METHODOLOGICAL APPROACH FOR LANDSLIDE ANALYSIS

IN A REGIONAL SCALE

Data collection, susceptibility models and precipitation thresholds. Application in Gipuzkoa Province (Basque Country).

Lur labainketen analisirako hurbilketa metodologikoa eskala erregionalean

Datuen bilketa, suszeptibilitate modeloak eta euri prezipitazioen atalaseak. Gipuzkoako Lurralde Historikoan aplikatua (Euskal Herria).

UPV/EHU

eman ta zabal zazu
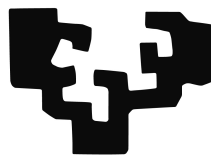
Universidad del País Vasco   Euskal Herriko Unibertsitatea

Txomin Bornaetxea Estela

2018

Tesiaren zuzendariak:

Iñaki Antiguedad Auzmendi

Orbange Ormaetxea Arenaza

Tesiaren epaimahaieko partaideak:

"If a landslide is, simply, a slide of land,

a cowboy is a male calf"

D.M. Cruden

Abstract

Worldwide, landslides happen every year and they cause undesired effects to the human being and its goods. The delimitation of the territory by the likelihood to experience landslides, by means of susceptibility maps, supposes the primary step for hazard and risk assessments in order to mitigate the damages caused by these geomorphological processes.

In the last 50 years, we passed from making landslide susceptibility maps based on the pure observation of the terrain to fully automated and sophisticated statistical procedures. But, this rapid increase of approaches resulted in a huge variety of different methodologies available and thus it complicates the definition of standards for the landslides susceptibility analysis. The current thesis presents a roadmap definition, starting from the scratch, for landslide susceptibility mapping in a regional scale. The aim is to define an updated methodological approach in which each decision in each step of the process would be justified and scientifically supported.

An administrative region of 1980 km^2 located in the north of the Iberian peninsula (Gipuzkoa Province, Basque Country) was used as test study area where several experiments and applications were carried out. Independent variables of different meaning and qualities, landslide inventories of divers types and sources, and already known methods together with innovative approaches were tested in order to strength the conclusions of this work.

Results indicate the need of the geomorphological inference when selecting the independent variables by statistically driven rules and the convenience of transforming the categorical variables into continues for the susceptibility models calculated by statistical approaches. Moreover, the use of an effective surveyed

area for calibrating the landslide susceptibility models proved to be positive when the landslide inventory is provided by geomorphological field work, whereas the use of slope units instead of the conventional pixel mapping units demonstrated the capacity of mitigating the uncertainty introduced by the field-based landslide inventory.

Additionally, the application of an algorithm for the automatic definition of precipitation thresholds responsible of landslides in the study area showed the necessary information to develop landslides occurrence forecasts basing in precipitation predictions, highlighting the available potential resources for advancing toward an integrated early warning system.

Laburpena

Mundu osoan zehar, lur labainketak etengabe gertatzen dira eta haien ondorioz gizarteak nahiz bere ondasunak kalteak pairatzen ditu. Lurraldearen zonazioa lur labainketak jasotzeko aukeren arabera da, suszeptibilitate mapen bitartez hain zuzen ere, mota honetako prozesu geomorfologikoek eragindako kalteak arindu eta mehatxua eta arrisku maila ebaluatu ahal izateko oinarrizko pausoa.

Azken 50 urteotan, lurraldearen suszeptibilitate mapetan, ingurunearen behaketa hutsean oinarritutako balorazioetatik prozedura erabat automatizatu eta sofistikatuetara igaro gara. Bestalde, planteamendu desberdinen hazkunde azkar horiek metodo desberdinen erabilgarritasun handia ekarri du; eta hortaz, lur labainketen suszeptibilitate analisiaren estandarizazioa zaildu. Tesi honek bide-orri baten definizioa aurkezten du, hutsetik abiatuta, lurraldearen suszeptibilitate mapen garapenerako eskala erregionalean. Helburua ikuspegi metodologiko eguneratu bat zehaztea da, prozeduraren pauso bakoitzean hartutako erabakiak zientifikoki justifikatuak eta onartuak izan daitezen.

Hainbat esperimentu eta aplikazio gauzatzeko, Iberiako penintsularen ipa-rraldean kokatua dagoen eskualde administratibo bat hautatu da (Gipuzkoako Lu-rralde Historikoa), 1980 km²-ko ikerketa eremua hain zuen ere. Ezaugarri eta izaera desberdinetako hainbat aldagai independente, mota eta iturri desberdinetako lur labainketa inbentarioak eta, metodo ezagun nahiz metodo berritzaileekin batera, ikuspegi desberdinak jorratu dira, azkenean lan honek aurkezten dituen ondorioak lortzeko.

Emaitzen arabera, inferentzia geomorfologikoaren beharra azpimarratu daiteke estatistikoki gidatutako arauen arabera aldagai independenteen hautaketa egiteko orduan, hala nola ondorioztatu da aldagai kategorikoen transformazioa aldagai

jarraietan, suszeptibilitate-modeloak garatzeko onuragarria dela. Gainera, lur labainketen suszeptibilitate modeloak kalibratzeko ikuskatutako eremu efektiboaren erabilera positiboa dela frogatu da, lur labainketen inbentarioa landa lanaren bitartez eskuratu den kasuetan. Bestalde, malda unitateen erabilerak lurralde unitate bezala, ohikoak diren pixel unitateak izan beharrean, landa laneko inbentario batek ezarri dezakeen ziurgabetasuna arintzeko ahalmena erakutsi du.

Horretaz gain, lur labainketak gertatzeko beharrezko prezipitazio atalasa definitzeko algoritmo baten aplikazioak ikerketa eremu berberean, aurreikuspenak egiteko beharrezkoa den informazioa erakutsi du, alerta goiztiarreko sistema baterantz aurreratzen joateko dauden aukerak zabalduz.

Dedication

Malda ezegonkorra da

horretan naiz jakituna

Bizitzak ere baditu

hainbat momentu astunak

Gurasoei doakie

nire lanaren bilduma

Haiek argitu bait dute

bide hontako iluna

Alde batetik bestera

eraman nauen zalduna

Gauero muxu batekin

ohera sartzen ninduna

Inoiz labaindu ezkero

zuengan daukat laguna

Zuek zuzentzen duzue

nire ezegonkortasuna

Fito eta Rosari

Declaration

I certify that I might have conferred with others in preparing for this assignment, and drawn upon a range of sources in this work, the content of this thesis work is my original work.

Indeed, the main results of the current thesis have led to the following publications as research article:

- Mapas de susceptibilidad de deslizamientos a partir del modelo de regresión logística en la cuenca del río Oria (Gipuzkoa). Estrategias de tratamiento de variables. https://doi.org/10.17735/cyg.v32i1-2.59493

- Effective surveyed area and its role in statistical landslide susceptibility assessments. https://doi.org/10.5194/nhess-18-2455-2018

and the following chapter in the book entitled **Education for Risk Reduction** of the **Risks and Catastrophes** series:

- Landslide & rainfalls: press inventory, conditioning factors characterization and precipitation thresholds for Gipuzkoa Province (Basque Country).

Acknowledgement

List of abbreviations and acronyms

List of abbreviations and acronyms.

Laburdura eta akronimoen zerrenda.

ADIF	Administración de Infraestucturas Ferroviarias
AEMET	Agencia Estatal de Meteorología
ALLISSA	Abridged Landslide Inventory of Spain for synoptic Susceptibility Assessment
AUC	Area Under the Curve
CART	Classification and Regression Trees
DEA	Datu Espazialen Azpiegitura
ELSUS	European landslide susceptibilitu mapa
ESA	Effective Surveyed Area
ESA-PM	Effective Surveyed Area, Pixel Map
ESA-SUM	Effective Surveyed Area, Slope Unit Map
EVE	Ente Vasco de Energía
GFA	Gipuzkoako Foru Aldundia
GIS	Giografiako Informazio Sistemak / Geographical Information Systems
GP	Gipuzkoa Province
IFFI	The Italian landslide inventory
IRC	Near Infra Red
LDA	Linear Discriminant Analysis
LH	Lurralde Historikoa
LLISCAT	Landslide data base of Catalonia
LR	Logistic Regression
LS	Landslide Susceptibility
MARS	Multivariate Adaptative Regression Splines
MAXENT	Maximum Entropy
NDVI	Normalized Difference Vetegation Index
NNA	Neural Network Analysis
QDA	Quadratic Discriminant Analysis
R	Red
ROC	Receiver Operating Characteristic
SAR	Surface Area Ratio
SU	Slope Unit
TWI	Topographic Wetness Index
WA-PM	Whole Area, Pixel Map
WA-SUM	Whole Area, Slope Unit Map

Contents

List of Figures

List of Tables

Chapter 1

Sarrera

Lur labainketak arrisku natural ohikoenetako bat dira. Esaterako, 1900 urteaz geroztik, eta mundu osoan, gutxienez 9.815 milioi US\$-eko galera ekonomikoak eta 68.098 hildako eragin dituztela zehazten du EM-DAT nazioarteko ezbeharren datu baseak (EM-DATA, 2017), nahiz eta kopuru horiek nabarmenki gutxietsita daudela kontsideratzen den. Izan ere, Petleyren arabera (Petley, 2012), 2004 eta 2012 urteen artean bakarrik 32.322 hildako kontabilizatu ziren lur labainketen ondorioz. Uholde (%44) eta ekaitzen (%35) atzetik, lur labainketak munduan erregistratutako ezbeharren %6 izan ziren 1970 eta 2012 urteen bitartean (CRED, 2014), baina prozesu horien eragina gizakiarengan ongi ezaguna den arren, uste da labainketa kartografiek munduko malda guztien %1 baino gutxiago hartzen dutela kontuan (Guzzetti *et al.*, 2012). Hala ere, fenomeno geomorfologiko horien jarraipen historikoa egitea oso zaila da, haiei buruzko datuen bilketa sistematikoen eskasia dela eta.

Gaur egun, horrelako prozesu dinamikoen ezagutzak erronka itzela suposatzen du oraindik. Prozesuen definiziotik hasita, haiek ikertzeko metodologietara arte, hainbat proposamen aurki daitezke literatura zientifikoan eta, beraz, lur labainketen analisiaren arloa, oraindik ere, fase esperimentalean dagoela esan genezake (Fell *et al.*, 2008).

Hala eta guztiz ere, tesi honen gaian zeharo murgildu aurretik, zenbait kontzeptu orokor aurkeztuko dira, testuaren irakurketan hainbatetan errepikatuko diren terminoak egoki definituta geratu daitezen.

Lehenik eta behin, **masa mugimenduaren** (*mass movement*) kontzeptua aipatu beharra dago. Lurrazaleko material bat grabitatearen ondorioz mobilizatua izatearen prozesuari esaten zaio. Kasu honetan, uraren, izotzaren edo haizearen garraio eragin zuzenik gabe gertatutako mugimendua da, eta mobilizazioak, grabitatearen ondoriozkoa izanda, norantza bertikalean izan behar du. Definizio honen barruan onartzen da, beraz, ibilbide bertikala duen edozein lurrazaleko mobilizazio, lurzoruaren subsidenzia prozesuak barne.

Lur labainketa (*landslide edo slope failure*), ordea, masa mugimendu mota bat da. Zehazki, mobilizatua den masa mendi hegal edo malda batean zehar higitzen denean erabiltzen da kontzeptu hau, eta ondo definitutako haustura azaleraren eta mobilizatutako masaren deformazio ertainak ezaugarritzen du. Edozer dela

ere, autoreen arabera kutsu desberdinetako definizioak aurki daitezke, eta definizio dibertsitate horrek, hain zuzen ere, fenomeno honen azterketan diharduten diziplinen aniztasuna islatzen du.

Hona hemen lur labainketa (*landslide*) terminoaren definizio batzuk:

- Arroka, lurzoru zein detritu (*debris*) masa baten mugimendua maldan behera (Cruden, 1991).

- Grabitatearen indarraren ondorioz maldan behera mobilizatzen den arroka edo lurzoru masa bat (Cruden & Varnes, 1996).

- Malda osatzen duten materialen beheranzko eta kanporanzko mugimendua. Mota askotariko prozesuak barneratzen ditu, eta mobilizatutako materiala arroka, lurzorua, betelan artifizialak edo haien arteko konbinaketak izan daitezke (USGS, 2004).

- Mendi hegaletan gertatzen diren lurzoru edo arroka mugimendu grabitazionalak dira. Mugimendu horiek maldaren egonkortasun baldintzen eraldaketaren ondorioz gertatzen diren orekatze prozesuak dira (Ferrer & García, 2009).

- Materialak grabitatearen eraginez mugiarazteko prozesua da, eta arriskutsua izan daiteke gizakiari edo haren ondasunei eragiten dienean (Gutiérrez-Elorza, 2008).

Ez da erraza euskarazko termino zehatza aurkitzea ingelesezko "*landslide*" hitz generikoa itzultzeko. Are gehiago, hainbat ikerlarik azpimarratu bezala (Crozier, 1986; Shanmugam & Wang, 2015), historikoki prozesu bera izendatzeko hainbat izen proposatu dira: *slope failure* (Ward, 1945), *mass wasting* (Yatsu, 2007), *mass movement* (Hutchinson, 1968), *landslides* (Varnes, 1958) edo *slope movement* (Varnes, 1978). Euskadiko Geologoen Elkargo ofizialak "lur irristatze" terminoa erabiltzen du (Aizpiri *et al.*, 2014), baina tesi honetan "lur labainketa" edo "*landslide*" terminoak erabiliko dira aurrerantzean prozesu horiei erreferentzia egiteko. Hala eta guztiz ere, aipatutako definizioen baitan izaera oso desberdinetako masa mugimendu ugari onartzen dira, harri jausiak edo fluxu korronteak adibidez, nahiz eta azken horietan labainketarik ez egon edo labainketa oso eskasa izan, zentzu zehatzean.

Halaber, errealitatean aurki daitezkeen lur labainketak hainbeste prozesu eta faktoreren ondorio dira, ezen klasifikazio aukera mugagabeak eskaintzen dituzten. Eragindako arroka edo lurzoruaren propietate fisikoen arabera (Terzaghi, 1943), mugimendu moten eta abiaduraren (Sharpe, 1938), ezaugarri geoteknikoen (Skempton & Hutchinson, 1969), klasifikazio morfometrikoaren (Crozier, 1975) edo prozesu eragileen araberako (Brunsden, 1993) proposamenak aurki daitezke gaur-gaurkoz. Baina, eskala erregionaleko suszeptibilitate, mehatxu eta arriskuaren ikerketetan Cruden eta Varnesen klasifikazioa da zabalduena (Cruden & Varnes, 1996; Corominas & Mavrouli, 2011), zeina hauetan oinarrituta dagoen: Varnes (1958), Varnes (1978) eta Casale *et al.* (1994). Gainera berriki eguneratua izan da Hungren proposamenean (Hungr *et al.*, 2014). Beraz, klasifikazio horien arabera, ondorengo lur labainketa mota nagusi hauek definitzen dira (ikusi 1.1 Irud.):

- **Erorketak edo jausiak** (*falls*): Haustura azalera baten ondorioz, ezponda oso malkartsuetatik masa bat banatzean gertatzen den prozesuari esaten zaio. Normalean arrokaz osatutako materialak izaten dira, nahiz eta metakinek edo lurzoruek ere jasan ditzaketen horrelako prozesuak. Pareta bertikal eta sub-bertikaletan banandutako fragmentuak erorketa librean jausten dira, momenturen batean behintzat; baina, ezpondaren angelua murrizten doan heinean, materialok saltoka nahiz errodatuz garraia daitezke. Mota honetako lur labainketen eragile nagusiak izozte-urtze zikloak (krioklastia), oinarriaren itsas edo ibai higadura, lurrikarak eta intentsitate handiko eurijasak izan ohi dira.

- **Iraulketak** (*topples*): Iraulketa gertatzen da mendi hegalaren kanpo aldera arroka masa bat biratzen denean, desplazatutako masaren grabitate zentruaren azpitik dagoen puntu edo ardatz baten arabera. Estruktura bertikaleko materialetan gertatzen dira horrelako prozesuak, zeinak diskontinuitate planoez osatuta dauden. Mobilizatutako materialen eta dimentsioen arabera mota desberdinak aurki daitezke. Eragile nagusiak deskonpresioa, izozte-urtutze zikloak eta gatz meteorizazioa dira.

- **Labainketak edo irristatzeak** (*slides*): Labainketa bat, zentzu zehatzenean, zizailatze mugimenduaren ondorioz haustura azalera batean zehar mugitzen

5

den gutxi deformatutako arroka edo lurzoru masa bat da. Sarritan, mobilizatutako materialaren sakoneraren arabera, **azaleko labainketa** (*shallow slide*) eta **sakoneko labainketa** (*deep slide*) desberdintzen dira. Bestalde, haustura azaleraren formaren arabera bi mota nagusi desberdintzen dira. **Labainketa errotazionaletan** (*rotational slide*) zizailatze mugimendua azalera ahur batean zehar gauzatzen da. **Labainketa translazionaletan** (*translational slide*), berriz, desplazamendua azalera lau edo apur bat ondulatuan zehar egiten da, diskontinuitate planoen ahulezia puntuak baliatuz normalean (estratifikazioak, failak, diaklasak). Mota honetako mugimenduak sortzeko eragile nagusien artean eurijasak, lurrikarak, elurraren urtze azkarra edo gizakiaren aktibitatea daude.

- **Alboko hedadura** (*lateral spreading*): Kasu honetan haustura azalera ez da bat bateko zizailatze mugimendu baten ondorioa. Bigunagoa den azpiko material baten azaleko arrokaren albo baterako desplazamenduari esaten zaio, eta hainbat prozesuren eraginez gerta daitezke horrelako mugimenduak (hausturen hedadura, geruzen tolestea, higadura, failak) (Gutiérrez-Santolalla *et al.*, 2005). Zenbait autoreren arabera *sackung* fenomenoa alboko hedaduraren fenomenoen barruan sartuko litzateke (Jahn, 1964).

- **Fluxuak** (*flows*): Airearen edo uraren ondorioz mobilizatutako materialaren barne deformazioa dela eta, desplazatutako masak fluxu portaera erakusten duenean gertatzen dira. Orokorrean materialak lurrazalean zehar mugitzen dira eta oso distantzia luzeetan barrena mobiliza daitezke. Kasu honetan ere, desplazatutako material motaren arabera hainbat fluxu mota desberdindu daitezke. Kontsolidatu gabeko materialak izan daitezke, hots **detritu fluxuak** (*debris flow*) edo **lurzoru fluxuak** (*soil/earth/mud flow*); edo ondo kontsolidatutako materialak, hots **arroka fluxuak** (*rock flow*).

- **Mugimendu konplexuak** (*complex movements*): Errealitatean lur labainketa gehienek prozesu osoan zehar portaera bat baino gehiago erakusten dute, izan mobilizatutako masaren ataletan, izan prozesuaren etapetan. Beraz, oso ohikoa da hasieran mota zehatz bateko lur mugimendua denak, ondoren garraioan zehar beste mota bateko izaera hartzea. Horrelako kasuetan, fenomeno

6

hauek mugimendu konplexuen taldean sailkatu dira. **Arroka-jausiak** (*rock avalanche*) eta **fluxu labainketak** (*flow slide*) dira adibide ohikoenak.
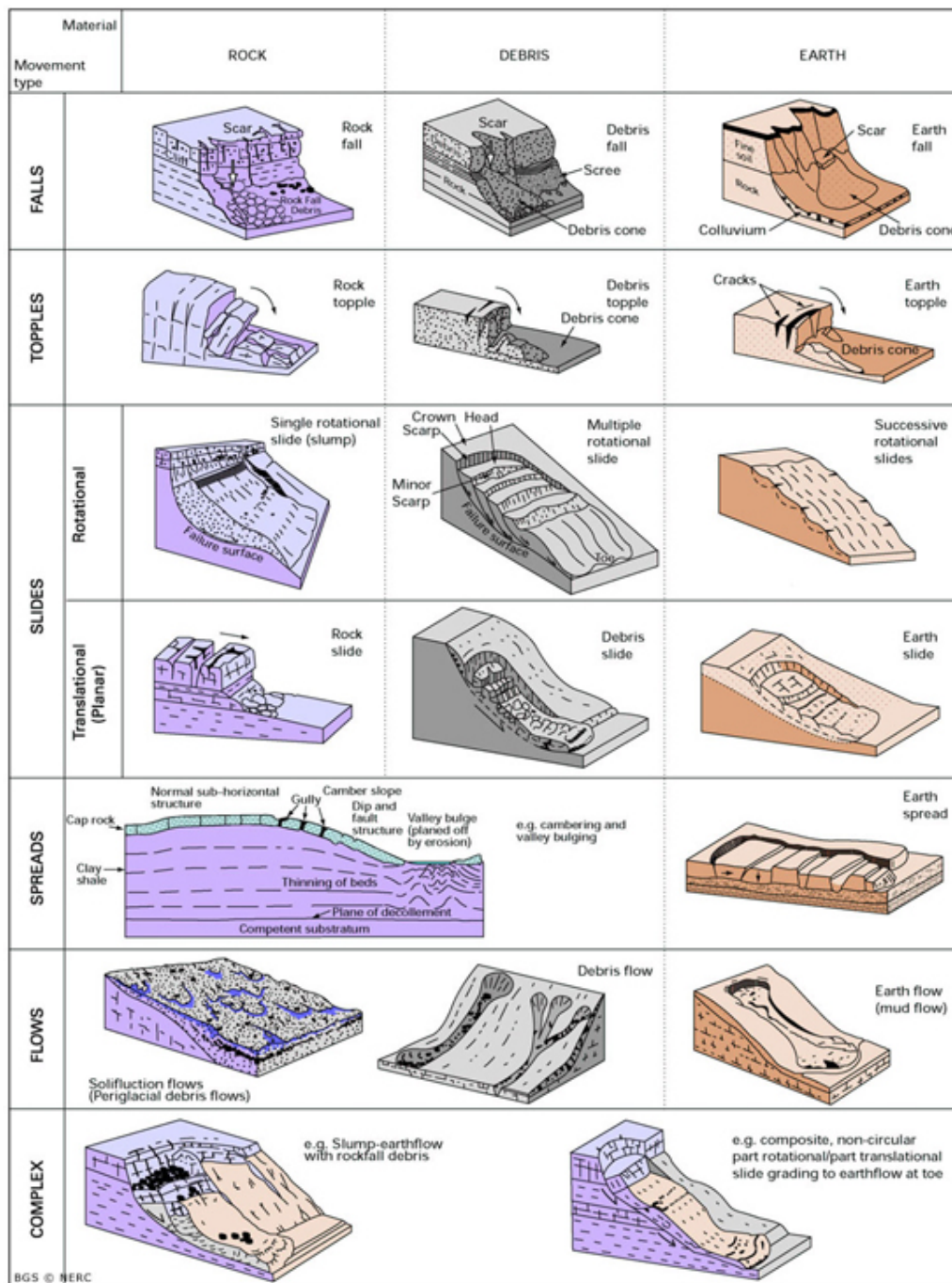
Figure 1.1: *Illustration of the abbreviated classification of landslides types proposed by Cruden & Varnes (1996). Image obtained from the British Geological Survey web site. www.bgs.ac.uk (last visit at 05-06-2017).*

1.1 Irudia: Cruden eta Varnesek proposatutako lur labainketa moten sailkapen sintetikoa (Cruden & Varnes, 1996). Irudia British Geological Survey-aren web orrialdetik aterata. www.bgs.ac.uk (*azkeneko bisita 2017-05-06*).

Lur labainketa mota hauen guztien artean, magnitude oso desberdinetako prozesuak aurkitzen dira. Badira urtean milimetro gutxi batzuk mugitzen diren lur

masak eta baita bat-batean gertatzen diren lur labainketak, badira m^3 gutxi batzuk mobilizatzen dituzten prozesuak eta baita milioika m^3 mugimenduak ere, eta badira metro gutxi batzuetako garraioa jasaten duten materialak eta baita ehundaka, eta zenbaitetan milaka, metrotako irismena dutenak ere. Baina, edozein dela ere lur labainketa bakoitzaren magnitudea, lurraldean eraldaketa bat eragiten du, eta beraz, lurraldea erabiltzen duen edozeini hots, gizakiari zuzenean edo zeharka, ondorioren bat ekarriko dio.

Hain zuzen ere, prozesu natural horiek, gizakiari, bere ondasunei edo inguruneari kalte egiten dietenean, arrisku (*risk*) bilakatzen dira, eta prozesu natural horietakoren bat gertatzeko probabilitatea eta intentsitatea ezagutuz gero, P mehatxua (*hazard level*) defini daiteke (Ayala-Carcedo & Olcina Cantos, 2002). Beraz, aski jakina den bezala, lur labainketa bat jasateko aukera duten pertsonak eta ondasunak identifikatuz gero, hots E esposaketa (*expoure*), eta ondasun horiek galtzeko onargarritasun maila zehaztuz gero, hots V ahultasuna (*vulnerability*), lur labainketen R arriskua definituko litzateke jarraian aurkezten den formula matematikoaren bitartez (Ayala-Carcedo & Olcina Cantos, 2002):

$$R = PxExV \qquad (1.1)$$

Ekuazio hau (Eq. 1.1) edozein motatako arrisku naturalen analisian aplika daiteke eta bertan desberdindu daitezke arriskuaren ebaluaketarako beharrezkoak diren bi oinarrizko faktore multzoak: prozesu naturala bera aztertzen dutenak (mehatxua) eta gizakiaren presentzia aztertzen dutenak (esposaketa eta ahultasuna).

Argi geratu da beraz, lur labainketak, kanpoko geodinamikako prozesu naturalak izateaz gain, gizakiarentzako arrisku potentzial bat direla. Labainketa horiek eragin ditzaketen kalteak ekidin edo, ahal den heinean, arintzeko bide eraginkorrena lurralde antolamendua da. Ondorioz, arriskuaren ebaluaketaren azkeneko emaitza, haren banaketa espaziala eremu geografiko batean zehar islatuko duen mapa bat izan ohi da. Horrek esan nahi du, azkeneko produktu horretara iritsi aurretik, oinarrizko faktore guztien banaketa espaziala ezagutu beharra dagoela. Alta, mehatxua, esposaketa nahiz ahultasuna kalkulatu eta espresatzeko metodologiak ugariak dira gaur egun, eta ez dago komunitate zientifikoak aho batez onartzen duen bide orririk.

Tesi hau mehatxuaren ebaluaketara bideratuta dago. Zehatzago esanda,

mehatxua definitzeko beharrezkoa den suszeptibilitatea kalkulatzeko metodologia proposamen bat aurkeztuko da. Izan ere, mehatxua denbora eta eremu zehatz batean potentzialki kaltegarria izan daitekeen fenomeno baten gertatzeko probabilitateri esaten zaio (Remondo, 2001). Beraz, definizio horren arabera, mehatxuak espazio probabilitatea eta denbora probabilitatea uztartzen ditu. Suszeptibilitatea ordea, labainketa bat ziurrenik non gerta daitekeen estimazio bat da, eta matematikoki ingurugiro baldintza zehatz batzuen arabera labainketa bat gertatzeko espazio probabilitatea bezala defini daiteke (Brabb, 1984; Guzzetti *et al.*, 2005).

Hori dela eta, edozein arrisku natural mota behar den bezala ebaluatzeko lehen pausoa arriskua sor dezakeen fenomenoaren suszeptibilitatea ezagutzean datza, eta tesi honek pauso konkretu hori aurrera eramateko metodologian sakonduko du, lur labainketen kasuan proposamen zehatz bat eskainiz. Azken finean, lur labainketak espazialki aurreikusten dituen modelo matematiko egoki bat sortzea da helburua, baina, helburu hori gauzatzeko, aldez aurretik gako batzuk argitu behar dira:

- Zein aldagai erabili behar dira modelo hori kalkulatzeko? Edo nola egin aldagaien hautaketa?

- Nola tratatu behar dira aldagai horiek?

- Zein mapa unitate erabiliko dira lurraldea banatzeko? (pixelak, malda unitateak, unitate homogeneoak, etab.)

Horretaz gain, lur labainketak sortzeko gertaera-faktorearen gaia (*triggering factor*) ere jorratuko da. Hau da, labainketak gauzatzeko probabilitate espaziala areagotzeko, lurraldearen eta inguruneko baldintzak zeintzuk izan daitezkeen ezagutzeaz gain, momentu zehatz batean maldaren haustura gertaraziko duen faktorea ere ikertuko da.

Helburu orokor horiek guztiak gauzatu ahal izateko ikerketa eremu esperimental bat aukeratu da, Gipuzkoako Lurralde Historikoa (hemendik aurrera Gipuzkoako LH), iberiar penintsulako iparraldean kokatutako 1980 km^2 inguruko lurraldea (ikusi 4 atala). Bertan, aukera metodologiko batzuen arteko froga desberdinak aplikatu dira eta, haien emaitza eta ondorioetan oinarrituta, proposamen zehatz bat definitu ahal izan da argitu gabeko gako horiei erantzun bat eman ahal izateko.

9

Testua zortzi atal nagusitan banatuta dago. Sarrera atalaren ostean (Chapter 1), bigarren atalean (Chapter 2), lur labainketen suszeptibilitate eta mehatxuaren analisiari buruzko berrikuspen historiko bat egingo da, nazioartetik hasita Gipuzkoako LHraino aurki daitezkeen aurrekariak aipatuz eta eskuragarri dauden metodologien aldeko eta kontrako ezaugarriak aurkeztuz.

Hirugarren atalean (Chapter 3), ikerketa honen hipotesia eta hura frogatu ahal izateko ezarri diren helburuak zehaztuko dira.

Laugarren atalean (Chapter 4), aurretik finkatutako helburuak betetzeko eta hipotesia egiaztatu ahal izateko gauzatu diren esperimentuak aplikatu diren ikerketa eremua deskribatuko da. Ikuspegi orokor batetik, lurraldearen ezaugarri geomorfologiko eta klimatikoak, ingurugiro zehaztapenak nahiz populazioaren banaketa aurkeztuko dira.

Bosgarren atalean (Chapter 5) ikerketa prozesuan zehar jarraitu den metodologia azalduko da. Erabili diren datuen jatorria edo haiek biltzeko metodoak, aplikatu diren modelo matematikoak eta softwareak eta, oro har, ikerketa honetan aurkeztuko diren esperimentu desberdinetan jarraitu diren pausoen zehaztapenak aurkeztuko dira.

Bestalde, seigarren atalean (Chapter 6), tesian zehar gauzatu diren hiru esperimentuen emaitzak aurkeztuko dira. Esperimentu bakoitza helburu zehatz batzuk lortzeko diseinatu da, eta hipotesian ezarritako ustezkoak frogatzeko egin da. Hori dela eta, 6-I atalean, lur labainketen suszeptibilitate mapak sortzeko beharrezkoak diren labainketa inbentarioen ezaugarriak aztertzen dira, modelo estatistikoetako aldagai independenteen tratamendurako estrategia desberdinen ebaluaketa jorratzen da eta aldagaien aukeraketarako bide desberdinez eztabaidatzen da; 6-II atalean, azterketa konparatibo baten bitartez, zuzeneko landa lanaren ondorioz sortutako labainketa inbentario bat erabiliz, suszeptibilitate mapak sortzeko metodologia berri bat proposatzen da, eta bi mapa unitate desberdinetan aplikatu egiten da; 6-III atalean, suszeptibilitate analisitikan, mehatxu analisira igarotzeko lehen pausoak ematen dira, denbora informazioa erabiliz, eta ikerketa eremurako labainketak gertatzeko euri atalase sorta bat kalkulatuz metodo probabilistikoaren bidez.

Zazpigarren atalean (Chapter 7), aurreko esperimentuetatik ateratako ondorio

nagusiak eta ondorio horietatik atera daitezkeen ideiak eztabaidatuko dira. Eta, azkenik, zortzigarren atalean (Chapter 8), ikerketa honek azaleratu dituen ezagutzak kontuan hartuta, lur labainketen suszeptibilitate eta mehatxu mapak egiteko orduan jarraitu beharreko irizpideen proposamena egingo da.

Bibliografia

Aizpiri, F. J., Guerrero, D., & Ormaetxea, V.: Lur irristatzeak gertatzeko arriskua. Zuk jakin beharrekoa, Ilustre Colegio Oficial de Geólogos del País Vasco, 2014.

Ayala-Carcedo, F. J. & Olcina Cantos, J.: Riesgos Naturales, Ariel Ciencia, Barcelona, 2002.

Brabb, E. E.: Innovative approaches to landslide hazard and risk mapping, in: 4th Int Symp Landslides, Toronto, vol. 1, pp. 307–323, 1984.

Brunsden, D.: Mass movement; the research frontier and beyond: a geomorphological approach, Geomorphology, 7, 85–128, 1993.

Casale, R., Fantechi, R., & Flageollet, J. C.: Temporal occurrence and forecasting of landslides in the European Community, Tech. rep., Community Research and Development Information Service, European Commission, 1994.

Corominas, J. & Mavrouli, O. C.: Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies, Tech. rep., SafeLand. 7th Framework Programme Cooperation Theme 6 Environment (including climate change) Sub-Activity 6.1.3 Natural Hazards, 2011.

CRED: Atlas of Mortality and Economic losses from Weather, Climate and Water Extremes (1970-2012), Tech. rep., World Meteorological Organization, Geneva, 2014.

Crozier, M. J.: Techniques for the morphometric analysis of landslides, Zeitscrift für Geomorphologie, 17, 78–101, 1975.

Crozier, M. J.: Landslides: causes, consequences & environment, Croom Helm, London, 252 pp, 1986.

Cruden, D. M.: A simple definition of a landslide, Bulletin of Engineering Geology and the Environment, 43, 27–29, 1991.

Cruden, D. M. & Varnes, D. J.: Landslide types and processes, in: Landslides: Investigation and Mitigation, edited by Turner, A. K. & Jayaprakash, G., 247, Transportation Research Board Special Report, Washington D.C., 1996.

EM-DATA: The Emergency Events Database - Université catholique de Louvain (UCL) -, URL `www.emdat.be`, Brussels, Belgium, 2017.

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, Engineering Geology, 102, 85–98, 2008.

Ferrer, M. & García, J. C.: Guia para la elaboración de mapas inventario y de susceptibilidad de movimientos de ladera a escala 1:50.000, Instituto Geológico y Minero de España (IGME), 2009.

Gutiérrez-Elorza, M.: Geomorfología, Ed. Pearson educación, S.A., Madrid, 920 pp, 2008.

Gutiérrez-Santolalla, F., Acosta, E., Ríos, S., Guerrero, J., & Lucha, P.: Geomorphology and geochronology of sackung features (uphill-facing scarps) in the Central Spanish Pyrenees, Geomorphology, 69, 298–314, 2005.

Guzzetti, F., Reichenbach, P., Cardinali, M., Galli, M., & Ardizzone, F.: Probabilistic landslide hazard assessment at the basin scale, Geomorphology, 72, 272–299, 2005.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Reviews, 112, 42–66, 2012.

Hungr, O., Leroueil, S., & Picarelli, L.: The Varnes classification of landslide types, an update, Landslides, 11, 167–194, 2014.

Hutchinson, J.Ñ.: Mass movements, in: The Encyclopedia of Geomorphology, edited by Goudie, A., pp. 688–695, Routledge, New York, 1968.

Jahn, A.: Slopes morphological features resulting from gravitation, Zeitscrift für Geomorphologie Supplement Band, 5, 59–72, 1964.

Petley, D.: Global patterns of loss of life from landslides, Geology, 40, 927–930, 2012.

Remondo, J.: Elaboración y validación de mapas de susceptibilidad de deslizamientos mediante técnicas de análisis espacial, Ph.D. thesis, Universidad de Oviedo, Oviedo, 2001.

Shanmugam, G. & Wang, Y.: The landslide problem, Journal of Palaeogeography, 4, 109–166, 2015.

Sharpe, C. F. S.: Landslides and related phenomena, Columbia University Press, New York, 138 pp, 1938.

Skempton, A. & Hutchinson, J.: Stability of natural slopes and embankment foundations, in: Seventh International Conference on Soil Mechanics and Foundation Engineering, pp. 291–340, Mexico City, 1969.

Terzaghi, K.: Theoretical soil mechanics, vol. 18, John Wiley and Sons, New York, 510 pp, 1943.

USGS: Landslide Type and Processes, U.S. Department of the Interior, 2004.

Varnes, D. J.: Landslide types and processes, Landslides and Engineering Practice, 24, 20–47, 1958.

Varnes, D. J.: Slope movements: types and processes, in: Landslides: Analysis and Control, edited by Schuster, R. L. & Krizek, R. J., 176, pp. 11–33, Transportation Research Board Special Report, Washington D.C., 1978.

Ward, W. H.: The stability of natural slopes, The Geographical Journal, 105, 170–197, 1945.

Yatsu, E.: 1962: Rock control in geomorphology. Tokyo, Sozosha., Progress in Physical Geography, 31, 199–202, 2007.

Chapter 2

Gaiaren testuingurua

Lur labainketen analisiak, eskala erregionalean, iraganera begira egiten dira. Izan ere, dinamika natural batean, lur labainketak eragiten dituzten faktore gehienak berdinak izango dira iraganean eta etorkizunean. Baina errealitatean lur labainketen problematika ulertzeko ideia horri, gaur egungo gizakiak ingurunean eragiteko daukan ahalmena gehitu behar zaio, dinamika natural horretan jokatzen duten faktoreei askoz anitzagoa eta kuantifikatzeko zailagoa den faktore bat gehituz, eragin antropikoa. Hala ere, lur labainketa analisiak errealitatearen gerturatze bat diren heinean, Fell eta beste aditu batzuek ezarri zuten bezala (Fell *et al.*, 2008), lur labainketen azterketan bi baldintza hauek ematen dira onartutzat:

- Iraganean lur labainketak izan diren lekuetan, gaur egun edo etorkizunean lur labainketak jasateko joera egongo da.

- Gaur egungo eta etorkizuneko lur labainketak, iraganean zeuden ingurune baldintza berdinen menpe gertatuko dira.

2.1 Aurrekariak

Aipatu den bezala, naturaren zientziaren hainbat arlotan mendetako garapena izan den arren, lur labainketen suszeptibilitatearen arloan hamarkada gutxi batzuetako ibilbidea besterik ez da egin oraindik. Baina urte gutxitako ibilbidea izanda ere, horrek ez du metodologia ugariren agerpena eragotzi, izan ere literaturan metodologia oso desberdinen proposamenak aurkitu daitezke. Metodo horiek guztiak hiru multzo nagusitan banatu daitezke: (i) ezagutzan oinarritutakoak, (ii) datuetan oinarritutakoak edo (iii) deterministikoak (Soeters & Van Westen, 1996; Fell *et al.*, 2008).

Lur labainketen zonazioaren lehen proposamen formalak 70. hamarkadan plazaratu ziren (Brabb *et al.*, 1972; Humbert, 1970, 1977; Kienholz, 1978; Nilsen, 1979). Garai hartan ezagutzan oinarritutako metodoak, edo metodo heuristikoak, jarraitzen ziren, zeinetan adituaren ezaguera eta esperientzia erabiltzen zen, zuzenean (metodo geomorfologikoa) edo indizeetan oinarrituta (metodo erdi-kuantitatiboa), lurraldearen ezegonkortasunerako joera maila ebaluatzeko. Kasu horietan, emaitzaren kalitatea adituaren ezagutza mailak erabat baldintzatuta geratzen da subjektibotasun oso altua emanez. Gainera, ezegonkortasun joera maila

ezartzeko orduan hartutako erabakiak ezin dira ez egiaztatu, ezta erreproduzitu ere (Van Westen, 1993; Cardinali *et al.*, 2002), eta horrek gabezia garrantzitsuak uzten ditu, batez ere egindako lanen kalitatea ebaluatu edo konparazioak egiterakoan. Hain zuzen ere gabezia horiei aurre egin nahian, 80. hamarkadatik aurrera (Brand & Bonnard, 1988), baina batez ere 90. hamarkadan, metodo kuantitatibo edo estatistikoak garatzen hasi ziren (Wong *et al.*, 1997; Hardingham *et al.*, 1998; Wong & Ho, 1998). Analisirako proposamen horiek metodo heuristikoetan aurki daitekeen subjektibitate maila ahalik eta gehien murrizteko sortu ziren. Horretarako, iraganeko ezegonkortasunak gertarazi zituzten faktoreak estatistikoki aztertzen dira lurraldearen unitate bakoitzerako, eta datu horietan oinarriturik, oraindik ezegonkortasunik egon ez den eremuetarako aurreikuspenak egin daitezke antzeko ezaugarriak mantentzen badira. Era berean, metodo deterministikoak (metodo fisikoak ere deitzen direnak) 80. hamarkadaz geroztik garatu ziren (Ward *et al.*, 1982; Okimura & Kawatani, 1986; Mulder & Van Asch, 1988; Mulder, 1991; Hammond *et al.*, 1992; Godt *et al.*, 2008; Gökceoglu & Aksoy, 1996). Malda batean aurkitzen diren, eta neur daitezkeen ezaugarrietan oinarriturik (malda, lurzoruaren sakonera, lur azpiko uren sakonera, etab.) ezponda baten masaren segurtasun muga (*safety factor*) kalkulatzen da, eta atalase hori gaindituz gero mobilizatzen hasiko da masa hori. Informazio horri esker, kalkulurako erabilitako faktore bakoitzaren mugako balioak defini daitezke modelo fisiko batzuen arabera (adibidez, ezponda infinituaren modeloa edo *infinite slope model*), eta beraz, maldaren portaera aurreikusi faktore bakoitzaren momentuko balioen arabera. Hala eta guztiz ere, 2.1 taulan islatuta geratzen den bezala, multzo bakoitzaren baitan metodologia desberdin ugari aurki daitezke literaturan.

Geografiako Informazio Sistemak (GIS) eta estatistikako softwareek azken urteotan izan duten garapenari esker, datuen tratamendua izugarri erraztu da. Horrek lur labainketen suszeptibilitatea, garai batean metodo heuristikoak jarraituz egiten zen denbora berean, estatistikoki edo deterministikoki baloratzea ahalbidetu du. Beraz, ezagutzan oinarritutako metodoak alde batera utzi, eta objektibotasuna eta erreproduzigarritasuna helburu, geroz eta modelo estatistiko zein fisiko sofistikatuagoak garatzen ari dira.

Historikoki lur labainketen suszeptibilitate eta mehatxuaren arloan egindako

Table 2.1: *List of some methods for landslide susceptibility mapping published since 1983 until 2014. Table modified by the author. Source: Malamud et al. (2014).*

2.1 Taula: 1983 eta 2014 urteen bitartean argitaratutako zenbait metodologien zerrenda lur labainketen arloan. Iturria: Malamud *et al.* (2014).

Knowledge based methods (Heuristic)			
Certainty factor model Knowledge based idex approach	Conventional weighting	Heuristic evaluation	Qualitative map combination

Deterministic methods (Factor of Safety calculation)			
Deterministic stability analysis Shallow Slope Stability Model (SHALSTAB)	Factor of safety Stability index mapping (SIN-MAP)	Infinite slope stability model Static factor of safety	Newmark slope stability model

Data driven methods (Statistical)			
Bayes' theorem	Bayesian logistic regression	Bayesian probability model	Bivariate analysis
Failure rate (FR)	Frequency ratio	GIS based	Landslide density
Overlay	Dempster-Shafer theory	Evidential belief function (EBF) model	Discriminant analysis
Kernel-based Fisher discriminant analysis	Linear discriminant analysis	Quadratic discriminant analysis	Quantification scaling type II
Index of entropy	Information entropy	Maximum entropy (MAXENT)	Shannon's entropy
Factor analysis	Fuzzy inference system	Fuzzy logic approach	Fuzzy-set membership function
Gamma Operation	Neuro-fuzzy approach	Zadeh's fuzzy set theory	InfoVal (Information Value) model
Landslide hazard evaluation factor (LHEF) rating scheme	Landslide nominal susceptibility factor (LNSF) methods	Landslide susceptibility index (LSI) model	Matrix method
Multiple factor model	Numerical rating scheme	Ordered weighted average	Ordinal scale approach using weighting rating system
Relative effect method	Soil stability value	Statistical index (Wi)	Surface cover index (SCI)
Vegetation influenced landslide index (VILI)	Weight index (Wi) model	Weighting factor (Wf) method	Generalised linear model (GLM)
Generalized additive model	Generalized linear regression (GLR)	Geographically weighted regression (GWR)	Least squares method
Linear regression	Weighted linear combination	Autologistic regression	Logistic regression
Rare events logistic regression	Analytical hierarchical process	Analytical network process	Kalman filter model
Multi-criteria decision analysis	Spatial multi-criteria evaluation	Genetic adaptive neural network (GANN)	Neural network
AdaBoost	First-order second-moment (FOSM)	Fractal statistics	Genetic programming
Geotechnical-based slope stability probability classification (SSPC)	Rock engineering system (RES)	Rough set theory	Conditional probability
Joint conditional probability	Likelihood ratio	Probabilistic analysis	Probabilistic likelihood ratio
Probability density function	Simple probabilistic method	Multivariate adaptive regression splines (MARS)	Multivariate regression
Spatial regression	Support vector machine (SVM)	Boosted regression trees (BRT)	Classification and regression trees (CART)
Decision tree	Random forest	Weight of evidence	

ekarpenen berrikusketa sakonagoak erreferentzia hauetan aurkitu daitezke: Soeters & Van Westen (1996); Van Westen *et al.* (2008, 1997); Van Westen (2000); Huabin *et al.* (2005); Van Westen *et al.* (2006); Chacón *et al.* (2006); Beguería (2006); Van Asch *et al.* (2007); Kanungo *et al.* (2009); Pardeshi *et al.* (2013); Reichenbach *et al.* (2018). Baina, laburbilduz, gaur egun metodo estatistikoak eta deterministikoak dira erabilienak, eta baten edo bestearen aukera, ikerketa bakoitzaren ikuspegi, helburu eta baliabideen arabera egingo da. Modelo fisikoetan oinarritutako metodoek labainketa bakoitzaren portaera mekanikoaren ikuspegitik aztertzen dituzte fenomenoak, eta eskala handiko eremuetan aplikatzeko egokiak dira bakarrik (<1:5000) (Corominas & Mavrouli, 2011), beti ere sarrera datuen kalitate altua eta eskuragarritasuna bermatzen baldin badira. Metodo estatistikoetan berriz, ikuspegi orokorrago bat hartzen da, eta lur labainketa bakoitza nola gauzatu den baino gehiago, noiz eta non gertatu den aztertzea izaten dute helburu, modelo matematikoetan izaera desberdineko informazio espazial eta denboralak erabiliz sarrera datu moduan. Mota horretako metodoen abantaila nagusia eskala lokal, erregional nahiz nazionalean aplikatzeko aukera da, datu nahikoa izanez gero beti ere. Izatez, azken urteotan sortutako bibliografiaren berrikusketa sakon baten arabera (Malamud *et al.*, 2014), 1983 eta 2014 urteen artean publikatutako lur labainketen suszeptibilitatearen eta mehatxuaren ikerketa estatistikoetan denetariko ikerketa eremuak jorratu dira, 10 km²-tik hasita 10.000 km² baino gehiagora arte (2.1 Irud.).

Bereziki lur labainketen mehatxuaren azterketari lotuta, prozesu hauek zein momentutan gertatzen diren ikertzen duten hainbat proposamen ere aurki daitezke literaturan. Zenbait ikerlarik lur labainketak sortzen dituzten euriteen atalaseak kalkulatzeko metodologien berrikuspen sakonak egin dituzte (Guzzetti *et al.*, 2007; Ramos-Cañón *et al.*, 2015). Suszeptibilitatearen kasuan bezala, hemen ere iraganeko datuen tratamendu estatistikoetan oinarritzen diren metodoak nagusitzen dira, baina prezipitazioen zein parametro erabiltzen den izaten da desberdintasun nagusia batetik bestera. Ramos-Cañónen taldeak gauzatutako berrikuspenean prezipitazioekin lotutako 32 parametro desberdin identifikatzen dituzte aztertutako 244 artikuluen artean.
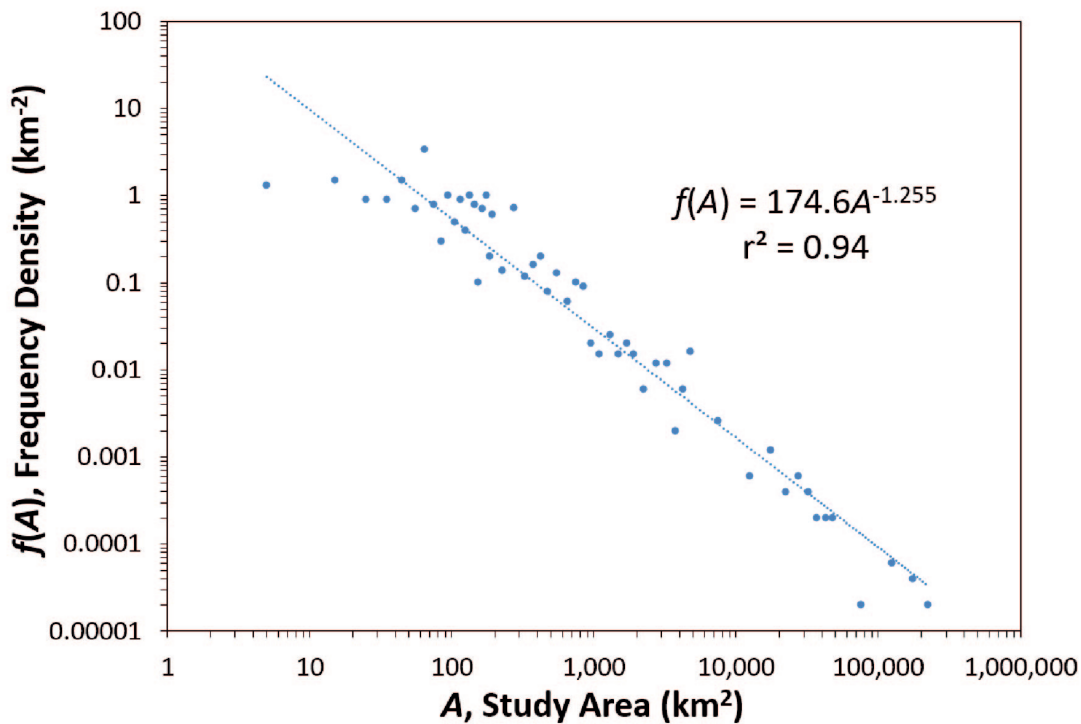
Figure 2.1: *Frequency density as a function of study areas (in km^2). This graph has been obtained from Malamud et al. (2014). Data serie 1983-2014.*

2.1 Irudia: Frekuentzia dentsitatea ikerketa eremuaren azaleraren funtzio bezala adierazita (km^2-tan). Grafiko hau Malamud *et al.* (2014) argitalpenetik lortu da. Datu seriea 1983-2014.

2.2 Zenbait erreferentziazko ikerketa

Lur labainketek gizartean sortzen duten eraina Brabbek argitaratutako "The World Landslide Problem" artikuluan maisuki azaldua geratu zen (Brabb, 1991). Bertan hitz hauek irakur daitezke:

"Landslides are generally more manageable and predictable than earthquakes, volcanic eruptions, and some storms, but only a few countries have taken advantage of this knowledge to reduce landslide hazards."

Eta handik urte gutxitara Soeters eta Van Westenek lehen berrikusketa sakona plazaratu zuten (Soeters & Van Westen, 1996), hain zuzen ere lur labainketen inguruko ezagutzari buruz hausnartu eta haien mehatxua eta lurraldearen kudeaketarako bide-orri bateratu bat definitzeko helburuarekin. Guzzeti-k, bide horretatik jarraituz, 20 urteko ibilbidean lortutako ezagutzak bildu zituen bere doktore tesian, "Dissertation on Landslide Hazard and Risk Assessment" (Guzzetti, 2006), eta Fell eta beste askoren lanaren ondorioz, 2008an, "Guidelines for landslide

susceptibility, hazard and risk zoning for land use planning" argitaratu zen (Fell *et al.*, 2008). Azken hori Europar Komisioak finantzatutako "Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies" dokumentuaren oinarri nagusia da, non lehen aldiz lur labainketen mehatxu eta arriskuari dagokienez jarraitu beharreko bide-orri bat eskaintzen den instituzio ofizial baten eskutik (Corominas & Mavrouli, 2011).

Bertan ematen diren gomendioen artean oinarrizkoena, ordea, lur labainketen gertaeren datu baseak gauzatzekoa da. Izan ere, edozein motatako analisietarako, iraganeko informazioa behar beharrezkoa da. Arlo horretan ahaleginak egiten ari dira mundu mailatik hasita (Petley, 2012; Petley *et al.*, 2005; EM-DATA, 2017), nazio nahiz eskualde mailara arte. Adibiderik esanguratsuena Italiako kasua da. IFFI proiektuaren bitartez (Trigila, 2007; Trigila *et al.*, 2010), urteak daramatzate lur labainketen gertaerak sistematikoki erregistratzen herrialde osoan, eta horrek ahalbidetu du alerta sistema aurreratu bat ezartzea horren guztiaren inguruan. Espainian ere, ALISSA proiektuaren bitartez, antzeko pausoak ematen ari dira (Hervás, 2014), nahiz eta oraingoz bibliografian oinarritutako datu base bat den, eta beraz, sistematikoki datuak biltzeko prozedura ezarri gabe dagoen. Katalunian ordea, beste adibide eredugarri bat aurki daiteke, non duela hamarkada batez geroztik, LLISCAT proiektuaren baitan, lur labainketen erregistroa sistematikoki gauzatzen ari den (i Planells, 2007). Hala ere, oraindik lan ugari dago egiteko arlo honetan Van den Eeckhautek eta Hervásek beren "State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk" argitalpenean aipatzen duten moduan (Van Den Eeckhaut & Hervás, 2012).

