Doctoral Thesis

博士論文

Landslide susceptibility assessment using machine learning with emphasis on scaling and topographic representation issues

（空間スケールと地形の代表性を重視した機械学習による斜面崩壊の発生しやすさの評価）

Uttam PAUDEL

ウッタム　ポウデル

Doctoral Thesis

博士論文

Landslide susceptibility assessment using machine learning with emphasis on scaling and topographic representation issues

（空間スケールと地形の代表性を重視した機械学習による斜面崩壊の発生しやすさの評価）

Uttam PAUDEL

ウッタム　ポウデル

Department of Natural Environmental Studies,

Graduate School of Frontier Sciences,

The University of Tokyo

東京大学大学院新領域創成科学研究科

自然環境学専攻

Acknowledgments

This thesis could not have been completed without the great support and encouragement of a number of people. I would like to express my gratitude to all.

My first and foremost gratitude to my supervisor, Prof. Takashi Oguchi, for his guidance, support, and for the freedom I was given throughout these years. Every word of encouragement and criticism from him has contributed to the successful culmination of this dissertation. I must have sorely tested his patience many times; yet, his unabridged support was coming during the entire course of my research.

I am also grateful to Assoc. Prof. Yuichi Hayakawa for his guidance and support and for multiple conversations in diverse fields, all of which I believe will be instrumental in future endeavors. I would also like to thank him for the opportunity to work in the analysis of high-resolution DEMs and the financial support.

The comments and suggestions made by the Ph.D. thesis evaluation committee were helpful in improving this thesis. I would like to thank Prof. Toshihiko Sugai, Prof. Kaoru Saito, and Assoc. Prof. Kei Yoshimura for their time and appraisal.

Special thanks to my colleagues and seniors at Oguchi-Hayakawa laboratory. Song-san, Yunus-san, Chen-san, Dou-san, Tuba-san, Hao-san, Hagar-san, and lab-mates; thank you for your friendship and help. I would also like to thank all my friends in Nepal for their support in this special journey.

My sincere gratitude to the Government of Japan for the Monbukagakusho (MEXT) scholarship and Japan Student Services Organization (JASSO) and Graduate School of Frontier Sciences (GSFS) for the multiple academic grants. I would also like to

Abstract

Landslides are naturally occurring complex geological phenomena that cause significant damage in mountainous regions. Landslide susceptibility (LS) mapping is a requisite for safety against such sediment disasters caused by landslides, and considerable efforts have been exerted in this discipline. However, some key issues concerning the availability of thematic data, scaling of landslide causative factors, and the representation of landslide events still impose challenges. This research aimed to identify viable solutions to these issues.

Different parameters are generally used in LS studies (e.g., geology, soil depth, soil type, and land use) but limitations of data availability and resolution often restrict the replicability of such studies. Among the data required for LS studies, digital elevation models (DEMs) are currently the only dataset available globally at fine scales suitable for LS studies. This study examined the usefulness of a DEM-based LS analysis. One of the major challenges with this type of analysis is selecting an appropriate scale for LS studies due to the size heterogeneity and distribution of landslides, which requires identification of an optimal scale for landslide causative parameters. This study proposed a method to identify the optimum scale for each parameter and use multiple optimal parameter-scale combinations for LS mapping. Furthermore, the issue of topographic representation in a grid-based analysis arises because no raster cells within a landslide are equally responsible for landslide occurrence. Representation determines the sampling way of causative factors and thus affects further analysis. This study compared five different representation techniques: seed cells, center-cells, cells within the landslide boundary, cells within the depletion zone, and the dominant cells within the depletion zone (DCD), to

identify the best representation technique. DCD is a new method of representation proposed in this study.

This study utilized Random Forest, a relatively new machine learning technique in the field of landslide research, together with 16 geomorphological parameters extracted from 10, 30, 60, 90, 120, 150, and 300 m DEMs and an inventory of historical landslides. The geomorphological parameters employed are those frequently used in existing landslide research. Two equal-sized (625 km^2) areas in Niigata and Ehime, Japan, with different geological and environmental settings and landslide density, were selected as study areas. The methodology was developed using high-resolution data available for Japan, and was successfully applied in a study area in Nepal where the quality of data is relatively low.

