Land Use Change and Landslide Susceptibility Assessment Using GIS and Multivariate Quantitative Predictive Models for Mountainous Disaster Mitigation in South Sulawesi Indonesia

アンダン スルヤナ ソマ

Land Use Change and Landslide Susceptibility Assessment Using GIS and Multivariate Quantitative Predictive Models for Mountainous Disaster Mitigation in South Sulawesi Indonesia

ANDANG SURYANA SOMA

2018

Land Use Change and Landslide Susceptibility Assessment Using GIS and Multivariate Quantitative Predictive Models for Mountainous Disaster Mitigation in South Sulawesi Indonesia

A dissertation submitted

by

Andang Suryana Soma

In partial fulfillment of
the requirements for the degree of
Doctor of Agriculture at Department of
Forest and Forest Product Sciences, Graduate School of Bioresources
and Bioenvironmental Sciences

Kyushu University

Fukuoka, Japan

2018

署名

主査　　久保田哲也

副査　　水野秀明

副査　　磯辺篤彦

Table of Contents

List of Tables

List of Figure

"Learn never known the age, time, and place. Moreover, life is a sharing," My Parent said.

The voice still echoes in the mind

Chapter 1 Introduction

1.1 Background

A natural disaster is a significant adverse event resulting from natural processes of the Earth, i.e., flooding, landslide, tsunami, earthquake, and drought. As the most Natural disaster in Indonesia based on Indonesian National Disaster Management (BNPB) from 2011 – 2015, there are drought, forest fire, earthquake, flooding, and Landslide. Landslide is the natural hazard which causes a significant number of human lost in Indonesia. Landslide is a significant geological hazard worldwide, accounts for a high number of human casualties and an enormous amount of property loss, and causes severe damage to natural ecosystems and human-built infrastructures. Both environmental and triggering factors control landslide events. The environmental factors comprised of topography (e.g. elevation, slope, aspect, and curvature), geological settings (e.g. rock types, faults, and structural aspects), hydrological regimes (e.g. proximity to stream and drainage density), geomorphological situation (i.e. physiographic unit, terrain mapping units and geomorphological units) and human (e.g. land use change and distance to road). Eight hundred and fifty events of landslide caused 462 people die (Table 1). The landslide in Indonesia caused by heavy rainfall, weak material, steep slope and land-use change (change on upland change from forest area to farming area).

Table 1 Distribution of Natural Disasters Based on Indonesian National Disaster Management Agency in Indonesia (DIBI, 2016)

No	Natural Disaster	Event	People die
1	Flooding	1112	343
2	Drought	36	0
3	Landslide	850	462
4	Earthquake	1024	0
5	Forest Fire	81	0

Land use change (LUC) is a process by which human activities transform the landscape. LUC has been recognized throughout the world as one of the most critical factors influencing the occurrence of rainfall-triggered landslides (Glade, 2003), and LUC can have implication to

landslide occurrence on a steep slope (Mugagga et al., 2012). The correlation between intense rainfall and landslide initiation has been examined by many scholars (Crosta et al., 2004; Hasnawir et al., 2017), and triggering thresholds have been determined. In South Sulawesi Indonesia, LUC has been translated into numerous landslide incidents triggered by the intensity of rainfall compared to other factors such as earthquakes, especially in Ujung-Loe upper watershed. The topography is naturally very steep and mountainous (38.8% class slope >20 degrees) and has a very high level of instability, especially during the rainy season (Rainfall: 2,976 to 7,114 mm/year with average annual rainfall 4,524 mm/year; (Makassar Meteorology, Climatology, and Geophysical Agency, 2016). Moreover, the primary occupation of social community in that area is farming and located in the mountainous area and steep slope. It is hard to avoid this agricultural practices because this has become people's culture for agriculture in mountainous regions and have made it hereditary. This primary characteristic of location makes different from the other location and need to analysis for mitigation disaster especially by introduce LUC as a new causative factor to produce susceptibility map.

Geographic information system (GIS) as a tool used to build a map of landslide inventory and causative factor. Moreover, GIS is used to build landslide susceptibility. Quantitative methods employ mathematical models to estimate the probability of landslide occurrence in a region and thus define hazard zones on a continuous scale. To achieve an accurate estimation of the probability of slope failure, a recent landslide inventory map, and complete information on the past mass movements are necessary. Quantitative methods include bivariate statistical models such as frequency ratio, multivariate statistical techniques such as discriminant analysis, and linear and logistics, as well as expert choices such as certainty factor and non-linear methods such as artificial neural networks.

In this research, we divide the processes into three-steps to compare the performance of LUC as a new causative factor. There three-steps were derivate factor for LUC to landslide occurrence; performance of LUC causative factor on landslide susceptibility map (LSM) by

using frequency ratio (FR) and logistics regression (LR) with 9 causative factor, and performance of LUC using FR, LR, and introduce expert choice analysis with certainty factor (CF) with 11 causative factor; and optimization of causative factor using LR and artificial neural network (ANN) method for landslide susceptibility assessment. Furthermore, we also conduct four landslide susceptibility maps from different landslide susceptibility models in GIS environment (FR, LR, CF, and ANN).

1.2 Disaster management in Indonesia

The shift in disaster management paradigm from focusing on disaster response to enhancing disaster risk reduction was started in 2007 by the enactment of Undang-Undang (Law) 24/2007. Scientific society and government awareness drove it after post-tsunami emergency response and subsequent rehabilitation and reconstruction phase. The momentum was also appeared by the experiences of Nabire Earthquake 2004, Nias Earthquake 2005, and Yogyakarta Earthquake 2006 emergency responses. However, the initiative to reform the Disaster Management Law has been started before the earthquake and tsunami of 26 December 2004. There was a discussion forum between BAKORNAS PB (National Disaster Management Coordination and Agency), NGO's and MPBI (Indonesian Society for Disaster Management) to promote national disaster management. Before the enactment Law 24/2007, the disaster management in Indonesia was focusing on crisis management and disaster response coordinated by BAKORNAS PB.

The Disaster Management Law 24/2007 enforces a systematic approach to disaster risk reduction that contains three phases of the disaster management cycle as follows:

1. pre-disaster planning and preparedness, including disaster risk reduction, mitigation, preparedness, risk assessment and contingency planning
2. emergency response, including evacuation, search and rescue, providing immediate assistance, assessing damage and disaster relief
3. post-disaster management, including rehabilitation and reconstruction.

The law also mandates the creation of the "new BAKORNAS PB," later called as BNPB (National Disaster Management Agency), as a national coordinating agency for disaster management that is responsible for pre-disaster planning, emergency response, and post-disaster management. BNPB must coordinate all contingencies, preparedness, mitigation, prevention, disaster management training, risk assessment and risk zoning. In the emergency response phase, BNPB has a responsibility to coordinate government, NGO's and international organization during the emergency response phase. BNPB must also coordinate damage and loss assessment and coordinate rehabilitation and reconstruction in the post-disaster phase.

However, with the high responsibility for conducting disaster management, BNPB needs partners to provide all the technical support, to train technical personnel, and to create preventive disaster risk reduction culture in Indonesia. One of the representative partners to provide technical expertise in the full spectrum of disaster-related fields is the university partner. It is expected to be an intellectual capital, which can provide technical assistance in disaster risk reduction including the research and technology development of early warning systems, damage assessment, and risk analysis.

Risk analysis, as a basis for disaster risk reduction, is an essential issue in the Law 24/2007. Disaster prevention planning should include disaster risk data documentation and risk analysis. Development activities which may have high risk must be equipped with risk analysis. The implementation of risk analysis is closely related to spatial planning or land use planning. Two other laws were also enacted in 2007, i.e., Law 26/2007 spatial planning and Law 27/2007 on coastal zone management and small islands. Both have a secure attachment to disaster mitigation. Law 26/2007 dictates that spatial plan documents should be based on the consideration of disaster mitigation measures. Law 27/2007 states that disaster reduction strategy has to be included in the coastal zones and small islands spatial plan. Spatial planning at national, provincial and regency levels is developed for 20 years and can be reviewed once in 5 years. If a disaster happens due to the development in a high-risk area which is not equipped

with disaster risk analysis, the responsible parties can be fined for up to US$ 26000 or jailed up to 3 years.

Thus, spatial planning based on disaster risk reduction is one of the primary issues of Indonesia's national development agenda to promote sustainable development due to the increasing frequency of disasters and continuing environmental degradation. Regarding landslide disaster risk reduction, regional development and disaster mitigation are well approached by landslide susceptibility, hazard and risk zoning.

1.3 Landslide susceptibility analysis in developing country: lack of data availability

Landslide risk analysis involves several steps, i.e., scope definition, landslide hazard identification and risk estimation. Scope definition addresses several issues including delineating the study area, elements at risk identification, and methodology selection. Landslide hazard identification addresses several issues on understanding the physical characteristic of study area regarding landslide processes such as understanding geology, geomorphology, hydrogeology, and climate. It also includes collecting landslide data, such as landslide classification, area, volume, travel distance, date occurrence, and elements at risk. Hazard identification activities are mostly related to landslide inventory. Risk estimation deals with consequence analysis and frequency analysis.

Landslide inventory is critical in the landslide risk analysis because it gives information related to the frequency of occurrences, landslide typology, landslide extents and damage of elements at risk. Estimation of spatial probability, temporal, probability and magnitude probability is not possible without landslide inventory containing sufficient data of past landslide events. In Indonesia, especially where this research was undertaken, adequate landslide inventory is not available because the landslide locations were very remote area. It is a central problem of quantitative landslide risk analysis in Indonesia. Thus, producing landslide inventory maps and developing approaches of using those maps for landslide risk zoning in Indonesia are challenging tasks that this research focuses.

1.4 Research Scope and Objectives

Landslide is defined, as general terminology, to describe the movement of rock, debris or soil down a slope due to gravitational process (Fell et al., 2008). However, the terminology of landslide, in this research, is used interchangeably to define shallow and deep-seated slide. Landslide hazard and risk analysis, as a soft preventive countermeasure, is a vital tool for disaster risk reduction in Indonesia because of the shifting paradigm of its disaster management from focusing on disaster response to enhancing disaster risk reduction. However, the primary drawback of generating landslide risk analysis is the unavailability of landslide inventory data, which makes difficulties in estimating the spatial probability, temporal probability, and magnitude probabilities.

The landslide in Indonesia caused by heavy rainfall, weak material, steep slope and LUC (change on upland change from forest area to farming area). LUC has been recognized throughout the world as one of the most critical factors influencing the occurrence of rainfall-triggered landslides (Thomas Glade, 2003), and LUC can have implication to landslide occurrence on a steep slope (Mugagga et al., 2012). The correlation between intense rainfall and landslide initiation has been examined by many scholars (Hasnawir et al., 2017; Kubota, 2010), and triggering thresholds have been determined. In South Sulawesi Indonesia, LUC has translated into numerous landslide incidents triggered by the intensity of rainfall compared to other factors such as earthquakes, especially in Ujung-Loe upper watershed. The topography is naturally very steep and mountainous (38.8% class slope >20 degrees) and has a very high level of instability, especially during the rainy season (Rainfall: 1,436 to 5,052 mm/year with average annual rainfall 3,739 mm/year on Apparang Hulu rain gauge (Agency for Meteorology, 2016). Moreover, the primary occupation of social community in that area is farming and located in the mountainous area and steep slope. It is hard to avoid this agricultural practices because this has become people's culture for agriculture in mountainous regions and have made it hereditary. Based on this primary characteristic of location make different from the other location and need

to analysis for mitigation disaster especially by introducing LUC as a new causative factor to produce susceptibility map.

This research will distinguish its analysis based on the availability of landslide causative factors. Thus, the objectives of this research are:

(1) To examining to quantify the LUC in the study area during 2004 to 2011 and establish the relationship between LUC, topographic parameters (slope) and landslide occurrence.

(2) To examining the performance of land use change as a causative factor to produce landslide susceptibility map using frequency ratio, and logistic regression and comparing with 9 causative factor

(3) To examining the effect of land use change to produce landslide susceptibility map using frequency ratio, logistic regression, and certainty factor method and comparing with 11 causative factor

(4) To optimize causative factors by using logistic regression and artificial neural network (ANN) and combination to produce landslide susceptibility map.

1.5 Thesis Organization

The thesis comprises of the following 7 chapters (Figure 1). Chapter 1 is an introduction; chapter 2 is study area. Chapter 3 is preliminary analysis to see the relationship between land use change and landslide occurrence. Chapter 4 and 5 are a central body to see the performance of land use change to produce landslide susceptibility map (LSM). Chapter 6 is an optimized causative factor to produce LSM. Chapter 7 is a summary and conclusion.

Chapter 1 introduces (1) The important of land use change causative factor for analysis for landslide susceptibility map in Ujung-Loe Watershed, South Sulawesi Indonesia, (2) the shifting disaster mitigation policy in Indonesia, (3) the problems in landslide risk

zoning in Indonesia, (4) the scope and objectives of this study, and (5) the organization of the thesis.

Chapter 2 introduces the condition of the study area, i.e., soil and land use/land cover, geology and geomorphology, rainfall condition and socio-economic conditions.

Chapter 3 consist of preliminary analysis to see the relationship between land use change and landslide events and slope factor in Ujung-Loe Watersheds, South Sulawesi, Indonesia.

Chapter 4 consists of analysis the performance of land use change causative factor to produce landslide susceptibility map in comparason with two different landslide susceptibility analysis using bivariate frequency ratio, and multivariate logistic regression with using 9 causative factors (elevation, slope, aspect, curvature, lithology, distance to faults, distance to river, drainage density and land use change). Landslide data were separated into training data (70%) and validation data (30%). In an analysis to produce landslide susceptibility map, we use causative factors with and without LUC to see the performance. The produced landslide susceptibility maps were compared to evaluate the accuracy of each map in the study area of Ujung-Loe Watersheds, South Sulawesi, Indonesia.

Chapter 5 consists of analysis performance of land use change causative factor to produce landslide susceptibility map in comparison with three different landslide susceptibility analysis using bivariate frequency ratio, multivariate logistic regression and introduce expert choice analysis with certainty factor (CF) with 11 causative factors (elevation, slope, aspect, curvature, lithology, distance to faults, distance to river, drainage density, land use change, precipitation and distance to road). Landslide data were separated into training data (70%) and validation data (30%). The produced landslide susceptibility maps were compared to evaluate the accuracy of each map in the study area of Ujung-Loe Watersheds, South Sulawesi, Indonesia

Chapter 6 consists of optimized landslide causative factors (input data) to increase the accuracy of the landslide susceptibility map in Ujung-Loe Watershed, South Sulawesi, Indonesia by using logistic regression and artificial neural network approach. There is 3 type of optimized, i.e., optimized by logistic regression (forward stepwise), artificial neural network (ANN) and combination between logistic regression (forward stepwise) and ANN.

Chapter 7 summarizes and concludes the results and achievements of the study. Problems are also highlighted for future studies.

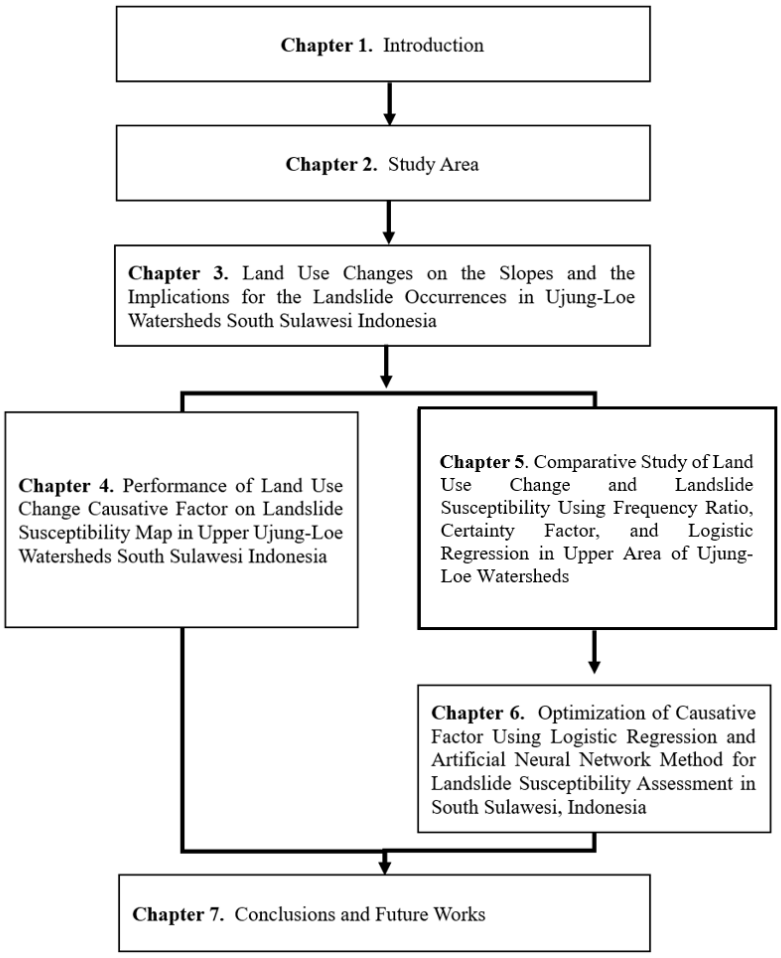


Figure 1 Flowchart of thesis organization

1.6 References

Agency for Meteorology, C. and G., 2016. Rain fall Data from 2002 until 2015. Indonesian Agency for Meteorology, Climatology and Geophysics, Makassar.

Anonym. 2007. Undang-Undang Nomor 24 Tahun 2007 tentang Penanggulangan Bencana, Lembaran Negara Republik Indonesia tahun 2007 nomor 66, Jakarta.

Anonym. 2007. Undang-Undang Nomor 26 Tahun 2007 tentang Penataan Ruang, Lembaran Negara Republik Indonesia tahun 2007 nomor 68, Jakarta.

Anonym. 2007. Undang-Undang Nomor 27 Tahun 2007 tentang Pengelolaan Wilayah Pesisir dan Pulau-pulau Kecil, Lembaran Negara Republik Indonesia tahun 2007 nomor 84, Jakarta.

DIBI. 2016. Indonesian Landslide Disaster Database 2011 - 2015. Accessed on August 2016 from http://dibi.bnpb.go.id/DesInventar/dashboard.jsp

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan 65, 15–31. doi:10.1016/j.geomorph.2004.06.010

Chauhan, S., Sharma, M., Arora, M.K., Gupta, N.K., 2010. Landslide susceptibility zonation through ratings derived from the artificial neural network. Int. J. Appl. Earth Obs. Geoinf. 12, 340–350. doi:10.1016/j.jag.2010.04.006

Crosta, G.B., Chen, H., Lee, C.F., 2004. Replay of the 1987 Val Pola Landslide, Italian Alps. Geomorphology 60, 127–146. doi:10.1016/j.geomorph.2003.07.015

Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One 10. doi:10.1371/journal.pone.0133262

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land use planning. Eng. Geol. 102, 85–98. doi:10.1016/j.enggeo.2008.03.022

Glade, T., 2003. Landslide occurrence as a response to land use change: A review of evidence from New Zealand. Catena 51, 297–314. doi:10.1016/S0341-8162(02)00170-4

Hasnawir, Kubota, T., Sanchez-Castillo, L., Soma, A.S., 2017. The Influence of Land use change and rainfall on shallow landslide in Tanralili Sub-watwrshed, Indonesia. J. Fac. Agric. Kyushu Univ. 62, 171–176.

Kubota, T., 2010. Shallow Landslides Distribution and Rainfall Threshold in Kelara Watershed, Indonesia 1–8.

Mugagga, F., Kakembo, V., Buyinza, M., 2012. Land use changes on the slopes of Mount Elgon and the implications for the occurrence of landslides. Catena 90, 39–46. doi:10.1016/j.catena.2011.11.004

.

Chapter 2 Study Area

2.1 Introduction

Upper of Ujung-Loe Watersheds was a mountainous area and located in Bulukumba and Sinjai Regency, South Sulawesi Province, Indonesia (figure 2). It opposites with Jeneberang watershed with the big caldera of bawakaraeng. This location provides a fertile land but frequently suffers from landslide disasters. Landslide disasters occur almost every year, especially during the rainy season, which induces flash floods and debris flows in the upstream.

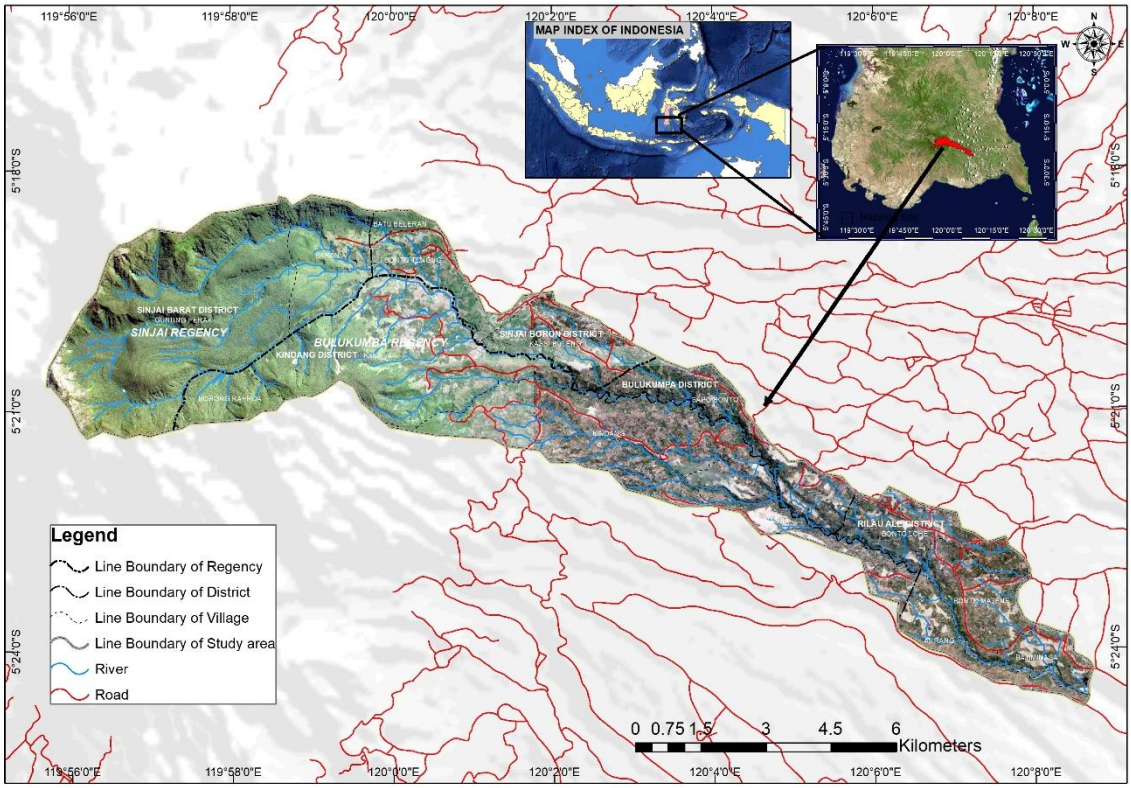


Figure 2 Study Area

The upper of Ujung-Loe Watersheds is located at 119° 55' 42.34"E to 120° 8' 43.12"E and 5° 18' 19.07" S to 5° 24' 43.33" S with the altitude of 255 – 2,860 meters above sea level with areas of 79.79 km2. It provided forests covering and area of cultivation and farming. Some areas are, particularly in the upstream part. The slope is around 38.8% with slope class >20 degrees including a particular area at the upstream with very steep (>40 degrees).

2.2 Soil and Land Use/Land Cover

According to map land system from Departemen Transmigrasi Indonesia (1987), soil type of the study area is dystrandepts and ustropepts (Table 2 and Figure 3).

Table 2 Soil (great group) according to land system Sulawesi 1982 in the study area

Number	Soil (great group)	Area (hectare)
1	Dystrandepts	4,061
2	Ustropepts	3,918
	Grand Total	7,979

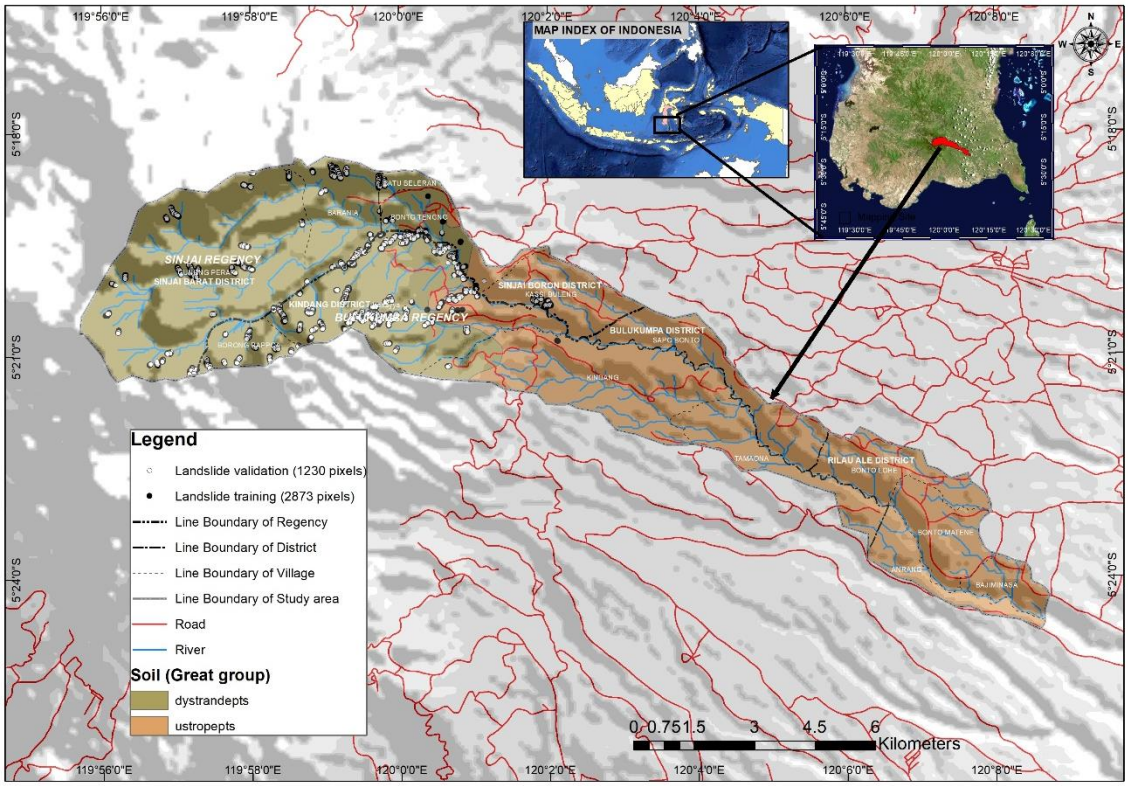


Figure 3 Soil Map

Dystrandepts and ustropepts is an inceptisol soil. Inceptisols commonly occur on landscapes that are relatively active, such as mountain slopes, where erosional processes are actively exposing unweathered materials, and river valleys, where relatively unweathered sediments are being deposited. The different is Dystrandept with andepts formed chiefly in volcanic ash or regoliths with high componence of ash and ustropets with tropepts formed an ustic moisture regime and receive dominantly summer precipitation, or they have an isomesic, hyperthermic, or warmer temperature regime. They formed mostly in Pleistocene or Holocene

deposits. Some of the soils that have steep slopes formed in older deposits (Resources and Service, 1999)

Table 3 Land use/Land cover 2015 in the study area

Number	Land use/Land cover	Area (hectare)	Percentage
1	Primary Forest	2,527	31.67
2	Secondary Forest	2,176	27.27
3	Savanna	1,342	16.82
4	Scrub	87	1.09
5	Farming Area	1,410	17.67
6	Open Area/Paddy Field include settlement	159	1.99
7	Open Area/Paddy Field	251	3.15
8	Open Area	28	0.35
	Total	7,979	

Source: interpretation of Landsat 8 image recorded 2015

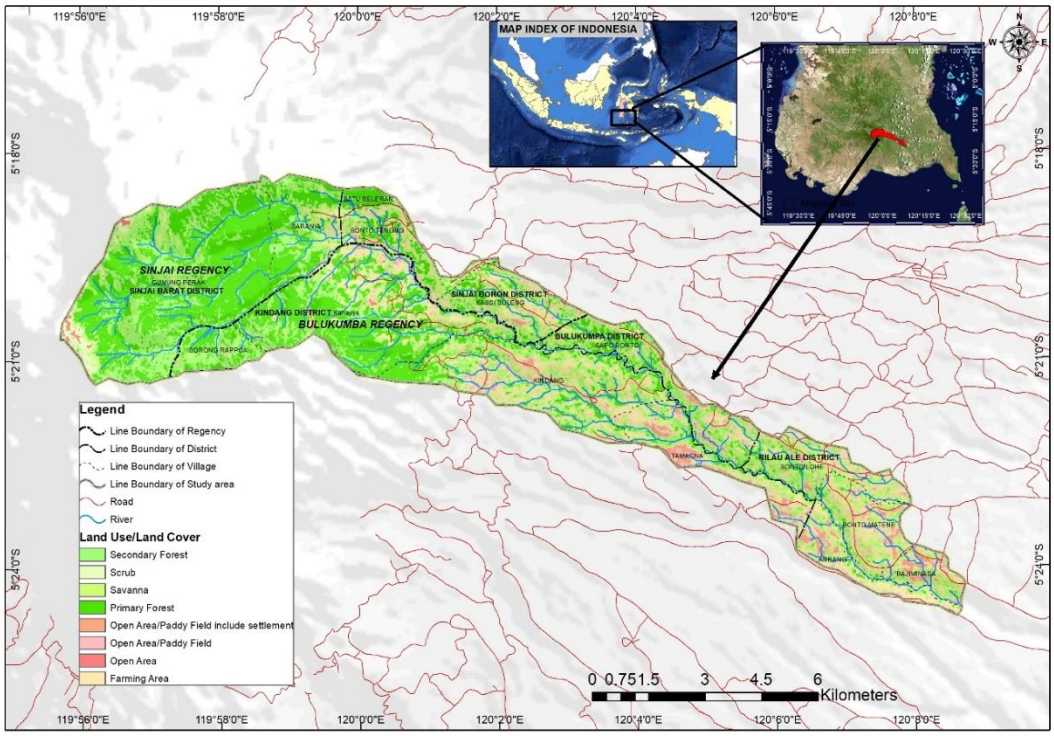


Figure 4 Land use/Land cover 2015 in study area

Concerning to the type of land use/land cover in the current research, 31,67% of the land surface has a primary forest cover, and 27.27 % has a secondary forest in the upper to the middle of watersheds. Moreover, land surface has a cover by savanna, farming area, paddy field and open area with 41.07% from upper to downstream of watersheds (Table 3 and Figure 4).

2.3 Geology and General Geomorphology

High relief characterizes the morphology of Upper Ujung-Loe Watershed, extreme slope (Figure 5), a high degree of weathering as well as erosion activities as soil movement and landslides. The base map on the location according to geological maps of Sulawesi consists of Quarter Lompobattang Vulcanic (Qlv), Lompobattang Vulcanic center (Qlvc), and Quarter Lompobattang Vulcanic Breccia (Qlvb)(Sukamto and S. Supriatna, 1982).

Table 4 Formation of Lithology according to land system Sulawesi 1982 in the study area

Number	Lithology	Area (hectare)	Percentage
1	Quarter Lompobattang Vulcanic Breccia (Qlvb)	1,965	24.63
2	Quarter Lompobattang Vulcanic (Qlv)	5,643	70.72
3	Quarter Lompobattang Vulcanic center (Qlvc)	371	4.65
	Grand Total	7,979	

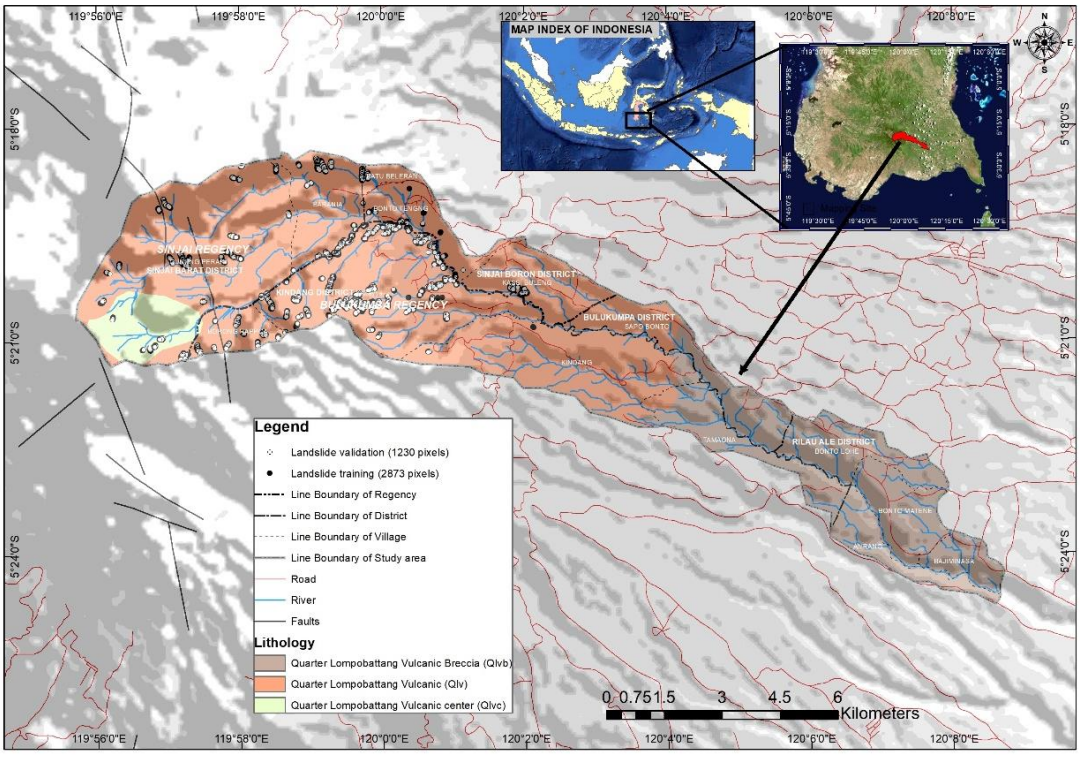


Figure 5 Geology Map

The formation of lithology in current research was dominated by Quarter Lompobattang Vulcanic (Qlv) with 70.72%, then Quarter Lompobattang Vulcanic Breccia (Qlvb) and Quarter Lompobattang Vulcanic center (Qlvc) with 24.63% and 4.65% respectively (Table 4 and Figure

5). The Qlv and Qlvc consist of extrusive, mafic, and polymict, which form a broad stratovolcano and quarter lompobattang volcanic were estimated starting from the last of Pliocene to early Pleistocene of volcanic rock. Qlvb consists of extrusive, mafic, and polymict, and are estimated to start from first Pleistocene to early Holocene of volcanic rock (Sukamto and S. Supriatna, 1982).

2.4 Rainfall condition

In the current research, there are three rain gauge stations, i.e., Bulo-bulo, Apparang Hulu and Malino. In Malino rain gauge station, the monthly rainfall ranges from 41 mm in August to 772 in December with period 2002 until 2015. The intense rainfall usually occurs from November to June (rainy season). There are two distinct seasons, i.e., dry and rainy. The dry season from July to October. The yearly rainfall ranges from 319 mm in 2012 to 5474 mm in

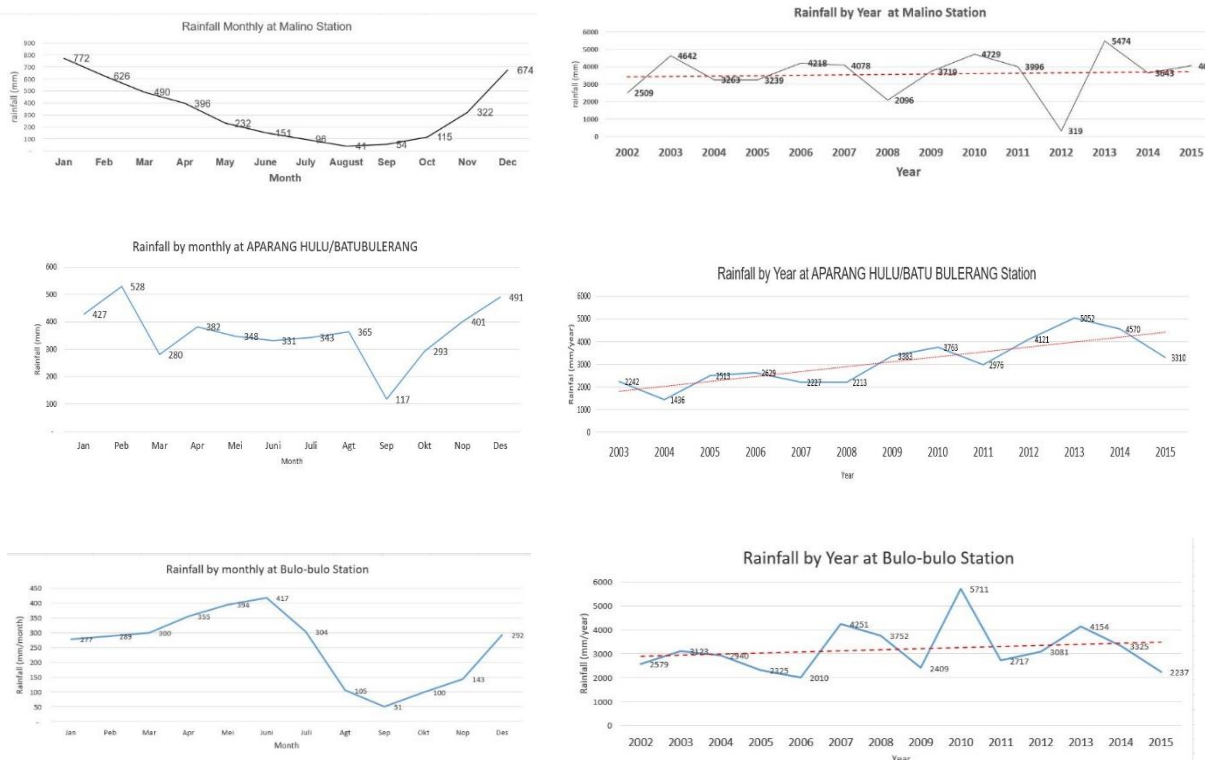


Figure 6 Graph of Rainfall by Year and Average Rainfall by Month From 2002 Until 2015 in Three Rain Gauge Station, i.e., Malino, Apparang Hulu, And Bulo-Bulo

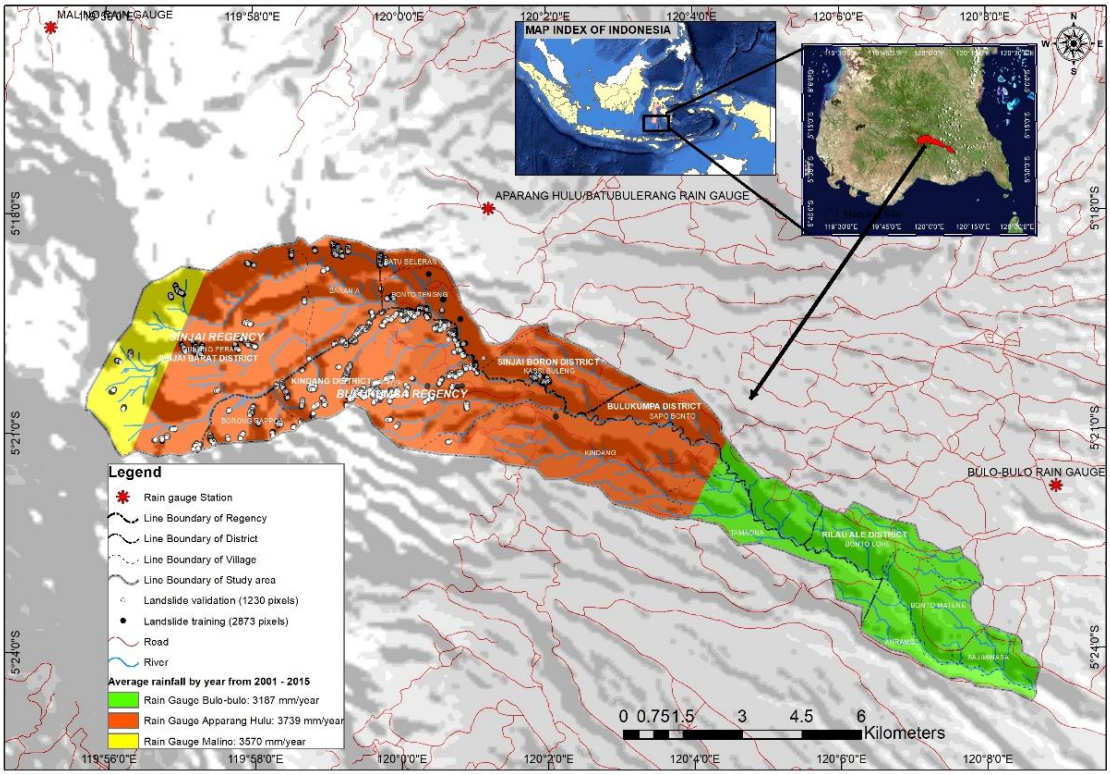


Figure 7 Rainfall Map with Polygon Thiessen

2013 with average 3570 mm/year. Its annual rainfall tends to be stable in the period 2002 to 2015. In Apparang Hulu rain gauge station, the monthly rainfall ranges from 117 mm in August to 528 in December with period 2002 until 2015. The intense rainfall usually occurs from October to August (rainy season). The dry season is only in September. The yearly rainfall ranges from 1436 mm in 2004 to 5052 mm in 2013 with average 3739 mm/year. Its annual rainfall tends to increase in the period 2002 to 2015. In Bulo-bulo rain gauge station, the monthly rainfall ranges from 51 mm in September to 417 in June with period 2002 until 2015. The intense rainfall usually occurs from November to July (rainy season). The dry season from August to October. The yearly rainfall ranges from 2010 mm in 2012 to 5711 mm in 2010 with average 3187 mm/year. Its annual rainfall tends to be stable in the period 2002 to 2015. Its annual rainfall tends to increase in the period 2002 to 2015 (Figure 6) (Agency for Meteorology, 2016).

According to polygon Thiessen analysis, Apparang Hulu station covered area 68.81% area, Bulo-bulo station covered 23.14%, and Malino station covered 7.75% (Table 5 and Figure 8).

2.5 Socio-economic conditions

The study area is located in Bulukumba, and Sinjai regency, South Sulawesi Province, Indonesia. Socio-economic of intermediate status comparing to nearby city and regency including Makassar, Bantaeng, and Gowa. Bulukumba regency with three districts (Bulukumpa, Kindang, and Rilau Ale district) with five villages. Sinjai regency with two districts (Sinjai Barat and Sinjai borong district) with five villages. In this current research area, the highest population located in Anrang village (1,963 people/km2) and Borong Rappoa village (1,553 people/km2) (Table 6 and Figure 8). There is 70.09% of the total population located in Bulukumba regency and only 29.91% in Sinjai regency (Badan Pusat Statistik, 2016a, 2016b).