Izatez, azken urteotan lur labainketen suszeptibilitatea ikertzeko aplikazio zuzen ugari argitaratu dira, baina jarraian lan honekin erlazio zuzena daukaten erreferentzia esanguratsuenak bakarrik aipatuko dira. Esaterako, 2013an, ELSUS, Europa osoko lur labainketen suszeptibilitate mapa, plazaratu zen lehen aldiz 1:1M eskalan (Günther *et al.*, 2014). Eta horren bertsio eguneratua 2018an argitaratu da maparen erresoluzio maila 1000 m-tik 200 m-ra handituz besteak beste (Wilde *et al.*, 2018).

Bestalde, ikerlari batzuen ahaleginak suszeptibilitate mapa horiek sortzeko

metodologia espezifiko batzuetara bideratzen diren bitartean, metodo bat baino gehiago aplikatzeko aukerak ematen dituzten softwareak sortzen konzentratzen ari dira beste batzuk (Rossi *et al.*, 2010; Rossi & Reichenbach, 2016; Fabbri *et al.*, 2004). Kasu bakoitzean espezifikazio desberdinak aurki daitezke, baina antzematen denaren arabera, tendentzia, datuen bilketa sistematizatzeaz gain, haietatik sortutako suszeptibilitate mapak sistematizatzea ere badela ematen du.

Modeloak kalkulatzeko orduan eta modelo horiek mapa batean adierazteko garaian erabiliko diren unitate espazialei dagokienez, gaur egun, hainbat aukera daude eztabaidan. Izan ere lurraldea ikuspegi desberdinen arabera banandu daiteke: unitate topografikoak, unitate hidrologikoak, adibidez, malda unitateak edo unitate erregularrak erabiliz. Eta kasu bakoitzean emaitza nahiz emaitzen beraren interpretazioa desberdina izango da. Hausnarketa esanguratsuak aurki daitezke mapa unitate bakoitzaren erabilerak izan dezakeen eraginari buruz (Carrara *et al.*, 1995; Reichenbach *et al.*, 2018).

Tesi hau Gipuzkoako LHan gauzatu da (ikusi 4 atala). Alta, hau ez da inguru hauetan lur labainketen suszeptibilitatea eta mehatxua lantzen den lehen aldia. 80. hamarkadan Gipuzkoako Foru Aldundiak (GFA) lurraldearen ikerketa geomorfologikoa enkargatu zuen eta, horren ondorioz, "Gipuzkoako Mapa Geomorfologikoa" argitaratu zen 1:25000 eskalan (GFA, 1991). Orduko baliabideak erabiliz eta metodo heuristikoa aplikatuz, mapa haietan identifikatutako lur labainketa moten banaketa espaziala erakusten zen. Gaur egun, mapa haiek eskuragarri daude formatu digitalean Euskadiko Datu Geoespazialen Azpiegituran[1].

Beste aurrekari garrantzitsu batzuk Deba bailararen beheko aldean (Gipuzkoako mendebaldean) egindako ikerketak dira. Remondok, 1954 eta 1997 urteen arteko aireko argazkiak erabiliz eta landa lan bitarteko egiaztapenak eginez (Remondo, 2001), Deba ibaiaren behe ibilgua osatzen duten 4 udalerrietarako lur labainketen inbentario multi-tenporala gauzatu zuen, eta metodo estatistiko bibarianteak aplikatu zituen GISak erabiliz inguru haietako lehen lur labainketen suszeptibilitate mapa sortzeko. Ondoren, Bonacheak inbentario hori eguneratu eta metodo bera erabiliz (Bonachea, 2006), suszeptibilitate mapaz gain, lur labainketen mehatxu eta arrisku mapak garatu zituen. Eta azkenik, Felicísimok eta beste batzuk, inbentario

---

[1] www.geo.euskadi.eus

berbera aprobetxatuz, metodo estatistiko multibariante batzuen arteko konparazioa (*multiple logistic regression*, *multivariate adaptive regression splines*, *classification and regression trees* eta *maximum entropy*) aurkeztu zuten (Felicísimo *et al.*, 2013).

Gipuzkoako beste eremu batean, metodo heuristikoa aplikatu da lur labainketen mehatxua ebaluatzeko: Oiartzun ibaiaren erdi ibilguan hain zuzen (Etxeberria *et al.*, 2005).

2007an GFAk lurralde historiko osoa bere gain hartzen zuen lehen lur labainketen suszeptibilitate mapa argitaratu zuen (GFA, 2007). Bertan, 80. hamarkadan bildutako lur labainketen kokalekuen informazioa erabiliz eta *discriminant analysis* metodo multibariantea aplikatuz, 1:25000 eskalako mapa digital sorta aurkeztu zen 10x10 m-ko pixel erresoluzioarekin.

Beste alde batetik, lur labainketen suszeptibilitate ikerketan gertatzen den bezala, labainketak gertatzeko prezipitazio atalaseen ikerketak ere baditu aurrekari garrantzitsu batzuk.

Wieczorekek ireki zuen bidea (Wieczorek, 1987), detritu fluxuen (*debris flow*) gertaeren eta euriteen intentsitatearen eta iraupenaren arteko erlazioa definitu zuenean metodo enpirikoa erabiliz. Ideia horri buruz sakonago hausnartzen dute beranduago Wieczorek eta Gladek "Climatic Factors and Debris Flows" argitalpenean (Wieczorek & Glade, 2005).

Aplikazio zehatzetara jotzen badugu, adibide bikainak aurki daitezke nazioarteko ikerketetan (Cuesta *et al.*, 1999; Glade *et al.*, 2000; Zêzere & Rodrigues, 2002; Jakob & Weatherly, 2003; Li *et al.*, 2011; Bui *et al.*, 2013; Zêzere *et al.*, 2015; Vaz *et al.*, 2018) baina batez ere Portugalen eta Italian, gai hau gehien landu den herrialdetan. Brunettiren lan taldeak metodo Bayesiarra eta Frekuentista aplikatu zituzten Abruzzo eskualdean, baina herrialde osorako aplikatzeko aukera azpimarratu zuten (Brunetti *et al.*, 2010). Peruccaccik eta bere kideek ikerketa eremua zabaldu zuten eta, euri intentsitateak erabili beharrean, akumulatutako prezipitazioa erabili zuten parametro bezala (Peruccacci *et al.*, 2012). Beste kasu batzuetan modelo hidrologiko bat proposatu da, algoritmo automatiko baten bitartez (Terranova *et al.*, 2015), lur labainketak sorrarazteko prezipitazio atalaseak definitzeko; bide horretatik jarraituz, Melilloren lan taldeak lur labainketak sorrarazten dituzten eurite motak identifikatu eta klasifikatzeko algoritmo bat proposatu zuen (Melillo *et al.*, 2015).

Azkenik, baliabide hauetan guztietan oinarrituta, prezipitazio atalaseak modu automatizatuan kalkulatzeko gai den algoritmo bat proposatu zuten (Melillo *et al.*, 2016; Peruccacci *et al.*, 2017). Bestalde, kontzeptualki, lurzoruaren ur saturazio maila kontuan hartzea egokiagoa dela proposatzen duten argitalpenak aurki daitezke (Bogaard & Greco, 2018) baita, kontzeptu horren aplikazio metodologikoak ere (Valenzuela *et al.*, 2018).

Gipuzkoako kasura bueltatuz, prezpitazioen eta lur labainketen arteko erlazioa ezer gutxi landu den gai bat da, momentuz, hasierako ezaugarriketa bat bakarrik argitaratuz (Bonachea *et al.*, 2017).

2.3   Motivation

According to the general description of the state of the art presented in the previous sections, the investigations displayed in the following chapters were carried out with the motivation of filling some gaps on the field of landslide susceptibility modelling by statistical methods as well as for landslide causing precipitation threshold calculation. Thereby, the current thesis can be considered in two levels: the production of new or updated results for a territory related to landslide susceptibility maps and precipitation thresholds; and the development of innovative methods and results that contributes to progress toward a definitive methodological approach within this scientific field.

Concerning the first level, the motivation for the work presented here comes from the following gaps of knowledge: i) the need for an updated database for landslide analysis in Gipuzkoa Province; ii) the necessity to produce, compare and check new landslide susceptibility maps carried out following innovative approaches and in different scales, in order to improve the previously existing maps (GFA, 2007); and iii) the lack of information on the landslide responsible precipitation thresholds.

Furthermore, starting from the mentioned gaps of knowledge for landslide analysis in our study area, divers methodological questions were raised related to the susceptibility modelling. To begin, there is still no clear the most appropriate way to work with both, categorical and continuous, explanatory variables. In addition, the selection of these explanatory variables to be included in the statistical models

is usually too subjective and, as a consequence, the objectivity and reproducibility of the approach is reduced. Also other aspects such as, the uncertainty that results from the aleatory sampling of stable (or no-landslide) places; or the advantages or drawbacks related to the cartographic mapping unit applied to the models, needs to be investigated more in depth.

So, the search to solve all this gaps of knowledge resulted in the current thesis, which involves three sequentially designed investigations.

Bibliografia

Beguería, S.: Validation and evaluation of predictive models in hazard assessment and risk management, Natural Hazards, 37, 315–329, 2006.

Bogaard, T. & Greco, R.: Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Natural Hazards and Earth System Sciences, 18, 31–39, 2018.

Bonachea, J.: Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debidos a procesos geomorfológicos, Ph.D. thesis, Universidad de Cantabria, Santander, 2006.

Bonachea, J., Remondo, J., Rivas, V., González Díez, A., & Sánchez-Espeso, J.: Modelización de la peligrosidad de deslizamientos para diferentes escenarios empíricos en el Cantábrico oriental, in: IX Simposio sobre Taludes y Laderas Inestables, edited by Alonso, E., Corominas, J., & Hürlimann, M., pp. 407–418, International Centre for Numerical Methods in Engineering (CIMNE), Santander, 2017.

Brabb, E. E.: The World Landslide Problem, Episoded, 14, 52–61, 1991.

Brabb, E. E., Bonilla, M. G., & Pampeyan, E. H.: Landslide susceptibility in San Mateo County, California: US Geol. Survey Misc, Field Studies Map MF-360, scale 1:62500, 1972.

Brand, E. W. & Bonnard, C.: Special lecture: landslide risk assessment in Hong Kong, in: V International Symposium on Landslides, vol. 2, pp. 1059–1074, Lausanne, 1988.

Brunetti, M., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F.: Rainfall thresholds for the possible occurrence of landslides in Italy, Natural Hazards and Earth System Science, 10, 447–458, 2010.

Bui, D. T., Pradhan, B., Lofman, O., Revhaug, I., & Dick, Ø. B.: Regional prediction of landslide hazard using probability analysis of intense rainfall in the Hoa Binh province, Vietnam, Natural Hazards, 66, 707–730, 2013.

Cardinali, M., Reichenbach, P., Guzzetti, F., Ardizzone, F., Antonini, G., Galli, M., Cacciano, M., Castellani, M., & Salvati, P.: A geomorphological approach to the estimation of landslide hazards and risks in Umbria, Central Italy, Natural Hazards and Earth System Science, 2, 57–72, 2002.

Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P.: GIS technology in mapping landslide hazard, in: Geographical Information Systems in Assessing Natural Hazards, edited by Carrara, A. & Guzzetti, F., pp. 135–175, Springer, 1995.

Chacón, J., Irigaray, C., Fernandez, T., & El Hamdouni, R.: Engineering Geology maps: landslides and geographical information systems, Bulletin of Engineering Geology and the Environment, 65, 341–411, 2006.

Corominas, J. & Mavrouli, O. C.: Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies, Tech. rep., SafeLand. 7th Framework Programme Cooperation Theme 6 Environment (including climate change) Sub-Activity 6.1.3 Natural Hazards, 2011.

Cuesta, M. J. D., Sánchez, M. J., & García, A. R.: Press archives as temporal records of landslides in the North of Spain: relationships between rainfall and instability slope events, Geomorphology, 30, 125–132, 1999.

EM-DATA: The Emergency Events Database - Université catholique de Louvain (UCL) -, URL `www.emdat.be`, Brussels, Belgium, 2017.

Etxeberria, P., Edeso, J. M., & Brazaola, A.: Evaluación espacial de los peligros naturales en el valle de Oiartzun (Guipuzkoa), Munibe Ciencias Naturales. Natur Zientziak, 1, 5–20, 2005.

Fabbri, A., Chung, C. J., & Jang, D. H.: A software approach to spatial predictions of natural hazards and consequent risks, WIT Transactions on Ecology and the Environment, 77, 2004.

Felicísimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10, 175–189, 2013.

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, Engineering Geology, 102, 85–98, 2008.

GFA: Estudio geomorfológico y edafológico de Gipuzkoa, Tech. rep., Gipuzkoako Foru Aldundia, 1991.

GFA: Elaboración de tres modelos de predicción de riesgos naturales de incendios, deslizamientos e inundaciones en el Territorio Histórico de Gipuzkoa, Tech. rep., Gipuzkoako Foru Aldundia, 2007.

Glade, T., Crozier, M., & Smith, P.: Applying probability determination to refine landslide-triggering rainfall thresholds using an empirical antecedent daily rainfall model, Pure and Applied Geophysics, 157, 1059–1079, 2000.

Godt, J., Baum, R., Savage, W., Salciarini, D., Schulz, W., & Harp, E.: Transient deterministic shallow landslide modeling: requirements for susceptibility and hazard assessments in a GIS framework, Engineering Geology, 102, 214–226, 2008.

Gökceoglu, C. & Aksoy, H.: Landslide susceptibility mapping of the slopes in the residual soils of the Mengen region (Turkey) by deterministic stability analyses and image processing techniques, Engineering Geology, 44, 147–161, 1996.

Günther, A., Van Den Eeckhaut, M., Malet, J. P., Reichenbach, P., & Hervás, J.: Climate-physiographically differentiated Pan-European landslide susceptibility assessment using spatial multi-criteria evaluation and transnational landslide information, Geomorphology, 224, 69–85, 2014.

Guzzetti, F.: Disertation on Landslide Hazard and Risk Assessment, Ph.D. thesis, University of Bonn, 2006.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P.: Rainfall thresholds for the initiation of landslides in central and southern Europe, Meteorology and Atmospheric Physics, 98, 239–267, 2007.

Hammond, C. J., Prellwitz, R. W., & Miller, S. M.: Landslide hazard assessment using Monte Carlo simulation, in: 6th international symposium on landslides, Christchurch, New Zealand, vol. 2, pp. 251–294, 1992.

Hardingham, A., Ditchfield, C., Ho, K., & Smallwood, A.: Quantitative risk assessment of landslides. A case history from Hong Kong, in: Seminar on Geotechnical Risk Management, Geotechnical Division, edited by of Engineers, H. K. I., pp. 145–152, 1998.

Hervás, J.: ALISSA: Abridged Landslide Inventory of Spain for synoptic Susceptibility Assessment, in: EGU General Assembly Conference Abstracts, vol. 16, 2014.

Huabin, W., Gangjun, L., Weiya, X., & Gonghui, W.: GIS-based landslide hazard assessment: an overview, Progress in Physical Geography, 29, 548–567, 2005.

Humbert, M.: Les mouvements de terrains: Principes de réalisation d'une carte prévisionnelle dans les Alpes, pp. 13–28, 1970.

Humbert, M.: La cartographie zermos-modalites d'etablissement des cartes des zones exposees a des risques lies aux mouvements du sol et du sous-sol, Bulletin du BRGM, Section III, pp. 5–8, 1977.

i Planells, L. M.: Determinació de llindars de pluja desencadenants d'esllavissades a Catalunya, Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, 2007.

Jakob, M. & Weatherly, H.: A hydroclimatic threshold for landslide initiation on the North Shore Mountains of Vancouver, British Columbia, Geomorphology, 54, 137–156, 2003.

Kanungo, D., Arora, M., Sarkar, S., & Gupta, R.: Landslide Susceptibility Zonation (LSZ) Mapping–A Review., Journal of South Asia Disaster Studies, 2, 81–105, 2009.

Kienholz, H.: Maps of geomorphology and natural hazards of Grindelwald, Switzerland: scale 1: 10000, Arctic and Alpine Research, pp. 169–184, 1978.

Li, C., Ma, T., Zhu, X., & Li, W.: The power–law relationship between landslide occurrence and rainfall level, Geomorphology, 130, 221–229, 2011.

Malamud, B. D., Reichenbach, P., Rossi, M., & Mihir, M.: D6.3 Report on standards for landslide susceptibility modelling and terrain zonations, Tech. rep., LAMPRE Project, Seventh Framework Program. European Comission, 2014.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., & Guzzetti, F.: An algorithm for the objective reconstruction of rainfall events responsible for landslides, Landslides, 12, 311–320, 2015.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., & Guzzetti, F.: Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events, Landslides, 13, 165–172, 2016.

Mulder, H. & Van Asch, T.: A stochastical approach to landslide hazard determination in a forested area, in: V International Symposium on Landslides, pp. 1207–1210, A.A. Balkema Publisher, 1988.

Mulder, H. F. H. M.: Assessment of landslide hazard, Ph.D. thesis, University of Utrecht, Netherlands, 150 pp, 1991.

Nilsen, T. H.: Relative slope stability and land-use planning in the San Francisco Bay region, California, vol. 944, US Govt. Print. Off., 1979.

Okimura, T. & Kawatani, T.: Mapping of the potential surface-failure sites on granite mountain slopes, International Geomorphology, 1, 121–138, 1986.

Pardeshi, S. D., Autade, S. E., & Pardeshi, S. S.: Landslide hazard assessment: recent trends and techniques, SpringerPlus, 2, 1–11, 2013.

Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., & Guzzetti, F.: Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy, Geomorphology, 139, 79–90, 2012.

Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., & Guzzetti, F.: Rainfall thresholds for possible landslide occurrence in Italy, Geomorphology, 290, 39–57, 2017.

Petley, D.: Global patterns of loss of life from landslides, Geology, 40, 927–930, 2012.

Petley, D., Dunning, S., Rosser, N., & Hungr, O.: The analysis of global landslide risk through the creation of a database of worldwide landslide fatalities, Landslide Risk Management. Balkema, Amsterdam, pp. 367–374, 2005.

Ramos-Cañón, A. M., Trujillo-Vela, M. G., & Prada-Sarmiento, L. F.: Niveles umbrales de lluvia que generan deslizamientos: una revisión crítica, Ciencia e Ingeniería Neogranadina, 25, 61–80, 2015.

Reichenbach, P., Rossi, M., Malamud, B., Mihir, M., & Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60–91, 2018.

Remondo, J.: Elaboración y validación de mapas de susceptibilidad de deslizamientos mediante técnicas de análisis espacial, Ph.D. thesis, Universidad de Oviedo, Oviedo, 2001.

Rossi, M. & Reichenbach, P.: LAND-SE: a software for landslide statistically-based susceptibility zonation, Version 1.0, Geosci. Model Dev., pp. 3533–3543, 2016.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C., & Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, Geomorphology, 114, 129–142, 2010.

Soeters, R. & Van Westen, C.: Slope stability recognition analysis and zonation, in: Landslides: Investigation and Mitigation, edited by Turner, A. K. & Jayaprakash, G., 247, pp. 129–177, Transportation Research Board Special Report, Washington D.C., 1996.

Terranova, O. G., Gariano, S. L., Iaquinta, P., & Iovine, G. G.: GA SAKe: forecasting landslide activations by a genetic-algorithms-based hydrological model, Geoscientific Model Development, 8, 1955–1978, 2015.

Trigila, A.: Rapporto sulle frane in Italia: il Progetto IFFI: metodologia, risultati e rapporti regionali, APAT, 2007.

Trigila, A., Iadanza, C., & Spizzichino, D.: Quality assessment of the Italian Landslide Inventory using GIS processing, Landslides, 7, 455–470, 2010.

Valenzuela, P., Domínguez-Cuesta, M. J., García, M. A. M., & Jiménez-Sánchez, M.: Rainfall thresholds for the triggering of landslides considering previous soil moisture conditions (Asturias, NW Spain), Landslides, 15, 273–282, 2018.

Van Asch, T. W., Malet, J. P., van Beek, L. P., & Amitrano, D.: Techniques, issues and advances in numerical modelling of landslide hazard, Bulletin de la Société Géologique de France, 178, 65–88, 2007.

Van Den Eeckhaut, M. & Hervás, J.: State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk, Geomorphology, 139, 545–558, 2012.

Van Westen, C., Van Asch, T. W., & Soeters, R.: Landslide hazard and risk zonation. Why is it still so difficult?, Bulletin of Engineering Geology and the Environment, 65, 167–184, 2006.

Van Westen, C. J.: Application of geographic information systems to landslide hazard zonation, Ph.D. thesis, Delft University of Technology, 1993.

Van Westen, C. J.: The modelling of landslide hazards using GIS, Surveys in Geophysics, 21, 241–255, 2000.

Van Westen, C. J., Rengers, N., Terlien, M., & Soeters, R.: Prediction of the occurrence of slope instability phenomenal through GIS-based hazard zonation, Geologische Rundschau, 86, 404–414, 1997.

Van Westen, C. J., Castellanos, E., & Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview, Engineering Geology, 102, 112–131, 2008.

Vaz, T., Zêzere, J. L., Pereira, S., Oliveira, S. C., Garcia, R. A., & Quaresma, I.: Regional rainfall thresholds for landslide occurrence using a centenary database, Natural Hazards and Earth System Science, 18, 1037–1054, 2018.

Ward, T. J., Li, R. M., & Simons, D. B.: Mapping landslide hazards in forest watersheds, Journal of Geotechnical and Geoenvironmental Engineering, 108, 319–324, 1982.

Wieczorek, G. F.: Effect of rainfall intensity and duration on debris flows in central Santa Cruz Mountains, California, Reviews in Engineering Geology, 7, 93–104, 1987.

Wieczorek, G. F. & Glade, T.: Climatic factors influencing occurrence of debris flows, in: Debris-flow hazards and related phenomena, edited by Jakob, M., Hungr, O., & Jakob, D. M., pp. 325–362, Springer, 2005.

Wilde, M., Günther, A., Reichenbach, P., Malet, J. P., & Hervás, J.: Pan-European landslide susceptibility mapping: ELSUS Version 2, Journal of Maps, 14, 97–104, 2018.

Wong, H. & Ho, K.: Overview of risk of old man-made slopes and retaining walls in Hong Kong, in: Seminar on Slope Engineering in Hong Kong, pp. 193–200, A.A. Balkema Publisher, 1998.

Wong, H., Ho, K., & Chan, Y.: Assessment of consequence of landslides, in: International Workshop on Landslide Risk Assessment, edited by Cruden, D. & Fell, R., pp. 111–149, 1997.

Zêzere, J. & Rodrigues, M.: Rainfall thresholds for landsliding in Lisbon Area (Portugal), in: Landslides: Proceedings of the First Europpean Conference on Landslides, edited by Rybar, J., Stemberk, J., & Wagner, P., pp. 333–338, AA Balkema: Lisse, Prague, Czech Republic, 2002.

Zêzere, J., Vaz, T., Pereira, S., Oliveira, S., Marques, R., & Garcia, R.: Rainfall thresholds for landslide activity in Portugal: a state of the art, Environmental Earth Sciences, 73, 2917–2936, 2015.

Chapter 3

Hypothesis and objectives

3.1 Hypothesis

The possibility of establishing significant and quantifiable relationships between different environmental and atmospheric factors associated with the landslide occurrence is something already accepted and demonstrated in several scientific studies (see chapter 2). It is also well known that an enormous variety of methods exist (above all statistical methods), which permit to carry out a landslide susceptibility and hazard analysis, and, it is assumed that the choice between one or another presents particular advantages or disadvantages against the rest. But, no matter the statistical method applied, there are always some common steps. So, the general hypothesis of this work maintains that **there are some crucial decisions during the evaluation process of the landslide susceptibility and hazard in a region, that can affect directly the results**. And this statement can be ordered as follow:

- The landslide inventory, and moreover, the type of data represented in the landslide inventory affects the results of the analysis.

- The only usage of statistical indicators does not completely ascertain the suitability of an explanatory variable for landslide modelling.

- The choice of the mapping unit is relevant to mitigate some uncertainties related to the input data when modelling landslide susceptibility.

- The usage of direct field work based landslide inventories in statistical landslide susceptibility assessments can introduce considerable uncertainties, unless an specific treatment is applied.

  In addition, in order to advance toward a landslide hazard assessment in the experimental study area, further experiments were applied under the following hypothesis:

- Landslide occurrence is directly related to certain type of rainfalls, and this relation can be expressed by a landslide responsible precipitation threshold.

3.2   Objectives

The main objective of the thesis is **to assess landslide susceptibility at regional scale, using a justified and scientifically supported method, and incorporating precipitation thresholds responsible for landslide occurrence**. Indeed, the development of a landslide susceptibility map as well as the calculation of precipitation thresholds responsible of landslide occurrence are full of critical steps in which a decision has to be taken.

For this reason, and following the previously cited hypothesis order, below is shown the specific objectives list, each of them related with one of the most critical steps during the statistical analysis:

- To test different landslide inventories for landslide analysis in a regional scale in order to detect the most suitable features necessary to run susceptibility models.

- To experiment with the available spatial digital layers, for their usage as independent explanatory variables in landslide susceptibility analysis, as well as to test different ways for selecting, in an objective way, only the most convenient to build the model.

- To observe and recognize the advantages and drawbacks of different mapping units in landslide susceptibility mapping.

- To prove that during the calibration, the restriction of the area in which no-landslide data are sampled to the visible portion of the territory during the survey (called Effective Surveyed Area), in place to the entire area under investigation (called Whole Area), enhances the quality of the model, in the cases where the landslide inventory was carried out by direct field work.

- To detect relations between the inventoried instabilities and the rainfall events for the calculation of landslides responsible precipitation thresholds in Gipuzkoa Province.

Moreover, from the point of view of reproducibility, it was fixed an additional objective related to the implementation of new and updated technologies. So, the

usage of digital spatial data and statistical software's was prioritized in order to ease the application of the presented approach in the future or in any other study area.

Chapter 4

Ikerketa eremua

Nahiz eta lan honetan aurkeztuko diren kontzeptualizazioak ez diren esparru geografiko bakar batera mugatzen, aldez aurretik formulatutako hipotesi eta helburuak lortu nahian aurkeztuko diren lan-metodo eta emaitzak ikerketa eremu zehatz batean gauzatu dira, Gipuzkoako Lurralde Historikoan. Lurralde hau aukeratzeko arrazoiak, batez ere, ondorengo hauek izan dira:

(i) Eguneroko giza jardueretan eragin zuzena suposatzen duten mota desberdine-tako lur labainketen gertakariak egotea. Hots, errepide mozketak, etxebizitzen ustutzeak, etab.

(ii) Lur labainketen suszeptibilitatearekin erlazionatutako ikerketen aurrekariak egotea.

(iii) Lurralde osoa, estaltzen duen oinarrizko informazio espazial eta tematikoaren eskuragarritasuna formatu digitalean.

(iv) Ikerketa honetatik aterako diren ondorioak kontutan hartu eta gomendioak aplikatzeko gai den entitate administratibo propioaren izaera.

(v) Landa laneko jarduera erraztuko duen kokapen geografikoa.

Beraz, jarraian Gipuzkoako LH ezaugarritzen duten, eta lur labainketa prozesuekin erlazionatuta dauden, arlo nagusiak deskribatuko dira, ondorengo ataletan aurkeztuko diren emaitzen testuinguru espazial bezala.

4.1 Kokalekua

Gipuzkoako LH Iberiar penintsularen iparraldean kokatuta dago Bizkaiko Golkoaren hego-ekialdeko erpinean (4.1 Irud.). Piriniar mendikatearen mendebaldeko ertzak mugatzen du ekialdetik eta Aralar eta Aizkorri-Aratz mendi zerrak hegoaldean. Mendebaldean berriz, Deba eta Artibai ibai arroen arteko mendiek definitzen dute ondoan kokatuta dagoen Bizkaiko LHrekiko muga. Azkenik, iparraldean 66 km-ko kostaldea zabaltzen da Kantauri Itsaso aldera. 1980 km^2 inguruko azalerarekin, ondorengo erpin koordenatuek definitzen duten lauki zuzenaren baitan kokatua geratuko litzateke:

$X_{min} = 532269; X_{max} = 603631; Y_{min} = 4749441; Y_{max} = 4805769.$

88 udalerriz osatuta dagoen 717,832[1] biztanleko eremu administratibo propioa da. Bere kokalekua kontuan izanda, izaera estrategikoa dauka Iberiar penintsula eta Europako beste herrialdeen arteko pasabide bezala (Urrestarazu & Galdos, 2008). Izan ere, Pirinioekin alderatuta, bertan aurkitzen diren erliebeek erakusten dituzten altuera apalagoek (0 m eta 1550 m artean) bailaren arteko komunikazioa erraztu egiten dute, berez, ingurune menditsu bat izan arren.
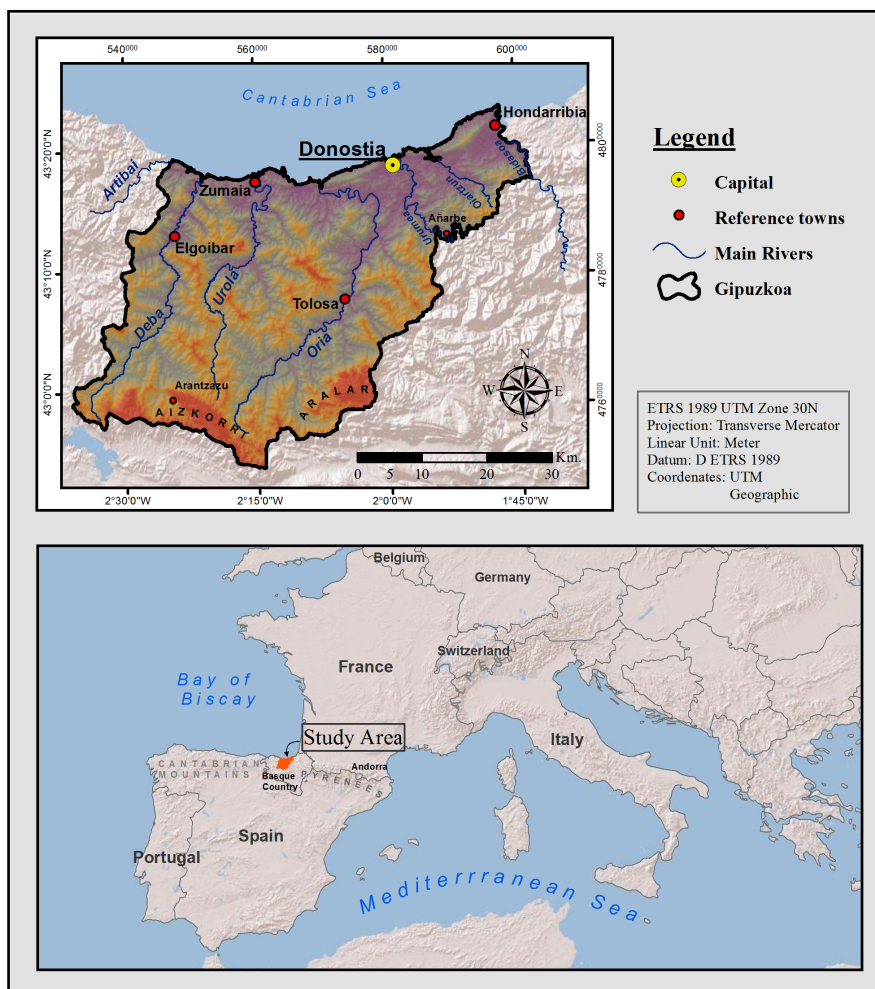
Figure 4.1: *Location of the Gipuzkoa Province.*
4.1 Irudia: Gipuzkoako LHren kokalekua.

4.2 Ezaugarri geologikoak

Ikerketa eremua Piriniar mendi katearen mendebaldeko luzapena den heinean, lurraldearen gehiengoa Euskal-Kantauriar Arroa izeneko unitate geologikoaren

---

[1]2016ko datuak www.datuak.net-en arabera

baitan kokatzen da, Euskal Arkua deritzon azpi unitatean zehatzago izateko (Barnolas & Pujalte, 2004; Ábalos, 2016). Orogenesi Hertziniar edo Variskarraren ondoren eremu guzti honek zartatze prozesu bat jasan zuen norabide ezberdinetako urratzeak emanez. Besteak beste, Iruñako (IIE-HHM norabidean) eta Bilboko (IIM-HHE norabidean) failak. Bi urratze hauek Euskal Arkuaren izaera eta bilakaera baldintzatu dute, ondorengo Mesozoikoko estentsio prozesuan eta alpetar konpresioan (Martínez-Torres, 1997; Pedreira, 2004)(4.2 Irud.).
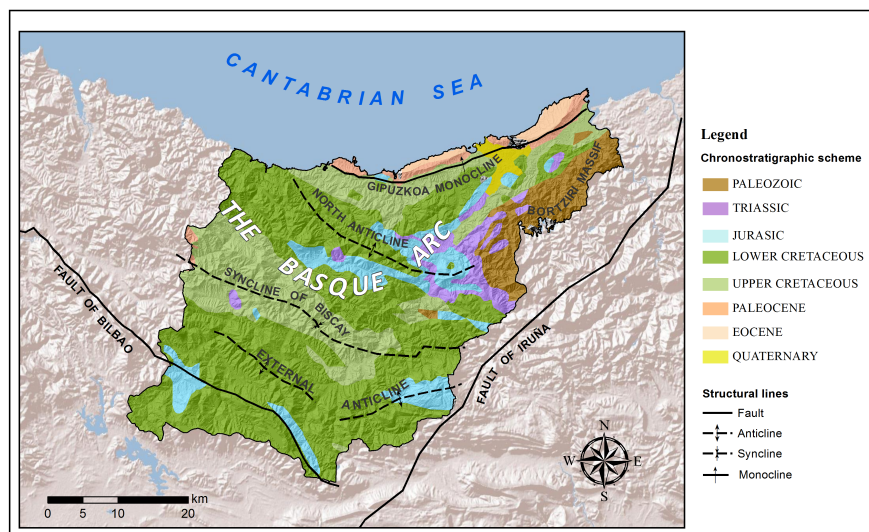
Figure 4.2: *Chronologic units and principle structural lines of Gipuzkoa. Modified on the basis of EVE (1991) and Ábalos (2016).*

4.2 Irudia: Gipuzkoako unitate kronologikoak eta lerro estrukturalen antolaketa nagusia. EVEk egindako mapan eta Ábalosen proposamenean oinarritua (EVE, 1991; Ábalos, 2016).

Bertako ezaugarri nagusia egitura erregionalen arku forma da, iparralderantz orientatutako geruzaz, labainketa faila azpi bertikalez eta toles handiz osatua dagoena (Barnolas & Pujalte, 2004). Egitura nagusiak, iparraldetik hegoaldera antolatuta, Gipuzkoako Monoklinala, Iparraldeko Antiklinala, Bizkaiko Sinklinala eta Kanpoko Antiklinala dira (Ábalos, 2016). Laburbilduz, Euskal Arkuaren ezaugarri nagusiak ondorengo hauek dira (Barnolas & Pujalte, 2004): (i) Kretazeoan zehar gertatutako magmatismo prozesuen existentzia; (ii) batez ere, izaera termikoko metamorfismoen lekukoak diren azaleratzeak material Mesozoikoetan; (iii) itsas zabaleko baldintzetan izandako potentzia handiko higakinen suzesioak Jurasiko, Kretaziko eta Behe Paleogenoan; (iv) Euskal Kantauriar Arroan metatutako Flysch motako turbiditen metakin garrantzitsuak.

Euskal Arkuaren ekialdean Pirinieo Axialari dagokion unitate geologikoa ere

aurki daiteke; Bortziriko Mazizoa (González & Serrano, 1996). Piriniar mendikatea sortu zuten mugimendu orogenikoen bultzadak, Deboniar garaian sortutako plutoiaren azaleramendua eragin zuen, gaur egun 83 km²-ko erliebe granitiko menditsua eratuz ipar-ekialdean (Denèle *et al.*, 2012).
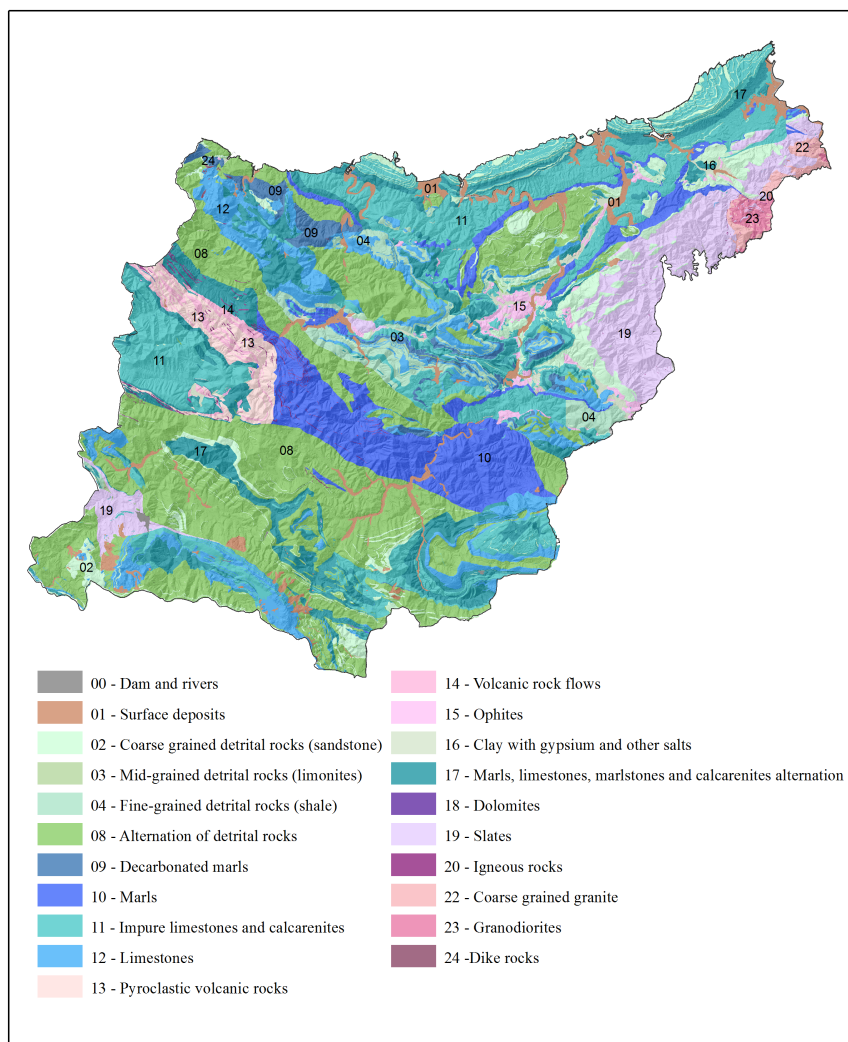
Figure 4.3: *Lithological map of Gipuzkoa (Euskadiko DEA, 2014).*
4.3 Irudia: Gipuzkoako mapa litologikoa (Euskadiko DEA, 2014).

Mesozoiko eta Zenozoikoan izandako bilakaera geologikoaren emaitza dira gaur egun lurralde historikoan aurki daitezkeen material litologikoak (4.3 Irud.). Hegoaldeko mendietan (Aralar eta Aizkorrin) tuparri (marls), kareharri (limestone), eta kalkarenitak agertzen dira, zenbait arroka detritikoren alternantziarekin batera. Lurraldearen erdialdean Piriniar orogenesiaren ondorioz deformatutako deskarbonatatutako tuparriak (decarbonated marls), kareharri inpuruak (impure limestones), ofitak (ophites), buztinak (clay), igeltsua (gypsium) eta ale lodiko

arroka detritikoak tartekatuz doaz hainbat toles eta zamalkadura osatuz. Ekialdean, ukimenezko metamorfismoaren ondorioz sortutako arbela (slate) nabari da, azaleramendu paleozoikoen (granitoen) inguruan. Kosta lerroaren paraleloan, tuparriak (marls), kareharri inpuruak (impure limestones), kalkarenitak (calcarenites), kareharri margatsuak (marly limestones) eta ale lodiko arroka detrikikoak (hareharriak gehien bat) tartekatuz doaz. Ekialdeko kostaldea (Hondarribia eta Getaria artean) hareharriz eta kareharri hareatsuz osatuta dago; mendebaldeko ertzean (Zumaia eta Mutriku artean) flysch formakuntzaren azaleramendua aurkitzen den bitartean. Kuaternarioko azaleko metaketak (surface deposits) ibai nagusien aldeetan eta itsasoratze eremuetan aurkitzen dira (Campos & García-Dueñas, 1972; Campos *et al.*, 1983).

4.3 Ezaugarri klimatikoak

Isurialde atlantikoan kokatuta, Gipuzkoako LH klima ozeanikoaren baitan defini daiteke (Urrestarazu & Galdos, 2008). Izan ere, ipar-mendebaldetik iristen diren aire masak, ozeanoko ur epelen gainetik igarotzean hezetasun handiarekin iristen dira, lur barneratzean ur hori prezipitatuz. Gainera, Kantauri itsasoarekin duen gertutasunak tenperatura moderatuak eragiten ditu, urteko batez besteko tenperaturak 10°C-14°C artean mantenduz (Ikusi Donostiako klimograma 4.4 Irud.).

Kostaldetik gertuko orografiak jasotzen dituen fronte eta aire masa hezeen talkak eraginda, urteko batez besteko prezipitazioa 1000 eta 2200 mm artekoa da (Uriarte, 1996). Euri hori guztia urtean zeharreko bi periodo maximoetan banatzen da (4.4 Irud.): azaro eta urtarril artean urteko prezipitazioen %34 jasotzen da, eta apirilean %10 (González-Hidalgo *et al.*, 2011; Fdez-Arroyabe & Martin-Vide, 2012). Orokorrean, hiru motako euriteak dira ohikoenak (Corominas, 2006; Eusko Jaurlaritza, 2015; Ormaetxea & Sáenz de Olazagoitia, 2017): i) Intentsitate altuko eta iraupen motzeko ekaitzak, 10 l/m²-ko neurketak utz ditzazketenak; ii) Ekaitz estazionarionak, intentsitate handia izatez gain, ordu bat baino gehiagoko iraupena dutena eta iii) Fronteek ekarritako euriteak, udazken eta neguan intentsitate ertaineko prezipitazioak uzten dituztenak, baina luzaroan iraun dezaketenak.

Lurraldean zehar, prezipitazioari dagokionez, desberdintasun esanguratsu batzuk
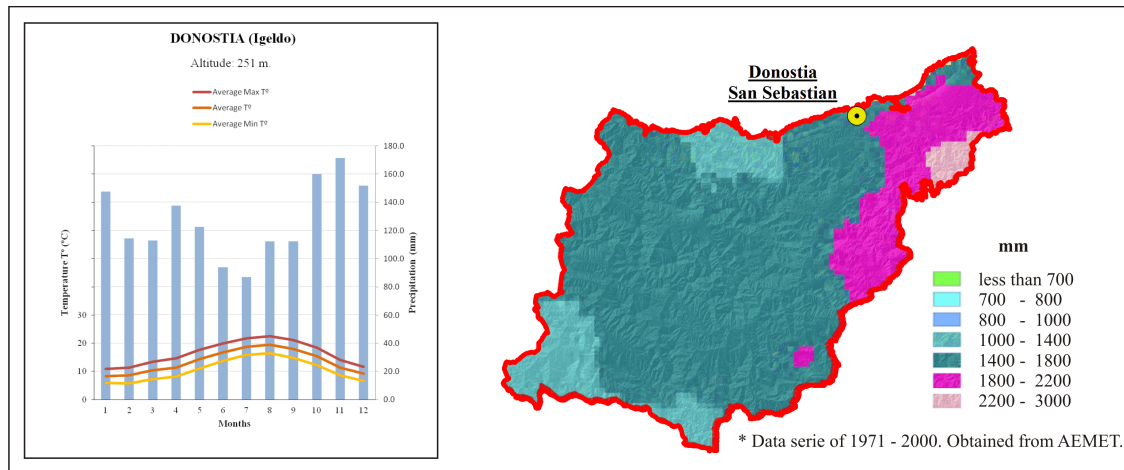
Figure 4.4: *Climographs for the capital of Gipuzkoa Province (data serie 1981-2013) and annual mean precipitation map. All the data were obtained from AEMET.*

4.4 Irudia: Gipuzkoako LHko hiriburuko klimograma (datu seriea 1981-2013) eta batez besteko urteko prezipitazioen mapa. Datu guztiak AEMETetik lortu dira.

antzematen dira (4.4 Irud.). Kostaldean urte osoan zehar prezipitazio ugaria jasotzen da beherakada lauso batekin udako hilabeteetan zehar. Kostaldetik gertuko mendi inguruetan antzeko banaketa dute prezipitazioek, baina metatutako euria nabarmen handitzen da kostaldetik sartzen diren aire masek, hego-ekialdeko norantzan, mendiak gainditzean deskargatutako euriak direla eta.

Horretaz gain, urtean, batez bestean, 200 egunetan prezipitatzen du; eta beraz, eguneroko batez besteko tasa 7.5 mm-koa da. Gainera, prezipitariorik gabeko epealdi luzeak ez ohikoak dira, zeinak 50 urte inguruko itzulera denbora bait duten (Borja & Collins, 2004).

4.4   Ezaugarri hidrografiko eta hidrologikoak

Gipuzkoako LHren erliebea, batez ere, ibai dinamikak markatutako prozesu geomorfologikoen ondorioa da (CGS *et al.*, 1991). Drainatze sarea bereziki dentsoa duten sei arro hidrografikotan banonduta aurkitzen da ikerketa eremua (4.5 Irud.). Arro horietan, 1550 m-ko altueratik (hegoaldeko mendi zerretan) itsas maila arteko trantsizio azkarra gertatzen da. Prezipitazioek sortutako grabitatezko prozesuek, higadura hidrikoak eta ibai higaduraren eraginez erliebe erabat aldapatsua eratu da, izan ere, lurralde osoko azaleraren erdia baino gehiagok (%55) 15°tik gorako malda

dauka. Ezaugarri hauen ondorioz isurketa koefiziente altuak (0.5 eta 0.7 artean) eta kontzentrazio denbora motzak ematen dira kasu guztietan (Ibisate *et al.*, 2000).

Kantauri itsasora urak garraiatzen dituzten 64 eta 900 km² arteko arroez hitz egiten ari gara (4.1 Tau.), Ebro arrora isurtzen duen hegoaldeko 19 km² eremu txiki baten kasuan izan ezik. Ur emariek erregimen plubial-ozeanikoa erakusten dute (Ibisate *et al.*, 2000) neguko emari altuekin eta udako beherakada esanguratsuekin, nahiz eta ibai nagusiek urte osoan ura garraiatzen duten.
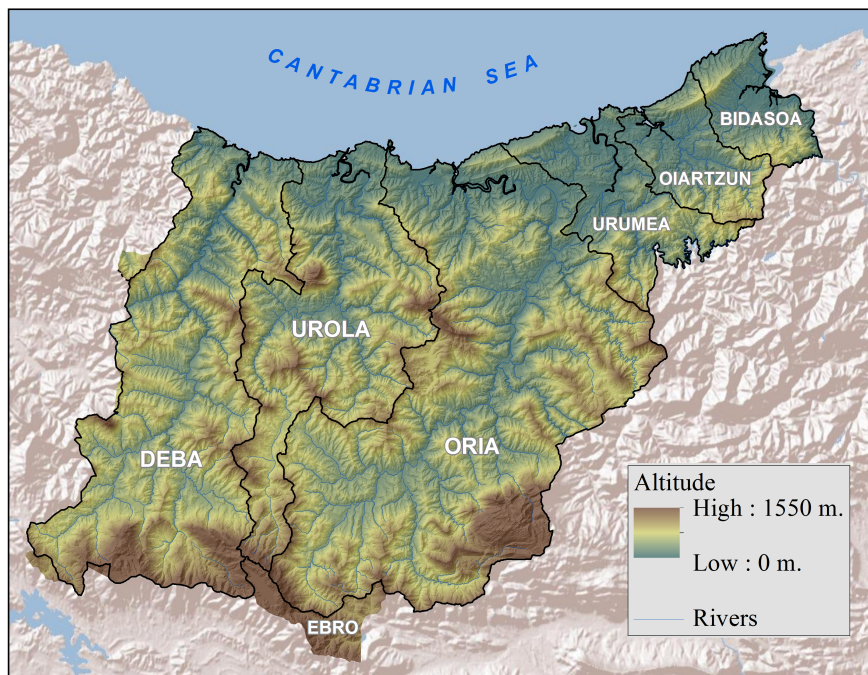
Figure 4.5: *Location of the main river basins within the study area.*
4.5 Irudia: Ikerketa eremuko ibai arro nagusien kokalekua.

Sei unitate hidrografiko hauek, hegoaldetik iparraldera norabidetutako eta malda handiko bailaraz osatutako egitura hidrografiko dendritikoa osatzen dute, kostaldeko hainbat errekak zuzenean itsasora isurtzen dituzten sare kataklinalekin batera. Horien ezaugarri hidrologiko garrantzitsuenak (4.1 Tau.) torrentzialitate handiarekin lotuta daude, prezipitazio altuen eta material iragazkaitzen gehienezko presentziaren ondorioz (Borja & Collins, 2004). Horrek ur emarien igoerak tartekatzea dakar maiztasun handiz. Ibaiak nahiko ahokatuta zirkulatzen dute, ubide bakarrekoak dira eta ez oso meandriformeak; beraz, malda handiko zirkulazioak dira eta garraio energia kontzentratuta doa (4.1 Tau.) (Ibisate *et al.*, 2000).

Hala eta guztiz ere, azpimarratzekoa da Gipuzkoako arro bat berak ere ez

duela bere osotasunean dinamika %100 naturala aurkezten. Izan ere arro guztietan badira ur emariak erregulatzen dituzten urtegiak, esaterako: Urkulu eta Aixola, Deba bailaran; Barrendiola eta Ibai-Eder, Urolan; Arriaran eta Lareo, Oria arroan; Artikutza eta Añarbe, Urumean eta San Anton, Bidasoan.

Table 4.1: *Summary table of the average hydrological and hydrographical features for the 6 principal river basins covering the study area (URA, 2017; Rallo et al., 1992).*

4.1 Taula: Ikerketa eremuko 6 ibai arro nagusien ezaugarri hidrologiko eta hidrografikoak laburtzen dituen taula (URA, 2017; Rallo *et al.*, 1992).

Name	Catchment surface (km^2)	Mean precipitation (mm)	Water contribution (hm^3)	Runoff Coef.
Bidasoa	64.71	1895	75.5	0.62
Oiartzun	85.78	1775	92.3	0.61
Urumea	290.8	1967	410.2	0.72
Oria	899.35	1642	800.3	0.54
Urola	342.21	1486	260.5	0.51
Deba	537.46	1552	441	0.53
Name	**Hmax (m)**	**Length (km)**	**Mean slope (%)**	
Bidasoa	900	66	1.36	
Oiartzun	320	18.5	1.73	
Urumea	600	46.5	1.29	
Oria	1000	70	1.43	
Urola	620	55.7	1.11	
Deba	860	54	1.59	

4.5   Lurraldearen beste ezaugarri batzuk

Ingurune fisikoa ezaugarritzeaz gain, lurzoruaren estaldurak eta biztanleriaren banaketak ere, eragin zuzena izan dezakete aztergai diren lur labainketen banaketa espazialean. Kasu honetan, ikerketa eremua baso (%62.6), larre-belardiz (%25.5) eta laborantza lurrez (%1.2) estalia dagoen lurralde bat izan arren gehien batean (Eusko Jaurlaritza, 2016), gizakiak okupatutako eremu antropizatu oso kontzentratuak agertzen dira bailara nagusien ibilguetan zehar (%6.5) (4.6b Irud.). Baso naturalen artean espezie nagusienak pagadiak (*Fagus silvatica*) eta ariztiak dira (*Quercus robur*, *Q. pirenaica* eta *Q. petrea* gehien bat). Landatutako basoetan berriz, Gipuzkoako azaleraren %37.8 izanda (Eusko Jaurlaritza, 2016), espezie koniferoak gailentzen dira (*Pinus radiata*, *Larix europaea*, *Larix leptolepis* edo *P. laricio*).

Bestalde, Gipuzkoako LHren kokaleku estrategikoak Europarako pasabide bezala, komunikabide sare biziki dentso baten garapena ekarri du (4.6c Irud.). AP-1 autopistak (111.5 km) eta autobide nazional bik (A-1, 44 km eta A-15, 19.2 km) gurutzatzen dute, eta ADIF[2] (240 km) eta EUSKOTREN (93.1 km) trenbide sareak zabaltzen dira bai kostaldeko eta bai barnealdeko eremuetan zehar. Horretaz gain, bigarren mailako errepide sarea bailara eta azpi bailara guztietara hedatuta dago.

Gainera, 365.4 biz/km^2-ko batez besteko populazio dentsitatea izan arren, biztanleria hetereogeneoki sakabanatuta aurkitzen da. 1000 biz/km^2-tik gorako udalerri gehienak ekialdean kontzentratzen dira, herri guztien erdiek 150 biz/km^2 baino gutxiago dituen bitartean (4.6d Irud.). Izan ere, urbanizatutako lurzoruak korridore nagusietan zehar kokatuta badaude ere, etxebizitza ugari urbanizatu gabeko landa lurretan kokatuta aurkitzen dira.