The usefulness of a DEM-based LS analysis was examined using two sets of models - with and without geological information. The results suggest that the addition of geological information leads only to a small increase in the prediction accuracy of the LS model in an area of high seismicity, and that the geological parameters are consistently ranked lower in importance than most other topographic parameters. Such an observation seems to reflect the coarser scale of the geological information used, and that the topography represented by a detailed DEM may include the effects of local geology. Accordingly, a DEM-based LS study is useful even if other high-quality datasets are unavailable, at least for rapidly eroding mountainous areas like those in Japan and Nepal.

A multi-resolution LS analysis technique was proposed to address the scaling issues. The method first determines the optimum scales for all parameters to best represent the conditions of slope failure. The parameters at different optimum scales are then brought together for the final LS mapping. The analysis of the scale and importance

of the DEM-derived parameters revealed that while some parameters show similar importance and scale dependency for different regions, environmental differences result in variability between the study regions. The performance of LS models also suggests that the finest scale of analysis is not always the best. The proposed multi-resolution LS analysis permits higher accuracy LS mapping than any single-scale analysis.

The multi-resolution LS modeling was used together with five different representation techniques to evaluate the appropriateness of the techniques. The results indicate that the newly proposed DCD method always leads to higher performance in corresponding susceptibility models. This is logical because the dominant cell within the depletion zone probably represents the dominant process governing landslide initiation. The use of the proposed multi-scale approach together with the proposed DCD led to high prediction accuracies: 81.2% in Niigata and 83.27% in Ehime.

In summary, this dissertation work has suggested the usefulness of a DEM-based LS study with two newly developed methodologies for landslide modeling. The broad applicability of these findings should be examined in future research.

Keywords: multi-resolution, landslide susceptibility, DEM, Random Forest.

Table of contents

List of figures

List of tables

List of acronyms

AHP	Analytical hierarchic process
ANN	Artificial Neutral Networks
ASTER	Advanced Spaceborne Thermal Emission and Reflection Radiometer
AUC	Area under the ROC curve
DCD	Dominant cells within the depletion
DEM	Digital elevation model
GSI	Geospatial Information Authority of Japan
GSJ	Geological Survey of Japan
LS	Landslide susceptibility
LSM	Landslide susceptibility mapping
ML	Machine learning
NIED	National Research Institute for Earth Science and Disaster Prevention
OOB	Out-of-bag
RF	Random Forest
ROC	Receiver operating characteristic
SPI	Stream power index
SRTM	Shuttle Radar Topography Mission
STCI	Sediment transport capacity index
SVM	Support Vector Machine
TCI	Terrain characterization index
TWI	Topographic wetness index
USGS	United States Geological Survey

Chapter 1. Introduction

1.1. Background

Landslides are naturally occurring complex geological phenomena marked by a downslope movement of slope-forming materials under the action of gravity. They are caused by a range of external triggers such as rainfall, rapid snowmelt, water level change, storm waves, rapid stream erosion, and earthquakes (Guzzetti et al., 2002; Keefer, 1984; Malamud et al., 2004a, 2004b) under the influence of numerous other factors such as topography, lithology, geological structure, soil moisture content, and anthropogenic activities (Crozier, 1999; Glade et al., 2000). The most important trigger, however, is intense rainfall (Glade et al., 2000; Guzzetti et al., 2008) followed by earthquakes (Chigira and Yagi, 2006; Kargel et al., 2016; Yin et al., 2009). Irrespective of the causes, landslides are serious geological hazards for human activities throughout the world, causing catastrophic loss of life and property, and worryingly an increasing trend has been observed for these disasters (Petley, 2011). This may be linked with the persistent environmental degradation as a consequence of urbanization, economic development and deforestation, which may be further exasperated by projected changes in the climate especially global warming (Brardinoni et al., 2003; Guthrie, 2002; Huggel et al., 2012; Van Asch et al., 1999). Landslide mitigation and risk reduction require mapping of susceptible areas and estimating the likelihood of landslide occurrences (Guzzetti et al., 1999). Landslide susceptibility (LS) mapping is therefore a requisite, and considerable effort has been exerted in this discipline.