Table 5 Rain gauge station in the study area

Number	Rain Gauge Station	Area (hectare)	Percentage
1	Bulo-bulo	1,870	23.44%
2	Apparang Hulu	5,490	68.81%
3	Malino	619	7.75%
	Grand Total	7,979	

Table 6 Population data of in current research location

Number	Regency	District	Village	Area (Ha)	Percentage Area (%)	Population (People)	Percentage Population(%)	Population (People/Km2)
1	Bulukumba	Bulukumpa	Sapo Bonto	527	6.61	3,738	10.93	361
		Kindang	Borong Rappoa	221	2.77	3,430	10.03	1,553
			Kahayya	1,276	15.99	1,238	3.62	97
			Kindang	1,085	13.60	3,032	8.86	279
			Tamaona	417	5.23	2,567	7.50	616
		Rilau Ale	Anrang	259	3.25	2,493	7.29	1,963
			Bajiminasa	192	2.41	3,165	9.25	331
			Bonto Lohe	410	5.14	2,384	6.97	254
			Bonto Matene	386	4.83	1,929	5.64	232
	Sub Total Of Bulukumba Regency			4,773	59.82	23,976	70.09	632
2	Sinjai	Sinjai Barat	Barania	477	5.97	1,965	5.74	105
			Gunung Perak	1,931	24.20	3,115	9.11	136
		Sinjai Borong	Batu Belerang	69	0.86	1,772	5.18	197
			Bonto Tengnga	313	3.92	1,423	4.16	211
			Kassi Buleng	417	5.23	1,956	5.72	369
	Sub Total Of Bulukumba Regency			3,206	40.18	10,231	29.91	204
	Grand Total			7,979	100.00	34,207	100.00	418

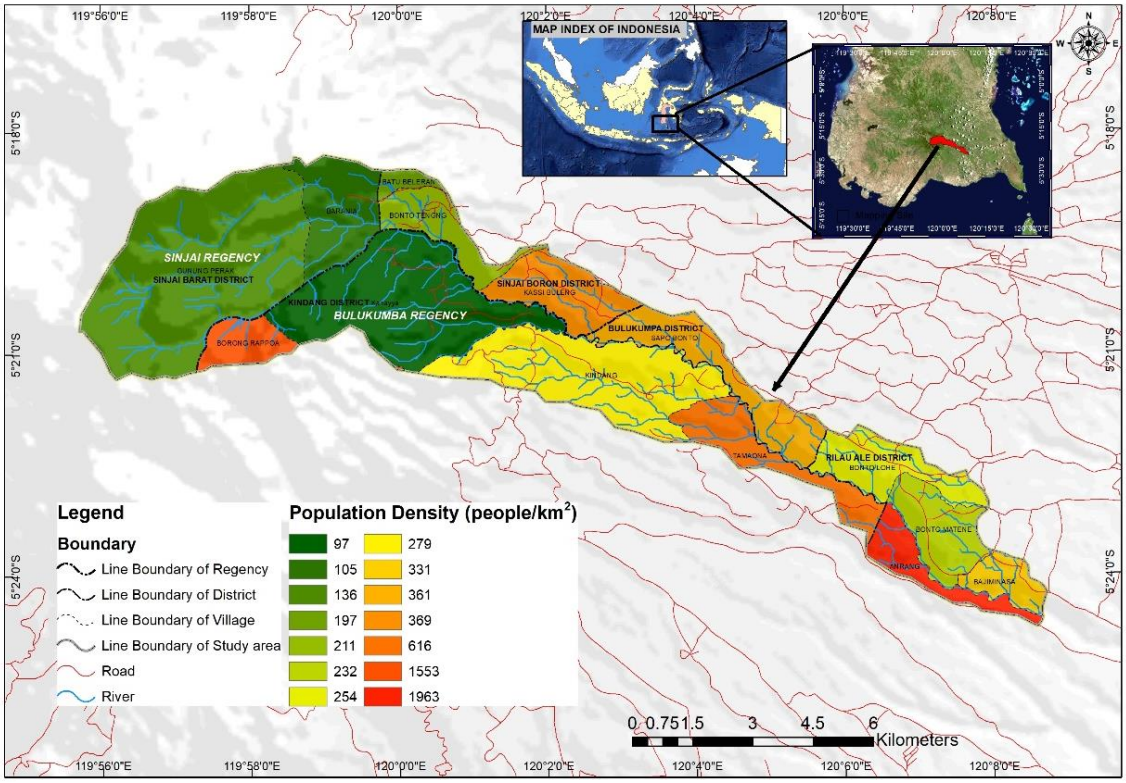


Figure 8 Population Density Map

The topography of the disaster-prone area in the upstream of Ujung-loe watershed is mountainous. Paddy and corn fields are limited only along the river and in terrace area. Plantations of coffee, coconut, cacao, clove, candlenut, cashew, vanilla, pepper as well as paddy and corn fields dominate the village industry. The vegetable is grown mostly for self-consumption. They also raise cow, horses, and goats (Badan Pusat Statistik, 2016a, 2016b)

2.6 References

Agency for Meteorology, C. and G., 2016. Rain fall Data from 2002 until 2015. Indonesian Agency for Meteorology, Climatology and Geophysics, Makassar.

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains , Central Japan 65, 15–31. doi:10.1016/j.geomorph.2004.06.010

Badan Pusat Statistik, 2016a. Sinjai dalam Angka. Badan Pusat Statistik Kabupaten Sinjai, Sinjai.

Badan Pusat Statistik, 2016b. Bulukumba dalam Angka. Badan Pusat Statistik Kabupaten Bulukumba, Bulukumba.

Chauhan, S., Sharma, M., Arora, M.K., Gupta, N.K., 2010. Landslide susceptibility zonation through ratings derived from artificial neural network. Int. J. Appl. Earth Obs. Geoinf. 12, 340–350. doi:10.1016/j.jag.2010.04.006

Crosta, G.B., Chen, H., Lee, C.F., 2004. Replay of the 1987 Val Pola Landslide, Italian Alps. Geomorphology 60, 127–146. doi:10.1016/j.geomorph.2003.07.015

Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One 10. doi:10.1371/journal.pone.0133262

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., 2008. Guidelines for landslide

susceptibility, hazard and risk zoning for land use planning. Eng. Geol. 102, 85–98. doi:10.1016/j.enggeo.2008.03.022
Glade, T., 2003. Landslide occurrence as a response to land use change: A review of evidence from New Zealand. Catena 51, 297–314. doi:10.1016/S0341-8162(02)00170-4
Hasnawir, Kubota, T., Sanchez-Castillo, L., Soma, A.S., 2017. The Influence of Land use change and rainfall on shallow landslide in Tanralili Sub-watwrshed, Indonesia. J. Fac. Agric. Kyushu Univ. 62, 171–176.
Kubota, T., 2010. Shallow Landslides Distribution and Rainfall Threshold in Kelara Watershed , Indonesia 1–8.
Mugagga, F., Kakembo, V., Buyinza, M., 2012. Land use changes on the slopes of Mount Elgon and the implications for the occurrence of landslides. Catena 90, 39–46. doi:10.1016/j.catena.2011.11.004
Resources, N., Service, C., 1999. Soil Taxonomy. doi:10.1017/S0016756800045489
Soma, A.S., Kubota, T., 2017. The performance of Land Use Change Causative Factor on Landslide Susceptibility Map In Ujung-Loe Watersheds South Sulawesi Indonesia. Geoplanning J. Geomatics Plan. 4.
Sukamto, R., S. Supriatna, 1982. Geologic Map of The Ujungpandang, Benteng and Sinjai Quadrangles, Sulawesi. Bandung, Indonesia. Geological Research and Development Centre, Bandung, Indonesia.

Chapter 3 Land Use Changes on the Slopes and the Implications for the Landslide Occurrences in Ujung-Loe Watersheds South Sulawesi Indonesia

3. 1 Introduction

Land use change (LUC) is a process by which human activities transform the landscape. LUC has been recognized throughout the world as one of the most critical factors influencing the occurrence of rainfall-triggered landslides (Thomas Glade, 2003), and LUC can have implication to landslide occurrence on a steep slope (Mugagga et al., 2012). The correlation between intense rainfall and landslide initiation has been examined by many scholars (Bacchini and Zannoni, 2003; Crosta and Frattini, 2003), and triggering thresholds have determined.

In South Sulawesi Indonesia, LUC has been translated into numerous landslide incidents triggered by the intensity of rainfall compared to other factors such as earthquakes, especially in Ujung-Loe upper watershed. The topography is naturally very steep and mountainous (38.8% class slope >20 degrees) and has a very high level of instability, especially during the rainy season (annual rainfall 1436 mm/year – 5052 mm/year period 2002 to 2015 at Apparang Hulu rain gauge (Meteorology, climatology and geophysics Makassar, 2016)). The primary occupation of social community in that area is farming and located in the mountainous area. It is hard to avoid this agricultural practices because this has be become people's culture for agriculture in mountainous regions and have made it hereditary. Rudiarto and Doppler (2013) said that in Indonesia, where many upland areas can found, land use/cover change for the extension of agriculture activity commonly occurs.

The objectives of the study are examining to quantify the LUC in the study area during 2004 to 2011 and establish the relationship between LUC, topographic parameters (slope) and landslide occurrence.

3. 2 Material and Method

This study utilized Landsat satellite images on scene path 114 row 064 of 2004 and 2011. Landsat 5 TM (date recorded September, 21th 2004) and Landsat 7 ETM+ (date recorded October, 11th 2011) images, each with a 30 m resolution, and Aster DEM 30m, are collected from United States Geological Survey (USGS) at website http://earthexplorer.usgs.gov/. Resolution of pixel was 30 x 30 meter with total pixel 88,879 and 383 pixels for a landslide.

Landsat images have been registered and geo-corrected from the source. The atmospheric correction has been conducted with the use of vegetation delineation function was through to the ArcGIS software package version 10.3. Also, the radiometric correction has been carried out by using the Landsat calibration function within the tools of Arc GIS 10.3 software.

Remote sensing and GIS (RSGIS) techniques have shown great potential in land use mapping and monitoring due to its advantages over traditional procedures regarding cost-effectiveness and timeliness in the availability of information over larger areas (Armentaras et al.,2003; Franklin, 2001). RSGIS techniques were employed to classify land use. The unsupervised classification method is applied to classify land use. The unsupervised classification consists of three step: (1) Creation of N spectral-class maps using Iterative Self-Organizing Data Analysis Technique Algorithm; (2) development of Land use (LU) maps with assistance of reference data; and (3) accuracy assessments of all the LU maps using independent reference data and selection of one LU map with the highest accuracy (Ruili et. al., 2008). This method is applied to classify land use into seven classes, i.e., open area, paddy field, farming area, scrub, savanna, secondary forest and primary forest. The certified maps, which were generated, classified and validated by using ground control points method in the same year of Landsat images and google earth pro imagery map, were used to measure the accuracy assessment. Overall accuracy values of 86%, 90% and Kappa values of 0.83, 0.88 were achieved for the unsupervised classified maps of 2004 and 2011, respectively. In this research, land use was analysis by unsupervised classification and classify in five classes. The five class

were no vegetation (open area), sparse vegetation (paddy field), medium vegetation (farming area and shrub, savana), high vegetation (secondary forest) and dense vegetation (primary forest) is conducted. Slope was divided into 6 clases i.e. 0 -10, 10-20, 20-30, 30-40, 40-50 and >50 degrees by analysis in software ArcGIS 10.3 (Fig. 3b). Historical landslide inventory was delineated in google earth pro with high-resolution image time series from 2012 to 2014 (Figure 10 and Figure 11). Moreover, the framework of this research follows Figure 9. To see the correlation between the LUC, slope and landslide occurrence was conducted by statistical analysis (binary logistic regression). Categorical data were arbitrarily coded to convert it to nominal data. Binary logistic regression formula is Equation 3.1;

$$P = \frac{1}{1+\exp^{-\sum \beta i \, x \, ij}} \tag{3.1}$$

where P is Probability, β is regression coefficients, and ij is variable at observation (LUC and slope).

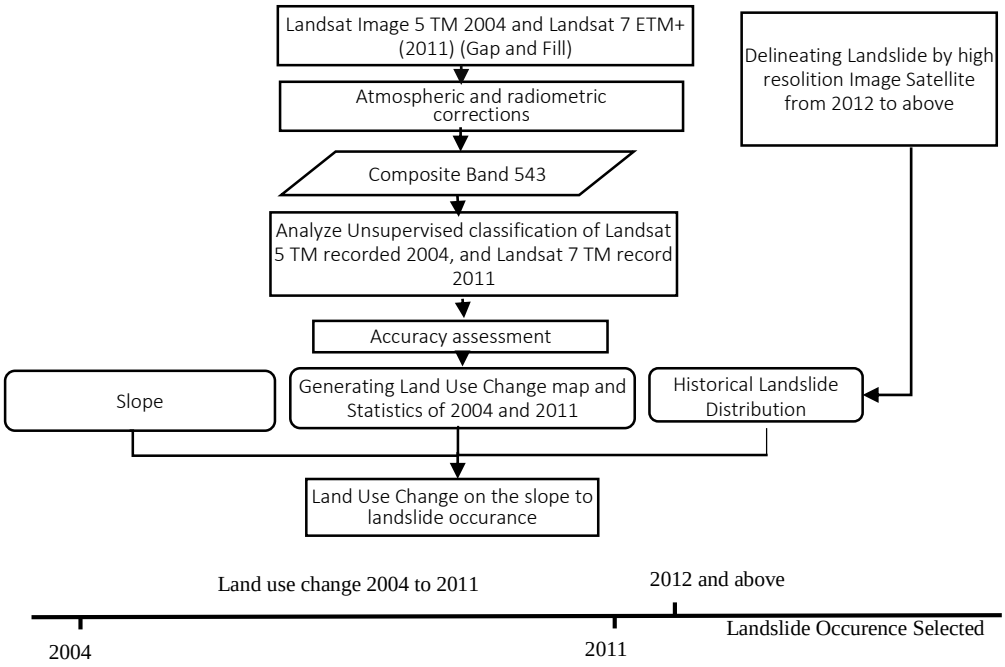


Figure 9 Research framework

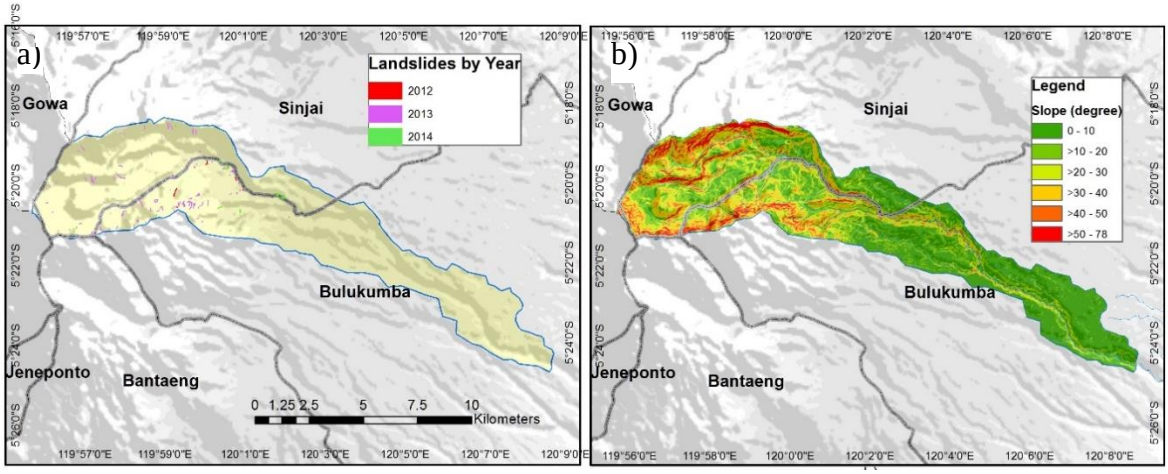


Figure 10 a) Landslide inventory b) Slope study area in degree

3. 3 Results and Discussion

Based on GIS and remote sensing analysis, it is found that the highest increasing land use change (LUC) was medium-vegetation with 45.29% from 2004 to 2011, follow by high vegetation and sparse vegetation, with increasing 41.64% and 19.08 respectively. Moreover, the highest decreasing was no vegetation with 60.69%. Then dense vegetation with 45.72% is declining (see Table 7, Figure 12). The significant trend of increasing average rate of LUC

Figure 11 Landslide photo

were medium vegetation and high vegetation with the mean rate of change 5.66% and 5.21% respectively. Moreover, the significant trend of decreasing average rate of LUC was no vegetation and dense vegetation with the mean rate of change -7.59% and -5.71% respectively. LUC confirmed the growth of the medium-vegetation class from 25.22% from total area in 2004 to 36.64% in 2011 with the highest change from high vegetation; it happened because of the increased rates of farming areas (Figure 13). This rising rate can attribute to the growing population in need of agricultural land in the study area. Moreover, the increasing population rate was 1.15%, primarily by farm workers and people with education below junior high school (Central Bureau of Statistics of South Sulawesi Indonesia, 2013). On the other hand, the class of dense vegetation decreased from 1,685 ha in 2004 to 914 ha in 2011 or decreased 96.27 ha per year because of open land for farming and illegal logging in dense forest areas. Perceived susceptibility and severity of land degradation strongly influence farmers' awareness of and attitude toward environmental problems (Bayard et. Al., 2006).

Landslides have occurred 128 times in 2012 to 2014. The highest landslides were in 2013 with 93 times. They were happening with rainfall with 5,052 mm/year and intensity 27.8 mm/day in 2013 (Meteorology, climatology and geophysics Makassar, 2016). The significant derivate factor of land use/cover change to landslides in the study area was shown in Figure 14. Results of the land use/cover to landslides occurrence analysis confirmed the highest of LUC from high vegetation to medium vegetation on the slope >30 degrees with 82 landslides occurrence with the high intensity of rainfall, particularly in 2013. It happens because the decrease in the vegetation can make adverse influence to the stability slope as Kubota et al. (2007) said that land with forest by the root system would reinforce the soil strength and stabilize the slope. Moreover, the probability of landslide occurrence, particularly shallow landslides increases and is very sensitive to short-lasting high intensive rainfall (Hasnawir and Kubota, 2012; Aditian and Kubota, 2017).

Table 7 Land Use Change from 2004 to 2011

Land Use Density Class	2004 Land Use Density Class area			2011 Land Use Density Class area			Change between 2004 to 2011		Average rate of change	
	pixels	ha	%	pixels	ha	%	ha	%	ha/year	%
No Vegetation	15470	1392	17.41	6081	547	6.84	-845	-60.69	-105.63	-7.59
Sparse Vegetation	25021	2252	28.15	29794	2681	33.52	430	19.08	53.70	2.38
Medium Vegetation	22416	2017	25.22	32568	2931	36.64	914	45.29	114.21	5.66
High Vegetation	7255	653	8.16	10276	925	11.56	272	41.64	33.99	5.21
Dense Vegetation	18717	1685	21.06	10160	914	11.43	-770	-45.72	- 96.27	-5.71
Grand Total	88879	7999	100.00	88879	7999	100.00				

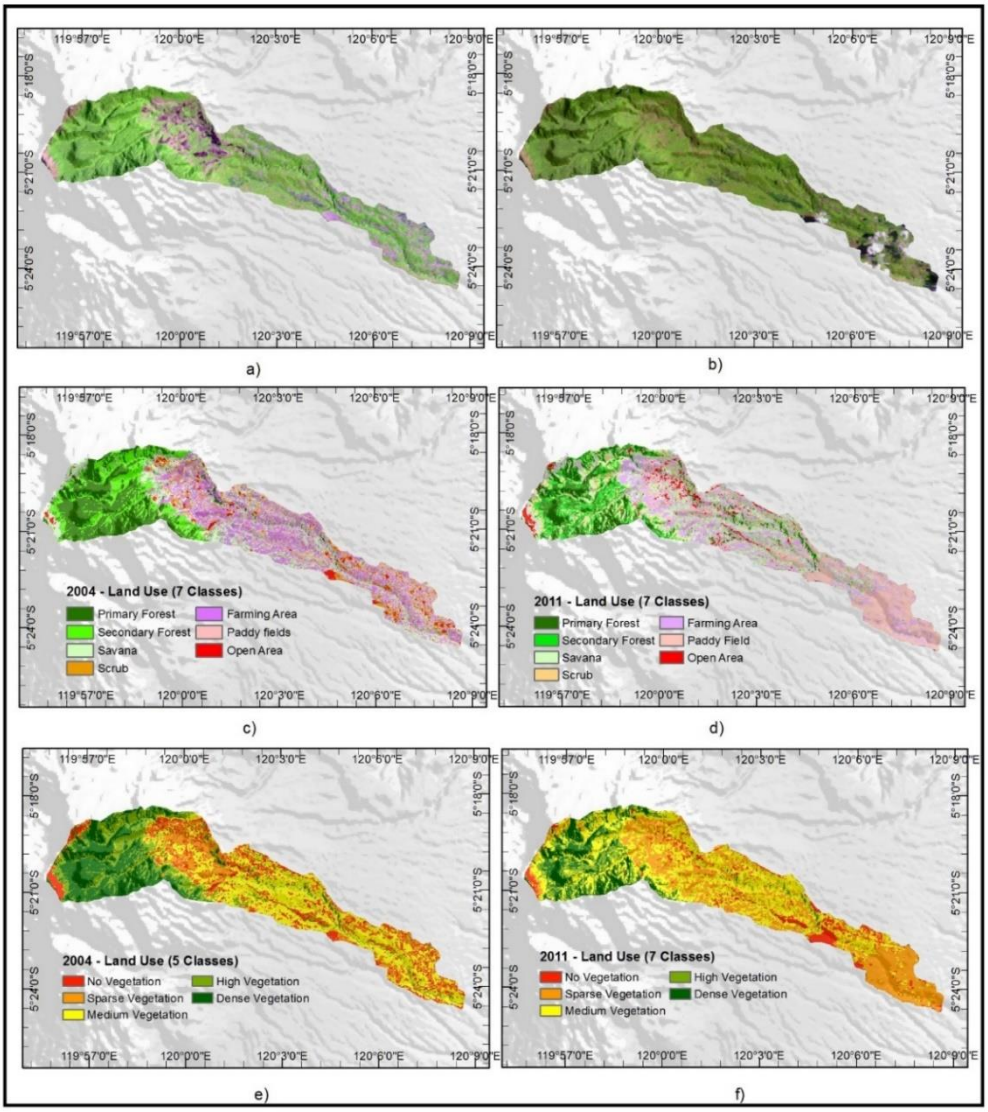


Figure 12 Landsat 5 TM recorded 2004 band 543; b) Landsat 5 TM recorded 2011band 543; c) Land use 2004 (7 classes); d) Land use 2011 (7 classes); e) Land use 2004 (5 classes); and f) Land use 2004 (5 classes)

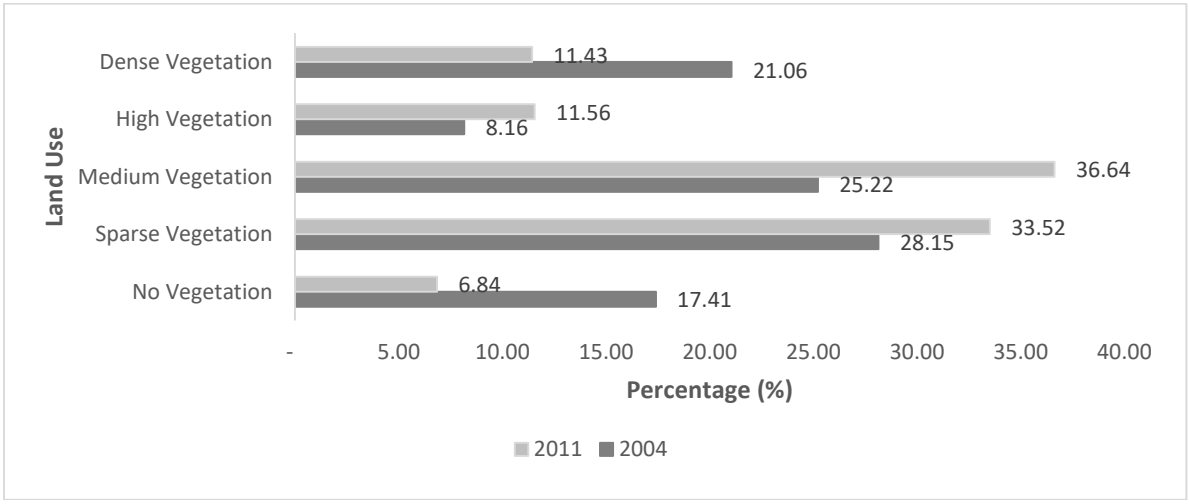
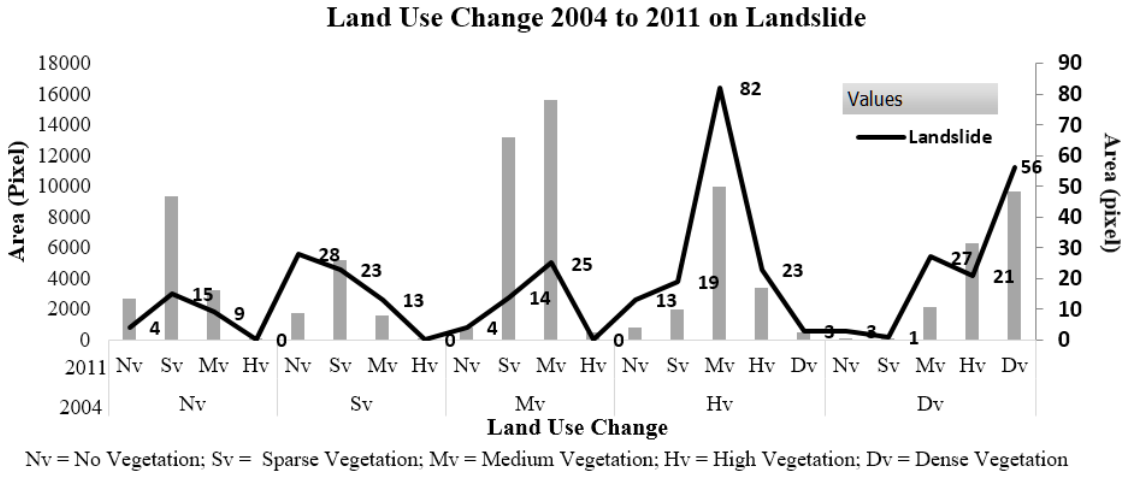
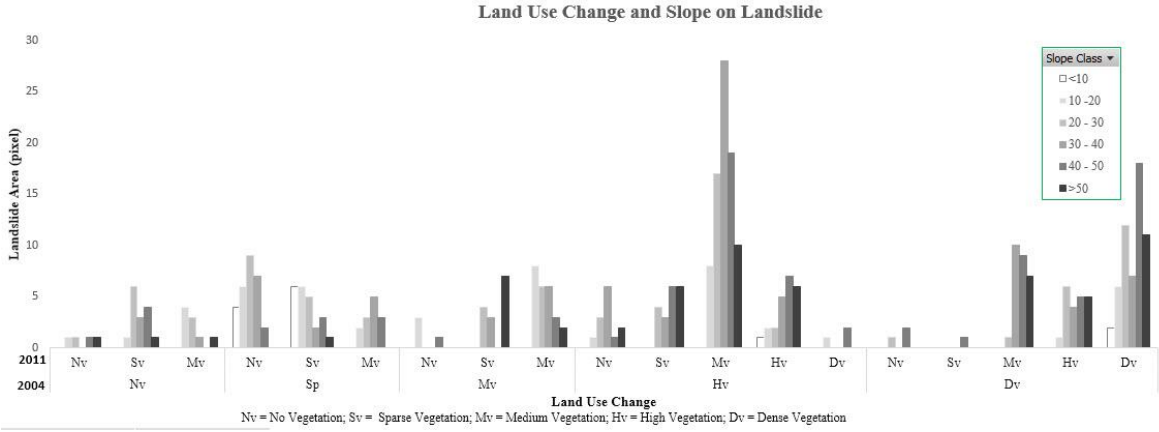


Figure 13 Histogram of Land Use Change from 2004 to 2011



Nv = No Vegetation; Sv = Sparse Vegetation; Mv = Medium Vegetation; Hv = High Vegetation; Dv = Dense Vegetation

Figure 14 Land Use Change from 2004 to 2011 on Landslide Occurrence



Nv = No Vegetation; Sv = Sparse Vegetation; Mv = Medium Vegetation; Hv = High Vegetation; Dv = Dense Vegetation

Figure 15 Slope and Land Use Change from 2004 to 2011 on Landslide Occurrence

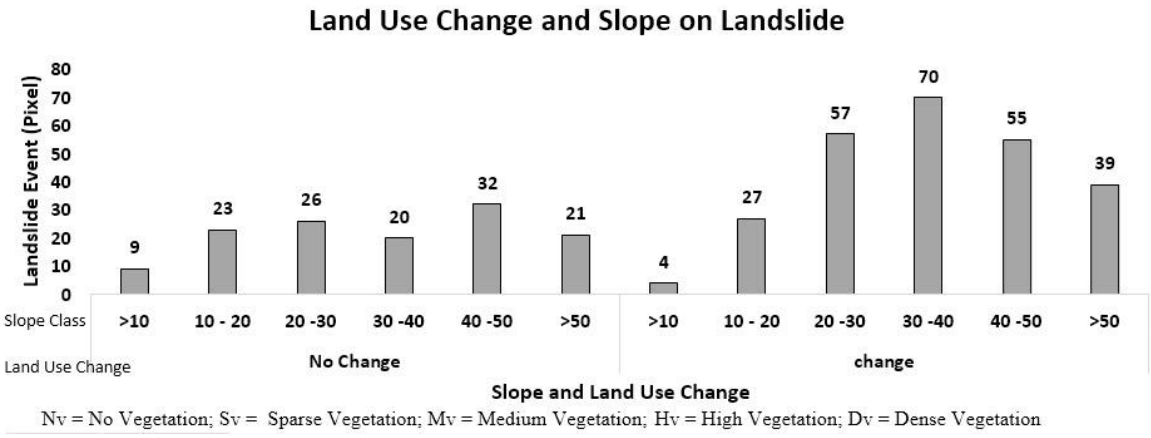


Figure 16 General land use change from 2004 to 2011 on landslide occurrence

Table 8 Variables in the Equation of Slope and Land Use Change

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	slope	.662	.034	384.242	1	.000	1.939
	Land use change	.312	.108	8.327	1	.004	1.367
	Constant	-7.729	.160	2319.998	1	.000	.000

a. Variable(s) entered on step 1: slope, land use change.

Logistic regression analysis indicates that LUC (from the dense vegetation to high vegetation, high vegetation to medium vegetation respectively, and so on) has influences on landslide occurrences. Further statistical analysis showed that in term of influence to landslide occurrences, slope showed greater influence (B-intercept: 0.662) compared to individual LUC (B-intercept: 0.312) (Table 8 and Figure 16). It was like as Bergueria and Santiago (2005) that LUC affects with the landslide occurrences.

3. 4 Conclusion

Based on the obtained results and subsequent discussions, the following conclusions are presented:

1. The significant decrease of land use change (LUC) in 2004 to 2011 was observed at Ujung-Loe watersheds in no vegetation (-7.59%), and dense vegetation class (-5.7%) while the increased LUC was found in the class of medium vegetation (5.66%) and high vegetation (5.21%).

2. Landslides have occurred most dominantly in the one with the LUC from high vegetation to medium vegetation on slope 30 – 40 degrees.

In general, LUC in Ujung-Loe watershed indicates a significant effect on landslides occurrence and slope instability.

3. 5 References

Aditian, Aril, and Kubota, Tetsuya, 2017. The influence of Increasing Rainfall Intensity on Forest Slope Stability in Aso Volcanic Area, Japan. International Journal of Ecology and Development vol. 32 number 1 page 66-74

Armentaras, D., Gast, F., Villareal, H., 2003. Andean forest fragmentation and the representativeness of protected areas in the Eastern Andes, Colombia. Biological Conservation 113, page 245–256.

Bacchini, M., Zannoni, A., 2003. Relations between rainfall and triggering of debris flows: case study of Cancia (Dolomites, north-eastern Italy). Natural Hazards and Earth System Sciences 3, page 71–79.

Bergueria, Santiago, 2005. Changes in land cover and shallow landslide activity: A case study in the Spanish Pyrenees. Elsevier: Geomorphology 74, page 196– 206.

Buddy Bayard, Curtis M. Jolly, Dennis A. Shannon, Alejandro A. Lazarte, 2006. Low-income Farmers' Behavior Toward Land Degradation: The Effects of Perceptions, Awareness, Attitude, and Land Use. International Journal of Ecological Economics and Statistics, vol. 6 number 3 page 56-63.

Central Bureau of Statistics of South Sulawesi Indonesia, 2013. Sulawesi Selatan 2013 in Figures. Badan Pusat Statistik Provinsi Sulawesi Selatan, Makassar.

Crosta, G.B., Frattini, P., 2003. Distributed modeling of shallow landslides triggered by intense rainfall. Natural Hazards and Earth System Sciences 3, 81–93.

Franklin, S., 2001. Remote sensing for forest management. Lewis, FL.

Hasnawir and Tetsuya Kubota, 2012. Rainfall threshold for shallow landslide in Kelara Watersheds, Indonesia. Internasional Journal of Japan Erosion Control Engineering Technical Note 5(No.1):86-92

Agency of Meteorology, climatology and geophysics Makassar, Indonesia, 2016. Rainfall data from 2002 to 2016. BMKG, Makassar, Indonesia

Mugagga, F., Kakembo, V., Buyinza, M., 2012. Land use changes on the slopes of Mount Elgon and the implications for the occurrence of landslides. Catena 90, page 39-46.

Rudiarto I, Doppler W.,2013. Impact of land use change in accelerating soil erosion in Indonesian upland area: A case of Dieng Plateau, Central Java – Indonesia. International Journal of AgriScience Vol. 3(7): 558-576.

Ruili Lang, Guofan Shao, Bryan C. Pijanowski, Richard L. Farnsworth, 2008. Optimizing unsupervised classifications of remotely sensed imagery with a data-assisted labeling approach. Computer and Geoscience vol 34 page 1877-1885.

Tetsuya Kubota, Omura, H., and Devkota, B.D., 2007. Influence of The Forest on Slope Stability with Different Forest felling Condition. EGU General Assembly 2007 Vienna.

Thomas Glade, 2003. Landslide occurrence as a response to land use change: a review of evidence from New Zealand. Catena vol 51 page 297 – 314.

Chapter 4 Performance of Land Use Change Causative Factor on Landslide Susceptibility Map in Upper Ujung-Loe Watersheds South Sulawesi Indonesia

4. 1 Introduction

Land use changes (LUC) has increased the level of vulnerability to landslides, especially in mountainous regions. It is recognized throughout the world as one of the most critical factors influencing the occurrence of rainfall-triggered landslides (Glade, 2003). It implies to landslide occurrence on a steep slope (Mugagga et al., 2012).

In South Sulawesi Indonesia, especially in Ujung-Loe upper watershed, LUC has been translated into numerous landslide incidents triggered by the intensity of rainfall compared to other factors such as earthquakes. The topography is extremely steep and naturally mountainous (38.8% class slope >20 degrees). It has a very high level of instability, especially during the rainy season. The annual rainfall can reach 1,436 to 5052 mm/year with average annual rainfall of 3,739 mm/year. The main occupation of social community in that area is farmer. Most of them live and do their activity in mountainous area. Avoiding this agricultural practice is hard. It has become people's culture for agriculture in mountainous regions, and they have made it hereditary (Soma and Kubota, 2017).

Landslide susceptibility was defined as quantitative or qualitative assessment classification, volume (or area), and the spatial distribution of landslides or potentially may occur in the zone. Susceptibility can also include a description of the speed and intensity of existing or potential landslides (Fell et al., 2008). Using scientific analysis of landslides, we can assess and predict landslide susceptibility and decrease landslide damage through proper preparation (Lee et al., 2002).

A few studies have evaluated land use change (LUC) that contributed to landslide occurrence (García-Ruiz et al., 2010; Glade, 2003; Mugagga et al., 2012; Soma and Kubota,

2017). However, the use of LUC as a causative factor to see the performance of LUC and to build landslide susceptibility has not yet been implemented. Therefore, the LUC will be as a new causative factor to change a land use as a human factor. Land use is implemented merely to look at the current time and it is different with the LUC. LUC can prepare what land has used before. For example, first case, LUC was from primary forests to farmland, and second case, LUC was from open area to farmland, from these cases have different slope stability. Based on these factors, previous researchers have not drawn the performance of land use changes as a causative factor. The objective of the study was to examine the performance of land use change as a causative factor to produce landslide susceptibility map using frequency ratio, and logistic regression and comparison each models.

4. 2 Data and Methods

4. 2. 1 Preparation of data

Data selection is the crucial thing in the preparation of the landslide susceptibility map (LSM). The excellent data selection for analysis helps to find satisfactory results. Management and collection or selection using Arc GIS© 10.3 must be accurate in establishing a spatial data landslide inventory and also a causative factor. The analysis of the frequency ratio (FR) calculation was carried out using Microsoft Excel©, while the logistic regression (LR) used the program Statistical Package for Social Sciences (SPSS©). More detail of the research, it is shown in Figure 17.

4. 2. 2 Landslide inventory

Landslides inventory can involve field surveys, expression of morphological, and interpretation of remote sensing images based on spectral characteristics, shape, and contrast (Kanungo et al., 2006). This study used data landslide from 2012 to 2016 using air photography of Google Earth Pro© and ground survey (Figure 18). The purpose was to find a correlation

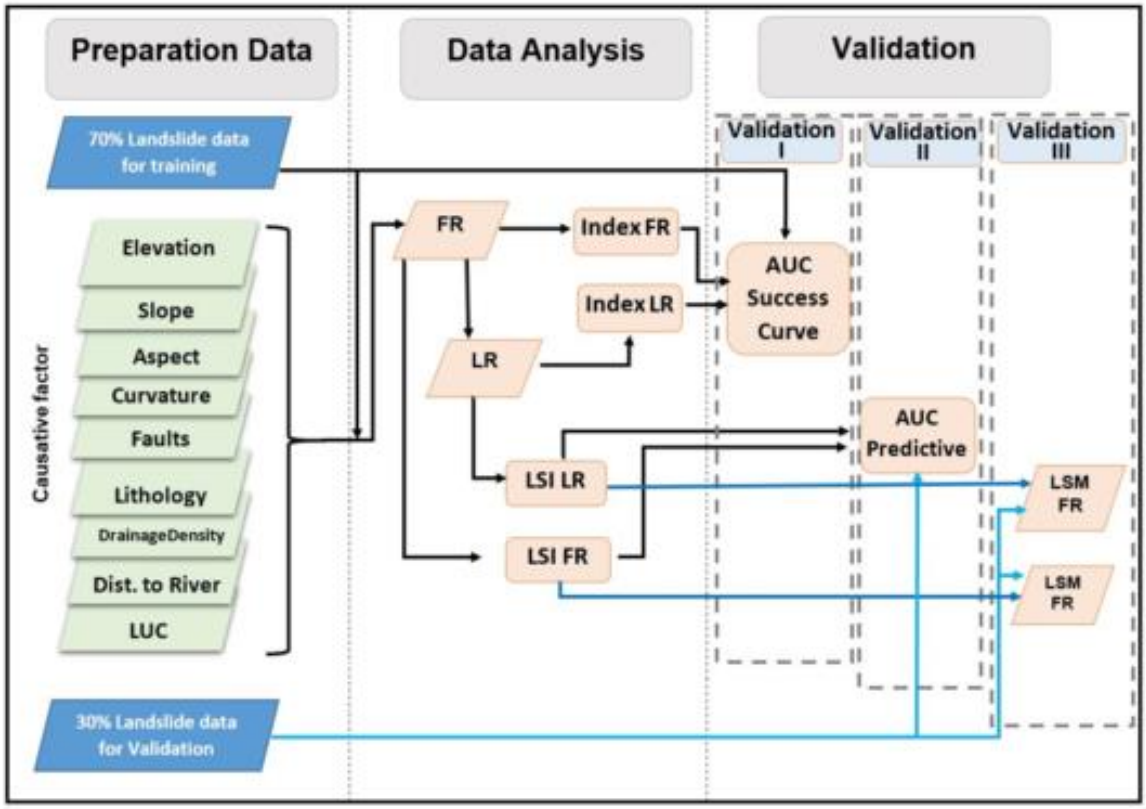


Figure 17 Research framework

between the occurrence of landslides and land use change from 2004 to 2011. The study area was limited to the upper of Ujung-Loe Watersheds. A total of 188 landslides were identified covering an area of 43.65 hectares (0.44 km^2). Most of the landslides are of the shallow type with minimum and maximum landslide area of 137 m^2 and 15,600 m^2, respectively. Using the landslide data from the survey and digitizing high-resolution from Google Earth Pro© to Arc GIS© 10.3, we digitized the time series imaging data by image interpretation landslide, and these files were saved as GIS compatible format as extension kml. Then, the data was again subsequently changed into shapefile and raster format 10x10 meter. Figure 19 shows the location of all landslide data divided into two group, i.e., landslide for training at 2,873 pixels (70%) and a landslide for validation at 1,230 pixels (30%).

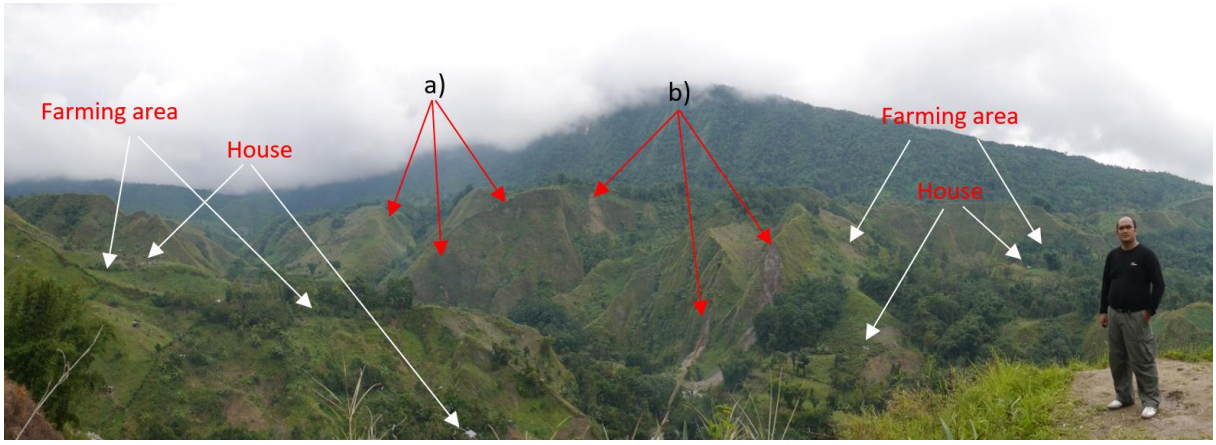


Figure 18 Landslide inventory a) old landslide, b) new landslide

4. 2. 3 Landslide causative factors

In susceptibility map, the most critical assumption that the incidence of landslides will occur in the same condition is affected by the cause of the landslides that have been occurred. There are no strict guidelines for the selection of causative factors to be used in logistic regression analysis and certainty factor, and as such, the selected covariates vary widely between studies (Ayalew and Yamagishi, 2005; Dou et al., 2015). Correspondingly, the determination of landslide causative factors was associated with the availability of data. Therefore, we selected causative factors based on the general knowledge found in previous studies and its availability in the target location. The entire landslide causative factors have been used for the independent variable in the landslide susceptibility mapping (Figure 20). The independent variable was nine causative factors including elevation, slope, aspect, curvature, lithology, distance from fault, distance to river, drainage density, and land use change (LUC).