Ondorioz, aipatutako baldintza klimatikoak, erliebearen ezaugarriak eta gizakiaren esku hartzearen ondorioz, paisaia unitate desberdinetan konfiguratuta dagoen lurralde batetaz hitz egiten ari gera. Baina, izan bailara fondoetako paisaian, izan atlantiar mendiez ezaugarritutako paisaietan lurraldearen eraldaketa garrantzitxu bat gauzatu da, batez ere, garraio azpiegituren eta hiri hazkundeari erantzuna emateko. Horrek Gipuzkoako mendi hegaletan eragin handia izan du,

---

[2]ADIF sarearen barne kontutan hartu dira abiadura handiko trenaren azpiegitura lanen trazatuak

malda askoren egonkortasuna desorekatuz eta lur labainketen gertaerak emanez, besteak beste (Corominas *et al.*, 2017).
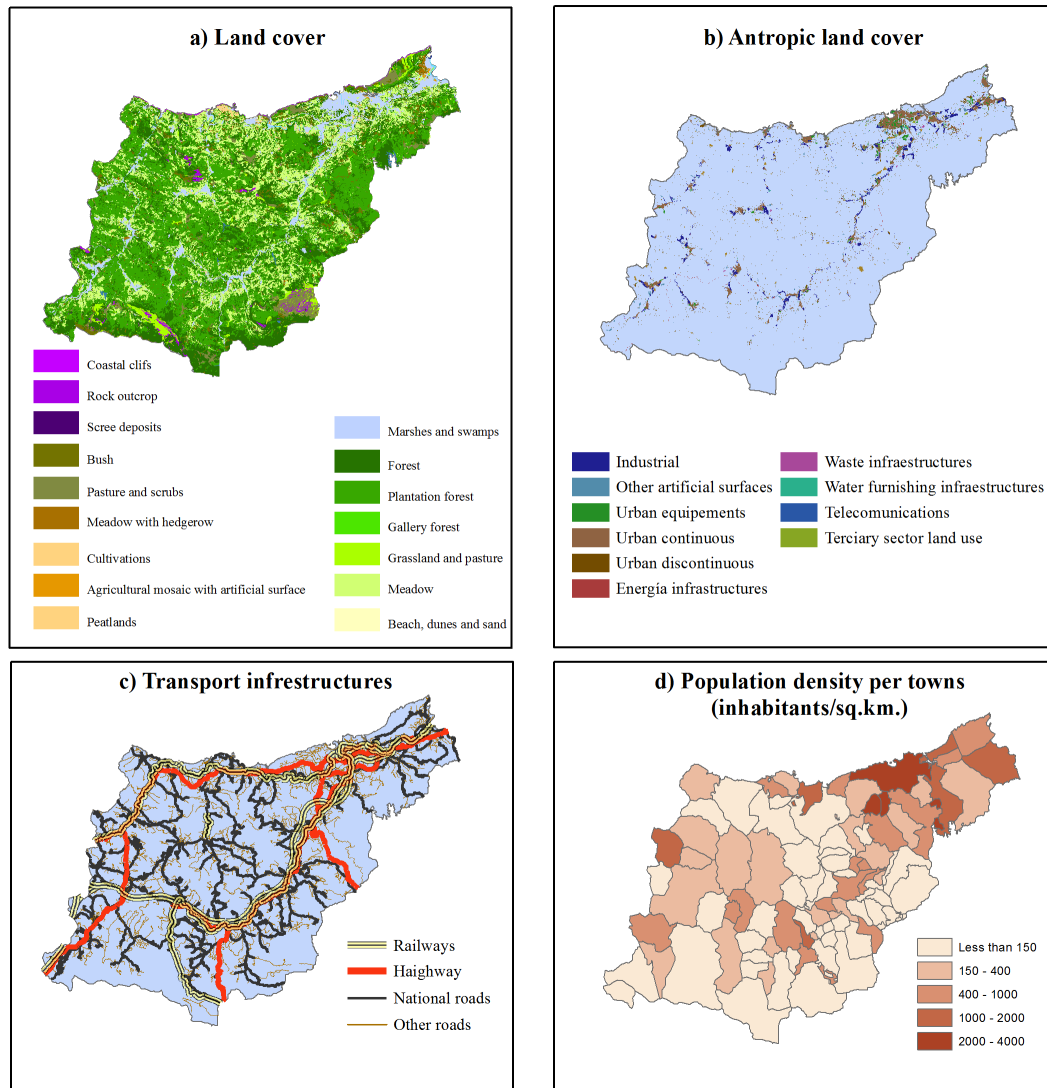
Figure 4.6: *Description of the land cover and population distribution. a) Land cover distribution according to the National Forest Inventory of the 2010; b) Artificial land cover distribution; c) Communication network obtained from Euskadiko DEA (2014); d) Population density map. Data form 2016 obtained from Gaindegia (2018).*
4.6 Irudia: Lurzoruaren estaldura eta biztanleriaren banaketaren deskribapena. a) Lurzoruaren estaldura 2010eko Baso Inbentario Nazionalaren arabera; b) Lurzoru artifizialaren estaldura; c) Komunikabide sarea (Euskadiko DEA, 2014); d) Biztanleriaren dentsitate mapa. 2016ko datuak Gaindegia (2018).

Bibliografia

Ábalos, B.: Geologic map of the Basque-Cantabrian Basin and a new tectonic interpretation of the Basque Arc, International Journal of Earth Sciences, 105, 2327–2354, 2016.

Barnolas, A. & Pujalte, V.: La Cordillera Pirenaica, in: Geología de España, edited by Vera, J., pp. 233–241, SGE-IGME, Madrid, 2004.

Borja, A. & Collins, M.: Oceanography and marine environment in the Basque Country, vol. 70, Elsevier, 2004.

Campos, J. & García-Dueñas, V.: Mapa Geológico de España 1:50000, hoja 64. San Sebastián, IGME, 1972.

Campos, J., Olivé, A., Ramírez, J. I., Solé, J., & Villalobos, L.: Mapa Geológico de España 1:50000, hoja 89. Tolosa, IGME, 1983.

CGS, Salazar, A., de Alba, S., Gallardo, J., Portero, G., Pascual, M. H., & Olivé, A.: Geomorfología y edafología de Guipúzcoa, Diputación Foral de Gipuzkoa, 1991.

Corominas, J.: El clima y sus consecuencias sobre la actividad de los movimientos de ladera en España, Cuaternario & Geomorfología, 20, 89–113, 2006.

Corominas, J., Mavrouli, O., & Ibarbia, I.: Metodología integrada para la evaluación de riesgos en la Red de Carreteras de Gipuzkoa, in: IX Simposio sobre Taludes y Laderas Inestables, edited by Alonso, E., Corominas, J., & Hürlimann, M., pp. 395–406, International Centre for Numerical Methods in Engineering (CIMNE), Santander, 2017.

Denèle, Y., Paquette, J. L., Olivier, P., & Barbey, P.: Permian granites in the Pyrenees: the Aya pluton (Basque Country), Terra Nova, 24, 105–113, 2012.

Euskadiko DEA: Infraestructura de datos espaciales de Euskadi, URL `www.geo.euskadi.eus`, 2014.

Eusko Jaurlaritza: Plan especial de emergencias ante el riesgo de inundaciones de la Comunidad Autónoma de Euskadi, Tech. rep., Segurtasun Saila, 2015.

Eusko Jaurlaritza: EAEko baso inbentarioa 2016, Tech. rep., Ekonomia Garapen eta Azpiegitura Saila, 2016.

EVE: Mapa Geológico del País Vasco Escala 1:200000, Ente Vasco de Energía Instituto Tecnológico Geominero de España, 1991.

Fdez-Arroyabe, P. & Martin-Vide, J.: Regionalization of the probability of wet spells and rainfall persistence in the Basque Country (Northern Spain), International Journal of Climatology, 32, 1909–1920, 2012.

Gaindegia: Euskal Herriko adierazle galeria nagusia, Tech. rep., Euskal Herriko Ekonomia eta Gizarte Garapenerako Behategia, 2018.

González, M. J. & Serrano, E.: El Relieve, in: Geografía de Euskal Herria, edited by Meaza, G. & Urrestarazu, E., vol. 2, p. 238, Editorial Ostoa, 1996.

González-Hidalgo, J. C., Brunetti, M., & de Luis, M.: A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945–November 2005), International Journal of Climatology, 31, 715–731, 2011.

Ibisate, A., Ollero, A., & Ormaetxea, O.: Las inundaciones en la vertiente cantábrica del País Vasco en los últimos veinte años: principales eventos, consecuencias territoriales y sistemas de prevención, Serie Geográfica, 1, 177–186, 2000.

Martínez-Torres, L.: Transversal a la Cuenca Vasco-Cantábrica: introducción a la estructura y evolución geodinámica, Euskal Herriko Unibertsitateko Argitalpen Zerbitzua, 1997.

Ormaetxea, O. & Sáenz de Olazagoitia, A.: Análisis y caracterización de los factores que intervienen en los movimientos de ladera y aproximación de la susceptibilidad en el País Vasco, Lurralde: Investigación y Espacio, pp. 81–109, 2017.

Pedreira, D.: Estructura cortical de la zona de transición entre los Pirineos y la Cordillera Cantábrica, Ph.D. thesis, Universidad de Oviedo, Oviedo, 2004.

Rallo, A., Docampo, L., García de Bikuña, B., Rico, E., & Sevillano, M.: Caracterización hidrobiológica de la red fluvial de Álava y Gipuzkoa, Servicio Central de Publicaciones del Gobierno Vasco, 1992.

URA: Mapa hidrológico de la Comunidad Autónoma del País Vasco, Escala 1: 150000, Uraren Euskal Agentzia, 2017.

Uriarte, A.: El clima, in: Geografía de Euskal Herria. Clima Y Aguas, edited by Meaza, G. & Urrestarazu, E., vol. 3, chap. 1, pp. 8–83, Editorial Ostoa, 1996.

Urrestarazu, E. & Galdos, R.: Geografía del País Vasco, Editorial NEREA, 2008.

Chapter 5

Methodology

This chapter aims to provide the complementary information about the basic data and methods used in this thesis. However, it is not intended to offer a deep review of the different methodologies applied, considering that all of them were sufficiently discussed in previous studies (see chapter 2). Instead, it is intended to contribute with additional material to ease the comprehension of the following chapter 6, where results of three investigations will be presented.

These researches were sequentially designed and carried out in order to confirm or reject the hypothesis of this thesis and to achieve the proposed objectives. So, chapter 5, provides the details about the data and methods that support what is presented in sections 6-I, 6-II and 6-III.

The chapter is organized as follow. First, information about all the data used for subsequent experiments is shown, as well as their sources or the work flow followed up during their collection (section 5.1). Then, details about the methods followed up at different steps for developing landslide susceptibility maps are presented in section 5.2. This one, refers to the two different approaches showed in section 6-I and 6-II. And finally, the methods applied for the calculation of landslides responsible precipitation thresholds and the qualitative relation between rainfalls and landslides (section 6-III) are explained in section 5.3.

5.1   Data collection

5.1.1   Landslide inventories

Statistical landslide susceptibility models are performed according to the basic information of presence or absence of landslides, usually encoded with 1 and 0 values respectively. Consequently, there is no doubts about the fact that, the landslide inventory is the first critical step during the landslide susceptibility modelling, because the model equation will be completely dependant on this data. Nevertheless, according to Guzzetti *et al.* (2012), the quality of a landslides map (or landslides inventory) depends on its accuracy, and on the type and certainty of the information shown in the map. In addition, considering that standards do not exist, the definition of that accuracy is not straightforward.

For this reason, different landslide inventories were tested in this study.

Three of them were obtained through the bibliographical review, and after a preliminary assessment of their accuracy, a fourth inventory was produced by direct geomorphological field work. Additionally, an independent landslide inventory was also carried out by means of press review, for the purpose of obtaining a set of landslides with temporal information. The following sections explain the details about all these inventories.

5.1.1.1 Bibliographical landslide inventories

The bibliographical landslide inventory was performed by collecting already existing and available landslide registrations coming from three different sources.

Inventory of the Basque Government

It is about a research carried out in 1995 for Gipuzkoa, Bizkaia and Araba Provinces (INGEMISA, 1995), where by means of an exhaustive bibliographical review of the scientific articles and technical reports published until that date, the inventory of the landslides was obtained. Landslides were classified as *slides*, *falls*, *flows*, *topples* or *complex*, and a detailed sheet was provided for each of them.

The UTM coordinates as well as the type of landslide information were extracted from the sheets and a data table was created in order to summarize all the attributes. Then, thanks to the ArcGIS 10. software of Geographic Information Systems (GIS), only those points concerning Gipuzkoa Province (GP) were extracted. The resulting 425 points can be seen in figure 5.1.

Inventory of the road network

This work was carried out in 2013 by order of the Provincial Council of Gipuzkoa and was executed by the IKERLUR company (GFA, 2013). The objective was the inspection and control of the stability along the whole road network of the territory, providing detailed information about the "conflictiv" points. In this case, 117 unstable points (and its UTM coordinates) were detected, classified as *slow slides*, *falls* and *debris flows* (Fig. 5.2).
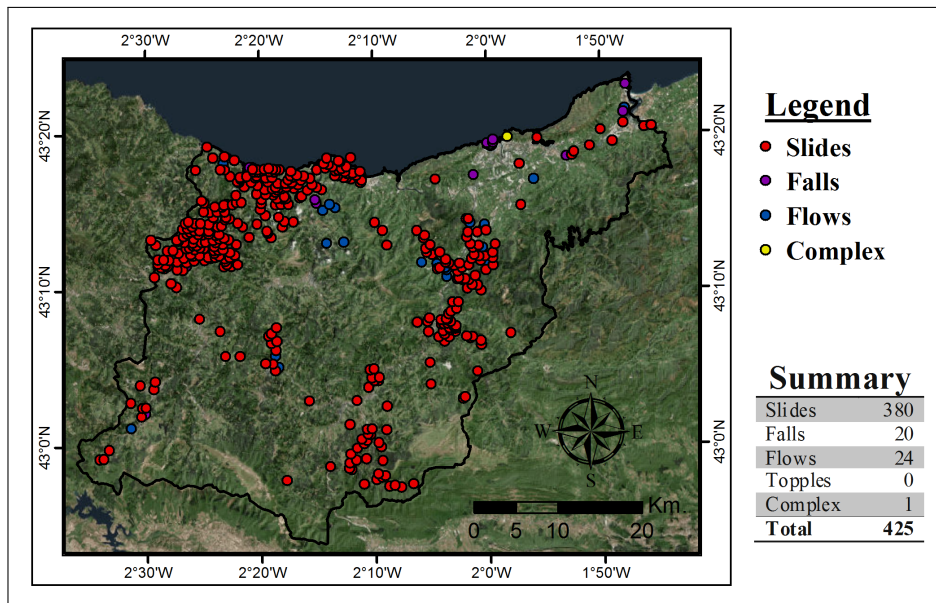
Figure 5.1: *Distribution of the bibliographic landslide inventory from INGEMISA (1995).*

5.1 Irudia: Bibliografiatik lortutako lur labainketen banaketa espaziala. Iturria: INGEMISA (1995).
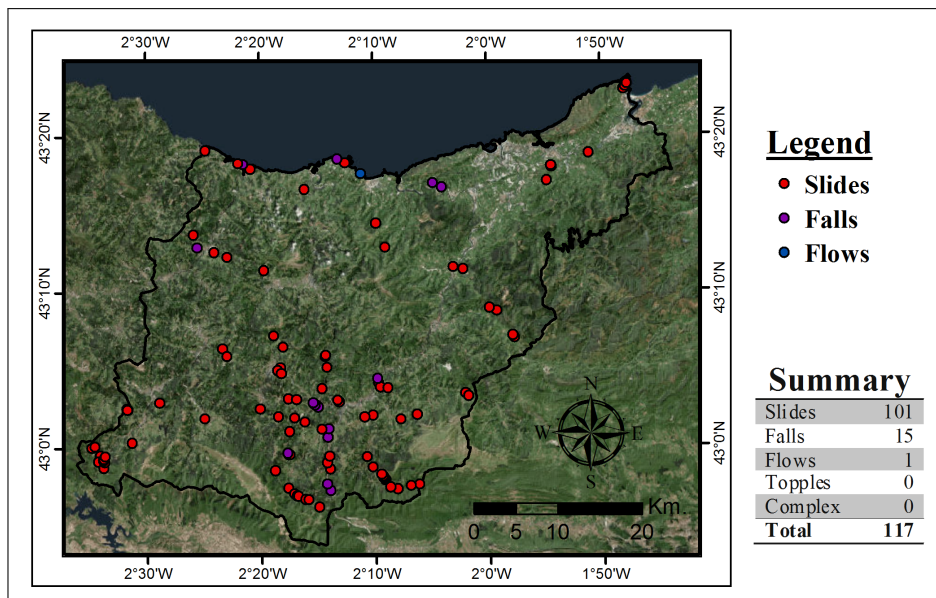
Figure 5.2: *Distribution of the bibliographic landslide inventory from GFA (2013).*

5.2 Irudia: Bibliografiatik lortutako lur labainketen banaketa espaziala. Iturria: GFA (2013).

Inventory of the geomorphological map

The geomorphological map of Euskadi in 1:25000 scale is available in digital format from the Infrastructure of Spatial Data Service of the Basque Council (Euskadiko DEA, 2014) and it provides the spatial distribution of three different landslides features:

Landslide scarps: they are the main scarps of the landslides drown by polygons. 93 scarps were detected within GP.

Rock mass deposits: they are rock masses that were travelled from their original state delimited by polygonal areas. In this case 38 polygons were located in our study area.

Shalow slides: it shows the areas of the shallow landslides by means of polygons. 88 cases were found.

Furthermore, in order to ease the location of all these features, the central point of each polygon was calculated (Fig. 5.3).
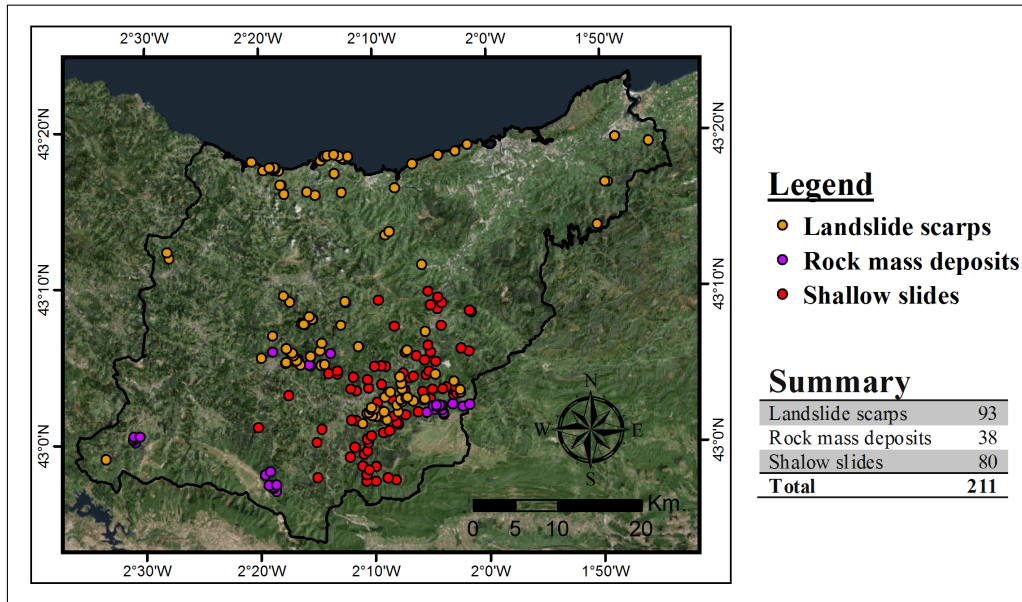
Figure 5.3: *Distribution of the bibliographic landslide inventory from Euskadiko DEA (2014).*

5.3 Irudia: Bibliografiatik lortutako lur labainketen banaketa espaziala. Iturria: Euskadiko DEA (2014).

5.1.1.2   Field-based landslide inventory

It concerns a single landslide inventory coming from a field survey (carried out during summer 2015 and 2016) in which every landslide was documented considering its shape and dimensions. Although the visual interpretation of aerial photographs may be the most suitable option in order to obtain a multi-temporal landslide inventory (Santangelo *et al.*, 2015), the dimension of the study area together with the time limitations concerning this project makes this technique not applicable. Other interesting options could be the more recently developed automatic and semi-automatic techniques such as: (i) analysis of high and very-high resolution (VHR) digital elevation models (DEM) derived from LIDAR, (ii) visual analysis of monoscopic high and VHR satellite images, or (iii) automatic and semi-automatic analysis of high and VHR satellite images (Murillo-García *et al.*, 2015). But, the lack of resources and the need of expert management for carrying out such new methods, made not possible their application with guaranties. Thus, it was decided to produce the own landslide inventory by direct geomorphological field work.

To do so, the study area was divided in 6 portions corresponding to the six main watersheds, and a field trip that spent between one or two weeks (depending on the size of each sub-area) was carried out for each one. During the field trip, a random sub-set of the bibliographical landslide points as well as some newspaper references were used as guide-points, but every landslide found along the field work was documented in a field-sheet as the example shown in figure 5.4 (all the field-sheets together with the digital layer of the complete field-based landslide inventory can be seen in supplementary material, see Appendix A). There, apart from some general information like the watershed and the municipality or other possible relevant factors observed on the field, at least, the following basic data were collected:

The specific ID: a unique identifier was given to each single landslide, where the first three characters means the abbreviation of the watershed in which it was found, followed by the identifier number.

The type of landslide: according to the Cruden & Varnes (1996) classification shown in section 1, the typology of the slope movement was registered.

Occurrence date: in case of knowing the date in which the landslide happened,

it was also documented, though this information was scarce.

Date in which the landslide was visited: indicates, when was documented this information.

UTM coordinates: using the GPS the location of the landslide was registered by means of the UTM coordinates.

Pictures of the landslide: for all the landslides found during the field trips, pictures were captured for further verifications in case of doubts.

LANDA LANEKO ORRIA		
Unitate hidrografikoa:	Udalerria:	zk:

Lur labainketa mota:

Gertaera eguna:

Egilea:	Data:

1:5.000 orriaren zk:	Koordenatuak (UTM):
Hegaldia:	X:
Fotograma zk:	Y:
	Z:

BALDINTZA FAKTOREAK

Litologia:	Malda (°):
Lurzoruaren lodiera:	Orientazioa:
Landaretza:	Makurdura:

Hausturak:

Lur erabilpenak (egungo eta iraganekoak):

Azpiegiturekin erlazioa:

ERAGIN FAKTOREAK

Ez ohiko euri-jasa:

Besteak (higadura, ibai dinamikak …):

BESTELAKO EZAUGARRIAK

Dimentsioak:

Egoera eta garapena:

Kalteak:

Zuzenketa neurriak:

Argazkiak:

Besterik:

Informazio iturriak: Landa lana

Figure 5.4: *Example of a field-sheet. Compiled in Basque language.*
5.4 Irudia: Landa laneko fitxaren adibidea. Euskaraz.

Figure 5.5 summarizes the methodology used for the field-work based landslide inventory collection. First, all the paths travelled during the field work were tracked with the GPS. Then, apart from the data collected in the field sheets, the GPS waypoints of the upper part of the crown (a), the lower part of the toe (b) and both sides (c, d) were saved using the OruxMaps app (version 6.5.10) of the smartphone (Fig. 5.5 II). After that, the waypoints were exported to the Google Earth application and the landslide polygons were digitalized and named with their ID code. Nevertheless, it has to be pointed out that a little portion of landslides could not be visualized in the Google Earth's satellite imagery, because they were very recent, they were undetectable due to re-vegetation or they were removed and reconstructed before a new satellite image was taken. In such case, the landslide digitalisation was carried out approximately using the four waypoints dimension and the pictures captured on the field. Finally, the polygons were exported into the GIS software and transformed into shapefiles (*.shp) for their further processing.
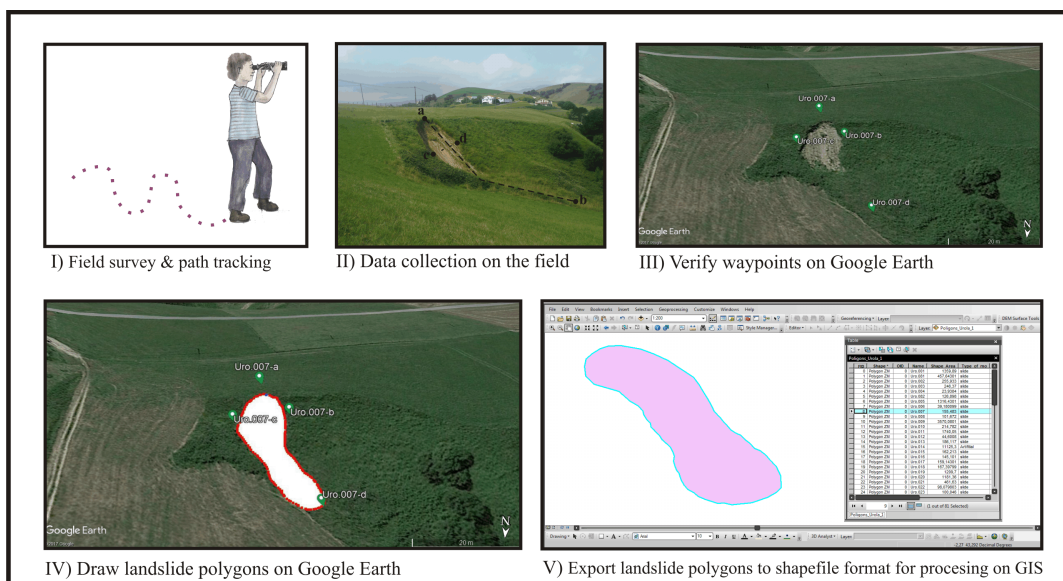
Figure 5.5: *Methodological work flow scheme for the field-work based landslide inventory collection.*
5.5 Irudia: Landa lanean oinarritutako lur labainketen inbentarioa gauzatzeko jarraitutako prozedura.

5.1.1.3   Press-based landslide inventory

Landslide location and temporal information can be collected by means of the newspaper review (Cuesta *et al.*, 1999). In this work, the most sold newspaper in Gipuzkoa Province (El Diario Vasco) was chosen and was collected every report

in which the key words "deslizamiento", "desprendimiento" or "derrumbe" (slide, landslide and crumble as the typical colloquial words in Spanish for referring to landslides, respectively) was cited (see example in Fig. 5.6). Because the digital newspaper library did not offer any information previous to 2006, the time lap covered by this review starts the first January of this year and finishes the thirty one December of 2015 (detailed information about all the press-based landslide inventory is available in Appendix A).

Figure 5.6: *Example of one recorded press report offering information about a landslide event. El Diario Vasco 13-02-2013.*

5.6 Irudia: Lur labainketa bati buruz jasotako berriaren adibide bat. El Diario Vasco 2013-02-13.

As a result, 2005 reports were obtained from the newspapers review. Later on, considering this information source, landslides that occurred within the administrative boundaries of the GP were selected and the following information was summarized in a data base:

Occurrence date: the date in which the landslide happened.

Occurrence moment: the exact time at which the landslide occurred. If the exact time was not known, the approximated time was registered depending on the accuracy of the information (night 6:00; morning 12:00; afternoon 18:00; evening 23:59). And if only the date was known 23:59 time was registered.

Occurrence moment accuracy: the accuracy of the information about the occurrence time of the landslide (Exact time; Relative time; Only the day is known). Notice that the time was transformed into Coordinated Universal Time (UTC).

Location accuracy: the accuracy of the information about the location of the reported landslide. At least the municipality has to be known to accept the event in the inventory (Exact location; Relative location with less than 1 km of error; Relative location with less than 10 km of error; Only the municipality is known).

Type of movement: when information about the characteristics of the movement was reported, or if the news was supported with pictures, the process was classified according to the (Cruden & Varnes, 1996) classification (Slide; Rock fall; Flow; Mixed movement). If no information was provided in this regards, the generic word "landslide" was assigned.

Cause: if this information was reported on the news it was introduced in the database (Rainfall; Human activity; Waves; Fluvial erosion; Wind; Unknown).

Damage type: if this information was reported on the news it was also introduced in the database the type of damage caused by the movement (Personal; Buildings; Communication network; Parks; Undefined).

However, it has to be pointed out the meaning of the landslide inventory. Each landslide refers to a single case of terrain instability occurred at a given moment and in a given place. Thus, it frequently happens that more than one landslide occurred very close in time and space, due to the same rainfall event.

5.1.2 Explanatory variables

Quoting to Van Westen, after the landslide inventory, "*the next crucial input data for susceptibility, hazard and risk assessment consists of the spatial representation of the factors that are considered relevant for the prediction of the occurrence of future landslides*" (Van Westen *et al.*, 2008). The usage of such variables can change depending on the type of landslide, the scale and the method in which they are

applied, and they directly depend on the availability of existing data and resources (Van Westen *et al.*, 2008; Süzen & Kaya, 2012; Budimir *et al.*, 2015; Malamud *et al.*, 2014).

For an overview about the trending on the spatial variables used for landslide susceptibility assessment, in Van Westen *et al.* (2008) and more recently in Malamud *et al.* (2014) detailed reviews based on papers survey are available. Moreover, Budimir *et al.* (2015) also offers the review of the most used explanatory variables in the specific case of logistic regression (LR) method driven susceptibility analysis.

As a matter of fact, despite the big amount of different explanatory variables tested in the scientific literature, the biggest part of them can be classified in one of the following groups:

DEM and derived variables: the Digital Elevation Model (DEM) is the digital representation of the earth surface elevation which shows, with different level of details (depending on its spatial resolution), the topography of a given study area. Thanks to this information and the developed functionalities of the GIS computer programs, they can be obtained divers derived variables such as the slope gradient or the aspect, and additionally other morphometric parameters such as the flow accumulation or the drainage density.

Geology and Soil related variables: as landslides are movements of the terrain down the slope, all kind of features of the terrain itself were historically used for landslide susceptibility modelling. Starting from the lithological classification until the soil typology, going through the depth of the surface formation or the stratigraphic orientation and so on.

Land cover variables: the *land cover* is another common variable which can be represented in form of land use maps or vegetation maps, as well as, for example, more dynamic variables such as land use change maps or *normalized difference vegetation index* (*NDVI*) evolution maps.

But, as Ayalew & Yamagishi (2005) already pointed out, the selection of the spatial factors with major role is a difficult task because neither universal criteria nor guidelines exists about this issue. However, according to the same authors, there is a consensus about the minimum conditions that every variable must meet: (i)

to have a certain degree of affinity with the dependent variable; (ii) to be fairly represented all over the study area; (iii) it has to be non-uniform; (iv) it must be expressed by any of the different types of measuring scales and (v) its effect should not account for double consequences in the final result.

The aim of this section is to present the spatial variables considered in the current study and the method followed for their final selection as explanatory variables of the landslide susceptibility models, so as to be sure that the used variables fulfil the above mentioned conditions.

From the Infrastructure of Spatial Data Service of Euskadi (Euskadiko DEA, 2014), that compiles the basic, photographic and geoscientific digital cartography, the vector layers of Lithology, Permeability, Regolith Thicknesses, Land Uses (two layers from two different sources), Vegetation, Drainage network and Transport network by road and train were downloaded. Likewise, the DEM with 5x5 m spatial resolution as well as SPOT 5 satellite multi-spectral imagery were used to produce derived variables.

Taking into account the available spatial information, for this work a set of 20 original environmental variables was considered (Tab. 5.1). These below described variables can be continuous, which means that each pixel of the layer represent a numerical value between the minimum and maximum of the range of the variable, or categorical, which means that each pixel belongs to a category between the different classes of the variable. Nevertheless, some of them were afterwords transformed in order to test all the different forms of presentation of variables found in the bibliography (see section 6.1.3).

5.1.2.1 Continuous variables

Elevation

The DEM used for the *elevation* representation comes from the LIDAR (Light Detection and Ranging) data capture carried out during 2012 and re-sampled in 5x5 meters cell resolution. The source raster file is available in Euskadiko DEA (2014) named under the code **md_IDEEu_MDT_LIDAR_5M_2013.tiff**, and it covers the entire administrative boundaries of Euskadi, so before using it, the corresponding area of the GP was clipped.

Table 5.1: *List of the original environmental variables. Land cover 3 refers to the vegetation map.*

5.1 Taula: Ingurugiroko aldagi originalen zerrenda. Lurzoruaren estaldura 3 landaretza mapari dagokio.

Continuous variables		
elevation	*slope*	*sinusoidal slope*
surface area ratio	*topographic wetness index*	*curvature*
profil curvature	*planform curvature*	*distance to the river*
distance to the transport network	*normalized difference vegetation index*	
Categorical variables		
lithology	*permeability*	*regolith thickness*
land cover 1	*land cover 2*	*land cover 3*
aspect	*distance to the main riverstreams CAT*	*distance to transport network CAT*

Theoretically, it is worth the usage of a DEM previous to any slope instability event in order to represent, in a more faithful way, the geometry of the surface before the landslide occurrence. Nevertheless, the landslide inventories that will be used in this study comes from divers sources and the occurrence moment of the most part of events was uncertain. Only the dates of publication of the bibliographic sources and the dates of the field trips were available as temporal references, and in such case, there were landslide occurred before 1995 (section 5.1.1.1) as well as others occurred before 2016, because the field trips were carried out during the summer of 2015 and 2016. Thereby, considering that at the beginning of this thesis (2014) they were only available DEM layers of 2008 and 2013, the most recent was chosen in order to have a well known and observable starting point.

The variable *elevation* (Fig. 5.7) is quantitative and continuous, and it represents in meters the altitude above the sea level of each regular cell. Although, the altitude itself is hardly justifiable as landslide influencing factor, many authors like Corominas (2000) found relations, usually more related with derived features from the altitude, such as the higher precipitation or the non-existence of soils in high altitudes. Moreover, as the DEM was necessary for other derived variables production, the decision was taken to also test the *elevation* as explanatory variable.
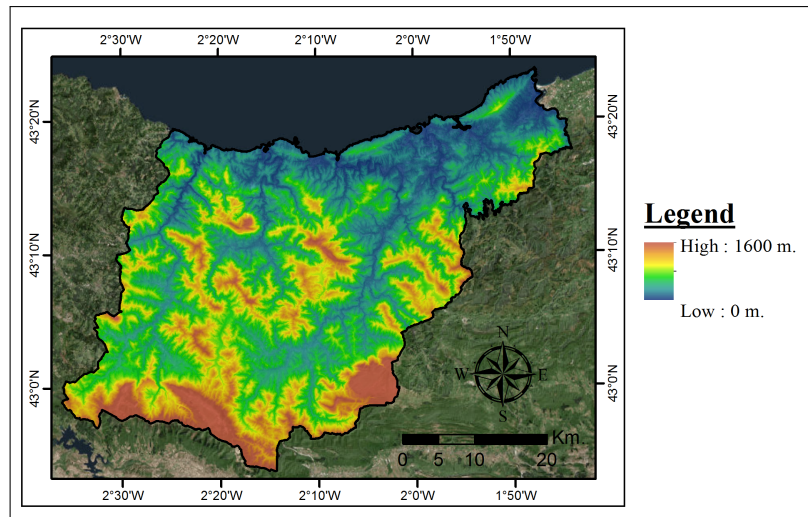
Figure 5.7: *Elevation variable's spatial distribution.*
5.7 Irudia: Altuera aldagaiaren banaketa espaziala.

Slope

The *slope* (Fig. 5.8) was digitally calculated as a derived product of the DEM. By means of the ArcGIS 10 software, the maximum rate of change in value from each cell to its neighbours was calculated, assuming that the maximum change in elevation over the distance between the cell and its eight neighbours identifies the steepest downhill descent from the cell. For this reason, the calculations were carried out over the original DEM, and then the resulting layer was clipped in order to fit the study area. This way, border errors were avoided.

This quantitative and continuous variable is one of the basic factors considered in every landslide susceptibility analysis. It represents the existing angle between the terrain surface and the horizontal plane in degrees, and it shows a direct relation with the tangent and normal cutting stress of the surface formation (Amorim, 2012). In addition, the *slope* influences very importantly the water flow velocity and distribution.

Sinusoidal slope

According to the observations carried out by Santacana *et al.* (2003) and Amorim (2012), there are some types of landslides, like shallow slides, that usually are concentrated in medium slope areas, decreasing their presence from 45° of slope on. Such a behaviour can be explained with the lack of surface formations in very steeply
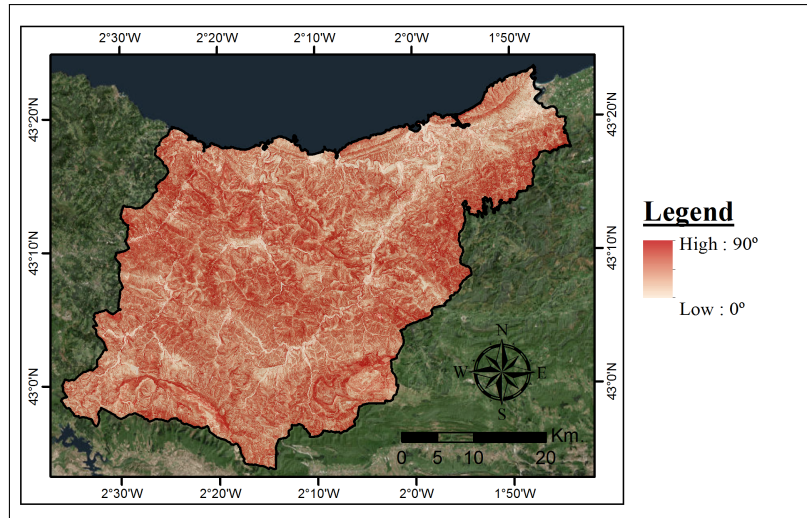
Figure 5.8: *Slope variable's spatial distribution.*
5.8 Irudia: Malda aldagaiaren banaketa espaziala.

areas, being, there, more frequent the underlying rock outcrops. Thus, depending on the type of landslides considered to the susceptibility analysis, the relation between them and the *slope* may not be completely positive, because in some cases, from 45° of slope, the more is the slope the less is the probability of finding landslides (Amorim, 2012).

In order to cover as much as possible variables, a mathematical transformation proposed in Santacana (2001), and shown in equation 5.1, was applied to the *slope* variable, so as to increase its value until the 45°, and then decrease it gradually until 90°. This way, the value of the *sinusoidal slope* is high for medium values of slope, and it is lower for flat and very steeply areas.

$$Sinusoidal slope = sin(2 \cdot Slope) \tag{5.1}$$

It has to be noticed that the slope values were previously transformed into radians, because the raster calculator of ArcGIS 10 considers the sine function in that unit of measurement.

As a result, a quantitative and continuous variable with values between 0 and 1 was obtained (Fig. 5.9).

Surface area ratio (SAR)

This variable represents the surface roughness, i.e. the variation in a real surface respect to its ideal form of a given area. Rough or smooth surfaces, apart from
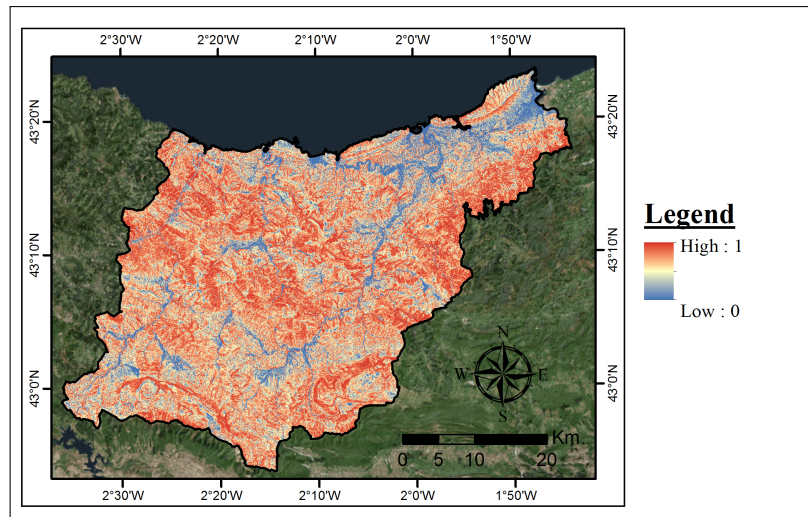
Figure 5.9: *Sinusoidal slope variable's spatial distribution.*
5.9 Irudia: Malda sinusoidala aldagaiaren banaketa espaziala.

being directly related with the superficial water runoff, can be indicators of internal deformation structures, fissures, tension cracks, flow lobes, step like morphology, scarps, or semi circular features (Van Westen *et al.*, 2008).

For this work, the *surface area ratio* index was used as indicator of the level of roughness, which calculates, basing on the DEM, the ratio of the theoretical volume of each cell respect to the surface occupied by it. The derived layer was produced thanks to the *DEM Surface Tools* plug in available for the ArcGIS 10 software [1].

As this quantitative and continuous variable is a ratio, its theoretical value ranges between 1 and 100, though in the case of study area the maximum SAR value reaches 53.

Topographic wetness index (TWI)

It is a concept develloped by Kirkby & Beven (1979) on the field of basin hydrological modelling and used as explanatory variable for landslide susceptibility modelling by Yilmaz in several studies (Yilmaz, 2009, 2010a,b). This quantitative and continuous variable is about a topographic index used to describe the spatial soil moisture patterns, and according to Yilmaz (2010a) it is defined as follow:

$$TWI = ln(\frac{a}{tan\beta}) \tag{5.2}$$

---

[1]http://www.jennessent.com/arcgis/surface_area.htm
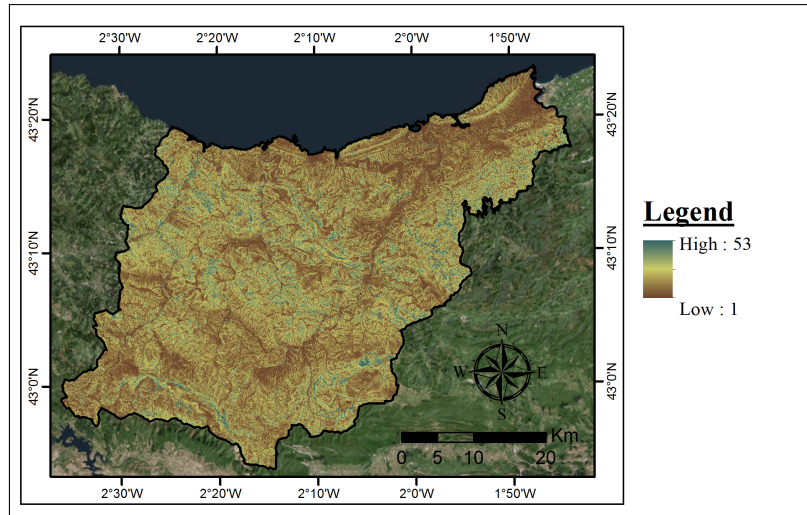
Figure 5.10: *Surface area ratio (SAR) variable's spatial distribution.*
5.10 Irudia: SAR aldagaiaren banaketa espaziala.

where a is the local upslope area draining through a certain pixel per unit contour length and $tan\beta$ is the local slope.

High values of *TWI* (Fig. 5.11) signifies a higher amount of water collected in each point, which may imply a big infiltration of the surface water flow into the surface materials, increasing the pore water pressure and inciting the decrease of the shear strength. The values on the following map are normalized on the range 0-100.
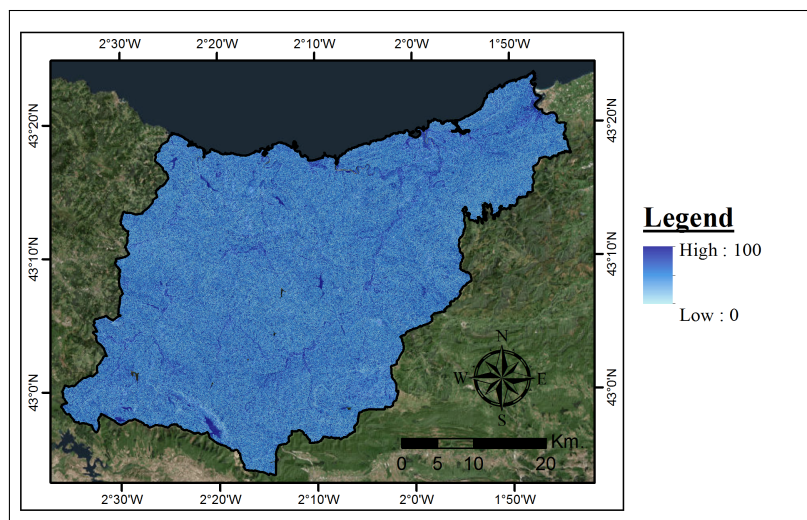
Figure 5.11: *Topographic wetness index (TWI) variable's spatial distribution.*
5.11 Irudia: TWI aldagaiaren banaketa espaziala.

Profile curvature, Planform curvature and Curvature

The curvature represents the change of slope angle within a surface and it shows the degree of concavity or convexity of a given area. In this regards, this feature has widely been used as explanatory variable for landslide susceptibility modelling (Biswajeet & Saro, 2007; Van Westen *et al.*, 2008; Nefeslioglu *et al.*, 2011; Felicísimo *et al.*, 2013; Alvioli *et al.*, 2016).

The curvature layers (Fig. 5.12) were produced by the *curvature* tool available in the ArcGIS 10 software, which is calculated by computing the second derivative of the surface in each cell in a 3x3 matrix. However, the tool offers to the user three different options, and in this study all of them were considered with the objective of testing every possibility. Following are listed the definitions for each one according to the ArcGIS users guide:

Profile curvature: it shows the curvature value parallel to the slope and indicates the direction of maximum slope. It affects the acceleration and deceleration of flow across the surface. A negative value indicates that the surface is upwardly convex at that cell, and flow will be decelerated. A positive value indicates that the surface is upwardly concave at that cell, and the flow will be accelerated. A value of zero indicates that the surface is linear, so no acceleration neither deceleration is expected.

Planform curvature: it shows the curvature value perpendicular to the direction of the maximum slope. It is related to the convergence and divergence of flow across a surface. In this case, a positive value indicates the surface is laterally convex at that cell and negative values indicate the surface is laterally concave at that cell.

Curvature: the standard curvature combines both the profile and planform curvatures.

Distance to the main river-streams

The euclidean distance in meters from each mapping unit to the closest river was computed on the GIS by means of the *proximity* tool. This time, as well as in *slope*
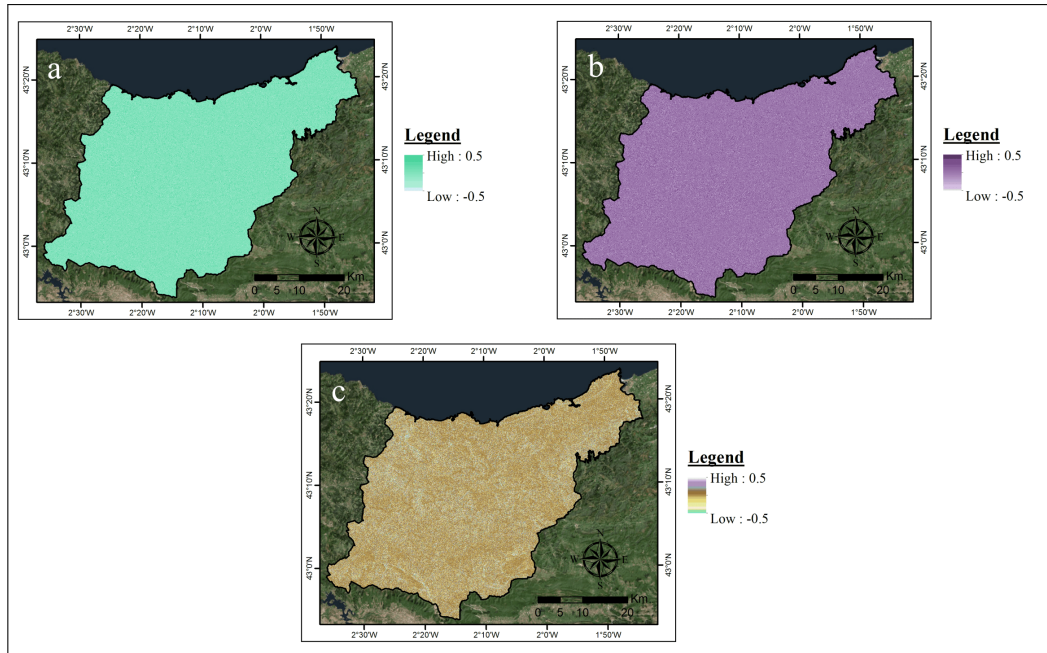
Figure 5.12: *Spatial distribution of a) Profile curvature; b) Planform curvature and c) Curvature.*
5.12 Irudia: a) Profil kurbatura; b) Kurbatura planarra eta c) Kurbatura aldagaien banaketa espaziala.

variable, the original DEM of Euskadi was used for distance calculations, and then the study area of interest was clipped in order to avoid border errors (Fig. 5.13).

Authors like Dai & Lee (2002), Lee (2005) or Bonachea (2006) state that the proximity to a river-stream could bring on landslides due to the lateral erosion caused by the rivers, and thus weakening the base of the slope.
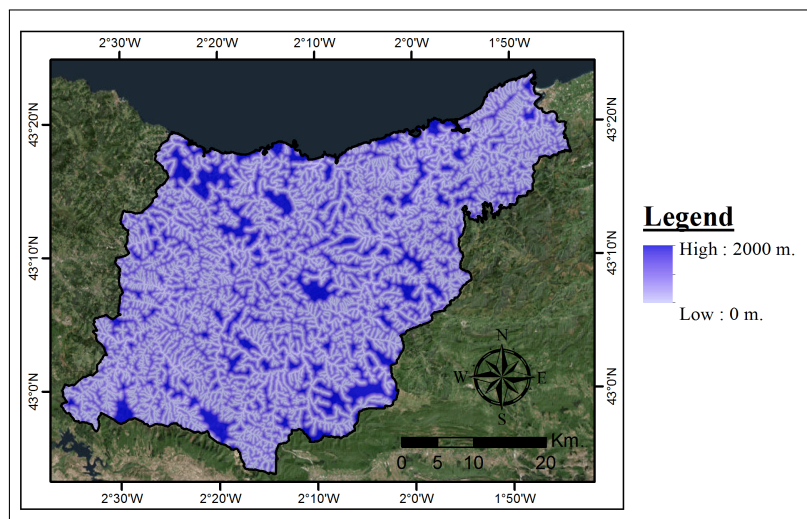
Figure 5.13: *Spatial distribution of distance to the main river-streams variable.*
5.13 Irudia: Ibai gertuenarekiko distantzia aldagaiaren banaketa espaziala.

Distance to the transport network

Following the same method applied in the previous variable, the euclidean distance in meters from each mapping unit to the closest transport infrastructure was computed. In this case, highways, main road network as well as all the train network were considered, although all the tunnel and bridge segments were previously removed in order to only take under consideration the superficial segments (Fig. 5.14).

Similar variables were used before for landslide susceptibility modelling (Van Westen *et al.*, 2003; Pradhan & Lee, 2010; Akgun, 2012) arguing that landslides may be more frequent along roads, due to inappropriate cut slopes and drainage from the roads or other lineal transport infrastructures like railways.
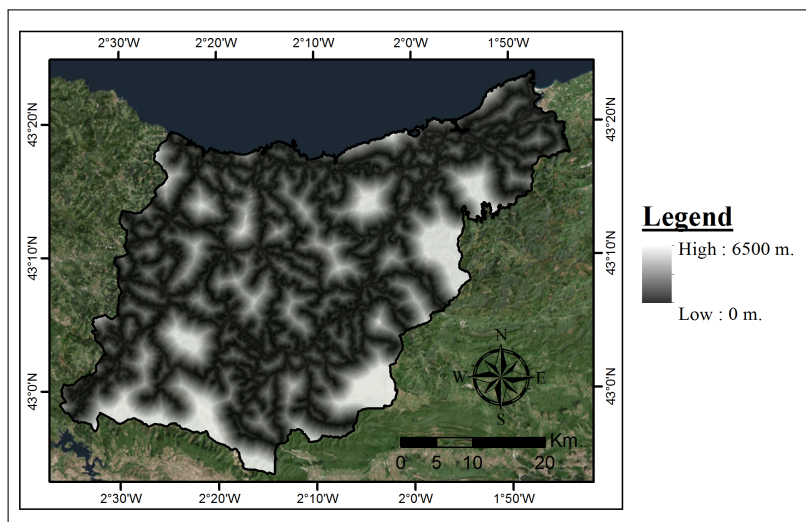
Figure 5.14: *Spatial distribution of distance to the transport network variable.*
5.14 Irudia: Garraio sarearekiko distantzia aldagaiaren banaketa espaziala.

Normalized difference vegetation index (NDVI)

It is about a quantitative and continuous estimate of the vegetation growth and biomass measured by means of the surface reflectance captured by satellite sensors. Using the satellite images in 2.5x2.5 meters of resolution from the SPOT 5 sensors available thanks to the Spanish National Remote Sensing Plan[2], the *NDVI* was calculated applying the following formula, as Yilmaz (2010a) already did:

$$NDVI = \frac{IRC - R}{IRC + R} \tag{5.3}$$

---

[2]ftp.pnt.ign.es

where *IRC* is the near-infrared portion of the electromagnetic spectrum, while *R* is the red portion.

Calculations and spatial resolution re-sampling in order to stay consistent with the rest of the variables were carried out in the ArcGIS 10 software basing on the following two images:

- *Scene ID* 5 035-263 13/09/05 10:22:28

- *Scene ID* 5 035-264 13/08/14 10:46:46

However, the available images only made possible to partially cover the study area, leaving the south-west corner without data. For this reason, this variable could only be used in one part of this study (see section 6-I).

