

Dou, J., Yunus, A.P., Merghadi, A., Shirzadi, A., Nguyen, H., Hussain, Y., Avtar, R., Chen, Y., Pham, B.T. & Yamagishi, H. (2020) Different sampling strategies for predicting landslide susceptibilities are deemed less consequential with deep learning. *Science of The Total Environment*, 720, 137320, <https://doi.org/10.1016/j.scitotenv.2020.137320>.

ESRI, (2020a) *ArcGIS Desktop Release 10.8.1*. Environmental Systems Research Institute, Redlands, CA. Available at: <https://www.esri.com/en-us/arcgis/products/arcgis-desktop/resources>. Accessed: 15.1.2025.

ESRI, (2020b) *ArcGIS PRO 2.7-3.2.*, Environmental System Research Institute, Redlands, CA. Available at: <https://www.esri.com/en-us/arcgis/products/arcgis-pro/overview>. Accessed: 15.1.2025.

Farooq, S. & Akram, M.S. (2021) Landslide susceptibility mapping using information value method in Jhelum Valley of the Himalayas. *Arabian Journal of Geosciences*, 14, 824, <https://doi.org/10.1007/s12517-021-07147-7>.

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

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E. & Savage, W.Z. (on behalf of the JTC-1 Joint Technical Committee on Landslides and Engineering Slopes) (2008a). Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. *Engineering Geology*, 102, 85-98, <https://doi.org/10.1016/j.enggeo.2008.03.022>.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E. & Savage, W.Z. (2008b) Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. *Engineering Geology*, 102, 99–111, <https://doi.org/10.1016/j.enggeo.2008.03.014>.

Fu, Z., Wang, F., Dou, J., Nam, K. & Ma, H. (2023) Enhanced Absence Sampling Technique for Data-Driven Landslide Susceptibility Mapping: A Case Study in Songyang County, China. *Remote Sensing*, 15, 3345, <https://doi.org/10.3390/rs15133345>.

Gaidzik, K. & Ramírez-Herrera, M.T. (2021) The importance of input data on landslide susceptibility mapping. *Scientific Reports*, 11, 19334, <https://doi.org/10.1038/s41598-021-98830-y>.

Golden Software (2018) *Grapher* ® 13.2.734 (64-bit). Available at: <https://www.goldensoftware.com/products/grapher/>. Accessed: 15.1.2025.

Guzzetti, F., Carrara, A., Cardinali, M. & Reichenbach, P. (1999) Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. *Geomorphology*, 31, 181–216, [https://doi.org/10.1016/S0169-555X\(99\)00078-1](https://doi.org/10.1016/S0169-555X(99)00078-1).

Guzzetti, F., Cardinali, M., Reichenbach, P. & Carrara, A. (2000) Comparing landslide maps: a case study in the upper Tiber River Basin, Central Italy. *Environmental Management*, 25, 247-263, <https://doi.org/10.1007/s002679910020>.

Guzzetti, F., Galli, M., Reichenbach, P., Ardizzone, F. & Cardinali, M. (2006a) Landslide hazard assessment in the Collazzone area, Umbria, Central Italy. *Natural Hazards and Earth System Sciences*, 6, 115–131, <https://doi.org/10.5194/nhess-6-115-2006>.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M. & Galli, M. (2006b) Estimating the quality of landslide susceptibility models. *Geomorphology*, 81, 166–184, <https://doi.org/10.1016/j.geomorph.2006.04.007>.

Guzzetti, F., Mondini, A.C., Cardinali, M., Federica, F., Santangelo, M. & Chang, K.-T. (2012) Landslide inventory maps: New tools for an old problem. *Earth-Science Reviews*, 112, 1-2, 42-66, <https://doi.org/10.1016/j.earscirev.2012.02.001>.

Hemasinghe, H., Rangali, R.S.S., Deshapriya, N.L. & Samarakoon, L. (2018) Landslide susceptibility mapping using logistic regression model (a case study in Badulla District, Sri Lanka). *Procedia Engineering*, 212, 1046-1053, <https://doi.org/10.1016/j.proeng.2018.01.135>

Ho, T.K. (1995) Random decision forests. In: M., Kavanaugh & P., Storms (eds) *Proceedings of 3rd International Conference on Document Analysis and Recognition*. Montreal, IEEE Computer Society Press, 278–282.

Hong, H., Miao, Y., Liu, J. & Zhu, A.-X. (2019) Exploring the effects of the design and quantity of absence data on the performance of random forest-based landslide susceptibility mapping. *Catena*, 176, 45-64, <https://doi.org/10.1016/j.catena.2018.12.035>.

Huang, F., Tao, S., Li, D., Lian, Z., Catani, F., Huang, J., Li, K. & Zhang, C. (2022) Landslide Susceptibility Prediction Considering Neighborhood Characteristics of Landslide Spatial Datasets and Hydrological Slope Units Using Remote Sensing and GIS Technologies. *Remote Sensing*, 14, 4436, <https://doi.org/10.3390/rs14184436>.

Hussin, H. Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C. & Bălteanu, D. (2016) Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. *Geomorphology*, 253, 508-523, <https://doi.org/10.1016/j.geomorph.2015.10.030>.

Jaboyedoff, M., Oppikofer, T., Abellán, A., Derron, M.-H., Loye, A., Metzger, R. & Pedrazzini, A. (2012) Use of LIDAR in landslide investigations: a review. *Natural Hazards*, 61, 5–28, <https://doi.org/10.1007/s11069-010-9634-2>.

Jacobs, L., Kervyn, M., Reichenbach, P., Rossi, M., Marchesini, I., Alvioli, M. & Dewitte, O. (2020) Regional susceptibility assessments with heterogeneous landslide information: Slope unit- vs. pixel-based approach. *Geomorphology*, 356, 107084, <https://doi.org/10.1016/j.geomorph.2020.107084>.

Jagodnik, P., Jagodnik, V., Arbanas, Ž. & Mihalić Arbanas, S. (2020) Landslide types in the Slani Potok gully, Croatia, *Geologia Croatica*, 73, 1, 13-28, <https://doi.org/10.4154/gc.2020.04>.