1.2. Review of practices for landslide susceptibility assessment

Landslides are spatially discrete and temporally dynamic events controlled by a number of stochastic geo-environmental factors, represented by surface and subsurface variables. On the basis of such local terrain conditions, LS deals with the likelihood of landslide occurrence in an area (Brabb, 1984). Most LS studies follow a simple principle: the past and the present are the keys to the future. The conditions leading to the past and present failures will help in estimating the style, frequency, extent, and consequences of failures in the future (Varnes, 1984). Over the years, the number of studies concerning methods and progress in LS has grown rapidly. They involve either qualitative or quantitative modeling (Hussin et al., 2015). Many pioneer works in this field concern qualitative studies where the judgment established by experts, based on the data investigated, was used to produce susceptibility maps (Atkinson and Massari, 1998; Brabb et al., 1972). The subjectivity of these methods was addressed by the adoption of quantitative assessment methods such as bivariate or multivariate statistical analysis (Nandi and Shakoor, 2010; Yalcin et al., 2011), logistic regression (Ayalew and Yamagishi, 2005; Guzzetti et al., 1999), likelihood ratio (Akgun, 2011; Kanungo et al., 2011), weight-of-evidence (Lee and Choi, 2004; Regmi et al., 2010), discriminant analysis (Guzzetti et al., 2006), frequency ratio (Choi et al., 2012), and analytical hierarchy processing (Yalcin et al., 2011; Yoshimatsu and Abe, 2006).

Recently, various machine learning (ML) techniques have been used for statistical modeling in landslide research, more often because of their robustness in handling large and complex data associated with regional environmental conditions. ML techniques learn from the available data in order to perform processing tasks such as classification, prediction, and clustering. These approaches include Artificial

Neutral Networks (ANN; Conforti et al., 2014; Zare et al., 2012), Support Vector Machine (SVM; Tien Bui et al., 2012), Decision Trees (Saito et al., 2009; Yeon et al., 2010), and Random Forest (RF) (Catani et al., 2013; Cutler et al., 2007; Paudel et al., 2016).

The large array of modeling methods makes it challenging to choose one for a particular application. The simplest approach to select an optimal model for prediction is to compare their prediction accuracies (Goetz et al., 2015). Numerous comparisons of LS modeling methods have been conducted; yet no single best method has been concluded (Akgun, 2011; Brenning, 2005; Goetz et al., 2015; Lee and Talib, 2005; Pradhan, 2013; Yalcin et al., 2011; Yesilnacar and Topal, 2005; Yilmaz, 2009, 2010a).

RF is a relatively new ML technique (Breiman, 2001) that utilizes an ensemble learning approach for classification, regression, and other functions; its use in landslide research is still limited to a few examples (Brenning, 2005; Catani et al., 2013; Goetz et al., 2015; Paudel et al., 2016; Trigila et al., 2015; Vorpahl et al., 2012). However, it is widely used in various other fields of data mining and has been demonstrated to have excellent performance in comparison to similar ML algorithms and traditional statistical models (Breiman, 2001; Lee et al., 2005; Meyer et al., 2003; Svetnik et al., 2003; Wu et al., 2003). Fernández-Delgado et al. (2014) compared most relevant classifiers available today (179 classifiers) using 121 data sets and found that RF was the best followed by SVM. Another comparative LS study involving several ML and statistical techniques reported RF to have the best predictive performance (Goetz et al., 2015). RF is utilized in this study based on its superiority as suggested by several similar empirical studies, and

to utilize numerous model outputs unique to RF that provide valuable insight into LS.

1.3. Existing problems

1.3.1. Availability of thematic information

The dynamics and interactions of the different factors affecting landslide occurrence are important for an effective LS assessment (Kawabata and Bandibas, 2009). As a result, different intrinsic and extrinsic parameters are used to analyze LS. However, many parameters such as geology, soil depth, soil type, and land use usually have limitations of availability and relevant scale (Coe et al., 2011; Hasegawa et al., 2009). Such limitations make LS assessments feasible only when adopting a qualitative evaluation technique (Günther et al., 2013), thereby restricting the replicability of many quantitative LS methodologies in areas where high-quality data sets are absent (Cascini et al., 2010).