The resulting *NDVI* variable (Fig. 5.15) shows values between -1 and 1, where values below 0.1 correspond to barren areas, sand, or snow. Moderate values represent scrubs and grassland (0.2-0.3), while high values indicate temperate and tropical rainforests (0.6-0.8) (Weier & Herring, 2000).
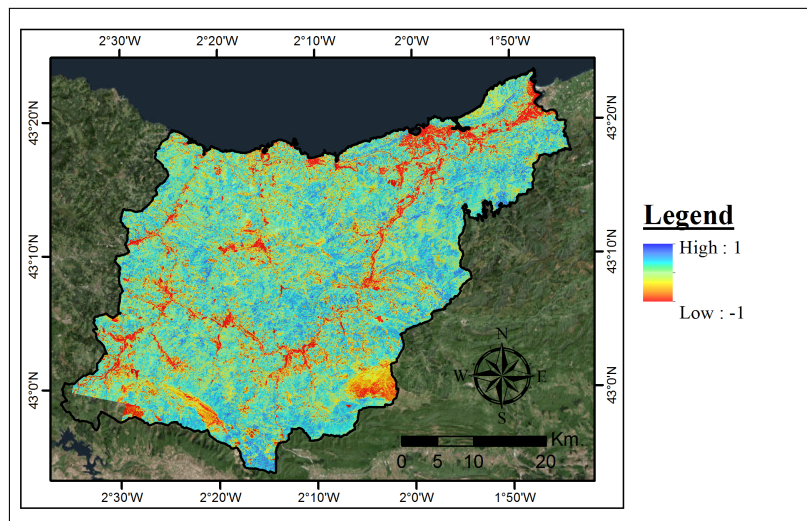
Figure 5.15: *Normalized difference vegetation index (NDVI) variable's spatial distribution.*
5.15 Irudia: NDVI aldagaiaren banaketa espaziala.

5.1.2.2 Categorical variables

Lithology

This is one of the most used variables in landslide susceptibility studies. Moreover, in Felicísimo *et al.* (2013) it was highlighted (by means of a study carried out within a

smaller portion of our study area) that better results of the models can be expected when one takes into account the *lithology*.

Lithology, or rock typology, constitutes the geological substrate of the territory classified according to its composition and its physic-chemical behaviour. The original layer, available in Euskadiko DEA (2014) and named **CT_LITOLOGICO_25000 _ETRS89**, was carried out in 1999 based on the contributions of the Geological Map of Euskadi developed by the EVE (Ente Vasco de Energia)(Fig. 4.3).

According to this map, 21 different rock typologies can be found in study area, but knowing that the resistance against the shear tension as well as the water infiltration capacity could be considered similar in some of those lithologies, the original 21 classes were re-classified in 7 classes, following the expert criteria.

Table 5.2 summarizes the original lithological classes and their simplified reclassification, whose geographical distribution can be observed in the simplified lithological map (Fig. 5.16).
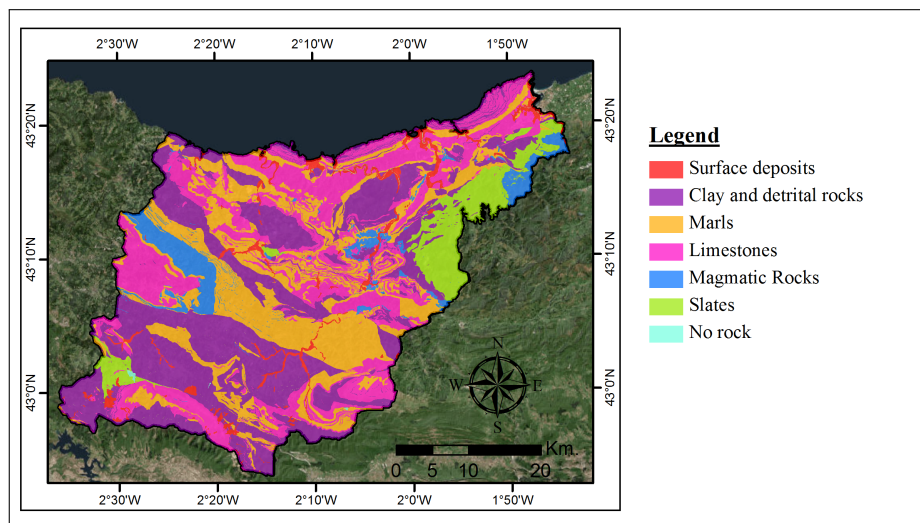
Figure 5.16: *Spatial distribution of the simplified lithological classes.*
5.16 Irudia: Litologia mota sinplifikatuen banaketa espaziala.

Permeability

Permeability refers to the capability of surface rocks or sediments to permit the flow of water through its pore spaces. This feature is directly related to the hydrological response of the slopes allowing the infiltration or the superficial flow of precipitated rain. Consequently, it shows how easy is to reach the ground saturation, and thus, the building up of the pore water pressure, which is considered one of the key conditions

Table 5.2: *Original lithological typologies reclassification table.*

5.2 Taula: Litologia mota originalen birklasifikazio taula.

Original classes	Simplified classes
Dam and rivers	No rock
Surface deposits	Surface deposits
Alternation of detrital rocks Fine-grained detrital rocks (lutites) Mid-grained detrital rocks (limonites) Coarse grained detrital rocks (sandstone)	Clay and Detrital rock
Marls, limestones, marlstones and calcarenites alternation Marls Decarbonated marls Gypsium, clay and other salts	Marls
Impure limestones and calcarenites Limestones Dolomites	Limestones
Granodiorites Coarse grained granite Igneous rocks Dike rocks Ophites Volcanic rock flows Pyroclastic volcanic rocks	Magmatic rocks
Slates	Slate

that favours slope instabilities (Guzzetti *et al.*, 2007). The usage of *permeability* as explanatory variable was more extended on the field of the geotechnical, or deterministic, analysis such as in Cho (2014), though some statistical approaches also applied it in a regional scale, highlighting the difficulty of obtaining such information that covers the entire study area (Duman *et al.*, 2006; Nefeslioglu *et al.*, 2010).

According to Bogaard & Greco (2018), the infiltration capacity of the soil is related more with the type of landslides than with their probability of occurrence, considering that impermeable surfaces increases the superficial flow, what is favourable for shallow landslides, and highly permeable surfaces allow the infiltration of the water, producing more probably deep-seated landslides.

For the current study, the *permeability* layer was obtained from the lithological map of Euskadi. It offers the *permeability* distribution of the study area basing in the porosity and the degree of fissuration associated with the lithologies. It is about a simplified classification in low, medium or high permeability together with impermeable areas and water covered areas (Fig. 5.17).
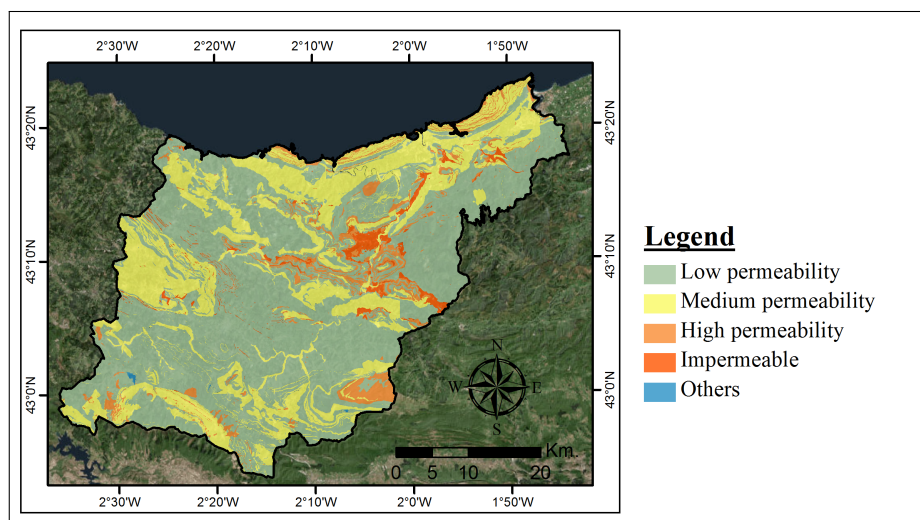
Figure 5.17: *Spatial distribution of the simplified permeability classes.*
5.17 Irudia: Sinplifikatutako permeabilitate moten banaketa espaziala.

Regolith thickness

The *regolith thickness* map of Euskadi is available in Euskadiko DEA (2014) under the code of **CT_ESPESOR_REGOLITO_25000_ETRS89** in shapefile format. The corresponding area to the GP was clipped for the analysis carried out in this work (Fig. 5.18). Notice that the cited map was drawn as an inference of the *lithology*, the

slope and only some punctual direct measures along the studied area. As a result, the territory was divided (apart from the water cover class) in 5 thickness classes: 0-0.5 m; 0.5-1 m; 1-2 m; 2-4 m; more than 4 m.

The more is the thickness of the altered bed rock layer, the more is the amount of material susceptible to be moved. That is way Remondo *et al.* (2003), Felicísimo *et al.* (2013) or Jaiswal *et al.* (2010) used this feature as a possible explanatory variable in their landslide susceptibility studies.
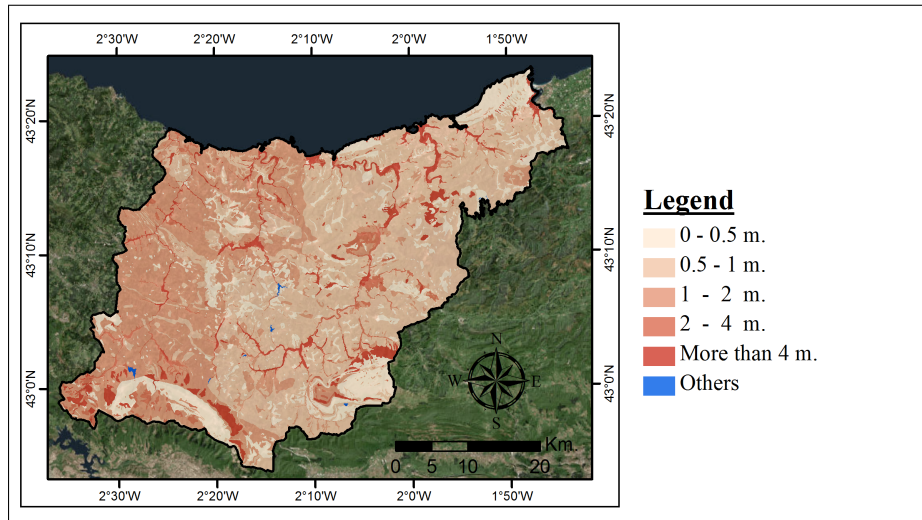
Figure 5.18: *Spatial distribution of the regolith thickness classes.*
5.18 Irudia: Erregolitoaren sakonera klaseen banaketa espaziala.

The surface cover

The *land cover*, or some times only the vegetation, is another typical explanatory variable used on the field of landslide susceptibility assessment. Fell *et al.* (2008) and Van Westen *et al.* (2008) agree that the knowledge of the surface cover is of critical relevance in landslide spatial distribution and, somehow, the most part of the landslide susceptibility studies include at least one variable like that (Atkinson & Massari, 1998; Dai & Lee, 2002; Carrara *et al.*, 2008; Van Den Eeckhaut *et al.*, 2012; Trigila *et al.*, 2015; Wang *et al.*, 2017).

Its relevance respect to the landslide occurrence is, once again, related to the hydrological behaviour of the slope. The roots of a given type of vegetations give to the ground a mechanical protection against the external triggering factors, increasing the resistance to failure, while other land uses leave the surface naked in front of the external agents. Moreover, the differences in the evapo-transpiration capacity of the
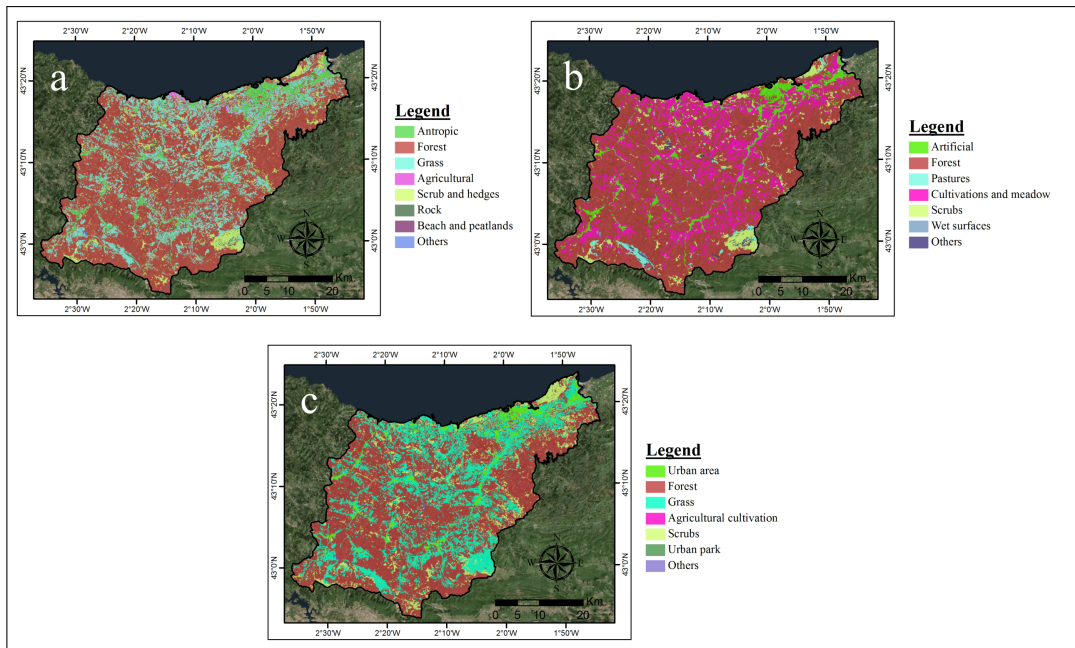
Figure 5.19: *Maps of the different land cover variables. a) Land cover 1; b) Land cover 2; c) Land cover 3.*
5.19 Irudia: Lurzoruaren estaldura aldagaien mapak. a) Lurzoruaren estaldura 1; b) Lurzoruaren estaldura 2; c) Lurzoruaren estaldura 3.

soils are considerable if it is an urbanized or forested land cover, and the impact of the precipitation also changes between non-vegetated and vegetated areas, due to the leafs interception.

Nevertheless, the surface cover could be changed and in some territories such changes can be of a considerable relevance. For this reason, Van Westen *et al.* (2008) suggested the systematic update of the land-use maps with a frequency of 1-10 years, in order to ensure the correct *land cover* class corresponding to each landslide at the moment of its trigger. In our case, despite the landslides occurrence moments were not available, we acknowledge that the land use have changed during the last decades in the study area. So, instead of using only one surface cover layer, that would show the situation of a given moment not necessarily according to the landslide inventory, three different *land cover* maps were considered in order to finally use the most fitting one with our landslide inventory.

All of the following surface cover categorical variables are available in Euskadiko DEA (2014) and they were clipped in order to extract only the area of interest.

Land cover 1: it is the corresponding part to the study area of the National Forest Inventory of the 2010 (IFN 4), which is an update of IFN 3, carried out in

2005. The update was carried out by aerial photo-interpretation of images obtained in 2009 and with 25 cm of spatial resolution. The original 32 classes were re-classified by expert criteria as it is shown in table 5.3 and in figure 5.19 a. The source layer can be found under the code **INV_FORESTAL_2010_10000_ETRS89**.

Land cover 2: it is about an alternative land use layer available as part of the Harmonised Topographical Base (BTA) which summarizes in 7 classes the land use spatial distribution (Fig. 5.19 b). It is based in the CORINE Land Cover map of 2006, and the original classes were maintained for the analysis.

Land cover 3: it is the vegetation map available under the code **CT_VEGETACION_25000_ETRS89**, whose original 11 vegetation typologies were reclassified by expert criteria in 7 simplified classes (Tab. 5.4 and Fig. 5.19 c). This project was the updated version in 2007 of the previous vegetation map carried out during the 1990's decade.

Table 5.3: *Reclassification of the original land use classes of the Land cover 1 variable.*

5.3 Taula: Lurzoruaren estaldura 1 aldagaiaren banaketa espaziala birklasifikatua.

Original classes in *land cover 1*	Simplified classes
Marshes and swamps	
Water streams	
Sea and Oceans	Others
Estuaries	
Dams and lakes	
Industrial	
Other artificial surfaces	
Urban equipments	
Urban continuous	
Urban discontinuous	
Energy infrastructures	
Waste infrastructures	Antropic
Water furnishing infrastructures	
Transport infrastructures	
Telecommunications	
Terciary sectory land use	
Forest	
Plantation forest	Forest
Gallery forest	
Grassland and pastures	Grass
Meadow	
Cultivations	Agricultural
Agricultural mosaic with artificial surface	
Bush	
Pasture and scrubs	Scrubs and hedges
Meadow with hedgerow	
Coastal clifs	
Rock outcrop	Rock
Scree deposits	
Beach, dunes and sand	Beach and peatlands
Peatlands	

Table 5.4: *Reclassification of the original typologies of the land cover 3 variable.*

5.4 Taula: Lurzoruaren estaldura 3 aldagaiaren mota originalen birklasifikazioa.

Original *land cover 3* classes	Simplified classes
No vegetation	
Dam	
Estuary	Others
Erosion	
Grassland	Grassland
Agricultural cultivation	Agricultural cultivation
Forest	
Forest plantation	Forest
Scrubs	Scrubs
Urban	Urban
Urban park	Urban Park

Aspect

According to different authors, the slope *aspect* could be a meaningful variable to influence landslide initiation. On one hand, moisture retention and vegetation is reflected by slope *aspect*, which in turn may affect soil strength and susceptibility to landslides (Dai & Lee, 2002). But on the other hand, in some study areas precipitations present a pronounced directional component by influence of a prevailing wind, where the amount of rainfall falling on a slope may vary depending on its *aspect* (Wieczorek *et al.*, 1997).

For the current study, the slope *aspect* was derived from the DEM by means of the specific tool available in ArcGIS 10 software (*aspect*) for this purpose. As a result the *aspect* value in degrees for each mapping unit was obtained, where both 0° and 360° represent the north, and the value of -1 represents flat areas. In this case, despite its numerical nature, those values does not represent any magnitude which make the treatment of this variable as continuous variable meaningless. For this reason it was re-classified (Fig. 5.20). The layer was divided in 9 classes according to the 8 main cardinal orientations (*North*; *North-East*; *East*; *South-East*; *South*; *South-West*; *West* and *North-West*) plus the *Flat* areas.
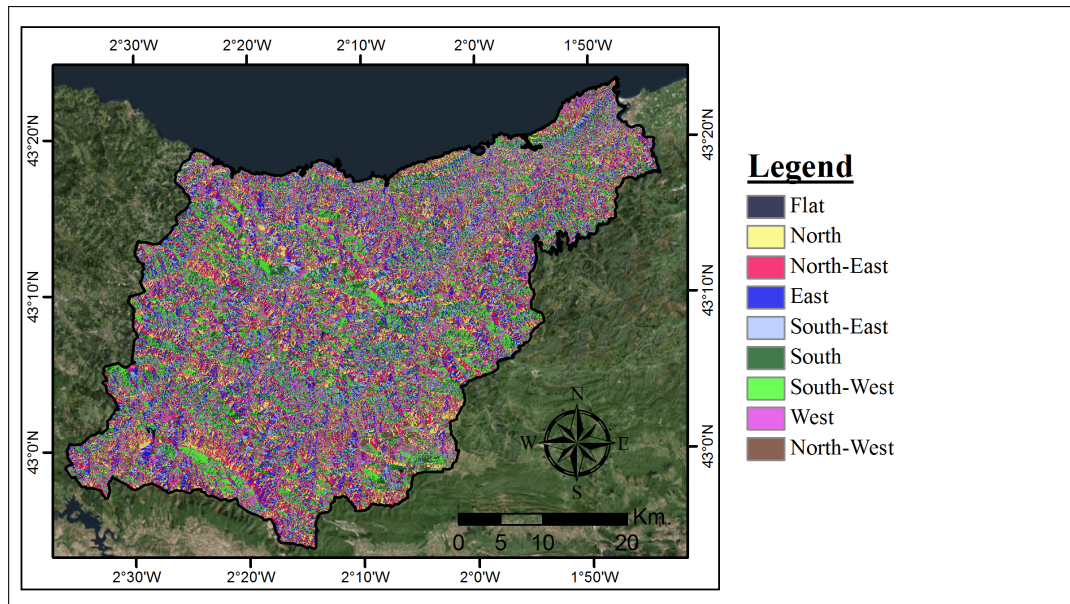
Figure 5.20: *Spatial distribution of the slope aspect classes.*

5.20 Irudia: Malda orientazioa aldagaiaren banaketa espaziala.

Modified variables

Distance to the main river-streams CAT: the original continuous layer was re-classified in 7 categories, following what previously other authors did (Pourghasemi *et al.*, 2012; Xu *et al.*, 2012), and basing on the expert criteria together with observations on the field. Each class represents the range of distance from a given point to the closest main river-stream in meters. The classification was set as follows: *0-50*; *50-100*; *100-150*; *150-200*; *200-250*; *250-300* and *more than 300*.

Distance to transport network CAT: basing on observations on the field, the original continuous layer was re-clasified in 8 classes representing the range of distance in meters from a given point to the closest transport infrastructure as follow: *0-20*; *20-50*; *50-100*; *100-150*; *150-200*; *200-250*; *250-300* and *more than 300*.

5.1.3 Precipitation data

In section 6-III, precipitation data from 2006 to 2015 were used for the landslides responsible precipitation thresholds definition. Such information was collected from the Meteorological Agency of the Basque Country[3] which provided the precipitation records of each 10 minutes for 24 rain gauges within the study area (Fig. 5.21).

---

[3]www.euskalmet.euskadi.eus
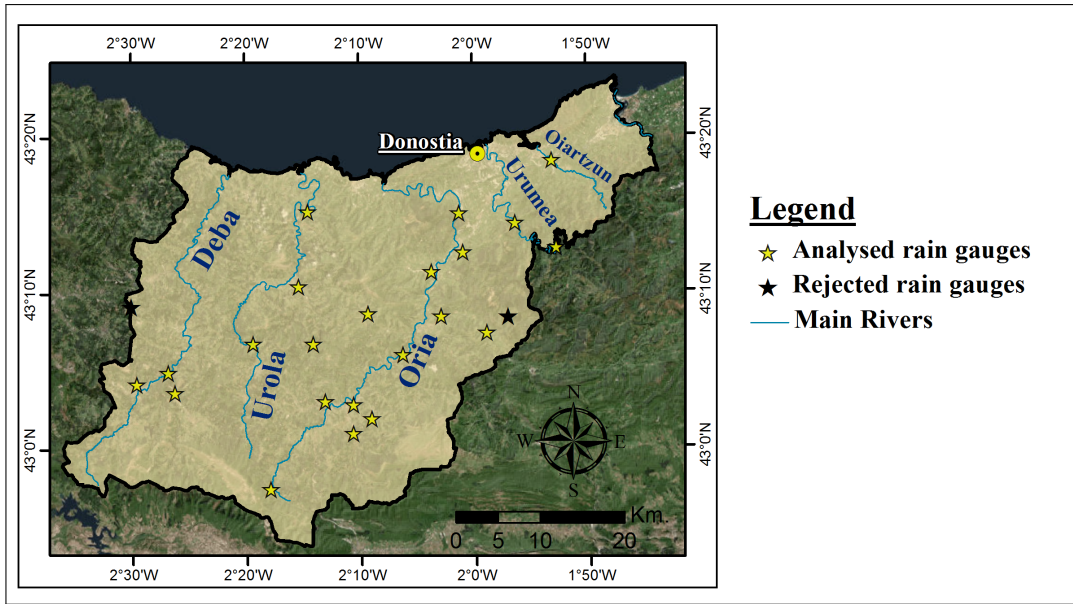
Figure 5.21: *Spatial distribution of the rain gauges used to collect precipitation data.*
5.21 Irudia: Prezipitazio datuak jasotako plubiometroen banaketa espaziala.

Before starting to analyse the precipitation, the original 10 minutes records were verified to ensure the completeness of the data series, and they were transformed into hourly data for this application. In this regards, whenever a rain gauge presented more than one month of missing values, this rain gauge was removed from the analysis.

5.2 Susceptibility models

5.2.1 The logistic regression model

Mathematical models are simplified representations of complex phenomena that aim to explain by means of equations a given data set. Taking into account this statement, it seems logical the fact that the accuracy of the model to simulate the reality depends primarily on (i) the type of mathematical function chosen to explain the data set, and (ii) the degree of representativeness of the data set respect to the reality.

Landslides are geomorphological phenomena that occur worldwide and have occurred during all the earth's history. However, being very local and rapid processes their marks are often disappeared (by natural erosion and sedimentation, vegetation coverage or anthropic reconstruction of slopes), which complicates the

acknowledgement about all the landslides happened during the whole earth's history for a given study area. This becomes hardly quantifiable the degree of representativeness of any data set so, the election of the appropriate model results even more crucial (Zêzere, 2002).

The huge increase of the computational capacity together with the socialization of the statistical software's permeated the development of a considerable range of functions available to model landslide susceptibility (Tab. 2.1 in chapter 2). Among all of then, for the analysis carried out in this thesis the Logistic Regression (LR), by means of the logit function (Hosmer Jr & Lemeshow, 2004), was chosen.

The LR is an statistical and multivariate method which searches to analyse and interpret the data coming from some given observations of $v > 1$ variables over a set of n cases. According to Malamud *et al.* (2014) (Fig. 5.22), this model is the most used for landslide susceptibility mapping, among other reasons, because in this case, through the addition of an appropriate link function to the usual linear regression model, the variables may be either continuous or discrete (categorical) or any combination of both types, and they do not necessarily have normal distributions (Lee *et al.*, 2007). Moreover, several comparative studies in which different models were applied to the same data set concluded the LR as the most appropriate, or at least, one of the most suitable option.

Rossi *et al.* (2010) carried out a comparison between LR, linear discriminant analysis (LDA), quadratic discriminant analysis (QDA) and neural network analysis (NNA) in a 78.9 km^2 study area and arrived to the conclusion that LR and LDA produced superior predictions and less uncertain zonations when compared to the QDA and NNA models. Additionally, they suggested that "*the combination of landslide susceptibility zonations developed by different models can provide optimal susceptibility assessments*".

Amorim (2012) compared the results obtained by applying the LR, LDA and NNA to the same study area of 40 km^2, and concluded that "*the results obtained with the three methods are similar, with the LDA model being generally the best performance, followed closely by the other models*".

In Felicísimo *et al.* (2013), LR, multivariate adaptive regression splines (MARS), classification and regression trees (CART) and maximum entropy (MAXENT)

models were compared. According to the authors, although MAXENT and CART presented the best prediction results, "*the confidence intervals show that MAXENT and LR are the most stable methods, while, CART is the most unstable*".

Model Type	n	% (of Model Types)
Logistic regression analysis	111	17.7%
Neural network analysis	64	10.2%
Data overlay analysis	59	9.4%
Index based analysis	53	8.5%
Fuzzy sets analysis	42	6.7%
Multi-criteria decision analysis	39	6.2%
Probability based analysis	37	5.9%
Weight of evidence analysis	37	5.9%
Statistical analysis of factor of safety	26	4.1%
Linear regression analysis	25	4.0%
Discriminant analysis	23	3.7%
Heuristic analysis	23	3.7%
Bivariate analysis	17	2.7%
Support vector machine	13	2.1%
Tree based analysis	12	1.9%
Dempster-shafer analysis	10	1.6%
Bayesian analysis	8	1.3%
Entropy based analysis	8	1.3%
Regression analysis	6	1.0%
Factor analysis	4	0.6%
Other analysis	10	1.6%
Total # of Model Types Applied (out of 413 articles)	**627**	**100.0%**
Review article (no model type)	10	

Figure 5.22: *Ranking of the most used methodologies for landslide susceptibility modelling according to the review carried out by Malamud et al. (2014).* **n** *is the number of research papers that use a given method.*
5.22 Irudia: Lur labainketa suszeptibilitate modeloak garatzeko erabilitako metodologien urrenkera. Iturria: Malamud *et al.* (2014). **n** metodo jakin bat erabili duten artikulu zientifikoen zenbatekoa da.

5.2.1.1 The logit function

From the mathematical point of view, the LR is a regression method in which the dependant variable is dichotomous, i.e. it can only have two possible results (for example true or false; success or failure; yes or no). This result depends on some given explanatory variables that can contain categorical or continuous values. The objective of the regression is to estimate the parameters of the model in order to best fit the observation set. Unlike in linear regressions (where the least square method is commonly used), in this case the maximum likelihood method is used to carry out the estimate of the model parameters.

What is interesting in binomial models, is not the result of a given value, but how probable is a given outcome respect to the other. This relation is named **odds**, and can be represented as follow,

$$odds = \frac{P}{1-P} \tag{5.4}$$

where P is considered, for example, the success probability and thus, $1 - P$ would be the failure probability. But the problem of this expression is that it is limited to values greater than 0 so, in order to enable working with all the real values the logarithmic transformation is applied to the expression, resulting in,

$$log(odds) = log(\frac{P}{1-P}) \tag{5.5}$$

In such a situation, if it is considered that the probability of success (or failure) can be explained by one or more variables, it could be modelled following the next equation,

$$log(\frac{P}{1-P}) = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n \tag{5.6}$$

which is equivalent to,

$$P = \frac{exp(\beta_0 + \beta_1 X_1 + ... + \beta_n X_n)}{1 + exp(\beta_0 + \beta_1 X_1 + ... + \beta_n X_n)} \tag{5.7}$$

This is known as the **logit** function, whose application permit to model the

probability of a given dichotomous result considering one or more explanatory variables.

In this regards, $exp(\beta_0)$, represents the value of **odds** when the explanatory variable (or variables) Xn takes the value of 0, showing how much more probable is the success respect to the failure when $Xn = 0$. On the other hand, $exp(\beta_n)$, which is the estimate coefficient, represents the **odds ratio** per increased unity of the Xn variable. Thus, the estimates show how much increases (or decreases) the probability of success according to the increase (or decrease) of the explanatory variable.

However, it has to be pointed out that being exponential functions, positive values of the β_n estimates, means an **odds ratio** grater than 1, which imply a rise in the success probability, while negative values signifies an **odds ratio** lower than 1, and so, the greater probability of the failure.

In order to model the landslide spatial susceptibility, the dichotomous variable considered in this study was the presence or absence of landslides, encoded

respectively with 1 and 0 values. So as that, applying the **logit** function, the result would indicate the $P_{(1)}$ probability, i.e. the probability of landslide presence.

5.2.1.2   Statistical software

As it was previously stated, the development of the statistical software allowed the experimentation of landslide susceptibility modelling by means of a huge range of mathematical options. The LR analysis could not be done without the computational systems support, so as that the β_n estimates resolution, as well as many other calculations, run during this work were carried out by two of the most powerful statistical packages available: **IBM SPSS Statistics** (SPSS, 2011) and **R Project for Statistical computing** (R Core Team, 2016).

The usage of those tools, allowed also, apart from the general objectives of this thesis, to highlight operational advantages and drawbacks between them.

IBM SPSS Statistics is about a commercial software which offers to the user a large range of statistical procedures ready to apply and largely tested and validated, by means of a users friendly interface. Apart from the analytical outcomes, it also gives the option to very easily plot and export the results in graphic and table format.

R Project for Statistical computing is a free software environment for statistical computing and graphics which also offers all kinds of statistical approaches. Unlike SPSS, this tool presents a very simple interface in which all the processes are called by command sequences, which allows the design of customized codes.

5.2.2   Assessment of the bibliographical landslide inventory

Regarding to the landslide inventory coming from bibliographical sources (see section 5.1.1.1), a preliminary assessment was carried out in order to evaluate the accuracy level of them. To do so, the Oria river basin (see section 6.1.1) was selected as a reduced portion of the study area and a small set of landslides (around 10 %) was chosen at random for checking them on the field.

With the help of the GPS, the exact location of 23 landslides coming from the bibliographic review were visited and the presence of slope instabilities or signs of

instabilities occurred in the past such as convex shapes resulting from a landslides, gravitational deposits or contention measures were surveyed.

In the meanwhile, all the landslides found during the same field trips were inventoried and documented. Thereby, later it was ascertained if those "newly" inventoried landslides were already part of the bibliographic inventory or not. Results of this assessment are shown in section 6.2.

5.2.3   Methods for explanatory variables selection

As it was previously mentioned, the selection of the explanatory variables with major role is a difficult task because neither universal criteria nor guidelines exists about this issue. Depending on the statistical method adopted to calculate the landslide susceptibility, one can choose among different options such as the principal component analysis (Baeza & Corominas, 2001), stepwise approach (Brenning, 2005) or more sophisticated statistical analysis like in Lombardo *et al.* (2016). In this work, simplified and statistically oriented work flows were proposed in two alternative applications (see sections 6-I and 6-II), where the usage of only significant variables was ensured as well as the non-redundancy of the contributed information by each covariate.

Nevertheless, although the rationale behind both approaches was always the fulfilment of the conditions cited in section 5.1.2, the discover of new tools along the development of this thesis allowed to afford this question in two slightly different ways for each step of the project. Hence, next lines are dedicated to the detailed explanation of each variables selection approach applied in the Oria river catchment and in Gipuzkoae Province.

5.2.3.1   Variables selection approach applied in the Oria river catchment

In this experiment, all the statistical calculations were carried out by the SPSS XXII package, and before running the LR model, some descriptive statistics were computed individually for each variable. Once the dependant variable, i.e., landslide presence and absence sample, was prepared, the significance level of each categorical variable respect to the dependant variable was computed. This was done by means of the Chi-Square (Chi^2) test, which tabulates a variable into categories and

calculates the Chi^2 statistic based on the differences between observed and expected frequencies (IBM Corporation, 1988). All classes with less than 5 cases were removed to avoid over-estimations (Rana *et al.*, 2015). The starting hypothesis assumes equal expected frequencies among the categories. Accordingly, values under 0.05 signifies the rejection of the hypothesis, what means that the dependant variable is unequally distributed among the categories of the analysed variable, thus this variable can be considered significantly relevant respect to the distribution of the presence and absence of landslides.

In a similar way, the significance level of each continuous variables respect to the dependant variable was also computed. However, depending on the distribution of the continuous variable (normal or non-normal) two different statistics are usually suggested (Pardo & Ruiz, 2002). So, first the Kolmogorove-Smirnof ($K-S$) test was applied, where values under 0.05 implies that the variable does not follow a normal distribution, and vice versa. According to De Winter & Dodou (2010), if a variable presents a normal distribution, then the t-Student test should be performed. But, if it does not present a normal distribution, then the Mann-Whitney test should be computed. Consequently, we applied this rule, but in any case, values under 0.05 would imply the significant relevance of the continuous variable respect to the distribution of the presence and absence of landslides.

Then, those variables with significance values above 0.05 were rejected, and the independence of the remaining variables was tested by means of the correlation test of Spearman, which is a non-parametric test that allows to highlight associations between variables (Pardo & Ruiz, 2002). Thus, it was considered that two variables were highly correlated if their correlation coefficient overcame the absolute value of 0.5 with a significance level of 0.01. In such a case, this couple of variables would not be introduced together into the LR analysis, because we would be introducing redundant information.

So, knowing the significance level as well as the existing correlation between all variables, different combinations of explanatory variables were tested in the LR using the backward Wald stepwise method (Pardo & Ruiz, 2002). The software builds the equation starting with all variables and then removes them one by one if their Wald statistic significance value is higher than 0.1. After trying with all

possible combinations, the most suitable set of variables was selected basing on the classification results and the discarded variables in each case by the backward Wald method.

5.2.3.2 Variables selection approach applied in Gipuzkoa Province

During this experiment, the variable's selection procedure was also carried out searching an objective way to choose only the most relevant explanatory variables, regarding the landslide spatial distribution, while the independence between them was ensured. But, this time, the LAND-SE software was used for all the calculations (Rossi & Reichenbach, 2016). It is about an open source code developed in R environment (R Core Team, 2016), designed for the specific purpose of landslide susceptibility statistical assessment. Its structure in command lines allowed to modify the original version of the code, so an additional module, called LAND-SVA, was developed in order that the software gave statistical descriptive information about the introduced explanatory variables. This new module allowed, among other things, to compute Spearman's correlation coefficients and to plot them graphically (information about the availability of these codes is available in Appendix B).

This tool permitted carrying out the following approach, and furthermore, it presents the advantage of optionally automatize the whole procedure.

To begin, all the available variables were introduced into the LR analysis using the modified LAND-SE software, and then, their pairwise collinearity was checked. Once again, we considered collinear two variables when their correlation coefficient was greater than 0.5 with a significance level of 0.01. At the same time, the significance p-values of the LR estimates (see section 6.8.1) were also considered for each variable. Values higher than 0.05 indicate a weak contribution of the explanatory variable to the model performance. So, in such a case, the variable would not be considered statistically significant and it would be removed from the analysis (Schlögel *et al.*, 2018). Conversely, in statistical terms, those predictors with p-values under the threshold of 0.05 were all significant. Thus, the p-value was used as an objective indicator for the selection of the most relevant variables to be used in the second run of the LR model. That is, only the variable with lowest p-value would be selected in case of pairwise collinearity.

5.2.4 Susceptibility model's validation procedures

As far as we are using simplified representations in order to predict complex phenomena like landslides, it is compulsory to evaluate the performance of every model, before considering it as a faithful representative of the real phenomena. That is why the validation procedure is considered one of the most important steps in landslide susceptibility mapping (Duman *et al.*, 2006).

In this case, thanks to the LR, it was obtained a mathematical function that was supposed to be able of simulating the spatial distribution of the future landslides according to a probability of its occurrence. And consequently, the assessment of the resulting susceptibility maps should be done considering two decision rules (Can *et al.*, 2005):

- On the map, most of the future landslides should have to be located in areas with high probability of landslide occurrence i.e., in high susceptibility classes.

- On the map, these high susceptibility classes should have to cover smaller areas than low susceptibility classes. Because, if high susceptibility classes cover large areas, all described landslides will be, logically, included within them.

In order to check the fulfilment of one, or even both, decision rules, different validation tests were applied to the models carried out during the current thesis. But whatever was the test, there was always ensured the independence of the validation data against the calibration data. Because, quoting Frattini *et al.* (2010), the correct assessment of the model accuracy should be performed by analysing the agreement between the model results and the observed data. And, in the case of landslide susceptibility models, the observed data comprise the presence or absence of landslides within a certain terrain unit of the studied area.

This agreement can be performed considering different classification features when obtaining the independent landslide set. One of the most used option is the random selection of a given percentage of landslides up to the complete inventory, reserving them for the validation procedure. This way, validation data are scattered within the same domain as the calibration data, that is why some authors like Brenning (2005) named it *spatial intra-domain* validation. Other option could be

the *spatio-temporal intra-domain* validation, which consist in test data within the same spatial domain of the training data, but taking the landslides occurred previous to a given date for training, and those produced later as validation (Remondo *et al.*, 2003). The inconvenient in this case is the difficulty to obtain temporal information about the slope failures. Additionally, there is a third option named *spatial extra-domain* validation, that performs the assessment of a model calibrated in a given area with validation landslides occurred in a different area. But the usage of such approach should be justified with a considerable similitude between both, training and validation areas (Domínguez Cuesta, 2003).

The lack of sufficient temporal information about our landslide inventory made no possible the adoption of the *spatio-temporal intra domain* approach, although authors like Chung & Fabbri (2003) strongly recommend it. At the same time, the moderate spatial heterogeneity due to the dimensions of our study area suggested to avoid using the *spatial extra-domain* approach. So, for all the landslide susceptibility models developed in further analysis the *spatial intra domain* approach was adopted. Below are detailed all the validation tests applied to the resulting models.

Cumulative percentage curves: this is a validation test proposed by Duman *et al.* (2006) where two curves are drown according to the landslide susceptibility classes (x axis) and the cumulative percentage (y axis). One curve (curve-a in Fig. 5.23), represents the landslide susceptibility class versus observed cumulative percentage of the number of mapping units that include only landslides within this class. It shows how well results satisfy the first decision rule previously cited. For example, in figure 5.23, it can be stated that considering the cut-off value of 0.5, approximately 80% of the observed landslides (validation landslides) are locate in the high susceptibility values. The other curve (curve-b in Fig. 5.23), represents landslide susceptibility class versus cumulative percentage of the number of mapping units representing the same landslide susceptibility class. It defines the areal distribution of susceptibility classes in the studied area so, in order to satisfy the second decision rule, the cumulative area obtained from this curve for high susceptibility classes should be small as possible. In the same example, considering the cut-off value of 0.5, the cumulative area of the high susceptibility values is obtained as 37%, which is considered as satisfactory result in Duman *et al.* (2006).
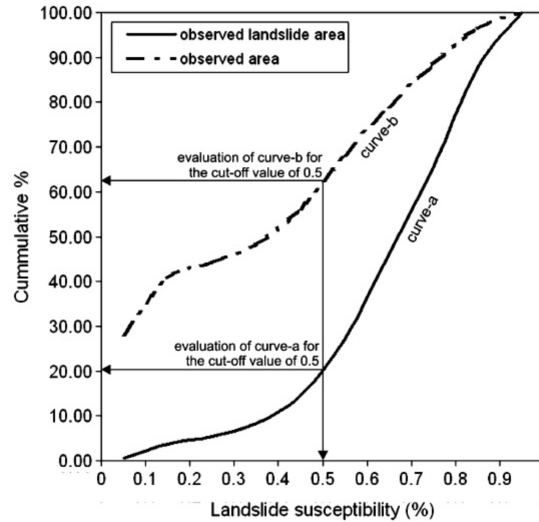
Figure 5.23: *Example of a cumulative percentage curves obtained from Duman et al. (2006).*

5.23 Irudia: Akumulatutako portzentaien kurbaren adibide bat. Iturria: Duman *et al.* (2006).

Prediction rate curve: Chung & Fabbri (2003) proposed to plot the cumulative percentage of validation landslides occurred within each predicted susceptibility class, inversely ordered (y axis), together with the cumulative percentage area covered by each susceptibility class (x axis) (see Fig. 6.6 in section 6.4). Thereby, if a landslide susceptibility map was generated randomly, then the prediction-rate curve should be the straight line connecting two points, (0, 0) and (1, 1), showing an area under the curve (AUC) of 0.5. On the other hand, if the prediction has any significance, then the prediction rate curve should be far above the straight line, showing an AUC closer to 1, the higher it is its prediction capacity. Hence, in a random prediction, if the most susceptible class cover only 10% of the area, then it should predict only 10% of the validation landslides, whereas in a satisfactory susceptibility map the most part of the validation landslide should be within this class. This validation test has widely been used in landslide susceptibility modelling (Bonachea *et al.*, 2009; Amorim, 2012; Trigila *et al.*, 2015; Estela *et al.*, 2018).

Confusion matrices: they are very commonly applied error tests in mathematical modelling, in which the percentage of the true positive (TP), true negative (TN), false positive (FP) and false negative (FN) are computed. It is named TP when the modelled result and the observed result (validation data) are positive, whereas, if both, prediction and validation results are negative, it is called

TN. When the prediction shows a positive result, but observations show negative, then it is about a FP, and inversely, if the prediction launch a negative result, while observations indicate the opposite, it is FN. On the field of the landslide susceptibility modelling, susceptibility values that overcome a given cut (usually 0.5) are usually considered as indicator of presence of landslides and considered positive results, whereas low susceptibility values (usually under 0.5 cut of value) are interpreted as indicator of absence of landslides, and thus, are considered negative results. Hence, this validation test, unlike the previously mentioned ones, takes under consideration not only how well predicts each model the presence of landslides, but also the absence of them. Thereby, the outcomes of this test can be displayed as a matrix, with absolute or percentage values, or as four-fold plots (also called contingency plots) (Rossi *et al.*, 2010).

Receiver operating characteristic (ROC) curves: they are two-dimensional graphs in which TP rate is plotted on the y axis and FP rate is plotted on the x axis (see Fig. 6.11 in section 6.9.1). A ROC graph depicts relative trade-offs between benefits (true positives) and costs (false positives) (Fawcett, 2006). As well as in prediction rate curves, the area under the curve is usually calculated in order to reduce ROC performance to a single scalar value representing expected performance, and thereby, allow the numerical comparison between several models. That is the case of the most part of landslide susceptibility models carried out by statistical multivariate methods such as Carrara *et al.* (2008); Amorim (2012); Trigila *et al.* (2015); Yilmaz (2009); Van Den Eeckhaut *et al.* (2012) or Schlögel *et al.* (2018), among others.

Cohen's Kappa index (κ): it is another commonly used test (Guzzetti *et al.*, 2006; Van Den Eeckhaut *et al.*, 2009; Rossi *et al.*, 2010), also called Heidke skill score (Cohen, 1960; Corominas & Mavrouli, 2011). It is about a combination of correct and incorrect classified positives and negatives results as follow:

$$\kappa = \frac{P_C - P_E}{1 - P_E}, \qquad \{-\infty < \kappa < 1\} \tag{5.8}$$

where, P_C is the proportion of observations correctly classified as presence or absence of landslides by the model: $P_C = (TP + TN)/N)$, and P_E is the proportion

of observations for which the agreement is expected by chance: $P_E = ((TN + FN)x(TN + FP) + (TP + FP)x(TP + FN))/N^2$, where N is the total number of observations $N = TP + TN + FP + FN$ (Rossi *et al.*, 2010). Thereby, according to this calculations, higher values also indicate a more accurate prediction capacity.

Model error plots: to estimate the uncertainty associated with the landslide susceptibility value assigned to each mapping unit, it is possible to run multiple instances of the model varying, randomly, the input data. In each run, the input is obtained sampling the original training dataset (i.e. the landslide inventory sample reserved for calibration) with a bootstrap technique (Rossi & Reichenbach, 2016), which means a random sampling with replacement. Thereby, model error plots summarize the distribution of multiple results and show the mean probability estimate of landslide spatial occurrence for each mapping unit (x-axis), ranked from low (left) to high (right) values, related to the variation of the model estimate (y-axis), measured by 2 standard deviations (2σ) of the probability estimates obtained by the different model runs (Guzzetti *et al.*, 2006). Additionally, the parabolic model fitting (i.e. using a non-linear least square method) of the resulting point cloud describes analytically the overall model variability (see Fig. 6.11 in section 6.9.1).

Mismatch maps: additionally to the statistical or numerical performance indicators, a graphical test was applied to some susceptibility maps, in order to objectively compare the differences respect to the spatial distribution of the susceptibility along the study area (Amorim, 2012). To do so, each mapping unit was re-classified as stable or unstable considering the cut-off value of 0.5 of landslide occurrence probability, and then, the compared maps were overlapped so as to identify matches and mismatches (see Fig. 6.14 in section 6.10). Thereby the mismatch degree between different susceptibility maps was quantified in terms of amount of mismatched mapping units, and also considering the overall mismatched area. Although this test does not show which of the models performs the best prediction capacity, it is useful to detect changes (or agreements) in the spatial distribution of the landslide susceptibility between two different approaches performing similar prediction capacity.

5.2.5   Mapping units

The selection of an appropriate terrain subdivision is another critical phase in landslide susceptibility analysis. The land surface can be divided in portions following geomorphologic features using terrain units, topographic units, geo-hydrological units or slope units, but also considering thematic layers resulting in unique condition units or administrative units, as well as regular grid cells partitions (Van Den Eeckhaut *et al.*, 2006). As Carrara *et al.* (2008) pointed out, selection of different mapping units can result in considerable differences in the susceptibility assessment.

According to Hansen (1984), a mapping unit is defined as the portion of land surface which contains a set of ground conditions which differ from the adjacent units across definable boundaries. Two of the most extended mapping units used in landslide susceptibility modelling are regular grid cells (pixels) and slope units (SU), so in this thesis, both pixels and SU were tested to prepare divers susceptibility maps. More detailed information about the issue of defining proper mapping units for different natural hazard analysis was widely discussed in Carrara *et al.* (1995) and more recently in Reichenbach *et al.* (2018).

5.2.5.1   Regular grid cells

Regular grid cells consist in the subdivision of the region into pixels characterized by their size, typically but not necessarily coinciding with the digital elevation model (DEM) grid cells, which are used as reference mapping units. This approach presents operational advantages in raster-based GIS applications and it allows fast processing. That is why the most part of GIS driven analysis used this mapping unit (Lee, 2005; Godt *et al.*, 2008; Trigila *et al.*, 2010; Grozavu *et al.*, 2013).

Nevertheless, since grid-cell boundaries do not bear any relation to geological, geomorphological or other environmental features, some authors argued that this subdivision is not the most suitable for mapping and modelling geomorphological landforms and processes (Carrara *et al.*, 1991; Van Den Eeckhaut *et al.*, 2009; Alvioli *et al.*, 2016).

5.2.5.2   Slope units

The other terrain subdivision technique used in this thesis was based on the partition of a region into sub-basins or slope units. Since a clear physiographic relation exists between landsliding and the fundamental morphological elements of a hilly or mountainous region, namely drainage and divide lines, this technique seems appropriate for landslide susceptibility assessment. Nevertheless, the difficulty of consistently drawing divide lines on topographic maps covering large areas calls for an automatic procedure for their delineation. Moreover, many different slope unit subdivisions can be obtained for the same territory depending on the type and degree of required homogeneity (i.e., *slope* gradient, *elevation, curvature, aspect*). Hence, it is important to take into account the slope unit sizes to be consistent with the characteristics of the landslide inventory and of the study area. A more detailed discussion about such mapping units and their application in landslide susceptibility studies can be found in Carrara *et al.* (1995); Alvioli *et al.* (2016); Schlögel *et al.* (2018).

5.3   Precipitation thresholds

In the case of the study area, rainfalls are the principal triggering factor of landslides (Remondo *et al.*, 2005; Bonachea, 2006; Felicísimo *et al.*, 2013; Remondo *et al.*, 2017) so, precipitation data and temporal information about landslide occurrence were used in order to determine how much and in which conditions should rain to trigger landslides.

In Chapter 6-III, it is presented the application of a methodology to search qualitative relations between slope instabilities and rainfall events, together with the definition of landslides responsible precipitation thresholds. In this regards, details about data and methods followed in this experiment are explained in the following sections.

5.3.1 The algorithm for the objective reconstruction of rainfall events and precipitation thresholds calculation

Once the precipitation data set was checked, it was carried out the reconstruction of the rainfall events that took place in Gipuzkoa Province between 2006 and 2015, as well as the calculation of the precipitation thresholds responsible for landslides by means of the algorithm proposed in Melillo *et al.* (2015) (information about the repository in which the original code can be found is available in Appendix B). According to the same authors, a rainfall event was defined as a period of continuous rainfall or a chronological ensemble of periods of continuous rainfall, separated from preceding and successive rainfall events by periods with no rainfall.

For this case study, two seasonal periods were defined: (1) a "dry" period, when there is a generalized decrease of the precipitation, between the months of June and August (Urrestarazu & Galdos, 2008), and (2) a "wet" period, when frontal systems which traverse the study area are more frequent and precipitation probability higher, between September and May (Uriarte, 1996). Accordingly, the minimum dry period to differentiate between independent rainfall events for each season was set, by default, to 48 hours during "dry" season and 96 hours during "wet" season (Melillo *et al.*, 2015).

The algorithm gives the metrics for each of the detected rainfall event, including the rainfall duration D (h); the event total cumulated precipitation E (mm); the mean rainfall intensity I (mm/h); the peak hourly rainfall intensity I_p (mm/h) and the maximum precipitation in 24 hours E_{max24} (mm).

Additionally, as it is explained more in detail by Brunetti *et al.* (2010) and Peruccacci *et al.* (2017), an empirical definition approach was adopted for the landslides responsible precipitation thresholds calculation. Each landslide from the inventory was linked to the closest rain gauge and the metrics of the rainfall event related to each instability were used for further calculations. Applying the Frequentist methodology (Brunetti *et al.*, 2010), the algorithm gives different precipitation threshold curves for different exceedance probability levels, which assuming that the catalogue of rainfall events is sufficiently complete and representative, it can be stated that the probability of experiencing a landslide

triggered by rainfall below this threshold is less than the given exceedance value.

The Frequentist method consists in plotting the log-transformed values of cumulated precipitation amount (E) respect to the duration (D) for each known rainfall event that caused a landslide. In this regards, only the cumulated precipitation and the duration until the landslide occurrence exact time are considered. Thus, although more than one landslide could be related to the same rainfall event, they do not necessarily have to present the same E and D coordinated, unless they happened at the same exact time. Later on, the distribution of the rainfall conditions, $\log(E)$ vs. $\log(D)$, that resulted in landslides is fitted (least square method) with a linear equation of the type (Brunetti *et al.*, 2010),

$$Log(E) = Log(\alpha) + \gamma \cdot Log(D) \tag{5.9}$$

or which is equivalent,

$$E = \alpha \cdot D^{\gamma} \tag{5.10}$$

Then, the precipitation threshold for a given exceedance level T_i corresponds to the parallel of the fitting curve in which the i % of the observations stay below the curve.

Finally, basing on the rainfall duration and the total cumulated precipitation for the landslide associated events, the definition of thresholds and their associated uncertainties for different exceedance probabilities are provided applying the bootstrapping statistical technique (Peruccacci *et al.*, 2012).

5.3.2   Landslides and rainfalls characterization

In order to check the features of the landslides that were reported on the newspaper during the analysed period of time, the frequencies of slope movements belonging to each class of *slope*, *lithology* and *land cover* were calculated using the Quantum GIS 2.14.7 software (QGIS Development Team, 2009). For this analysis it was necessary to ensure the location accuracy of the landslides, thus, this characterization was carried out only taking into account those landslides where the exact location was known.

On the other hand, a descriptive analysis of the total reconstructed rainfall conditions was also carried out in order to compare them with the landslide

associated rainfall conditions. To begin, the rainfall events were classified in different typologies following two different approaches. The first one defines the type of rainfall depending on the maximum precipitation in 24 hours as it was proposed by Alpert *et al.* (2002) (Tab. 5.5). The second is a classification developed by the authors which considers the combination of different rainfall duration classes and different cumulated precipitation classes (Díaz *et al.*, 2012) (Tab. 5.6). This alternative classification was proposed because it was considered interesting the characterization of the rainfall conditions taking into account the totality of each event and not only the maximum intensity. This way, complementary information was added for further analysis.

Then, the relative frequencies of each rainfall type were presented for the total rainfall events detected in the study area during the studied period, and also for the rainfalls that cause landslides, in order to observe the differences.

Table 5.5: *Rainfall classification suggested by Alpert et al. (2002). E_{max}: the maximum precipitation in 24 hours in mm.*

5.5 Taula: Alpertek proposatutako euriteen klasifikazioa Alpert *et al.* (2002). E_{max}: 24 ordutan prezipitatutako maximoa mm-tan.