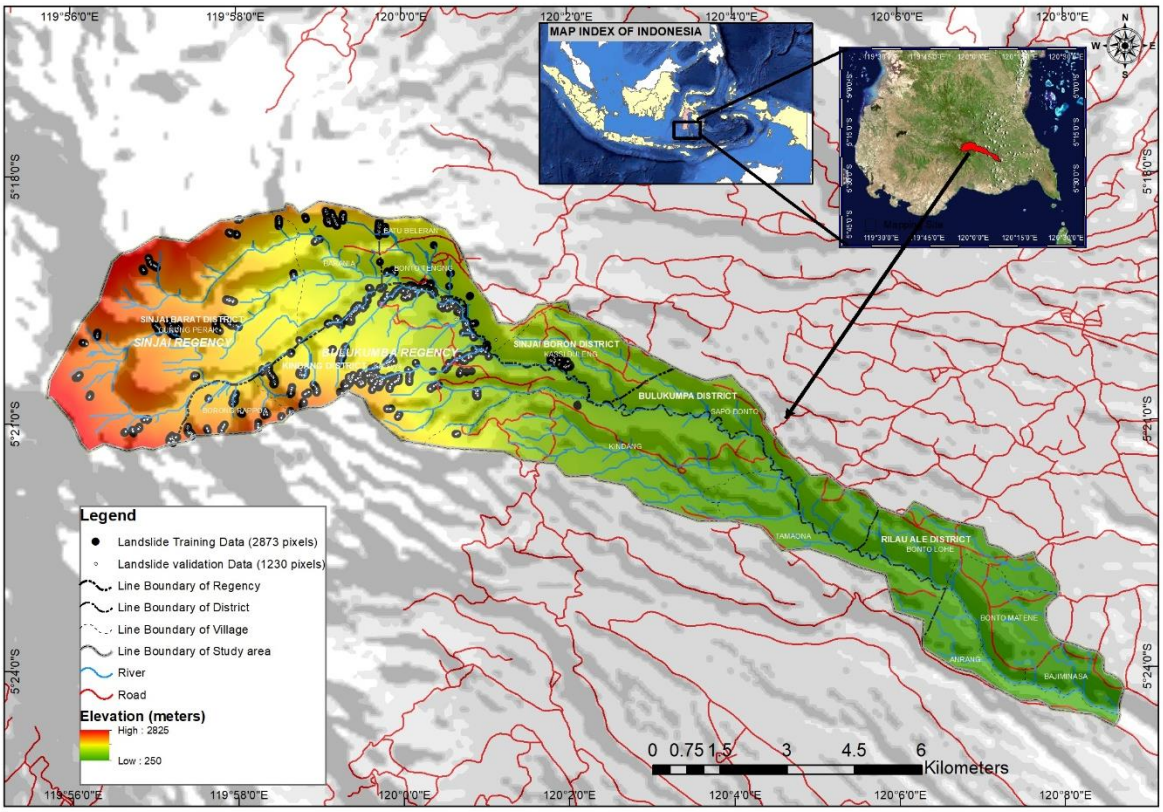


Figure 19 Map of Landslide Occurrence

Elevation, slope, aspect, and curvature were extracted from contour data with digital interval 12.5 meters. Contour data of Rupa Bumi Indonesia (RBI) on map scale 1: 25000 from Badan Geospatial Indonesia (BIG) was obtained using arc toolbox raster surface in ArcGIS 10.3 and elevation, slope, aspect, and curvature were extracted from contour. By using the uniform isotropic material, increased slope correlates with increased likelihood of failure. In this study, we have used six slope categories (0–10°, 10–20°, 20–30°, 30–40°, 40–50°, and above 50°) which were considered and represented in the form of slope thematic data layer. Likewise, the aspect map plays a significant role in slope stability assessment (Chauhan et al., 2010). In this study, aspect was divided into nine classes namely, flat, North, Northeast, East, Southeast, South, Southwest, West, and Northwest. Curvature was classified using the curvature of the profiles into three categories: concave, flat and convex. In the case of profile curvature, it was related to the puddle condition after heavy rain. Moreover, the reason is that, following heavy

rainfall, a more upwardly concave or convex slope has more water and retains it longer (Lee and Lee, 2006).

The geology of the area was obtained digital data produced by Indonesia Government, namely Geology Map by Geological Research Institute, at a scale of 1:250000. This map includes the current study area. The geology data consist of lithology, structure (fault or lineament), and rock type. Lithology is the primary data or parameters for analysis of the landslide map. Lithology is a standard variable that controls the landslide danger. It related to the strength of the material because lithological composition and structure vary for different types of rocks (Kanungo et al., 2006). In addition, resistance to the driving force depends on the strength of rocks and stones that will be more resistant. Faults are structural features, which describes the zones/areas of weakness, fractures, and lineament going higher susceptibility to landslides. It has been observed that the increased probability of landslide occurrence in a location close to faults not only affects the surface structure of the material but equally contributes to the permeability and cause slope instability. For this purpose, the distance from faults was used to analyze the incidence of landslides at a distance of faults. The proximity of the fault was obtained by buffering the map of faults (Rasyid et al., 2016).

Both drainage lines and landslide occurrence in the hilly area had a strong association due to erosional activity in this location. Buffering analysis of streamlines has calculated the distance from the river. This information was derived from a topographic map of scale 1:25000 called Peta Rupa Bumi Indonesia ((RBI) prepared by Badan Informasi Geospasial (BIG) Indonesia at 2012. The class of distance to river was grouped in five classes i.e. 0 – 100 m, 100 – 200 m , 200 – 300 m, 300 – 400 m and >400m. Similarly, drainage density was calculated using Arc Toolbox kernel density in km/km^2. The class of drainage density was grouped in five classes i.e. 0 – 1 km/km^2, 1 – 2 km/km^2 , 2 - 3 km/km^2 , 3 – 4 km/km^2 and >4 km/km^2.

Besides topographic factors and geology, land use (cover) is an essential element/factor responsible for landslide occurrences. The incidence of the landslide is inversely related to the

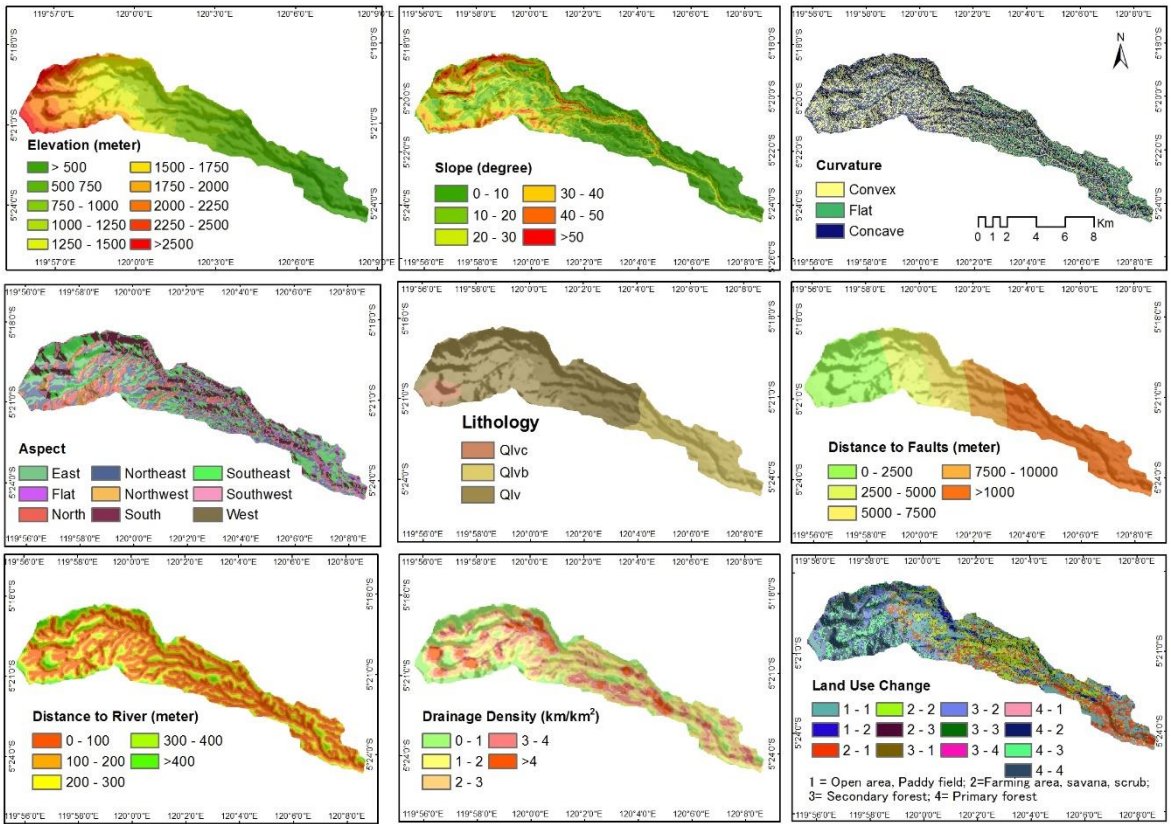


Figure 20 Eleven causative factor of landslide

vegetation density. This research used land use change (LUC) as vegetation density. LUC was as a new causative factor to change land use pattern to build a landslide susceptibility mapping. To the critical slope, LUC triggered a series of shallow and profound landslides (Mugagga et al., 2012). The LUC map was derived from overlaying land use 2004 and land use 2011. Land use was derived from interpretation Landsat 5 TM© (date recorded on September, 21th 2004) and Landsat 7 ETM+© (date recorded on October, 11th 2011) images, each with a 30 m resolution, collected from United States Geological Survey (USGS). The unsupervised classification method was applied to classify land use. Unsupervised classification consists of three steps: (1) the map creation of N spectral class using Iterative Self-Organizing Data Analysis Technique Algorithm, (2) the development of land use (LU) map with the help of reference data, and (3) the accuracy measurement of the ratings of all LU reference map using independent data and selection of a map LU with the highest accuracy (Lang et al., 2008). This method was applied to classify land use into seven such as Soma and Kubota (2017) types:

open area, paddy field, farming area, scrub, savanna, secondary forest and primary forest. Overall accuracy values of LU 2004 and LU 2011 were 86% and 90%, respectively. Kappa values of 0.83 and 0.88 were achieved for the unsupervised classified maps of LU 2004 and LU 2011, respectively. Moreover, LUC was built by classifying once more LU 2004 and 2011 in four classes: one (open area, paddy field), two (farming area and shrub, savanna), three (secondary forest) and fourth (primary forest). In the next step, each other was overlaid using ARC GIS 10.3 and were founded 13 classes as LUC. They were 1 – 1 (from open area and paddy field to open area and paddy field) , 1 – 2 (from open area and paddy field to farming area and scrub, savanna), 2 – 1(from farming area and scrub, savanna to open area and paddy field), 2 – 2 (from farming area and scrub, savanna to farming area and scrub, savanna), 2 – 3 (from farming area and scrub, savanna to secondary forest), 3 – 1 (from secondary forest to open area and paddy field), 3 – 2 (from secondary forest to farming area and scrub, savanna), 3 – 3(from secondary forest to secondary forest), 3 – 4(from secondary forest to primary forest), 4 – 1 (from primary forest to open area and paddy field), 4 – 2, 4 – 3 (from primary forest to secondary forest), and 4 - 4 (from primary forest to secondary forest), and 4 - 4 (from primary forest to primary forest). LUC was downgraded from pixel size of 30 x 30-meter to pixel size of 10 x 10 meter.

Landslide was described as the dependent variable, and causative factor, i.e., elevation, slope, aspect, curvature, distance to river, drainage density, lithology, distance to faults and LUC were described as the independent variables (Figure 20). Independent and dependent variables were used as a map input and then processed to turn it into a raster map with a pixel size of 10 m × 10 m. The study area included 795,227 pixels and the landslide data used in the model included 2,873 pixels (70% of Landslide) and 1,230 pixels (30%) for validation.

4. 2. 4 Data Analysis

There are two analyses methods to understand the performance of each causative factor (frequency ratio, and logistic regression) to produce landslide susceptibility map. Frequency ratio analysis was implemented to define the performance of class of each causative factor and logistic regression methods could describe the performance of each causative factor to the susceptibility of landslide occurrence.

4.2.4.1 Frequency Ratio

The landslide and the causes were related, and it can be concluded areas of the landslide occurs with the causative factors of landslides. Simple statistical method to determine the closeness of the relationship have been applied to the frequency ratio (FR) approach. FR for each causative factor was calculated by dividing the landslide occurrence rate by the area ratio. If the ratio is more significant than 1.0, the relationship between the landslide and the causative factor is higher, and, if the relationship is less than 1, the connection is low (Lee and Lee, 2006). A ratio value in each class shows the level of relationship and the frequency ratio value is calculated by the following Equation 4.1,

$$FR = \frac{\mathrm{PxcL(nm)}/\sum \mathrm{PnxL}}{\mathrm{Pixel(nm)}/\sum \mathrm{Pnx}} \tag{4.1}$$

where, Pxcl(nm) number of pixel with landslide within class n of j parameter, Pixel(nm) Number of pixel in class n of m parameter, ΣPnxL total pixel of m parameter, and ΣPnx whole pixel of the area).

To create an index of landslides susceptibility, all causative factors in the form of raster maps of the value FR then summed by using Equation 4.2.,

$$\mathrm{LSI} = \mathrm{FR1} + \mathrm{FR2} + \ldots + \mathrm{FRn} \tag{4.2}$$

where FR1, FR2, FR3… FRn is the frequency ratio raster maps of landslide causative factors.

4.2.4.2 Logistic regression

Logistic regression resulted in landslide susceptibility index. A simple introduction to logistic regression is available in Chau and Chan (2005) which defines the probability occurrence of landslides divided by the probability of no occurrence of landslides. It is useful to predict the presence or absence of a characteristic or outcome based on values of a set of variable predictors. Generally, in the logistic regression, spatial prediction can be modeled using the independent and the dependent variables (Shirzadi et al., 2012). It is useful when the variable is a binary or dichotomous. Variables can be continuous, or discrete, or a combination of the two types and they do not always have a normal distribution. The probability of regression can be understood as the possibility of state-dependent variables. Data analysis created iteration in ten tests using same proportion data of landslide and no landslide. Using an equal proportion of data occurrence of landslide and no landslide will result better and fair for logistic regression analysis (Rasyid et al., 2016). They were restricted to fall within a range of values from 0 to 1 (Xu et al., 2013). The value of zero shows the probability of 0% landslide occurrences, and one shows a 100% probability (Dai et al., 2004). The logistic regression followed on logistic function –z expressed by the following Equation 4.3.,

$$P = \frac{1}{1+\exp^{-Z}} \tag{4.3}$$

$$Z = C_0 + C_1CF_1 + C_2CF_2 + \ldots + C_nCF_n \tag{4.4}$$

where: P is the probability of landslide occurrence that estimated values vary from 0 to 1. Variable Z is landslide causative factors and assumed as a linear combination of the causative factors i (i = 1, 2,…n). Moreover, Z is calculated by using Equation 4.4. C0 is the intercept, and C1, C2, ., Cn are coefficients, which measure the contribution of independent factors (CF1, CF2, . . ., CFn) to the variations in Z.

4. 2. 5 Validation and verification

In addition to increase the prediction of accuracy and probability, validation can improve the reliability. During the modeling predictions, the most essential and critical component is to carry out the validation of the results of prediction (Chung and Fabbri, 2003). In this study, the landslide inventories were divided into two parts; one for training and the other for validation. This study used a 2,873 pixel (70%) inventories landslides to produce models and 1,230 (30%) of pixels for validation. The assumptions to selection landslide of data for training and validation of the model were randomly taken on each part of landslide occurrence in the area of research and also based on the representation of the landslide area. By illustrating the procedure, a small portion of the landslide-prone areas was selected as the data for validation. Size, area, depth of landslides and distribution significantly varies from place to place. Also, we used the ROC curve to plot predicted probabilities in order to understand the problem of accuracy, selection criteria, and interpretation. For validating the landslide susceptibility map, AUC curve was used as a measure of overall fit and comparison of modeled prediction. The area determines the success rate under the curve (AUC) of the training dataset, and predictable level calculated from the AUC of the validation dataset. ROC curves were used to evaluate the predictive accuracy of the model selected in the statistical approach, such as logistic regression (Gorsevski et al., 2006). The AUC obtained from the ROC plot statistics is the most preferred type that can influence rating (Akgun et al., 2012). Predicted probabilities generated by logistic regression can be seen as an indicator to be continuously compared with a binary response variable observed. In this study, the validation process further demonstrates the level of accuracy of landslide susceptibility map to calculate the ratio of the data for validation of landslides that fall into each class of vulnerability. It was assumed that most of the landslides for validation must occur on a high to very higher susceptibility class (H + VH).

Table 9 The value of Frequency Ratio and Certainty Factor for each landslide causative factors

Factor	Class	Pixel Class*	% Class	Landslide pixel**	% Landslide	Frequency Ratio
Elevation (meter)	<500	126010	15.85	0	0.00	0.00
	500 – 750	113821	14.31	0	0.00	0.00
	750 – 1000	117886	14.82	382	13.30	0.90
	1000 – 1250	99735	12.54	544	18.93	1.51
	1250 – 1500	80401	10.11	446	15.52	1.54
	1500 – 1750	73551	9.25	665	23.15	2.50
	1750 – 2000	62583	7.87	452	15.73	2.00
	2000 - 2250	63418	7.97	202	7.03	0.88
	2250 – 2500	38830	4.88	180	6.27	1.28
	>2500	18992	2.39	2	0.07	0.03
Slope (degree)	0 -10	277391	34.88	233	8.11	0.23
	10 – 20	193490	24.33	504	17.54	0.72
	20 – 30	142736	17.95	519	18.06	1.01
	30 – 40	114954	14.46	613	21.34	1.48
	40 – 50	56795	7.14	819	28.51	3.99
	>50	9861	1.24	185	6.44	5.19
Curvature	Concave	335269	42.16	1,614	56.18	1.33
	Flat	100826	12.68	158	5.50	0.43
	Convex	359132	45.16	1,101	38.32	0.85
Aspect	Flat	48980	6.16	43	1.50	0.24
	North	105139	13.22	963	33.52	2.54
	Northeast	140313	17.64	599	20.85	1.18
	East	128555	16.17	311	10.82	0.67
	Southeast	155292	19.53	191	6.65	0.34
	South	127354	16.01	515	17.93	1.12
	Southwest	48881	6.15	43	1.50	0.24
	West	11324	1.42	23	0.80	0.56
	Northwest	29389	3.70	185	6.44	1.74
Lithology	Qlvb	195818	24.62	0	0.00	0.00
	Qlv	562441	70.73	2,826	98.36	1.39
	Qvlc	36968	4.65	47	1.64	0.35
Distance to Faults (meter)	0 – 2500	228372	28.72	913	31.78	1.11
	2500 -5000	123498	15.53	1,333	46.40	2.99
	5000 – 7500	106243	13.36	472	16.43	1.23
	7500 – 10000	92127	11.58	155	5.40	0.47
	>10000	244987	30.81	0	0.00	0.00
Distance to River (meter)	0 - 100	325991	40.99	1,489	51.83	1.26
	100 – 200	240871	30.29	726	25.27	0.83
	200 – 300	139539	17.55	397	13.82	0.79
	300 – 400	59549	7.49	189	6.58	0.88
	400 – 500	19942	2.51	42	1.46	0.58
	>500	9335	1.17	30	1.04	0.89
Drainage Density (km/km^2)	0 – 1	147677	18.57	698	24.30	1.31
	1 - 2	228100	28.68	635	22.10	0.77
	2 - 3	252005	31.69	829	28.85	0.91
	3 - 4	121676	15.30	512	17.82	1.16
	>4	45769	5.76	199	6.93	1.20
LUC (1=Open area, Paddy area; 2=Farming area, savanna, scrub; 3=Secondary Forest; 4=Primary Forest)	1 - 1	167966	21.12	608	21.16	1.00
	1 – 2	44883	5.64	276	9.61	1.70
	2 – 1	127015	15.97	134	4.66	0.29
	2 - 2	140425	17.66	215	7.48	0.42
	2 - 3	3971	0.50	2	0.07	0.14
	3 – 1	24542	3.09	157	5.46	1.77
	3 - 2	88061	11.07	513	17.86	1.61
	3 - 3	30715	3.86	158	5.50	1.42
	3 – 4	4602	0.58	26	0.90	1.56
	4 – 1	954	0.12	30	1.04	8.70
	4 – 2	19912	2.50	177	6.16	2.46
	4 – 3	55800	7.02	180	6.27	0.89
	4 – 4	86381	10.86	397	13.82	1.27

*Total pixel area 795,227 **Landslide Training 2,873

4. 3 Result and Discussion

4. 3. 1 Frequency ratio

Table 9 indicates a correlation between landslide occurrence and each class of landslide causative factors. In the case of the relationship between landslide occurrence and LUC, class of primary forest to open area and paddy field (4-1) had the highest probability of landslide occurrence with frequency ratio 8.70. Moreover, class of secondary forest to the farming area, savanna, scrub (4-2) had a frequency ratio 2.20. The vegetation which causes this frequency ratio affects the stability of the slope. Land with forest having the root system would reinforce the soil strength and stabilize the slope (Kubota et al., 2015). Forest clearance seems to have manifested primarily through increased rates of landslide activity (Glade, 2003).

In slope class, slope above 20° has a ratio of >1 which indicates a high probability of landslide occurrence. Moreover, slope below 20° has a ratio of <1, which shows a very low probability of landslide occurrence.

In class of elevation, the values between 1000 to 2000 meters (m) have indicated a high degree of likelihood of the landslide occurrence. In the class of curvature, the concave class has a higher probability of landslide occurrence with ratio value >1. In the case of the class aspect, north, northwest, south and northeast-facing slopes have frequency ratio > 1, which shows a high rate of probability of the landslides occurrence.

In the case of lithology, Quarter lompobattang volcanic (Qlv) has a ratio of >1, which indicates a high probability of landslide occurrence. Qlv is one of the volcanic and sediment formations in South Sulawesi.

Causative factor, i.e., distance to fault and rivers, the ratio of the distance/proximity is used to understand the degree of influence on the landslide. Distance to faults below 7500 m has a ratio > 1. It shows that more close distance to the fault, the probability of landslide occurrence will

increase. Similarly, the distance to the river below 100 m has frequency ratio > 1. It indicates the probability of landslide will increase if the distance to the river is nearer.

To create an index of susceptibility to landslides, all causative factors were mapped in the form of raster maps of the value FR then summed using Equation 4.2. The index value of frequency ratio of with LUC was in the range of 2.70 to 25.41 and without LUC was in the range of 2.46 to 17.97. The higher landslide susceptibility index (LSI) value showed higher susceptibility to landslides. The results showed that LUC change creates a higher value than without LUC meaning that LUC is better to predict of landslide occurrence (Table 13).

4. 3. 2 Logistic regression

Hence, this study proposes to investigate ten tests to acquire best result and sense of fairness as shown in Table 10 and Table 11. LUC as a new causative factor for landslide had a value of 0.589 (number test seventh) that affects landslide occurrence. Forest land with root system would reinforce the soil strength and stabilize the slope to reduce surface erosion or shallow landslides (Kubota et al., 2015). The highest value of 3.081 shows the distance to the river having the most significant effect on landslide occurrence. Moreover, the lowest value of elevation (0.353) indicated a small effect on landslide occurrence in this research.

Table 10 Logistic regression coefficient of landslide causative factors using an equal proportion of landslide and a non-landslide pixel with LUC causative factor

Number Test	Variable in the equation									
	Elevation	Slope	Aspect	Curvature	Lithology	Distance to Faults	Distance to River	Drainage Density	LUC	Constant
1	0.344	0.553	0.576	0.659	1.681	0.400	3.044	1.214	0.512	-10.496
2	0.353	0.562	0.548	0.534	1.696	0.476	3.081	0.995	0.551	-10.355
3	0.274	0.475	0.624	0.605	1.631	0.473	2.630	1.184	0.518	-9.882
4	0.302	0.533	0.561	0.452	1.612	0.439	2.817	0.703	0.425	-9.304
5	0.335	0.532	0.590	0.391	1.684	0.459	2.934	1.175	0.521	-10.136
6	0.307	0.535	0.627	0.376	1.722	0.371	2.914	0.907	0.497	-9.741
7	0.245	0.571	0.551	0.524	1.682	0.448	2.818	0.897	0.481	-9.693
8	0.317	0.511	0.525	0.388	1.572	0.506	2.986	0.814	0.539	-9.638
9	0.400	0.552	0.498	0.430	1.541	0.445	3.119	0.597	0.378	-9.421
10	0.348	0.563	0.498	0.329	1.723	0.446	2.985	0.803	0.355	-9.530

Table 11 Table 3 Logistic regression coefficient of landslide causative factors using an equal proportion of landslide and non-landslide pixel without LUC causative factor

Number Test	Variable in the equation								
	Elevation	Slope	Aspect	Curvature	Lithology	Distance to Faults	Distance to River	Drainage Density	Constant
1	0.486	0.587	0.554	0.671	1.618	0.376	2.983	1.317	-10.047
2	0.509	0.600	0.533	0.537	1.643	0.442	3.026	1.130	-9.927
3	0.409	0.509	0.603	0.635	1.566	0.448	2.573	1.325	-9.479
4	0.417	0.562	0.544	0.472	1.578	0.419	2.791	0.814	-9.014
5	0.471	0.568	0.576	0.406	1.617	0.428	2.902	1.317	-9.740
6	0.448	0.571	0.611	0.393	1.669	0.342	2.892	1.031	-9.397
7	0.372	0.605	0.533	0.538	1.647	0.422	2.793	1.032	-9.368
8	0.464	0.546	0.504	0.411	1.503	0.481	2.916	0.956	-9.200
9	0.504	0.579	0.485	0.446	1.517	0.422	3.090	0.698	-9.160
10	0.453	0.591	0.491	0.341	1.680	0.424	2.985	0.931	-9.339

4. 3. 3 Validation

Table 12 and Figure 21 show results of AUC curve for both success rate and predictive rate for each test. Some landslide and non-landslide pixels were used to obtain AUC success and predictive rate.

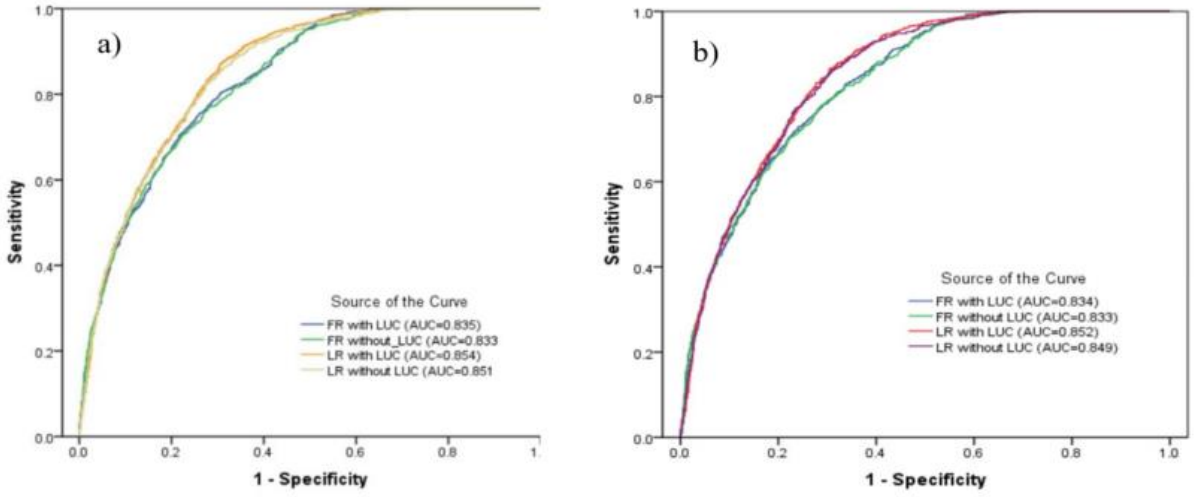


Figure 21 AUC of ROC of landslide susceptibility with and without LUC causative factor using FR, and LR method; a) success rate and b) predictive rate

Table 12 AUC of ROC curve of success and predictive rate and ratio of landslide validation on landslide susceptibility map using FR, and LR Method

	Method	FR	Number Test of LR									
			1	2	3	4	5	6	7	8	9	10
With LUC:	AUC Success rate	0.835	0.854	0.854	0.853	0.854	0.854	0.854	0.854	0.854	0.854	0.854
	AUC Predictive rate	0.834	0.852	0.852	0.851	0.852	0.852	0.852	0.852	0.852	0.851	0.852
	H+VH (%)	79.35	80.08	80.24	78.70	79.43	79.92	80.00	80.00	79.19	79.67	79.76
Without LUC:	AUC Success rate	0.833	0.850	0.851	0.850	0.,851	0.850	0.850	0.851	0.850	0.850	0.851
	AUC Predictive rate	0.833	0.848	0.849	0.848	0.848	0.848	0.848	0.849	0.848	0.848	0.849
	H+VH (%)	78.46	77.97	77.56	77.24	78.38	77.40	77.97	79.19	77.07	79.19	78.94

In general, the AUC of ROC curves representing excellent, good, and valueless tests were plotted on the graph. It classifies the accuracy of a diagnostic test i.e. the value ranges from 0.50 to 0.60 (fail), 0.60–0.70 (poor), 0.70–0.80 (fair), 0.80–0.90 (good), and 0.90–1.00 (excellent) (Rasyid et al., 2016). The results showed that the entire test of FR and LR methods both of with and without LUC are included in the good category. The value ranged from 0.833 to 0.854 in success rate and 0.833 to 0.852 in predictive rate, respectively (Table 12 and Figure 21). Moreover, success rate and predictive rate value for all methods were near to the interval of 0.02 indicating that all the methods were more reliable to a predictive landslide in the future. The proximity of success rate and predictive rate values show how the method helps in landslide prediction in the future (Meten et al., 2015).

In this study, LR method conducted one more validation to choose the best statistical model for creating landslide susceptibility map and the best equation in logistic regression approach from the ten tests. The sum of FR value and equation of the LR models were used to create landslide susceptibility map (LSM). All LSM classes were created by reclassifying LSI of the models using natural breaks method. Overlaid landslide data validation on LSM described another level of accuracy besides AUC curve. The natural breaks method or Jenks optimization method has been widely used mainly by planners. It is designed to determine the best arrangement of values into different classes. This approach maximizes the variance between classes and reduces the variation within classes. The description of landslide susceptibility level on location was grouped into five categories, namely very low, low, medium, high and very high. A Landslide susceptibility map verified the accuracy of landslide susceptibility map by overlaying it with 30% of landslide data validation. Validation on LSM for the LR model was better than FR model, and causative factor with LUC was better than without LUC (Figure 21). Validation of FR method with LUC (0.835) in success rate value had slightly higher accuracy than without LUC (0.833). Similarly, the LR method LUC (0.854) had slightly higher accuracy than without LUC (0.851). These show that the FR and LR model

with LUC are useful model for identifying landslide as opposed to the model without LUC. In the case of AUC curve for predictive rate, FR method with LUC (0.834) value had slightly higher accuracy than without LUC (0.833), and LR method LUC (0.852) had slightly higher than without LUC (0.849). Both of FR and LR model with LUC is better without LUC. The curve of the model and validation proves that the susceptibility model is acceptable and the model could be applied to predict the potential landslides in the future. As an interesting point to be noticed in Table 12, the seventh tests for LR had a good result in AUC curve, which is 0.857 in success rate and 0.856 in predictive rate, respectively.

Figure 22 shows the landslide susceptibility map with and without LUC causative factor using FR, and the second test equation of LR model with LUC (Table 10) and seventh test equation of LR without LUC (Table 11). The LSM by LR model with LUC was obtained using the coefficient values of landslide causative factors as in the equation below;

Z = -10.355 + 0.353 Elevation + 0.562 Slope + 0.548 Aspect + 0.534 Curvature + 1.696 Lithology+ 0.476 faults + 3.081 Distance to River + 0.995 Drainage density + 0.551 LUC

The LSM by LR model without LUC was obtained using the coefficient values of landslide causative factors as in the equation below;

Z = -9.368 + 0.372 Elevation + 0.605 Slope + 0.533 Aspect + 0.538 Curvature + 1.647 Lithology+ 0.422 distance to faults + 2.793 Distance to River + 1.032 Drainage density

The ranges of the index value of each model in five categories were established using natural breaks method. Can et al. (2005) and Bai et al. (2010) stated two crucial guidance for validating landslide susceptibility map, i.e. (1) the high to very high classes should cover only small areas and 2) landslide data validation should lie in high or very high classes.

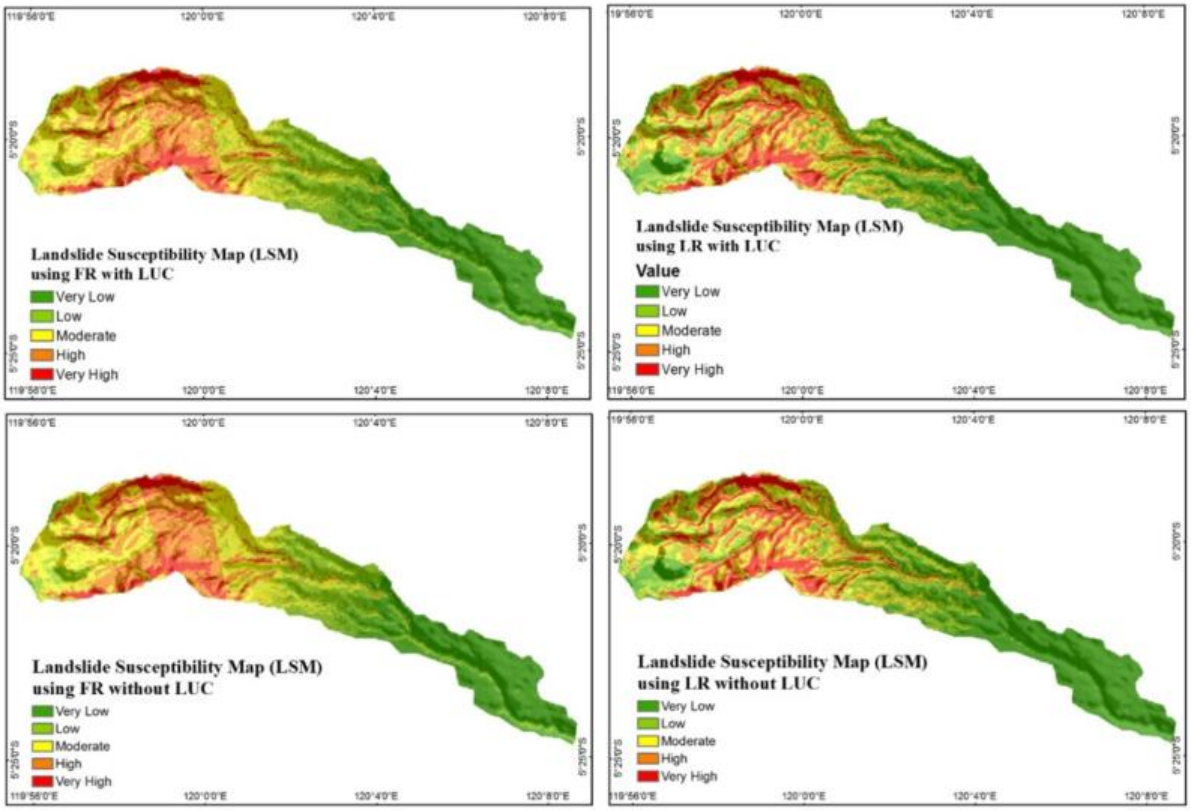


Figure 22 Landslide susceptibility map of with and without LUC causative factors using FR, and LR method

Table 13 The characteristic of susceptibility classes on landslide susceptibility map using FR, and LR method with and without LUC causative factor

Class Number	Reclassified index value	Vulnerability class	Number of pixels	% area covered	Number of landslide validation pixel	% area of landslide validation covered
Frequency Ratio With LUC						
1	2.700 - 5.906	Very Low	187187	23.54	0	0.00
2	5.906 - 8.578	Low	153294	19.28	19	1.54
3	8.578 - 10.893	Moderate	208585	26.23	235	19.11
4	10.893 - 13.476	High	181945	22.88	452	36.75
5	13.476 - 25.410	Very High	64216	8.08	524	42.60
Logistic Regression with LUC						
1	0.0019 - 0.1231	Very Low	292856	36.83	3	0.24
2	0.1231 - 0.3069	Low	162173	20.39	62	5.04
3	0.3069 - 0.5025	Moderate	133513	16.79	178	14.47
4	0.5025 - 0.7137	High	113755	14.30	331	26.91
5	0.7137 - 0.9992	Very High	92930	11.69	656	53.33
Frequency Ratio Without LUC						
1	2.460 - 5.1971	Very Low	188730	23.73	0	0.00
2	5.1971 - 7.630	Low	150267	18.90	24	1.95
3	7.630 - 9.820	Moderate	220097	27.68	241	19.59
4	9.820 - 12.253	High	178338	22.43	466	37.89
5	12.253 - 17.970	Very High	57795	7.27	499	40.57
Logistic Regression without LUC						
1	0.0027 - 0.1228	Very Low	272313	34.24	4	0.33
2	0.1228 - 0.3126	Low	171069	21.51	60	4.88
3	0.3126 - 0.5063	Moderate	148090	18.62	192	15.61
4	0.5063 - 0.7116	High	112403	14.13	322	26.18
5	0.7116 - 0.9904	Very High	91352	11.49	652	53.01

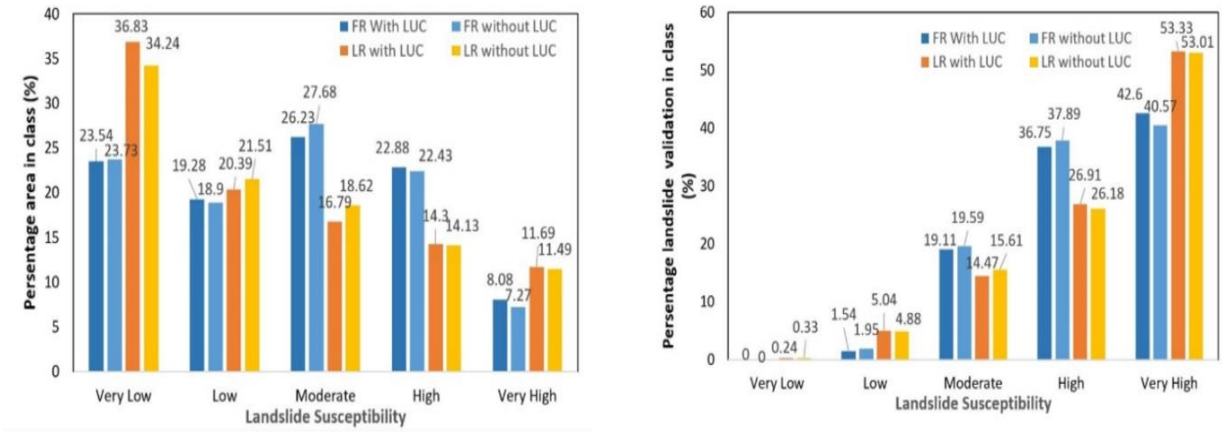


Figure 23 Percentage of landslide susceptibility classes and rate of landslide susceptibility validation on landslide susceptibility of FR, and LR method

Table 13 shows the characteristics of susceptibility class for FR and LR models with and without LUC causative factors. It indicates that the ratio of high to very high vulnerability classes covers a small area or less than 32% for FR and less than 26% of the total area for LR. The data validation landslide included in the class shows the ratio below 10%. The accuracy of the predicted future landslide from the LSM should have a lower ratio in low to very low classes and higher in the high to very high classes (Rasyid et al., 2016). Figure 23 shows high to very high vulnerability classes for LR with LUC (80.24%) having a higher value for validation than LR without LUC (79.19%). Moreover, FR with LUC (79.35%) had a higher value than FR without LUC (78.46%). It indicates that the performance of LUC as a causative factor both using FR and LR model gives a good result. Taken together, the results suggested that changing the vegetation to another landscape causes slopes unstable and increases the probability of landslide occurrence.

4. 4 Conclusion

In conclusion, land use change (LUC) showed a good demonstration as a new causative factor to build landslide susceptibility map. The result indicated that LUC has the effect to produce LSM. Validation of landslide susceptibility was carried out in this study at both with

and without LUC causative factors. Firstly, performances of each landslide model were tested using AUC curve for success and predictive rate. The highest value of predictive rate was at with LUC in both FR and LR methods (83.4% and 85.2%, respectively). Secondly, the ratio of landslides on high to very high classes of susceptibility was obtained, which indicates the accuracy level of the method. LR method with LUC had the highest accuracy of 80.24 %. These results suggested that changing the vegetation to another landscape causes slopes unstable and increases the probability of landslide occurrence.

4. 5 References

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C., Pradhan, B., 2012. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput. Geosci. 38, 23–34. doi:10.1016/j.cageo.2011.04.012

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains, Central Japan 65, 15–31. doi:10.1016/j.geomorph.2004.06.010

Agency of Climatology and Geophysical, Makasar, Indonesia, 2016. Rainfall Data.

Bai, S.B., Wang, J., Lü, G.N., Zhou, P.G., Hou, S.S., Xu, S.N., 2010. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. Geomorphology 115, 23–31. doi:10.1016/j.geomorph.2009.09.025

Can, T., Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., Duman, T.Y., 2005. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology 72, 250–271. doi:http://dx.doi.org/10.1016/j.geomorph.2005.05.011

Chau, K.T., Chan, J.E., 2005. Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. Landslides 2, 280–290. doi:10.1007/s10346-005-0024-x

Chauhan, S., Sharma, M., Arora, M.K., Gupta, N.K., 2010. Landslide susceptibility zonation through ratings derived from artificial neural network. Int. J. Appl. Earth Obs. Geoinf. 12, 340–350. doi:10.1016/j.jag.2010.04.006

Chung, C.-J.F., Fabbri, A.G., 2003. Validation of Spatial Prediction Models for Landslide Hazard Mapping. Nat. Hazards 30, 451–472. doi:10.1023/B:NHAZ.0000007172.62651.2b

Dai, F.C., Lee, C.F., Tham, L.G., Ng, K.C., Shum, W.L., 2004. Logistic regression modeling of storm-induced shallow landsliding in time and space on natural terrain of Lantau Island, Hong Kong. Bull. Eng. Geol. Environ. 63, 315–327. doi:10.1007/s10064-004-0245-6

Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One 10. doi:10.1371/journal.pone.0133262

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., Savage, W.Z., 2008. Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning. Eng. Geol. 102, 99–111. doi:10.1016/j.enggeo.2008.03.014

García-Ruiz, J.M., Beguería, S., Alatorre, L.C., Puigdefábregas, J., 2010. Land cover changes and shallow landsliding in the flysch sector of the Spanish Pyrenees. Geomorphology 124, 250–259. doi:10.1016/j.geomorph.2010.03.036

Glade, T., 2003. Landslide occurrence as a response to land use change: A review of evidence from New Zealand. Catena 51, 297–314. doi:10.1016/S0341-8162(02)00170-4

Gorsevski, P. V., Gessler, P.E., Boll, J., Elliot, W.J., Foltz, R.B., 2006. Spatially and temporally distributed modeling of landslide susceptibility. Geomorphology 80, 178–198. doi:10.1016/j.geomorph.2006.02.011

Hasnawir, Kubota, T., Sanchez-Castillo, L., Soma, A.S., 2017. The Influence of Land use change and rainfall on shallow landslide in Tanralili Sub-watwrshed, Indonesia. J. Fac. Agric. Kyushu Univ. 62, 171–176.