Jebur, M.N., Pradhan, B. & Tehrany, M.S. (2014) Optimization of landslide conditioning factors using very high-resolution airborne laser scanning (LiDAR) data at catchment scale. *Remote Sensing of Environment*, 152, 150–165, <https://doi.org/10.1016/j.rse.2014.05.013>.

Jolliffe, I. T. & Stephenson, D. B. (eds) (2003) *Forecast verification, A Practitioner's Guide in Atmospheric Science*. Chichester, John Wiley.

Kavzoglu, T., Sahin, E.K. & Colkesen, I. (2014) Landslide susceptibility mapping using GIS-based multi-criteria decision analysis, support vector machines, and logistic regression. *Landslides*, 11, 425–439, <https://doi.org/10.1007/s10346-013-0391-7>.

Krkač M., Špoljarić D., Bernat S. & Mihalić Arbanas S. (2017) Method for prediction of landslide movements based on random forests. *Landslides*, 14, 3, 947–960, <https://doi.org/10.1007/s10346-016-0761-z>.

Krkač, M., Bernat Gazibara, S., Sećanj, M., Sinčić, M. & Mihalić Arbanas, S. (2021) Kinematic model of the slow-moving Kostanjek landslide in Zagreb, Croatia. *The Mining-Geological-Petroleum Engineering Bulletin*, 36, 2, 59-68, <https://doi.org/10.17794/rgn.2021.2.6>.

Krkač, M., Bernat Gazibara, S., Sinčić, M., Lukačić, H., Šarić, G. & Mihalić Arbanas, S. (2023) Impact of input data on the quality of the landslide susceptibility large-scale maps: A case study from NW Croatia. In: Alcántara-Ayala, I., Arbanas, Ž., Cuomo, S., Huntley, D., Konagai K, Mihalić Arbanas, S., Mikoš, M., Sassa, K., Tang, H. & Tiwari, B. (eds) *Progress in Landslide Research and Technology Volume 2 Issue 1*. Springer, Cham, 135-146. https://doi.org/10.1007/978-3-031-39012-8_4.

Lee, S. (2007) Landslide susceptibility mapping using an artificial neural network in the Gangneung area, Korea. *International Journal of Remote Sensing*, 28, 4763–4783, <https://doi.org/10.1080/01431160701264227>.

Lee, S. (2019) Current and Future Status of GIS-based Landslide Susceptibility Mapping: A Literature Review. *Korean Journal of Remote Sensing*, 35, 1, 179-193, <https://doi.org/10.7780/kjrs.2019.35.1.12>.

Lukačić, H., Bernat Gazibara, S., Sinčić, M., Krkač, M., Arbanas, Ž., Jagodnik, P., Damjanović, V. & Mihalić Arbanas, S. (2022) Influence of expert knowledge on completeness and accuracy of landslide inventory maps – Example from Istria, Croatia. In: Peranić, J., Vivoda Prodan, M., Bernat Gazibara, S., Krkač, M., Mihalić Arbanas, S. & Arbanas, Ž. (eds) *Landslide Modelling & Applications: Proceedings of the 5th Regional Symposium on Landslides in the Adriatic-Balkan Region*. Rijeka, Faculty of Civil Engineering, University of Rijeka, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, 87-92.

Malamud, B.D., Turcotte, D.L., Guzzetti, F. & Reichenbach, P. (2004) Landslide inventories and their statistical properties. *Earth Surface Processes and Landforms*, 29, 6, 687–711, <https://doi.org/10.1002/esp.1064>.

Mathew, J., Jha, V.K. & Rawat, G.S. (2007) Weights of evidence modelling for landslide hazard zonation mapping in part of Bhagirathi valley, Uttarakhand. *Current Science*, 92, 5, 628-638.

Merghadi, A., Yunus, A.P., Dou, J., Whiteley, J., Pham, B.T., Bui, D.T., Avtar, R. & Abderrahmane, B. (2020) Machine learning methods for landslide susceptibility studies: A comparative overview of algorithm performance. *Earth-Science Reviews*, 207, 103225, <https://doi.org/10.1016/j.earscirev.2020.103225>.

Microsoft Corporation (2018) *Microsoft Excel*. Available at: <https://www.office.com/>. Accessed: 15.1.2025.

Mihalić, S. & Arbanas, Ž. (2012) The Croatian-Japanese Research Project on Landslides: Activities and Public Benefits. In: Sassa, K., Rouhban, B., Briceño, S., McSaveney, M. & He,

B. (eds) *Landslides: Global Risk Preparedness*. Springer, Berlin, Heidelberg, 333-349, https://doi.org/10.1007/978-3-642-22087-6_24.

Mihalić Arbanas, S., Arbanas, Ž., Bernat Gazibara, S., Ljubičić, G., Krkač, M. & Jagodnik, P. (2023b) *Guidelines for Application of Landslide Maps in Croatia*. Zagreb, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, Faculty of Civil Engineering, University of Rijeka. Available at: <https://pri-mjer.hr/pilot-podrucje/smjernice-za-ublazavanje-posljedica-od-klizanja-za-primjenu-na-regionalnoj-i-lokalnoj-razini/>. Accessed: 15.1.2025.

Mihalić Arbanas, S., Bernat Gazibara, S., Krkač, M., Sinčić, M., Lukačić, H., Jagodnik, P. & Arbanas, Ž. (2023a) Landslide Detection and Spatial Prediction: Application of Data and Information from Landslide Maps. In: Alcantara-Ayala, I., Arbanas, Ž., Huntley, D., Konagai, K., Mikoš, M., Sassa, K., Sassa, S., Tang, H. & Tiwari, B. (eds) *Progress in Landslide Research and Technology, Volume 1 Issue 2*. Springer, Cham, 195-212, https://doi.org/10.1007/978-3-031-18471-0_16.

Mihalić Arbanas, S., Krkač, M. & Bernat, S. (2016) Application of advanced technologies in landslide research in the area of the City of Zagreb (Croatia, Europe), *Geologia Croatica*, 69, 2, 231-243, <https://doi.org/10.4154/gc.2016.18>.