P. Alpert's classification

$E_{max}24$ *(mm)*	*Class*
≤ 4	Light
4 to 16	Light - Moderate
16 to 32	Moderate
32 to 64	Moderate - Heavy
64 to 128	Heavy
128 to 256	Heavy - Torrential
$256 \geq$	Torrential

Table 5.6: *Rainfall classification proposed by the authors. D: Duration of the rainfall event in hours; E: Cumulated precipitation in mm.*

5.6 Taula: Autoreek proposatutako euriteen klasifikazioa. D: Euritearen iraupena orduetan; E: Akumulatutako prezipitazioa mm-tan.

Authors classification

D (h)	*Class*	*E (mm)*	*Class*
≤ 24 (1 day)	a	≤ 15	A
24 to 72 (1 - 3 days)	b	15 to 30	B
72 to 144 (3 - 6 days)	c	30 to 60	C
144 to 288 (6 - 12 days)	d	60 to 120	D
288 to 432 (12 - 18 days)	e	120 to 180	E
431 to 719 (18 - 30 days)	f	$180 \geq$	F
$720 \geq$ (30 days)	g		

References

Akgun, A.: A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey, Landslides, 9, 93–106, 2012.

Alpert, P., Ben-Gai, T., Baharad, A., Benjamini, Y., Yekutieli, D., Colacino, M., Diodato, L., Ramis, C., Homar, V., Romero, R., *et al.*: The paradoxical increase of Mediterranean extreme daily rainfall in spite of decrease in total values, Geophysical Research Letters, 29, 1–4, 2002.

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F.: Automatic delineation of geomorphological slope units with *r.slopeunits v1.0* and their optimization for landslide susceptibility modeling, Geoscientific Model Development, 9, 3975–3991, 2016.

Amorim, S. F.: Estudio comparativo de métodos para la evaluación de la susceptibilidad del terreno a la formación de deslizamientos superficiales: Aplicación al Pirineo Oriental, Ph.D. thesis, Universidad Politécnica de Catalunya, Barcelona, 2012.

Atkinson, P. M. & Massari, R.: Generalised linear modelling of susceptibility to landsliding in the central Apennines, Italy, Computers & Geosciences, 24, 373–385, 1998.

Ayalew, L. & Yamagishi, H.: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan, Geomorphology, 65, 15–31, 2005.

Baeza, C. & Corominas, J.: Assessment of shallow landslide susceptibility by means of multivariate statistical techniques, Earth Surface Processes and Landforms, 26, 1251–1263, 2001.

Biswajeet, P. & Saro, L.: Utilization of optical remote sensing data and GIS tools for regional landslide hazard analysis using an artificial neural network model, Earth Science Frontiers, 14, 143–151, 2007.

Bogaard, T. & Greco, R.: Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Natural Hazards and Earth System Sciences, 18, 31–39, 2018.

Bonachea, J.: Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debidos a procesos geomorfológicos, Ph.D. thesis, Universidad de Cantabria, Santander, 2006.

Bonachea, J., Remondo, J., Terán, D., Díaz, J. R., González Díez, A., & Cendrero, A.: Landslide risk models for decision making, Risk Analysis, 29, 1629–1643, 2009.

Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation, Natural Hazards and Earth System Science, 5, 853–862, 2005.

Brunetti, M., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F.: Rainfall thresholds for the possible occurrence of landslides in Italy, Natural Hazards and Earth System Science, 10, 447–458, 2010.

Budimir, M., Atkinson, P., & Lewis, H.: A systematic review of landslide probability mapping using logistic regression, Landslides, 12, 419–436, 2015.

Can, T., Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H., & Duman, T. Y.: Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses, Geomorphology, 72, 250–271, 2005.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., & Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes and Landforms, 16, 427–445, 1991.

Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P.: GIS technology in mapping landslide hazard, in: Geographical Information Systems in Assessing Natural Hazards, edited by Carrara, A. & Guzzetti, F., pp. 135–175, Springer, 1995.

Carrara, A., Crosta, G., & Frattini, P.: Comparing models of debris-flow susceptibility in the alpine environment, Geomorphology, 94, 353–378, 2008.

Cho, S. E.: Probabilistic stability analysis of rainfall-induced landslides considering spatial variability of permeability, Engineering Geology, 171, 11–20, 2014.

Chung, C. J. & Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping, Natural Hazards, 30, 451–472, 2003.

Cohen, J.: A coefficient of agreement for nominal scales, Educational and Psychological Measurement, 20, 37–46, 1960.

Corominas, J.: Landslides and climate. Keynote lecture, in: 8th International Symposium on Landslides, edited by Bromhead, E and Dixon, N and Ibsen, ML), vol. 4, pp. 1–33, A.A. Balkema, Cardiff, 2000.

Corominas, J. & Mavrouli, O. C.: Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies, Tech. rep., SafeLand. 7th Framework Programme Cooperation Theme 6 Environment (including climate change) Sub-Activity 6.1.3 Natural Hazards, 2011.

Cruden, D. M. & Varnes, D. J.: Landslide types and processes, in: Landslides: Investigation and Mitigation, edited by Turner, A. K. & Jayaprakash, G., 247, Transportation Research Board Special Report, Washington D.C., 1996.

Cuesta, M. J. D., Sánchez, M. J., & García, A. R.: Press archives as temporal records of landslides in the North of Spain: relationships between rainfall and instability slope events, Geomorphology, 30, 125–132, 1999.

Dai, F. & Lee, C.: Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, Geomorphology, 42, 213–228, 2002.

De Winter, J. C. & Dodou, D.: Five-point Likert items: t test versus Mann-Whitney-Wilcoxon, Practical Assessment, Research & Evaluation, 15, 1–12, 2010.

Díaz, E., Sáenz de Olazagoitia, A., Ormaetxea, O., & Ibisate, A.: Análisis de factores de desestabilización de laderas en dos cuencas del ámbito atlántico: Sollube-mape (Bizkaia) y Ramaio (Alava), Cuaternario & Geomorfología, 26, 171–190, 2012.

Domínguez Cuesta, M. J.: Geomorfología e inestabilidad de laderas en la Cuenca Carbonífera Central (Valle del Nalón, Asturias). Análisis de la susceptibilidad ligada a los movimientos superficiales del terreno, Ph.D. thesis, Universidad de Oviedo, Oviedo, 2003.

Duman, T. Y., Can, T., Gokceoglu, C., Nefeslioglu, H. A., & Sonmez, H.: Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey, Environmental Geology, 51, 241–256, 2006.

Estela, T. B., Antigüedad, I., & Ormaetxea, O.: Mapas de susceptibilidad de deslizamientos a partir del modelo de regresión logística en la cuenca del río Oria (Gipuzkoa). Estrategias de tratamiento de variables, Cuaternario & Geomorfología, 32, 7–29, 2018.

Euskadiko DEA: Infraestructura de datos espaciales de Euskadi, URL `www.geo.euskadi.eus`, 2014.

Fawcett, T.: An introduction to ROC analysis, Pattern Recognition Letters, 27, 861–874, 2006.

Felicísimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10, 175–189, 2013.

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, Engineering Geology, 102, 85–98, 2008.

Frattini, P., Crosta, G., & Carrara, A.: Techniques for evaluating the performance of landslide susceptibility models, Engineering Geology, 111, 62–72, 2010.

GFA: Evaluación y gestión integada de riesgos geotécnicos en la red de carreteras de la Diputación Foral de Gipuzkoa, Tech. rep., Mugikortasun eta Bide Azpierituren Saila, 2013.

Godt, J., Baum, R., Savage, W., Salciarini, D., Schulz, W., & Harp, E.: Transient deterministic shallow landslide modeling: requirements for susceptibility and hazard assessments in a GIS framework, Engineering Geology, 102, 214–226, 2008.

Grozavu, A., Pleşcan, S., Patriche, C. V., Mărgărint, M. C., & Roşca, B.: Landslide susceptibility assessment: GIS application to a complex mountainous environment, in: The Carpathians: Integrating Nature and Society Towards Sustainability, edited by Kozak, J., Ostapowicz, K., Bytnerowicz, A., Wyżga, B., *et al.*, pp. 31–44, Springer, 2013.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81, 166–184, 2006.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P.: Rainfall thresholds for the initiation of landslides in central and southern Europe, Meteorology and Atmospheric Physics, 98, 239–267, 2007.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Reviews, 112, 42–66, 2012.

Hansen, A.: Landslide hazard analysis, Slope instability., pp. 523–602, 1984.

Hosmer Jr, D. W. & Lemeshow, S.: Applied Logistic Regression, John Wiley and Sons, 2004.

IBM Corporation: SPSS-X User's Guide, SPSS International B.V., 3 edn., 1988.

INGEMISA: Inventario y Análisis de las Áreas sometidas a Riesgo de Inestabilidades del Terreno de la C.A.P.V., Tech. rep., Eusko Jaurlaritza, 1995.

Jaiswal, P., van Westen, C. J., & Jetten, V.: Quantitative landslide hazard assessment along a transportation corridor in southern India, Engineering Geology, 116, 236–250, 2010.

Kirkby, M. & Beven, K.: A physically based, variable contributing area model of basin hydrology, Hydrological Sciences Journal, 24, 43–69, 1979.

Lee, S.: Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data, International Journal of Remote Sensing, 26, 1477–1491, 2005.

Lee, S., Ryu, J. H., & Kim, I. S.: Landslide susceptibility analysis and its verification using likelihood ratio, logistic regression, and artificial neural network models: case study of Youngin, Korea, Landslides, 4, 327–338, 2007.

Lombardo, L., Fubelli, G., Amato, G., & Bonasera, M.: Presence-only approach to assess landslide triggering-thickness susceptibility: a test for the Mili catchment (north-eastern Sicily, Italy), Natural Hazards, 84, 565–588, 2016.

Malamud, B. D., Reichenbach, P., Rossi, M., & Mihir, M.: D6.3 Report on standards for landslide susceptibility modelling and terrain zonations, Tech. rep., LAMPRE Project, Seventh Framework Program. European Comission, 2014.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., & Guzzetti, F.: An algorithm for the objective reconstruction of rainfall events responsible for landslides, Landslides, 12, 311–320, 2015.

Murillo-García, F. G., Alcántara-Ayala, I., Ardizzone, F., Cardinali, M., Fiourucci, F., & Guzzetti, F.: Satellite stereoscopic pair images of very high resolution: a step forward for the development of landslide inventories, Landslides, 12, 277–291, 2015.

Nefeslioglu, H., Sezer, E., Gokceoglu, C., Bozkir, A., & Duman, T.: Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey, Mathematical Problems in Engineering, 2010, 2010.

Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H., & Gorum, T.: Medium-scale hazard mapping for shallow landslide initiation: the Buyukkoy catchment area (Cayeli, Rize, Turkey), Landslides, 8, 459–483, 2011.

Pardo, A. & Ruiz, M. A.: SPSS 11. Guía para el análisis de datos, McGraw-Hill/Interamericana de España, Madrid, 2002.

Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., & Guzzetti, F.: Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy, Geomorphology, 139, 79–90, 2012.

Peruccacci, S., Brunetti, M. T., Gariano, S. L., Melillo, M., Rossi, M., & Guzzetti, F.: Rainfall thresholds for possible landslide occurrence in Italy, Geomorphology, 290, 39–57, 2017.

Pourghasemi, H. R., Pradhan, B., & Gokceoglu, C.: Application of fuzzy logic and analytical hierarchy process (AHP) to landslide susceptibility mapping at Haraz watershed, Iran, Natural Hazards, 63, 965–996, 2012.

Pradhan, B. & Lee, S.: Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling, Environmental Modelling & Software, 25, 747–759, 2010.

QGIS Development Team: QGIS Geographic Information System, Open Source Geospatial Foundation, URL `http://qgis.osgeo.org`, 2009.

R Core Team: R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria, URL `https://www. R-project.org/`, 2016.

Rana, R., Singhal, R., *et al.*: Chi-square test and its application in hypothesis testing, Journal of the Practice of Cardiovascular Sciences, 1, 69–71, 2015.

Reichenbach, P., Rossi, M., Malamud, B., Mihir, M., & Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60–91, 2018.

Remondo, J., González-Díez, A., De Terán, J. R. D., & Cendrero, A.: Landslide susceptibility models utilising spatial data analysis techniques. A case study from the lower Deba Valley, Guipúzcoa (Spain), Natural Hazards, 30, 267–279, 2003.

Remondo, J., Soto, J., González-Díez, A., de Terán, J. R. D., & Cendrero, A.: Human impact on geomorphic processes and hazards in mountain areas in northern Spain, Geomorphology, 66, 69–84, 2005.

Remondo, J., Bonachea, J., Rivas, V., Sánchez-Espeso, J., Bruschi, V., Cendrero, A., de Terán, J. R. D., Fernández-Maroto, G., Gómez-Arozamena, J., González-Díez, A., & Sainz, C.: Landslide hazard scenarios based on both past landslides and precipitation, in: Workshop on World Landslide Forum, pp. 981–988, Springer, Cham, 2017.

Rossi, M. & Reichenbach, P.: LAND-SE: a software for landslide statistically-based susceptibility zonation, Version 1.0, Geosci. Model Dev., pp. 3533–3543, 2016.

Rossi, M., Guzzetti, F., Reichenbach, P., Mondini, A. C., & Peruccacci, S.: Optimal landslide susceptibility zonation based on multiple forecasts, Geomorphology, 114, 129–142, 2010.

Santacana, N.: Análisis de la susceptibilidad del terreno a la formación de deslizamientos superficiales y grandes deslizamientos mediante el uso de sistemas de información geográfica. Aplicación a la cuenca alta del río llobregat, Ph.D. thesis, Universidad Politécnica de Catalunya, Barcelona, 2001.

Santacana, N., Baeza, B., Corominas, J., De Paz, A., & Marturiá, J.: A GIS-based multivariate statistical analysis for shallow landslide susceptibility mapping in La Pobla de Lillet area (Eastern Pyrenees, Spain), Natural Hazards, 30, 281–295, 2003.

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., Natural Hazards and Earth System Sciences, 15, 2111–2126, 2015.

Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M., & Malet, J. P.: Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models, Geomorphology, 301, 10–20, 2018.

SPSS, I.: IBM SPSS statistics base 20, Chicago, IL: SPSS Inc, 2011.

Süzen, M. L. & Kaya, B. Ş.: Evaluation of environmental parameters in logistic regression models for landslide susceptibility mapping, International Journal of Digital Earth, 5, 338–355, 2012.

Trigila, A., Iadanza, C., & Spizzichino, D.: Quality assessment of the Italian Landslide Inventory using GIS processing, Landslides, 7, 455–470, 2010.

Trigila, A., Iadanza, C., Esposito, C., & Scarascia-Mugnozza, G.: Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy), Geomorphology, 249, 119–136, 2015.

Uriarte, A.: El clima, in: Geografía de Euskal Herria. Clima Y Aguas, edited by Meaza, G. & Urrestarazu, E., vol. 3, chap. 1, pp. 8–83, Editorial Ostoa, 1996.

Urrestarazu, E. & Galdos, R.: Geografía del País Vasco, Editorial NEREA, 2008.

Van Den Eeckhaut, M., Vanwalleghem, 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, 76, 392–410, 2006.

Van Den Eeckhaut, M., Reichenbach, P., Guzzetti, F., Rossi, M., & Poesen, J.: Combined landslide inventory and susceptibility assessment based on different mapping units: an example from the Flemish Ardennes, Belgium, Natural Hazards and Earth System Sciences, 9, 507–521, 2009.

Van Den Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J. P., Montanarella, L., & Nadim, F.: Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data, Landslides, 9, 357–369, 2012.

Van Westen, C., Rengers, N., & Soeters, R.: Use of geomorphological information in indirect landslide susceptibility assessment, Natural Hazards, 30, 399–419, 2003.

Van Westen, C. J., Castellanos, E., & Kuriakose, S. L.: Spatial data for landslide susceptibility, hazard, and vulnerability assessment: an overview, Engineering Geology, 102, 112–131, 2008.

Wang, F., Xu, P., Wang, C., Wang, N., & Jiang, N.: Application of a GIS-Based Slope Unit Method for Landslide Susceptibility Mapping along the Longzi River, Southeastern Tibetan Plateau, China, ISPRS International Journal of Geo-Information, 6, 172, 2017.

Weier, J. & Herring, D.: Measuring vegetation (ndvi & evi), Earth Observatory Library of NASA, 2000.

Wieczorek, G. F., Mandrone, G., & DeCola, L.: The influence of hillslope shape on debris-flow initiation, in: Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment, pp. 21–31, ASCE, 1997.

Xu, C., Dai, F., Xu, X., & Lee, Y. H.: GIS-based support vector machine modeling of earthquake-triggered landslide susceptibility in the Jianjiang River watershed, China, Geomorphology, 145, 70–80, 2012.

Yilmaz, I.: Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat Turkey), Computers & Geosciences, 35, 1125–1138, 2009.

Yilmaz, I.: Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial neural networks, and support vector machine, Environmental Earth Sciences, 61, 821–836, 2010a.

Yilmaz, I.: The effect of the sampling strategies on the landslide susceptibility mapping by conditional probability and artificial neural networks, Environmental Earth Sciences, 60, 505–519, 2010b.

Zêzere, J.: Landslide susceptibility assessment considering landslide typology. A case study in the area north of Lisbon (Portugal), Natural Hazards and Earth System Science, 2, 73–82, 2002.

Chapter 6

Results

I Landslide susceptibility maps using logistic regression model for the Oria river catchment (Gipuzkoa Province). Strategies for variables processing

Landslide susceptibility modelling demands to take many decisions that at the end will have direct effects in the resulting simulation. The following section shows the results of different tests carried out in an experimental zone within our study area that gives objective arguments to support some crucial decisions. In particular, the first two objectives of this thesis were addressed since they are considered two of the basic issues in any statistical modelling approach: Is the dependant variable's quality reliable enough to model the landslide susceptibility? and How can be determined in which explanatory variables depends the landslide occurrence on? or How should these independent variables be processed?

The dependant variable, in this case, was the presence and absence of landslides along the territory that depending on its spatial distribution the probability of occurrence of the future landslides could be modelled. There are different options in order to obtain this information. Some authors collect the published information (research articles, technical reports, etc ...) about the location of landslides so as to produce a bibliography-based landslide inventory. Others dedicate long time surveying aerial photographs and satellite imagery to get multi temporal landslide inventories. And some others got the landslide inventories by direct observations on the field (Guzzetti *et al.*, 2012; Van Den Eeckhaut & Hervás, 2012; Hervás, 2014; Santangelo *et al.*, 2015; Fiorucci *et al.*, 2018). Considering the time and resources available for this thesis project, photo-intepretation and remote sensing techniques were not an option, thus some tests were performed in order to asses if the bibliographical sources available were accurate enough, or instead other options such as field work were more suitable.

However, the scale of the project as well as the origin of the landslide inventory can condition the suitability of certain independent variables (Corominas & Mavrouli, 2011), which implies the need of finding the right set of variables. But not only that, because if the final goal is the design of an standardised methodology for landslide susceptibility, then this variables selection should be done in as much

as objective way possible in order to ensure its reproducibility.

In addition to that, according to the consulted references (Dai & Lee, 2002; Lee, 2005; Duman *et al.*, 2006; Van Den Eeckhaut *et al.*, 2012; Amorim, 2012; Grozavu *et al.*, 2013), there are different possibilities for introducing the same variable in an statistical model like the LR. One independent variable being spatially continuous, such as the slope, can be transformed into categorical variable by grouping its infinite values in different classes. Or inversely, a categorical variable, such as *lithology*, could be processed as a continuous by giving to each class a numerical value by means of a certain weighting criteria. Consequently, depending on the type of processing applied to the independent variables, the resulting model will probably be affected in a given way, and so, also the finale susceptibility map.

On this context, it was carried out the assessment of the different bibliographical sources of landslide inventories. Additionally, an experimental variables selection procedure was tested and the comparison between several susceptibility models in which the explanatory variables followed different processing strategies was studied.

6.1   Methodological approach

6.1.1   Experimental zone

It was decided to carry out the following tests in a representative basin of the study area in order to save time and resources, but above all, to evaluate how to proceed at least in the first phase of the landslide susceptibility assessment. The selected experimental zone was a portion of the Oria river catchment, specifically the part included in the study area. The Oria river is a 78.49 km long stream with a 882.5 km^2 drainage basin and a drainage density of 1,6 km·km^{-2} (Fig. 6.1). It is the biggest catchment of the study area occupying the 40% of it and the 60% of its surface presents more than 20° of slope gradient, with a maximum slope gradient of 88°.

As in the rest of the study area (see chapter 4) there is a very dense communication network (1.94 km·km^{-2}) due to the high density of households in non urban land. Concerning the geology, the region is structurally complex and lithologically very diverse, with materials from Paleozoic rocks to Quaternary sediments (see section 4.2). Thereby, from a general point of view, it corresponds
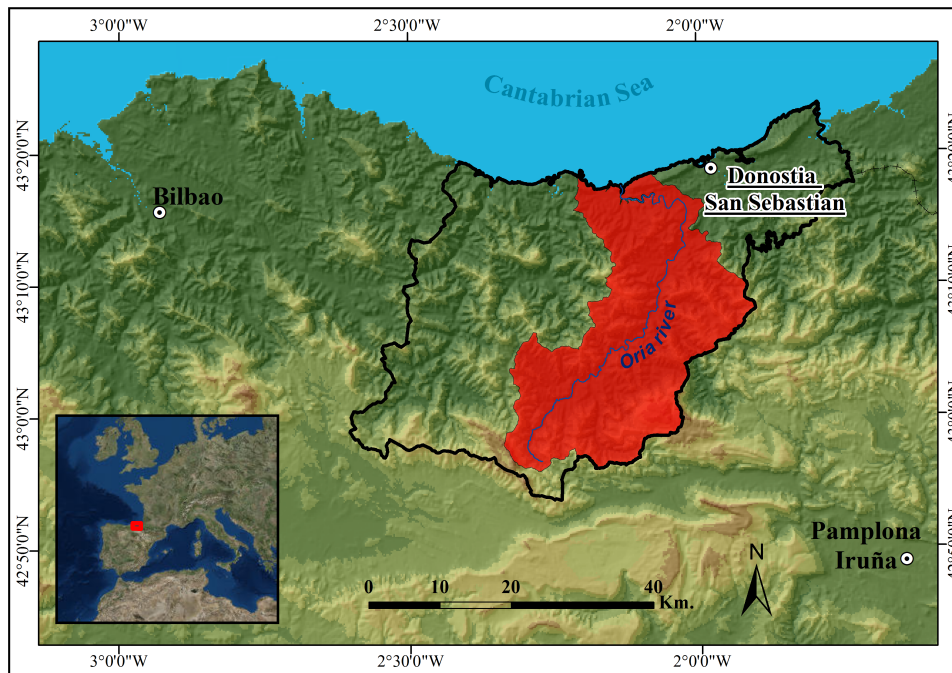
Figure 6.1: *Location of the Oria river basin (in red) which belongs to the Gipuzkoa Province (black lines).*
6.1 Irudia: Gipuzkoako LHren barne (marra beltza) kokatzen den Oria ibai arroaren zatia (gorriz).

to a hilly and mountainous atlantic landscape (Mücher *et al.*, 2010), where an average annual precipitation of 1600 mm is registered with two maximum periods (November-January and April) (González-Hidalgo *et al.*, 2011) and with a persistent daily rains as a typical feature (Fdez-Arroyabe & Martin-Vide, 2012).

6.1.2 Framework

Even though the details about the methodological procedures were explained in chapter 5, the overall work flow followed during this research is explained in the next lines.

In this work, a regular grid cell of 5x5 meters of spatial resolution was used as reference mapping unit (coinciding with the DEM, see section 5.1.2.1). This was chosen as the optimal grid cell size considering the balance between the amount of data to be processed (number of pixels) and the fairest possible representation of the surface conditions of the studied area. Thus, all the further GIS analysis were carried out on the basis of this reference mapping unit.

First, the available bibliographical landslide inventories were assessed visiting on the field a random set of their landslide points, as it was detailed in section 5.2.2.

In the meanwhile, the field trip was profited to obtain landslide location data of every slope instability found. To do so, the UTM coordinates of the top of main scarp (not movilized ground) were collected from the observed landslides. With the Google Earth app each point's position was verified and corrected, in order to reflect the environmental conditions prior to landsliding (Wang *et al.*, 2015). Then, using all this information it was decided weather the available data were reliable enough to continue and develop the landslide susceptibility models.

After the assessment, taking into account that the statistical model chosen to develop the susceptibility models was the LR, a similar number of "landslide free" points were obtained (Felicísimo *et al.*, 2013; Costanzo *et al.*, 2014). In this case, those places without any type of landslide inventoried were considered stable areas (or landslide free areas)(Nefeslioglu *et al.*, 2008; Pourghasemi *et al.*, 2013; Wang *et al.*, 2015). Therefore, using the ArcGIS software, a random sample of "landslide free" points was created (the same amount as the inventoried landslides) by means of an spatially uniform sampling scheme but excluding a 30 m buffer zone for all landslides so as to minimize the impact of their size (Dai & Lee, 2002).

Once the dependant variable was defined, these data were used to perform some descriptive statistics and apply the variables selection approach detailed in section 5.2.3.1. For that, all the collected spatial variables described in section 5.1.2 were tested except *land cover 3* -this variable was obtained after this research was carried out-.

Finally, in order to assess the performance of the models and to provide objective metrics for their comparison, the 20% of the final landslide inventory was reserved for validation. Crossing this validation sample with the results of each susceptibility map, Cumulative Percentage Curves (Duman *et al.*, 2006) and the area under the curve (AUC) of the prediction rate curves (Chung & Fabbri, 2003) were calculated. Given the limited extent of the landslide inventory available, this calibration/validation division was chosen (80% and 20%) in order to ensure the maximum amount of data in the calibration of the statistical model without compromising the availability of data for validation (Nefeslioglu *et al.*, 2011).

6.1.3 Variables processing strategies

One of the objectives of the study was to experiment with the available explanatory variables, and after the bibliographic revision, it was discovered the differences in the way that some variables were treated during the statistical modelling. Therefore, the variables resulting from the selection process were transformed from continuous to categorical, and vice versa, in order to test their suitability.

The originally continuous variables were divided and grouped in classes of a given range in order to process them as categorical variables, according to what it was seen in other author's works (Dai & Lee, 2002; Lee, 2005; Bonachea, 2006). Variables of elevation and sinusoidal slope were re-classified automatically using the ArcGIS software, and divided in 5 classes applying the "equal intervals" option (i.e. all the classes are of the same range size). NDVI was manually divided in 3 single classes representing the water covered or no vegetation area (between -0.6 and -0.2), poor vegetation area (between -0.2 and 0.2) and sparse vegetation area (between 0.2 and 0.4)(Weier & Herring, 2000). Concerning the categorical variables, in order to reduce as much as possible the subjective decisions, the original classes were maintained instead of using the reclassified layers, which were done by expert criteria. The only exception was *land cover 1*, where due to the big number of categories closely linked to urban areas, they were grouped together into a new class named *Urban area* (all the original classes are shown in Appendix C).

The transformation of categorical variables into continuous variables had to be done by giving to each category a numerical value according to a given range of magnitude. These values should be given in an objective way to avoid subjectively driven decisions. So, in this study the landslide density (LD) approach was applied. This approach, applied previously in Bai *et al.* (2010) and Grozavu *et al.* (2013), considers the density of landslides presence in each category in terms of area and it can be computed as follow:

$$LD_i = \frac{LA_i/A_i}{LA/A} \tag{6.1}$$

Where, LD_i is the landslide density value for class i, LA_i and A_i are the landslide area in class i and the total area of class i, respectively, and LA and A are the total

landslide area in the studied area and the total area of the studied area, respectively.

Hence, considering different combinations between the variables processing strategies four models of landslide susceptibility were prepared. In three of them (model A, B and C), only the processing strategy varies, using the same set of stable and unstable points and the same explanatory variables (chosen in the selection phase of variables). For the fourth model (model D), the explanatory variables were selected taking into account the results of the first three models and following the expert criterion. Next, the strategies applied for each model are detailed.

Model A

Model A was calculated by transforming categorical variables into binary code confusion variables, always using the first class as a reference. More specifically, for each of these variables, the *k* categories were replaced by *k-1* dummy variables, each one with values of 1 or 0, which indicates the presence or absence of one of the categories *k-1* (Van Den Eeckhaut *et al.*, 2012). It is an automatic procedure that provides the SPSS XXII software as an alternative to enable the introduction of categorical variables in the RL (Pardo & Ruiz, 2002). This means that, in this case, each class of the categorical variables will obtain its own coefficient β. On the other hand, the continuous variables were not transformed at all.

Model B

The continuous variables were transformed into categorical dividing their infinite values in grouped classes (Dai & Lee, 2002). Then the LR was carried out applying to all the variables the same replacement of dummy variables as in model A.

Model C

In model C all the variables were processed as continuous. The landslide densities for the classes of each categorical variables were calculated and these values were used as relative numerical values for their transformation into continuous variables (Zhu & Huang, 2006; Grozavu *et al.*, 2013; Trigila *et al.*, 2015). This methodology avoids the creation of an excessive number of dummy variables, but requires the previous step of the calculation of landslide densities.

Model D

Model D was calculated following the same processing strategy of variables as in model C. However, in this case, the set of variables introduced in the LR was modified by expert criteria taking as a reference the statistical results of the selection phase of the variables, as well as the results of models A, B and C.

6.2 The landslide inventories

6.2.1 Assessment of the bibliographical sources

Among all the landslide data collected from three different bibliographical sources (see section 5.2.2 and Fig. 6.2a) only those located inside the experimental zone were extracted (Fig. 6.2b). The 74% of this subset was considered as slide or shallow slide type of movement, 11% were rock falls or rock mass deposits and 4% were flows or complex movements, adding to the 11% of the subset that was labelled as landslide scarp but without specifying the type of landslide. Thus, taking into account that different typologies of slope instabilities respond to different mechanisms, only the most extended type of landslides was used for the research, i.e. slides and shallow slides (Fig. 6.2c).

After selecting at random 23 slide locations from all the bibliographical sources (Fig. 6.2d), they were checked on the field and the results are summarized in table 6.1. 10 of the check points came from the *Inventory of the Basque Government* and after their surveillance on the field, it was ascertained that 6 of them didn't present any slope instability's evidence even in the surroundings, while in the rest the location was not exact though landslide evidences were found next to them. Other 7 points came from the *Inventory of the road network*, and in this case all of the surveyed locations corresponded to the exact location of slide movements. The last 6 check points came from the *Inventory of the geomorphological map*, where 5 field visits resulted in true slides locations, although in 1 case the location was inaccurate.

All bibliographical data
(753 landslides)

Bibliographical data
within the Oria river basin
(377 landslides)

Only slide type data
within the Oria river basin
(277 slides)

Random check points
within the Oria river basin
(23 slides)
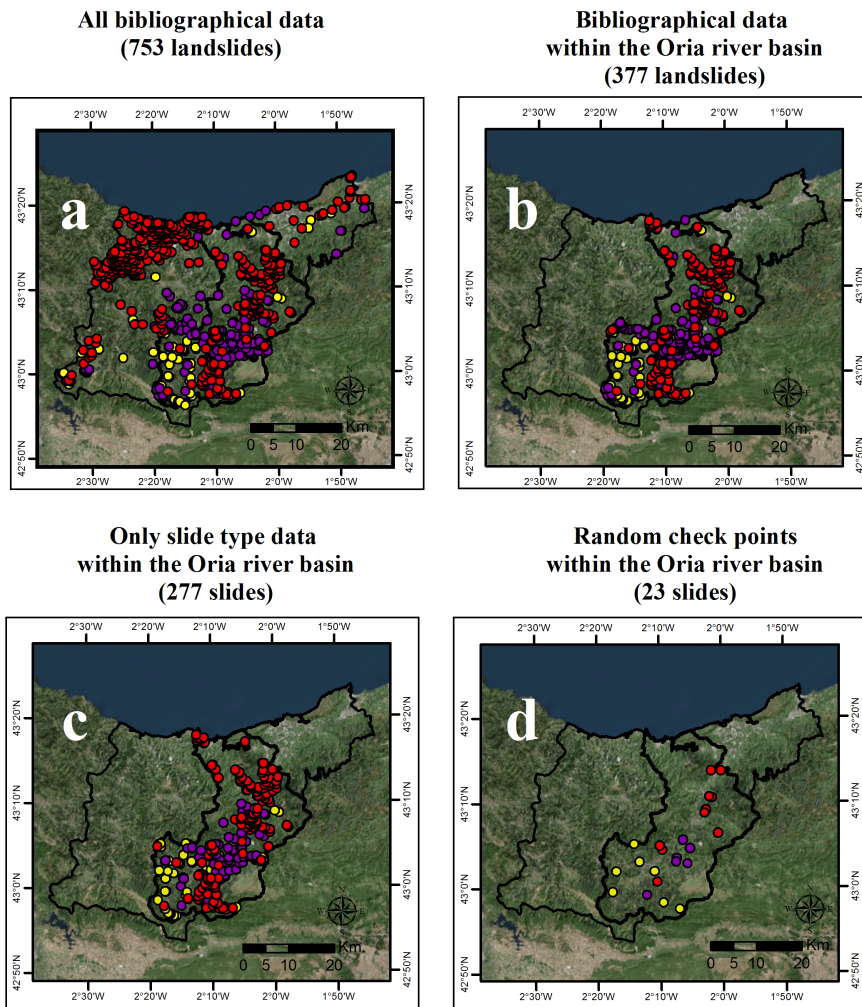
<u>**Bibliographical sources**</u>
- ● Inventory of the Basque Government
- ● Inventory of the geomorphological map
- ● Inventory of the road network

Figure 6.2: *Sampling steps for the bibliographic landslide data assessment.*

6.2 Irudia: Bibliografiatik hartutako lur labainketen laginketarako pausoak.

Table 6.1: *Results of the bibliographical landslide data assessment.*

6.1 Taula: Bibliografiatik hartutako lur labainketa datuen balioespen emaitzak.

Data source	Instability evidences	No evidences of instability	Total
Inventory of the Basque Government	4	6	10
Inventory of the road network	7	0	7
Inventory of the geomorphological map	5	1	6

6.2.2 The field based landslide inventory in the Oria river catchment

In view of the preliminary results of the bibliographical data sources, it was decided to collect every slide location observed during the field trips. This way, 325 points corresponding to shallow and deep slides (both rotational and translational) were inventoried by field work (Fig. 6.3a). Then, all the data collected on the field were compared with the bibliographical landslide inventory in order to check if some of those observed landslides were already inventoried by other sources.

Figure 6.3b shows the frequency distribution of the distance between the slope instabilities inventoried on the field and the previously collected bibliographical landslide points. It was observed that among the 325 field-work-based landslide points, only the 5% was located to less than 10 meters of distance from landslide points of bibliographical origin. Moreover, the most part of them corresponded to the data coming from the *Inventory of the road network*. The 13% were placed in the surroundings of the already inventoried slides, i.e between 10 and 50 meters of distance to them, which suggests a correct identification of the instability though with not enough spatial accuracy. The 9% was inventoried in the vicinities of the points offered by the bibliographical sources (to less than 100 meters of distance). And all the rest (73% of the landslides collected on the field) was located to more than 100 meters of distance from any other landslide already inventoried.
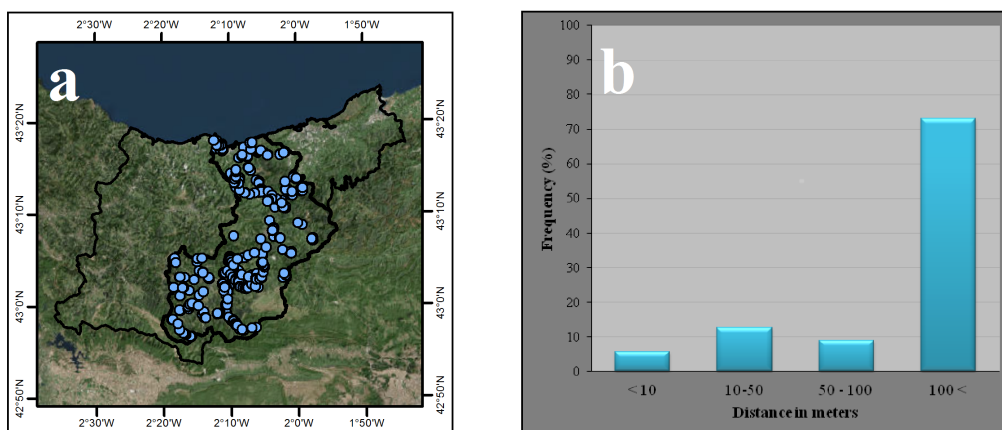
Figure 6.3: *a) Spatial distribution of the field-work-based landslide inventory; b) Frequency distribution of the distance between field-work-based landslide points and bibliographical source landslide points.*

6.3 Irudia: a) Landa laneko lur labainketen inbentarioa; b) Landa laneko lur labainketen eta bibliografiatik ateratako datuen arteko distantziaren frekuentzia banaketa.

All the results presented in this section support the idea that the bibliographical sources of information did not offer accurate enough data for the objectives of this study. So, considering that the landslide inventory was the primary information in which all the further analysis would be based on, it was taken the decision to continue our analysis using the field based landslide inventory. To do so, a validation subset was reserved for validation tests procedures (65 slide points selected at random) and a similar amount of the rest of the field- work-based landslide inventory, 260 points, was created to use them as "landslide free" points for the explanatory variables analysis as well as for the landslide susceptibility models calibration (Fig. 6.4).
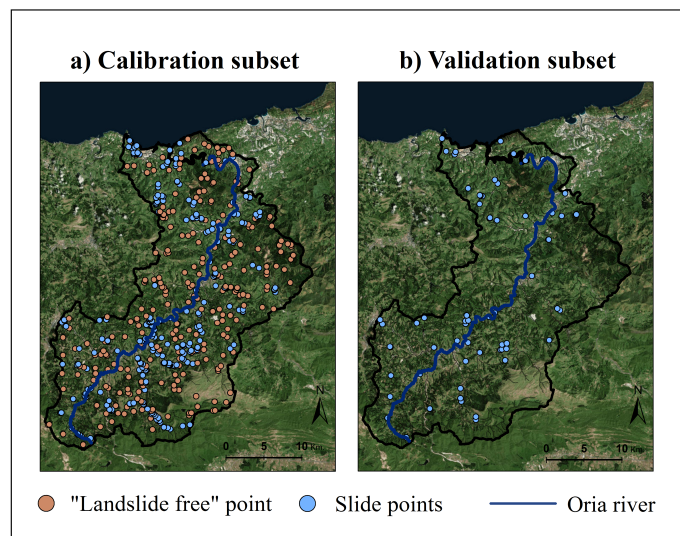
Figure 6.4: *Spatial distribution of a) Calibration and b) Validation samples.*
6.4 Irudia: a) Kalibrazio eta b) Balidazio laginen banaketa espaziala.

6.3 Independent variables. Analysis and selection.

Following the procedure explained in section 5.2.3.1, some descriptive statistics were computed in order to select the explanatory variables with major role for slope movements occurrence in our experimental zone.

As it is summarized in table 6.2, all the continuous variables performed values under the threshold of 0.05 in the Kolmogorove-Smirnof test, which means that they did not follow a normal distribution. Consequently, the significance tests were carried out with the Mann-Whitney test. In the case of the categorical variables, the Chi^2 test was performed. Among the original 19 explanatory variables, 4 were excluded

because they exceeded the significance threshold value (marked with an asterisk in Tab. 6.2). So, *permeability*, *curvature*, *profil curvature* and *planform curvature* did not present an statistically significant difference between the stable and unstable points.

Table 6.2: *Results of the statistical tests of Kolmogorov Smirnov (K-S), Mann Whitney and Chi^2 of the explanatory variables. Marked with asterisk (*): variables that did not show statistical significant difference with the dependent variable.*

6.2 Taula: Kolmogorov Smirnov (K-S), Mann Whitney eta Chi^2 estatistikoen emaitzak aldagai eragileetarako. Asteriskoaz markatuta (*): aldagai dependentearekiko estatistikoki diferentzia esanguratsurik azaldu ez duten aldagaiak.

Continuous variable	K-S	Mann-Whitney	Categorical variable	Chi^2
elevation	0.05<	0	*lithology*	0
slope	0.05<	0	* *permeability*	0.09
sinusoidal slope	0.05<	0	*regolith thickness*	0.01
SAR	0.05<	0	*land cover 1*	0
TWI	0.05<	0.01	*land cover 2*	0
* *curvature*	0.05<	0.41	*distance to the main river-streams CAT*	0
* *planform curvature*	0.05<	0.26	*distance to transport network CAT*	0
* *profile curvature*	0.05<	0.34	aspect	0
distance to the river	0.05<	0.01		
distance to the transport network	0.05<	0		
NDVI	0.05<	0		

Among the remaining 15 variables, the correlation test showed high collinear relationship between different groups of variables. The continuous variables *slope*, *sinusoidal slope* and *SAR* presented positive results (Tab. 6.3), thus, they were introduced separately to the LR due to their high correlation. Likewise, in categorical variables, understandably, both *land cover 1* and *land cover 2* represent the same information, so they could not be taken into account together in the modelling process (Tab. 6.4). Additionally, *distance to the river* and *distance to the transport network* were also computed separated from *distance to the main river-streams CAT* and *distance to transport network CAT* respectively, since the latter are the result of a reclassification of the former.

Table 6.3: *Correlation matrix of the continuous explanatory variables. The Spearman coefficients (Coef.) and their related significance level (Sig.) are showed. High correlation coefficients are highlighted in grey.*

6.3 Taula: Aldagai jarraien korrelazio matrizea. Spearman koefizientea (*Coef.*) eta bere esangura maila (*Sig.*) erakusten dira. Korrelazio koefiziente altuak grisez azpimarratu dira.

Variable	Measure	1	2	3	**4**	**5**	6	7	8
1. elevation	*Coef*	1							
	Sig.	0							
2. TWI	*Coef*	-0.04	1						
	Sig.	0.304	0						
3. NDVI	*Coef*	0.236	-0.04	1					
	Sig.	0	0.313	0					
4. slope	*Coef*	-0.076	-0.275	0.061	1				
	Sig.	0.054	0	0.117	0				
5. sinusoidal slope	*Coef*	0.076	-0.272	0.066	0.998	1			
	Sig.	0.052	0	0.095	0	0			
6. SAR	*Coef*	-0.077	-0.28	0.068	0.987	0.984	1		
	Sig.	0.048	0	0.083	0	0	0		
7. distance to the river	*Coef*	0.216	-0.001	0	-0.134	-0.136	-0.139	1	
	Sig.	0	0.972	0.998	0.001	0.001	0	0	
8. distance to the transport network	*Coef*	0.36	0.013	0.293	-0.074	-0.076	-0.084	0.161	1
	Sig.	0	0.749	0	0.06	0.052	0.032	0	0

Table 6.4: *Correlation matrix of the categorical explanatory variables. The Spearman coefficients (Coef.) and their related significance level (Sig.) are showed. High correlation coefficients are highlighted in grey.*

6.4 Taula: Aldagai kategorikoen korrelazio matrizea. Spearman koefizientea (*Coef.*) eta bere esangura maila (*Sig.*) erakusten dira. Korrelazio koefiziente altuak grisez azpimarratu dira.

Variable	Measure	1	2	**3**	**4**	5	6	7
1. lithology	*Coef*	1						
	Sig.	0						
2. regolith thickness	*Coef*	0.089	1					
	Sig.	0.023	0					
3. land cover 1	*Coef*	0.039	0.071	1				
	Sig.	0.324	0.071	0				
4. land cover 2	*Coef*	0.029	0.094	0.548	1			
	Sig.	0.455	0.017	0	0			
5. aspect	*Coef*	-0.002	0.006	0.101	0.086	1		
	Sig.	0.964	0.886	0.01	0.028	0		
6. distance to transport network CAT	*Coef*	-0.108	-0.158	-0.187	-0.236	0.024	1	
	Sig.	0.006	0	0	0	0.54	0	
7. distance to the main river-streams CAT	*Coef*	-0.064	-0.134	0.029	0.082	0.072	0.191	1
	Sig.	0.102	0.001	0.459	0.037	0.067	0	0

As a result of such inter relations between variables, 24 different combinations of 10 variables were tested in a preliminary LR run, by means of the SPSS XXI software (see tables 6.5 and 6.6). In all the computed regressions, the *regolith thickness,aspect* and *distance to the river* were eliminated from the final set of estimate variables of the model. The variables *TWI* and *distance to the main river-streams CAT* were selected in some cases and rejected in others, but each time they were introduced to the equation the classification performance decreased, so they were not considered as suitable explanatory variables. At the end, the highest classification performance (84.3 in Tab. 6.6) was set by the combination composed by the *elevation, NDVI, lithology, sinusoidal slope, land cover 1* and *distance to transport network CAT*. Nevertheless, it can be pointed out that the usage of *slope* instead of *sinusoidal slope* only decreases the performance in 0.1, which goes in consonance with the very high correlation coefficient between them. This suggests that probably the introduction of any of those variables would not result in relevant changes in the susceptibility model, but in order to maintain the statistically driven approach, only the most performing combination was considered as the most suitable. In the same way, even though *land cover 1* and *NDVI* represent similar information about the terrain surface, the former concerns a typological aspect and the later a numerical indicator, apart from the fact that the automatic software does not reject none of them among the 24 tests. Therefore, these were the explanatory variables selected for the development of the first three susceptibility models.

Table 6.5: *Part I. Summary of the 24 combinations tested in a preliminary LR model run. The explanatory variables introduced in the model in each run are marked with X. The explanatory variables composing the final equation of each run are highlighted in grey.*

6.5 Taula: I zatia. LR modeloaren kalkulu preliminarrean testatutako 24 konbinazioren laburpena. Kalkulu bakoitzean erabilitako aldagaiak X bitartez adierazi dira. Azken ekuazioa osatzen duten aldagaiak grisez azpimarratu dira.

	1	2	3	4	5	6	7	8	9	10	11	12
elevation	X	X	X	X	X	X	X	X	X	X	X	X
TWI	X	X	X	X	X	X	X	X	X	X	X	X
NDVI	X	X	X	X	X	X	X	X	X	X	X	X
lithology	X	X	X	X	X	X	X	X	X	X	X	X
regolith thickness	X	X	X	X	X	X	X	X	X	X	X	X
aspect	X	X	X	X	X	X	X	X	X	X	X	X
slope	X			X			X			X		
sinusoidal slope		X			X			X			X	
SAR			X			X			X			X
land cover 1	X	X	X				X	X	X			
land cover 2				X	X	X				X	X	X
distance to the river	X	X	X	X	X	X						
distance to the main riverstreams CAT							X	X	X	X	X	X
distance to the transport network	X	X	X	X	X	X	X	X	X	X	X	X
distance to transport network CAT												
Overall classification index	82.0	82.0	81.5	80.5	80.6	78.0	82.0	82.3	80.3	79.4	79.8	79.8

Table 6.6: *Part II. Summary of the 24 combinations tested in a preliminary LR model run. The explanatory variables introduced in the model in each run are marked with X. The explanatory variables composing the final equation of each run are highlighted in grey.*

6.6 Taula: II zatia. LR modeloaren kalkulu preliminarrean testatutako 24 konbinazioren laburpena. Kalkulu bakoitzean erabilitako aldagaiak X bitartez adierazi dira. Azken ekuazioa osatzen duten aldagaiak grisez azpimarratu dira.

	13	14	15	16	17	18	19	20	21	22	23	24
elevation	X	X	X	X	X	X	X	X	X	X	X	X
TWI	X	X	X	X	X	X	X	X	X	X	X	X
NDVI	X	X	X	X	X	X	X	X	X	X	X	X
lithology	X	X	X	X	X	X	X	X	X	X	X	X
regolith thickness	X	X	X	X	X	X	X	X	X	X	X	X
aspect	X	X	X	X	X	X	X	X	X	X	X	X
slope	X			X			X			X		
sinusoidal slope		X			X			X			X	
SAR			X			X			X			X
land cover 1	X	X	X				X	X	X			
land cover 2				X	X	X				X	X	X
distance to the river	X	X	X	X	X	X						
distance to the main river-streams CAT							X	X	X	X	X	X
distance to the transport network												
distance to transport network CAT	X	X	X	X	X	X	X	X	X	X	X	X
Overall classification index	**84.2**	**84.3**	**82.6**	**82.6**	**82.5**	**80.8**	**84.2**	**84.3**	**82.6**	**82.6**	**82.5**	**80.8**

6.4   Models A, B and C. Results and comparison

Tables 6.7 and 6.8 show the coefficients assigned to each explanatory variable by means of which the different susceptibility models were defined. The signs and absolute values of the coefficients indicate which of the environmental factors play a decisive role in the occurrence of landslides. In categorical variables, a positive and high value indicates an important role in favour of slope movement of that class, while a negative value means an important role in favour of stabilization. However, it should be noted that due to the dummy reclassification applied in models A and B, the coefficients of the categorical classes were obtained in reference to the first class of each variable, so they are about values of relative magnitude, and consequently they are not comparable with values of other models. On the other hand, in continuous variables, a high and positive coefficient means that the greater the pixel value in that variable, the greater would be the probability of occurrence of a landslide. Conversely, if the coefficient is negative, the presence of high absolute values would represent a lower probability of landslide occurrence.

In view of the β values in tables 6.7 and 6.8, it can be seen that *distance to transport network CAT*, *sinusoidal slope* and *elevation* present similar relative results in models A, B and C. Although *elevation* did not imply almost any relevance (coefficient close to 0), the other two variables proved to be really significant for the three models, at least, in our experimental zone and with our landslide inventory. Thus, the *distance to the transport network CAT* and the *sinusoidal slope* are considered the main factors that explain the distribution of the dependent variable. Likewise, according to these results, it could be inferred that the following values describe the most common features on the slopes affected by landslides in the Oria river basin (see Tabs. 6.7 and 6.8): *sinusoidal slope* values close to 1 (around 45° of slope); *elevation* between 0 and 300 m; *NDVI* between 0.2 and 0.4 (sparse vegetation); *distances from the transports network CAT* between 0 and 20 m; *land cover 1* covered by scrub, pasture or meadows; and coarse grained detrital rocks or marls *lithology*.

The cumulative percentage curves (Fig. 6.5) showed high coincidence with respect to the instability points reserved for validation. In all cases the great

Table 6.7: *Part I: Estimate coefficients (β) for each model result and the landslide density (LD) values for each class. Only categorical variables.*

6.7 Taula: I Atala: Modelo bakoitzaren β koefiziente estimatuak eta *LD* lur labainketa dentsitate balioa klase bakoitzerako. Aldagai kategorikoak bakarrik.

Models		A	B	C		D
		β coef.	β coef.	*LD*	β coef.	β coef.
lithology	Coarse grained detrital rocks (sandstone)	0	0	*1.346*	0.615	0.827
	Surface deposits	-3.721	-3.207	*0.5356*		
	Alternation of detrital rocks	-1.334	-1.371	*0.5994*		
	Fine-grained detrital rocks (lutites)	-4.142	-3.803	*0.4593*		
	Marls, limestones, marlstones and calcarenites alternation	-1.552	-1.611	*13.561*		
	Marls	-0.884	-0.961	*20.422*		
	Decarbonated marls	-2.771	-4.429	*0.2845*		
	Clay with gypsium and other salts	-3.221	-2.849	*11.025*		
	Impure limestones and calcarenites	-1.516	-1.54	*10.961*		
	Limestones	-4.039	-3.906	*0.0503*		
	Ophites	-1.669	-1.224	*26.592*		
	Slates	-3.008	-3.671	*0.0674*		
land cover 1	Urban area	0	0	*0.32*	0.559	0.782
	Transport infraestructures	0.002	0.576	*19.193*		
	Forest	-0.332	-1.047	*0.4361*		
	Plantation forest	0.247	-0.334	*0.5323*		
	Gallery forest	23.014	22.302	*0.9212*		
	Grassland and pastures	-17.491	-15.376	*0*		
	Meadow	1.374	1.038	*28.446*		
	Cultivations	-2.373	-2.594	*0.4198*		
	Bush	1.528	0.96	*14.334*		
	Pasture and scrubs	1.471	1.591	*0.1995*		
	Meadow with hedgerow	-20.967	-20.61	*0*		
distance to transport network CAT	0 - 20 m	0	0	*5.745*	0.499	- -
	20 - 50 m	-0.52	-0.752	*46.428*		
	50 - 100 m	-1.712	-1.759	*30.149*		
	100 - 150 m	-1.471	-1.491	*25.151*		
	150 - 200 m	-2.395	-2.536	*12.551*		
	200 - 250 m	-3.391	-3.365	*0.6732*		
	250 - 300 m	-2.112	-1.991	*0.6295*		
	> 300 m	-3.329	-3.569	*0.2682*		

Table 6.8: *Part II: Estimate coefficients (β) for each model result and the landslide density (LD) values for each class. Only continuous variables.*

6.8 Taula: Modelo bakoitzaren β koefiziente estimatuak eta *LD* lur labainketa dentsitate balioa klase bakoitzerako. Aldagai jarraiak bakarrik.

Models		A	B	C		D
		β coef.	β coef.	*LD*	β coef.	β coef.
NDVI	-0.6 to -0.2	-5.342	0	*16.226*	-4.574	-4.587
	-0.2 to 0.2		-1.124	*0.9534*		
	0.2 to 0.4		-0.216	*10.395*		
sinusoidal slope	0 - 0.2	5.975	0	*0.0595*	5.312	4.853
	0.2 - 0.4		2.526	*0.2702*		
	0.4 - 0.6		2.516	*0.3784*		
	0.6 - 0.8		4.559	*12.982*		
	0.8 - 1		5.265	*14.433*		
elevation	0 - 300 m	-0.002	0	*16.888*	-0.002	- -
	300 - 600 m		-0.226	*0.8119*		
	600 - 900 m		-1.934	*0.0979*		
	900 - 1200 m		-20.5	*0*		
	1200 - 1550 m		-18.328	*0*		
Intercept		-0.829	0.431		-5.85	-5.75

majority of the validation sample (between 90% and 95%) was located in landslide susceptibility classes greater than 0.5 probability, as it was noticed in the a-curves of each model. In this regards, model B stood out, whose a-curve showed that less than 5% of the landslides coincided with the lower susceptibility classes, while in models A and C this value increased up to about 10%. Nevertheless, b-curves indicated that not all models presented the same discrimination capacity. In model B, high probability classes covered greater percentage of the experimental zone (approximately 30%) than in models A and C (near 15%).