Kanungo, D.P., Arora, M.K., Sarkar, S., Gupta, R.P., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Eng. Geol. 85, 347–366. doi:10.1016/j.enggeo.2006.03.004

Kubota, T., Sanchez-castillo, L., Soma, A.S., 2015. ROOT STRENGTH OF UNDERSTORY VEGETATION FOR EROSION CONTROL ON.

Lang, R., Shao, G., Pijanowski, B.C., Farnsworth, R.L., 2008. Optimizing unsupervised classifications of remotely sensed imagery with a data-assisted labeling approach. Comput. Geosci. 34, 1877–1885. doi:10.1016/j.cageo.2007.10.011

Lee, S., Lee, M.J., 2006. Detecting landslide location using KOMPSAT 1 and its application to landslide-susceptibility mapping at the Gangneung area, Korea. Adv. Sp. Res. 38, 2261–2271. doi:10.1016/j.asr.2006.03.036

Meten, M., PrakashBhandary, N., Yatabe, R., 2015. Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia. Geoenvironmental Disasters 2, 1–17. doi:10.1186/s40677-015-0016-7

Mugagga, F., Kakembo, V., Buyinza, M., 2012. Land use changes on the slopes of Mount Elgon and the implications for the occurrence of landslides. Catena 90, 39–46. doi:10.1016/j.catena.2011.11.004

Rasyid, A.R., Bhandary, N.P., Yatabe, R., 2016. Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. Geoenvironmental Disasters 3, 19. doi:10.1186/s40677-016-0053-x

Shirzadi, A., Saro, L., Hyun Joo, O., Chapi, K., 2012. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. Nat. Hazards 64, 1639–1656. doi:10.1007/s11069-012-0321-3

Soma, A.S., Kubota, T., 2017. Land Use Changes on the Slopes and the Implications for the Landslide Occurrences in Ujung-Loe Watersheds South Sulawesi Indonesia. Int. J. Ecol. Dev. 32, 33–42.

Xu, C., Xu, X., Dai, F., Wu, Z., He, H., Shi, F., Wu, X., Xu, S., 2013. Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake in China. Nat. Hazards 68, 883–900. doi:10.1007/s11069-013-0661-7

.

Chapter 5 Comparative Study of Land Use Change and Landslide Susceptibility Using Frequency Ratio, Certainty Factor, and Logistic Regression in Upper Area of Ujung-Loe Watersheds

5. 1 Introduction

Landslide susceptibility is defined as a "quantitative or qualitative assessment of the classification, volume (or area), and spatial distribution of landslides which exist or potentially may occur in an area" (Fell et al., 2008). Using scientific analysis of landslides, we can assess and predict landslide susceptibility and decrease landslide damage through proper preparation (Lee et al., 2002).

Landslide susceptibility can be used to identify one or more landslide causes and triggers. The landslide causative factors are the reasons that a landslide occurred at that location and at that time. Several factors such as geomorphological, geological, hydrological and anthropogenic factors in addition to rainfall affect landslides occurrence. Geomorphology factors are elevation, slope, aspect, and curvature. Geology factors are lithology and structure (fault of lineament). Hydrology factors are river and density of the river. Human factor are land use and road construction.

A few studies have evaluated land use change (LUC) contribution to landslide (Glade, 2003; García-Ruiz et al., 2010; Mugagga et al., 2012). However, none of these researchers used LUC as a causative factor to build landslide susceptibility. Therefore, LUC will be used as a new causative factor to produce landslide susceptibility map. LUC is different with land use. Land use only looks at the current state, but it did not see the past land cover. LUC will not affect directly to landslide occurrence in one or two years but will be the effect after a few years. For example change the primary forest to farming area in steep slope, when the clear cutting the tree in primary forest to convert to farming area, in one or two year the stability slope still good, but after a few years when the roots were decay and make hole in soil, then the

rainfall fill the hole. It causes the slope unstable and landslide will occur. Kubota et al. (2007) point out that land with forest cover would reinforce the soil strength and stabilizes the slope due to its root system.

LUC is a process by which human activities transform the landscape. The latest intensification of land use changes has increased the level of susceptibility to landslides, especially in mountainous areas. LUC influences the occurrence of rainfall-induced landslides (Glade, 2003) and LUC can be related to landslide occurrence on a steep slope (Mugagga et al., 2012). LUC influences the occurrence of rainfall-induced landslides (Glade, 2003) and LUC can be related to landslide occurrence on a steep slope (Mugagga et al., 2012)

In South Sulawesi Indonesia, LUC has caused 31 landslide incidents from 2011 to 2015 triggered by the intensity of rainfall (BNPB Indonesia, 2016), especially in the upper area of Ujung-Loe watershed. The topography is naturally very steep and mountainous (38.8% slope class of >20 degrees) and has a very high level of instability, especially during the rainy season (Rainfall: 2,976 to 5,052g mm/year with average annual rainfall 3,965 mm/year; (Meteorology, Climatology, and Geophysical Agency Makassar, 2016). The primary occupation of social community in that area is farming in the steep mountainous area. It is hard to avoid this agricultural practice because this has become people's hereditary for practicing agriculture in the mountainous region. In Indonesia, where many upland areas can be found, land use/cover change due to farming activity commonly occurs (Rudiarto and Doppler, 2013).

The objectives of the study are to examine the effect of land use change in producing landslide susceptibility map to provide a comparative evaluation of the models.

5. 2 Material and method

This research divided into three main stages, i.e., data preparation, data analysis and validation (Figure 24).

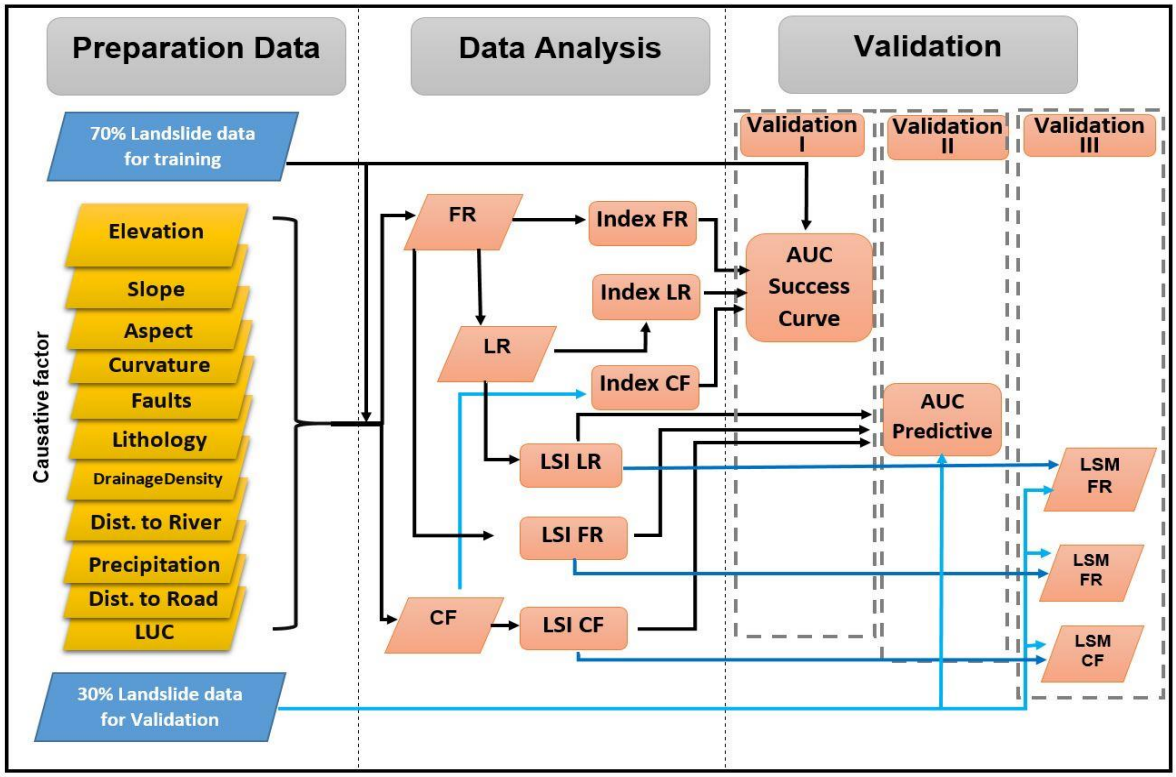


Figure 24 Research framework

5. 2. 1 Preparation of data

In preparation data, management, collection, and selection must be accurate in establishing a spatial data landslide inventory and a causative factor. This preparation data using GIS tools with ArcGIS© 10.3. For the analysis of the frequency ratio (FR) and the certainty factor (CF) calculation is done by Microsoft Excel©, while for the logistic regression (LR) using the Statistical Package for Social Sciences (SPSS©) software.

5.2.1.1 Landslide inventory

Inventory of landslides can include field surveys and interpretation of remote sensing images based on spectral characteristics, shape, contrast and morphological expression (Kanungo et al., 2006). This study used landslide events during 2012-2016 to quantitatively evaluate the influences of land use change and landslide occurrences during 2004-2011. Landslides from 2012 to 2016 were collected by using air photography from Google Earth Pro© and ground survey (Figure 26). To identify landslides occurrence by year, we delineated image according to year, started from 2012 until 2016. One hundred and eighty-eight landslides were identified,

covering area of 43.65 hectares (0.44 km^2). Most of the landslides are of the shallow type with minimum and maximum landslide area of 137 m^2 and 15,600 m^2 respectively. The study area was limited to upper of Ujung-Loe Watersheds. Figure 25 shows the location of all landslide data that were divided into two groups, i.e., landslide for training 2,873 pixels (70%) and a landslide for validation 1,230 pixels (30%). The selection of training and validation data was using ARC GIS tool by random selection.

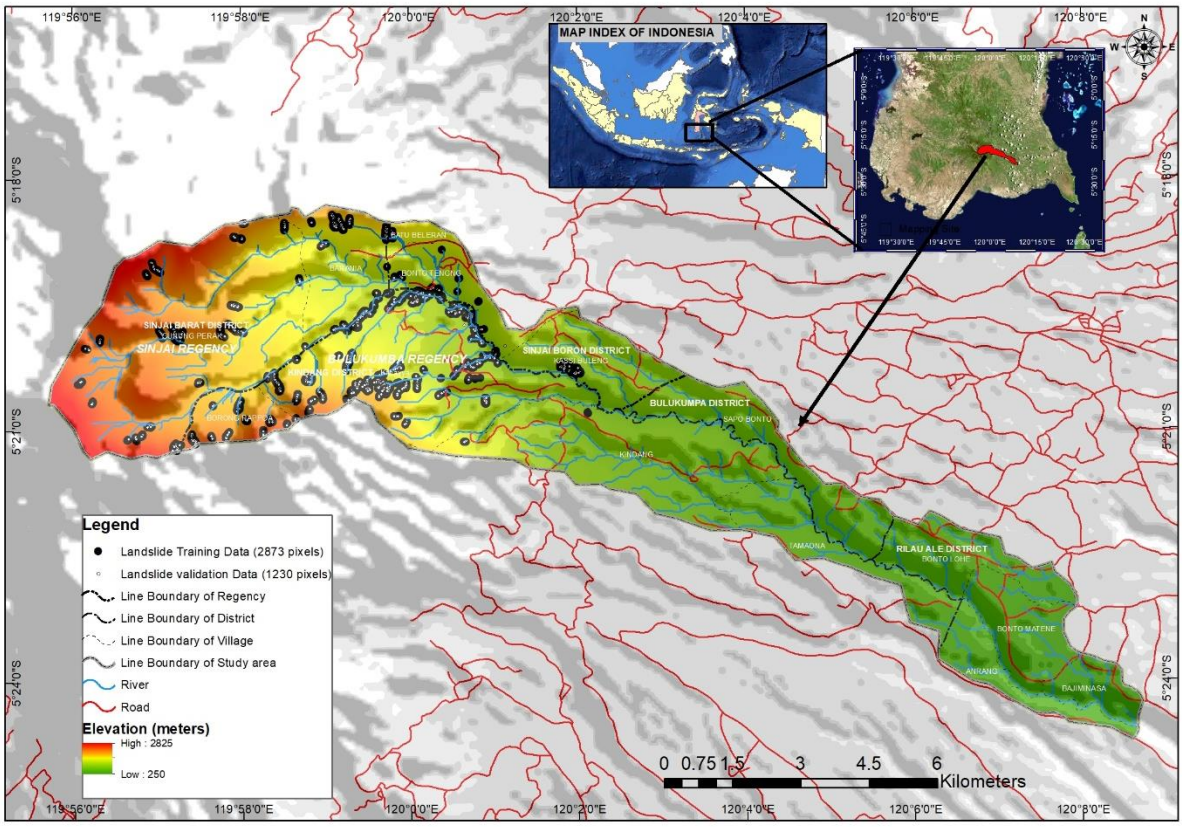


Figure 25 Map of Landslide Distribution

5.2.1.2 Landslide causative factors

In landslide susceptibility map, the most critical assumption that the incidence of landslides that will occur in the same condition is affected by the cause of the landslides that have occurred. There are no strict guidelines for the selection of causal factors for use in logistic regression analysis and assurance factor analysis and have been widely using by many studies (Ayalew and Yamagishi, 2005; Dou et al., 2015). Also, the determination of landslide causative factors is heavily reliant on data availability. Therefore, we chose causative factors based on the general knowledge found in previous studies (Rasyid et al., 2016) and data availability in

the target area. So according to past research and data availability, we use eleven causative factor, i.e., elevation, slope, aspect, curvature, lithology, distance from fault, distance to river, drainage density, precipitation, distance from the road, and land use change (LUC) (Figure 27).

Causative factors, i.e., elevation, slope, aspect, curvature were extracted from digital contour data with an interval of 12.5 meters. Digital contour data was derived from RBI (Indonesia Terrain) map with a scale of 1: 25,000 from Badan Informasi Geospasial (Geospatial Information Agency).

In this study, we used six classes of slope, i.e., 0–10°, 10–20°, 20–30°, 30–40°, 40–50°, and above 50°, which considered and represented in the form of slope thematic data layer. Likewise, the aspect map plays a significant role in slope stability assessment (Chauhan et al., 2010). Aspect was divided into nine classes namely, flat, north, northeast, east, southeast, south, southwest, west, and northwest. Profile curvature was classified into three categories; concave, convex, and flat. The value of the arch represents topographic morphology. In the case of profile curvature, it is associated with inundation conditions after heavy rains. Curvature slope profiles contain more water and hold water from high rainfall for more extended periods (Lee and Lee, 2006).

The geology data consists of lithology and fault lines. It is related to the strength of the material, because lithologic composition and structure vary for different types of rocks (Kanungo et al., 2006), and resistance to the driving force depends on the strength of rocks. Faults are structural features, which describes the zones/areas of weakness, fractures, and among lineament going higher susceptibility to landslides. It has been observed that the probability of landslide occurrence increased in a location close to faults, and it was not only affect the surface structure of the material but also contributes to the permeability and cause slope instability (Rasyid et al., 2016). For this purpose, the distance to faults was used to analyze the incidence of landslides occurrence. The distance to fault is done by buffering the map of faults.

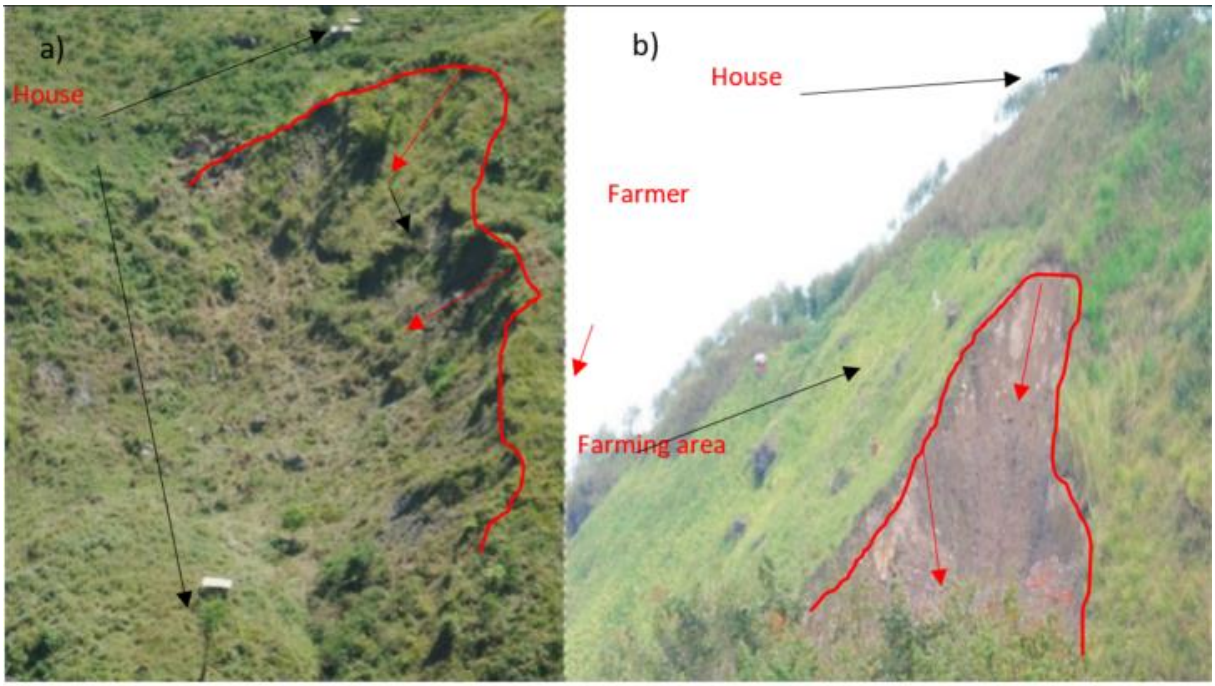


Figure 26 Landslide inventory a) old landslide, b) new landslide

Distance to river and landslide occurrence in the hilly area have strong association due to erosion process. Closer to the river, the soil conditions will be more humid, and with soil moisture, soil fertility will be high so that the soil bonds are not stable so erosion and landslides will quickly occur, especially during the rainy season. The distance from the river was calculated by buffering the map of the river in ARC GIS 10.3. River layer derived from a topographic map of scale 1:25.000. The classification of distance to river starts from 0 to 100 m and ends with > 500 m. Similarly, distance from the river, distance from the road also derived from the topographic map by interval 500 m in nine classes, and the class starts from 0 to 500 meter and ends with >4000 meters. Moreover, drainage density calculated by using Arc Toolbox kernel density in km/km^2. Drainage density was classified into five classes and start from 0 to 1 km/km^2 and ends with >4 km/km^2.

In addition to topographic and geological factors, land use change is a crucial element/factor responsible for landslide events. The incidence of landslides is inversely proportional to the density of vegetation. This research used land use change (LUC) factor as

identification of vegetation density. Change in land use to the critical slope triggered a series of shallow and profound landslides (Mugagga et al., 2012). The LUC map was derived from overlaying land use 2004 and land use 2011. Land use data were derived from interpretation Landsat 5 TM (date recorded September, 21th 2004) and Landsat 7 ETM+ (date recorded October, 11th 2011) images, each with a 30 m resolution, collected from United States Geological Survey (USGS). The unsupervised classification method is applied to classify land use. The unsupervised clasification was consisted of three steps: 1) Creation of N spectral class map using the self-organizing iterative data analysis algorithm; 2) development of land use map (LU) with the aid of reference data; and 3) the accuracy of all LU assessments which will validate with reference maps using independent data (Lang et al., 2008). This method is applied to classify land use into seven classes, i.e., open area, paddy field, farming area, scrub, savanna, secondary forest and primary forest. LU was validated by using ground control points method in the same year of Landsat images, and Google Earth Pro© imagery map was used to measure the accuracy. Accuracy assessment was using random sampling. Overall accuracy values of LU 2004 and LU 2011 were 86% and 90% respectively. Kappa values of 0.83, 0.88 were achieved for the unsupervised classified maps of LU 2004 and LU 2011, respectively. Moreover, LUC built by classifying LU 2004 and 2011 in four classes, i.e., one (open area, paddy field), two (Farming area and Shrub, Savana), three (secondary forest) and four (primary forest) again. Then, overlay each other using ArcGIS© 10.3 and founded 13 classes LUC. i.e., 1 – 1 (no change of open area and paddy field) , 1 – 2 (from open area and paddy field to farming area and scrub, savanna), 2 – 1(from farming area and scrub, savanna to open area and paddy field), 2 – 2 (no change on farming area and scrub, savanna), 2 – 3 (from farming area and scrub, savanna to secondary forest), 3 – 1 (from secondary forest to open area and paddy field), 3 – 2 (from secondary forest to farming area and scrub, savanna), 3 – 3(no change of secondary forest), 3 – 4 (from secondary forest to have similar density of primary forest), 4 – 1 (from primary forest to open area and paddy field), 4 – 2, 4 – 3(from primary forest to secondary

forest), and 4 - 4 (no change on primary forest). LUC in pixel 30 x 30-meter wasresampled to pixel 10 x 10 meter.

Landslide occurrence is described as the dependent variable, and causative factor, i.e., elevation, slope, curvature, distance to river, drainage density, lithology, distance to faults, precipitation, LUC and distance to roads are described as the independent variables. Independent variables and the dependent variable were used as an input for analysis landslide susceptibility map with a pixel resolution of 10 m × 10 m. We can see the causative factor map in Figure 27.

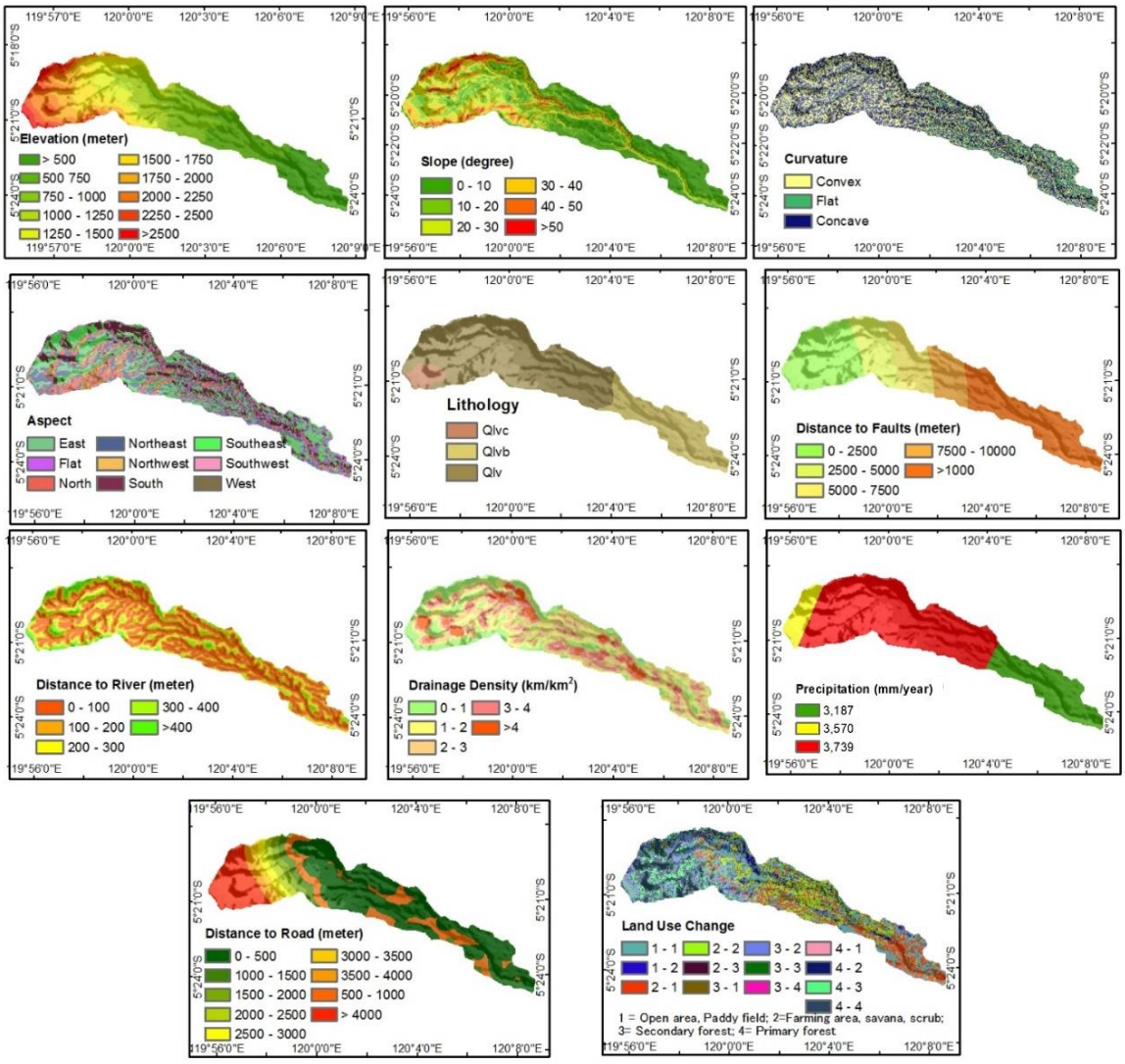


Figure 27 Eleven causative factor of landslide

5. 2. 2 Data Analysis

Three analyses methods were conducted to produce landslide susceptibility map, i.e., frequency ratio (FR), certainty factor (CF) and logistic regression.

5. 2. 2. 1 Frequency Ratio

The relationship between the area of the landslide and the causes can be inferred from the relationship between the area of the landslide and non-landslide and considering the causative factors. Frequency ratios for each causative factor type or range were calculated by dividing the landslide occurrence ratio by the area ratio. If the ratio is more significant than 1.0, the relationship between the landslide and the causative factor is higher, and, if the relationship is less than 1, the relationship between the landslide and each causative factor is low (Lee and Lee, 2006). A ratio value in each class shows the level of relationship and the frequency ratio value calculated by the following 5.1;

$$FR = \frac{\text{PxcL(ij)}/\sum\text{PixL}}{\text{Pixel(ij)}/\sum\text{Pix}} \tag{5.1}$$

where, PxcL(ij) number of pixel with landslide within class i of j parameter, Pixel(qij) number of pixel in class i of j parameter, ΣPixL total pixel of j parameter, and ΣPix total pixel of the area.

5. 2. 2. 2 Certainty factor

The certainty factor (CF) is a rule-based expert system method developed by Shortliffe and Buchanan (1975). The CF values range between -1 to 1, indicating a measure of belief and disbelief and can be calculated using the following function as Equation 5.2. Here, higher CF value indicates a higher relationship with landslide occurrences;

$$\text{CF} = \begin{cases} \dfrac{PPa - PPs}{PPa(1 - PPs)} \; if \; PPa \geq PPs \\ \dfrac{PPa - PPs}{PPs(1 - PPs)} \; if \; PPa < PPs \end{cases} \tag{5.2}$$

where; PPa is the probability of landslides in class and PPs is the prior probability of a total number of landslides in the study area.

5. 2. 2. 3 Logistic regression

A simple introduction to logistic regression is available in Chau and Chan (2005) which defines the probability occurrence of landslides divided by the probability of non-occurrence of landslides. It is useful to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. Generally, in the logistic regression, spatial prediction is modeled by the independent variables and the dependent variable (Shirzadi et al., 2012). It is useful when the variable is a binary or dichotomous. Variables can be continuous, or discrete, or a combination of the two types and they do not always have a normal distribution. The probability of regression can be understood as the probability of state-dependent variables. They are restricted to fall within a range of values from 0 to 1 (Xu et al., 2013); Zero shows probability of 0% landslide occurrences, and one showed a 100% probability (Dai et al., 2004). The logistic regression based on logistic function –z is expressed by the following Equation 5.3;

$$P = \frac{1}{1+\exp^{-Z}} \tag{5.3}$$

where P is the probability of landslide occurrence which varies from 0 to 1. Variable Z is landslide causative factors and assumed as a linear combination of the causative factors xi (i = 1,2,…n).

Fixing the sample size to create an equation in logistic regression analysis can be done using an equal number of landslide data and no landslide data to reduce bias in the sampling process. The constant and coefficient of independent variables provided by logistic regression analysis were estimated using SPSS (Rasyid et al., 2016).

5. 2. 3 Validation and verification

During the modeling predictions, the most essential and critical component is to carry out the validation of the results of prediction (Chung and Fabbri, 2003). Data for validation were selected randomly on each part of landslide occurrence without including the training dataset.

To illustrate the procedure, a small portion of the landslide- areas were selected as the data for validation. Size, area, depth of landslides and distribution significantly vary from place to place. Also, we used the ROC curve to plot predicted probabilities to estimate the model's accuracy. For validating the landslide susceptibility map, area under curve (AUC) was used as a measure of overall fit and comparison of modeled predictions. The model with higher AUC is considered to be the best. If the area under the AUC is close to 1, the result of the test is excellent. On the other hand, if the model does not predict well, then this value will be close to 0.5. The area determines the success rate AUC of the training dataset, and predictable level calculated from the AUC of the validation dataset. ROC curves are used to evaluate the predictive accuracy of the model selected in the statistical approach of dichotomous, such as logistic regression (Gorsevski et al., 2006), and AUC was obtained from the ROC plot statistics which give most preferred types and influence rating (Akgun et al., 2012). Predicted probabilities generated by logistic regression can be seen as an indicator continuously to compare with a binary response variable observed. Furthermore, further validation processes show the accuracy of landslide vulnerability maps is to calculate the ratio of landslide data that fall into each class of landslide susceptibility. It was assumed that most of the landslides for validation must occur on a high to very high susceptibility class (H + VH).

5. 3 Results and discussion

5. 3. 1 Frequency ratio

Table 14 indicates a correlation between landslide occurrence and each class of presence and absence landslides inventories for class of landslide causative factors. In the case of the relationship between landslide occurrence and LUC, LUC from primary forest to open area and paddy field (4-1) has the highest probability of landslide occurrence with frequency ratio 8.70. Moreover, LUC from primary forest to farming area, savanna, scrub (4-2) with frequency ratio 2.46. It is happening because the vegetation affects the stability of the slope. Land with forest by the root system would reinforce the soil strength and stabilizes the slope (Kubota et al.,

2007), and forest clearance seems to have manifested primarily through increased rates of landslide activity (Glade, 2003). Then, in slope class, slope above 20° has a ratio of >1 which indicates a high probability of landslide occurrence. Slope below 20° has a ratio of <1, which indicates a very low probability of landslide occurrence.

The class of elevation 1000 to 2000 meters (m) shows the highest correlation with landslide with landslide occurrences. In the class of curvature, the values represent the topographic morphology. A convex indicates a positive value, concave showed negative, and the value of zero indicates a flat surface. Accordingly, the value of the concave frequency ratio (1.33) has a higher likelihood of landslide than concave (0.85) and flat (0.43). In the case of the class aspect, north, northwest, south and northeast-facing slopes have the frequency ratio> 1, which shows a high rate of probability of the occurrence of landslides.

In the case of lithology classes, only Qlv has a ratio of >1 among the three lithology classes, which indicates a high probability of landslide occurrence. Quarter lompobattang volcanic (Qlv) is one of the volcanic and sediment formations in South Sulawesi area. In this case, the distance of the fault, rivers, and roads, the ratio of the distance/proximity are used to understand the degree of influence on the landslide. Distance from the faults below 7500 m has a ratio>1. It shows that as the distance from the fault decrease, the probability of landslide occurrence increases. Also the distance from the river below 100 m has frequency ratio >1. The distance from the road above 500 m has a ratio of >1. In the event of distance from roads, the landslide densities are higher for distance classes far away.

In precipitation class, class precipitation 4,528 mm/year has a ratio >1, which indicates a high probability of landslide occurrence. Moreover, it indicated more precipitation would induce more landslide occurrence. Aditian and Kubota (2017) point out that increasing rainfall rate; it is possible to become unstable and prone to landslide disaster.

To create an index of landslides susceptibility, causative factor in the form of raster maps of the value FR then summed by using Equation 5.4;

Table 14 The value of Frequency Ratio and Certainty Factor for each landslide causative factors

Causative factor	Class	Pixel Class*	% Class	Landslide pixel**	% Landslide	Frequency Ratio	PPa	PPs	CF
Elevation (meter)	<500	126010	15.85	0	0.00	0.00	0	0.003612805	-1.000
	500 – 750	113821	14.31	0	0.00	0.00	0	0.003612805	-1.000
	750 – 1000	117886	14.82	382	13.30	0.90	0.003240419	0.003612805	-0.103
	1000 – 1250	99735	12.54	544	18.93	1.51	0.005454454	0.003612805	0.339
	1250 – 1500	80401	10.11	446	15.52	1.54	0.005547195	0.003612805	0.350
	1500 – 1750	73551	9.25	665	23.15	2.50	0.009041345	0.003612805	0.603
	1750 – 2000	62583	7.87	452	15.73	2.00	0.007222409	0.003612805	0.502
	2000 - 2250	63418	7.97	202	7.03	0.88	0.003185216	0.003612805	-0.119
	2250 – 2500	38830	4.88	180	6.27	1.28	0.004635591	0.003612805	0.221
	>2500	18992	2.39	2	0.07	0.03	0.000105307	0.003612805	-0.971
Slope (degree)	0 -10	277391	34.88	233	8.11	0.23	0.00083997	0.003612805	-0.768
	10 – 20	193490	24.33	504	17.54	0.72	0.002604786	0.003612805	-0.280
	20 – 30	142736	17.95	519	18.06	1.01	0.003636083	0.003612805	0.006
	30 – 40	114954	14.46	613	21.34	1.48	0.005332568	0.003612805	0.324
	40 – 50	56795	7.14	819	28.51	3.99	0.014420283	0.003612805	0.752
	>50	9861	1.24	185	6.44	5.19	0.018760775	0.003612805	0.810
Curvature	Concave	335269	42.16	1,614	56.18	1.33	0.004814045	0.003612805	0.250
	Flat	100826	12.68	158	5.50	0.43	0.001567056	0.003612805	-0.567
	Convex	359132	45.16	1,101	38.32	0.85	0.003065725	0.003612805	-0.152
Aspect	Flat	48980	6.16	43	1.50	0.24	0.000877909	0.003612805	-0.758
	North	105139	13.22	963	33.52	2.54	0.009159303	0.003612805	0.608
	Northeast	140313	17.64	599	20.85	1.18	0.004269027	0.003612805	0.154
	East	128555	16.17	311	10.82	0.67	0.002419198	0.003612805	-0.331
	Southeast	155292	19.53	191	6.65	0.34	0.001229941	0.003612805	-0.660
	South	127354	16.01	515	17.93	1.12	0.004043846	0.003612805	0.107
	Southwest	48881	6.15	43	1.50	0.24	0.000879687	0.003612805	-0.757
	West	11324	1.42	23	0.80	0.56	0.002031084	0.003612805	-0.439
	Northwest	29389	3.70	185	6.44	1.74	0.006294872	0.003612805	0.428
Lithology	Qlvb	195818	24.62	0	0.00	0.00	0	0.003612805	-1.000
	Qlv	562441	70.73	2,826	98.36	1.39	0.005024527	0.003612805	0.282
	Qvlc	36968	4.65	47	1.64	0.35	0.00127137	0.003612805	-0.649
Distance to Faults (meter)	0 – 2500	228372	28.72	913	31.78	1.11	0.003997863	0.003612805	0.097
	2500 -5000	123498	15.53	1,333	46.40	2.99	0.010793697	0.003612805	0.668
	5000 – 7500	106243	13.36	472	16.43	1.23	0.004442646	0.003612805	0.187
	7500 – 10000	92127	11.58	155	5.40	0.47	0.00168246	0.003612805	-0.535
	>10000	244987	30.81	0	0.00	0.00	0	0.003612805	-1.000
Distance to River (meter)	0 - 100	325991	40.99	1,489	51.83	1.26	0.004567611	0.003612805	0.210
	100 – 200	240871	30.29	726	25.27	0.83	0.003014061	0.003612805	-0.166
	200 – 300	139539	17.55	397	13.82	0.79	0.002845083	0.003612805	-0.213
	300 – 400	59549	7.49	189	6.58	0.88	0.003173857	0.003612805	-0.122
	400 – 500	19942	2.51	42	1.46	0.58	0.002106108	0.003612805	-0.418
	>500	9335	1.17	30	1.04	0.89	0.003213712	0.003612805	-0.111
Drainage Density (km/km^2)	0 – 1	147677	18.57	698	24.30	1.31	0.004726532	0.003612805	0.236
	1 - 2	228100	28.68	635	22.10	0.77	0.002783867	0.003612805	-0.230
	2 - 3	252005	31.69	829	28.85	0.91	0.003289617	0.003612805	-0.090
	3 - 4	121676	15.30	512	17.82	1.16	0.004207896	0.003612805	0.142
	>4	45769	5.76	199	6.93	1.20	0.004347921	0.003612805	0.170
Precipitation (mm/year)	3187	186406	23.44	0	0.00	0.00	0	0.003612805	-1.000
	3570	61646	7.75	84	2.92	0.38	0.00509709	0.003612805	0.292
	3739	547175	68.81	2,789	97.08	1.41	0.001362619	0.003612805	-0.624
LUC (1=Open area, Paddy area; 2=Farming area, savanna, scrub; 3=Secondary Forest; 4=Primary Forest)	1 - 1	167966	21.12	608	21.16	1.00	0.00361978	0.003612805	0.002
	1 – 2	44883	5.64	276	9.61	1.70	0.006149322	0.003612805	0.414
	2 – 1	127015	15.97	134	4.66	0.29	0.001054994	0.003612805	-0.709
	2 - 2	140425	17.66	215	7.48	0.42	0.001531066	0.003612805	-0.577
	2 - 3	3971	0.50	2	0.07	0.14	0.000503651	0.003612805	-0.861
	3 – 1	24542	3.09	157	5.46	1.77	0.006397197	0.003612805	0.437
	3 - 2	88061	11.07	513	17.86	1.61	0.005825507	0.003612805	0.381
	3 - 3	30715	3.86	158	5.50	1.42	0.005144066	0.003612805	0.299
	3 – 4	4602	0.58	26	0.90	1.56	0.005649718	0.003612805	0.362
	4 – 1	954	0.12	30	1.04	8.70	0.031446541	0.003612805	0.888
	4 – 2	19912	2.50	177	6.16	2.46	0.008889112	0.003612805	0.596
	4 – 3	55800	7.02	180	6.27	0.89	0.003225806	0.003612805	-0.107
	4 – 4	86381	10.86	397	13.82	1.27	0.004595918	0.003612805	0.215
Distance to Road (meter)	0 – 500	407277	51.22	940	32.72	0.64	0.002308012	0.003612805	-0.362
	500 – 1000	115348	14.51	449	15.63	1.08	0.003892569	0.003612805	0.072
	1000 – 1500	34878	4.39	405	14.10	3.21	0.011611904	0.003612805	0.691
	1500 – 2000	24877	3.13	167	5.81	1.86	0.006713028	0.003612805	0.463
	2000 – 2500	23831	3.00	105	3.65	1.22	0.004406026	0.003612805	0.181
	2500 – 3000	23799	2.99	189	6.58	2.20	0.00794151	0.003612805	0.547
	3000 – 3500	23266	2.93	101	3.52	1.20	0.004341099	0.003612805	0.168
	3500 – 4000	23477	2.95	110	3.83	1.30	0.004685437	0.003612805	0.230
	>4000	118474	14.90	407	14.17	0.95	0.003435353	0.003612805	-0.049

*Total pixel area 795,227 **Landslide Training 2,873

$$LSI = FR_1 + FR_2 + \ldots + FR_n \quad (5.4)$$

where FR1, FR2, FR3… FRn is the frequency ratio raster maps of landslide causative factors.

The index value of frequency ratio falls in the range of 3.34 to 29.66. The higher LSI value showed a higher susceptibility to landslides and if the LSI lower showed lower susceptibility.

5. 3. 2 Certainty Factor

Table 14 indicates a correlation between landslide occurrence and each class of presence and absence landslides inventories for class of landslide causative factors. In the case of the relationship between landslide occurrence and LUC, LUC from primary forest to open area and paddy field (4-1) has the highest belief of probability of landslide occurrence with certainty factor (CF) 0.888. Moreover, LUC from primary forest to farming area, savanna, scrub (4-2) has CF value 0.596. It is also same with FR is happening because the vegetation affects the stability of the slope. Then in slope class, slope class above 50°, 40° - 50°, 30° - 40° and class 20° - 30° has a CF value 0.810, 0.752, 0.325 and 0.006 respectively, which indicates a significant belief of probability of landslide occurrence. Slope below 20° has a CF value <1, which indicates a very low probability of landslide occurrence.

In elevation class, elevation between 1000 to 2000 meters (m) stated belief to the likelihood of landslide occurrence. In curvature class, only concave has a conviction of probability of landslide occurrence with CF value 0.250. In the case of aspect class, the north, northwest, south, and northeast facing slopes, CF value is >0, which indicates a belief in the probability of landslide occurrence.

In the case of lithology classes, only Qlv has a ratio of >0 among the three lithology classes, which indicates a belief of probability of landslide occurrence. In the case of the distance from the fault, rivers, and roads used to understand the ratio of the distance/proximity

to the level of influence on the landslide. For distance from fault below 7500 m has a ratio of >0. It shows that as the distance from the fault decrease has a belief of probability of landslide occurrence. Also the distance from the river below 100 m has CF value >0. In the case of the distance from the road above 500 m has a CF value >1. In the event of distance from roads, the landslide densities are higher for distance classes far away, and its mean that distance to the road is not to effect to the landslide in this case. In precipitation class, only class precipitation 4528 mm/year has a CF >0, which indicates a belief of probability of landslide occurrence. Under increasing rainfall rate, it is possible for many forest slopes to become unstable and prone to landslide disaster shortly (Aditian and Kubota, 2017).

Table 15 Example is illustrating the calculation of certainty factor values for the combination of thematic layers using integration rules.

No	CF_e	CF_s	CF_{es}	CF_a	C_{fes-a}
1	0.600	0.320	0.728	0.110	0.758
2	0.220	-0.280	-0.077	-0.440	-0.483
3	0.340	-0.280	0.083	0.110	0.184
4	-0.100	-0.770	-0.793	0.110	-0.767
5	0.220	0.750	0.805	-0.660	0.426
6	0.220	0.010	0.228	-0.440	-0.275

CFe: Certainty factor value for elevation; CFs: Certainty factor value for slope; CFa: Certainty factor value for aspect; CFes: Combined certainty factor value of elevation and slope after integration for the various combination; Cfes-a: Combined certainty factor value combination of elevation-slope and aspect after integration for the various combination.

Landslide susceptibility index was created by combining pairwise layers according to the integration rules (Pourghasemi et al., 2013). The combination of CF values of two thematic layers 'Z' is expressed by the following equation as given by (Binaghi et al., 1998) as Equation 5.5.

$$\mathrm{Z} = \begin{cases} CF1 + CF2 - CF1CF2 \;\; CF1, CF2 \geq 0 \\ CF1 + CF + CF1CF2 \;\; CF1, CF2 < 0 \\ \dfrac{CF1 + CF2}{1 - \min(|\mathrm{CF1}|, |\mathrm{CF2}|)} \; CF1, CF2, opposite\ signs \geq 0 \end{cases} \tag{5.5}$$

The certainty factor values are computed by overlaying each thematic layer with the landslide map and calculate the landslide frequencies. Each thematic layer is reclassified according to the certainty factor value calculated and is combined pairwise to generate the

landslide susceptibility map using the integration rule of Equation 5.5. Table 15 illustrates the integration using a parallel combination.