Mihalić Arbanas, S., Sečanj, M., Bernat Gazibara, S., Krkač, M., Begić, H., Džindo, A., Zekan, S. & Arbanas, Ž. (2017) Landslides in the Dinarides and Pannonian Basin—from the largest historical and recent landslides in Croatia to catastrophic landslides caused by Cyclone Tamara (2014) in Bosnia and Herzegovina. *Landslides*, 14, 1861–1876, <https://doi.org/10.1007/s10346-017-0880-1>.

Mihalić, S., Oštrić, M. & Vujnović, T. (2008) Landslide susceptibility mapping in the Starca Basin (Croatia). In: *Proceedings of the 2nd European conference of International Association of Engineering Geology, Euroengeo 2008. Cities and their underground environment*. Madrid, 15-19 Sept 2008, 1-7.

Miklin, Ž., Mlinar, Ž., Brkić, Ž., Hečimović, I. & Dolić M (2007) *Detailed engineering geological map of Podsljeme urbanised zone at a scale of 1:5 000 (DIGK-Phase I), Books 1-4, 44 supplements*. Zagreb, Croatian Geological Survey.

Nefeslioglu, H.A., Gokceoglu, C. & Sonmez, H. (2008) An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. *Engineering Geology*, 97, 171–191, <https://doi.org/10.1016/j.enggeo.2008.01.004>.

Neuhäuser, B., Damm, B. & Terhorst, B. (2012) GIS-based assessment of landslide susceptibility on the base of the Weights-of-Evidence model. *Landslides*, 9, 511–528, <https://doi.org/10.1007/s10346-011-0305-5>.

Pascale, S., Parisi, S., Mancini, A., Schiattarella, M., Conforti, M., Sole, A., Murgante, B. & Sdao, F. (2013) Landslide Susceptibility Mapping Using Artificial Neural Network in the Urban Area of Senise and San Costantino Albanese (Basilicata, Southern Italy). In: Murgante, B., Misra, S., Carlini, M., Torre, C.M., Nguyen, H.-Q., Taniar, D., Opduhan, B.O. & Gervasi, O. (eds) *Computational Science and its Applications – ICCSA 2013*. Springer, Berlin, Heidelberg, 473–488, https://doi.org/10.1007/978-3-642-39649-6_34.

Petschko, H., Bell, R. & Glade, T. (2016) Effectiveness of visually analyzing LiDAR DMT derivatives for earth and debris slide inventory mapping for statistical susceptibility modeling. *Landslides*, 13, 5, 857–872, <https://doi.org/10.1007/s10346-015-0622-1>.

Posit Software, PBC (2022) *RStudio 2022.07.2+576*. Available at: <https://posit.co/download/rstudio-desktop/>. Accessed: 15.1.2025.

Pourghasemi, H. R. & Rahmati, O. (2018) Prediction of the landslide susceptibility: Which algorithm, which precision?. *Catena*, 162, 177–192, <https://doi.org/10.1016/j.catena.2017.11.022>.

Pradhan, B. (2013) A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. *Computers & Geosciences*, 51, 350–365, <https://doi.org/10.1016/j.cageo.2012.08.023>.

QGIS (2020) *QGIS Geographic Information System 3.14 – 3.8.3*. Available at: <https://www.qgis.org/en/site/>. Accessed: 15.1.2025.

R Core Team (2023) *R: A Language and Environment for Statistical Computing*. Available at: <https://www.R-project.org/>. Accessed: 15.1.2025.

Ray, R.L., Lazzari, M. & Olutimehin, T. (2020) Remote Sensing Approaches and Related Techniques to Map and Study Landslides. In: Ray, R. & Lazzari, M. (eds) *Landslides- Investigation and Monitoring*. London, IntechOpen, <https://doi.org/10.5772/intechopen.93681>.

Razak, K.A., Straatsma, M.W., van Westen, C.J., Malet, J.-P. & de Jong, S.M. (2011) Airborne Laser Scanning of Forested Landslides Characterization: Terrain Model Quality and Visualization. *Geomorphology*, 126, 186–200, <https://doi.org/10.1016/j.geomorph.2010.11.003>.

Razak, K.A., Santangelo, M., van Westen, C.J., Straatsma, M.W. & de Jong, S.M. (2013) Generating an Optimal DTM from Airborne Laser Scanning Data for Landslide Mapping in a Tropical Forest Environment. *Geomorphology*, 190, 112–125, <https://doi.org/10.1016/j.geomorph.2013.02.021>.

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>.

Ripley, B. D. (1996) *Pattern Recognition and Neural Networks*. Cambridge, Cambridge University Press.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A.C. & Peruccacci, S. (2010) Optimal landslide susceptibility zonation based on multiple forecasts. *Geomorphology*, 114, 129–142, <https://doi.org/10.1016/j.geomorph.2009.06.020>.

Rossi, M., Bornaetxea, T. & Reichenbach, P. (2022) LAND–SUITE V1.0: a suite of tools for statistically based landslide susceptibility zonation. *Geoscientific Model Development*, 15, 5651–5666, <https://doi.org/10.5194/gmd-15-5651-2022>.

Sandić, C., Marjanović, M., Abolmasov, B. & Tošić, R. (2023) Integrating landslide magnitude in the susceptibility assessment of the City of Doboj, using machine learning and heuristic approach. *Journal of Maps*, 19,1, <https://doi.org/10.1080/17445647.2022.2163199>.

Sarkar, S., Roy, A.K. & Martha, T.R. (2013) Landslide susceptibility assessment using Information Value Method in parts of the Darjeeling Himalayas. *Journal of the Geological Society of India*, 82, 351–362, <https://doi.org/10.1007/s12594-013-0162-z>.

Scaioni, M., Longoni, L., Melillo, V. & Papini, M. (2014) Remote Sensing for Landslide Investigations: An Overview of Recent Achievements and Perspectives. *Remote Sensing*, 6, 10, 9600-9652, <https://doi.org/10.3390/rs6109600>.