Concerning the prediction rate curves (Fig. 6.6), their shape revealed that the three different strategies for the explanatory variables processing were very satisfactory in terms of accuracy, although small differences could be highlighted between these susceptibility models. Numerically, with an AUC of 0.951, model C offered the best performance followed by model A (AUC = 0.948), while model B showed a slightly lower value (AUC = 0.938).

Apart from the validation tests of the statistical models, considering each model's equations, landslide susceptibility maps were displayed (Fig. 6.7). Among them,
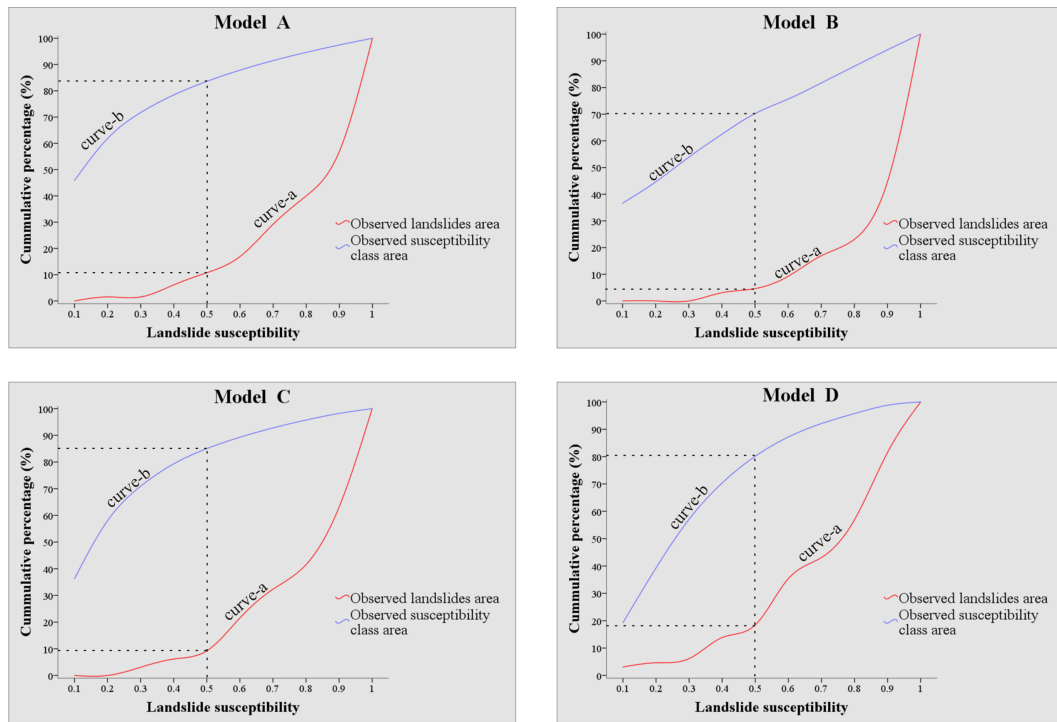
Figure 6.5: *Cummulative percentage curves for models A, B, C and D.*

6.5 Irudia: Akumulatutako portzentaien kurbak A, B, C eta D modeloetarako.
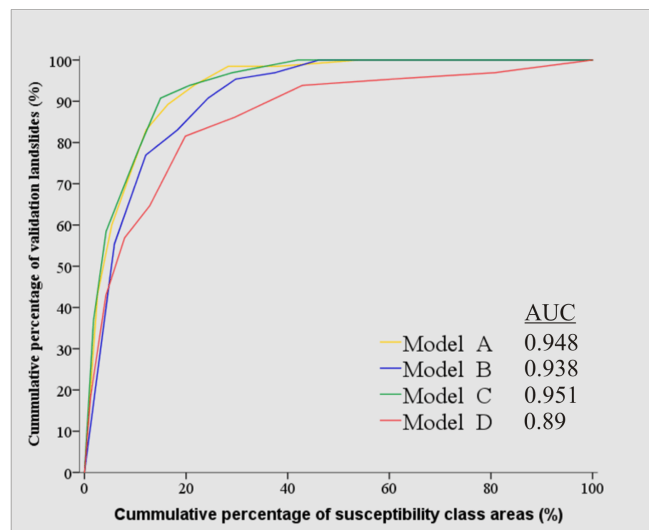
Figure 6.6: *Prediction rate curves for models A, B, C and D; and their corresponding area under the curve (AUC).*

6.6 Irudia: A, B, C eta D modeloetarako aurreikuspen tasa kurbak; eta horien kurba azpiko azalera (AUC).

the biggest difference was observed in model B, where, by means of a visual inspection, the uniformity of the non-susceptible zones in the eastern part of the basin (mountainous areas) stood out, compared to the higher variability showed by models A and C in the same area.

Additionally, it was highlighted the broad similarity of the three maps, in which the high susceptibility classes were distributed clearly following a linear pattern similar to the drainage network. As a matter of fact, more detailed inspection of the maps (Fig. 6.8) revealed the almost complete dependency of the distribution of susceptibility to the variable *distance to transport network CAT*. This could show some bias that the landslide inventory could have with respect to this variable, since the information of the locations of the landslides was obtained by direct field work.

6.5   Results of model D

Based on the results of the first three models, it was concluded that (i) the *elevation* had practically no effect on the presence or absence of the inventoried landslides (coefficient β very close to 0), and (ii) the distribution of the susceptibility maps was strongly conditioned by the variable *distance to transport network CAT*, which suggests that at least part of the inventory could be biased by this explanatory variable.

For this reason, an additional susceptibility model was carried out (model D), in which the same strategy of variables processing as in model C was applied. However, in this case *elevation* and *distance to transport network CAT* were removed as explanatory variables and only *lithology, land cover 1, NDVI* and *sinusoidal slope* were used.

The susceptibility map D (Fig. 6.7) shows notable differences with respect to the previous models. The probability values of landslide occurrence are not distributed following the same linear patterns present in models A, B and C, and such difference is even more evident if the zoomed portions are observed (Fig. 6.8).

Regarding the validation tests, the curve-a and curve-b of the cumulative percentage curves (Fig. 6.5) showed that less than 20% of the validation landslides showed susceptibility values under 0.5. Moreover, around 80% of the experimental

MODEL A

MODEL B

MODEL C

MODEL D

Susceptibility

Very high : 1

Very low : 0

0 10 20 40
Km.

N
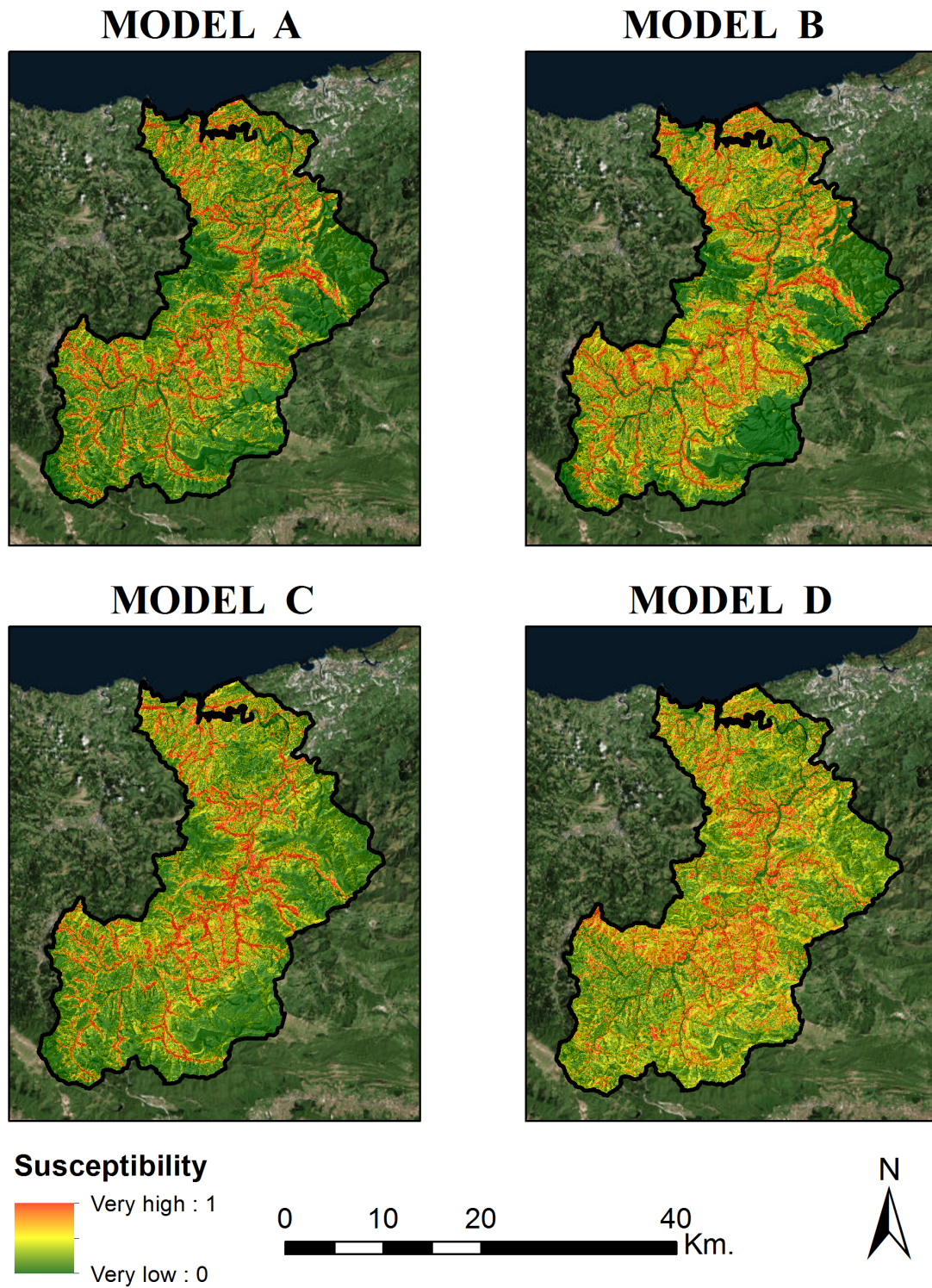
Figure 6.7: *Landslide susceptibility maps for models A, B, C and D.*

6.7 Irudia: Lur abainketa suszeptibilitate mapak A, B, C eta D modeloetarako.

zone was classified as a low susceptibility zone. The same information, represented in the prediction rate curve, performed an area under the curve of 0.89 (Fig. 6.6).

6.6   Discussion

The analysis carried out in a reduced experimental zone within the complete study area of the current thesis allowed to clarify some steps, that we considered crucial, for landslide susceptibility modelling: (i) the available data about the landslide locations were assessed; (ii) an explanatory variables selection approach was tested and; (iii) it was find out the most suitable strategy for explanatory variables processing. Thus, the conclusions as well as considerations obtained through the presented results are discussed in the following lines.
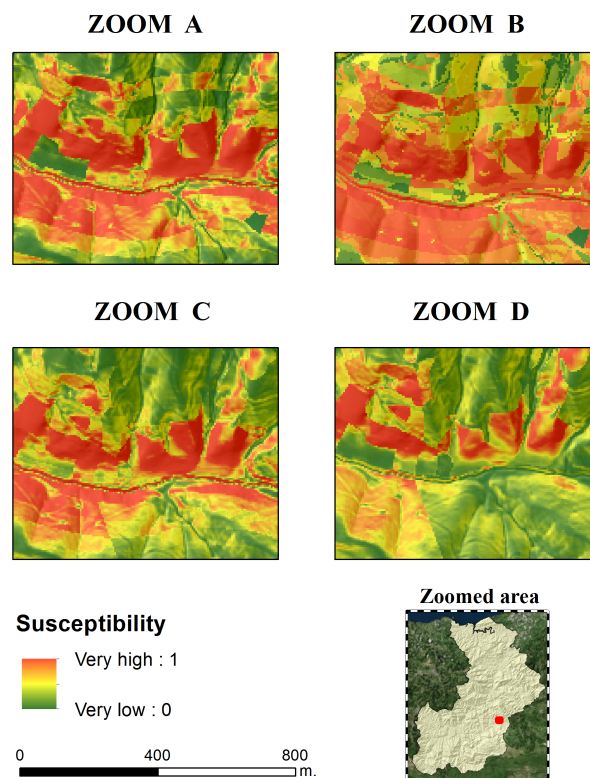
Figure 6.8: *Detailed zoom of susceptibility maps A, B, C and D.*
6.8 Irudia: Handitutako zatia A, B, C eta D suszeptibilitate mapetan.

The aleatory check points of the landslide location data coming from different bibliographical sources showed up the limited accuracy of the bibliographical inventory. To begin, the *Inventory of the Basque Government*, which is the most

abundant source of data, provided very inaccurate landslide locations according to the field visits carried out. *Inventory of the geomorphological map* provided right approximated coordinates in the most part of the checked points, but the comparison of the total set to the shallow slides directly observed on the field revealed a minimum mismatch of 10 m, which was considered unacceptable considering that the resolution of the raster layers used for the model construction, and consequently the final susceptibility maps, was of 5 m. The only bibliographical source that displayed landslide locations of enough accuracy was the *Inventory of the road network*, however, the amount of data was very limited besides the fact that this technical report only considered slope movements that affects the road network. So in view of such results it was considered that the bibliographical inventory was not a valid data set to model the landslide susceptibility in our experimental zone, and consequently, neither in the complete study area. Instead, a field work based landslide inventory was carried out and the data were used for all the further analysis of the research. Nevertheless, despite the reliability, in terms of spatial accuracy, provided by shallow slides locations obtained by direct observation on the field, some drawbacks were detected. On one hand, the usage of one single point to represent the theoretical environmental conditions that played a decisive role in favour of the landslide occurrence could be considered as a source of uncertainty, taking into account that shallow slides are processes that can cover from tens to thousands of square kilometres. So, even if this approach leads with good results, other techniques like the delimitation of the complete landslide area (instead of applying an average buffer to each point) worth being tested. On the other hand, the approach followed in the current investigation assumes that any place in which no landslide was inventoried is free of landslides. Such assumption is another source of uncertainty, unless a complete multi temporal landslides inventory was done, in which every places of the territory was surveyed and ascertained that no landslide evidences exist. In case of bibliographical inventories no-landslide places are not known at all, and in field work-based inventories such information can only be ensured in the specifically visited places, which usually takes only a portion of the real area under study.

The variables selection procedure demonstrated to be valid for the objective

choice of explanatory variables that are directly related to the spatial distribution of landslides and the identification of variables highly correlated to each other. The rejection of the first 4 variables by means of the statistical significance tests was an expected result, considering that all these predisposing factors presented very uniform spatial distribution along the study area. In addition, the fact that the software itself systematically eliminated other 3 variables was in agreement with the results of other investigations (Carrara, 1983; Guzzetti *et al.*, 1999; Remondo *et al.*, 2008; Yilmaz, 2009). In the case of the *distance to the river*, its elimination could be due to the high density of the drainage network, which makes it probable for most points, stable or unstable, to place near a river, and therefore the variability can be significantly reduced. Additionally, it has to keep in mind that some of the variables were simple reclassifications of others, which were transformed in order to cover the widest range of possibilities. But logically, in these cases the correlation was practically perfect and consequently this type of variables could not be introduced together in the model.

Thereby, it was possible to select, in an objective way, only 6 variables from a set of 19 initial options. However, this selection did not result entirely satisfactory, since once the LR models and the susceptibility maps were defined, the inadequate effect of some variables was detected. First, in the case of *elevation*, although the Man Witney test suggested a statistically significant relation with the presence or absence of landslides, once introduced in the LR it showed practically null effect within the model, with a β estimate value very close to 0. In fact, the geomorphological justification of this variable is not straightforward, since altitude, by itself, does not have a direct effect on slopes failures, though, it can represent the association of any other variable, such as lithology or land use. Moreover, the statistical tests also ensured the relationship between *distance to the transport network CAT* and the presence or absence of landslides, but, as the landslide inventory was obtained by field work, probably it had an intrinsic influence of this explanatory variable with respect to the dependant variable. As a matter of fact, it was only in view of the susceptibility maps A, B and C when the biased effect of *distance to the transport network CAT* was manifested (Fig. 6.7). In these maps it can be observed that the almost complete concentration of the highest susceptibility classes were found

around the communication routes.

Following this reasoning, it was concluded that the susceptibility maps A, B and C were not operationally acceptable. Even though the comparison of the validation results obtained by applying different strategies of explanatory variables processing allowed, actually, the identification of the procedure followed in model C as the most suitable one. Considering the prediction capacity of these 3 models, model A presented good results (Figs. 6.5 and 6.6), but due to the small size of some categories, the appearance of extreme values of the coefficients could strongly condition the results. The same happened in the case of model B, which showed slightly lower results, probably due to the subjective division that was applied to continuous variables for their conversion into categorical. As an example of this effect, stood out the high absolute values of the β estimates associated with the higher elevation classes, which mask any possible influence of the rest of the variables. However, despite its high prediction performance, it has to be taken into account the large extension covered by high susceptibility classes in map B (see Figs. 6.5 and 6.6), which strongly reduces the discrimination capacity of this model. In addition, processing all variables as categorical by means of the dummy codification supposes a considerable increase of calculation time and the interpretation of the β estimates turns more laborious. Finally, model C presented the best results with excellent balance between the prediction and discrimination capacity. So the transformation of categorical variables into continuous giving a relative value based on the presence of landslides to each of its class, such as the landslide density value, allowed mitigating the effect of the smaller categories offering a more robust susceptibility model, as suggested by Grozavu *et al.* (2013) and Trigila *et al.* (2015).

Once defined the most suitable strategy for the explanatory variables processing, and due to the detection of *elevation* and *distance to the transport network CAT* as inadequate variables, model D was carried out. As a result, it presented a good prediction and discrimination capacity with an AUC value of the prediction rate curve of 0.89 (Fig. 6.6). Moreover, the graphical representation of the model (Figs. 6.7 and 6.8) showed a spatial distribution of high susceptibility areas slightly marked by lithological types, but much more diffused by the rest of the explanatory variables, which offers a satisfactory landslides susceptibility map.

To conclude, the tests performed during this study reached answering the questions formulated at the beginning of this chapter, but also raised new issues to be consider in the application of the LR model for the definition of a landslide susceptibility map in general, and for the case of our study area in particular:

- In the particular case of the Oria river basin, and consequently neither in Gipuzkoa, bibliographical landslide inventories are not of enough accuracy for modelling the landslide susceptibility. Instead, direct geomorphological field survey demonstrated to be a suitable alternative.

- Statistically oriented explanatory variables selection approach is effective for eliminating no-significant variables, though being statistically significant does not ascertain the suitability of a given variable. It is compulsory to ensure the independence between the dependant and explanatory variable, as well as a geomorphological justification about their relationship.

- Transforming categorical variables into continuous is more advantageous for the proper execution of landslide susceptibility maps using the LR model, since this procedure avoids the creation of a large number of dummy variables, while it maintains the maximum prediction capacity.

- Landslide inventory could provide more accurate information if the areas of each landslide would be delimited.

- The definition of the places free of landslides could be a considerable source of incertitude, even more if the landslide inventory was carried out by direct geomorphological field survey.

References

Amorim, S. F.: Estudio comparativo de métodos para la evaluación de la susceptibilidad del terreno a la formación de deslizamientos superficiales: Aplicación al Pirineo Oriental, Ph.D. thesis, Universidad Politécnica de Catalunya, Barcelona, 2012.

Bai, S. B., Wang, J., Lü, G.Ñ., Zhou, P. G., Hou, S. S., & Xu, S.Ñ.: GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China, Geomorphology, 115, 23–31, 2010.

Bonachea, J.: Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debidos a procesos geomorfológicos, Ph.D. thesis, Universidad de Cantabria, Santander, 2006.

Carrara, A.: Multivariate models for landslide hazard evaluation, Mathematical Geology, 15, 403–426, 1983.

Chung, C. J. & Fabbri, A. G.: Validation of spatial prediction models for landslide hazard mapping, Natural Hazards, 30, 451–472, 2003.

Corominas, J. & Mavrouli, O. C.: Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies, Tech. rep., SafeLand. 7th Framework Programme Cooperation Theme 6 Environment (including climate change) Sub-Activity 6.1.3 Natural Hazards, 2011.

Costanzo, D., Chacón, J., Conoscenti, C., Irigaray, C., & Rotigliano, E.: Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy), Landslides, 11, 639–653, 2014.

Dai, F. & Lee, C.: Landslide characteristics and slope instability modeling using GIS, Lantau Island, Hong Kong, Geomorphology, 42, 213–228, 2002.

Duman, T. Y., Can, T., Gokceoglu, C., Nefeslioglu, H. A., & Sonmez, H.: Application of logistic regression for landslide susceptibility zoning of Cekmece Area, Istanbul, Turkey, Environmental Geology, 51, 241–256, 2006.

Fdez-Arroyabe, P. & Martin-Vide, J.: Regionalization of the probability of wet spells and rainfall persistence in the Basque Country (Northern Spain), International Journal of Climatology, 32, 1909–1920, 2012.

Felicísimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10, 175–189, 2013.

Fiorucci, F., Giordan, D., Santangelo, M., Dutto, F., Rossi, M., & Guzzetti, F.: Criteria for the optimal selection of remote sensing optical images to map event landslides, Natural Hazards and Earth System Sciences, 18, 405–417, 2018.

González-Hidalgo, J. C., Brunetti, M., & de Luis, M.: A new tool for monthly precipitation analysis in Spain: MOPREDAS database (monthly precipitation trends December 1945–November 2005), International Journal of Climatology, 31, 715–731, 2011.

Grozavu, A., Pleşcan, S., Patriche, C. V., Mărgărint, M. C., & Roşca, B.: Landslide susceptibility assessment: GIS application to a complex mountainous environment, in: The Carpathians: Integrating Nature and Society Towards Sustainability, edited by Kozak, J., Ostapowicz, K., Bytnerowicz, A., Wyżga, B., *et al.*, pp. 31–44, Springer, 2013.

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, 31, 181–216, 1999.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang,

K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Reviews, 112, 42–66, 2012.

Hervás, J.: ALISSA: Abridged Landslide Inventory of Spain for synoptic Susceptibility Assessment, in: EGU General Assembly Conference Abstracts, vol. 16, 2014.

Lee, S.: Application of logistic regression model and its validation for landslide susceptibility mapping using GIS and remote sensing data, International Journal of Remote Sensing, 26, 1477–1491, 2005.

Mücher, C. A., Klijn, J. A., Wascher, D. M., & Schaminée, J. H.: A new European Landscape Classification (LANMAP): A transparent, flexible and user-oriented methodology to distinguish landscapes, Ecological Indicators, 10, 87–103, 2010.

Nefeslioglu, H., Gokceoglu, C., & 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, Engineering Geology, 97, 171–191, 2008.

Nefeslioglu, H. A., Gokceoglu, C., Sonmez, H., & Gorum, T.: Medium-scale hazard mapping for shallow landslide initiation: the Buyukkoy catchment area (Cayeli, Rize, Turkey), Landslides, 8, 459–483, 2011.

Pardo, A. & Ruiz, M. A.: SPSS 11. Guía para el análisis de datos, McGraw-Hill/Interamericana de España, Madrid, 2002.

Pourghasemi, H., Moradi, H., & Aghda, S. F.: Landslide susceptibility mapping by binary logistic regression, analytical hierarchy process, and statistical index models and assessment of their performances, Natural Hazards, 69, 749–779, 2013.

Remondo, J., Bonachea, J., & Cendrero, A.: Quantitative landslide risk assessment and mapping on the basis of recent occurrences, Geomorphology, 94, 496–507, 2008.

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., Natural Hazards and Earth System Sciences, 15, 2111–2126, 2015.

Trigila, A., Iadanza, C., Esposito, C., & Scarascia-Mugnozza, G.: Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy), Geomorphology, 249, 119–136, 2015.

Van Den Eeckhaut, M. & Hervás, J.: State of the art of national landslide databases in Europe and their potential for assessing landslide susceptibility, hazard and risk, Geomorphology, 139, 545–558, 2012.

Van Den Eeckhaut, M., Hervás, J., Jaedicke, C., Malet, J. P., Montanarella, L., & Nadim, F.: Statistical modelling of Europe-wide landslide susceptibility using limited landslide inventory data, Landslides, 9, 357–369, 2012.

Wang, Y. T., Seijmonsbergen, A. C., Bouten, W., & Chen, Q. T.: Using statistical learning algorithms in regional landslide susceptibility zonation with limited landslide field data, Journal of Mountain Science, 12, 268–288, 2015.

Weier, J. & Herring, D.: Measuring vegetation (ndvi & evi), Earth Observatory Library of NASA, 2000.

Yilmaz, I.: Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat Turkey), Computers & Geosciences, 35, 1125–1138, 2009.

Zhu, L. & Huang, J. f.: GIS-based logistic regression method for landslide susceptibility mapping in regional scale, Journal of Zhejiang University-Science A, 7, 2007–2017, 2006.

II Effective surveyed area and its role in statistical landslide susceptibility assessment

After the tests carried out in the Oria river basin, the knowledge obtained by that research was applied in the complete study area, in order to create a landslide susceptibility map for the Gipuzkoa Province (see section 4). But considering the issues raised in the previous chapter, more than obtaining the definitive susceptibility map for our study area, the aim of this phase of the study was to continue investigating about the crucial methodological steps in landslide susceptibility modelling.

In statistical landslide susceptibility models, as the LR model adopted in this work, the preparation of the training dataset is a fundamental and critical step. Commonly, this requires the selection of a sample of stable (without landslides) and unstable (with landslides) mapping units. While assuring the presence of a landslide is straightforward, and it can be supported by the geomorphological signatures on the slope or by direct observation of the events, the selection of landslide-free areas is more critical. Assuming as landslide-free the locations of a study area where no landslides were reported in a field survey is correct only in the unlikely circumstance that the landslide inventory has been prepared surveying every single site of the study area, and following homogeneous criteria. In other words, any landslide-free location in an inventory map should have been explicitly checked to be free from landslides.

Nowadays, there are methods based on the visual interpretation of aerial photographs or digital processing of remotely acquired optical and radar imagery (Catani *et al.*, 2005; Herrera *et al.*, 2009; Fiorucci *et al.*, 2011; Casagli *et al.*, 2017; Mondini, 2017; Fiorucci *et al.*, 2018; Alvioli *et al.*, 2018) that allow to prepare historical and event landslide inventories. However, the adoption of such methods can be hampered by the lack of image accuracy classification performance due to uncertain factors. Alternatively, bibliographic sources like newspapers and news feeds, administrative reports or scientific literature can be used to obtain landslide information. Nevertheless, the downside of these type of data is that they hardly are as accurate as required by landslide susceptibility studies, like it was demonstrated in the Oria river basin. As a consequence, sometimes, like in this case, the best option

to obtain a reliable landslide inventory is a straightforward geomorphological field mapping. A detailed discussion about the characteristics, advantages and limitations of different approaches for landslide mapping can be found in Guzzetti *et al.* (2012); Santangelo *et al.* (2015) and Fiorucci *et al.* (2018).

An operational disadvantage of field-based landslide mapping is the difficulty in surveying the whole area where the susceptibility map must be carried out, since some places can be inaccessible or not visible from the accessible places. Difficulties in surveying the landscape affect the completeness and the spatial representativeness of the landslide inventory and, as a result, inclusion of non-visible areas within a landslide inventory introduces a bias, since presence or absence of landslides cannot be ascertained in such portions of landscape. This uncertainty has hardly been considered in existing studies that use field-based landslide inventories (Yesilnacar & Topal, 2005; Wang *et al.*, 2017).

In this work, we considered grid cells and slope units (Carrara *et al.*, 1991, 1995; Guzzetti *et al.*, 2006; Alvioli *et al.*, 2016; Zêzere *et al.*, 2017; Rosi *et al.*, 2018; Ba *et al.*, 2018), and investigated the effect of the different ways of training landslide susceptibility models within both types of mapping units.

We propose an automatic and reproducible procedure to delineate the actual area which was explicitly surveyed in preparing a landslide inventory by geomorphological field mapping, *i.e.* the effective surveyed area (ESA), and to use such relevant information in the statistical analysis. The procedure allows to carry out the calibration of the statistical model within the ESA and then to apply the resulting susceptibility model to the whole area (WA) under investigation. Moreover, we implemented an automatic approach for the delineation of the ESA in a newly developed module named *r.survey.py* (see section 6.7.3). The module delineates the theoretical visible areas from the points of view recorded during the field trip by the GPS tracks. And, most importantly, the ESA, as delineated by *r.survey.py*, is an objective and reproducible portion of the study area directly observed by the geomorphologist, thus allowing to avoid arbitrary assumptions about which sites were actually surveyed and which ones were not.

In particular, this research aims to achieve the third and fourth objectives of the current thesis. On one hand, it is intended to observe and recognize the advantages

and drawbacks of different mapping units in landslide susceptibility mapping. On the other hand we aim at demonstrating that the calibration of a landslide susceptibility model within the ESA, instead of the WA (the whole study area, encompassing the ESA), enhances the performance of model itself. Thus, we calibrated the multivariate logistic regression model for landslide susceptibility in four different ways, combining two different calibration areas (ESA and WA) with two different mapping unit types: (i) a regular grid cell partition with a ground resolution of 5 m x 5 m and (ii) an slope unit (SU) partition (consisting in irregular terrain subdivisions bounded by drainage and divided lines)(see section 6.7.4).

Detailed information about the area in which the following investigation was applied can be found in chapter 4. Thereby, the next sections are organized as follows. Section 6.7 shows the details about the data acquisition. Section 6.8 contains a general description about the software specifications used for applying the multivariate method to model the landslide susceptibility and the approach followed for validate it, as well as a detailed description about the set-up of the different models. Results are described in Section 6.9 and are further discussed in Section 6.10.

6.7 Data preparation

6.7.1 Landslide inventory

We prepared a landslide inventory by a direct geomorphological field survey, as it is detailed in section 5.1.1.2. Moreover, and also important to define the ESA, we digitalized the route followed during the field survey. This information was then elaborated using a module developed for the purpose.

As a result of several field trips, 793 individual landslides were collected; 746 of them were classified as shallow movements (Fig. 6.9a). Our observations together with the revised bibliographical sources (INGEMISA, 1995; GFA, 2013; Euskadiko DEA, 2014) confirmed that shallow slides are the most frequent type of landslide, just like in the Oria river basin. Consequently, in order to consider only landslides triggered by the same mechanisms, only shallow movements were used as landslide presence when defining the dependent variable in the susceptibility assessment.

Figures 6.9b and 6.9c show the distribution of landslide sizes, highlighting that a difference of five orders of magnitude exists between the smallest and the largest inventoried shallow slide.
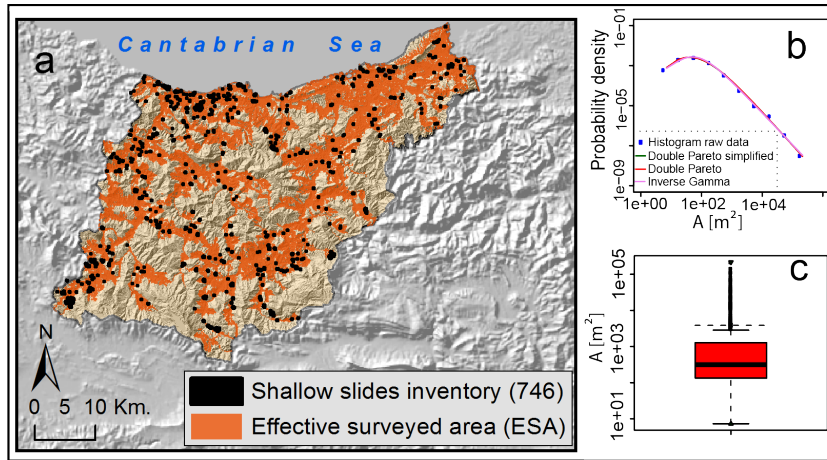
Figure 6.9: *(a) Distribution of the shallow slides inventory along the study area and extension of the Effective Surveyed Area (ESA); (b) Probability density plot of the shallow landslide size (Area in m^2) distribution; (c) Box plot of the same distribution.*
6.9 Irudia: (a) Azaleko lur labaiketen inbentarioaren banaketa espaziala eta ikuskatutako eremu efektiboaren (ESA) zabalkundea; (b) Lur labainketen azaleren dentsitate probabilitatearen banaketa kurba (m^2-tan); (c) Banaketa berdinaren bloke diagrama.

6.7.2 Explanatory variables

The group of explanatory variables used for this research, respectively to those used in the Oria river basin, changed due to different reasons. According to the results obtained in the previous chapter, *distance to transport network* and its reclassified version *distance to transport network CAT* were rejected in order to avoid probable biased effect due to the fact that the landslide inventory was done by direct field work. *elevation* showed negligible effect. *distance to the river*, as well as its reclassified version *distance to the main river-streams CAT*, were not chosen either because of the high drainage density, which reduces the variability between stable and unstable mapping units. In the previous research (see Tabs. 6.5 and 6.6), *land cover 2* showed always lower overall classification indexes in front of *land cover 1*, so it was decided that it was no reasons for still test its suitability. And finally, *NDVI* could not be used for the application in the Gipuzkoa Province, because this layer did not cover the complete study area.

As a result, a group of 13 explanatory variables were used for this investigation. As morphometric continuous variables we used *slope, sinusoidal slope, surface area ratio (SAR), topographic wetness index (TWI), curvature, planform curvature* and *profile curvature*. And as categorical variables *lithology, permeability, regolith thickness, land cover 1* and *aspect* were considered, besides an additional new variables which was not used in the Oria river tests, namely *land cover 3*.

For the categorical variables, we computed frequency ratio (FR) values for each class, and used them as an alternative relative value (instead of the LD value) for their transformation into continuous variables (Lee & Min, 2001; Yilmaz, 2009; Trigila *et al.*, 2015). FR is a concept introduced by Lee & Min (2001), which has been widely used in bivariate statistical approaches on the field of landslide susceptibility modelling (Süzen & Doyuran, 2004; Gorsevski *et al.*, 2006; Yilmaz, 2009). Also, Trigila *et al.* (2015) suggested its utilisation as relative numerical value for transforming categorical variables into continuous, in order to use them in multivariate statistical approaches. Although the concept behind is similar to the LD, the formula for computing it is slightly different.

$$FR_i = \frac{L_i/L_{tot}}{NL_i/NL_{tot}} \tag{6.2}$$

Where L_i and L_{tot} are the amount of mapping units with landslides in class i and in the total study area, respectively, while NL_i and NL_{tot} are the corresponding landslide-free quantities.

In the context of this research, where a comparison was carried out between two different calibration areas, we acknowledge that the FR values can vary depending on the portion of the territory considered as the total area (ESA or WA). However, in order to perform a direct comparison, we decided to maintain the same FR values (calculated considering the WA) in all regular grid cell-based susceptibility analysis (the summary table with all the FR values corresponding to each class of the categorical variables is shown in Appendix D).

The selection of the appropriate explanatory variables to build the landslide susceptibility models was also carried out, in this case, by an statistically oriented approach detailed in section 5.2.3.2. We first adopted grid cells as mapping units, and applied the cited approach to ensure that only significant variables were taken

into account as well as the non-redundancy of the contributed information by each covariate (Ayalew & Yamagishi, 2005). Then, considering the variables actually used for the application of the LR models with grid cells, we have further restricted the set of variables to be used with slope units (see section 6.9.2).

6.7.3 Definition of the effective surveyed area

In this work we suggest the concept of ESA, and training of statistical models therein, as an approach to be used to train a landslide susceptibility model avoiding assumptions about the presence or absence of landslides in areas not explicitly observed. We delimited the ESA by means of the newly developed python module, namely *r.survey.py*, which can be run by means of GRASS SIG[1] software (Bornaetxea *et al.*, 2018) (information about the repository in which the original code can be found is available in Appendix B). Input data to define the visible area (*i.e.* ESA in our case) are: i) a sample of points to be considered as points of view; ii) a DEM of the area; iii) the maximum visible distance. The sample of points of view, in our case, was defined re-sampling a given number of points along the recorded path during the field campaigns. This number of points depends on the maximum distance set between them, and together with the DEM resolution selected the results can be directly affected.

In a 10 km^2 subset of the study area, we tested the software output using: i) maximum distance between sampled points of 50, 100, 200 and 500 m; ii) the original DEM at 5 m resolution and resampled versions of the DEM at 20, 50 and 100 m resolution; and iii) maximum visible distance of 500 m (the later was dictated by the largest distance between the digitized field path and the farthest landslide pixel in the subset of the study area). Results of the test are summarized in Table 6.9.

As target criteria, we considered that the best setting option was the one which allows covering the totality of the landslides but using the less possible points (bigger D_{max} value) and the lower possible resolution in order to optimize the calculation time. In the case of the complete study area, the maximum visible distance was set in 1,100 m, in view that the largest distance between the digitized field path and

---

[1]https://grass.osgeo.org/

Table 6.9: *Results of the setting test of r.surbey in a 10 km^2 subset of the study area.*

6.9 Taula: Ezarpenen testaren emaitzak ikerketa eremuko 10 km^2-ko azpi-eremuan.

Name	Resolution (m)	D$_{max}$	Percentage of landslides within (%)
Survey 5	5	50	35
Survey 6	20	50	70
Survey 7	50	50	95
Survey 8	100	50	100
Survey 9	5	100	30
Survey 10	20	100	60
Survey 11	50	100	95
Survey 12	100	100	100
Survey 13	5	200	30
Survey 14	20	200	55
Survey 15	50	200	85
Survey 16	100	200	100
Survey 17	5	500	0
Survey 18	20	500	35
Survey 19	50	500	60
Survey 20	100	500	95

the farthest landslide pixel was 1,092 m. Then, and according to the results of Table 6.9, the rest of the settings were fixed: maximum sampling distance of 200 m, DEM resolution of 100 m.

We can make sense of the numerical values of the parameter used in the *r.survey.py* module considering that the minimum size A of an object visible from a distance Δ is given by Rodrigues *et al.* (2010) and Minelli *et al.* (2014):

$$A = \frac{25\,\Delta^2}{c}, \tag{6.3}$$

where c is a steradiant to square minutes conversion factor, $c \simeq 1.18{\cdot}10^7$. Using $\Delta = 1,100$ m in Eq. (6.3), we get $A = 2.6$ m^2, meaning that the smallest landslide in our inventory, with size 7.3 m^2, would actually be identifiable from at least one point along the route, if the landslide sits within the ESA. The resulting ESA covered 44.24% of the entire study area and it is shown in Fig. 6.9a.

6.7.4   Slope units delineation

For SU delineation we adopted the *r.slopeunit* module described in Alvioli *et al.* (2016) (information about the repository in which the original code can be found is available in Appendix B). The code provides a GRASS GIS module, as the *r.survey.py* code presented in this work, and it was designed for the automatic and adaptive delineation of SUs, given a DEM and a set of user-defined input parameters. The code can be used to produce several SU partitions, using different combinations of the input parameters, which can thus be tuned according to user-defined criteria. We partially followed Alvioli *et al.* (2016), in that we selected the best SU partition maximizing the quality of terrain aspect segmentation. In addition, we performed preliminary tests using the LR susceptibility model, showing that the use of very small SUs provides unrealistic results, which can be understood considering the limited variability of variables within such small SU polygons. We concluded that, in the case of the Gipuzkoa Province the most suitable SU partition for landslide susceptibility zonation should be obtained with the following *r.slopeunits* input parameters: flow accumulation area threshold $t = 1$ km^2; minimum SU planimetric area $a = 0.15$ km^2; minimum circular variance of terrain aspect within each SU $c = 0.2$; reduction factor $r = 5$; threshold value for the cleaning procedure *cleansize* =

0.025 km^2. As a result, we obtained a set of SUs which range in size from 0.026 km^2 to 3.6 km^2 with average 0.28 km^2. A discussion of SU delineation and optimization of input parameters can be found in Alvioli *et al.* (2016) and Schlögel *et al.* (2018), and it is out of the scope of this work.

6.8   Modelling framework

Four landslide susceptibility maps (LS maps) were prepared by means of a statistical approach. All the maps were obtained by means of a multivariate LR model. Classification performances were measured by means of a set of validation tests explained in the following sections. We prepared the first two maps using 5 m x 5 m regular grid cells as mapping units. These two maps differ because in one case the LR model was calibrated within the WA, and within the ESA in the other case (Figs. 6.14a, c). The third and fourth LS maps, instead, were prepared with different mapping units, namely with SUs instead of grid cells, where calibration data were also changed considering data within WA in one case and within ESA in the other (Figs. 6.14b, d). We end up with four maps, which we name as follows: WA-PM (whole area, pixel map), ESA-PM (effective surveyed area, pixel map), WA-SUM (whole area, slope unit map) and ESA-SUM (effective surveyed area, slope unit map).

6.8.1   Statistical analysis

All the statistical analysis carried out in this research were done by means of a R software's module designed for the specific purpose of landslide susceptibility statistical assessment, called LAND-SE (Rossi & Reichenbach, 2016). In particular, we used logistic regression (see section 5.2.1), one of the multivariate statistical approaches available in the LAND-SE module, to build the landslide susceptibility models in the test study area.

This tool allows to obtain the conventional β estimates, that maximize the agreement between the model equation, i.e. landslide probability, and empirical landslide data, in training area. But additionally, the implementation of the glm (general linear model) function by the LAND-SE software is such that it is possible

to investigate the estimated standard error of a t-statistic for the null hypothesis of each of the coefficients of the linear model. The p-value represents the probability that the parameter is zero: for p-values much smaller than 0.05 the null hypothesis (vanishing coefficient) is rejected, thus the associated variable is significant for the final result. So, p-value can be considered as an objective indicator for the selection of the most relevant variables to be used in the statistical model (Schlögel *et al.*, 2018).

6.8.2   Evaluation of model performance

The performance of statistical susceptibility models was evaluated comparing its predictions with independent landslide data. Concerning this point, the definition of training and validation input samples was crucial to detect how well fitted each model to the input data itself, but also how valid was each model to predict unknown data. So, these performances were used to evaluate the pairwise comparisons.

The statistical metrics used in this research for this purpose were: (i) confusion matrices (contingency tables) and their graphical representation (four-fold or contingency plots); (ii) Receiver Operating Characteristic (ROC) curves and their associated Area Under Curve (AUC) value; (iii) classification error plots; and (iv) Cohen's Kappa index. Detailed explanation about each test can be found in section 5.2.4.

In this study the probability of landslide occurrence resulting from each model estimate (trained either within the ESA or within the WA) and for each considered mapping unit (either grid cells or slope units), was reclassified in five landslide susceptibility classes which were labelled as *Very low* (for susceptibility values in the range 0-0.2), *Low* (0.2-0.45), *Medium* (0.45-0.55), *High* (0.55-0.8) and *Very high* (0.8-1).

Moreover, in order to spatially identify the pairwise matching degree between different model estimates, mismatch maps were prepared (see section 5.2.4). Each mapping unit was reclassified as stable or unstable considering a threshold value of 0.5. The different maps, all prepared with the same mapping unit partition, were overlapped. Then, the mismatch degree between grid cell and SU susceptibility maps was quantified in terms of number of mismatched mapping units and overall

mismatched area.

6.8.3   Data selection for landslide susceptibility

The DEM available for the study area consists of $7.91 \cdot 10^7$ cells with 5 m resolution. For landslide susceptibility assessment, both using grid cells (*i.e.* pixel based) and SUs, we prepared raster layers corresponding to each available explanatory variable, aligned to the DEM grid cells.

We established a rigorous sampling procedure to minimize possible statistical biases during training/validation partition. The procedure is slightly different for the grid cell and SU mapping units cases.

In the first case, a grid cell was considered unstable if it is located within any landslide area, and stable if it is outside the landslide boundaries. In the second case, an SU was considered unstable depending on the percentage of landslide area present within it. In any case, the 75% of the unstable mapping units together with a similar amount of stable mapping units were used to train the LR model, and the remaining 25%, also together with a similar amount of stable mapping units, for validation. The choice of an equal number of stable and unstable mapping units was done on purpose, and it is the standard procedure required by the LAND-SE software for landslide susceptibility assessment, because the LR model requires a balanced dataset, in which the number of stable and unstable cases are similar (Felicísimo *et al.*, 2013; Costanzo *et al.*, 2014).

For regular grid cell-based models, we selected at random 558 landslides (75%) for model training, and converted them into raster layers (84,623 unstable pixels). The remaining 188 landslides (25%), used for validation, were also rasterized (29,247 unstable pixels). This is at variance with the usual random selection of unstable pixels, in which a given percentage of grid cells are sampled within landslide. Here we select whole landslides, and consider all the pixels encompassed by the landslide bodies as training/validation samples. We ran the experiment with three different training/validation random sets, containing the above percentages, and selected the one with the best classification results. This exercise allowed us to confirm that the random selection of the landslide inventory would not affect the model results in a relevant way, because in all the cases the model classification performances were very

similar. In order to choose one single data set to be the same for further comparative analysis, the data set with the best classification result was selected. Then, training sets were selected as follow: 84,623 unstable pixels and an equal number of stable pixels. These two sets were selected at random first within WA and then within ESA. We ensured that unstable pixels were exactly the same in the two cases, because we wanted the only difference to be that the stable pixels were sampled within the WA, in the first case, and within the ESA, in the second case. Finally, in order to guarantee the comparability of the prediction performances, one unique validation sample was created as follow: the remaining 29,247 unstable pixels together with an equal number of stable pixels selected at random among the remaining stable pixels within the ESA.

Concerning the SU-based models, we first partitioned the study area in 6,907 SUs with the technique outlined in Section 6.7.4. SU boundaries do not match those of the dependent or explanatory variables layers, allowing the presence of different classes, or values, inside each SU. Moreover, the presence of one single landslide pixel within a slope unit was not considered enough to label this SU as unstable. Therefore, instead of arbitrarily defining a given threshold value in order to consider a SU as unstable, we decided to use the overall landslide density in the WA. For this reason, we considered as unstable those SUs containing equal or more than 0.15% of unstable pixels, and stable otherwise. We used as explanatory variables the mean and the standard deviation of the morphometric variables for each SU and the percentage of the area covered by each class of the categorical layers. In 304 cases the SU contained 0.15% or more unstable pixels, so we selected at random 228 of them (75%) for training, and the remaining 76 (25%) were used for validation. Like in grid cell approaches, we created two different training samples where unstable SUs were exactly the same, and only the stable SUs vary in each case. The first training sample includes 228 stable SUs selected at random along the WA. The second training sample includes an equal number of stable SUs units selected at random among those that at least partially overlap the ESA. It is true that considering all the SUs that only overlap the ESA could introduce into the model some that slightly are within, but whose most part stay outside. Though, in our case, we observed that this happened fewer times than the opposite. Thus,

considering that the ESA is an approximation of the real surveyed area, we decided, as a conservative measure, to take into account every SU that overlap the ESA, even though we acknowledge that it still can introduce some incertitude to the model. But in any case this incertitude should always be less than considering the WA, where SUs completely outside the ESA - thus, with big probability not observed -, can be selected to train the susceptibility model. Additionally, 76 SUs labelled as unstable were reserved from the total for validation. Then, the validation sample was completed by adding a random selection of the same number of SUs labelled as stable and which at least partially overlap the ESA. Thus, the validation sample contained 152 SUs (76 unstable + 76 stable).

6.9 Results

6.9.1 Susceptibility maps using grid cells

We ran the LR model using the pixel-based datasets twice: once using the entire training pixel sample and once using the effective training pixel sample as dependent variables. We defined the obtained results as whole area pixel map (WA-PM) and effective surveyed area pixel map (ESA-PM), respectively.

Both in WA-PM and ESA-PM, we first used the same 13 explanatory variables, listed in table 6.10, and then we selected for each model assessment, the most relevant explanatory variables considering the collinearity between each pair of variables and the significance (p-value) of the regression estimates (see section 6.7.2).

Figure 6.10 shows the result of the pairwise collinearity analysis among the 13 explanatory variables. In particular, the figure shows the values of the correlation coefficients and their graphical representation using ellipses with eccentricity and colour intensity proportional to the degree of mutual correlation and with ellipses orientation and colour indicating direct (rightward-increasing blue ellipses) or indirect correlation (leftward-increasing red ellipses). We flagged as collinear two variables, once again, when their correlation coefficient is greater than 0.5 with a significance level of 0.01. In such a case, as an objective criterion for variable selection, the variable with highest p-value (showed in Tab. 6.10) between the two, was rejected from the final run of the LR model.

Table 6.10: *Set of environmental variables introduced for WA-PM and ESA-PM models calculation, together with the significance p-value corresponding to each explanatory variable. The final predictors are labelled with an asterisk and their corresponding β estimate coefficient is shown.*

6.10 Taula: WA-PM eta ESA-PM modeloetan erabilitako aldagaien zerrenda eta hauen p-balioa. Azken kalkuluan erabilitako aldagaiak asterisko batez azpimarratu dira eta horien β koefizienteak erakusten dira.

Variable	WA-PM		ESA-PM	
	p-value	β coef.	p-value	β coef.
Continuous				
slope	$1.17 \cdot 10^{-189}$		$1.06 \cdot 10^{-111}$	
sinusoidal slope	$1.00 \cdot 10^{-155}$		$7.57 \cdot 10^{-134}$ *	0.418
surface area ratio	$3.743 \cdot 10^{-203}$ *	-0.242	$1.89 \cdot 10^{-99}$	
topographic wetness index	$9.864 \cdot 10^{-10}$ *	0.022	0.127	
curvature	0.909		0.526	
planform curvature	0.909		0.526	
profile curvature	0.909		0.526	
Categorical				
lithology	0 *	0.894	0 *	1.125
permeability	$1.496 \cdot 10^{-33}$ *	0.227	$7.632 \cdot 10^{-72}$ *	0.401
regolith thickness	0 *	0.58	0 *	0.378
land cover 1	$5.14 \cdot 10^{-291}$		$1.42 \cdot 10^{-87}$	
land cover 3	0 *	0.99	$1.596 \cdot 10^{-173}$ *	0.498
aspect	0 *	0.997	0 *	1.153
Intercept		-3.958		-4.157
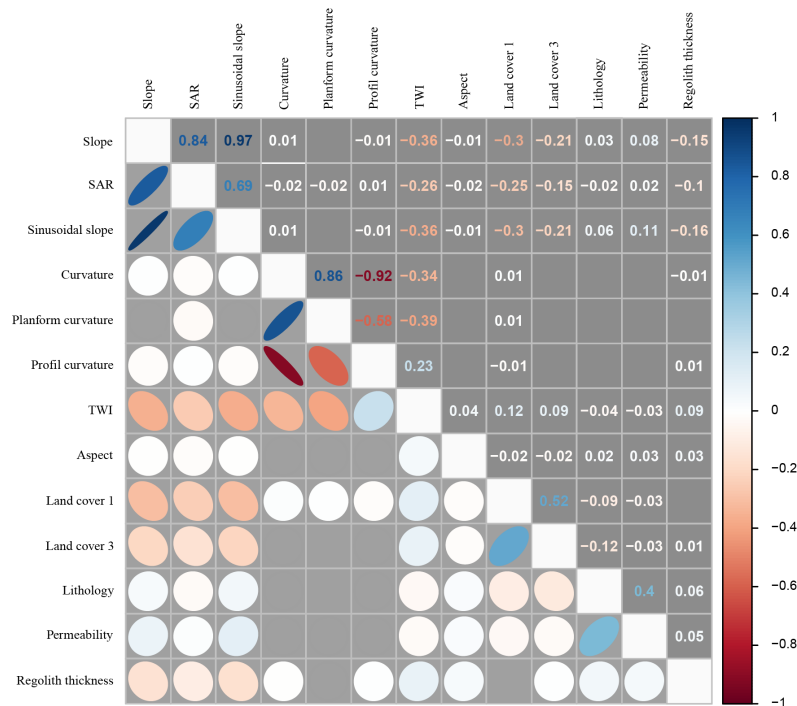
Figure 6.10: *Correlation matrix. The correlation coefficient is only shown if the corresponding significance level is lower than the threshold value of 0.01.*

6.10 Irudia: Korrelazio matrizea. Korrelazio koefizientea bere esangura maila 0.01eko atalasea baino baxuagoa denean bakarrik agertzen da.

As a result of the variable selection procedure we selected a slightly different set of explanatory variables as the most suitable predictors for each model (Tab. 6.10). *aspect, lithology, permeability* and *regolith thickness* presented no correlations between any other variable and they were always associated with a p-value lower or equal than 0.05, so we used all of them for building both WA-PM and ESA-PM. The three curvature variables (*curvature, planform curvature* and *profile curvature*) showed in all cases p-values over the threshold of 0.05, so they were rejected as final predictors. *land cover 1* and *land cover 3* performed a high correlation coefficient between them (0.52) (Fig. 6.10), but the latter always showed a lower p-value, so *land cover 3* was selected as a final explanatory variable. *slope, sinusoidal slope* and *surface area ratio* (*SAR*) are three highly correlated variables, but according to their p-value we chose each time a different option. In WA-PM, *SAR* was chosen as suitable predictor, whereas in ESA-PM, *sinusoidal slope*. Finally, *Topographic wetness index* (*TWI*) did not present high correlations with any other explanatory variable, but the relation with the variability of the dependent variable changed. Thereby, in WA-PM we considered it as one of the final predictors (p-value < 0.05),

while in ESA-PM, it was rejected (p-value > 0.05) (Tab. 6.10).