5. 3. 3 Logistic regression

Hence, this study proposes ten iterations logistic regression coefficient (Table 16). Ten interations for logistic regression analysis will obtain the best result and sense of fairness with the same proportion of data landslide between the landslide and non-landslide event like as Rasyid et. al. (2016). Logistic regression method was conducted to compute the landslide occurrence probability, and if values are closer to one, landslides are more likely to occur. Land use change as a new causative factor for value was the distance to the river with coefficient 2.837 in iteration 7th that shows the distance to the river has the highest effect on landslide occurrence. Moreover, the lowest value was the distance to the road with -0.148, that indicates this causative factor is not influential in landslide occurrences. Moreover, according to likelihood ratio test LR indicates slope has the highest effect to landslide occurrence with chi-square value 417.299 and LUC on the 5th place from eleven causative factors after distance to river, distance to faults and aspect with value 85.065 (Table 17).

Table 16 Logistic regression coefficient of landslide causative factors using equal proportion of landslide and non-landslide pixel

Number Test	Variable in the equation of LR of Each Causative Factor											
	Elevation	Slope	Aspect	Curvature	Lithology	Distance to Faults	Drainage Density	Distance to River	Precipitation	Distance to Road	LUC	Constant
1	0.261	0.593	0.571	0.429	1.453	0.469	0.977	2.84	0.597	-0.168	0.378	-9.97
2	0.248	0.574	0.576	0.408	1.37	0.457	0.653	2.989	0.839	-0.164	0.57	-10.177
3	0.195	0.593	0.525	0.503	1.439	0.573	0.862	3.05	0.749	-0.162	0.661	-10.661
4	0.401	0.554	0.526	0.445	1.219	0.489	0.794	2.632	0.722	-0.187	0.623	-9.825
5	0.351	0.561	0.498	0.634	1.535	0.483	0.901	2.624	0.638	-0.183	0.441	-10.126
6	0.22	0.617	0.513	0.554	1.374	0.491	0.786	2.92	0.708	-0.143	0.596	-10.275
7	0.332	0.548	0.501	0.449	1.184	0.52	1.031	2.837	0.734	-0.148	0.589	-10.175
8	0.314	0.538	0.507	0.473	1.167	0.546	0.684	2.647	0.801	-0.185	0.479	-9.561
9	0.383	0.572	0.568	0.379	1.235	0.5	0.733	2.738	0.644	-0.201	0.539	-9.674
10	0.312	0.545	0.478	0.539	1.212	0.484	0.97	2.775	0.76	-0.083	0.502	-10.088

Table 17 Likelihood Ratio Tests using Logistic Regression

Causative Factor	Likelihood Ratio Tests
	Chi-Square
Slope	417.229
Distance to River	291.508
Distance to Faults	164.011
Aspect	128.522
LUC	85.065
Lithology	53.832
Drainage Density	35.265
Precipitation	27.894
Elevation	22.557
Curvature	13.893
Distance to Road	7.405

5. 3. 4 Validation

Table 18 shows results of AUC curve for both success rate and predictive rate for each test. In general, the AUC of ROC curves representing excellent, good, and missing values tests were plotted on the graph. The classification of the accuracy diagnostic test i.e. the value ranges from 0.50 to 0.60 (fail), 0.60–0.70 (poor), 0.70–0.80 (fair), 0.80–0.90 (good), and 0.90–1.00 (excellent) (Rasyid et al., 2016). The results show that the entire test of FR, CF and LR methods fall into the good category because the value ranges from 0.828 to 0.857 in success rate and 0.827 to 0.856 in predictive rate. Moreover, success rate and predictive rate value for all method were a closeness with interval 0.01 that indicates all the method more reliable to a predictive landslide in the future. The closeness of success rate and predictive rate values show how the method helps in landslide prediction in the future (Meten et al., 2015).

Table 18 AUC of ROC curve of success and predictive rate and the ratio of landslide validation on landslide susceptibility map using FR, CF and LR Method

Method	FR	CF	Number Test of LR									
			1	2	3	4	5	6	7	8	9	10
AUC Success rate	0.828	0.831	0.857	0.857	0.857	0.836	0.856	0.857	0.857	0.857	0.857	0.857
AUC Predictive rate	0.827	0.83	0.855	0.856	0.856	0.835	0.855	0.856	0.856	0.855	0.855	0.856
H+VH (%)	81.46	85.28	81.63	82.03	80.98	72.85	81.87	80.73	82.11	80.65	80.57	79.59

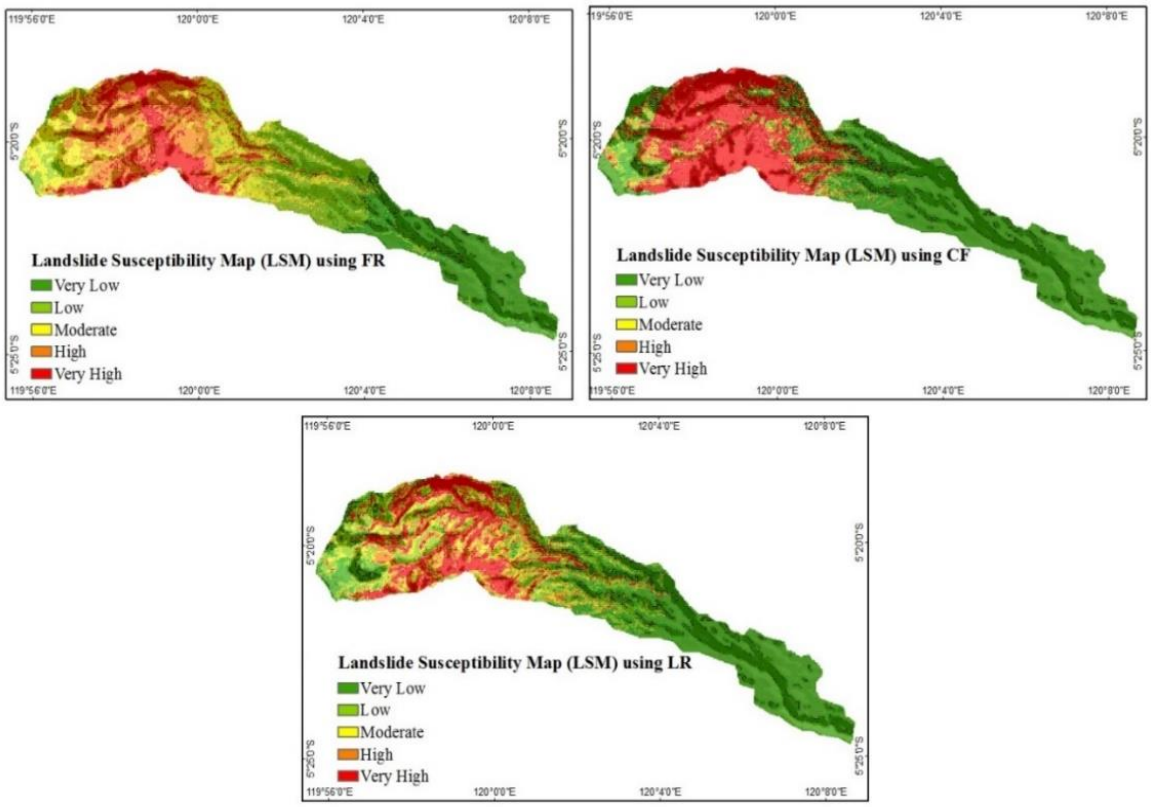


Figure 28 Landslide susceptibility map of FR, CF, and LR method 7th iterations

In this study, LR method conducts one more validation to choose the best statistical model for creating landslide susceptibility map and the best equation in logistic regression approach from the ten tests. The sum of FR value and equation of the LR models were used to create landslide susceptibility map (LSM). Moreover, CF was used Equation (5.5) to produce landslide susceptibility map (LSM). All LSM classes are created by reclassifying LSI of the models using natural breaks method, and overlaid landslide data validation on LSM will describe another level of accuracy besides AUC curve. The natural breaks method or Jenks optimization method has been used widely especially by planners, and it is designed to determine the best arrangement of values into different classes. This approach maximizes the variance between classes and reduces the variance within classes. The five classes include very low, low, moderate, high and very high describing the level of landslide susceptibility (proneness) in the study area. The level of accuracy of the landslide susceptibility map was verified by overlaying with the landslide data for validation. Table 18 shows the results of

overlaid landslide data for validation on LSM. LR method 7th iterations ratio (0.857) in AUC success rate were better than the ratio of FR (0.828) and CF (0.831), which shows that LR model is better of a model of identifying landslide. In predictive rate, LR method 7th iterations ratio was also better than the ratio of FR (0.827) and CF (0.830), which shows that LR model is better to predict of landslide occurrence. On the other hand, validation with the percentage of landslide falls into class high and very high of LSM, and CF model (85.28%) was better than the percentage of FR (81.46%) and LR model (82.1%). The curve of the model and validation proves that the susceptibility model is acceptable and the model could be applied to predict the potential landslides in future. As an interesting point to be noticed in Table 18, 0.857 is success rate and 0.856 is predictive rate.

Table 19 The Characteristic of susceptibility classes on landslide susceptibility map using FR, CF, and LR method

Class Number	Reclassified index value	Susceptibility class	Number of pixels	% area covered	Number of landslide validation pixel	% area of landslide validation covered
Frequency Ratio						
1	3.34 - 7.16	Very Low	177234	22.29	0	0.00
2	7.16 - 10.36	Low	143392	18.03	8	0.65
3	10.36 - 13.04	Moderate	201593	25.35	220	17.89
4	13.04 - 16.14	High	188458	23.70	428	34.80
5	16.14 - 29.66	Very High	84550	10.63	574	46.67
Certainty Factor						
1	-1 - -0.7647	Very Low	369510	46.47	31	2.52
2	-0.7647 - -0.2785	Low	75038	9.44	70	5.69
3	-0.2785 - 0.2626	Moderate	48592	6.11	80	6.50
4	0.2626 - 0.7096	High	80360	10.11	148	12.03
5	0.7096 - 0.9997	Very High	221727	27.88	901	73.25
Logistic Regression						
1	0.0011 - 0.1186	Very Low	282841	35.57	4	0.33
2	0.1186 - 0.3026	Low	163178	20.52	43	3.50
3	0.3026 - 0.5063	Moderate	138129	17.37	173	14.07
4	0.5063 - 0.7216	High	114794	14.44	336	27.32
5	0.7216 - 0.9997	Very High	96285	12.11	674	54.80

Figure 28 shows the landslide susceptibility map using FR, CF and LR method 7th iterations. The LSM by LR model was obtained using the coefficient values of landslide causative factors as in the equation below;

Z = -10.175 + 0.332 Elevation + 0.548 Slope + 0.501 Aspect + 0.449 Curvature + 1.184 Lithology+ 0.52 faults + 1.031 Drainage density + 2.837 Distance to River + 0.734 Precipitation - 0.148 Distance to Roads + 0.589 LUC

The ranges of the index value of each model in five classes were established using natural breaks method. Can et al. (2005) and Bai et al. (2010) stated two crucial guidance for validating landslide susceptibility map, i.e., 1) the high to very high class should cover only small areas and 2) landslide data validation should lie in high or very high classes.

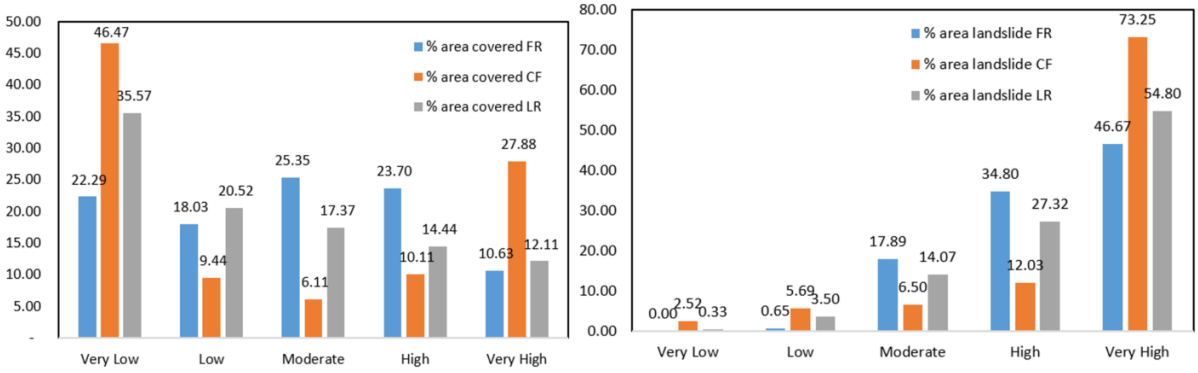


Figure 29 Percentage of landslide susceptibility classes and percentage of landslide susceptibility validation on landslide susceptibility of FR, CF and LR method

Table 19 shows the characteristics of susceptibility class for FR, CF and LR models. It indicates that the ratio of high to very high susceptibility class covers a small area. It was generated by dividing the number of pixels in each class on LSM to the total number of pixels. Furthermore, the ratio of landslide data for validation that fall on the LSM has a high value on high to very high class compared to very low to low class. The ratio calculated by dividing the number of a landslide for validation pixels, which lies on each susceptibility class to the total number of a landslide for validation pixels. This method is similar to FR, and CF model or the density method. In general, the procedure of creating landslide susceptibility map begins with the use of data of landslide occurrence as the dependent variable and landslide causative factors as the independent variables. Logically, landslide data covers a small area and occasionally in the form of scattered areas in the entire study area. The accuracy of the predicted future landslide that laid on the LSM should have a lower ratio in the class of low to very low class and higher in the high to very high class (Rasyid et al., 2016). Figure 29 shows, the ratio of the area with the classification of low to very low grade in LSM from FR, CF, and LR models have a total average of more than 40% of the total area, and the data validation landslide that fell on

the class shows the ratio below 10%. Also, the proportion of the area with the classification of high to very high grade in landslide susceptibility map from FR, CF, and LR model has a total average of more than 40% of the total area, and the data validation landslide that fell on the class shows the ratio of 80%. The highest values were 85.38% in the CF model. It indicates that the CF is better than others for validation by LSM with the landslide at high and very high-class.

5. 4 Conclusions

In conclusion, by using land use change (LUC) as a novel causative factor to produce landslide susceptibility map, LUC is influential in the creation of LSM. It can be inferred from the results of FR and CF, LUC has the highest value on both at LUC from primary forest to open area and paddy field. However, in logistic regression method, LUC has on the 5th place from eleven causative factor, according to likelihood ratio test with chi-square value 85.065 after slope, distance to river, distance to faults and aspect. This research comparatively evaluates the performance of FR, CF and LR models as well. Two-step of validation was carried out in this study. First, performances of each landslide model were tested using AUC curve for success and predictive rate, which is more than 82 % with the highest at LR Model. In the second, the ratio of landslides falling on high to a very high class of susceptibility was obtained, which indicates the level of accuracy of the model. The CF model has highest accuracy with 85.28 % landslides fall in the range of high to very high class while in LR and FR model, it is 82.11% and 81.46%. By the two-step of validation that LR shows highest accuracy in step 1 and CF show the highest in step 2.

5. 5 References

Aditian, A., and Kubota, T., 2017, The Influence of Increasing Rainfall Intensity on Forest Slope Stability in Aso Volcanic Area, Japan: International Journal of Ecology and Development, v. 32, p. 54-65.

Akgun, A., E.A.Sezer, H.A.Nefeslioglu, C.Gokceoglu, and B.Pradhan, 2012, An easy to use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm: Computers &Geosciences, v. 38, p. 23-34.

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

Badan Nasional Penangulangan Bencana (BNPB) Indonesia, 2016, Data dan Informasi Bencana Indonesia. http://dibi.bnpb.go.id/profil-wilayah/73/sulawesi-selatan.

Bai, S.B., Wang, J., Lü, G.N., Zhou, P.G., Hou, S.S., Xu, S.N., 2010. GIS-based logistic regression for landslide susceptibility mapping of the Zhongxian segment in the Three Gorges area, China. Geomorphology 115, 23–31. doi:10.1016/j.geomorph.2009.09.025

Binaghi, E., Luzi, L., Madella, P., Pergalani, F., and Rampini, A., 1998, Slope Instability Zonation: a Comparison Between Certainty Factor and Fuzzy Dempster–Shafer Approaches: Natural Hazards, v. 17, p. 77-97.

Can, T., Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., Duman, T.Y., 2005. Susceptibility assessments of shallow earthflows triggered by heavy rainfall at three catchments by logistic regression analyses. Geomorphology 72, 250–271. doi:http://dx.doi.org/ 10.1016/ j.geomorph.2005.05.011

Chau, K.T., Chan, J.E., 2005. Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. Landslides 2, 280–290. doi:10.1007/s10346-005-0024-x

Chauhan, S., Sharma, M., Arora, M.K., Gupta, N.K., 2010. Landslide susceptibility zonation through ratings derived from artificial neural network. Int. J. Appl. Earth Obs. Geoinf. 12, 340–350. doi:10.1016/j.jag.2010.04. 006

Chung, C. J. F., and Fabbri, A. G., 2003, Validation of spatial prediction models for landslide hazard mapping: Natural Hazards, v. 30, no. 3, p. 451-472.

Dai, F. C., Lee, C. F., Tham, L. G., Ng, K. C., and Shum, W. L., 2004, Logistic regression modelling of storm-induced shallow landsliding in time and space on natural terrain of Lantau Island, Hong Kong: Bull Eng Geol Environ, v. 63, p. 315-327.

Dou, J., Bui, D. T., Yunus, A. P., Jia, K., Song, X., Revhaug, I., Xia, H., and Zhu, Z., 2015, Optimization of Causative Factors for Landslide Susceptibility Evaluation Using Remote Sensing and GIS Data in Parts of Niigata, Japan: PLoS ONE, v. 10 (7), p. e0133262. doi:0133210.0131371/journal.pone.0133262.

Fell, R., Corominas, J., Bonnard, C., Cascini, L., Leroi, E., and Savage, W. Z., 2008, Guidelines for landslide susceptibility, hazard and risk zoning for land-use planning: Engineering Geology, v. 102, p. 99-111.

García-Ruiz, J. M., Beguería, S., Alatorre, L. C., and Puigdefábregas, J., 2010, Land cover changes and shallow landsliding in the flysch sector of the Spanish Pyrenees: Geomorphology, v. 124, p. 250-259.

Glade, T., 2003, Landslide occurrence as a response to land use change: a review of evidence from New Zealand: Catena, v. 51, no. 3-4, p. 297-314.

Gorsevski, P. V., Gessler, P. E., Foltz, R. B., and Elliot, W. J., 2006, Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis: Transactions in GIS, v. 10 (3), p. 395-415.

Hasnawir, and Kubota, T., 2012, Shallow Landslides Distribution and Rainfall Thresholdin Kelara Watershed, Indonesia: International Journal of Japan Erosion Control Engineering Technical note v. 5 (No 1), p. 86-92.

Kanungo, D. P., Arora, M. K., Sarkar, S., and Gupta, R. P., 2006, A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas: Engineering Geology, no. 85, p. 347-366.

Kubota, T., Omura, H., and Devkota, B. D., 2007, Influence of The Forest on Slope Stability with Different Forest felling Condition: EGU General Assembly 2007 Vienna.

Lang, R. L., Shao, G. F., Pijanowski, B. C., and Farnsworth, R. L., 2008, Optimizing unsupervised classifications of remotely sensed imagery with a data-assisted labeling approach: Computers & Geosciences, v. 34, no. 12, p. 1877-1885.

Lee, S., Chwae, U., and Min, K. D., 2002, Landslide susceptibility mapping by correlation between topography and geological structure: the Janghung area, Korea: Geomorphology, v. 46, no. 3-4, p. 149-162.

Lee, S., and Lee, M.-J., 2006, Detecting landslide location using KOMPSAT 1 and its application to landslide-susceptibility mapping at the Gangneung area, Korea: Advances in Space Research, v. 38, p. 2261-2271.

Meteorology, Climatology, and Geophysical Agency Makassar, 2016. Rainfall Data 2010-2015.

Matebie Meten, Netra PrakashBhandary, and Ryuichi Yatabe, 2015, Effect of Landslide Factor Combinations on the Prediction Accuracy of Landslide Susceptibility Maps in the Blue Nile Gorge of Central Ethiopia: Geoenvironmental Disasters, v. 2, p. 9. doi:10.1186/ s40677-40015-40016-40677.

Mugagga, F., Kakembo, V., and Buyinza, M., 2012, Land use changes on the slopes of Mount Elgon and the implications for the occurrence of landslides: Catena, v. 90, p. 39-46.

Pourghasemi, H.R., Pradhan, B., Gokceoglu, C., Mohammadi, M., Moradi, H.R., 2013. Application of weights-of-evidence and certainty factor models and their comparison in landslide susceptibility mapping at Haraz watershed, Iran. Arab. J. Geosci. 6, 2351–2365. doi:10.1007/s12517-012-0532-7

Rasyid, A. R., Bhandary, N. P., and Yatabe, R., 2016, Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia: Geoenvironmental Disasters, v. 3, p. 19.

Rudiarto, I., and Doppler, W., 2013, Impact of land use change in accelerating soil erosion in Indonesian upland area: a case of Dieng Plateau, Central Java - Indonesia: International Journal of AgriScience, v. 3(7), p. 558-576.

Shirzadi, A., Saro, L., Joo, O. H., and Chapi, K., 2012, A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran: Natural Hazards, v. 64, p. 1639-1656.

Shortliffe, E. H., and Buchanan, G. G., 1975, A model of inexact reasoning in medicine: Math. Biosci, v. 23, p. 351-379.

Xu, C., Xu, X., Dai, F., Wu, Z., He, H., Shi, F., Wu, X., and Xu, S., 2013, Application of an incomplete landslide inventory, logistic regression model and its validation for landslide susceptibility mapping related to the May 12, 2008 Wenchuan earthquake of China: Natural Hazards, v. 68, p. 883-900.

Chapter 6 Optimization of Causative Factor Using Logistic Regression and Artificial Neural Network Models for Landslide Susceptibility Assessment in the Mountainous area of Ujung Loe Watershed South Sulawesi Indonesia

6. 1 Introduction

Landslide susceptibility map (LSM) plays a vital role in assisting land-use planning and managing landslide hazard and their mitigation measures. The landslide susceptibility/hazard mapping relies on a thorough understanding of the complex slope movements and their controlling factors. The reliability of landslide susceptibility or hazard map is highly dependent on the number and quality of existing data, the scale of the input data, data analysis, and selection of appropriate models. The process of making this map involves several qualitative or quantitative approaches (Aleotti and Chowdhury, 1999; Cardinali et al., 2002; Dou et al., 2015). Quantitative methods investigate the relationship between landslide occurrence and causative factors to predict the occurrence probabilities.

A wide range of quantitative methods has been successfully used for landslide susceptibility mapping by researchers around the world. The widely used methods are bivariate and multivariate statistical methods including logistic regression (LR) (Ayalew and Yamagishi, 2005; Rasyid et al., 2016; Soma and Kubota, 2017a), fuzzy logic (Demicco and Klir, 2004; Ercanoglu and Gokceoglu, 2004; Tien Bui et al., 2017), support vector machines (Ballabio and Sterlacchini, 2012; Chen et al., 2017; Pourghasemi et al., 2013; Yao et al., 2008) and artificial neural network (Chauhan et al., 2010; Ermini et al., 2005; Pradhan et al., 2010; Pradhan and Lee, 2010). The statistical method of bivariate and multivariate to estimate landslide probabilities is based on correlation analysis between causative factors and historical landslide events, whereas the deterministic methods assess slope failures using the factor of safety (FoS) (Bahsan et al., 2014; Jamsawang et al., 2015). Logistic regression (LR) is considered to be the most commonly used methods for the assessment of the probability of occurrence of landslides

at medium and regional scales (Meinhardt et al., 2015; Shahabi et al., 2013). The advantage of logistic regression (LR) over other multivariate analysis methods is that LR is independent of data distribution and also it can handle a variety of data such as continuous, categorical, and binary data (Dou et al., 2015). Artificial neural network (ANN) is computational information processing units inspired by the structure and behavior of real biological neurons whose architecture mimics the knowledge acquisition and organizational skills of human brain cells. According to Yilmaz (2009), Artificial Neural Networks with their remarkable ability to derive meaning from complicated data, can be used to extract patterns and detect trends that are too complex to be noticed by either humans or other computer techniques. Artificial neural networks can be considered as 'experts' in the information category. It has been used to provide projections given new situations of interest and answer the question 'what if' and other advantages are adaptive learning, organizing their own real-time operation and fault tolerance through redundant coding of information.

Also, many scientists randomly and subjectively select factors such as geology, hydrology and human factor such as distance to road, and especially with land use change factors (Soma and Kubota, 2017a) to produce landslide susceptibility maps. Therefore, the selection of landslides causative factors is an essential point in the study of landslide susceptibility maps. In this research, we address this issue by proposing the LR, ANN and their Combination method that has rarely been used for feature selection of optimized causative factor to produce LSM studies.

The study area is located in the mountainous area of upper Ujung Loe watershed. The topography is naturally mountainous and steeper (38.8% of the slope class is higher than 20 degrees) and very high rainfall with 2,976 to 7,114 mm/year. These two landslide causative factors contribute significantly to the landslide occurrence in the study area. Besides this, the land use changes from primary forest to an open area, paddy field, and farming area has made the area more prone to landslides (Soma and Kubota, 2017b). The primary occupation of the

community in the study area is farming in this mountainous area, thereby enhancing erosion and groundwater infiltration followed by land sliding. It is hard to avoid this agricultural practices because this has become people's culture, means of existence and practice that has been inherited from their ancestors (Soma and Kubota, 2017b). Based on these premises, the primary objective of this study was to optimize the landslide causative factors using logistic regression and artificial neural network to produce the landslide susceptibility maps and select the best one among the two maps. Under this primary objective, the specific objectives include preparing landslide causative factor maps.

6. 2 Data and Methods

Landslide susceptibility analysis in the present study has been carried out in three main steps (Figure 30), i.e. (1) Data Preparation, (2) Data Analysis, (3) Validation.

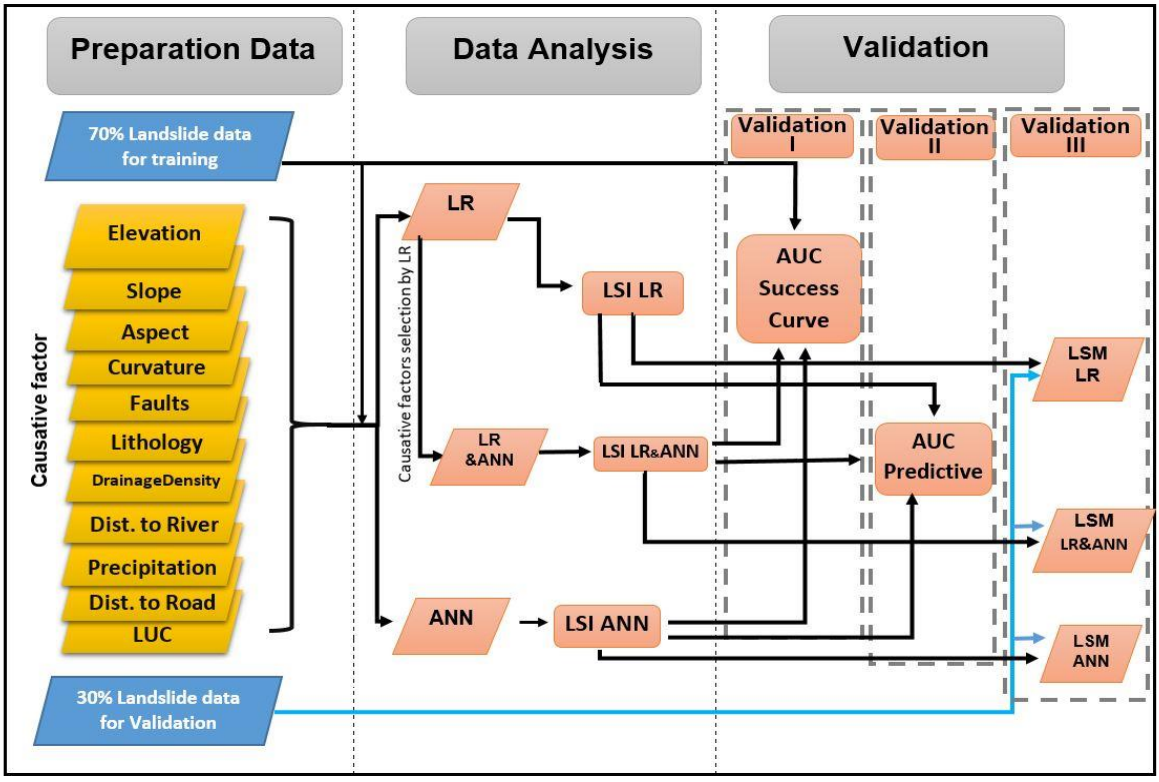


Figure 30 Research framework

6 . 2. 1 Data Preparation

Data selection is the critical thing in the preparation of the landslide susceptibility map. The excellent data selection for analysis helps to find satisfactory results. Management and collection or selection using Arc GIS© 10.3 must be accurate in establishing a spatial data landslide inventory and also a causative factor.

6. 2. 1. 1 Landslide inventory

Based on these premises, the primary objective of this study was to optimize the landslide causative factors using logistic regression and artificial neural network to produce the landslide susceptibility maps and select the best one among the two maps. Under this primary objective, the specific objectives include preparing landslide causative factor maps (Kanungo et al., 2006). This study used landslide data extracted from 2012 to 2016 Google Earth Pro© images and ground surveys (Figure 31 and Figure 32). The study area was limited to the upper Ujung Loe watershed. A total of 188 landslides were identified covering an area of 43.65 hectares (0.44 km2). Most of the landslides are of the shallow type with minimum and maximum landslide

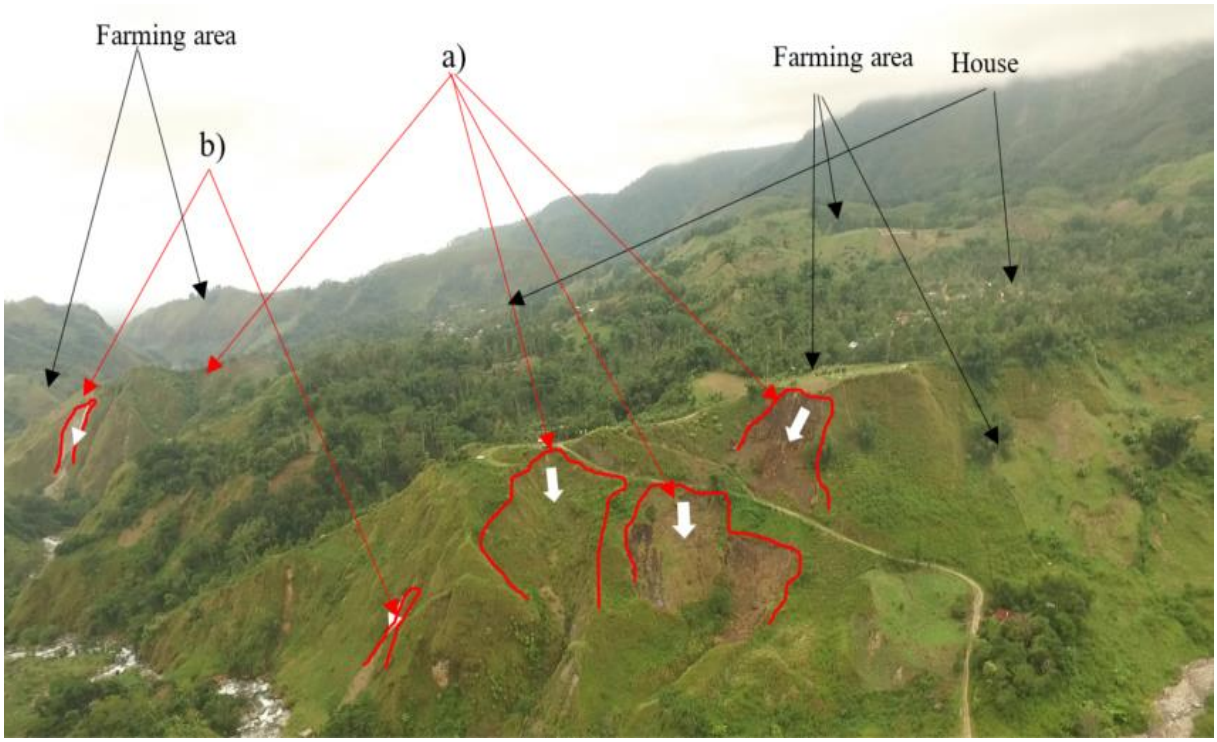


Figure 31 Landslide inventory a) old landslide, b) new landslide

area of 137 m2 and 15,600 m2, respectively. Landslide inventory generated from the survey and digitizing high-resolution from Google Earth Pro©, by digitized the time series imaging data by delineate the landslide, and these files were saved as GIS compatible format as extension kml. Then, the data was again subsequently changed into shapefile and raster format 10x10 meter.

6. 2. 1. 2 Landslide causative factors

The most critical assumption in landslide susceptibility map is that the landslide occurrence was caused by causative factors. There are no strict guidelines for the selection of causative factors to be used in logistic regression analysis and artificial neural network (Ayalew and Yamagishi, 2005; Dou et al., 2015; Ermini, 2005, Aditian and Kubota, 2017, Soma and Kubota, 2017b). Correspondingly, the determination of landslide causative factors was associated with the availability of data. Therefore, we selected causative factors based on the general knowledge found in previous studies and its availability in the target location. The entire landslide causative factors have been used for the independent variables in the landslide

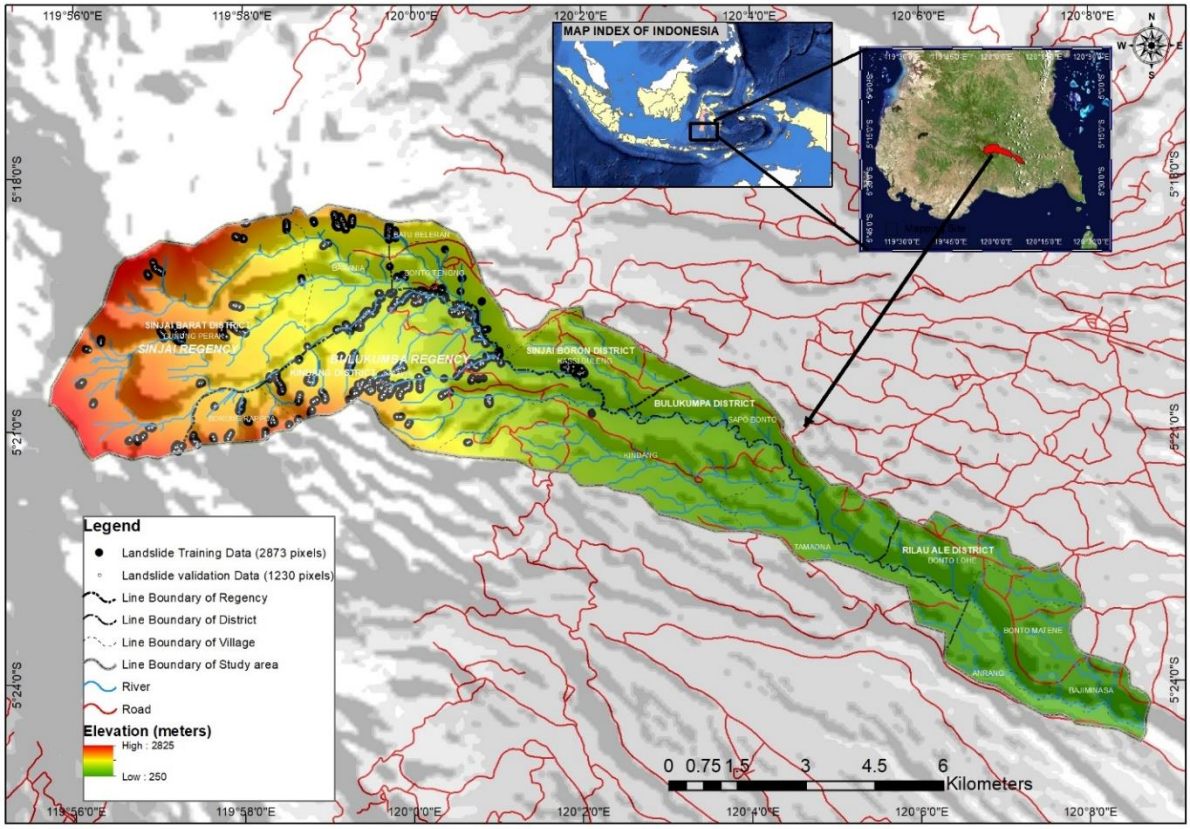


Figure 32 Map of Landslide Distribution

Table 20 Landside causative factors and their classes

No	Landside causative factors	Class
1	Elevation (m)	(1)[>500]; (2)[500,750]; (3)[750,1000]; (4)[1000,1250]; (5)[1250,1500]; (6)[1500,1750]; (7)[1750,2000]; (8)[2000, 2250]; (9)[2250, 2500]; (10)[>2500]
2	Slope (degree)	(1) [0,10]; (2) [10,20]; (3)[20,30]; (4) [30,40]; (5) [40,50]; (6) [>50] (Chauhan et al., 2010)
3	Aspect	(1) flat; (2) north; (3) northeast; (4) east; (5) southeast; (6) south; (7) southwest; (8) west; (9)northwest (Chauhan et al., 2010)
4	Curvature	(1) Convex; (2) Flat; (3) Concave
5	Lithology	(1) Quarter Lompobattang Vulcanic center (Qlvc); (2) Quarter Lompobattang Vulcanic (Qlv); (3) Quarter Lompobattang Vulcanic Breccia
6	Distance to Faults (m)	(1)[0,2500]; (2)[2500,5000]; (3)[5000,7500]; (4)[7500,10000]; (5)[>10000]
7	Distance to River (m)	(1)[0,100]; (2)[100,200]; (3)[200,300]; (4)[300,400]; (5)[>400]
8	Drainage Density	(1)[0,1]; (2)[1,2]; (3)[2,3]; (4)[3,4]; (5)[>4]
9	Precipitation (mm/year)	(1)[3538]; (2)[3933]; (3)[4528]
10	Distance to Road (m)	(1)[0,500]; (2)[500,1000]; (3)[1000,1500]; (4)[1500,2000]; (5) [2000,2500]; (6)[2500,3000]; (7)[3000,3500]; (8)[3500,4000] (9)[>4000]
11	Land Use Change (LUC) (1=Open area, Paddy area; 2=Farming area, savanna, scrub; 3=Secondary Forest; 4=Primary Forest)	(1)[1-1]; (2)[1-2]; (3)[2-1]; (4)[2-2]; (5) [2-3]; (6)[3-1]; (7)[3-2]; (8)[3-3] (9)[3-4];(10)[4-1]; (11) [4-2]; (12)[4-3]; (13)[4-4] (Soma and Kubota, 2017b)

susceptibility mapping (Figure 33). The independent variables were eleven of causative factors including elevation, slope, aspect, curvature, lithology, distance from fault, distance to river, drainage density, precipitation, distance to road and land use change (LUC) from 2004 to 2011.

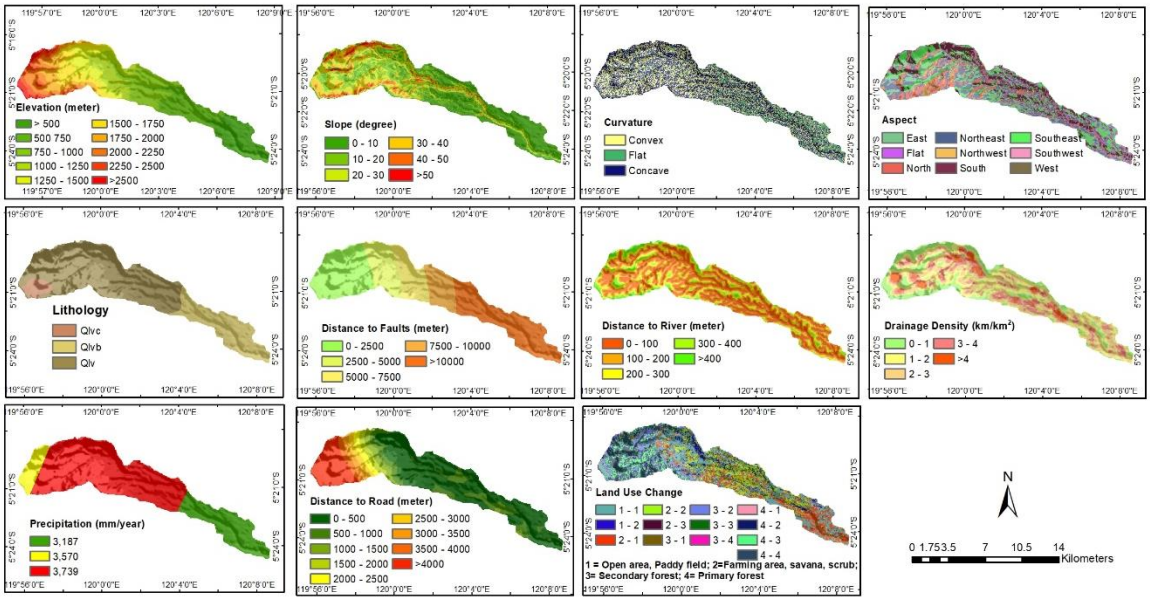


Figure 33 Eleven causative factors of landslide

Elevation, slope, aspect, and curvature were extracted from contour digital data with 12.5 meter interval. Contour data of Rupa Bumi Indonesia (RBI) on map scale 1: 25000 from Badan Geospatial Indonesia (BIG) were obtained using arc toolbox raster surface in ArcGIS 10.3©.

The geology of the area was obtained from the digital geologic map of Indonesia produced by Geological Research Institute at a scale of 1:250,000. The geologic data consists of lithology (rock type) and structures (fault and lineament). Lithology is the primary data or parameter for analyzing the probability of landslide occurrence. It related to the strength of the material because lithological composition and structure vary for different types of rocks (Kanungo et al., 2006). In addition, resistance to the driving forces also depends on the strength of rocks. Faults are structural features that describe the zones or areas of weakness following fractures or joints in a certain rock type posing a higher degree of susceptibility to landsliding. It has been observed that the probability of landslide occurrence will increase close in areas close to faults because faults not only affect the surface configuration and strength of the rocks but also increases the permeability and hence cause slope instability. For this purpose, the proximity faults were obtained by buffering the fault map of the area (Rasyid et al., 2016).

Both distances to river and drainage density in the hilly area had a strong association with landslide occurrence due to erosional activity in this location. The distance from the river was calculated by buffering analysis of streamlines and drainage density calculate by using line density in ArcGIS 10.3©. Distance to the road has good correlation with landslide occurrence when the road is under construction because the stability of the slope will be affected due to slope toe undercutting. Distance to the road has been calculated by Euclidean distance analysis in ArcGIS 10.3©. This information was derived from a topographic map of scale 1:25000 called Peta Rupa Bumi Indonesia (RBI) prepared by Badan Informasi Geospasial (BIG) Indonesia in 2012.