Shano, L., Raghuvanshi, T.K. & Meten, M. (2020) Landslide susceptibility evaluation and hazard zonation techniques – a review. *Geoenvironmental Disasters*, 7, 18, <https://doi.org/10.1186/s40677-020-00152-0>.

Süzen, M.L. & Doyuran, V (2004) Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. *Engineering Geology*, 71, 303-321, [https://doi.org/10.1016/S0013-7952\(03\)00143-1](https://doi.org/10.1016/S0013-7952(03)00143-1).

Sinčić, M., Bernat Gazibara, S., Krkač, M. & Mihalić Arbanas, S. (2022a) Landslide susceptibility assessment of the City of Karlovac using the bivariate statistical analysis. *The Mining-Geology-Petroleum Engineering Bulletin*, 37, 2, 149–170, <https://doi.org/10.17794/rgn.2022.2.13>.

Sinčić, M., Bernat Gazibara, S., Krkač, M., Lukačić, H. & Mihalić Arbanas, S. (2022b) The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments. *Land*, 11, 1360, <https://doi.org/10.3390/land11081360>.

Sinčić, M., Bernat Gazibara, S., Lukačić, H., Krkač, M. & Mihalić Arbanas, S. (2022c) Qualitative and quantitative assessments of input LiDAR data for landslide inventory mapping. In: Peranić, J., Vivoda Prodan, M., Bernat Gazibara, S., Krkač, M., Mihalić Arbanas, S. & Arbanas, Ž. (eds) *Book of Abstracts of the 5th ReSyLAB Landslide Modelling & Applications*.

Rijeka, Faculty of Civil Engineering, University of Rijeka, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb, 22-22.

Sinčić, M., Bernat Gazibara, S., Rossi, M., Krkač, M. & Mihalić Arbanas, S. (2024a). A Comprehensive Comparison of Stable and Unstable Area Sampling Strategies in Large-Scale Landslide Susceptibility Models Using Machine Learning Methods. *Remote Sensing*, 16, 16, 2923, <https://doi.org/10.3390/rs16162923>.

Sinčić, M., Bakula, L., Bernat Gazibara, S. & Mihalić Arbanas, S. (2024b) Influence of spatial landslide data sampling on large-scale landslide susceptibility modelling. In: Posarić, L., Gmižić, D., Ostojić, T. & Purgar Filjak, M. (eds) *8th Faculty of Science PhD Student Symposium: Book of Abstracts*. Zagreb, Faculty of Science, University of Zagreb, 109-109.

Sinčić, M., Bernat Gazibara, S., Rossi, M., Krkač, M., Lukačić, H. & Mihalić Arbanas, S. (2024c) Influence of the landslide inventory sampling on the accuracy of the susceptibility modelling using Random Forests: A case study from the NW Croatia. In: Marjanović, M. & Đurić, U. (eds) *Proceedings of the 6th Regional Symposium on Landslides In the Adriatic-Balkan Region*. Belgrade, Faculty of Mining and Geology, University of Belgrade, 143-148, <https://doi.org/10.18485/resylab.2024.6.ch19>.

Sinčić, M., Bernat Gazibara, S., Rossi, M. & Mihalić Arbanas, S. (2025) Comparison of conditioning factor classification criteria in large-scale statistically based landslide susceptibility models. *Natural Hazards and Earth System Sciences*, 25, 183-206, <https://doi.org/10.5194/nhess-25-183-2025>.

Soeters, R. & van Westen, C.J. (1996) Slope Instability Recognition Analysis and Zonation. In: Turner, K.T. & Schuster, R.L. (eds) *Landslides: Investigation and Mitigation*. Washington DC, Transportation Research Board National Research Council, 129–177.

Golden Software (2020) Surfer ® 19.1.189 (64-bit). Available at: <https://www.goldensoftware.com/products/>. Accessed: 15.1.2025.

Šikić, K., Basch, O. & Šimunić, A. (1972) *Basic geological map, Sheet 33-80 Zagreb, scale 1:100.000*. Belgrade, Geological Survey Zagreb.

Šimunić, A., Pikija, M. & Hečimović, I. (1982) *Basic geological map, Sheet 33-69 Varaždin, scale 1:100.000*. Belgrade, Geological Survey Zagreb.

Tarolli, P. (2014) High-Resolution Topography for Understanding Earth Surface Processes: Opportunities and Challenges. *Geomorphology*, 216, 295–312, <https://doi.org/10.1016/j.geomorph.2014.03.008>.

The MathWorks Inc (2021a) *Statistics and Machine Learning Toolbox 12.1 (R2021)*. Available at: <https://www.mathworks.com/products/matlab.html>. Accessed: 15.1.2025.

The MathWorks Inc (2021b) *MATLAB version 9.10.0.1602886 (R2021b)*. Available at: <https://www.mathworks.com/products/matlab.html>. Accessed: 15.1.2025.

Tien Bui, D., Tuan, T.A., Klempe, H., Pradhan, B. & Revhaug, I. (2016) Spatial prediction models for shallow landslide hazards: a comparative assessment of the efficacy of support vector machines, artificial neural networks, kernel logistic regression, and logistic model tree. *Landslides*, 13, 361–378, <https://doi.org/10.1007/s10346-015-0557-6>.

URL-1. Available at: <http://Geoportal.Dgu.Hr/Wms?Layers=TK25>. Accessed: 15.6.2022.

URL-2 Available at: <http://Geoportal.Dgu.Hr/Wms?Layers=DOF>. Accessed: 15.6.2022.

URL-3. Available at: <http://Download.Geofabrik.de/Europe.Html>. Accessed: 15.6.2022.

URL-4 Available at: http://Arhiva.Vzz.Hr/Images/Stories/Prostorni-Plan/LEPOGLAVA_ID/LEPOGLAVA_ID2/K1_KORISTENJE_I_NAMJENA_POVRSINA_HTRS.Pdf. Accessed: 15.6.2022.