As a result, for each case, only the variables marked with an asterisk in table 6.10 were introduced in the final LR models, whose corresponding β estimates allowed drawing the final landslide susceptibility maps (Figs. 6.14a and c).

Using the validation pixel sample, we evaluated the prediction skills of the pixel susceptibility maps (Fig. 6.11). Inspection of the four fold, or contingency, plots reveals that WA-PM predicted correctly the 63.58% (TP+TN) of the observed unstable and stable mapping units, whereas ESA-PM was capable to correctly predict a higher amount of mapping units (65.45%). The *ROC* curves also indicated better prediction skills in ESA-PM (AUC = 0.7) than in WA-PM (AUC = 0.68) and the same happened for the Cohen's Kappa index ($k = 0.309$ versus $k = 0.272$). Moreover, the classification error plots (Figs. 6.11c, f) provided an estimate of the error associated with the predicted susceptibility values, which did not exceed 0.1 standard deviations in any case, highlighting the reliability of the results. And finally, the mutual mismatch map (Fig. 6.14e) showed that 14.8% (corresponding to an extension of 293 km^2) of the mapping units flipped their landslide susceptibility class in WA-PM and ESA-PM.

6.9.2   Susceptibility maps using slope units

Due to the subdivision of categorical variables by means of its classes; and the mean and standard deviation calculations for morphometric variables, the introduction of the original 13 explanatory variables would result in 56 new variables in which many of them (all those classes belonging to the same categorical variable) would be highly correlated. For this reason, the variable selection approach used in the pixel-based case was not viable when working with SUs and a specific variable selection approach for SU models would require further investigation. Thus, for this work, the most appropriate set of explanatory variables, among those considered as the most relevant in pixel-based model assessment, was selected by expert criteria. Considering such set of variables as a starting point, we selected new sets of explanatory variables to evaluate landslide susceptibility using SUs, *i.e.* to calculate the whole area slope unit map (WA-SUM) and the effective area slope unit map (ESA-SUM). Taking into account that the automatic procedure for the

Whole Area Pixel Map (WA-PM)

Cohen's k	AUC$_{ROC}$	Overall Accuracy	Overall Error Rate
0.272	0.68	63.58%	36.42%
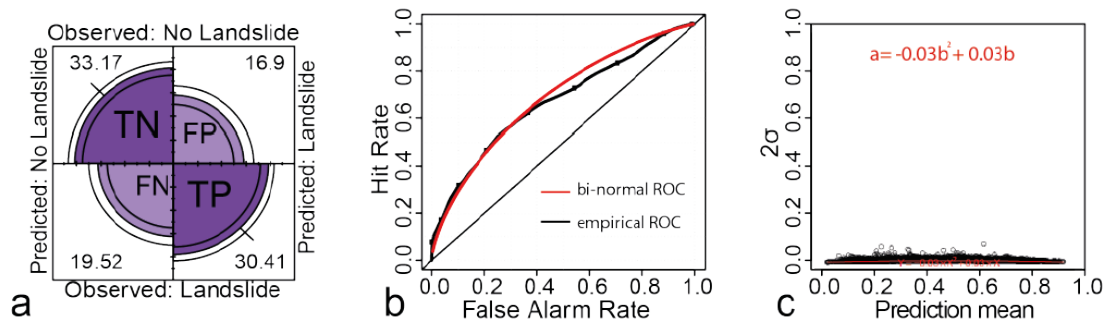
Effective Surveyed Area Pixel Map (ESA-PM)

Cohen's k	AUC$_{ROC}$	Overall Accuracy	Overall Error Rate
0.309	0.7	65.45%	34.55%
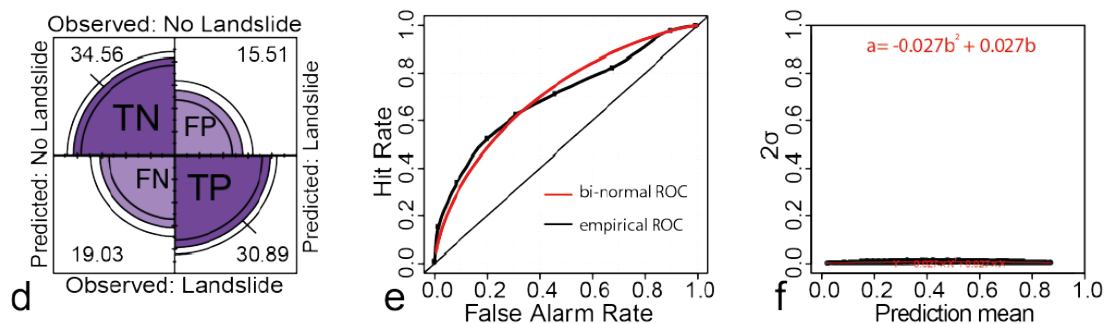
Figure 6.11: *Pixel-based LR models prediction performance results: summary tables of the Cohen's Kappa index, area under the ROC curve (AUC), overall accuracy ((TP+TN)/(TP+TN+FP+FN)) and overall error rate ((FP+FN)/(TP+TN+FP+FN)); (a,d) four fold or contingency plots; (b,e) ROC curves; (c,f) classification error plots and the quadratic regression fit curves (red line).*

6.11 Irudia: Pixeletan oinarritutako LR modeloen aurreikuspen emaitzak: Cohenen Kappa indizea, ROC kurbaren azpiko azalera (AUC), asmatze tasa eta errore tasa balioak; (a,d) kontingentzia diagramak; (b,e) ROC kurbak; (c,f) klasifikazio errore diagramak eta erregresio kuadratiko tendentzia kurba (gorriz).

SUs definition already included the flow accumulation calculation, used for *TWI* estimation, and the aspect component, we rejected *aspect* and *TWI* to avoid spurious correlations. We selected the following set of variables used to produce both pixel-based maps such as *lithology, permeability, regolith thickness* and *land cover 3*, and we added *slope*. The reason for choosing *slope* over *sinusoidal slope* or *SAR* is due to the fact that these two are derivative variables of the former. Moreover, we consider *slope* more suitable feature to describe the average morphology within SU than *sinusoidal* slope or *SAR*, so we decided to select it in order to simplify interpretation of the results.

Using the validation SU sample, we assessed the prediction skills of the SU maps. For the WA-SUM the 65.13% of the 152 validation mapping units were correctly classified (TP+TN) (Fig. 6.12a). The ROC curve provided an AUC value of 0.69, and the corresponding Cohen's Kappa was 0.302 (Fig. 6.12b). Concerning the classification error plot (Fig. 6.12c), it can be observed that in the SUs with high and low landslide susceptibility probability (probability > 0.8 and < 0.2) the 2σ value stayed below 0.2, but variability in the estimates became larger for intermediate susceptibilities. This reveals a considerable variation in the stable/unstable classification of the territory, which implies a low reliability, at least for the intermediate probabilities (Guzzetti *et al.*, 2006). For the ESA-SUM, the 63.82% of the 152 validation mapping units were correctly classified (TP+TN) (Fig. 6.12d) with AUC = 0.71, slightly larger with respect to the other SU model assessment, whereas, the Cohen's Kappa index performed slightly worse, being k = 0.276 (Fig. 6.12). The classification error plot showed a considerable variation in intermediate probabilities (Fig. 6.12f) while the uncertainty was lower for high and low probabilities. Nevertheless, the quadratic fit curves indicated a lower overall variability for ESA-SUM than for WA-SUM.

Visual inspection of the SU susceptibility maps (Figs. 6.14b, d) showed similarities between WA-SUM and ESA-SUM. The difference is graphically presented through the mismatch map (Fig. 6.14f), where 12.6% of the mapping units (corresponding to an extension of 247 km^2) changed their landslide susceptibility class, from WA-SUM to ESA-SUM.

Whole Area Slope Unit Map (WA-SUM)

Cohen's k	AUC$_{ROC}$	Overall Accuracy	Overall Error Rate
0.302	0.69	65.13%	34.87%
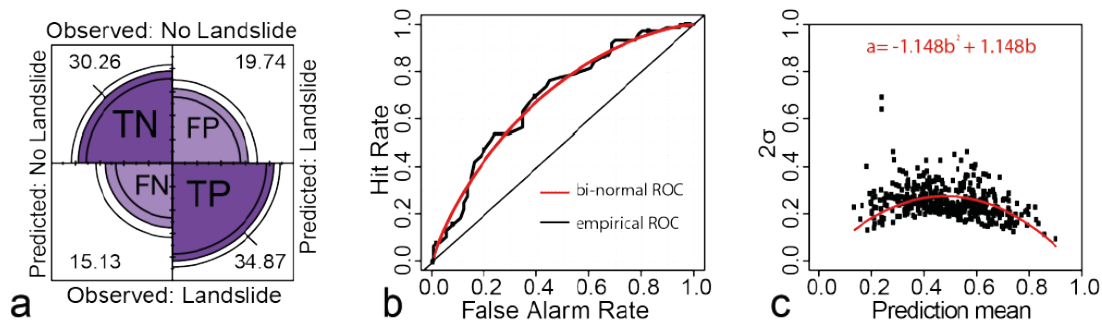
Effective Surveyed Area Slope Unit Map (ESA-SUM)

Cohen's k	AUC$_{ROC}$	Overall Accuracy	Overall Error Rate
0.276	0.71	63.82%	36.18%
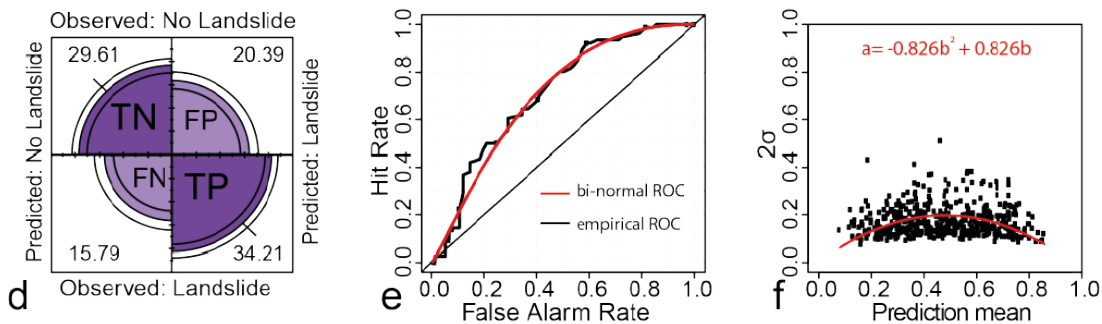
Figure 6.12: *SU-based LR models prediction performance results: summary tables of the Cohen's Kappa index, area under the ROC curve (AUC), overall accuracy ((TP+TN)/(TP+TN+FP+FN)) and overall error rate ((FP+FN)/(TP+TN+FP+FN)); (a,d) four fold or contingency plots; (b,e) ROC curves; (c,f) classification error plots and the quadratic regression fit curves (red line).*

6.12 Irudia: Malda unitatetan oinarritutako LR modeloen aurreikuspen emaitzak: Cohenen Kappa indizea, ROC kurbaren azpiko azalera (AUC), asmatze tasa eta errore tasa balioak; (a,d) kontingentzia diagramak; (b,e) ROC kurbak; (c,f) klasifikazio errore diagramak eta erregresio kuadratiko tendentzia kurba (gorriz).

6.9.2.1 Sensitivity test for landslide presence/absence threshold

The threshold limit to consider a SU as unstable can be a key issue in susceptibility maps carried out by this irregular mapping unit partition. In our particular case, the number of SUs containing at least one landslide pixel was 497, and the distribution of the landslide density area among all of them can be shown in the following figure 6.13. In order to confirm that the definition of the threshold did not affect the conclusions of this investigation, we carried out calculations using as a threshold the 5th percentile (P_5, threshold 0.013%), the 50th percentile (P_{50}, threshold 0.265%) and the 90th percentile (P_{90}, threshold 4.5%) of areal landslide distribution, along with the average landslide density calculated within the ESA, i.e. 0.33%. The resulting values of the area under the ROC curve between ESA and WA approaches for these tests are summarized in the table below (Tab. 6.11).
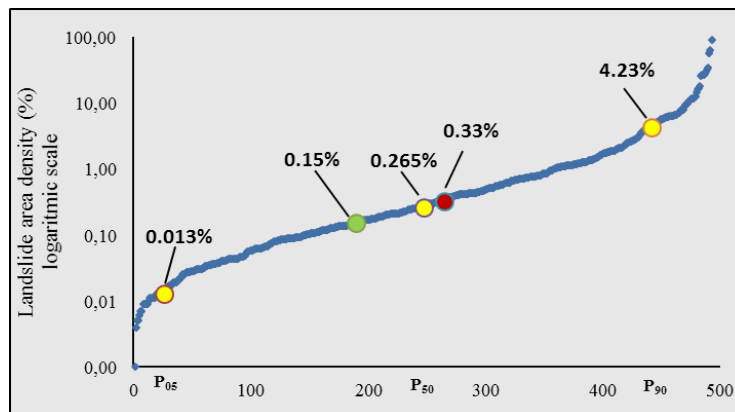
Figure 6.13: *Distribution of the landslide area density among slope units containing at least one landslide pixel, in logarithmic scale. The following points are highlighted: Overall landslide density in the study area in green; Landslide desnsity in ESA in red; 5th, 50th and 90th percentiles in yellow.*

6.13 Irudia: Malda unitateen barneko lur labainketen azalera dentsitatea eskala logaritmikoan. Ondorengo puntuak azpimarratuta daude: Ikerketa eremu osoko lur labainketa dentsitatea berdez; ESA barruko dentsitatea gorriz; 5., 50. eta 90. pertzentilak horiz.

6.10 Discussion

In this work, we showed that the information contained in a field-based landslide inventory for landslide susceptibility analysis should be critically examined, also in combination with the mapping unit of choice.

A field work-based landslide inventory is by definition a source of uncertainty in statistical analysis, owing to various reasons, including mapping errors, accuracy,

Table 6.11: *Comparison of AUC$_{ROC}$ values between ESA-SUM and WA-SUM for different landslide presence/absence thresholds, and the percentage of SUs classified as unstable for each threshold among SUs containing at least one landslide pixel.*

6.11 Taula: AUC$_{ROC}$ balioen konparaketa ESA-SUM eta WA-SUM artean lur labainketa presentzia/ausentzia atalase desberdinetarako eta ezegonkor bezala klasifikatutako malda unitateen portzentaia.

	P$_{05}$ (0.013%)	**WA$_{density}$** (0.15%)	**P$_{50}$** (0.265%)	**ESA$_{density}$** (0.33%)	**P$_{90}$** (4.496%)
ESA-SUM	0.659	0.71	0.679	0.619	0.556
WA-SUM	0.657	0.69	0.672	0.586	0.692
Unstable SUs	96.4%	61.2%	50.1%	46.5%	10.06%

subjectivity, and others. The focus of this work is the analysis of an additional uncertainty due to use of field mapping, namely the fact that it is impossible to ensure that the study area was surveyed in an homogeneous way. An objective delimitation of the surveyed area by means of the ESA, proposed in this work along with a module to objectively delineate the ESA, is one way to reduce this uncertainty.

The hypothesis tested in this work is that any statistical landslide susceptibility model trained inside the ESA is by definition more correct than considering the entire study area for training the model, if such ESA is representative of the WA. The statement was borne out by the results of multivariate LR model calculations. We acknowledge that the ESA is only an approximation of the real surveyed area, though a much more realistic one than using the whole study area. Our definition of the ESA depends on the maximum distance between points along the field, trips paths and the selected resolution of the DEM. Preliminary tests in a reduced portion of the territory provided the most suitable settings for a satisfactory definition of the ESA in the particular case of Gipuzkoa Province (section 6.7.3).

In the case of the pixel-based susceptibility maps, the metrics of model prediction performances were in agreement with our main statement about the relevance of ESA. As a matter of fact, all the validation performance tests (confusion matrix metrics, the area under the ROC curve and Cohen's Kappa index) presented an improvement if the stable pixels used for training the LR model are selected within the ESA (like in ESA-PM, Fig. 6.11a) than if they were taken from the WA (like in WA-PM, Fig. 6.11b). In addition, the almost flat classification error plots in both
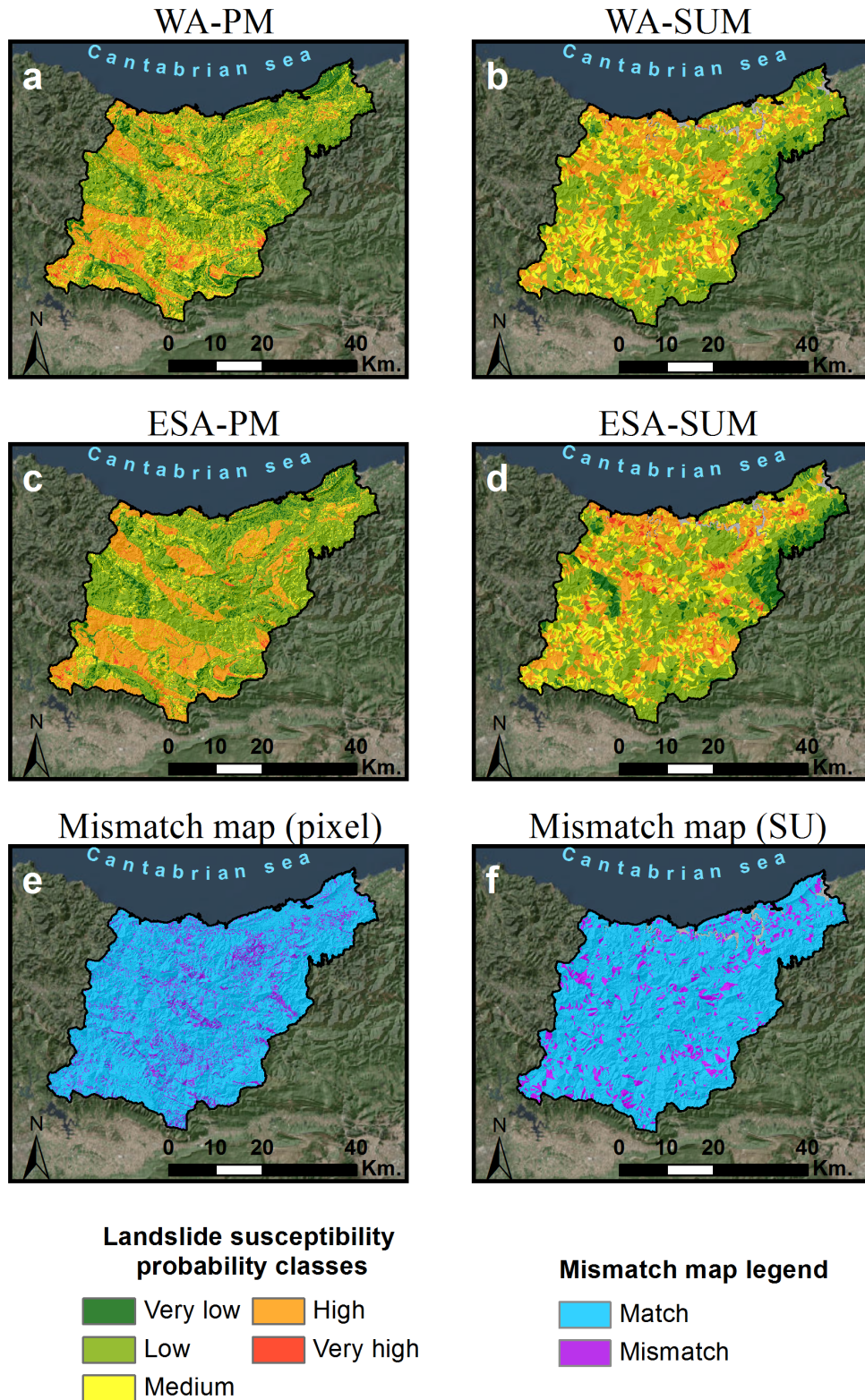
Figure 6.14: *(a-d) Landslide susceptibility maps represented in five classes for WA-PM, WA-SUM, ESA-PM and ESA-SUM; (e,f) Mismatch maps representing the spatial distribution of the mapping units differently classified using ESA between pixel models and slope unit models.*

6.14 Irudia: (a-d) WA-PM, WA-SUM, ESA-PM eta ESA-SUM suszeptibilitate mapak; (e,f) ESA erabiltzean pixel eta malda unitate mapen arteko klasifikazio desberdintasunak adierazten dituzten mapak.

cases (Figs. 6.11c, f) show high stability of model results. The spatial distribution of the susceptibility classes were different as well between ESA-PM and WA-PM (see Figs. 6.14a, c), and such differences were highlighted in the mutual mismatch map (Fig. 6.14e).

Another difference between the two pixel maps is the set of explanatory variables selected as predictors. The variables selection approach presented in this section and previously adopted in a similar way in Schlögel *et al.* (2018), demonstrated to be effective and capable to detect presence of redundant information, as well as offering an objective way to choose between collinear explanatory variables.

In the case of SU-based susceptibility maps, validation metrics do not present us with clear-cut results as in the pixel-based maps. As a matter of facts, AUC performs better in ESA-SUM while Confusion Matrix and Cohen's Kappa index present higher prediction performance in WA-SUM (Fig. 6.12). The classification error plots show considerable variations in intermediate susceptibility probability values, but the quadratic fit curves suggest a slightly lower variability in ESA-SUM (Figs. 6.12c, f). We interpret these results as an indication of a smaller effect that proper usage of the ESA can have in SU-based susceptibility maps, with respect to pixel-based maps. Despite the small difference in model prediction performance between WA-SUM and ESA-SUM, the reduction of the mismatch degree (Fig. 6.14f) suggests that the usage of the ESA is equally recommendable for SU susceptibility maps carried out by field work landslide inventories.

Moreover, since the threshold value for distinguishing stable and unstable SUs could affect the LR model performances, we performed a sensitivity test evaluating the LR models, for both the WA and ESA, using different presence/absence thresholds (Fig. 6.13 and Tab. 6.11). We observed that for all the cases, except in P_{90}, the model tests showed better performance for ESA-SUM than for WA-SUM, which is proof that the conclusions obtained following any approach were indistinguishable. We note that because of the high threshold defined in P_{90}, the model was trained with a very small sample of unstable SUs, which gives to the result a very poor reliability. On the other hand, in the P_5 case, the unbalance does not take place, since each SU where at least one landslide pixel exists belongs to the unstable class, resulting in minimum yet relevant number of unstable SUs. Therefore, we maintain

that results of the test confirm that SUs mitigate the relevance of the calibration area (ESA versus WA) when building an SU-based susceptibility model with a field-based landslide inventory, independently of the landslide presence threshold value. However, we acknowledge that the search of an optimal threshold value that ensures a balanced sample is a relevant point, though it is beyond the scope of this work.

The pixel- and SU-based maps obtained within the method presented in this work are inherently different from a conceptual point of view. We maintain that a SU-based map probably represents a better option, for SUs bear a clear relation with topography, they reduce mapping errors and are more useful for practical (planning) purposes. Nevertheless, for the sake of completeness and to show differences between the two approaches, we discussed pixel-based and SU-based maps independently. The uncertainty introduced by a field work-based landslide inventory can be mitigated by using SUs, resulting in more similar susceptibility maps and validation performances in WA-SUM and ESA-SUM than in pixel models.

We acknowledge that the overall performances of the landslide susceptibility maps presented in this section are of moderate to low prediction capacity, with AUC values ranging between 0.68 to 0.71 and an overall accuracy which hardly overcomes the 65% in the best case (Figs. 6.11 and 6.12). This could be due to (i) the lack of more complete landslide inventory -since the field work was not developed in the whole study area, besides the fact that completeness refers to the proportion of landslides shown in the inventory compared to the real (and most of the times unknown) number of landslides in the study area (Guzzetti *et al.*, 2012; Malamud *et al.*, 2004)- or (ii) the use of not up-to-date thematic layers.

Thereby, though the preparation of a definitive landslide susceptibility map for the study area was out of the scope of our investigation. We performed pairwise comparative analyses in which we only changed, across the compared model assessments, the region of logistic regression training, and the results illustrated by this investigation support the following conclusions:

- When working with pixel mapping units, training the LS model within the ESA is the correct approach to reduce the uncertainty inherent to the landslide inventory.

- When working with slope unit terrain partition this uncertainty can be

mitigated, even though it is still advantageous to train the LS model within the ESA.

- Use of ESA should be considered, if sufficient information is available, in preparing landslide susceptibility maps with any multivariate statistical model.

- Collecting information about the path followed during field campaigns for landslide mapping is a meaningful procedure for estimating the ESA, at model assessment time, using the module *r.survey.py.*

References

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F.: Automatic delineation of geomorphological slope units with *r.slopeunits v1.0* and their optimization for landslide susceptibility modeling, Geoscientific Model Development, 9, 3975–3991, 2016.

Alvioli, M., Mondini, A., Fiorucci, F., Cardinali, M., & Marchesini, I.: Topography-driven satellite imagery analysis for landslide mapping, Geomatics, Natural Hazards and Risk, 9, 544–567, 2018.

Ayalew, L. & Yamagishi, H.: The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan, Geomorphology, 65, 15–31, 2005.

Ba, Q., Chen, Y., Deng, S., Yang, J., & Li, H.: A comparison of slope units and grid cells as mapping units for landslide susceptibility assessment, Earth Science Informatics, pp. 1–16, 2018.

Bornaetxea, T., Rossi, M., Marchesini, I., & Alvioli, M.: Effective surveyed area and its role in statistical landslide susceptibility assessments, Natural Hazards and Earth System Sciences, 18, 2455–2469, 2018.

Carrara, A., Cardinali, M., Detti, R., Guzzetti, F., Pasqui, V., & Reichenbach, P.: GIS techniques and statistical models in evaluating landslide hazard, Earth Surface Processes and Landforms, 16, 427–445, 1991.

Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P.: GIS technology in mapping landslide hazard, in: Geographical Information Systems in Assessing

Natural Hazards, edited by Carrara, A. & Guzzetti, F., pp. 135–175, Springer, 1995.

Casagli, N., Frodella, W., Morelli, S., Tofani, V., Ciampalini, A., Intrieri, E., Raspini, F., Rossi, G., Tanteri, L., & Lu, P.: Spaceborne, UAV and ground-based remote sensing techniques for landslide mapping, monitoring and early warning, Geoenvironmental Disasters, 4, 9, 2017.

Catani, F., Farina, P., Moretti, S., Nico, G., & Strozzi, T.: On the application of SAR interferometry to geomorphological studies: estimation of landform attributes and mass movements, Geomorphology, 66, 119–131, 2005.

Costanzo, D., Chacón, J., Conoscenti, C., Irigaray, C., & Rotigliano, E.: Forward logistic regression for earth-flow landslide susceptibility assessment in the Platani river basin (southern Sicily, Italy), Landslides, 11, 639–653, 2014.

Euskadiko DEA: Mapa geomorfológico de Euskadi, URL `www.geo.euskadi.eus`, 2014.

Felicísimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10, 175–189, 2013.

Fiorucci, F., Cardinali, M., Carlà, R., Rossi, M., Mondini, A., Santurri, L., Ardizzone, F., & Guzzetti, F.: Seasonal landslide mapping and estimation of landslide mobilization rates using aerial and satellite images, Geomorphology, 129, 59–70, 2011.

Fiorucci, F., Giordan, D., Santangelo, M., Dutto, F., Rossi, M., & Guzzetti, F.: Criteria for the optimal selection of remote sensing optical images to map event landslides, Natural Hazards and Earth System Sciences, 18, 405–417, 2018.

GFA: Evaluación y gestión integada de riesgos geotécnicos en la red de carreteras de la Diputación Foral de Gipuzkoa, Tech. rep., Mugikortasun eta Bide Azpiegituren Saila, 2013.

Gorsevski, P. V., Gessler, P. E., Foltz, R. B., & Elliot, W. J.: Spatial prediction of landslide hazard using logistic regression and ROC analysis, Transactions in GIS, 10, 395–415, 2006.

Guzzetti, F., Reichenbach, P., Ardizzone, F., Cardinali, M., & Galli, M.: Estimating the quality of landslide susceptibility models, Geomorphology, 81, 166–184, 2006.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Reviews, 112, 42–66, 2012.

Herrera, G., Fernández-Merodo, J., Mulas, J., Pastor, M., Luzi, G., & Monserrat, O.: A landslide forecasting model using ground based SAR data: The Portalet case study, Engineering Geology, 105, 220–230, 2009.

INGEMISA: Inventario y Análisis de las Áreas sometidas a Riesgo de Inestabilidades del Terreno de la C.A.P.V., Tech. rep., Eusko Jaurlaritza, 1995.

Lee, S. & Min, K.: Statistical analysis of landslide susceptibility at Yongin, Korea, Environmental Geology, 40, 1095–1113, 2001.

Malamud, B. D., Turcotte, D. L., Guzzetti, F., & Reichenbach, P.: Landslide inventories and their statistical properties, Earth Surface Processes and Landforms, 29, 687–711, 2004.

Minelli, A., Marchesini, I., Taylor, F. E., De Rosa, P., Casagrande, L., & Cenci, M.: An open source GIS tool to quantify the visual impact of wind turbines and photovoltaic panels, Environmental Impact Assessment Review, 49, 70–78, 2014.

Mondini, A. C.: Measures of Spatial Autocorrelation Changes in Multitemporal SAR Images for Event Landslides Detection, Remote Sensing, 9, 554, 2017.

Rodrigues, M., Montañés, C., & Fueyo, N.: A method for the assessment of the visual impact caused by the large-scale deployment of renewable-energy facilities, Environmental Impact Assessment Review, 30, 240–246, 2010.

Rosi, A., Tofani, V., Tanteri, L., Stefanelli, C. T., Agostini, A., Catani, F., & Casagli, N.: The new landslide inventory of Tuscany (Italy) updated with PS-InSAR: geomorphological features and landslide distribution, Landslides, 15, 5–19, 2018.

Rossi, M. & Reichenbach, P.: LAND-SE: a software for landslide statistically-based susceptibility zonation, Version 1.0, Geosci. Model Dev., pp. 3533–3543, 2016.

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., Natural Hazards and Earth System Sciences, 15, 2111–2126, 2015.

Schlögel, R., Marchesini, I., Alvioli, M., Reichenbach, P., Rossi, M., & Malet, J. P.: Optimizing landslide susceptibility zonation: Effects of DEM spatial resolution and slope unit delineation on logistic regression models, Geomorphology, 301, 10–20, 2018.

Süzen, M. L. & Doyuran, V.: A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate, Environmental Geology, 45, 665–679, 2004.

Trigila, A., Iadanza, C., Esposito, C., & Scarascia-Mugnozza, G.: Comparison of Logistic Regression and Random Forests techniques for shallow landslide susceptibility assessment in Giampilieri (NE Sicily, Italy), Geomorphology, 249, 119–136, 2015.

Wang, F., Xu, P., Wang, C., Wang, N., & Jiang, N.: Application of a GIS-Based Slope Unit Method for Landslide Susceptibility Mapping along the Longzi River, Southeastern Tibetan Plateau, China, ISPRS International Journal of Geo-Information, 6, 172, 2017.

Yesilnacar, E. & Topal, T.: Landslide susceptibility mapping: a comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey), Engineering Geology, 79, 251–266, 2005.

Yilmaz, I.: Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: a case study from Kat landslides (Tokat Turkey), Computers & Geosciences, 35, 1125–1138, 2009.

Zêzere, J., Pereira, S., Melo, R., Oliveira, S., & Garcia, R.: Mapping landslide susceptibility using data-driven methods, Science of the Total Environment, 589, 250–267, 2017.

III Landslide and rainfalls. Press inventory, rainfalls characterization and precipitation thresholds for Gipuzkoa Province (Basque Country)

Investigation of the methodological practices on the field of landslide susceptibility are directed toward its application in the next logical steps, which are the definition of the hazard, and subsequently the risk that slope instabilities produce. This section presents the results of a methodological application where the relation between landslide occurrence and the precipitated rain was considered as a preliminary step toward landslide hazard and forecast studies.

As it was highlighted in the previous two studies presented in this thesis, there is a large amount of environmental conditions that made the slope vulnerable to failure. But, together with these conditioning factors, it can be assumed that it is a single event that finally a landslide is initiated at a given moment, which is called the triggering factor.

Rainfall is one of the most usual landslide triggering factors which affects to the terrain by building up of the water pressure into the ground and modifying the balance of the slope (Guzzetti *et al.*, 2007), and there are many researches that attempted to determine the amount of precipitation needed to the mobilization of the slopes (Wieczorek, 1987; Corominas & Moya, 1999; Dai & Lee, 2001; Li *et al.*, 2011; Brunetti *et al.*, 2010; Ramos-Cañón *et al.*, 2015; Zêzere *et al.*, 2015; Piciullo *et al.*, 2016; Valenzuela *et al.*, 2018). Nevertheless, for some specific cases, such as rock falls, quantitative relation with rainfalls are difficult to found citeplongoni2012definition. Moreover, the problem is that precipitation thresholds can vary a lot depending on the study area conditions and also depending on the type of data used for its definition (Zêzere *et al.*, 2015). Moreover, in the bibliography there are different approaches to calculate it. In Guzzetti *et al.* (2007) it is more in depth discussed the variety of the possibilities that are available in order to empirically define a landslide initiation rainfall threshold.

By means of this application we searched to discuss if it is possible to detect the principal features of rainfalls responsible of landslides, so we aimed at detecting qualitative relations between the inventoried landslides and the rainfall events, as well as the calculation of precipitation thresholds responsible of landslides in

Gipuzkoa Province. Thereby, a sequential work flow was defined in three levels:

(i) To carry out an inventory of those landslides that were reported in newspapers, as well as their description, localization and characterization considering some spatial conditioning factors.

(ii) To characterize the precipitation events occurred between 2006 and 2015 for different rain gauges within Gipuzkoa Province.

(iii) To detect qualitative relations between the inventoried instabilities and the mentioned rainfall events together with the calculation of a landslide responsible precipitation thresholds.

So, in order to achieve the presented objectives, we put in practice the approach proposed by Melillo *et al.* (2015) and Melillo *et al.* (2016) for automatically determining a landslide responsible precipitation threshold by means of an algorithm for the objective reconstruction of rainfall events responsible for landslides, created in the Istituto di Ricerca per la Protezione Idrogeologica of Perugia (IRPI). It is a tool programmed in R open-source software for advanced statistical computing and graphics, release 2.15.2 [2], which groups a continuous record of rainfall measurements considering a minimum dry period before and after each rainfall event, and then, each landslide in the temporal record can be associated with a single rainfall event. The algorithm was applied in Gipuzkoa Province (see section 4) considering a temporal precipitation data set of 10 years and a landslide inventory based on press reports whose temporal information were available for the same period of time. Additionally, spatial information about *slope*, *lithology* and *land cover* was crossed with the landslide inventory. And furthermore, the outputs of the software were also used to obtain an overview of the relation existing between the precipitation patterns of Gipuzkoa Province and the occurrence of landslides.

As in the previously presented studies, methodological details related to the results that are presented below can be found in chapter 5, and according to the mentioned work flow, results are shown up as follow. First the set of landslide points resulting from the press news survey is shown together with a general

---

[2]http://www.r-project.org

description of the obtained data base and a synthetic characterisation of them according to three typically used environmental variables, such as *slope*, *lithology* and *land cover* (section 6.11.1). After that, the set of rainfall events defined by the automatic algorithm for our study area is described and analysed (section 6.11.2). And then, landslide responsible precipitation thresholds resulting from the probabilistic analysis carried out by the algorithm are presented (section 6.11.3). Finally, all the results are discussed in section 6.12.

6.11   Results

6.11.1   Landslide inventory and its characterization

As a result of the press news survey detailed in section 5.1.1.3, an inventory of 339 different landslide events was obtained for a period of 10 years (Fig. 6.15). Notice that the frequency distribution of the inventoried landslides dates were mainly concentrated in January, February and November, even though we can find a few instability events in every month (Fig. 6.16). Among all these terrain instabilities, in 326 the cause of the event was defined as rainfall (77%) or unknown (19%), and all the rest was related to other triggering factors different from rainfalls, such as human activity (2%), waves (1%) or fluvial erosion (0.5%). For the sake of homogeneity, only the subset in which the cause was confirmed as rainfall or it was unknown was used as input for further analysis. We assumed that, with big probability, those news in which the cause of the landslide was not specified made reference, actually, to high precipitation events (see Cause of landslide occurrence summary graph in Fig. 6.16).

In the most part of the surveyed reports (49%), only the day of the landslide occurrence was detailed and in the 29% of cases the exact time was concreted. Nevertheless, the information about the type of movement was seldom provided, using in the very most reports the generic term of "landslide". About the damages produced by the inventoried landslides, it can be observed in the damage graph (Fig. 6.16) how the most part affected to human infrastructures such as the transportation network (56%), buildings (33%) or parks (6%), whereas, only 4% did not cause damage or they were undefined.

On the other hand, among all the data set, 44% of landslides were localized
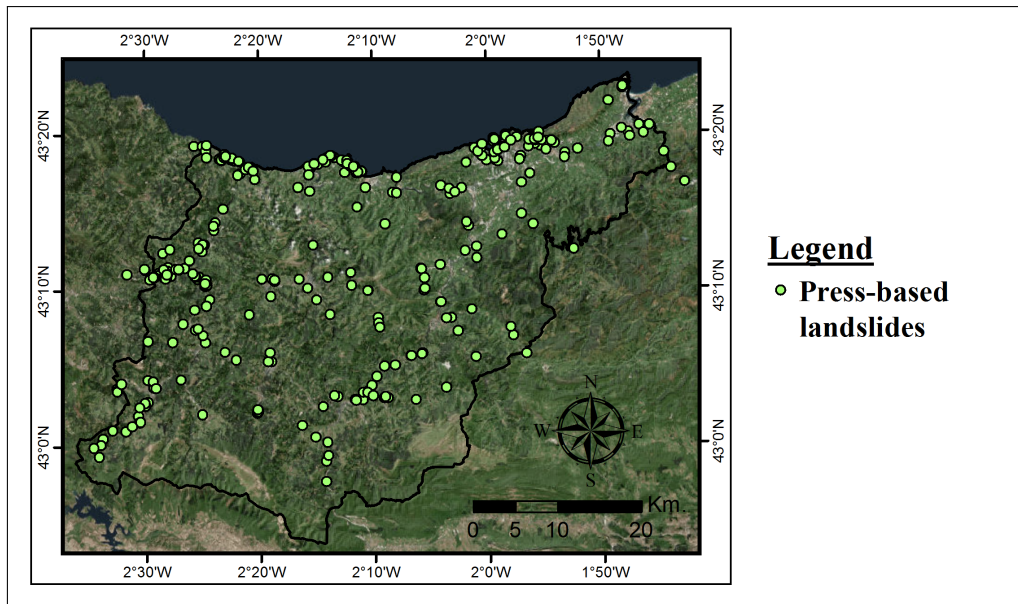
Figure 6.15: *Spatial distribution of the landslide inventory obtained by press news survey.*

6.15 Irudia: Egunkari berrien behaketaren bitartez lortutako lur labainketen inbentarioaren banaketa espaziala.

with exact accuracy, which were used to carry out the landslides characterization (see Location accuracy graph in Fig. 6.16). To do so, three environmental variables typically used for landslide susceptibility analysis were compared with landslides locations, and the relative frequencies of the instabilities belonging to each class were compared to the relative extension of the same class in the study area (Fig. 6.17). The selected environmental variables were the *slope*, the *lithology* and the *land cover*. In order to calculate the frequencies, *slope* was reclassified in 5 classes as follow: 0°-15°; 15°-30°; 30°-45°; 45°-60°; more than 60°. In case of *lithology*, the simplified reclassification was used (see section 5.1.2.2). And for *land cover* the simplified version of *land cover 3* layer was used (see section 5.1.2.2).

The *slope* distribution indicated that the majority of the inventoried terrain instabilities happened in areas with 15° to 60° of inclination. The most part of the study area presented moderate slope (15°-30°) where landslides were frequent, but the biggest amount of landslides was hosted in slopes between 30° and 45° of inclination, although this class covered less space than the former (Fig. 6.17a). It was also highlighted the big amount of landslides occurring in slopes between 45° and 60° of inclination, even though the extension of the study area in this range was scarce (1% of the study area). On the other hand, notice that a considerable part of the inventoried instabilities happened in slopes with an inclination under 15°.
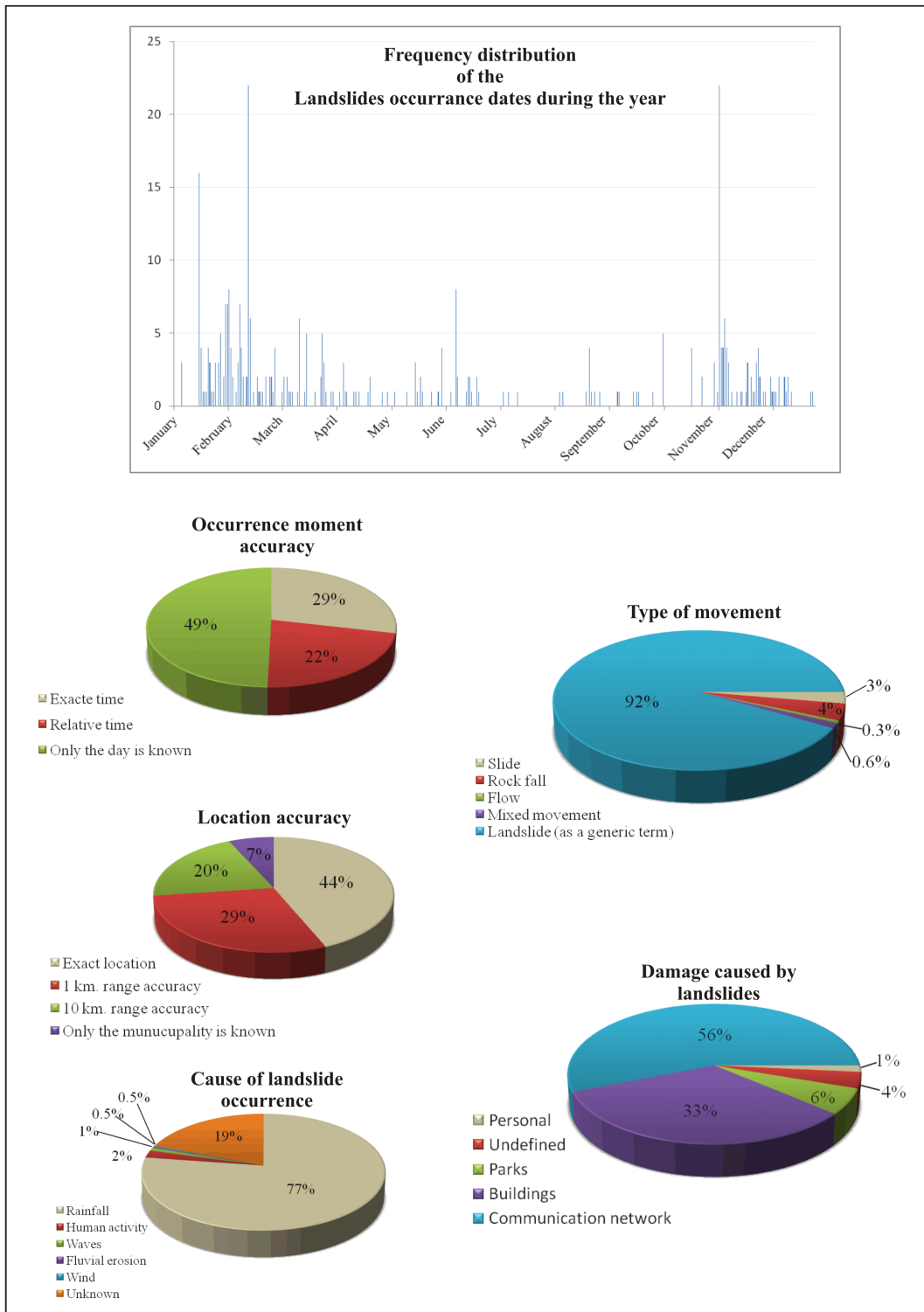
Figure 6.16: *Summary graphs of the press-based landslide inventory.*

6.16 Irudia: Egunkari berrietan oinarritutako lur labainketa inbentarioaren laburpen grafikoa.

If we take under consideration the lithological distributions, landslides belonging to limestone or surface deposits overcame the spatial distribution of these classes in the study area, which suggested an important susceptibility to landslide occurrence. In addition, clay and detrital rock together with marls also presented a considerable amount of landslides, but indeed, they were the most common lithologies along the study area, covering almost 60% of the territory (Fig. 6.17b).

Concerning the *land cover* distribution it was highlighted the big amount of slope instabilities falling in urban areas, whereas the presence of this class in the territory did not reach the 5%. Grassland also showed a greater relative frequency on landslides than in the study area for the analysed inventory. Unlike, forest displayed a lower relative frequency in landslides respect to the area covered by these classes, because although around the 30% of the landslides occurred in forested areas, this *land cover* class occupy almost the 60% of the territory (Fig. 6.17c).

6.11.2   Characterization of rainfalls

Concerning the precipitation data (obtained from the Meteorological Agency of the Basque Country[3]), preliminary controls pointed out the lack of too many records (more than one complete month missing) in two of the rain gauges, so finally only the information from 22 rain gauges out of the original 24 was used for this work (see Fig. 5.21).

Considering this precipitation information a comparison between the general rainfall characteristics in Gipuzkoa Province and the characteristics of the rainfalls that triggered landslides was carried out following the method explained in section 5.3.2. First, for each of the 22 rain gauges the relative frequency of the rainfall event classes defined in tables 5.5 and 5.6 was calculated (an average of 275 rainfall events per rain gauge). Then, the average of the 22 results for each class was represented in figures 6.18 and 6.19.

During the analysis ran by the algorithm, 23 landslides, among the first 326, were discarded because they could hardly be related with any rainfall event, and other 5 were manually discarded by the authors because of their confusing information, so at the end, 298 landslides were related with the rainfall events.

---

[3]www.euskalmet.euskadi.eus
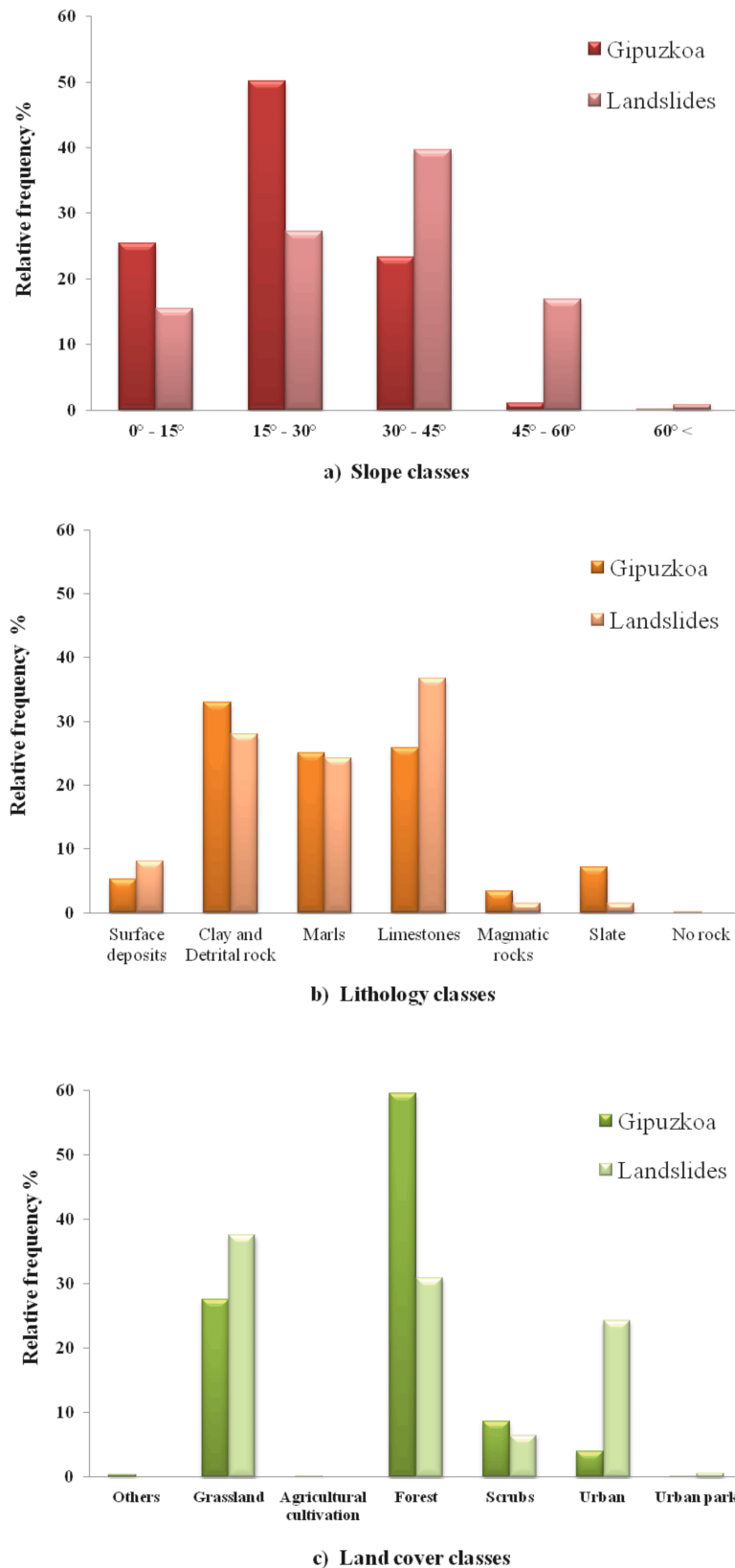
Figure 6.17: *Study area (Gipuzkoa Province) and landslides relative frequency distribution among (a) slope; (b) lithology and (c) land cover classes.*

6.17 Irudia: Ikerketa eremuaren (Gipuzkoako LH) eta lur labainketen frekuentzia erlatiboaren banaketa (a) Maldan; (b) Litologian eta (c) Lur estalduran.

Figure 6.18 shows the frequency distribution of the maximum 24 hours rainfall typologies according to the Alpert *et al.* (2002) classification for all the rainfall events and for rainfall events that trigger landslides. It can be observed that the most part of the total rainfall events (80%) belong to typologies from Light to Moderate-Heavy which represent a maximum precipitation in 24 hours between 0 and 32 mm, being the Light-Moderate typology (4 - 16 mm) the most usual (37%). Thus, it can be stated that the most common rainfall typologies for Gipuzkoa Province, according to this classification, was Light-Moderate. Nevertheless, if we consider only the rainfalls associated with landslides occurrence, then, the most frequent rainfall event typology was Heavy (45%), which represents a maximum precipitation in 24 hours between 64 and 128 mm.
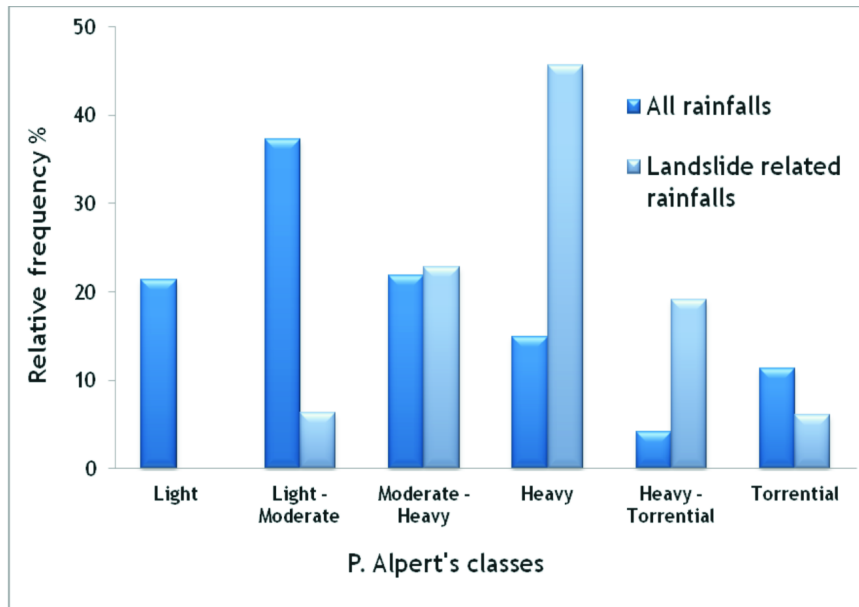
Figure 6.18: *Rainfall types relative distribution according to Alpert et al. (2002) classification.*
6.18 Irudia: Eurite moten banaketa erlatiboa Alperten klasifikazioaren arabera (Alpert *et al.*, 2002)

A similar comparison is displayed in figure 6.19, where two contingency graphics show the most frequent rainfall typologies for the total rainfall events and for the landslide associated rainfall events considering their duration (D) and their total cumulated rain (E).