Precipitation had a good correlation with landslide occurrence as a trigger of landslide. Precipitation map was generated by using "create Thiessen polygon" tool from three rain gauge

stations in the study area. Moreover, land use change has a good correlation with landslide occurrence because changing the vegetation to another landscape causes slopes unstable and increases the probability of landslide occurrence especially in Ujung Loe watershed (Soma and Kubota, 2017a). LUC causative factor were applied using the data from Soma and Kubota (2017a). The classification of each causative factor can be seen in Table 20.

The landslide was considered as a dependent variable while the causative factors like elevation, slope, curvature, distance to river, drainage density, lithology, distance to faults, precipitation, distance to road and LUC, was described as the independent variables. Independent and dependent variables were used as input maps and then processed to turn into a raster map with a pixel size of 10 m × 10 m. The causative factor maps can be seen in Figure 33. Generally, the study area contains 795,227 pixels. The landslides contain a total of 4,103 pixels in which 2,873 pixels (70% of the landslides) were used for training/prediction/ while 1,230 pixels (30% of the landslides) were used for validation purpose.

6 . 2. 2 Data Analysis

There are two main ideas in this analysis, i.e., first to understand the performance of each causative factor and second to optimize causative factors to produce landslide susceptibility maps. In order to understand the performance of each causative factor, the two statistical/ probabilistic methods, i.e., logistic regression (LR) and Artificial neural network (ANN) were used. However, to optimize causative factors LR, forward stepwise (likelihood ratio) logistic regression (FSLR), ANN and a combination between FSLR and ANN were used to eliminate the less critical causative factors and produce landslide susceptibility map using the optimized causative factors. The analysis of LR and ANN was also partly conducted in SPSS© software.

6. 2. 2. 1 Multivariate Logistic regression

A simple introduction to logistic regression is available in Chau and Chan (2005) which defines the probability occurrence of landslides divided by the probability of non-occurrence of landslides. It is useful to predict the presence or absence of a characteristic or outcome based

on values of a set of predictor variables. Generally, in the logistic regression, spatial prediction can be modeled using dependent and independent variables (Shirzadi et al., 2012). It is useful when the variable is a binary or dichotomous. Variables can be continuous, or discrete, or a combination of the two types and they do not always have a normal distribution. The probability of regression can be understood as the possibility of state-dependent variables. Data analysis created iteration in ten tests using same proportion of data landslide between landslide and no landslide occurrence. Using same proportion of data landslide between landslide and no landslide occurrence will result better and fair for logistic regression analysis (Rasyid et al., 2016). LR resulted in landslide susceptibility index. The logistic regression followed on logistic function –z expressed by the following equation 6.1 and 6.2;

$$P = \frac{1}{1+\exp^{-z}} \tag{6.1}$$

$$Z = C_0 + C_1CF_1 + C_2CF_2 + \ldots + C_nCF_n \tag{6.2}$$

where P is a probability of landslide occurrence, and its values are varying from 0 to 1. Variable Z is landslide causative factor and is assumed as a linear combination of the causative factors Xi (i = 1, 2, …). Moreover, Z calculates using Equation (6.2). C_0 is the intercept, and C1, C2, Cn are coefficient, which measures the contribution of independent factors (CF_1, CF_2, . . ., CF_n) to the variations in Z.

6. 2. 2. 2 Artificial Neural Network (ANN)

ANN is an information processing system which is inspired by the models of biological neural networks through the network during the training phase (Sheela and Deepa, 2013). ANN is widely used in many areas because of its essential features such as the high capacity of nonlinear mapping, high accuracy for learning, and good robustness (Kanungo et al., 2006; Sheela and Deepa, 2013). Artificial neural networks (ANN) are computational information processing units inspired by the structure and behavior of real biological neurons whose architecture mimics the knowledge which is an essential variable of different units. The normalization of data is essential as the variables of different units. The normalization data is

extended to allow it to accommodate new types of data, and the pre-existing aspects of the database structure can remain mostly or entirely unchanged. As a result, applications interacting with the database are minimally affected. The data are scaled within the range of 0 to 1. The scaling is carried out to improve the accuracy of subsequent numeric computation and obtain a better output (Sheela and Deepa, 2013). The normalization of data is obtained by the following equation 6.3;

$$X_i' = \left(\frac{x_i - x_{min}}{x_{max} - x_{min}}\right)(x'_{max} - x'_{min}) + x'_{min} \quad (6.3)$$

where X_i' is normalized input, X_i, $X_{\min}$, $X_{\max}$ are the actual input data, minimum and maximum input data. x'_{max}, x'_{min} be the minimum and maximum target value.

In this study, multi-layer perceptron (MLP) was applied. Each hidden and output layer neuron processes its inputs by multiplying each input (x_i) by a corresponding weight (w_i), summing the product with Equation 6.4;

$$net = \sum_{i=0}^{n} w_i \, x_i \quad (6.4)$$

Then processing the sum (if that exceeds the neuron threshold, the neuron is then activated) using a non-linear activation function to produce a result (y_i), which is the output node with Equation 6.5;

$$yi = G(b^2 + W^2(s(b^1 + W^1 i))) \quad (6.5)$$

A three-layer feed-forward network consisting of an input layer (11 neurons), one hidden layer (10 neurons) and two output layer was used as a network structure of 11-10-2 (Figure 34). There are many rule-of-thumb methods for determining an acceptable number of neurons to use in the hidden layers, such as the number of hidden neurons should be between the size of the input layer and the size of the output layer (Sheela and Deepa, 2013). The reason sigmoid function is chosen is that exponential functions are similar to handle mathematically and since learning algorithms involve lots of differentiation, thus choosing a function that is

computationally cheaper to handle is quite good. The sigmoid function has been using commonly for studies (Conforti et al., 2014; Kanungo et al., 2006; Nefeslioglu et al., 2008; Pham et al., 2017; Pradhan et al., 2010). The initial value of the learning level is the gradient descent algorithm. A higher learning rate means that the network will train faster, possibly at the cost of becoming unstable, for this analysis set to 0.65. The initial momentum parameter is the gradient descent algorithm. The momentum term helps to prevent instabilities caused by a too-high learning rate, for this analysis set to 0.95 with interval center set to 0 and interval offset set to 0.5. The number of iterations was set to 1,000, and the RMSE value used for the interrupt of the training phase was set to 0.001.

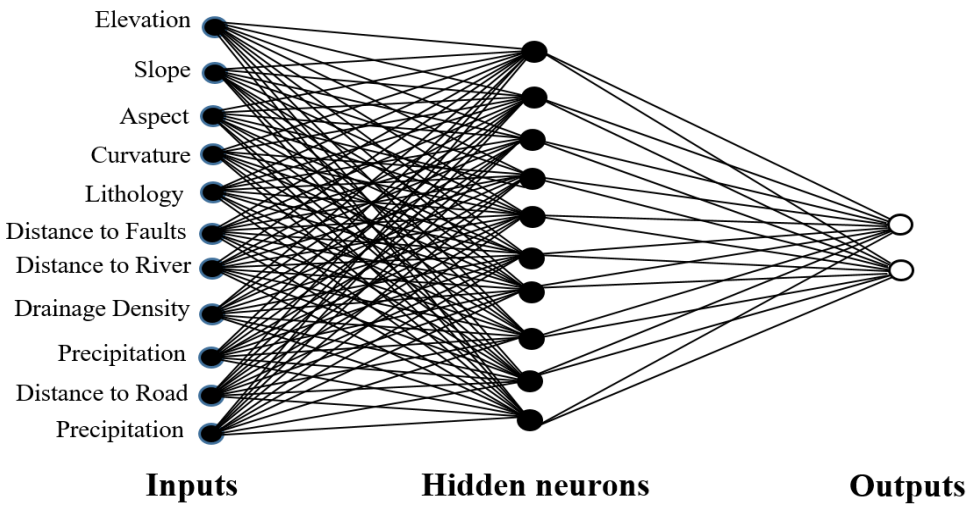


Figure 34 Architecture of artificial neural network in this research

6. 2. 2. 3 Optimized Causative Factor

Optimizing of causative factor was conducted by using 3 types of methods, i.e. (1) FSLR, (2) ANN; and (3) Combination FS-LR and ANN (FSLR-ANN). Optimizing causative factors with LR-FS was conducted by eliminating one by one from 11 factors to eight factors with the least essential effect by the likelihood ratio test values from FSLR analysis, and then the landslide susceptibility map was prepared using LR method. Optimizing causative factors with ANN was conducted by eliminating one by one factor from 11 factors to eight factors with least

significant effect by normalized importance value from ANN analysis and make landslide susceptibility map using ANN method. Optimizing causative factor using Combination FS-LR and ANN (FSLR-ANN) was conducted by eliminating one by one factor from 11 until eight factors with the least essential effect by the likelihood ratio test value from LR-FS analysis and make each susceptibility map used ANN method.

6 . 2. 3 Validation

During the modeling predictions, the most crucial component is to carry out a validation of the prediction result, and without some validation, the prediction model was useless and hardly any scientific significance (Chung and Fabbri, 2003). In this study, the landslide inventories were divided into two parts; one for training and the other for validation. 2,873 landslide pixels (70% of the landslide inventory) were used to produce the predictive models, and about 1,230 landslide pixels (30% of the landslide inventory) were used for validation purpose. Landslide inventory data was classified into training and validation datasets on a random basis from each landslide by taking the spatial distribution into account (Meten et al., 2015). Size, area, depth, and distribution of landslides vary from place to place. We also used the receiver operating characteristic (ROC) curve to plot predicted probabilities to understand the problem of accuracy, selection criteria, and interpretation. For validating the landslide susceptibility map, AUC curve was used as a measure of overall fit and comparison of the predictive model. The area determines the success rate under the curve (AUC) of the training dataset, and validation was possible from the AUC of the validation dataset. ROC curves were used to evaluate the predictive accuracy of the predictive accuracy of the selected statistical model like as logistic regression (Gorsevski et al., 2006). The AUC obtained from the ROC plot is the most preferred type that can influence the rating of model performance (Akgun et al., 2012). In this study, another possibility to validate the reliability of the predictive landslide susceptibility map from the training dataset was to overlay the validation landslides. This

procedure showed that most of landslides fall under high and very high susceptibility classes showing the high level of accuracy and performance of our models.

6. 3 Result and Discussion

6. 3. 1 Logistic regression

In Logistic regression (LR), it does little good to combine data with different measuring scales (Ayalew and Yamagishi, 2005), so to see the relationship between the landslide occurrence area and the landslide causative factors could be deduced from the relationship between areas where landslide had not occurred and the causative factors. Frequency ratio values show a correlation between landslides and each class of landslide causal factors in numerical format (Rasyid et al., 2016), so input data of independent variable for logistic regression method was using data frequency ratio in Table 21. In the present research, the highest frequency ratio was occurring in land use change (LUC) causative factor on LUC from primary forest to open area and paddy field (4-1) with ratio 8.70. Moreover, class of secondary forest to the farming area, savanna, scrub (4-2) had frequency ratio 2.20. The vegetation which causes this frequency ratio affects the stability of the slope. Forest clearance seems to have manifested primarily through increased rates of landslide activity because land with forest having the root system would reinforce the soil strength and stabilizes the slope (Glade, 2003; Hasnawir et al., 2015).

Hence, this study conducted ten tests in order to acquire the best result. Accordingly, the best validation result was found on the seventh test as can be seen from Table 22. LUC had a value of 0.589 (from the seventh test)) that affects landslide occurrence. Forest land with root system would reinforce the soil strength and stabilizes the slope to reduce surface erosion or shallow landslides (Hasnawir et al., 2015). The highest value of 3.081 shows the distance to the river having the most significant effect on landslide occurrence. Moreover, the lowest value of elevation (0.353) indicated a small effect on landslide occurrence in this research.

The logistic regression that was obtained from the seventh test can be expressed as follows:

Table 21 The value of Ratio and Normalized of Landslide Occurrences for each landslide causative factors

Factor	Class	Pixel Class*	% Class(a)	Landslide pixel**	% Landslide (b)	Frequency ratio (b/a)	Normalized of ratio
Elevation (meter)	<500	126010	15.85	0	0	0	0.000
	500 – 750	113821	14.31	0	0	0	0.000
	750 – 1000	117886	14.82	382	13.3	0.9	0.360
	1000 – 1250	99735	12.54	544	18.93	1.51	0.604
	1250 – 1500	80401	10.11	446	15.52	1.54	0.616
	1500 – 1750	73551	9.25	665	23.15	2.5	1.000
	1750 – 2000	62583	7.87	452	15.73	2	0.800
	2000 - 2250	63418	7.97	202	7.03	0.88	0.352
	2250 – 2500	38830	4.88	180	6.27	1.28	0.512
	>2500	18992	2.39	2	0.07	0.03	0.012
Slope (degree)	0 -10	277391	34.88	233	8.11	0.23	0.000
	10 – 20	193490	24.33	504	17.54	0.72	0.099
	20 – 30	142736	17.95	519	18.06	1.01	0.157
	30 – 40	114954	14.46	613	21.34	1.48	0.252
	40 – 50	56795	7.14	819	28.51	3.99	0.758
	>50	9861	1.24	185	6.44	5.19	1.000
Curvature	Concave	335269	42.16	1,614	56.18	1.33	1.000
	Flat	100826	12.68	158	5.5	0.43	0.000
	Convex	359132	45.16	1,101	38.32	0.85	0.467
Aspect	Flat	48980	6.16	43	1.5	0.24	0.000
	North	105139	13.22	963	33.52	2.54	1.000
	Northeast	140313	17.64	599	20.85	1.18	0.409
	East	128555	16.17	311	10.82	0.67	0.187
	Southeast	155292	19.53	191	6.65	0.34	0.043
	South	127354	16.01	515	17.93	1.12	0.383
	Southwest	48881	6.15	43	1.5	0.24	0.000
	West	11324	1.42	23	0.8	0.56	0.139
	Northwest	29389	3.7	185	6.44	1.74	0.652
Lithology	Qlvb	195818	24.62	0	0	0	0.000
	Qlv	562441	70.73	2,826	98.36	1.39	1.000
	Qvlc	36968	4.65	47	1.64	0.35	0.252
Distance to Faults (meter)	0 – 2500	228372	28.72	913	31.78	1.11	0.371
	2500 -5000	123498	15.53	1,333	46.4	2.99	1.000
	5000 – 7500	106243	13.36	472	16.43	1.23	0.411
	7500 – 10000	92127	11.58	155	5.4	0.47	0.157
	>10000	244987	30.81	0	0	0	0.000
Distance to River (meter)	0 - 100	325991	40.99	1,489	51.83	1.26	1.000
	100 – 200	240871	30.29	726	25.27	0.83	0.368
	200 – 300	139539	17.55	397	13.82	0.79	0.309
	300 – 400	59549	7.49	189	6.58	0.88	0.441
	400 – 500	19942	2.51	42	1.46	0.58	0.000
	>500	9335	1.17	30	1.04	0.89	0.456
Drainage Density (km/km2)	0 - 1	147677	18.57	698	24.3	1.31	1.000
	1 - 2	228100	28.68	635	22.1	0.77	0.000
	2 – 3	252005	31.69	829	28.85	0.91	0.259
	3 – 4	121676	15.3	512	17.82	1.16	0.722
	>4	45769	5.76	199	6.93	1.2	0.796
Precipitation (mm /year)	3187	186406	23.44	0	0	0	0.000
	3570	61646	7.75	84	2.92	0.38	0.270
	3739	547175	68.81	2,789	97.08	1.41	1.000
LUC (1=Open area, Paddy area; 2=Farming area, savanna, scrub; 3=Secondary Forest; 4=Primary Forest)	1 - 1	167966	21.12	608	21.16	1	0.100
	1 – 2	44883	5.64	276	9.61	1.7	0.182
	2 – 1	127015	15.97	134	4.66	0.29	0.018
	2 - 2	140425	17.66	215	7.48	0.42	0.033
	2 - 3	3971	0.5	2	0.07	0.14	0.000
	3 – 1	24542	3.09	157	5.46	1.77	0.190
	3 - 2	88061	11.07	513	17.86	1.61	0.172
	3 - 3	30715	3.86	158	5.5	1.42	0.150
	3 – 4	4602	0.58	26	0.9	1.56	0.166
	4 – 1	954	0.12	30	1.04	8.7	1.000
	4 – 2	19912	2.5	177	6.16	2.46	0.271
	4 – 3	55800	7.02	180	6.27	0.89	0.088
	4 – 4	86381	10.86	397	13.82	1.27	0.132
Distance to Road (meter)	0 – 500	407277	51.22	940	32.72	0.64	0.000
	500 – 1000	115348	14.51	449	15.63	1.08	0.171
	1000 – 1500	34878	4.39	405	14.1	3.21	1.000
	1500 – 2000	24877	3.13	167	5.81	1.86	0.475
	2000 – 2500	23831	3	105	3.65	1.22	0.226
	2500 – 3000	23799	2.99	189	6.58	2.2	0.607
	3000 – 3500	23266	2.93	101	3.52	1.2	0.218
	3500 – 4000	23477	2.95	110	3.83	1.3	0.257
	>4000	118474	14.9	407	14.17	0.95	0.121

$$Z_{seventh} = -10.175 + 0.332\,FR_{Elevation} + 0.332\,FR_{Slope} + 0.501\,FR_{Aspect} + 0.449\,FR_{Curvature} + 1.184\,FR_{Lithology} + 0.52\,FR_{Distance\ to\ faults} + 1.031\,FR_{Drainage\ Density} + 2.837\,FR_{Distance\ to\ river} + 0.734\,FR_{Precipitation} + 0.589\,FR_{LUC} - 0.148FR_{Distance\ to\ road}$$

where: $FR_{Elevation}$ is the frequency ratio value of elevation class. FR_{Slope} is the frequency ratio value of slope class. FR_{Aspect} is the frequency ratio value of slope aspect class. $FR_{Curvature}$ is the frequency ratio value of curvature class. $FR_{Lithology}$ is the frequency ratio value of lithology class. $FR_{Distance}$ to faults is the frequency ratio value of the distance to faults class. $FR_{Drainage\ Density}$ is the frequency ratio value of drainage density class. $FR_{Distance}$ to the river is the frequency ratio value of the distance to river class. $FR_{Precipitation}$ is the frequency ratio value of precipitation class. FR_{LUC} is the frequency ratio value of land use change class. $FR_{Distance\ to\ the\ road}$ is the frequency ratio value of the distance to road class. Then input $Z_{seventh}$ value into equation 1, and we got the probability of landslide occurrence with an index value from 0.001 to 0.999.

Table 22 Logistic regression coefficient of landslide causative factors using an equal proportion of landslide and non-landslide pixel

Number Test	Variable in the equation											
	Elevation	Slope	Aspect	Curvature	Lithology	Distance to Faults	Drainage Density	Distance to River	Precipitation	Distance to Road	LUC	Constant
1	0.261	0.593	0.571	0.429	1.453	0.469	0.977	2.84	0.597	-0.168	0.378	-9.97
2	0.248	0.574	0.576	0.408	1.37	0.457	0.653	2.989	0.839	-0.164	0.57	-10.177
3	0.195	0.593	0.525	0.503	1.439	0.573	0.862	3.05	0.749	-0.162	0.661	-10.661
4	0.401	0.554	0.526	0.445	1.219	0.489	0.794	2.632	0.722	-0.187	0.623	-9.825
5	0.351	0.561	0.498	0.634	1.535	0.483	0.901	2.624	0.638	-0.183	0.441	-10.126
6	0.22	0.617	0.513	0.554	1.374	0.491	0.786	2.92	0.708	-0.143	0.596	-10.275
7	0.332	0.548	0.501	0.449	1.184	0.52	1.031	2.837	0.734	-0.148	0.589	-10.175
8	0.314	0.538	0.507	0.473	1.167	0.546	0.684	2.647	0.801	-0.185	0.479	-9.561
9	0.383	0.572	0.568	0.379	1.235	0.5	0.733	2.738	0.644	-0.201	0.539	-9.674
10	0.312	0.545	0.478	0.539	1.212	0.484	0.97	2.775	0.76	-0.083	0.502	-10.088

The landslide susceptibility maps were reclassified using natural break classification system in 5 class. The outcome was an interpretable map showing increasing spatial possibility of future landslide incidence consisting of very low, low, moderate, high and very high susceptibility classes (Figure 35a).

The Hosmer–Lemeshow test revealed that the logistic regression equation's goodness of fit could be accepted if the significance of Chi-square is larger than 0.05. The value of Cox and Snell (R^2) and Nagelkerke (R^2) showed that the independent variables could explain the dependent variables (Table 23).

Table 23 Logistic regression model summary

-2 Log likelihood	Cox & Snell R^2	Nagelkerke R^2	Overall Percentage
5244.425[a]	.377	.503	78.2

a. Estimation terminated at iteration number 7 because parameter estimates changed by less than .001.

6. 3. 2 Artificial neural network

ANN was analyzed using normalized data in Table 21. This study carried out ten tests in order to acquire the best result, and this was found from the sixth test result analysis as can be seen from the sense of fairness and the normalized importance of each landslide causative factor as presented in Table 24. The final landslide susceptibility map was produced by multiplying each causative factor with the independent variable importance calculated through the ANN analysis, and then an overlay of these layers was performed using equation Equation 6.6;

$$LS\ ANN = \sum_{i=1}^{n} fw_i\, x_j \tag{6.6}$$

where LS ANN is the final landslide susceptibility map calculated for each pixel, fw_i is the weight of each causative factor and w_i, j is the normalized weight for the category j of factor i.

Table 24 The importance value derived from the artificial neural network (ANN)

Causative Factor	Number test of ANN									
	1	2	3	4	5	6	7	8	9	10
Elevation	0.057	0.073	0.053	0.075	0.052	0.045	0.067	0.041	0.072	0.060
Slope	0.231	0.197	0.207	0.178	0.193	0.192	0.181	0.166	0.203	0.178
Aspect	0.094	0.088	0.078	0.081	0.067	0.067	0.069	0.082	0.089	0.076
curvature	0.021	0.023	0.021	0.025	0.029	0.024	0.024	0.025	0.021	0.033
Lithology	0.094	0.094	0.101	0.101	0.123	0.128	0.091	0.068	0.088	0.107
Distance to Faults	0.125	0.125	0.131	0.123	0.114	0.115	0.128	0.140	0.123	0.116
Distance to River	0.142	0.137	0.137	0.133	0.121	0.131	0.124	0.116	0.144	0.125
Drainage Density	0.044	0.031	0.035	0.038	0.038	0.033	0.042	0.028	0.035	0.049
Precipitation	0.052	0.074	0.086	0.061	0.068	0.079	0.076	0.076	0.067	0.072
LUC	0.105	0.123	0.125	0.154	0.159	0.153	0.161	0.215	0.121	0.150
Distance to Road	0.036	0.036	0.029	0.032	0.036	0.034	0.036	0.042	0.038	0.034

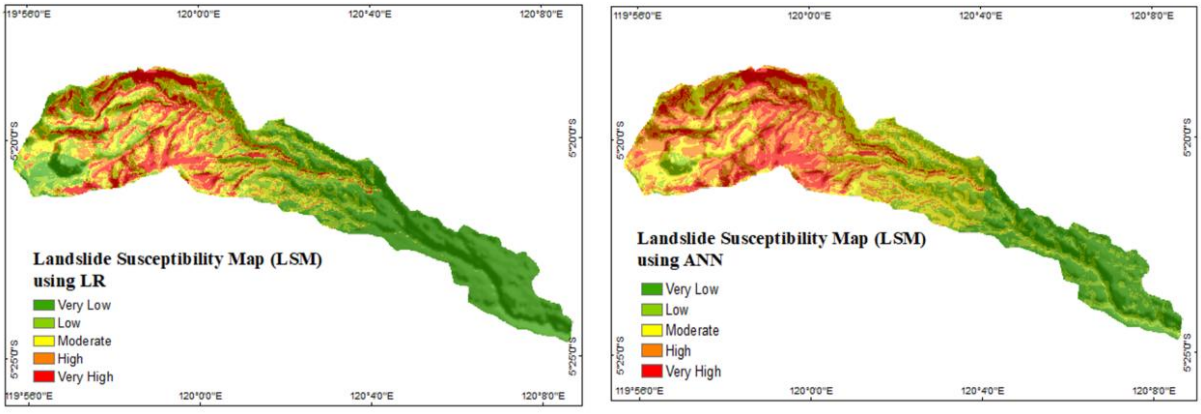


Figure 35 Landslide susceptibility maps (LSM). (a) LSM multivariate logistic on test seventh; (b) LSM artificial neural network (ANN) on sixth test models

The obtained ANN equation on test sixth is as follows:

$$LS\ ANN_{sixth} = 0.045\ ANN_{Elevation} + 0.192\ ANN_{Slope} + 0.067\ ANN_{Aspect} + 0.024\ ANN_{Curvature} + 0.128\ ANN_{Lithology} + 0.115\ ANN_{Distance\ to\ faults} + 0.033 ANN_{Drainage\ Density} + 0.131 ANN_{Distance\ to\ river} + 0.079\ ANN_{Precipitation} + 0.153\ ANN_{LUC} + 0.034\ ANN_{Distance\ to\ road}$$

The landslide susceptibility index values from LS ANN six ranges between 0.0026 to 0.8297 and this landslide susceptibility index map was reclassified into 5 classes using natural breaks classification system. The outcome was an interpretable map showing increasing spatial possibility of future landslide incidence ranging from very low, low, moderate, high and very high susceptibility to landslide (Figure 35b).

6. 3. 3 Comparison between LR and ANN

6. 3. 3. 1 The relationship between susceptibility maps and training/validation data

Landslide susceptibility maps were validated by comparing landslide areas to susceptibility classes that show the likelihood of landslide occurrence in a specific region. It was observed that the smaller goodness of fit in the low and very low susceptibility classes. The higher values of the goodness of fit were found to be in the high and very high susceptibility classes for the landslide susceptibility maps produced by the two models.

Analysis from logistic regression training dataset showed that about 82.67% of the total landslide pixels fall in the very high and high susceptibility classes while 17.33% of the total landslide pixels fall in the very low and moderate susceptibility classes (Figure 35a). The validation set for the LR model generated a good correlation with the occurrence of landslides. It is evident from the presence of landslides on high and very high susceptibility being 82.12% and 19.9% occurring in the very low, low and moderate susceptibility class. A reasonable degree of fit is obtained for the overlay analysis of the validation and training set for the LR-derived maps, which shows a constant increase from the very low to very high susceptibility classes.

Analysis data of the training and validation datasets over the ANN-derived susceptibility map showed better results among the susceptibility categories derived using other methods like

LR (Figure 35b). Among the training data set, 92.59% of the landslide pixels occur in the high and very high susceptibility classes and that only 7.41% in the very low, low and moderate susceptibility classes. However, in the case of validation data sets, 92.68% of the landslide pixels occur in the high and very high susceptibility classes while 7.32% of the landslide pixels occur in the very low, low and moderate susceptibility classes. It is evident that there is a variation between results produced by the validation and training datasets for the high and very high susceptibility classes showing a difference of 0.09%.

6. 3. 3. 2 Receiver Operating Curves

A standard validation analysis to compare prediction performance of various classifiers is the Receiver Operating Curve (ROC) and the calculation of Area Under Curve (AUC) (Akgun et al., 2012; Tien Bui et al., 2012). The ROC is a useful method for representing the quality of deterministic or probabilistic landslide susceptibility models. ROC graph is plotting pairs of sensitivity versus (1−specificity) at all possible values for the decision threshold when sensitivity and specificity are calculated non-parametrically, which represents the percent of correctly classified landslide pixels by the model against specificity, which is the proportion of predicted landslide pixels over the total study area. The AUC represents the quality of the models to predict the occurrence or the non-occurrence of landslides in a reliable manner.

A good fit model has an AUC value from 0.5 to 1. The ideal model performs an AUC value close to 1.0 (perfect fit), whereas a value close to 0.5 indicate inaccuracy in the model (random fit), (Carvalho et al., 2014). In general, the AUC of ROC curves representing excellent, good, and valueless tests were plotted on the graph. It classifies the accuracy of a diagnostic test i.e. the value ranges from 0.50 to 0.60 (fail), 0.60–0.70 (poor), 0.70–0.80 (fair), 0.80–0.90 (good), and 0.90–1.00 (excellent). Table 6 shows the ROC of LR and ANN models for the training and validation sets. The measurement of how well the model performs is represented in the success rate curve (training data) and the capability of the model to predict is represented in the prediction rate curve (validation data). It is observed that both LR and ANN models have

success rates expressed by their respective AUC values of 0.857 and 0.845. In the case of the prediction rate curve, LR and ANN model also show a similar result with AUC value of 0.856 and 0.844, respectively.

Based on each validation, it can be concluded that using ANN model produces a better landslide susceptibility map as compared to LR model. Aditian and Kubota (2017) has also proved that ANN model has superior accuracy and performance in landslide susceptibility mapping compared to LR and FR models. Artificial neural network (ANN) is useful for problem-solving and successfully applied in various science and engineering applications including the fields of landslide susceptibility, hazard, and risk mapping. ANN, is a machine learning method, has achieved more satisfactory results on landslide susceptibility mapping compared to other statistical models, e.g., logistic regression (Pradhan et al., 2010); Chauhan et al., 2010; Kanungo et al., 2006; Yilmaz, 2009).

6. 3. 4 Optimized Causative Factor

6. 3. 4. 1 Optimized using Logistic Regression (Forward stepwise) (FSLR)

Optimizing causative factors using FSLR helped us to eliminate the less effective causative factors using the likelihood ratio by observing the least coefficients from chi-square test and Akaike's information criterion (AIC) (Table 25). To produce LSM from LR model, first equation 2 was calculated using the coefficients of each causative factor followed by inserting the results of equation 6.2 in equation 6.1 (Table 26). Finally, the optimized 3 LSMs using LR model were obtained by using ten causative factors (Figure 37a), with nine causative factors (Figure 37b), and with eight causative factors (Figure 37c). Optimization using FSLR indicated more causative factors would give best optimization model.

Table 25 Likelihood Ratio Tests using Logistic Regression

Effect	Model Fitting Criteria	Likelihood Ratio Tests
	AIC of Reduced Model	Chi-Square
Slope	5433.213	417.229
Distance to River	5307.492	291.508
Distance to Faults	5179.995	164.011
Aspect	5144.506	128.522
LUC	5101.049	85.065
Lithology	5069.815	53.832
Drainage Density	5051.248	35.265
Precipitation	5043.878	27.894
Elevation	5038.540	22.557
Curvature	5029.877	13.893
Distance to Road	5023.388	7.405

Table 26 Coefficient of Each Causative Factor and Constant of Optimizing Causative Factor Using Forward Stepwise (Likelihood Ratio) Logistic Regression

Causative Factor	The coefficient value of causative factor		
	10	9	8
Elevation	0.271	0.271	eliminated
Slope	0.538	0.548	0.553
Aspect	0.489	0.501	0.515
Curvature	0.439	eliminated	eliminated
Lithology	1.196	1.206	1.302
Distance to Faults	0.495	0.493	0.556
Drainage Density	1.066	1.051	0.945
Distance to River	2.848	2.928	2.840
Precipitation	0.752	0.737	0.857
Distance to Road	eliminated	eliminated	eliminated
LUC	0.574	0.583	0.653
Constant	-10.248	-9.876	-9.771

Table 27 Important value of each causative factor and constant of optimizing causative factor using Artificial neural network (ANN)

Number	Causative Factor	Importance 11 causative factor by ANN		
		ten causative factor	nine causative factor	eight causative
1	Slope	0.187	0.221	0.260
2	Distance to River	0.141	0.138	0.185
3	Distance to faults	0.133	0.097	0.178
4	LUC	0.054	0.081	0.080
5	Lithology	0.083	0.074	0.074
6	Aspect	0.074	0.052	0.070
7	Precipitation	0.115	0.128	0.095
8	Elevation	0.149	0.156	0.059
9	Distance to Road	0.014	0.054	eliminated
10	Drainage Density	0.050	eliminated	eliminated
11	curvature	eliminated	eliminated	eliminated

6. 3. 4. 2 Optimized using artificial neural network (ANN)

Optimizing causative factors using ANN helped to eliminate causative factors with least ANN values as can be seen in Table 27. The causative factor eliminated in step one was

curvature; in step two were curvature and drainage density and in step 3 were curvature, drainage density, and distance to road. LSM from ANN model was produced using equation 5 for each step using the coefficient of each causative factor as can be seen in Table 27. By using the optimized causative factors, three LSMs were produced from the ANN model, i.e., LSM ANN using ten causative factors (Figure 37d), LSM ANN with nine causative factors (Figure 37e), and LSM ANN with eight causative factors (Figure 37f).

6. 3. 4. 3 Optimized using combination FSLR and ANN (FSLR-ANN)

Optimizing causative factors using a combination of FSLR and ANN technique can be done in two steps. Firstly, the least essential causative factors were eliminated until eight causative factors are remained by using FSLR and coefficient of the chi-square test, and Akaike's information criterion and secondly, landslide susceptibility maps were produced using ANN model in each step.

Table 28 Important value of each causative factor and constant of optimizing causative factor using Artificial neural network (ANN) after eliminating causative factor by Forwarding Stepwise Logistic Regression

Number	Causative Factor	Importance of causative factor by ANN		
		10	9	8
1	Slope	0.191	0.220	0.159
2	Distance to Fault	0.158	0.056	0.151
3	Distance to River	0.133	0.138	0.117
4	Precipitation	0.093	0.087	0.091
5	Lithology	0.091	0.065	0.106
6	Aspect	0.068	0.062	0.072
7	LUC	0.064	0.196	0.269
8	Drainage Density	0.041	0.030	0.035
9	Elevation	0.144	0.146	Eliminated
10	Curvature	0.017	Eliminated	Eliminated
11	Distance to Road	Eliminated	Eliminated	Eliminated

In the first step, distance to the road was eliminated. In the second step, distance to road and curvature were eliminated, and in the third step, distance to road, curvature, and elevation were eliminated using likelihood ratio with less coefficient of chi-square test and Akaike's information criterion(Table 25). LSM was produced using ANN method using equation 5 for

each step from the coefficient of each causative factor in Table 28. From the optimized causative factors, three LSMs were produced using ANN model, i.e., LSM FSLR-ANN with ten causative factors (Figure 37g), LSM FSLR-ANN with nine causative factors (Figure 37h), and LSM FSLR-ANN with eight causative factors (Figure 37i).

Optimization using FSLR-ANN model indicated collaboration between FSLR and ANN which give the best correlation of causative factor with landslide occurrence would result in the best optimization model likely in FSLR-ANN model with 9 causative factors.

Table 29 Validation data using Area under curve (AUC) values and percentage fall landslide occurrence on high and very high susceptibility class of the three landslide models for the training and validation dataset

Number	Optimized Landslide Susceptibility Model	Number of causative factors		
		10	9	8
AUC Value				
Training dataset				
1	Logistic regression	0.857	0.856	0.856
2	Artificial neural network	0.836	0.832	0.839
3	Combination FS-LR and ANN	0.838	0.847	0.839
Validation dataset				
1	Logistic regression	0.855	0.854	0.855
2	Artificial neural network	0.832	0.827	0.837
3	Combination FS-LR and ANN	0.835	0.844	0.8
Percentage Fall Landslide Occurrence into High and Very High Susceptibility Class (%)				
Training dataset				
1	Logistic regression	81.38	81.20	79.32
2	Artificial neural network	89.66	79.57	89.28
3	Combination FS-LR and ANN	88.97	91.09	92.69
Validation dataset				
1	Logistic regression	80.57	81.46	78.94
2	Artificial neural network	88.78	79.11	89.76
3	Combination FS-LR and ANN	87.97	91.30	90.08

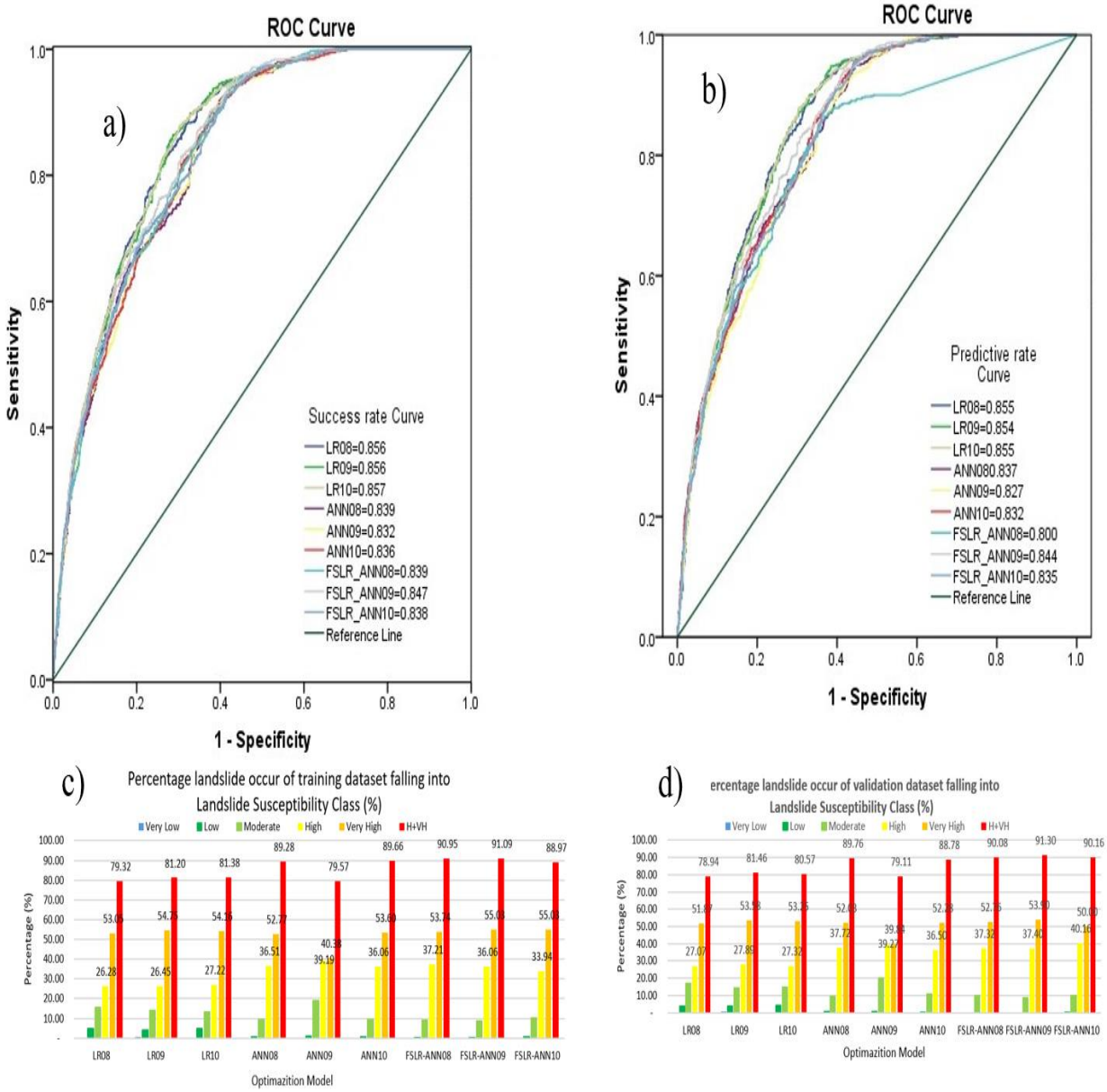


Figure 36 Validation data using Area under curve (AUC) and percentage of landslide occur falling into each class of landslide susceptibility of the three landslide models for the training and validation dataset. a) Success rate curve, b) Predictive rate curve, c) Percentage landslide occur of training dataset falling into Landslide Susceptibility Class (%) and d) Percentage landslide occur of validation dataset falling into Landslide Susceptibility Class (%).

6. 3. 4. 4 Models validation

In general, the AUC of ROC curves are produced from optimized causative factors of each model for both the training and validation datasets range between 0.8 and 0.9 (Table 29 and Figure 36) showing a high accuracy of prediction and validation. From overlay analysis, the comparison between validation and training datasets using FSLR-ANN susceptibility map shows the best result among all susceptibility categories, with an excellent validation accuracy in the range between 90% to 100% (Table 29 and Figure 36). In addition, the landslide susceptibility maps from LR and ANN models that were produced from their respective optimized causative factors, show landslide occurrence in the high and very high susceptibility classes for both the training and validation datasets with an AUC/ROC values ranging between 0.8 and 0.9 shown in Figure 36. However, the best landslide susceptibility map was obtained from the nine optimized causative factors (Figure 37h) using the combination method (FSLR-ANN) in which 91.09% of the training- and 91.3% of the validation landslide datasets are falling in the high and very high susceptibility classes.

In comparison of logistic regression with a bivariate statistics approach and found the logistic regression method to be the most accurate of these techniques like propose by Nandi and Shakoor(2010), Meten et al. (2015) and Rasyid et al. (2016), but several investigators have compared neural network models with logistic regression using different datasets with some researchers finding superior performance for the neural networks than logistic regression such as Yesilnacar and Topal (2005), Nefeslioglu et al. (2008), Yilmaz (2009) and Pradhan and Lee (2010b). In the current research, combine between Logistic regression and ANN model in this current research to optimization causative factor had given best result with 9 causative factor than the other models.