URL-5 Available at: http://Arhiva.Vzz.Hr/Images/Stories/Prostorni-Plan/BEDNJA/2-ID-PPUO-Bednja-05-2017/K1_KORISTENJE_I_NAMJENA_POVRSINA_ID2.Pdf. Accessed 15.6.2022.

URL-6 Available at: [https://Land.Copernicus.Eu/Imagery-in-Situ/Eu-Dem/Eu-Dem-v1.1?Tab=download\(EuDem_v11_E40N20.Tif](https://Land.Copernicus.Eu/Imagery-in-Situ/Eu-Dem/Eu-Dem-v1.1?Tab=download(EuDem_v11_E40N20.Tif). Accessed: 15.6.2022.

URL-7 Available at: <https://Scihub.Copernicus.Eu/Dhus/#/Home>. Accessed: 15.6.2022.

URL-8 Available at: <http://Geoportal.Dgu.Hr/Wms?Layers=HOK>. Accessed: 15.6.2022.

URL-9 Available at:
<https://Dgu.Gov.Hr/UserDocsImages/Dokumenti/Pristup%20informacijama/Zakoni%20i%20ostali%20propisi/Specifikacije/SP%20za%20izradu%20digitalnih%20ortofotokarata.ZIP> .
Accessed: 15.6.2022.

URL-10 Available at: <https://Land.Copernicus.Eu/Pan-European/Corine-Land-Cover/Clc2018?Tab=download> (CLC2018_CLC2012_V2018_20b2.Gdb). Accessed: 15.6.2022.

URL-11 Available at: <https://pri-mjer.hr/>. Accessed: 15.1.2025.

URL-12 Available at: <https://landslideplan.eu/en/home/>. Accessed: 15.1.2025.

van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M. & Poesen, J.A. (2009) Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium. *Natural Hazards and Earth Sciences*, 9, 2, 507–521, <http://doi.org/10.5194/nhess-9-507-2009>.

van Westen, C.J. (2002) Use of weights of evidence modeling for landslide susceptibility mapping. Enschede, ITC Publication. Available at:
https://filetransfer.itc.nl/pub/westen/GISSIZ/Chapter%206%20-%20Exercises/GIS%20exercises/Data%20driven%20models/Exercise%20data_driven%20models%202002.pdf. Accessed: 15.1.2025.

van Westen, C.J., Rengers, N. & Soeters, R. (2003) Use of Geomorphological Information in Indirect Landslide Susceptibility Assessment. *Natural Hazards*, 30, 399–419, <https://doi.org/10.1023/B:NHAZ.0000007097.42735.9e>.

van Westen, C.J., van Asch, T. & Soeters, R. (2006) Landslide hazard and risk zonation—why is it still so difficult?. *Bulletin of Engineering Geology and the Environment*, 65, 167–184, <https://doi.org/10.1007/s10064-005-0023-0>.

van Westen, C.J., Castellanos, E., Kuriakose, S.L. (2008) Spatial data for landslide susceptibility, hazard, and vulnerability assessment: An overview. *Engineering Geology*, 102, 112–131, <https://doi.org/10.1016/j.enggeo.2008.03.010>.

Vapnik, V.N. (1995) *The Nature of Statistical Learning Theory*. New York, Springer.

Varnes, D.J. & the IAEG Commission on Landslides and other Mass-Movements (1984) *Landslide hazard zonation: a review of principles and practice*. Paris, The UNESCO Press.

Venables, W. N. & Ripley, B. D. (2002) *Modern Applied Statistics with S*. Berlin, Springer.

Vujnović, T. (2005) Identification and mapping of landslide factors at the area of the Starča Basin, Master thesis, Faculty of Mining, Geology and Petroleum Engineering, University of Zagreb.

Wang, L.-J., Guo, M., Sawada, K., Lin, J. & Zhang, J. (2016) A comparative study of landslide susceptibility maps using logistic regression, frequency ratio, decision tree, weights of evidence and artificial neural network. *Geosciences Journal*, 20, 117–136, <https://doi.org/10.1007/s12303-015-0026-1>.

Wang, H., Zhang, L., Luo, H., He, J. & Cheung, R.W.M. (2021) AI-powered landslide susceptibility assessment in Hong Kong. *Engineering Geology*, 288, 106103, <https://doi.org/10.1016/j.enggeo.2021.106103>.

Yan, G., Liang, S., Gui, X., Xie, Y. & Zhao, H. (2019) Optimizing landslide susceptibility mapping in the Kongtong District, NW China: comparing the subdivision criteria of factors. *Geocarto International*, 34, 1408–1426, <https://doi.org/10.1080/10106049.2018.1499816>.

Yao, X., Tham, L.G., Dai, F.C. (2008) Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. *Geomorphology*, 101, 572–582, <https://doi.org/10.1016/j.geomorph.2008.02.011>.

Yin, K.L. & Yan, T.Z. (1988) Statistical prediction model for slope instability of metamorphosed rocks. In: A.A Bonnard (ed.) *Proceedings of Fifth International Symposium on Landslides*. Lausanne, Balkema, 1269–1272.

Youssef, A.M. & Pourghasemi, H.R. (2021) Landslide susceptibility mapping using machine learning algorithms and comparison of their performance at Abha Basin, Asir Region, Saudi Arabia. *Geoscience Frontiers*, 12, 2, 639–655, <https://doi.org/10.1016/j.gsf.2020.05.010>.

Yusof, N.M., Pradhan, B., Shafri, H.Z.M., Jebur, M.N. & Yusoff, Z. (2015) Spatial landslide hazard assessment along the Jelapang Corridor of the North-South Expressway in Malaysia using high resolution airborne LiDAR data. *Arabian Journal of Geosciences*, 8, 9789–9800, <https://doi.org/10.1007/s12517-015-1937-x>.

Zaninović, K., Gajić-Čapka, K., Perčec Tadić, M., Vučetić, M., Milković, J., Bajić, A., Cindić, K., Cvitan, L., Katušin, Z. & Kaučić, D. (2008) *Climate Atlas of Croatia 1961–1990, 1971–2000*. Zagreb, Croatian Meteorological and Hydrological Service. Available at https://klima.hr/razno/publikacije/klimatski_atlas_hrvatske.pdf. Accessed: 15.1.2025.