A clear difference can be observed between both figures 6.19a and 6.19b, and each one presented very clustered results. In figure 6.19a the very most rainfall events (33.4%) presented a typology with durations which can range between 0 h and 24 h where the precipitated rain did not exceed 15 mm (combination a-A). But

in figure 6.19b the most frequent (22.8%) rainfall type durations ranged between 24 h and 72 h where the cumulated rain was between 60 mm and 120 mm (combination b-D). Notice that there were also high frequencies (16.8%) in events of 72 h to 144 h of duration with cumulated rains between 60 mm and 120 mm (combination c-D). So, regarding this results, it can be stated that, in general, in Gipuzkoa Province rainfall events spent less than 24 h of time and they cumulate less than 15 mm, but in case that rainfalls become longer in time (1 to 5 days) and start to cumulate precipitation over 60 mm, then landslides start to trigger.
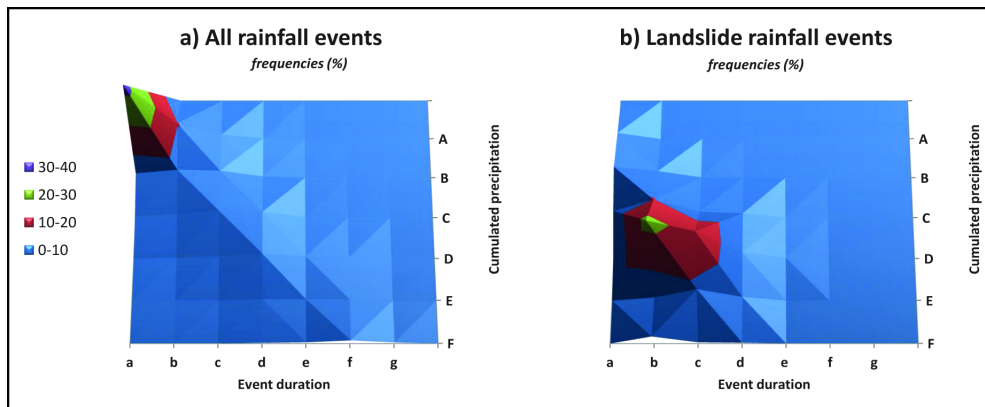
Figure 6.19: *Contingency graphics. Rainfall event relative distribution according to the event duration and event cumulated rain in: a) Rainfall events; b) Landslide trigger rainfall events, in Gipuzkoa Province (2006-2015). The meaning of the letters (a-g ; A-F) is summarized in Tab. 5.6*

6.19 Irudia: Kontingentzia taularen grafikoa. Euriteen banaketa erlatiboa euritearen iraupenaren eta akumulatutako euriaren arabera: a) Eurite guztietan; b) Lur labainkatek izandako euriteetan, Gipuzkoako LHan (2006-2015). Hizkien esan nahia 5.6 taulan ikus daiteke.

6.11.3   Landslides responsible rainfall threshold

We calculated objectively cumulated precipitation (E) and rainfall duration (D) based thresholds, for different exceedance probability levels, adopting the software described in Melillo *et al.* (2016). Each exceedance level represented the percentage of observations that stayed below a given threshold, so it can be interpreted as the error probability of such threshold. Below, we present the precipitation threshold equations proposed by the software considering the data obtained in our study area. These equations showed the minimum threshold of cumulated precipitation (E) that was expected to produce landslides for a given period (D). In this case, T_5 represents the curve above which landslide occurrence was expected with a 95% of probability, and T_1 represents so with a 99% of probability (Fig. 6.20).

$$T_5 : E = (4.4 \pm 0.7)D^{(0.47 \pm 0.04)} \tag{6.4}$$

$$T_1 : E = (3.0 \pm 0.5)D^{(0.47 \pm 0.04)} \tag{6.5}$$

Additionally, the uncertainty range of the thresholds (represented by the $\pm$ values), obtained by the bootstrap technique, was also provided (Peruccacci *et al.*, 2012). This shows the level of variability of the resulting curves respect to the introduced data (Figs. 6.20b and 6.20c).
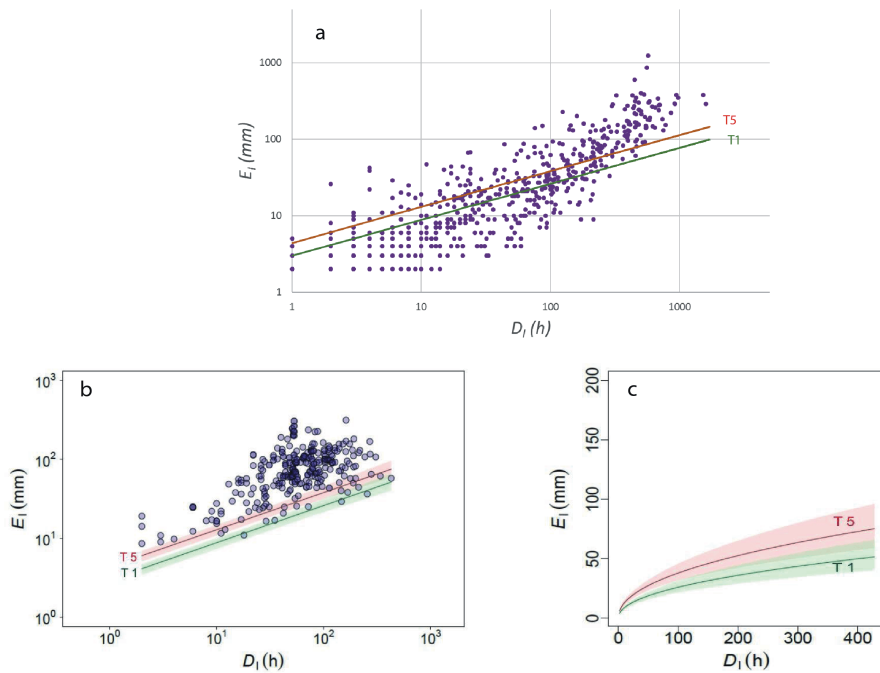
Figure 6.20: *Landslide responsible precipitation thresholds for Gipuzkoa Province: a) All rainfall events in logarithmic scale; b) Only rainfall events that trigger landslides in logarithmic scale (the difference between light and dark blue dots is result of the overlapping); c) T_1 and T_5 thresholds in decimal scale.*

6.20 Irudia: Lur labainketak sortzeko eurite atalaseak Gipuzkoako LHrako: a) Eurite guztiak eskala logaritmikoan; b) Lur labainketak izan diren euriteak bakarrik eskala logaritmikoan; c) T_1 eta T_5 atalaseak eskala dezimalean.

6.12   Discussion

The case study presented in this section shows the application of an approach that permitted the management of landslide occurrence data and precipitation data in order to obtain an overview of the behaviour of the slope instabilities in a given territory respect to their triggering because of rainfalls. By using the

approach proposed in Melillo *et al.* (2015), all the analytical part of the procedure was done automatically. In addition, landslides characterization and rainfall events characterizations tests were carried out, and during these analyses some important aspects were highlighted, which are detailed as follow.

First, it was highlighted the big number of landslides falling in urban *land cover* class (Fig. 6.17c). Considering that this class represents the spatial distribution of the most part of the human goods likely to be affected, it is logical to find that all the analysed landslide events except the 4% caused damages (Fig. 6.16). Furthermore, these results suggested that being a landslide inventory based on press reports, the landslide inventory was probably biased, because landslides that appear on the newspapers are mainly those that have caused some kind of negative effect to the humans, or their goods. So, the information provided by the landslide inventory cannot be considered as a representation of the general behaviour of landslides, but only of landslides which directly affects to humans or their activity.

Nevertheless, *slope* and *lithology* (Figs. 6.17a and 6.17b) are less related to the human activity than *land cover*. So, comparison of their relative distribution allowed to highlight if the presence of landslides in one or another class was given to the general presence of such class along the study area, or instead, the landslides are more likely to occur in one specific class than in others, concluding which are the most recurrent classes for the landslides with impact on humans. Thus, considering the limitation of the landslide inventory, the analysed data suggested that slopes with inclination between 15° and 60° with limestone or surface deposits are the most likely to be affected by rainfall induced landslides. Apart from that, clay and detrital rocks together with marls also presented considerable number of landslides, but it could be due to the large territory that they occupy in the study area. Indeed, this behaviour agrees with the results of Remondo *et al.* (2003) and Bonachea (2006), whose more detailed analysis in a smaller portion of our study area also reveals the importance of such materials respect to the landslide occurrence.

Referring to the rainfall data, the survey of the characterization tests (Figs. 6.18 and 6.19) showed that, for the studied 10 years, the most common rainfall event typology in Gipuzkoa Province and the typology responsible of landslide triggering were considerably different. This suggests that landslide occurrence is

closely related to a concrete type of rainfall. Furthermore, the classification of the rainfall events allowed the identification of precipitation types that were more likely to cause landslides in Gipuzkoa Province. According to the Alpert *et al.* (2002) classification, landslides in our study area were triggered mainly by rainfall events that contain a maximum cumulated precipitation in 24 hours of at least 64 mm and less that 128 mm (Fig. 6.18). Likewise, the alternative classification which considers the duration of the whole rainfall event and its cumulated precipitation presents similar results. According to this classification, the events of 1 to 4 days that can cumulate between 60 mm and 120 mm were the most likely of causing slope failures (Fig. 6.19).

The information provided by these classifications allows the responsible entities to obtain a general idea of the relation between rainfalls and landslide occurrence, although this generalization would be better supported by a longer data set to consider this rainfall pattern as a reliable trend. Moreover, although the extension of the studied area is relatively small, in section 4.3 is pointed out that the precipitation distribution presents slight spatial heterogeneities. So another improvement of this approach may include the division of the study area in more homogeneous portions in terms of precipitation, and perform the analysis separately.

Apart from the above mentioned general relationships, basing in 298 landslide observations and their special and temporal information, the automatic software proposed a set of landslides responsible precipitation thresholds for Gipuzkoa Province. Equations 6.4 and 6.5 indicated the threshold above which the 95% or 99% of the inventoried landslides were placed and the range of uncertainty associated with the input data (all the thresholds proposed by the algorithm for different exceedance levels are shown in Appendix E). So, it was considered reasonable to expect that, in the future and in this study area, the same kind of landslides will be mostly triggered if the threshold is overcome. Thereby, this information could be considered as an approximation toward a future data supported early warning system design, as suggested in Fell *et al.* (2008).

On the other hand, the analysis followed in this work should be interpreted considering some important topics. To begin with, rainfall events were defined automatically according to user set up values, such as the minimum interval without

precipitation (48 and 96 hours in this case). The rationale behind this option is the different rhythm of soils returning to under saturated state along the year (Melillo *et al.*, 2015, 2016), and though in this work these intervals were defined by expert criteria, we acknowledge that considering soil moisture data like in Valenzuela *et al.* (2018), uncertainties related to the antecedent soil conditions could be reduced, which can be important above all for medium and long duration rainfall events triggering landslides, as suggested by Zêzere *et al.* (2015). Furthermore, the empirical method applied by the software, presented several drawbacks. As it is more in depth discussed in Peruccacci *et al.* (2012), it has to be pointed out that the resulting threshold is highly dependent, not only from the quality and extension of the data, but also from its distribution, allowing the underestimation of large rainfall events with landslides. Thus, it has to be taken into account, that all the results are based on a given data set, and so we cannot do statistical inferences about the population outside the data set unless we know the relation between the data set (landslide inventoried) and the total population (the totality of occurred landslides). As we never know the totality of the occurred landslides, hence, we do not know how representative is our data set, and so, it is not possible in statistical terms to do a prediction. This threshold have to be considered as a suggestion of occurrence of landslides based on a given data set. And, for the same reason, if it would be used in a landslide warning system, the thresholds may result in false positives, as it is illustrated in the considerable amount of rainfall events plotted above the thresholds in figure 6.20a. Nevertheless, by means of the bootstrap approach, the statistical uncertainty associated with the data set was quantified.

To conclude, this study put in practice a very powerful tool which contributes to the definition of the rainfall needed (in quantity and quality) to trigger landslides, an information that, together with an appropriate susceptibility map, could be used to develop landslides occurrence forecasts. At the same time, some basic spatial data were added to the procedure in order to complement this outcome with the information about some physiographic features, such as *land cover*, *lithology* and *slope*, where landslides are most likely to occur. Additionally, the data about rainfall events produced by the software was profited to identify the most common rainfall typologies in Gipuzkoa Province for the studied period and based in two different

classifications, as well as the particular type of rainfalls responsible of landslides.

References

Alpert, P., Ben-Gai, T., Baharad, A., Benjamini, Y., Yekutieli, D., Colacino, M., Diodato, L., Ramis, C., Homar, V., Romero, R., *et al.*: The paradoxical increase of Mediterranean extreme daily rainfall in spite of decrease in total values, Geophysical Research Letters, 29, 1–4, 2002.

Bonachea, J.: Desarrollo, aplicación y validación de procedimientos y modelos para la evaluación de amenazas, vulnerabilidad y riesgo debidos a procesos geomorfológicos, Ph.D. thesis, Universidad de Cantabria, Santander, 2006.

Brunetti, M., Peruccacci, S., Rossi, M., Luciani, S., Valigi, D., & Guzzetti, F.: Rainfall thresholds for the possible occurrence of landslides in Italy, Natural Hazards and Earth System Science, 10, 447–458, 2010.

Corominas, J. & Moya, J.: Reconstructing recent landslide activity in relation to rainfall in the Llobregat River basin, Eastern Pyrenees, Spain, Geomorphology, 30, 79–93, 1999.

Dai, F. & Lee, C.: Frequency–volume relation and prediction of rainfall-induced landslides, Engineering Geology, 59, 253–266, 2001.

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, Engineering Geology, 102, 85–98, 2008.

Guzzetti, F., Peruccacci, S., Rossi, M., & Stark, C. P.: Rainfall thresholds for the initiation of landslides in central and southern Europe, Meteorology and Atmospheric Physics, 98, 239–267, 2007.

Li, C., Ma, T., Zhu, X., & Li, W.: The power–law relationship between landslide occurrence and rainfall level, Geomorphology, 130, 221–229, 2011.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., & Guzzetti, F.: An algorithm for the objective reconstruction of rainfall events responsible for landslides, Landslides, 12, 311–320, 2015.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., & Guzzetti, F.: Rainfall thresholds for the possible landslide occurrence in Sicily (Southern Italy) based on the automatic reconstruction of rainfall events, Landslides, 13, 165–172, 2016.

Peruccacci, S., Brunetti, M. T., Luciani, S., Vennari, C., & Guzzetti, F.: Lithological and seasonal control on rainfall thresholds for the possible initiation of landslides in central Italy, Geomorphology, 139, 79–90, 2012.

Piciullo, L., Gariano, S. L., Melillo, M., Brunetti, M. T., Peruccacci, S., Guzzetti, F., & Calvello, M.: Definition and performance of a threshold-based regional early warning model for rainfall-induced landslides, Landslides, pp. 1–14, 2016.

Ramos-Cañón, A. M., Trujillo-Vela, M. G., & Prada-Sarmiento, L. F.: Niveles umbrales de lluvia que generan deslizamientos: una revisión crítica, Ciencia e Ingeniería Neogranadina, 25, 61–80, 2015.

Remondo, J., González-Díez, A., De Terán, J. R. D., & Cendrero, A.: Landslide susceptibility models utilising spatial data analysis techniques. A case study from the lower Deba Valley, Guipúzcoa (Spain), Natural Hazards, 30, 267–279, 2003.

Valenzuela, P., Domínguez-Cuesta, M. J., García, M. A. M., & Jiménez-Sánchez, M.: Rainfall thresholds for the triggering of landslides considering previous soil moisture conditions (Asturias, NW Spain), Landslides, 15, 273–282, 2018.

Wieczorek, G. F.: Effect of rainfall intensity and duration on debris flows in central Santa Cruz Mountains, California, Reviews in Engineering Geology, 7, 93–104, 1987.

Zêzere, J., Vaz, T., Pereira, S., Oliveira, S., Marques, R., & Garcia, R.: Rainfall thresholds for landslide activity in Portugal: a state of the art, Environmental Earth Sciences, 73, 2917–2936, 2015.

Chapter 7

General discussion

The current thesis presents a depth reflection about divers methodological approaches applied for landslide analysis at different levels in a regional scale. The principal objective was the definition of an updated methodological approach in which each decision in each step of the process would be scientifically supported. The research project was based on three sequential approaches applied in a test study area, where different options about some steps, that are considered of relevant importance, were experimented: i) the landslide inventory assessment and the explanatory variables processing strategies issue addressed in the tests carried out in the Oria river catchment; ii) the proper usage of the field-based landslide inventory as well as the mapping unit partition issue tackled in the applications carried out in Gipuzkoa Province; and iii) the definition of a preliminary precipitation threshold responsible of landslide triggering as a sample of the direction that landslide analysis in a regional scale should go, in order to advance from susceptibility assessment toward hazard knowing.

This chapter brings together the knowledge obtained by these three approaches showed in previous sections and discuss the most important findings in order to offer impartial answers to the objectives of this work, considering them as questions to be solved.

Nonetheless, though the usage of statistical and data mining methodologies are widely recognized as the most suitable approaches for landslide susceptibility and hazard assessments, mainly because of their capacity of ensuring the objectivity and reproducibility of the models, the huge amount of existing mathematical variants leaves the discussion about the definitive mathematical model yet far to be solved. For this reason, in this thesis other more basic questions were addressed, and even though only the Logistic Regression model was tested, the main conclusions obtained through this work should be considered equally applicable for any other statistical or data mining method, since the management of the introduced data is similar in the most part of them.

To begin with, landslide data issue was tackled. It is the very basic information source, necessary for every landslide susceptibility and hazard modelling independently of the mathematical approach used (Corominas & Mavrouli, 2011). That is why the first objective of the thesis was **to test different landslide**

inventories for landslide analysis in a regional scale in order to detect the most suitable features necessary to run susceptibility models. We start from the idea that the knowledge of every landslides occurred along the history of a given landscape is fairly impossible, so the completeness and representativeness of each landslide inventory is usually assumed, which gives an unknown uncertainty from the very beginning (Guzzetti *et al.*, 2012). Once this is assumed, the first condition that a landslide inventory must fulfil is the spatial accuracy, because logically if the spatial information extracted from a given landslide location is incorrect, the complete model will show erroneous results. In this regards, the assessment tests carried out in the Oria river catchment with the available landslide data coming from the bibliographical sources showed that, for the particular case of Gipuzkoa, such source of information was not suitable. On the other hand, spatial accuracy can be ensured by carrying out a field-work based inventory, as it was done in both tests showed in sections 6-I and 6-II, but in such case some other important aspects have to be considered. First, if the data acquisition was done by single points, like in section 6-I, the uncertainty related to the landslide size have to be offset by the application of a buffer to the inventoried points, whereas if the areas of the complete landslide were collected, then this uncertainty could be eliminated, like in section 6-II. Second, the fact of being a landslide information obtained by direct field survey affects to the approach that should be followed during the statistical analysis, where the usage of an Effective Surveyed Area (ESA) for calibrating the model was proved to be a suitable option in order to reduce uncertainties about the location of landslide-free places. Nonetheless, there are still other uncertainties to be solved in further studies. It is acknowledged that the discrimination between the source area and the run-of area for each inventoried landslide would provide even more accurate information, but these information was not collected during the field work.

Apart from the spatial accuracy, the temporal information can also play a key roll (Soeters & Van Westen, 1996; Zêzere, 2002). Above all in landslide analysis related to the hazard modelling, this is an essential data that usually presents difficulties to obtain (Bonachea *et al.*, 2017). Field-work based landslide inventory can not offer such information, thus, in section 6-III press reports were used to collect landslide

occurrence dates information, which in some cases was very accurate. It has to be acknowledge the potential of press reports as data sources, but apart from being very time consuming, frequently additional information such as the type of landslide or its magnitude is lacking and the spatial accuracy is not always ensured.

As a consequence, as shown in the applications carried out during the thesis, the collection of a landslide inventory with exact spatial and temporal accuracy is a very difficult task, even more if they are landslides occurred in the past. So the balance between location accuracy and occurrence dates information should be assessed depending on the purposes of each research. For susceptibility modelling spatial accuracy should be prioritized, even though the temporal information could allow *spatio-temporal intra-domain* approaches to validate the models, instead of the *spatial-intra-domain* approach followed in this thesis. In hazard related studies, such as the definition of the precipitation threshold responsible for landslides, temporal data becomes more import than the landslide occurrence location.

One possible solution to fulfil both requirements could be the development of a multi temporal landslide inventory by means of aerial photo interpretation of flights performed in different years (Guzzetti *et al.*, 2012; Santangelo *et al.*, 2015). This would offer relative temporal information to the very accurate landslide locations allowing *spatio-temporal intra-domain* validations, apart from the fact that uncertainty about landslide-free places could be solved avoiding the usage of the ESA. As a drawback, this method is even more time consuming than the press reports survey considering the extension of our study area, and the temporal accuracy does not reach enough resolution for precipitation thresholds calculations (Remondo, 2001). Such level of detail could only be reached with a systematic collection of data about landslide occurrences in real time, where government administration, civil protection and even the citizens themselves should take part (Trigila *et al.*, 2010; i Planells, 2007).

Automatic and semi-automatic techniques for landslide detections by means of high and very-high resolution satellite imagery worth taking under consideration, besides their applicability has still not been enough tested (Alvioli *et al.*, 2018). In addition, such imagery are relatively recent and in some cases involves the need of high budget, which difficult their usage for long terms multi-temporal inventories in

a regional scale, being more suitable for event related inventories in a local scale.

The second objective of this thesis was related to the other essential ingredient necessary to build landslide statistical models, that is the management of the information about the spatial conditions that brings on the destabilisations of a slope. Thus, this thesis aimed **to experiment with the available spatial digital layers, for their usage as independent explanatory variables in landslide susceptibility analysis, as well as to test different ways for selecting, in an objective way, only the most convenient to build the model**. Thereby, 20 original environmental variables were tested in different applications, 11 of them were continuous layers of raster type and 9 categorical layers displayed in vector form. So, the first doubt that arises in this regards is usually if all these variables can be used together in statistical models independently of their continuous or categorical condition (Amorim, 2012; Felicísimo *et al.*, 2013; Pourghasemi & Rahmati, 2018). This depends on the mathematical model used for the project. In particular, the Logistic Regression applied in this work is able to admit both types of explanatory variables, however, the tests carried out in section 6-I showed that the transformation of categorical variables into continuous giving a relative value based on the presence of landslides to each of its class, allowed mitigating the effect of the smaller categories offering a more robust susceptibility model without loosing the model performance. Thus, this strategy could be considered as a standard approach which permits the adaptation of the explanatory variables to any type of mathematical modelling. Moreover, two options were tested to give a relative value based on the presence of landslides to each class, landslide density value in section 6-I and frequency ratio in 6-II, both of them with satisfactory results.

Nevertheless, in the case of susceptibility models, the introduction of all the variables together was not considered a correct approach, because if no exploration is done about the individual relation of each variable respect to the presence or absence of landslides or even the pairwise relationship of the variables is not tested, irrelevant explanatory variables could be introduced to the models or on the other hand, redundant information could be taken under consideration. In both cases the results would be aggravated. As an objective way to solve this issue, two statistically driven variables selection approaches were applied depending on the statistical software

used for the modelling process (see sections 6-I and 6-I). Such procedures permitted, in both cases, the objective selection of a reduced set of explanatory variables among a wider initial list of available spatial information, which suggests this practice as highly recommendable for landslide susceptibility modelling. However, results of the applications carried out in the Oria river catchment highlighted that statistical techniques, by their own, are not entirely capable of detecting inappropriate variables, becoming imperative the surveillance of a geomorphologist that conceptually justify the usage of such variables.

As a matter of fact, it was pointed out that even though the variables overcome the statistical tests, the application of certain approaches condition the availability of some explanatory variables. Such is the case of *distance to the transport network*, whose usage in a susceptibility model carried out by field-work based landslide inventories resulted in highly biased susceptibility maps. Likewise, *aspect* and *topographic wetness index* (*TWI*) were considered unsuitable explanatory variables if the model was built using an slope unit mapping partition. Anyway, this does not mean that these variables has nothing to do with the presence or absence of landslides, but only that they can not be used if these specific approaches are applied.

Moving on to the third objective, it concerns, above all, the cartographic aspect of the models mapping, though it also involves important considerations to take into account from the conceptual point of view. In this regards, it was stated as objective **the observation and recognition of the advantages and drawbacks of different mapping units in landslide susceptibility mapping**. Thus, in section 6-II two of the most used terrain partitions were compared, regular grid cells (also called pixels) and slope units.

In the case of regular grid cells, the main advantage of its usage is related to the operational facilities that this partition offers for spatial digital information processing. As raster files are already organized following the same regular structure and vectorial layer can easily be transformed into raster, the exact coincidence between pixels in different layers can be ensured if the same resolution and alignment is provided to each spatial variable. This allows the easy application of statistically driven variables analysis before defining the final set of variables to be introduced in the model. Moreover, if the performance tests are observed, validation results showed

"better" values in pixel based models than in slope unit based modes (such as higher AUC and k values), however, it doesn't necessarily mean that regular grid partition models perform better than slope unit models, since this apparent improvement could only be due to the differences in the amount of data processed by the model. Notice that in section 6-II pixel models were calibrated with 169,246 data whereas slope units only with 456. Indeed, this could be considered as a drawback, since the regular grid partition of the terrain is actually an arbitrary subdivision of the relieve that has nothing to do with its topographical nor morphological features, and this could result in the overestimation of the model performance comparing to the real behaviour of the slopes (Carrara *et al.*, 1995; Alvioli *et al.*, 2016; Reichenbach *et al.*, 2018).

In addition, using pixel partition it has to paid attention to the calibration/validation sampling, because as a single landslide area will be partitioned in several pixels, the random selection of landslide-presence data could result in that some pixels that belonging to the same landslide are used for calibration and others for validation, which is not acceptable from the conceptual point of view (Brenning, 2005). For this reason, the calibration/validation partition was carried out previously to the raster transformation of the landslide inventory in section 6-II. Furthermore, in the specific case of models run by field-based landslide inventories, regular grid cells showed higher sensibility to the incertitude introduced by the non verification of landslide-free data, becoming necessary the usage of an effective surveyed area delimitation.

Such drawbacks can be solved by using slope units. As this irregular terrain partition bear a clear relation with topography, it can be considered more suitable for modelling the future behaviour of slopes from a conceptual point of view. Incertitudes introduced by the usage of a field-based landslide inventory can be mitigated and the sampling of complete landslides boundaries in calibration or in validation sets is ensured. Nevertheless, there are still some issues that demands further investigations in order to use the slope unit partition with all the guaranties in landslide susceptibility models. The definition of an appropriate threshold of landslide density to classify slope units as landslide-presence or landslide-free worth being explored, in favour of standardisation of this approach and in order to

ease comparisons between different results. Also, objective selection of explanatory variables for landslide susceptibility models carried out by slope units is still an unsolved question, that for the moment demands the usage of pixel partition for a preliminary analysis.

The fourth objective of this thesis refers to an specific test carried out during the project. Namely, **to prove that during the calibration, the restriction of the area in which no-landslide data are sampled to the ESA, in place to the WA, enhances the quality of the model, in the cases where the landslide inventory was carried out by direct field work**. As it is more in depth discussed in section 6.10, it was proved that the usage of an Effective Surveyed Area enhances the performances of the model, if a field-based landslide inventory is used for its calibration. Thereby, implications that this new approach involves at different levels of the landslide susceptibility modelling has already been discussed in the previous paragraphs.

The fifth objective set on this work is related to the application of an automatized methodology that defines the minimum precipitation needed to generate landslides. As rainfalls are the principal triggering factor of slope instabilities in our study area, it was attempted **to detect relations between the inventoried instabilities and the rainfall events for the calculation of landslides responsible precipitation thresholds in Gipuzkoa Province**. This preliminary experience allowed to assess the potentiality of such methodology and to train the necessary practices for carrying it out correctly in the future, in order to develop hazard analysis or even, combined with susceptibility maps, the design of an integrated early warning system.

The approach followed in section 6-III showed promising results, above all because it was verified the accessibility of the necessary precipitation data and the suitability of the press-based landslide inventories, though the quality of the later could be considerably improved with the systematic data collection network that reports the landslide events. By means of such data, the used algorithm proved to be able, in a mathematically consistent and reproducible way, of defining relations between precipitation and landslide occurrence, which could imply several applications apart from calculating precipitation thresholds, such

as landslide occurrence frequency analysis, rainfall typology characterisations or landslide occurrence temporal predictions based on meteorological forecasts (Zêzere *et al.*, 2015; Bogaard & Greco, 2018).

Nevertheless, it has to be pointed out that the results displayed in section 6-III are only experimental results, so they can not be considered yet entirely applicable. On the one hand, because the precipitation and landslide occurrence data set only covers a period of 10 years, thus the real climatic behaviour of the precipitations was not properly represented, allowing the unbalanced influence of climatically anomalous years. On the other hand, the real performance of the thresholds was not validated, for which data of the landslide occurrences and precipitations of successive years would be needed in order to compare the cumulated precipitation and the duration of each new rainfall event that triggered a landslide with the defined threshold in a given exceedance probability level. Furthermore,the division of the study area in homogeneous precipitation portions could allow more accurate thresholds; the consideration of the antecedent rainfalls or the soil moisture state would improve the knowledge of the slopes response against rainfalls; or the more appropriate selection of the rain gauge (not necessarily the closest one) would also enhance the approach.

Indeed, during the development of the current thesis, a newly version of the algorithm tested in section 6-III has been published (Melillo *et al.*, 2018), where several improvements such as the consideration of the soil water saturation or the automatic selection of the representative rain gauge has been included. Hence, it is worth doing efforts collecting proper data in order to profit the potentials of such promising tool.

Besides the main objectives discussed until now, this work also followed a secondary goal related to the implementation of new and updated technologies. For the analysis carried out, two principal types of software were used: statistical analysis software (such as SPSS or R project) and geographical information systems (such as ArcGIS, QGIS or GRASS). These can be classified as commercial software (like SPSS and ArcGIS) or free software (like R project, QGIS and GRASS). According to the experience acquired during the current thesis, it was concluded that actually the most part of approaches followed could be carried out with any software, be commercial or free. It is only a matter of time dedication until understanding the

specific functionalities of each one. However, the main drawback of commercial programs, apart from the necessary budget for their utilization, is their hardly customizable condition, which difficult the development of new approaches or work flows if they were not previously available among the options of each software. In addition to that, if the work flow is organised in different phases, in most of the cases it is not possible to automatize the complete process, obligating the user to carry out the analysis step by step, which increases the risk of error as well as the processing time.

So, in favour of objective and reproducible methodologies it is worth the usage of command lines like free software, such as R project or GRASS, which allows the development of completely customizable codes like LAND-SE, *r.survey* or *r.slopunit*, to advance toward the definition of standards in landslide analysis.

To finish, as was already stated, the focus of this work was directed to the methodological approaches more than to the production of definitive landslide susceptibility maps or precipitation thresholds, which implies that there is still considerable work to do, above all in what concerns the data collection. Results suggest that the approaches tested in the current thesis could give really useful and applicable results if more accurate and extended landslide inventories were available for the analysed study area. In this regards, in the next chapter some recommendations are suggested in order to obtain more performing results.

References

Alvioli, M., Marchesini, I., Reichenbach, P., Rossi, M., Ardizzone, F., Fiorucci, F., & Guzzetti, F.: Automatic delineation of geomorphological slope units with *r.slopeunits v1.0* and their optimization for landslide susceptibility modeling, Geoscientific Model Development, 9, 3975–3991, 2016.

Alvioli, M., Mondini, A. C., Fiorucci, F., Cardinali, M., & Marchesini, I.: Automatic landslide mapping from satellite imagery with a topography-driven thresholding algorithm, no. e27067v1. PeerJ Preprints, 2018.

Amorim, S. F.: Estudio comparativo de métodos para la evaluación de la susceptibilidad del terreno a la formación de deslizamientos superficiales: Aplicación al Pirineo Oriental, Ph.D. thesis, Universidad Politécnica de Catalunya, Barcelona, 2012.

Bogaard, T. & Greco, R.: Invited perspectives: Hydrological perspectives on precipitation intensity-duration thresholds for landslide initiation: proposing hydro-meteorological thresholds, Natural Hazards and Earth System Sciences, 18, 31–39, 2018.

Bonachea, J., Remondo, J., Rivas, V., González Díez, A., & Sánchez-Espeso, J.: Modelización de la peligrosidad de deslizamientos para diferentes escenarios empíricos en el Cantábrico oriental, in: IX Simposio sobre Taludes y Laderas Inestables, edited by Alonso, E., Corominas, J., & Hürlimann, M., pp. 407–418, International Centre for Numerical Methods in Engineering (CIMNE), Santander, 2017.

Brenning, A.: Spatial prediction models for landslide hazards: review, comparison and evaluation, Natural Hazards and Earth System Science, 5, 853–862, 2005.

Carrara, A., Cardinali, M., Guzzetti, F., & Reichenbach, P.: GIS technology in mapping landslide hazard, in: Geographical Information Systems in Assessing Natural Hazards, edited by Carrara, A. & Guzzetti, F., pp. 135–175, Springer, 1995.

Corominas, J. & Mavrouli, O. C.: Living with landslide risk in Europe: Assessment, effects of global change, and risk management strategies, Tech. rep., SafeLand. 7th Framework Programme Cooperation Theme 6 Environment (including climate change) Sub-Activity 6.1.3 Natural Hazards, 2011.

Felicísimo, Á. M., Cuartero, A., Remondo, J., & Quirós, E.: Mapping landslide susceptibility with logistic regression, multiple adaptive regression splines, classification and regression trees, and maximum entropy methods: a comparative study, Landslides, 10, 175–189, 2013.

Guzzetti, F., Mondini, A. C., Cardinali, M., Fiorucci, F., Santangelo, M., & Chang, K. T.: Landslide inventory maps: New tools for an old problem, Earth-Science Reviews, 112, 42–66, 2012.

i Planells, L. M.: Determinació de llindars de pluja desencadenants d'esllavissades a Catalunya, Ph.D. thesis, Universitat Politècnica de Catalunya, Barcelona, 2007.

Melillo, M., Brunetti, M. T., Peruccacci, S., Gariano, S. L., Roccati, A., & Guzzetti, F.: A tool for the automatic calculation of rainfall thresholds for landslide occurrence, Environmental Modelling & Software, 105, 230–243, 2018.

Pourghasemi, H. R. & Rahmati, O.: Prediction of the landslide susceptibility: Which algorithm, which precision?, Catena, 162, 177–192, 2018.

Reichenbach, P., Rossi, M., Malamud, B., Mihir, M., & Guzzetti, F.: A review of statistically-based landslide susceptibility models, Earth-Science Reviews, 180, 60–91, 2018.

Remondo, J.: Elaboración y validación de mapas de susceptibilidad de deslizamientos mediante técnicas de análisis espacial, Ph.D. thesis, Universidad de Oviedo, Oviedo, 2001.

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., Natural Hazards and Earth System Sciences, 15, 2111–2126, 2015.

Soeters, R. & Van Westen, C.: Slope stability recognition analysis and zonation, in: Landslides: Investigation and Mitigation, edited by Turner, A. K. & Jayaprakash, G., 247, pp. 129–177, Transportation Research Board Special Report, Washington D.C., 1996.

Trigila, A., Iadanza, C., & Spizzichino, D.: Quality assessment of the Italian Landslide Inventory using GIS processing, Landslides, 7, 455–470, 2010.

Zêzere, J.: Landslide susceptibility assessment considering landslide typology. A case study in the area north of Lisbon (Portugal), Natural Hazards and Earth System Science, 2, 73–82, 2002.

Zêzere, J., Vaz, T., Pereira, S., Oliveira, S., Marques, R., & Garcia, R.: Rainfall thresholds for landslide activity in Portugal: a state of the art, Environmental Earth Sciences, 73, 2917–2936, 2015.

Chapter 8

Recommendations

Basing on the results showed up and discussed during the current thesis some recommendations are suggested in order to ensure good practices in some crucial steps when carrying out landslide susceptibility models and precipitation thresholds. In particular the following list of recommendation is adapted to the specific case of Gipuzkoa Province, although they are equally recommended to any other study area:

- The very first condition to obtain reliable landslide susceptibility models as well as precipitation thresholds is the availability of landslide events data with enough spatial ant temporal accuracy. After assessing the available data bases that cover the entire territory of Gipuzkoa, it has been concluded that at the moment, only field work based landslide inventory offers enough spatial accuracy, and press-based data do so offering temporal accuracy. Although these data sources allowed the relatively rapid collection of information, in a regional scale there is high probability of underestimating the real magnitude of the events. Thus, for the sake of the development of as much as accurate possible landslide inventory, it is recommended **the preparation of a multi-temporal inventory of the past landslides by means of aerial photo interpretation and field survey, while a systematic data collection network is organized to register the information about new landslides**. Even if this work could spent several years or even decades, it is of major priority if good performing results are aimed, independently of the mathematical approach used to construct the model.

- It is equally important to update and upgrade the spatial digital layers freely available, in order to always have access to as widest as possible set of explanatory variables. In this regards, **the transformation of the categorical variables into continuous** is recommended by giving a relative value based on the presence of landslides to each class, like for example landslide density value or frequency ratio.

- Concerning the previous two points, the roll of the administrative councils is of major importance. Above all, the systematic collection of new landslide events data by means of standardised protocols should take part of the activities developed after the knowledge of events like landslides. This way

215

such information can be available to the scientific community in order to carry out reliable susceptibility, hazard and risk models.

- The selection of the explanatory variables should be carried out in as much as objective way possible, for which two statistically driven approaches are proposed in this work. However, **this statistically driven decision rules should always be supervised by a geomorphologist in order to avoid eventual conflicts between any explanatory variable and the followed approach to construct the model**.

- Although this explanatory variables selection can only be applied using grid cell terrain unit partition, **it is recommended to carry out susceptibility maps by means of slope unit terrain partition**. If possible, it is worth using landslide inventory from multi temporal aerial photo interpretation in order to allow spatio temporal intra domain validation and to avoid uncertainties related to landslide-free location assumption. If field-based landslide inventory is used, the application of the Effective Surveyed Area during the calibration of the model is proposed.

- As data about the moment of landslide occurrence increase, it is worth **calculating new precipitation thresholds using the algorithm used in this work** and carry out temporal validations partitioning the data set in two subsequent period of time.

To finish, as an approximation to the recommendations proposed in this chapter, the current thesis leaves to the public administrations the following relevant outcomes: i) a new landslide data base collected by means of the field work; ii) a landslide susceptibility map carried out for the Oria river catchment as a test study area; iii) two landslide susceptibility maps performed for the complete area of Gipuzkoa Province (one using grid cells and the other using slope units as cartographic mapping units), which imply a difference in terms of the used data as well as the methods followed comparing to the previously existing map; and iv) the first data driven landslide responsible precipitation thresholds for the Gipuzkoa Province.

Appendices

APPENDIX A: Supplementary material about landslide inventories

All the landslide inventory data produced in the current thesis are available in the supplementary material folder, in the CD attached to the current thesis. The following files can be found there:

- *Field_work_landslide_inventory_Gipuzkoa_2018.shp*:

  Contains all the georeferenced polygons concerning each landslide localized by field work, as well as the corresponding attributes table.

- *Field_work_sheets.pdf*:

  Contains every single field sheet corresponding to each localized landslide by field work.

- *Landslide_occurrence_time_data.xlsx*:

  Contains a data table in which the landslide information obtained by press-report has been compiled.

Any utilisation of these data must be followed by the following reference:

Bornaetxea, T.: Methodological approach for landslide analysis in a regional scale. Data collection, susceptibility models and precipitation thresholds. Application in Gipuzkoa province (Basque Country), PhD. thesis, Euskal Herriko Unibertsitatea (EHU/UPV), 2018.

APPENDIX B: Software codes

The following web sites offer the direct access to the original and open source codes used during different phases of the studies developed in the current thesis.

LAND-SE

https://www.geosci-model-dev.net/9/3533/2016/gmd-9-3533-2016.html

The additional module LAND-SVA can be found in:
https://github.com/maurorossi/LAND-SVA/

r.slopeunits

https://www.geosci-model-dev.net/9/3975/2016/gmd-9-3975-2016.html

r.survey

https://www.nat-hazards-earth-syst-sci.net/18/2455/2018/

Algorithm for precipitation thresholds

http://geomorphology.irpi.cnr.it/tools/rainfall-events-and-landslides-thresholds/ctrl-t-algorithm

APPENDIX C: Urban area class reclassification

Table A1: *Land cover 1 classes reclassified as Urban area in section 6-I*

A1 Taula: Hiritar bezala birklasifikatuak izen diren lur estaldura 1eko klase originalak 6-I atalean.

Industrial	Other artificial surfaces	Energy infrastructures
Urban equipments	Urban continuous	Urban discontinuous
Waste infrastructures	Water furnishing infrastructures	Telecommunications
Terciary sectory land use		

APPENDIX D: Frequency Ratio values

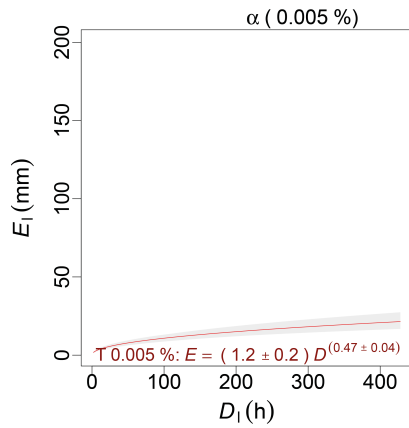**Table A2:** *Frequency Ratio values for each class of categorical variables used in section 6-II*

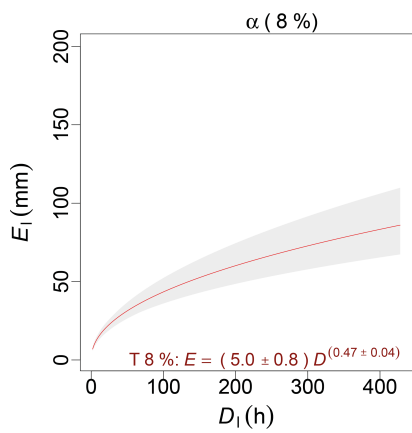**A2 Taula:** 6-II atalean erabiltzen diren FR balioen taula aldagai kategorikoetarako.

Variable	FR	Variable	FR
Lithology		*Permeability*	
No rock	0.013	Water	0.018
Surface deposits	0.247	Low	1.223
Clay and Detrital rock	1.785	Medium	0.621
Marls	0.764	High	0.777
Limestones	0.468	Impermeable	0.963
Magmatic rocks	0.678		
Slate	0.876		
Land cover 1		*Regolith thickness*	
Water	0.007	Water	0.233
Antropic	0.957	0 - 0.5 meters	0.463
Beach and turberas	0	0.5 - 1 meters	0.616
Forest	0.756	1 - 2 meters	1.514
Crops	1.725	2 - 4 meters	0.902
grassland	1.833	More than 4 meters	2.901
Scrub and hedges	0.798		
Rock	0.809		
Land cover 3		*Aspect*	
No vegetation	0.616	Flat	0.123
Grassland	1.719	North	1.152
Agricultural cultivation	3.122	North - East	1.041
Forest	0.804	East	0.84
Scrubs	0.278	South - East	0.536
Urban	0.585	South	0.918
Urban Park	0.086	South - West	1.026
		West	1.015
		North - West	1.338

APPENDIX E: Landslides responsible precipitation thresholds

The following graphs show the precipitation threshold calculated for Gipuzkoa Province and for different α exceedance levels following the procedure explained in section 6-III.

α (35 %)

E_l (mm)

D_l (h)

T 35 %: $E = (\,8.8 \pm 1.3\,)\,D^{(0.47 \pm 0.04)}$

α (50 %)

E_l (mm)

D_l (h)

T 50 %: $E = (\,10.9 \pm 1.7\,)\,D^{(0.47 \pm 0.04)}$
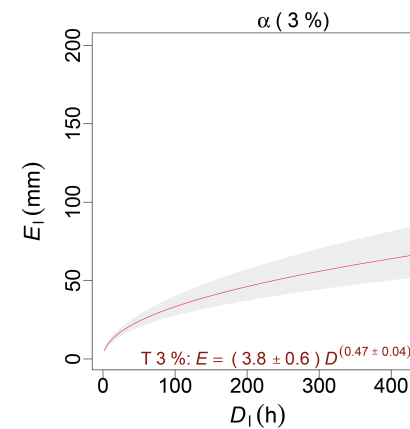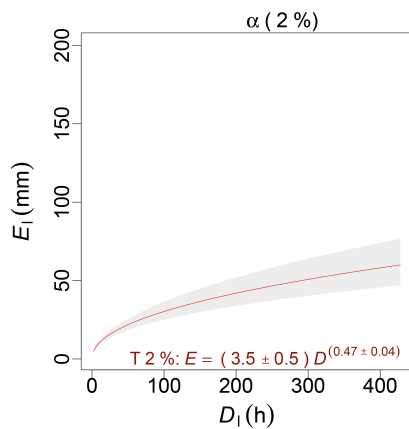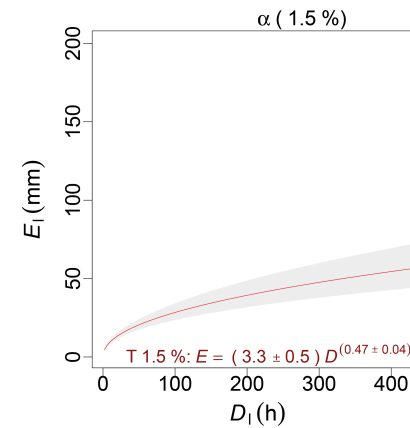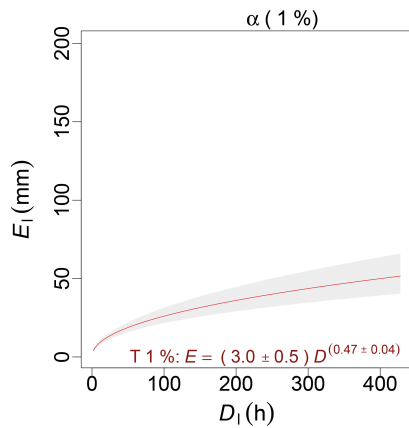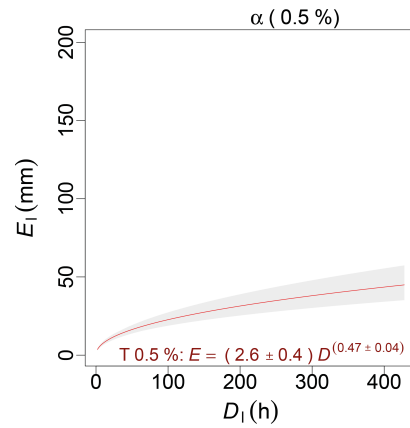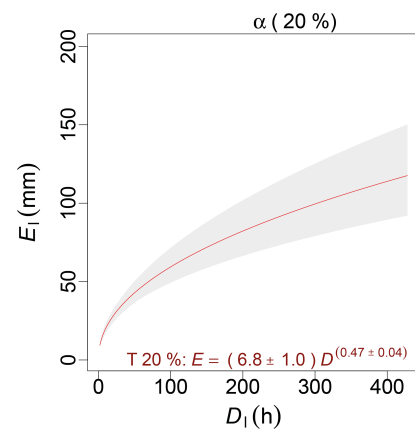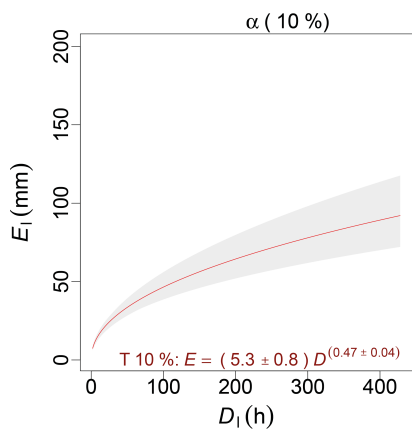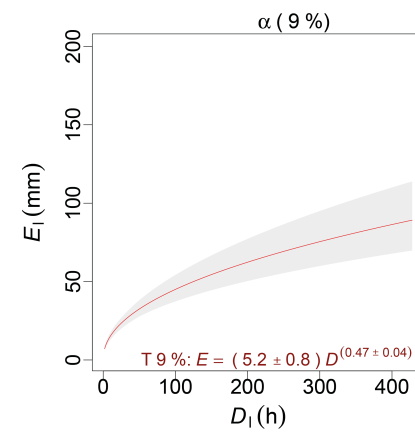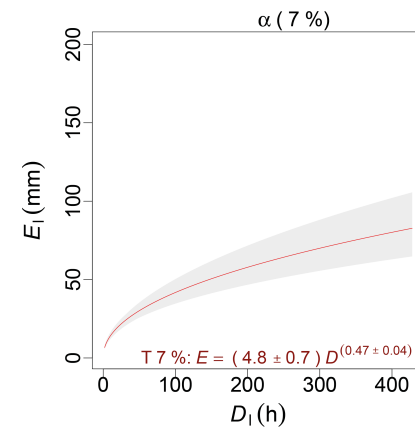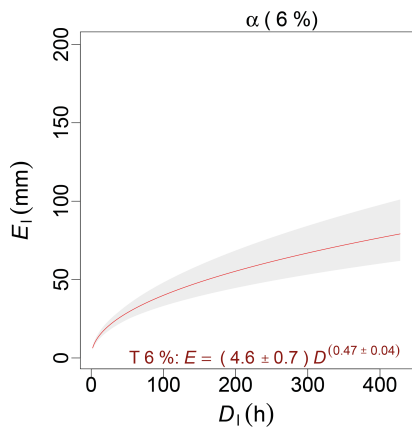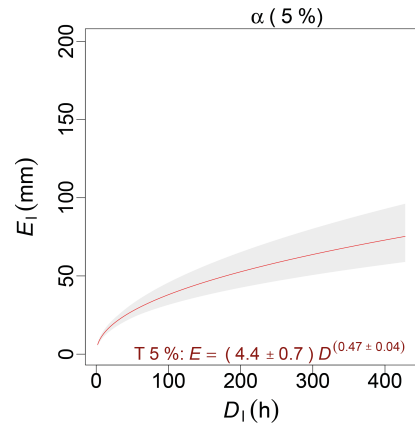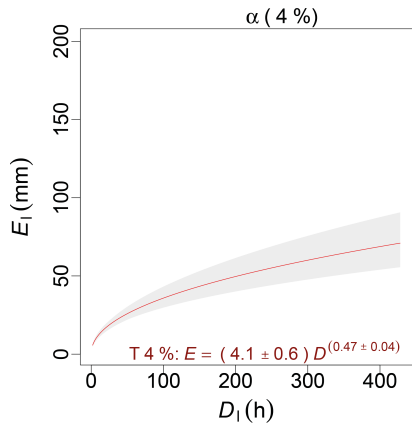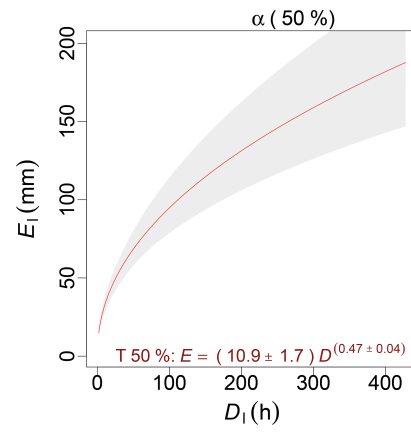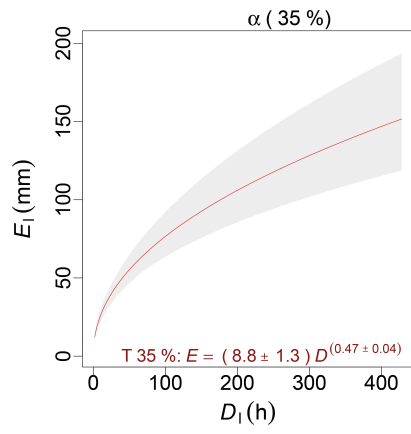
Glossary

Euskera - Ingelera glosarioa.

Euskera	Ingelera	Euskera	Ingelera
Ahultasuna	Vulnerability	Iraulketa	Topple
Alboko hedadura	Lateral spreading	Jausia	Fall
Arrisku	Risk	Kokalekua	Location
Arroka	Rock	Konplexu	Complex
Atalase	Threshold	Labainketa	Slide
Azaleko	Shallow	Lohi	Mud
Detritu	Debris	Lur	Earth
Erorketa	Fall	Lur labainketa	Landslide / Slope failure
Errotazional	Rotational	Lurzorua	Soil
Espesaketa	Exposure	Lurzoruaren erabilera	Land use
Eurite	Rainfall	Lurzoruaren estaldura	Land cover
Ezaugarri geologikoak	Geological features	Mada unitate	Slope unit
Ezaugarri hidrografikoak	Hydrographic features	Malda	Slope
Ezaugarri hidrologikoak	Hydrologic features	Masa mugimendu	Mass movement
Ezaugarri klimatikoak	Climatic features	Mehatxua	Hazard
Fluxu	Flow	Modelo	Model
Gertaera faktorea	Triggering factor	Sakona	Deep
Gipuzkoako Foru Aldundia	Provincial Council of Gipuzkoa	Sarrera	Introduction
Gipuzkoako Lurralde Historikoa	Gipuzkoa Province	Suszeptibilitate	Susceptibility
Ikerketa eremua	Study area	Translazional	Translational
Ikuskatutako eremu efektiboa	Effective surveyed area		

Glossary English - Basque.

Ingelera - Euskera glosarioa.

English	Basque	English	Basque
Climatic features	Ezaugarri klimatikoak	Mass movement	Masa mugimendu
Complex	Konplexu	Model	Modelo
Debris	Detritu	Mud	Lohi
Deep	Sakona	Provincial Council of Gipuzkoa	Gipuzkoako Foru Aldundia
Earth	Lur	Rainfall	Eurite
Effective surveyed area	Ikuskatutako eremu efektiboa	Risk	Arrisku
Exposure	Espesaketa	Rock	Arroka
Fall	Erorketa/Jausia	Rotational	Errotazional
Flow	Fluxu	Shallow	Azaleko
Geological features	Ezaugarri geologikoak	Slide	Labainketa
Gipuzkoa Province	Gipuzkoako Lurralde Historikoa	Slope	Malda
Hazard	Mehatxua	Slope unit	Mada unitate
Hydrographic features	Ezaugarri hidrografikoak	Soil	Lurzorua
Hydrologic features	Ezaugarri hidrologikoak	Study area	Ikerketa eremua
Introduction	Sarrera	Susceptibility	Suszeptibilitate
Land cover	Lurzoruaren estaldura	Threshold	Atalase
Landslide / Slope failure	Lur labainketa	Topple	Iraulketa
Land use	Lurzoruaren erabilera	Translational	Translazional
Lateral spreading	Alboko hedadura	Triggering factor	Gertaera faktorea
Location	Kokalekua	Vulnerability	Ahultasuna