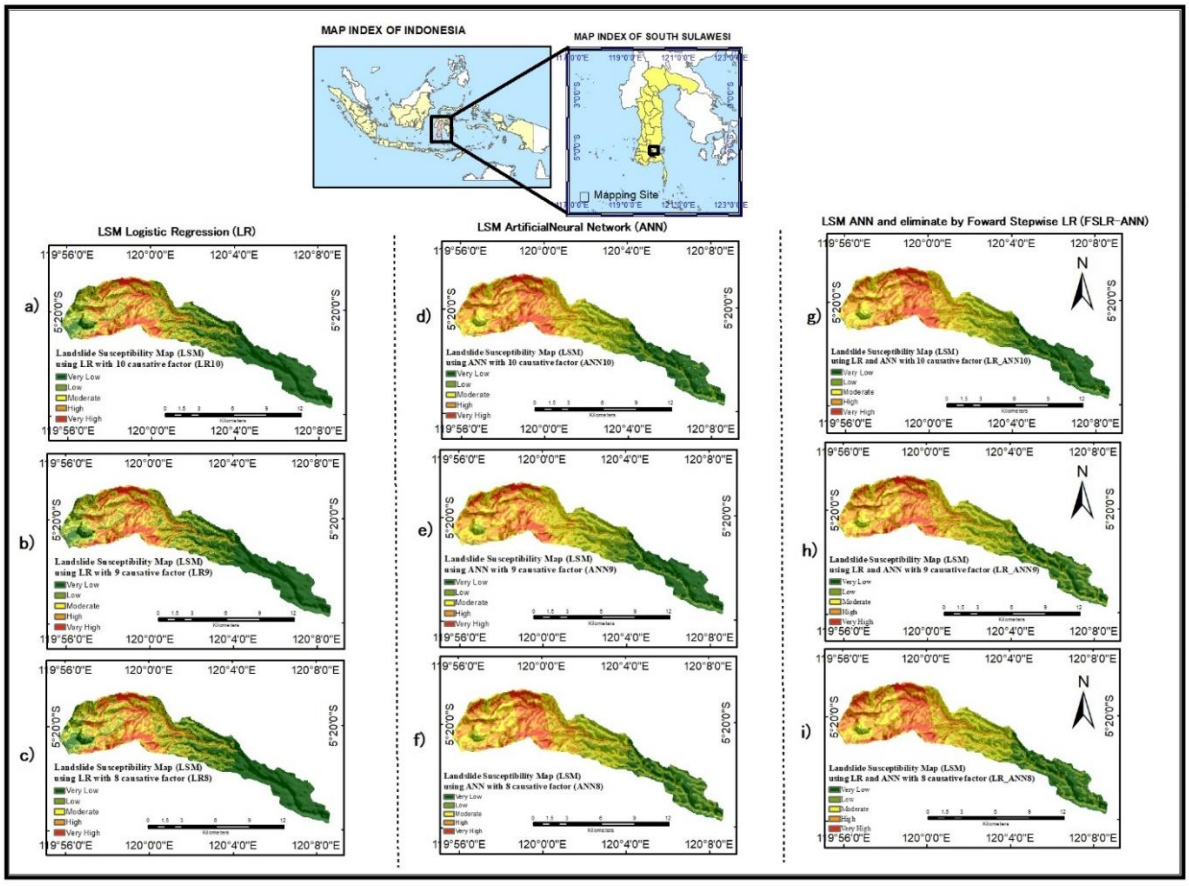


Figure 37 Landslide susceptibility map (LSM) of optimized causative factors using LR, ANN and combination FSLR-ANN. a) LSM LR with ten causative factor; b) LSM LR with nine causative factor; c) LSM LR with eight causative factor; d) LSM ANN with ten causative factor; e) LSM ANN with nine causative factor; f) LSM ANN with eight causative factor; g) LSM Combination FSLR-ANN with ten causative factor; h) LSM Combination FSLR-ANN with nine causative factor; i) LSM Combination FSLR-ANN with eight causative factor

6. 4 Conclusion

The preparation of accurate landslide susceptibility maps and landslide susceptibility assessments according to specific susceptibility indices is essential for supporting disaster warning systems. Studies have suggested numerous statistical methods and landslide causative factors to optimize when producing landslide susceptibility maps. In this study, the effectiveness of landslide susceptibility map production was assessed using Artificial Neural Network (ANN) and Logistic Regression (LR). We found that ANN outperforms LR when producing landslide susceptibility maps with eleven causative factors. Furthermore, optimizing of causative factors was done using three distinct methods: FSLR, ANN, and a combination of

the these (FSLR-ANN). The causative factors were optimized using each method by eliminating factors one by one, from 11 factors to eight, eliminated with the least value of correlation of causative factors to landslide occurrence, then producing a landslide susceptibility map using each method. Optimizing of causative factors using the combination FS-LR and ANN (FSLR-ANN) model with nine causative factors (slope, distance to faults, distance to river, precipitation, lithology, aspect, land use change (LUC), drainage density, and elevation) gave the best result. In conclusion, Artificial Neural Network (ANN) was the best method to produce landslide susceptibility map. The best Optimization of Causative Factors was a combination of FSLR -ANN with nine causative factors with AUC success rate 0.847, predictive rate 0.844 and validation with landslide fall into high and very high class with 91.30%. For this reason, this is an encouraging preliminary model towards a systematic introduction of FSLR-ANN model for optimization causative factors in landslide susceptibility assessment in the mountainous area of Ujung Loe Watershed.

6. 5 References

Akgun, A., Sezer, E.A., Nefeslioglu, H.A., Gokceoglu, C., Pradhan, B., 2012. An easy-to-use MATLAB program (MamLand) for the assessment of landslide susceptibility using a Mamdani fuzzy algorithm. Comput. Geosci. 38, 23–34. doi:10.1016/j.cageo.2011.04.012

Aleotti, P., Chowdhury, R., 1999. Landslide hazard assessment: summary review and new perspectives. Bull. Eng. Geol. Environ. 58, 21–44. doi:10.1007/s100640050066

Ayalew, L., Yamagishi, H., 2005. The application of GIS-based logistic regression for landslide susceptibility mapping in the Kakuda-Yahiko Mountains , Central Japan 65, 15–31. doi:10.1016/j.geomorph.2004.06.010

Ayalew, L., Yamagishi, H., Marui, H., Kanno, T., 2005. Landslides in Sado Island of Japan: Part II. GIS-based susceptibility mapping with comparisons of results from two methods and verifications. Eng. Geol. 81, 432–445. doi:10.1016/j.enggeo.2005.08.004

Bahsan, E., Liao, H.J., Ching, J., Lee, S.W., 2014. Statistics for the calculated safety factors of undrained failure slopes. Eng. Geol. 172, 85–94. doi:10.1016/j.enggeo.2014.01.005

Ballabio, C., Sterlacchini, S., 2012. Support Vector Machines for Landslide Susceptibility Mapping: The Staffora River Basin Case Study, Italy. Math. Geosci. 44, 47–70. doi:10.1007/s11004-011-9379-9

Cardinali, M., Reichenbach, P., Guzzetti, F., Ardizzone, F., Antonini, G., Galli, M., Cacciano, M., Castellani, M., Salvati, P., 2002. A geomorphological approach to the estimation of landslide hazards and risks in Umbria, Central Italy. Nat. Hazards Earth Syst. Sci. 2, 57–72. doi:10.5194/nhess-2-57-2002

Chau, K.T., Chan, J.E., 2005. Regional bias of landslide data in generating susceptibility maps using logistic regression: Case of Hong Kong Island. Landslides 2, 280–290. doi:10.1007/s10346-005-0024-x

Chauhan, S., Sharma, M., Arora, M.K., Gupta, N.K., 2010. Landslide susceptibility zonation through ratings derived from artificial neural network. Int. J. Appl. Earth Obs. Geoinf. 12, 340–350. doi:10.1016/j.jag.2010.04.006

Chen, W., Pourghasemi, H.R., Naghibi, S.A., 2017. A comparative study of landslide susceptibility maps produced using support vector machine with different kernel functions and entropy data mining models in China. Bull. Eng. Geol. Environ. 1–18. doi:10.1007/s10064-017-1010-y

Chung, C.-J.F., Fabbri, A.G., 2003. Validation of Spatial Prediction Models for Landslide Hazard Mapping. Nat. Hazards 30, 451–472. doi:10.1023/B:NHAZ.0000007172.62651.2b

Conforti, M., Pascale, S., Robustelli, G., Sdao, F., 2014. Evaluation of prediction capability of the artificial neural networks for mapping landslide susceptibility in the Turbolo River catchment (northern Calabria, Italy). Catena 113, 236–250. doi:10.1016/j.catena.2013.08.006

Demicco, R. V., Klir, G.J., 2004. Fuzzy Logic in Geology. Elsevier Science (USA), California.

Dou, J., Bui, D.T., Yunus, A.P., Jia, K., Song, X., Revhaug, I., Xia, H., Zhu, Z., 2015. Optimization of causative factors for landslide susceptibility evaluation using remote sensing and GIS data in parts of Niigata, Japan. PLoS One 10. doi:10.1371/journal.pone.0133262

Ercanoglu, M., Gokceoglu, C., 2004. Use of fuzzy relations to produce landslide susceptibility map of a landslide prone area (West Black Sea Region, Turkey). Eng. Geol. 75, 229–250. doi:10.1016/j.enggeo.2004.06.001

Ermini, L., Catani, F., Casagli, N., 2005. Artificial Neural Networks applied to landslide susceptibility assessment. Geomorphology 66, 327–343. doi:10.1016/j.geomorph.2004.09.025

Glade, T., 2003. Landslide occurrence as a response to land use change: A review of evidence from New Zealand. Catena 51, 297–314. doi:10.1016/S0341-8162(02)00170-4

Gorsevski, P. V., Gessler, P.E., Boll, J., Elliot, W.J., Foltz, R.B., 2006. Spatially and temporally distributed modeling of landslide susceptibility. Geomorphology 80, 178–198. doi:10.1016/j.geomorph.2006.02.011

Hasnawir, Kubota, T., Sanchez-castillo, L., Soma, A.S., 2015. Root Strength of Understory Vegetation For Erosion Control on Forest Slopes of Kelara Watershed, Indonesia, in: Proceedings of the 2nd Makassar International Conference on Civil Engineering(MICCE 2015) Makassar, Indonesia August 11-12,2015.

Jamsawang, P., Voottipruex, P., Boathong, P., Mairaing, W., Horpibulsuk, S., 2015. Three-dimensional numerical investigation on lateral movement and factor of safety of slopes stabilized with deep cement mixing column rows. Eng. Geol. 188, 159–167. doi:10.1016/j.enggeo.2015.01.017

Kanungo, D.P., Arora, M.K., Sarkar, S., Gupta, R.P., 2006. A comparative study of conventional, ANN black box, fuzzy and combined neural and fuzzy weighting procedures for landslide susceptibility zonation in Darjeeling Himalayas. Eng. Geol. 85, 347–366. doi:10.1016/j.enggeo.2006.03.004

Lee, S., Talib, J.A., 2005. Probabilistic landslide susceptibility and factor effect analysis. Environ. Geol. 47, 982–990. doi:10.1007/s00254-005-1228-z

Lin, G., Chang, M., Huang, Y., Ho, J., 2017. Assessment of susceptibility to rainfall-induced landslides using improved self-organizing linear output map , support vector machine , and logistic regression. Eng. Geol. 224, 62–74. doi:10.1016/j.enggeo.2017.05.009

Meinhardt, M., Fink, M., Tünschel, H., 2015. Landslide susceptibility analysis in central Vietnam based on an incomplete landslide inventory: Comparison of a new method to calculate weighting factors by means of bivariate statistics. Geomorphology. doi:10.1016/j.geomorph.2014.12.042

Meten, M., Bhandary, N.P., Yatabe, R., 2015. GIS-based frequency ratio and logistic regression

modelling for landslide susceptibility mapping of Debre Sina area in central Ethiopia. J. Mt. Sci. 12, 1355–1372. doi:10.1007/s11629-015-3464-3

Nandi, A., Shakoor, A., 2010. A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng. Geol. 110, 11–20. doi:10.1016/j.enggeo.2009.10.001

Nefeslioglu, H.A., Gokceoglu, C., Sonmez, H., 2008. An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng. Geol. doi:10.1016/j.enggeo.2008.01.004

Pham, B.T., Tien Bui, D., Prakash, I., Dholakia, M.B., 2017. Hybrid integration of Multilayer Perceptron Neural Networks and machine learning ensembles for landslide susceptibility assessment at Himalayan area (India) using GIS. Catena 149, 52–63. doi:10.1016/j.catena.2016.09.007

Pourghasemi, H.R., Jirandeh, A.G., Pradhan, B., Xu, C., Gokceoglu, C., 2013. Landslide susceptibility mapping using support vector machine and GIS at the Golestan Province, Iran. J. Earth Syst. Sci. Indian Acad. Sci. 122, 349–369. doi:10.1007/s12040-013-0282-2

Pradhan, B., Lee, S., 2010a. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ. Model. Softw. 25, 747–759. doi:10.1016/j.envsoft.2009.10.016

Pradhan, B., Lee, S., 2010b. Landslide susceptibility assessment and factor effect analysis: backpropagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling. Environ. Model. Softw. 25, 747–759. doi:10.1016/j.envsoft.2009.10.016

Pradhan, B., Lee, S., Buchroithner, M.F., 2010. A GIS-based back-propagation neural network model and its cross-application and validation for landslide susceptibility analyses. Comput. Environ. Urban Syst. 34, 216–235. doi:10.1016/j.compenvurbsys.2009.12.004

Rasyid, A.R., Bhandary, N.P., Yatabe, R., 2016. Performance of frequency ratio and logistic regression model in creating GIS based landslides susceptibility map at Lompobattang Mountain, Indonesia. Geoenvironmental Disasters 3, 19. doi:10.1186/s40677-016-0053-x

Shahabi, H., Ahmad, B.B., Khezri, S., 2013. Evaluation and comparison of bivariate and multivariate statistical methods for landslide susceptibility mapping (case study: Zab basin). Arab. J. Geosci. 6, 3885–3907. doi:10.1007/s12517-012-0650-2

Sheela, K.G., Deepa, S.N., 2013. Review on methods to fix number of hidden neurons in neural networks. Math. Probl. Eng. 2013. doi:10.1155/2013/425740

Shirzadi, A., Saro, L., Hyun Joo, O., Chapi, K., 2012. A GIS-based logistic regression model in rock-fall susceptibility mapping along a mountainous road: Salavat Abad case study, Kurdistan, Iran. Nat. Hazards 64, 1639–1656. doi:10.1007/s11069-012-0321-3

Soma, A.S., Kubota, T., 2017a. The Performance of Land Use Change Causative Factor on Landslide Susceptibility Map in Upper Ujung-Loe Watersheds South Sulawesi, Indonesia. Geoplanning J. Geomatics Plan. 4, 157–170. doi:10.14710/geoplanning.4.2.157-170

Soma, A.S., Kubota, T., 2017b. Land Use Changes on the Slopes and the Implications for the Landslide Occurrences in Ujung-Loe Watersheds South Sulawesi Indonesia. Int. J. Ecol. Dev. 32, 33–42.

Tien Bui, D., Nguyen, Q.P., Hoang, N.D., Klempe, H., 2017. A novel fuzzy K-nearest neighbor inference model with differential evolution for spatial prediction of rainfall-induced shallow landslides in a tropical hilly area using GIS. Landslides 14, 1–17. doi:10.1007/s10346-016-0708-4

Yao, X., Tham, L.G., Dai, F.C., 2008. Landslide susceptibility mapping based on Support Vector Machine: A case study on natural slopes of Hong Kong, China. Geomorphology 101, 572–582. doi:10.1016/j.geomorph.2008.02.011

Yesilnacar, E., Topal, T., 2005. Landslide susceptibility mapping: A comparison of logistic regression and neural networks methods in a medium scale study, Hendek region (Turkey). Eng. Geol. 79, 251–266. doi:10.1016/j.enggeo.2005.02.002

Yilmaz, I., 2009. Landslide susceptibility mapping using frequency ratio, logistic regression, artificial neural networks and their comparison: A case study from Kat landslides (Tokat-Turkey). Comput. Geosci. 35, 1125–1138. doi:10.1016/j.cageo.2008.08.007

Chapter 7 Conclusions and Future Works

7. 1 **Conclusions**

The increasing development in South Sulawesi as a gate to the east of Indonesia will increase pressure to the population and economic which will bring the finding of this study to the most relevance. Due to this, land use change as one impact of exploitation of the natural resources impacted to the landslide evident. In South Sulawesi, the land use change was many cases located in the mountainous area, like in current research in upper Ujung-loe Watersheds. Increasing the land use change in mountainous area affects the slope stability and landslide occurrence. It is also worthy to be mentioned that landslide susceptibility study is still limited in the study area as a result of many unreported landslide cases. Due to this, current research conducted assessment on the close correlation of land use change with landslide occurrence and the performance of LUC to produce landslide susceptibility using geographic information system and multivariate qualitative prediction model

The produced landslide susceptibility map is expected to be useful for government officials and urban planner in planning the development of the region. Geographic information systems are tools in disaster risk reduction planning which is one of the main issues of Indonesia's national development agenda to promote sustainable development and reduce the frequency of disasters and environmental degradation. Regarding landslide disaster risk, reduction and disaster mitigation are well approached by landslide susceptibility, hazard, and risk.

Based on the previous chapters and the subsequent discussions the following significant conclusions are presented:

1. Significant land use changes from 2004 to 2011 observed in the Ujung-Loe watershed that experienced a decline were no vegetation and dense vegetation classes, while those that have increased are medium vegetation and high vegetation classes. Landslides have

occurred 128 times during 2012 to 2014, and the most frequently occurred in 2013 and is mostly dominated by the one with the land use change from high vegetation to medium vegetation. The general land use change in Ujung Loe watershed indicates significant effect to landslides occurrence and slope instability.

2. Using land use change (LUC) as a new causative factor to produce landslide susceptibility map (LSM), LUC has a good effect. The result indicated that producing LSM with LUC was better than without LUC. Performances of each landslide model were tested using AUC curve for success and predictive rate, which had the highest value of predictive rate with LUC in both frequency ratio (FR) and logistic regression (LR) method (83.4% and 85,2%, respectively) and 80.24% of landslides validation fell in the class of high to very high. These results suggested that changing the vegetation to another landscape causes slopes unstable and increases the probability of landslide occurrence. LR method is better than FR to produce LSM. LUC affects landslide susceptibility in the study area; it was observed that the change in vegetation type to another landscape destabilized slopes. Validation of landslide susceptibility was carried out by calculating the area under the curve (AUC) of receiver operating characteristic curve (ROC). Firstly, LR shows the highest accuracy in both success and predictive rate (85.6%). Secondly, the frequency of landslides in high to a very high class of susceptibility was calculated, which indicates the level of accuracy of the method. CF returns the highest accuracy of 85.28 %.
3. Using LUC have an excellent effect to produce landslide susceptibility map. When we use FR and certainty factor (CF), LUC has the highest value on both at LUC from primary forest to open area and paddy field. In logistic regression method, LUC has the effect of landslide occurrence with significant value 0.589. Besides creating landslide susceptibility maps, this research illustrates the performance of frequency ratio (FR), certainty factor (CF) and logistic regression (LR) models as well. Two-steps of

validation were carried out in this study. First, performances of each landslide model were tested using AUC curve for success and predictive rate, which is more than 82 % with the highest at LR Model. In the second, the ratio of landslides falling on high to a very high class of susceptibility was obtained, which indicates the level of accuracy of the model. The CF model have highest accuracy with 85.28 % landslides fall in the range of high to very high class while in LR and FR model, it is 82.11% and 81.46%.

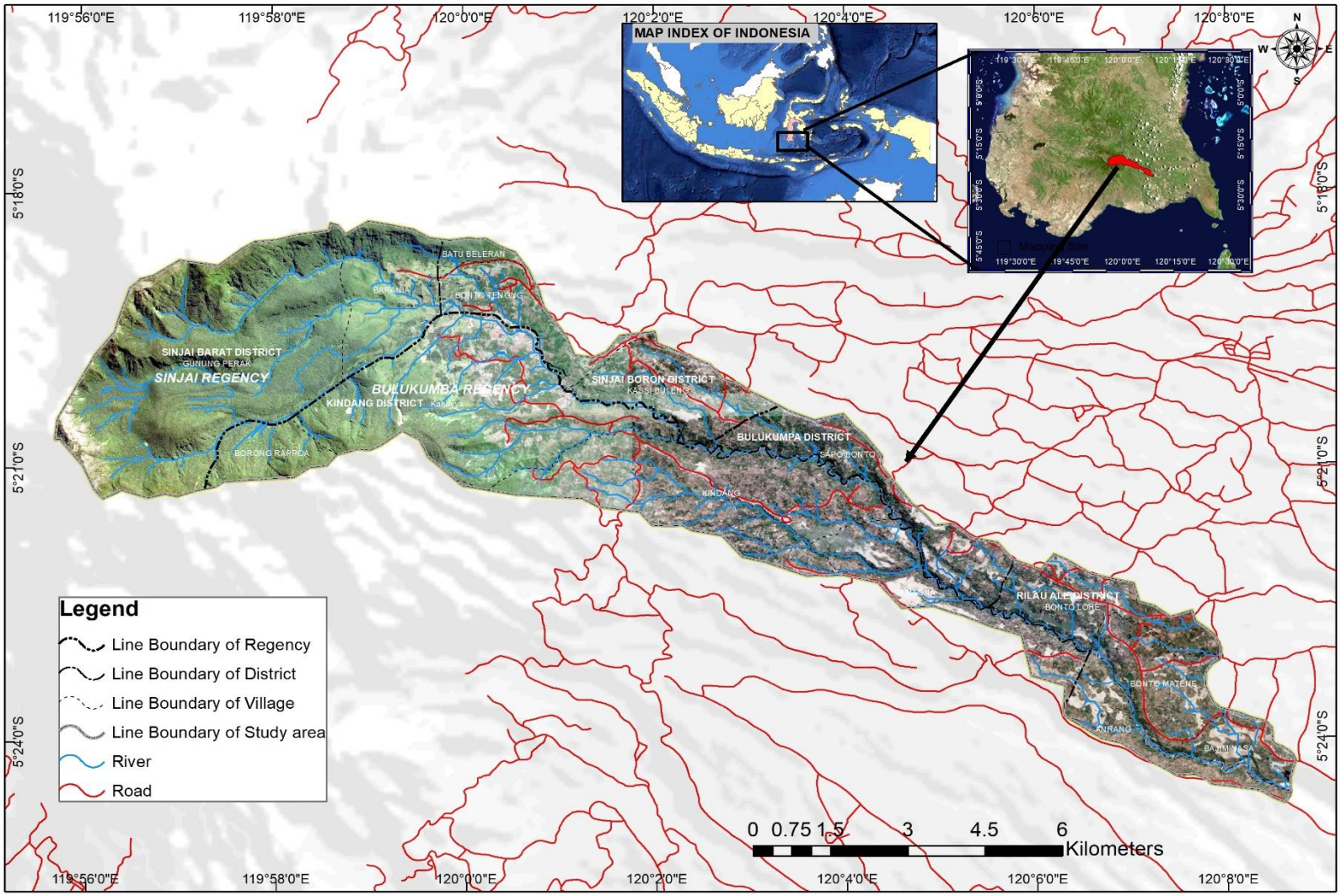
4. Artificial Neural Network (ANN) was the best method to produce landslide susceptibility map. The best optimization of causative factors was a combination of forward stepwise logistic regression (FSLR) - ANN with nine causative factors with AUC success rate 0.847, predictive rate 0.844 and validation with landslide fall into high and very high class with 91.30%. For this reason, this is an encouraging preliminary model towards a systematic introduction of FSLR-ANN model for optimization causative factors in landslide susceptibility assessment in the mountainous area of Ujung Loe Watershed.

7. 2 **Future works**

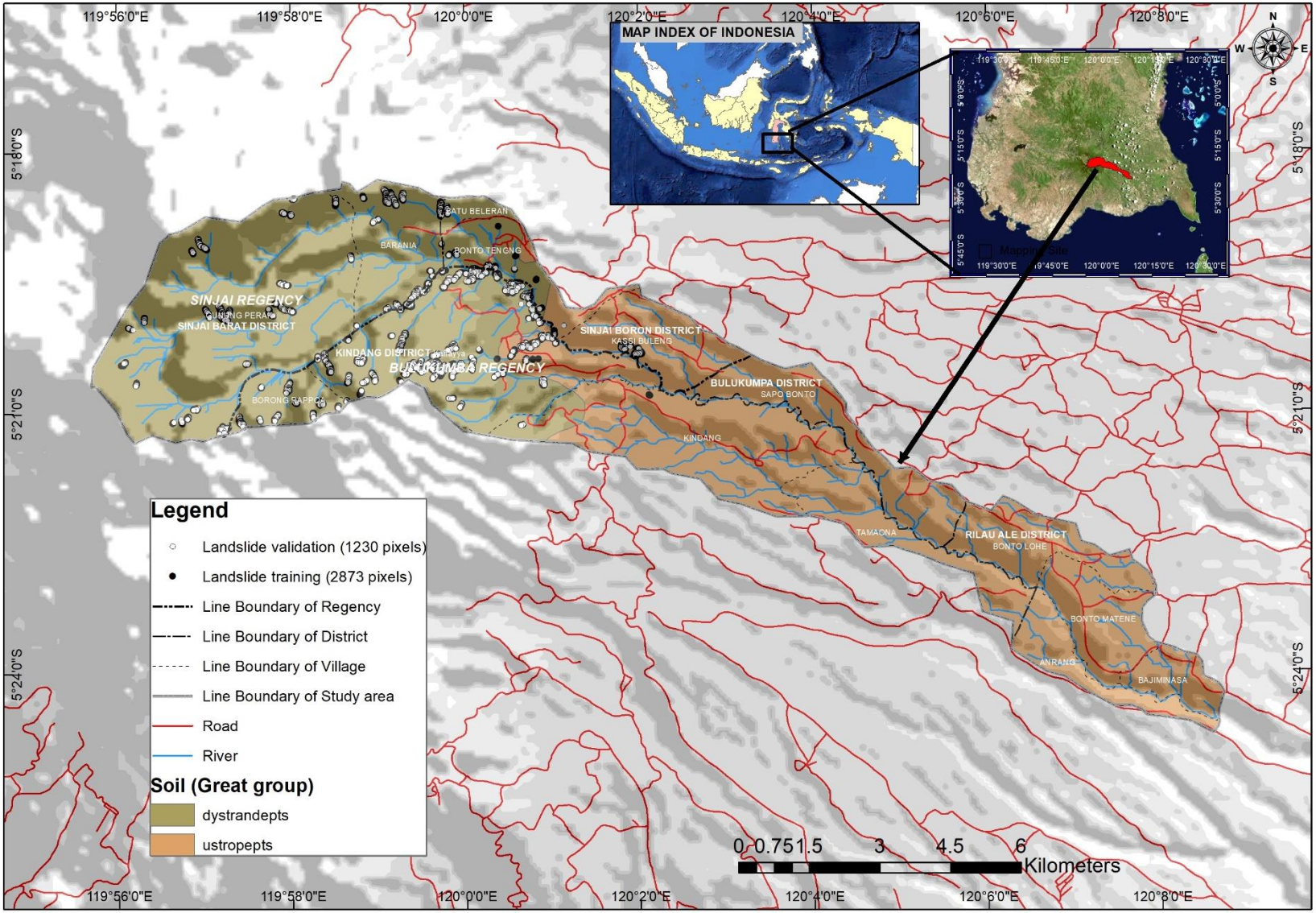
Based on the present study, further improvement could be included such as:

1. Improving the landslide inventory as the base data for landslide susceptibility assessments in the study area by using remote sensing analysis. Landslide inventory is one of the key input in landslide susceptibility mapping.
2. Further studies should also employ the high accuracy of DTM, i.e., LIDAR data to obtain better accuracy of simulation and zoning in a landslide.

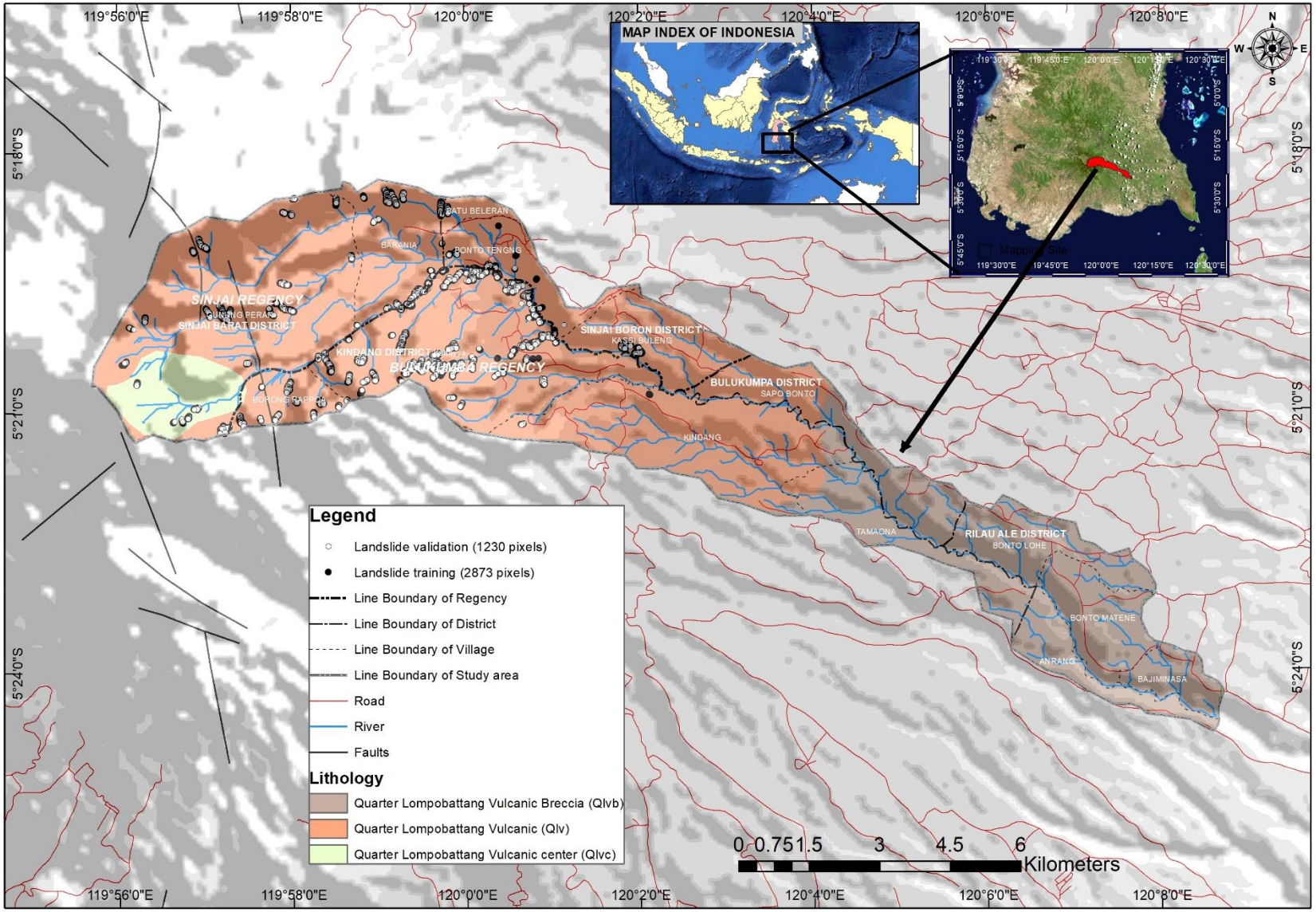
Appendices



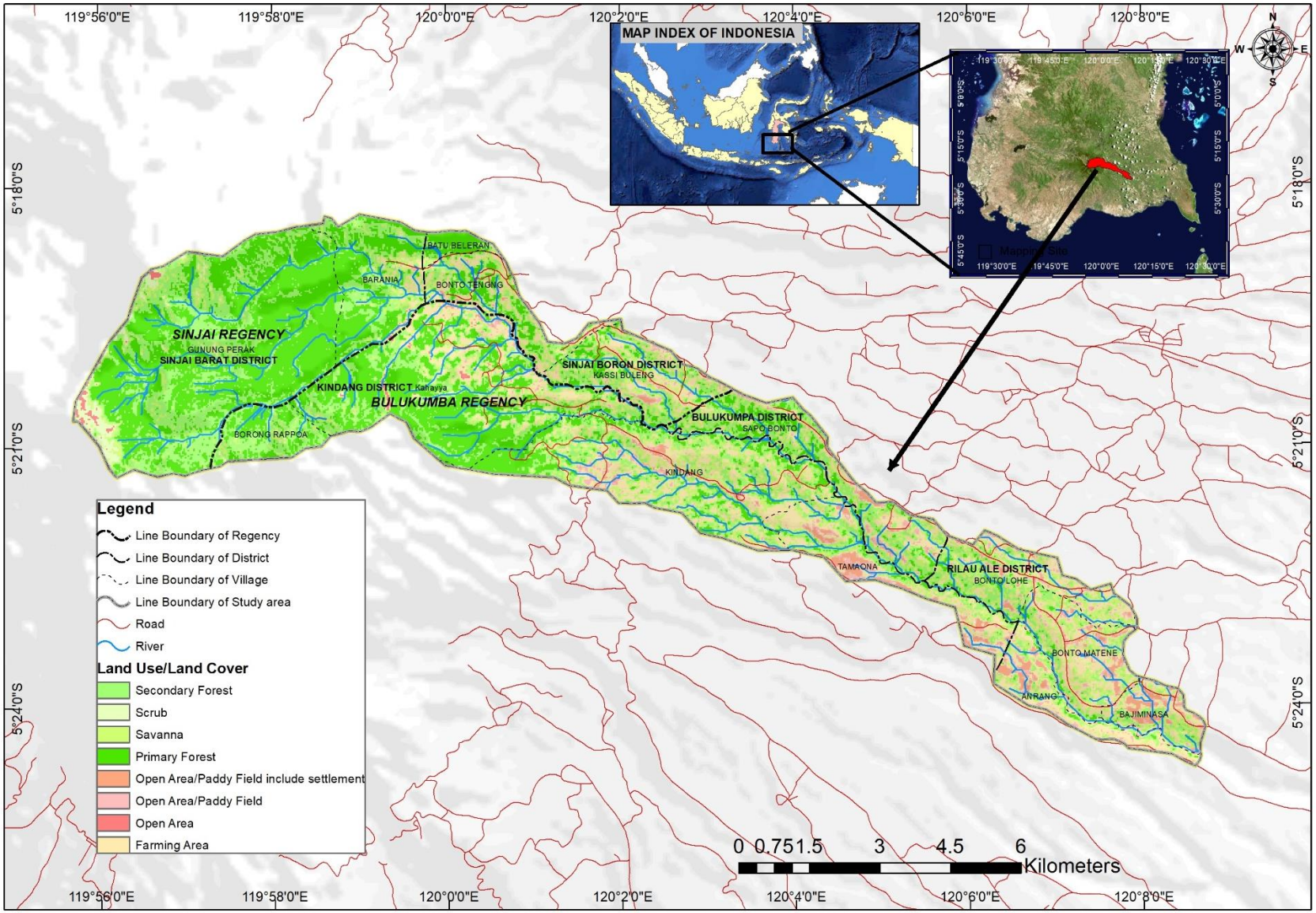
Appendix 1. Study area in Ujung Loe Watershed, South Sulawesi, Indonesia



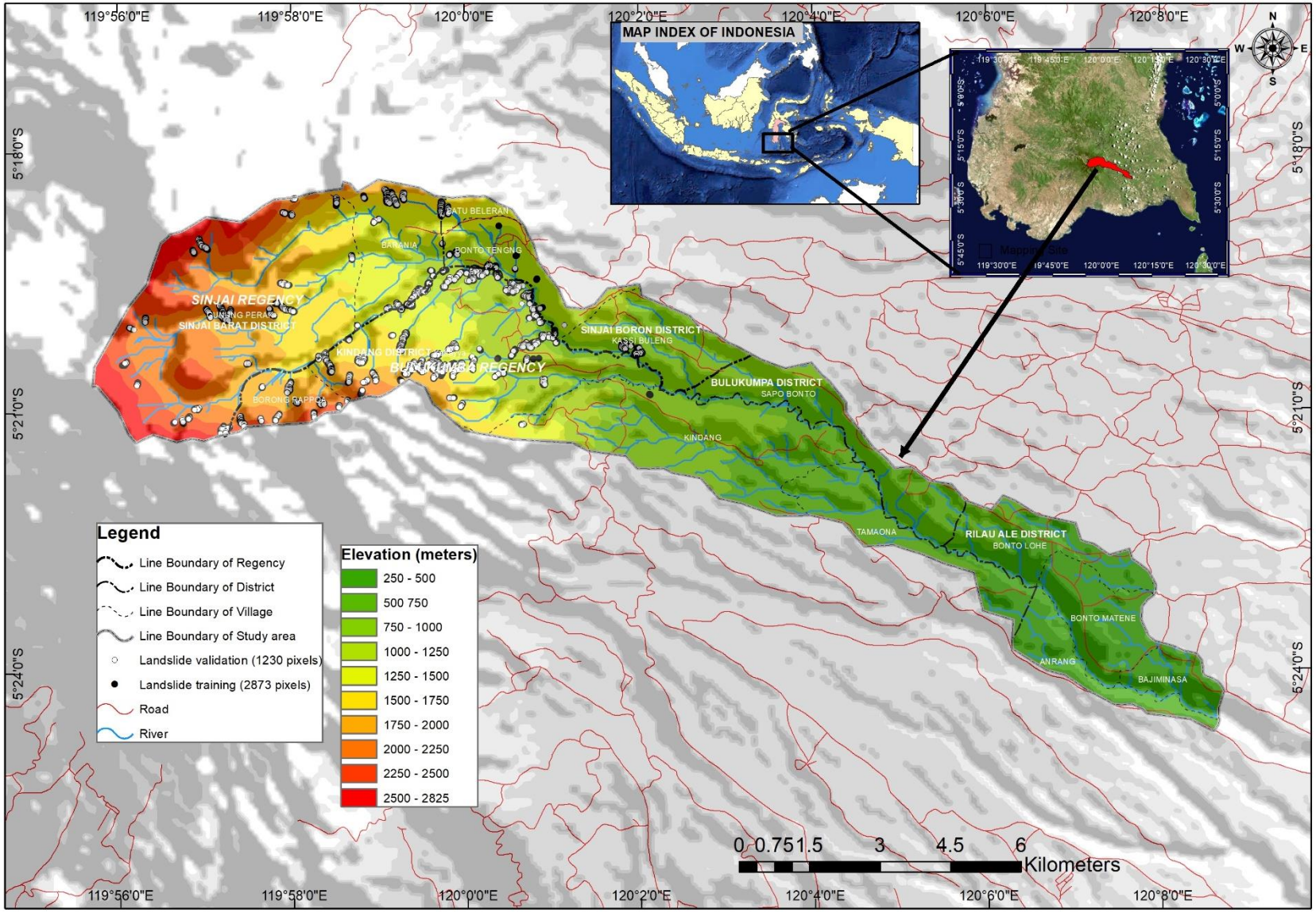
Appendix 2. Soil map in Ujung Loe Watershed, South Sulawesi, Indonesia



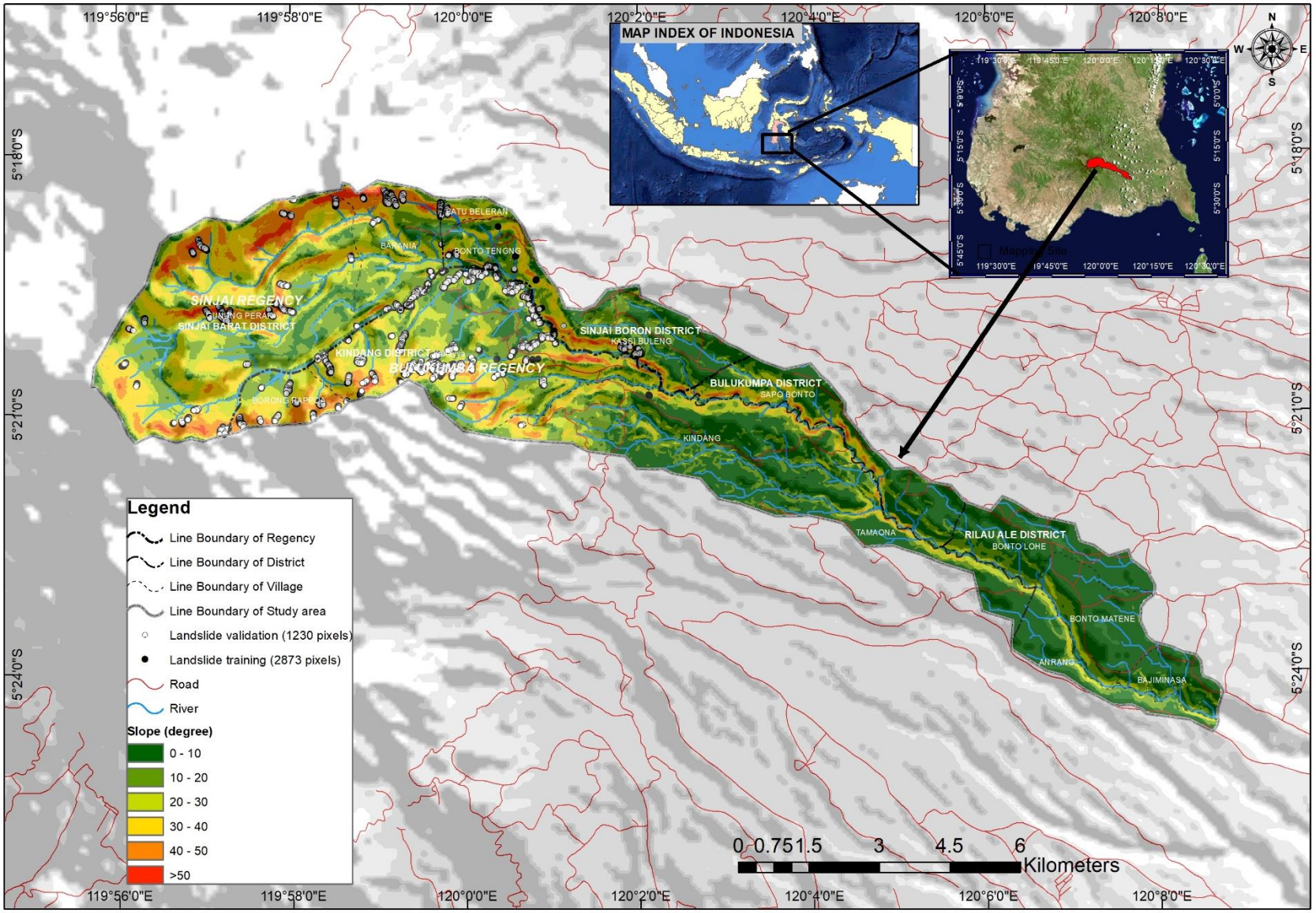
Appendix 3. Geology map in Ujung Loe Watershed, South Sulawesi, Indonesia



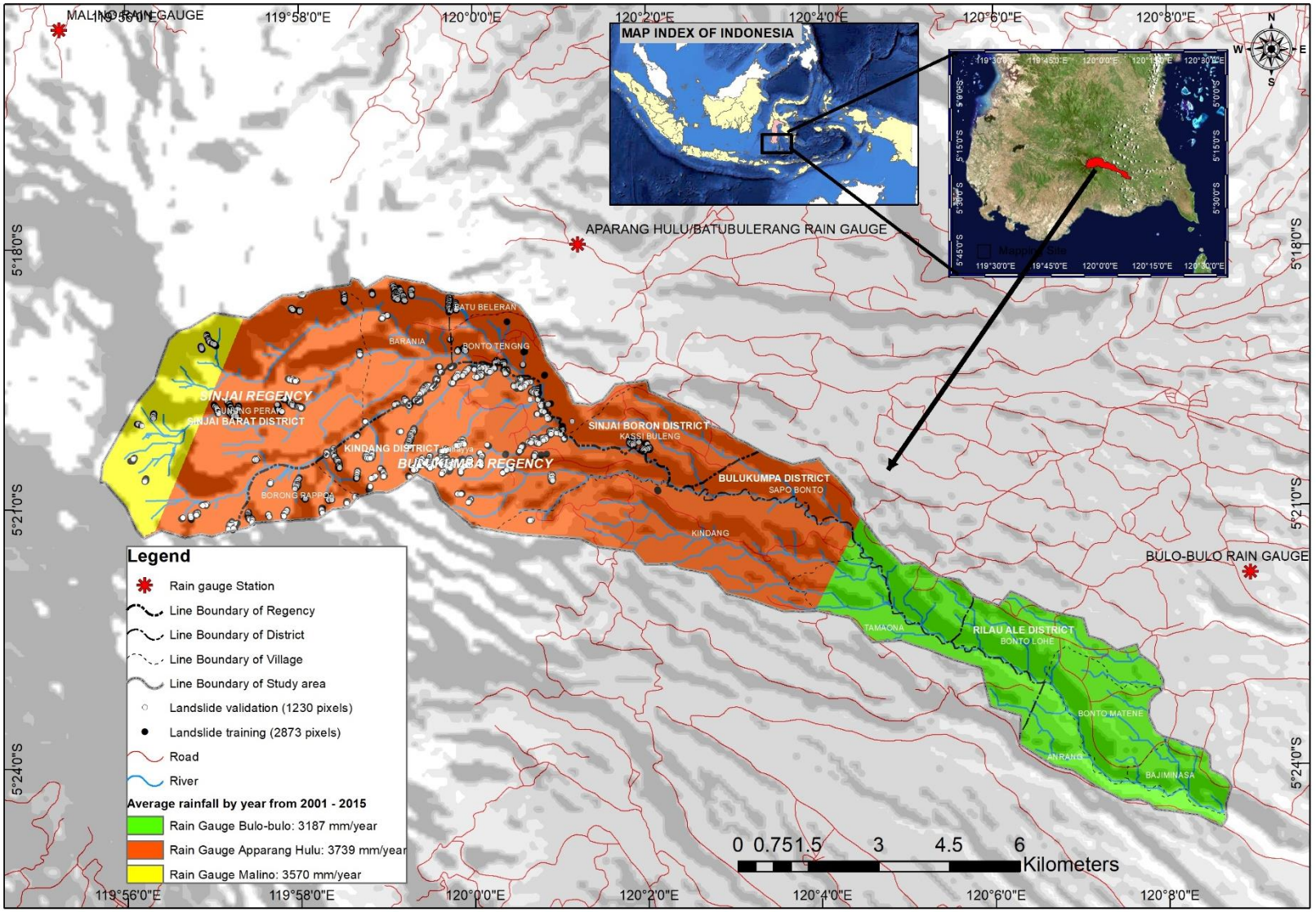
Appendix 4. Land use/land cover in 2015 at Ujung Loe Watershed, South Sulawesi, Indonesia