6. BIOGRAPHY OF THE AUTHOR

Marko Sinčić, born on October 15, 1995, in Karlovac, completed his elementary education at “PŠ Velika Jelsa” and “OŠ Dubovac,” followed by “Gimnazija Karlovac” course Natural Sciences and Mathematics which he finished in 2014. He then enrolled in the Undergraduate studies of Geological Engineering at the Faculty of Mining, Geology, and Petroleum Engineering, completing his degree in 2017. Afterward, he pursued his Graduate studies in Geological Engineering (Hydrogeology and Engineering Geology) from 2017 to 2020, defending his thesis “Landslide susceptibility assessment of the Karlovac City using bivariate statistics methodology” under the supervision of Professor Snježana Mihalić Arbanas, PhD. During this period, he completed internships in Croatia and Bosnia and Herzegovina, received the Dean’s Award for academic achievement and a Rector’s Award for individual scientific work. Since June 2020, Marko has worked at the Faculty, initially as a project administrator on the PRI-MJER project, followed by a Junior Researcher position in October 2020 for the Croatian Science Foundation's LandSlidePlan project. He also enrolled in doctoral studies in Applied Geosciences, Mining, and Petroleum Engineering under the Croatian Science Foundation's Young Researchers' career development project. During his studies, he assisted several Master’s students, authored five and coauthored 11 scientific papers. Marko attended six international scientific conferences, completed the “LARAM” summer school in Italy and has been involved in a dozen workshops and round tables as a member of the Croatian Landslide Group.

LIST OF PUBLICATIONS (15/1/2025)

Journal papers

Alvioli, Massimiliano; Loche, Marco; Jacobs, Liesbet; Grohmann, Carlos H.; Abraham, Minu Treesa; Gupta, Kunal; Satyam, Neelima; Scaringi, Gianvito; Bornaetxea, Txomin; Rossi, Mauro et al.: A benchmark dataset and workflow for landslide susceptibility zonation // *Earth-science reviews*, 258 (2024), 104927-0. doi: 10.1016/j.earscirev.2024.104927

Bernat Gazibara, Sanja; Sinčić, Marko; Krkač, Martin; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: Landslide susceptibility assessment on a large scale in the Podsljeme area, City of Zagreb (Croatia) // *Journal of Maps*, 19 (2023), 1, 11. doi: 10.1080/17445647.2022.2163197

Krkač, Martin; Bernat Gazibara, Sanja; Sečanj, Marin; Sinčić, Marko; Mihalić Arbanas, Snježana: Kinematic model of the slow-moving Kostanjek landslide in Zagreb, Croatia // *Rudarsko-geološko-naftni zbornik*, 36 (2021), 2; 59-68. doi: 10.17794/rgn.2021.2.6

Sinčić, Marko; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas, Snježana: Landslide susceptibility assessment of the City of Karlovac using the bivariate statistical analysis // *Rudarsko-geološko-naftni zbornik*, (2022), doi: 10.17794/rgn.2022.2.13

Sinčić, Marko; Bernat Gazibara, Sanja; Krkač, Martin; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: The Use of High-Resolution Remote Sensing Data in Preparation of Input Data for Large-Scale Landslide Hazard Assessments // *Land (Basel)*, 11 (2022), 8; 1360, 37. doi: 10.3390/land11081360

Sinčić, Marko; Bernat Gazibara, Sanja; Rossi, Mauro; Krkač, Martin; Mihalić Arbanas, Snježana: A Comprehensive Comparison of Stable and Unstable Area Sampling Strategies in Large-Scale Landslide Susceptibility Models Using Machine Learning Methods // *Remote sensing*, 16 (2024), 16; 1-26. doi: 10.3390/rs16162923

Sinčić, Marko; Bernat Gazibara Sanja; Rossi Mauro; Mihalić Arbanas Snježana: Comparison of conditioning factor classification criteria in large-scale statistically based landslide susceptibility models // *Natural Hazards and Earth System Sciences*, 25 (2025), 183-206. doi: <https://doi.org/10.5194/nhess-25-183-2025>

Book chapters

Bernat Gazibara, Sanja; Jagodnik, Petra; Lukačić, Hrvoje; Sinčić, Marko; Krkač, Martin; Šarić, Gabrijela; Arbanas, Željko; Mihalić Arbanas, Snježana: Landslide and Soil Erosion Inventory Mapping Based on High-Resolution Remote Sensing Data: A Case Study from Istria (Croatia) // Progress in Landslide Research and Technology Volume 2 Issue 1, 2023 / Alcántara-Ayala, Irasema; Arbanas, Željko; Cuomo, Sabatino et al. (Eds.). Switzerland: Springer Cham, 2023. p. 363-375 . doi: 10.1007/978-3-031-39012-8_18

Bernat Gazibara, Sanja; Sinčić, Marko; Rossi, Mauro; Reichenbach, Paola; Krkač, Martin; Lukačić, Hrvoje; Jagodnik, Petra; Šarić, Gabrijela; Mihalić Arbanas, Snježana: Application of LAND-SUITE for Landslide Susceptibility Modelling Using Different Mapping Units: A Case Study in Croatia // Progress in Landslide Research and Technology, Volume 2 Issue 2, 2023 / Alcántara-Ayala, Irasema; Arbanas, Željko; Huntley, David et al. (Eds.). Cham: Springer Cham, 2023. p. 343-354

Krkač, Martin; Bernat Gazibara, Sanja; Sinčić, Marko; Lukačić, Hrvoje; Šarić, Gabrijela; Mihalić Arbanas, Snježana: Impact of Input Data on the Quality of the Landslide Susceptibility Large-Scale Maps: A Case Study from NW Croatia // Progress in Landslide Research and Technology Volume 2 Issue 1, 2023 / Alcántara-Ayala, Irasema; Arbanas, Željko; Cuomo, Sabatino et al. (Eds.). Switzerland: Springer Cham, 2023. p. 135-146 . doi: 10.1007/978-3-031-39012-8_4

Mihalić Arbanas, Snježana; Bernat Gazibara, Sanja; Krkač, Martin; Sinčić, Marko; Lukačić, Hrvoje; Jagodnik, Petra; Arbanas, Željko: Landslide Detection and Spatial Prediction: Application of Data and Information from Landslide Maps // Progress in Landslide Research and Technology, Volume 1 Issue 2, 2022. Fullerton, CA, USA: Springer, 2023. p. 195-212 . doi: 10.1007/978-3-031-18471-0_16

Conference proceedings

Bernat Gazibara, Sanja; Mihalić Arbanas, Snježana; Sinčić, Marko; Krkač, Martin; Lukačić, Hrvoje; Jagodnik, Petra; Arbanas, Željko: LandSlidePlan - Scientific research project on landslide susceptibility assessment in large scale // 5th Regional Symposium on Landslides in

the Adriatic-Balkan Region, Rijeka, Croatia, 23.03.-2022.-26.03.2022. // Landslide Modelling & Applications: Proceedings of the 5th Regional Symposium on Landslides in the Adriatic-Balkan Region. / Peranić, Josip; Vivoda Prodan, Martina; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas, Snježana; Arbanas, Željko (Eds.). Rijeka: Rudarsko-geološko-naftni fakultet Sveučilišta u Zagrebu; Građevinski fakultet Sveučilišta u Rijeci, 2022. p. 99-106

Bernat Gazibara, Sanja; Sinčić, Marko; Rossi, Mauro; Krkač, Martin; Lukačić, Hrvoje; Jagodnik, Petra; Mihalić Arbanas, Snježana: Large-scale landslide susceptibility models: Examples and conclusions from the modelling of small and shallow landslides in the continental part of Croatia // 6th Regional Symposium on Landslides in the Adriatic-Balkan Region, Belgrade, Serbia, 15.05.2024.-18.05.2024. // Proceedings of the 6th Regional Symposium on LANDSLIDES In the Adriatic-Balkan Region / Marjanović, Miloš; Đurić, Uroš (Eds.). Beograd: University of Belgrade, Faculty of Mining and Geology, 2024. p. 149-155. doi: 10.18485/resylab.2024.6.ch20

Krkač, Martin; Bernat Gazibara, Sanja; Sinčić, Marko; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: Landslide inventory mapping based on LiDAR data: A case study from Hrvatsko Zagorje (Croatia) // 5th Regional Symposium on Landslides in the Adriatic-Balkan Region, Rijeka, Croatia, 23.03.-2022.-26.03.2022. // Landslide Modelling & Applications: Proceedings of the 5th Regional Symposium on Landslides in the Adriatic-Balkan Region. / Peranić, Josip; Vivoda Prodan, Martina; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas, Snježana; Arbanas, Željko (Eds.). Rijeka: Rudarsko-geološko-naftni fakultet Sveučilišta u Zagrebu; Građevinski fakultet Sveučilišta u Rijeci, 2022. p. 81-86

Lukačić, Hrvoje; Bernat Gazibara, Sanja; Sinčić, Marko; Krkač, Martin; Arbanas, Željko; Jagodnik, Petra; Damjanović, Vedran; Mihalić Arbanas, Snježana: Influence of expert knowledge on completeness and accuracy of landslide inventory maps – Example from Istria, Croatia // 5th Regional Symposium on Landslides in the Adriatic-Balkan Region, Rijeka, Croatia, 23.03.-2022.-26.03.2022 // Landslide Modelling & Applications: Proceedings of the 5th Regional Symposium on Landslides in the Adriatic-Balkan Region. / Peranić, Josip; Vivoda Prodan, Martina; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas, Snježana; Arbanas, Željko (Eds.). Rijeka: Rudarsko-geološko-naftni fakultet Sveučilišta u Zagrebu; Građevinski fakultet Sveučilišta u Rijeci, 2022. p. 87-92

Sinčić, Marko; Bernat Gazibara, Sanja; Rossi, Mauro; Krkač, Martin; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: Influence of the landslide inventory sampling on the accuracy of the susceptibility modelling using Random Forests: A case study from the NW Croatia // 6th Regional Symposium on Landslides in the Adriatic-Balkan Region, Belgrade, Serbia, 15.05.2024.-18.05.2024. // Proceedings of the 6th Regional Symposium on Landslides In the Adriatic-Balkan Region / Marjanović, Miloš; Đurić, Uroš (Eds.). Beograd: University of Belgrade, Faculty of Mining and Geology, 2024. p. 143-148. doi: 10.18485/resylab.2024.6.ch19

Conference abstracts

Bernat Gazibara, Sanja; Damjanović, Vedran; Krkač, Martin; Sinčić, Marko; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: Landslide susceptibility map of Croatia based on limited data and Fuzzy logic approach // 5th Regional Symposium on Landslides in the Adriatic-Balkan Region, Rijeka, Croatia, 23.03.-2022.-26.03.2022. // Book of Abstracts 5th ReSyLAB 'Landslide Modelling & Applications' / Peranić, Josip; Vivoda Prodan, Martina; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas, Snježana; Arbanas, Željko (Eds.). Rijeka: Rudarsko-geološko-naftni fakultet Sveučilišta u Zagrebu; Građevinski fakultet Sveučilišta u Rijeci, 2022. p. 27-28

Bernat Gazibara, Sanja; Sinčić, Marko; Krkač, Martin; Jagodnik, Petra; Arbanas, Željko; Mihalić Arbanas, Snježana: Landslide and erosion inventory mapping based on LiDAR data: A case study from Istria (Croatia) // 10th International Conference on Geomorphology, Coimbra, Portugal, 12.09.2022-16.09.2022. // doi: 10.5194/icg2022-408, 2022