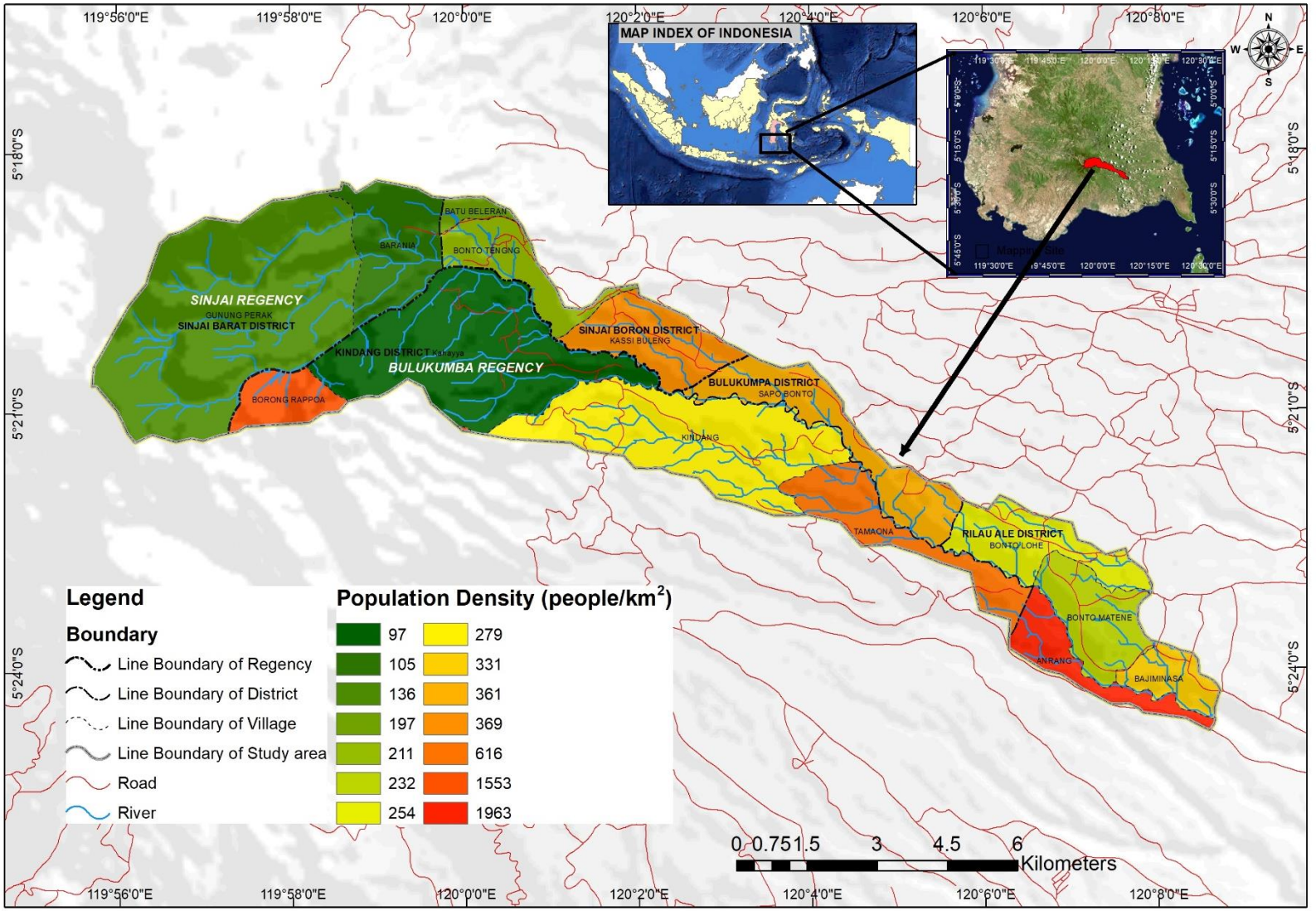
Appendix 5. Elevation map in Ujung Loe Watershed, South Sulawesi, Indonesia



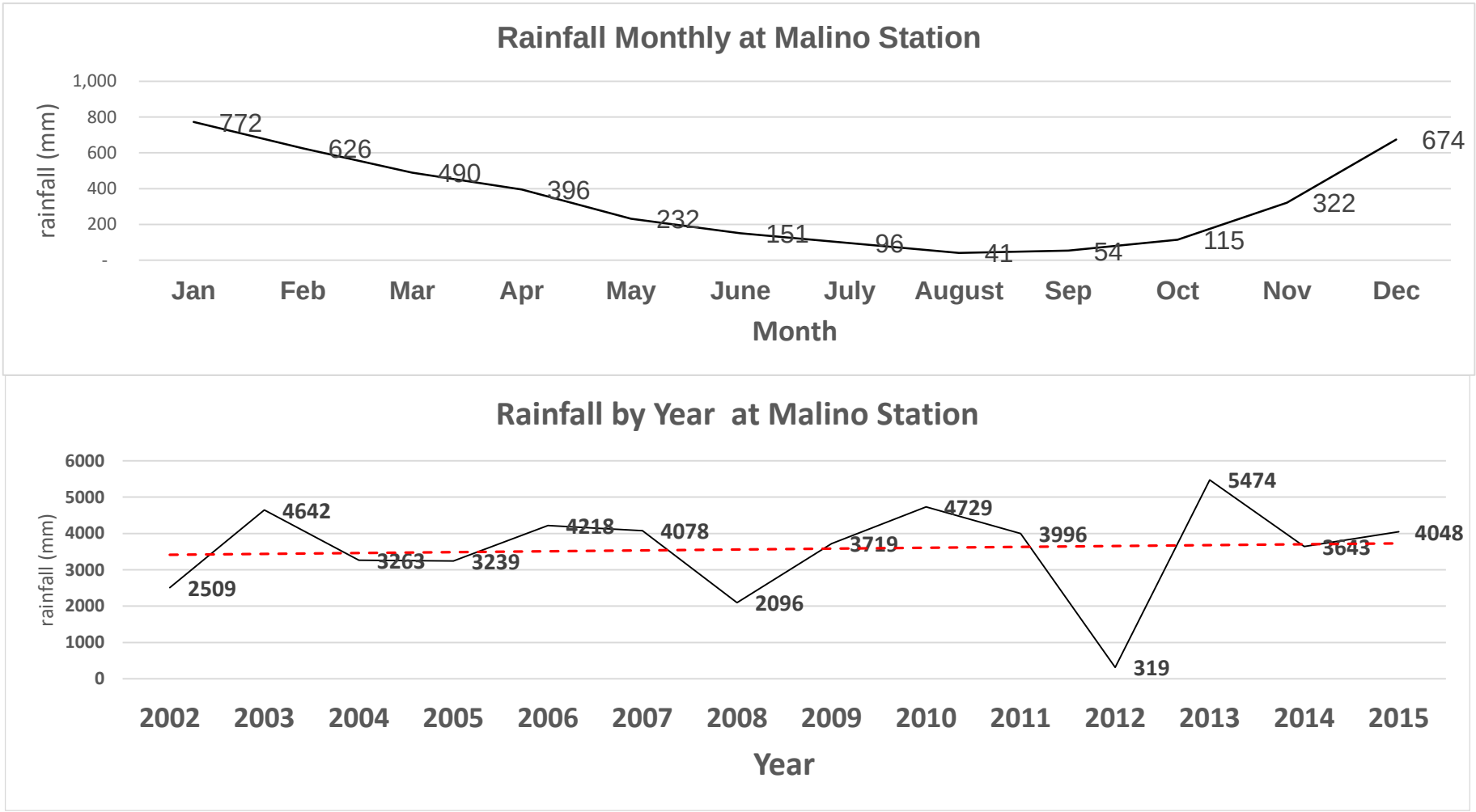
Appendix 6. Slope map in Ujung Loe Watershed, South Sulawesi, Indonesia



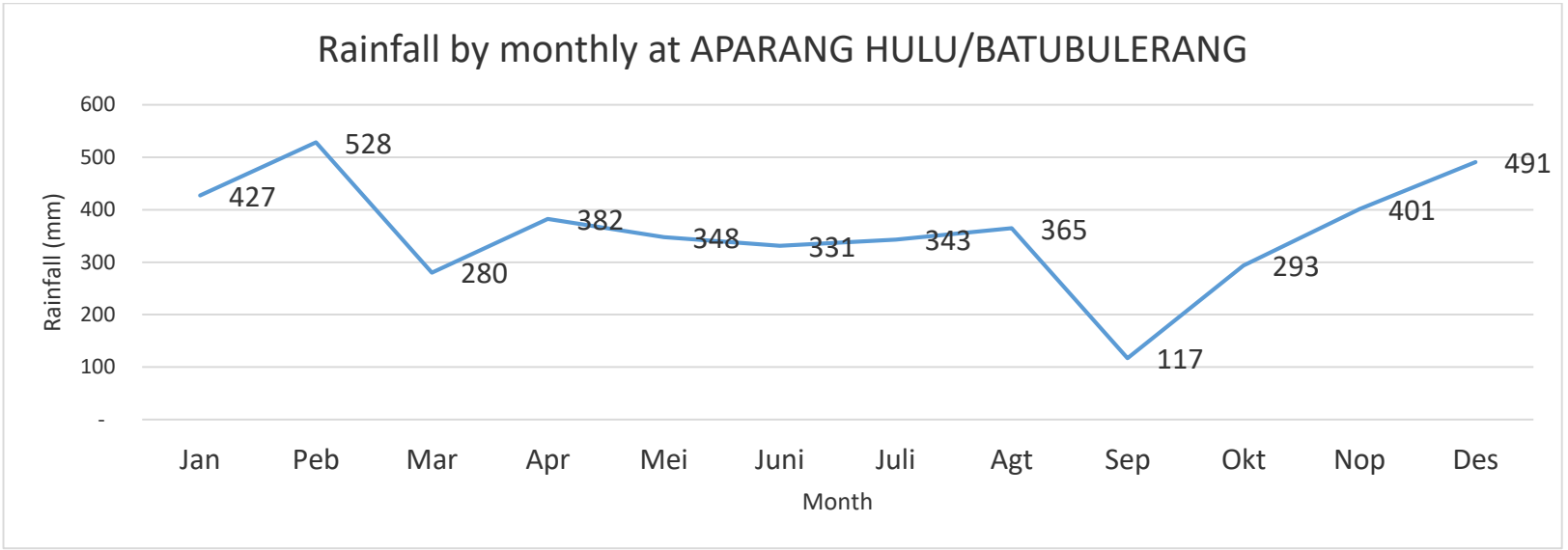
Appendix 7. Map of Polygon Thiessen of Rainfall in Ujung Loe Watershed, South Sulawesi, Indonesia

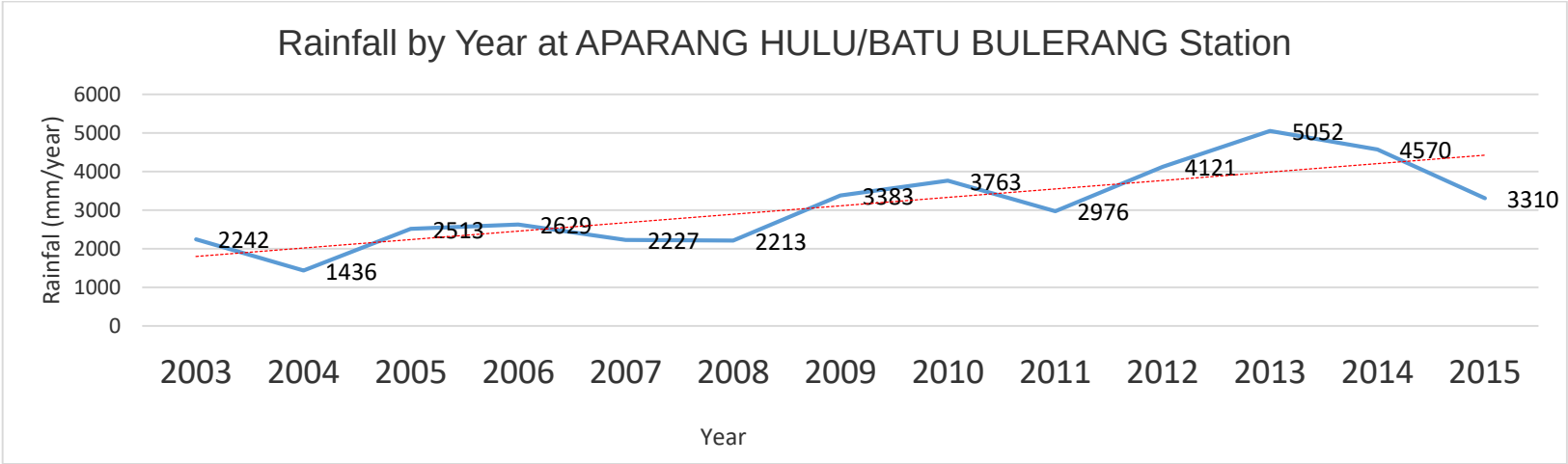


Appendix 8. Map of Population Density in Ujung Loe Watershed, South Sulawesi, Indonesia

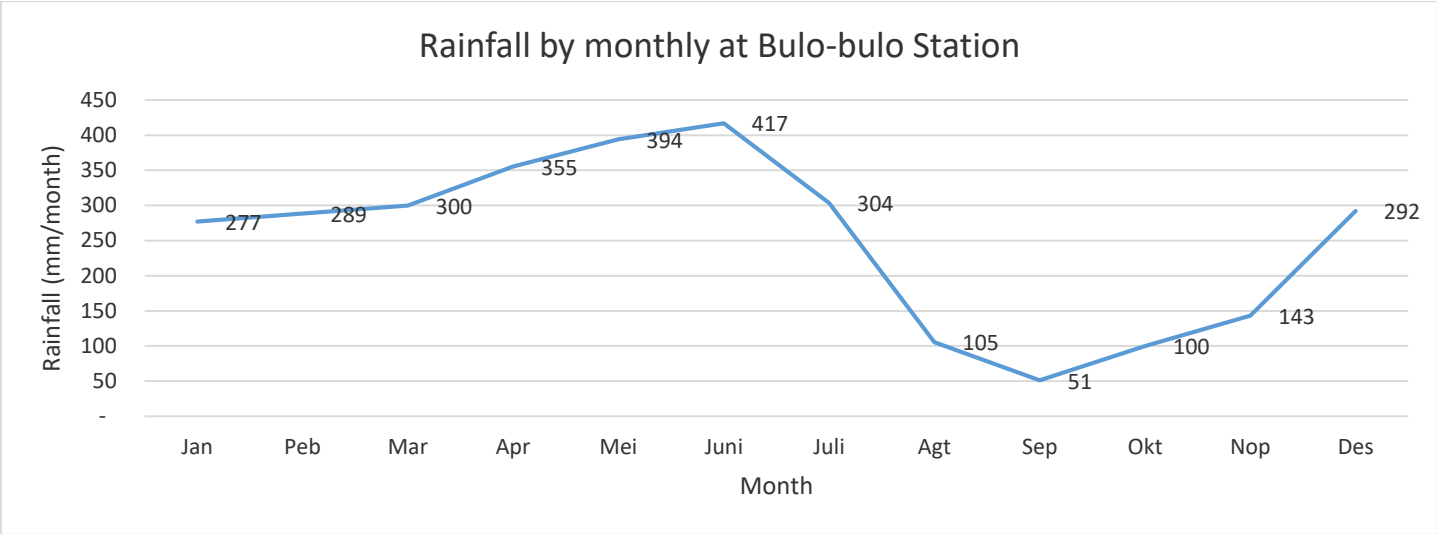


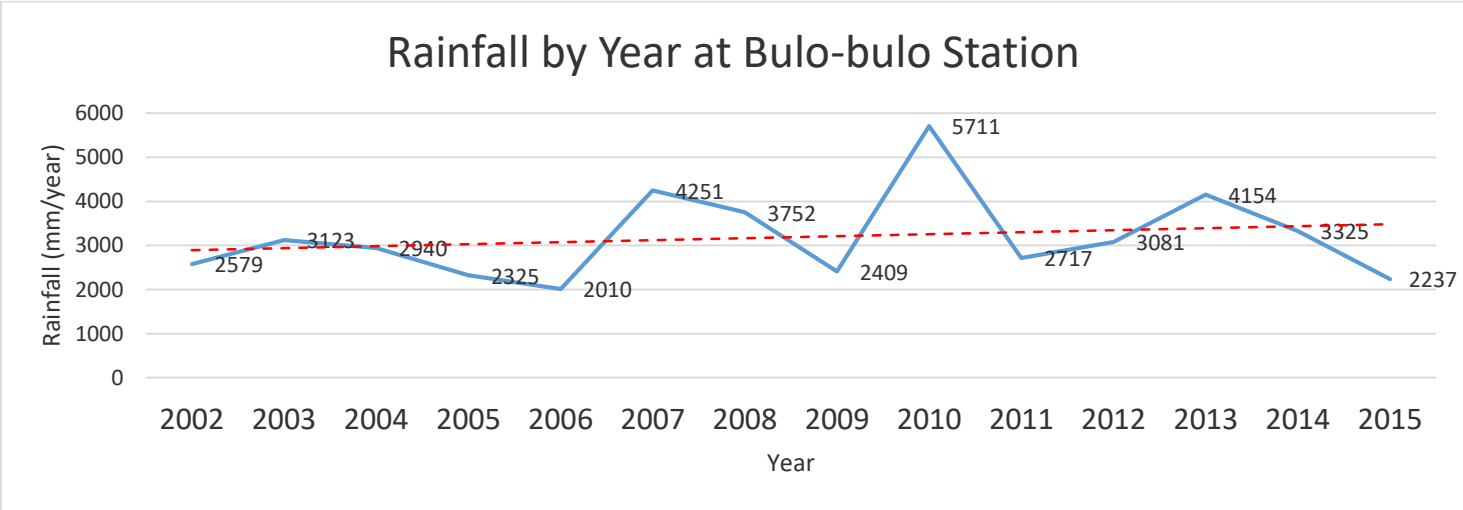
Appendix 9. Graph of Rainfall data period from 2002 to 2015 in Malino Rain Gauge Station



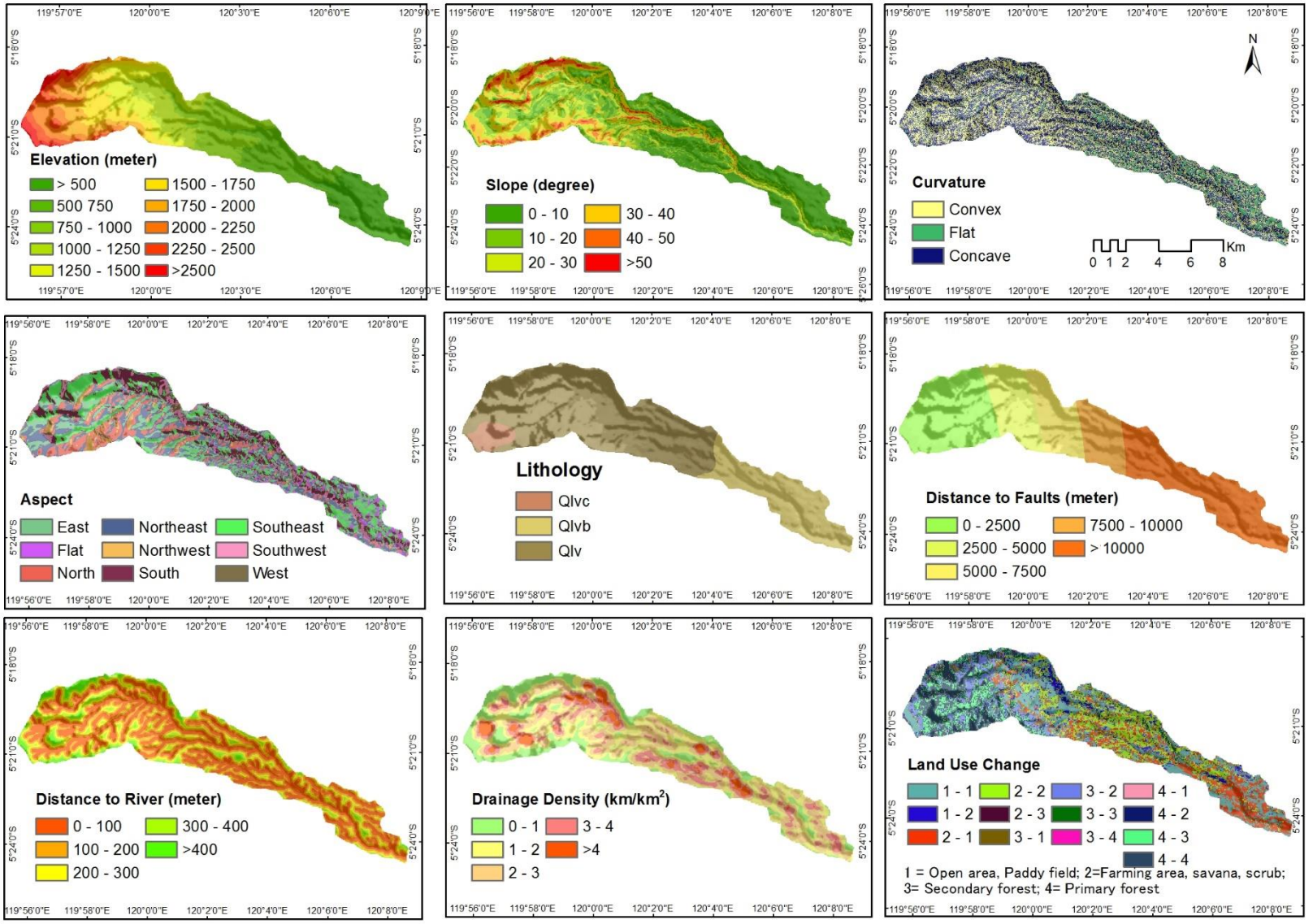


Appendix 10. Graph of Rainfall data period from 2003 - 2015 in Apparang Hulu Rain Gauge Station

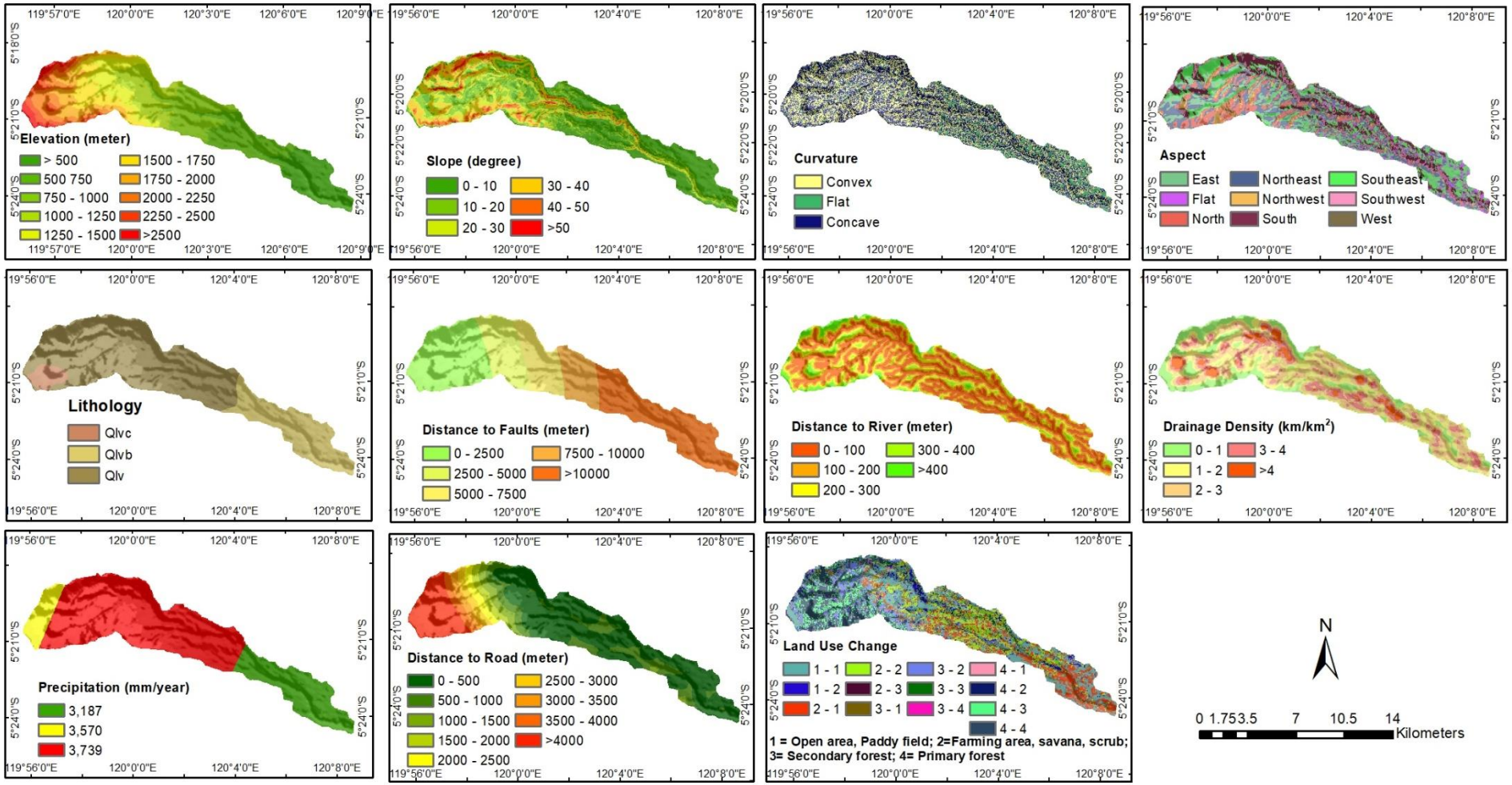




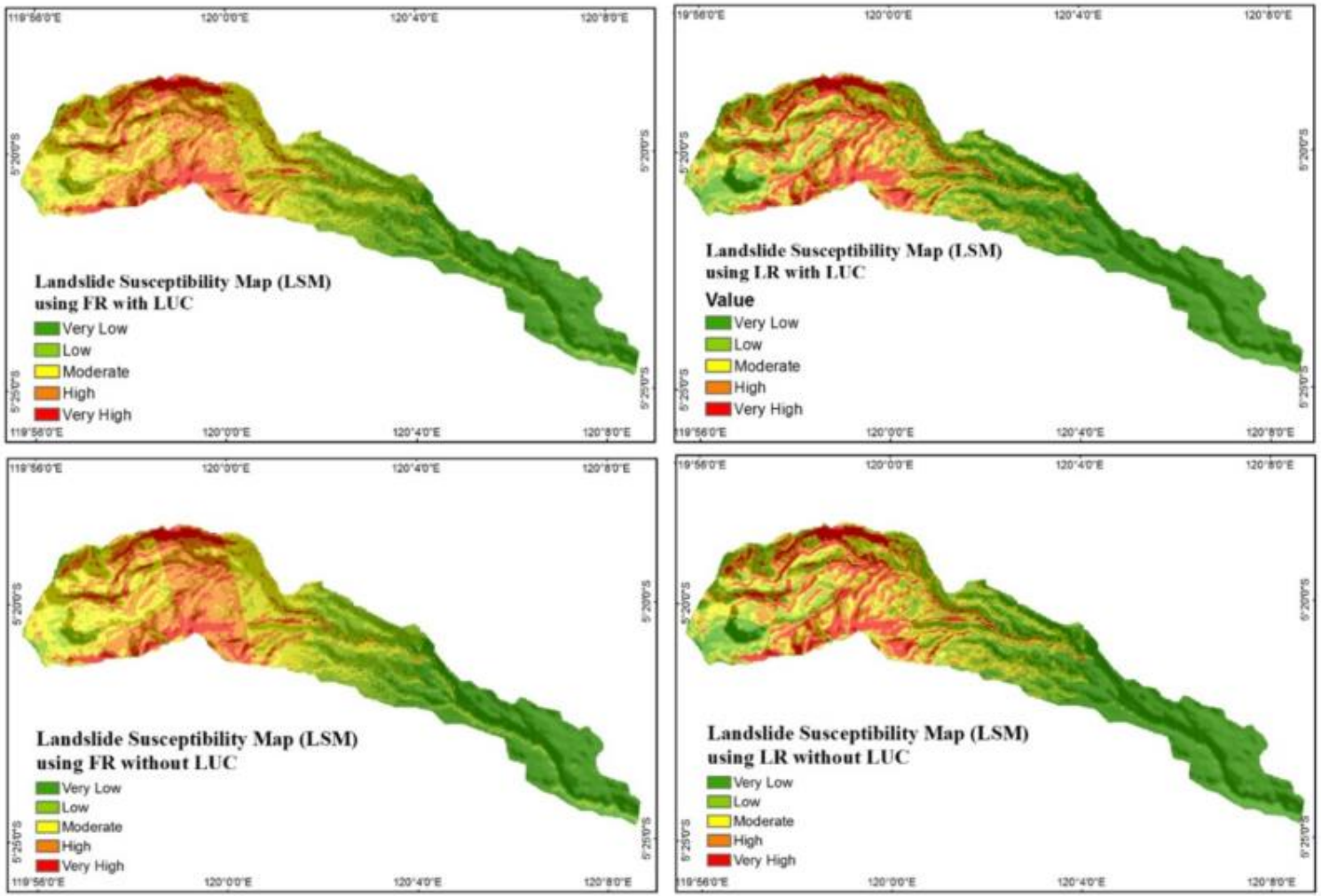
Appendix 11. Graph of Rainfall data period from 2002 - 2015 in Bulo-bulo Rain Gauge Station



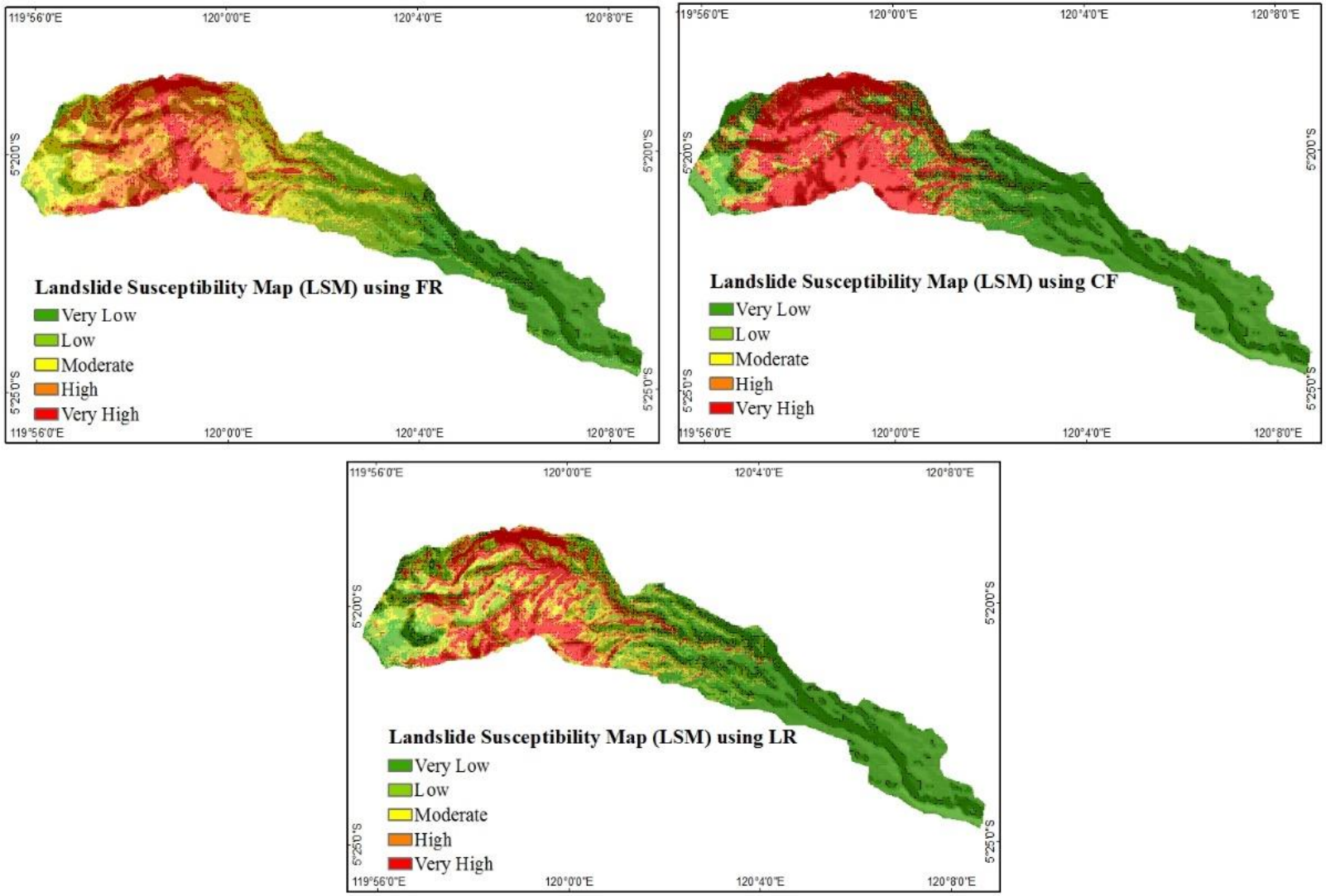
Appendix 12. Map of Nine Causative Factors



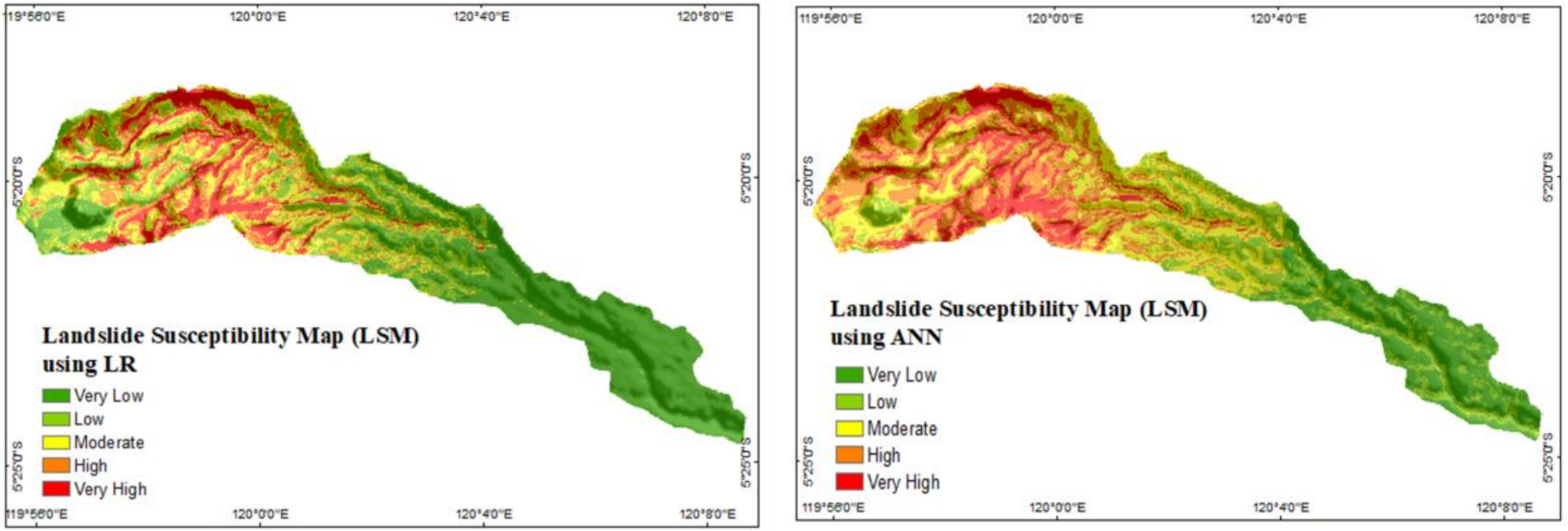
Appendix 13. Map of Eleven Causative Factors



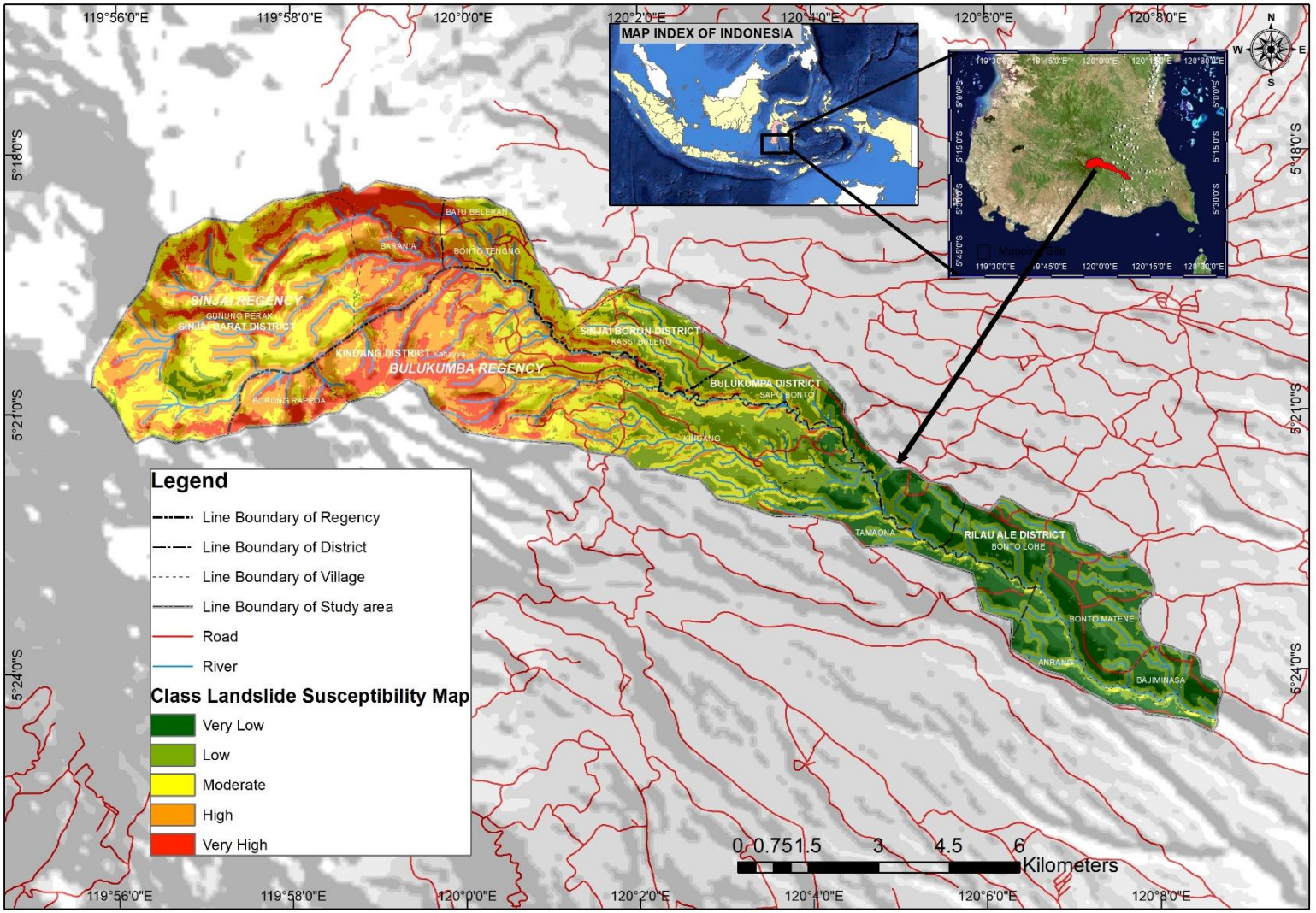
Appendix 14. Landslide susceptibility map of with and without LUC causative factors using FR, and LR method with nine causative factor in Ujung Loe Watershed, South Sulawesi, Indonesia



Appendix 15. Landslide susceptibility map of with eleven causative factors using FR, CF, and LR method in Ujung Loe Watershed, South Sulawesi, Indonesia



Appendix 16. Landslide susceptibility maps (LSM). (a) LSM multivariate logistic on test seventh; (b) LSM artificial neural network (ANN) with eleven causative factor



Appendix 17. Landslide susceptibility maps of the best Optimized causative factor using a combination of forward stepwise (likelihood ratio) logistic regression to eliminated causative factor and artificial neural network (FSLR-ANN) with nine causative factor




Appendix 18. Image of survey location

Appendix 19. Image of survey location

Appendix 20. Image of survey location

Appendix 21. Image of survey location

Acknowledgments

First of all, I would like to express my sincere gratitude to supervisor Professor Dr. Tetsuya Kubota for his wise advise and supervision, forehanded discussion, kind encouragement and continued support throughout the stages of my study. I feel a deep sense of gratitude co-supervisor, Associate Professor, Dr. Nobuya Mizoue and Associate Professor, Dr. Hideaki Mizuno for providing useful and constructive comments and suggestion in this dissertation.

Thanks also to all laboratory colleagues at Kyushu University for providing friendly environment throughout this research, especially to Dr. Aril Aditian and Putri Fatimah Nurdin.

I am much indebted to the government of Indonesia through DIKTI Scholarship Batch 2, 2015 awarded by the Ministry of Research, Technology and Higher Education of Indonesia. I am indebted also to Forest Faculty, Hasanuddin University for administrative support.

Finally, I would like to thank parents, Soma Saputra and Dina Tangketasik (who was pass away), for the encouragement my life and to almighty, Allah SWT, for the strength, nourishment, and opportunities. I would like to thank my wife, Nugrahwati Abbas, and sons, Andra and Danu for fulfilling supporting me in my studies in Japan. Last but not least, I am also indebted to my brother Kolonel Atjep Mihardja Soma, Ardi Suryawinata Soma and sister, Anni Junaningsih Soma, and father and mother-in-law, Abbas Makkanyuma and Hj. Nurhayati Enteng for their love and support during the study.