Bernat Gazibara, Sanja; Sinčić, Marko; Krkač, Martin; Lukačić, Hrvoje; Šarić, Gabriijela; Jagodnik, Petra; Mihalić Arbanas, Snježana: The landslide susceptibility assessment for application in the spatial planning system: from national to local // 7. hrvatski geološki kongres, Poreč, Hrvatska, 02.10.2023.-04.10.2023. // 7. HGK - Knjiga sažetaka (Book of Abstracts) / Fio Firi, Karmen (Ed.). Zagreb: Hrvatski geološki institut, 2023. p. 22-23

Bernat Gazibara, Sanja; Sinčić, Marko; Krkač, Martin; Jagodnik, Petra; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: Influence of the landslide inventory completeness on the accuracy of the landslide susceptibility modelling: A case study from the City of Zagreb (Croatia) // 6th

World Landslide Forum, Florence, Italy, 14.11.2023.-17.11.2023 // Proceedings of the 6th World Landslide Forum- Abstract book / Veronika, Tofani; Nocola Casagli (Eds.). Florence, Italy: OIC S.r.l., Florence, Italy, 2023. p. 644-644

Bernat Gazibara, Sanja; Štefičar, Iva; Sinčić, Marko; Krkač, Martin; Mihalić Arbanas, Snježana: Large-scale landslide susceptibility zonation: classification methods and influence on a further application in spatial planning system // 4th European Regional Conference of IAEG, Dubrovnik, Croatia, 08.10.2024.-12.10.2024. // 2024

Jagodnik, Petra; Bernat Gazibara, Sanja; Sinčić, Marko; Lukačić, Hrvoje; Arbanas, Željko; Mihalić Arbanas, Snježana: Geomorphological settings and types of landslides in the City of Buzet identified using LiDAR Digital Terrain Model // 7. hrvatski geološki kongres, Poreč, Hrvatska, 02.10.2023.-04.10.2023. // 7. HGK - Knjiga sažetaka (Book of Abstracts) / Fio Firi, Karmen (Ed.). Zagreb: Hrvatski geološki institut, 2023. p. 79-79

Jagodnik, Petra; Bernat Gazibara, Sanja; Sinčić, Marko; Lukačić, Hrvoje; Šarić, Gabrijele; Krkač, Martin; Arbanas, Željko; Mihalić Arbanas, Snježana: Landslides, gully erosion and badlands as associated geological hazards in flysch environment - analysis of geomorphological inventories and LiDAR DTM at a large scale // 6th World Landslide Forum, Florence, Italy, 14.11.2023.-17.11.2023. // Proceedings of the 6th World Landslide Forum- Abstract book / Veronika, Tofani; Nocola Casagli (Eds.). Florence, Italy: OIC S.r.l., Florence, Italy, 2023. p. 112-112

Krkač, Martin; Bernat Gazibara, Sanja; Sinčić, Marko; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: Geomorphological characteristics of landslides in the Hrvatsko Zagorje (NW Croatia) // 10th International Conference on Geomorphology, Coimbra, Portugal, 12.09.2022-16.09.2022. // doi: 10.5194/icg2022-410, 2022.

Mihalić Arbanas, Snježana; Bernat Gazibara, Sanja; Krkač, Martin; Sinčić, Marko; Lukačić, Hrvoje; Damjanović, Vedran; Jagodnik, Petra; Arbanas, Željko: Landslide evidence and spatial prediction - Application of data and information from landslide maps // 5th Regional Symposium on Landslides in the Adriatic-Balkan Region, Rijeka, Croatia, 23.03.-2022.-26.03.2022. // Book of Abstracts 5th ReSyLAB 'Landslide Modelling & Applications' / Peranić, Josip; Vivoda Prodan, Martina; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas,

Snježana; Arbanas, Željko (Eds.). Rijeka: Rudarsko-geološko-naftni fakultet Sveučilišta u Zagrebu; Građevinski fakultet Sveučilišta u Rijeci, 2022. p. 7-8

Sinčić, Marko; Bernat Gazibara, Sanja; Lukačić, Hrvoje; Krkač, Martin; Mihalić Arbanas, Snježana: Qualitative and quantitative assessments of input LiDAR data for landslide inventory mapping // 5th Regional Symposium on Landslides in the Adriatic-Balkan Region, Rijeka, Croatia, 23.03.-2022.-26.03.2022. // Book of Abstracts 5th ReSyLAB 'Landslide Modelling & Applications' / Peranić, Josip; Vivoda Prodan, Martina; Bernat Gazibara, Sanja; Krkač, Martin; Mihalić Arbanas, Snježana; Arbanas, Željko (Eds.). Rijeka: Rudarsko-geološko-naftni fakultet Sveučilišta u Zagrebu; Građevinski fakultet Sveučilišta u Rijeci, 2022. p. 22-22

Sinčić, Marko; Bernat Gazibara, Sanja; Krkač, Martin; Lukačić, Hrvoje; Mihalić Arbanas, Snježana: A slope units based landslide susceptibility analyses using Weight of Evidence and Random Forest // EGU General Assembly 2023, Vienna, Austria, 24.04.2023.-28.04.2023. // doi: 10.5194/egusphere-egu23-5755, 2023

Sinčić, Marko; Bernat Gazibara, Sanja; Šarić, Gabrijela; Krkač, Martin; Mihalić Arbanas, Snježana: Influence of landslide conditioning factor selection on landslide susceptibility modelling in large scale // 6th World Landslide Forum, Florence, Italy, 14.11.2023.-17.11.2023. // Proceedings of the 6th World Landslide Forum- Abstract book / Veronika, Tofani; Nicola Casagli (Eds.). Florence, Italy: OIC S.r.l., Florence, Italy, 2023. p. 651-651

Sinčić, Marko; Bakula, Lucija; Bernat Gazibara, Sanja; Mihalić Arbanas, Snježana: Influence of spatial landslide data sampling on large-scale landslide susceptibility modelling // 8th Faculty of Science PhD Student Symposium, Zagreb, Croatia, 26.04.2024.-27.04.2024. // 8th Faculty of Science PhD Student Symposium: Book of Abstracts / Posarić, Laura; Gmižić, Daria; Ostojić, Tea et al. (Eds.). Zagreb: Faculty of Science, University of Zagreb, Zagreb, Croatia, 2024. p. 109-109