Therefore, LS evaluation solely based on a digital elevation model (DEM) has been conducted in this study, assuming that topography reflects other factors such as geology, land use, and moisture content (Beven, 1997; Birkeland, 1984; Coblentz et al., 2014; Moore et al., 1993; Prima and Yoshida, 2010). Studies on terrain characterization suggest that DEMs can be utilized for geologic assessment (Coblentz et al., 2014; Iwahashi et al., 2001; Iwahashi and Pike, 2007). Differences in mechanical and chemical properties of subsurface lithologies as well as surface conditions such as land use/cover lead to variable weathering/erosion resulting in the variations in surface landforms (Germanoski, 2001; Moore et al., 1993; Prima and Yoshida, 2010; Wade, 1935). Studies have also shown the importance of topographic indices in characterizing terrain conditions. For example: soil moisture

4

content, an important consideration in LS, can be estimated by the topographic wetness index (Beven, 1997).

A DEM-based methodology is also relevant because, DEMs are the only data source available globally at finer spatial scales significant for quantitative LS assessment (e.g., SRTM and ASTER GDEM). Increased availability of high-resolution global DEMs (e.g., SRTM 1 Arc-Second Global, ASTER GDEM, ALOS world 3D – 30 m) and recent advances in DEM acquisition techniques facilitate this approach.

1.3.2. Scaling of topographic parameters

Most land-surface parameters and objects vary with spatial scale, a function of cell size or grid resolution (Wilson and Gallant, 2000). However, in absence of scale optimization techniques (Zhilin, 2008), many analysis are conducted at arbitrary scales often depending on data availability without much concern for scale effects in the analysis (Drăguţ and Eisank, 2011).

The heterogeneity in size and distribution of landslides requires the selection of an appropriate DEM scale to achieve high precision in LS research. However, the finest available DEM resolution is often utilized without this consideration. At coarser scales, terrain presentation may be too smoothed. Therefore, Keijsers et al. (2011) suggest the use of fine resolution DEMs for LS which provide better representation of slope morphology and hydrological patterns. However, Tarolli and Tarboton (2006) found that LS prediction performance decreases at finer resolutions because too localized topography does not represent the processes governing landslide initiation. Catani et al. (2013) found that the importance of landside predicting parameters changed with spatial scale, and concluded that for some parameters, scale representing not local values but their trends should be

evaluated. However, they did not conduct a concrete study to incorporate the variability of parameter importance at different scales for landslide susceptibility mapping (LSM).

This DEM-based study proposes a novel approach to identify the optimal resolution of each topographic parameter and use those parameters at multiple-optima for an LS study using an RF model. The output of variable importance, unique to RF models, was utilized for sensitivity analysis.

1.3.3. Topographic representation of landslides

Numerous studies have used different mathematical techniques in LSM with excellent results; however, only few of those studies have considered the way modeling datasets are prepared. Generally, LS models involve a sampling strategy to construct training and validation datasets as it is not possible to include all the data in an LS model (Atkinson and Massari, 2011; Kawabata and Bandibas, 2009). In other words, the topographic representation of landslide determines how the causative factor information for individual landslide is extracted in the dataset used for susceptibility modeling (Hussin et al., 2015) and it greatly affects the outcome (Wang et al., 2013)

Landslides are commonly mapped as vectors and demarcated either by points, polygons, or lines (Bai et al., 2010; Brenning, 2005; Galli et al., 2008; Malamud et al., 2004a; Xu et al., 2013, 2014). For landslides mapped as points, the representative attributes correspond to the cell value underneath the point. However, for landslides mapped as polygons, literature suggests the use of several representation strategies. Use of center-cell is the most common method for landslide representation in regional studies (Atkinson and Massari, 1998, 2011; Bai

et al., 2009; Dong et al., 2009; Mathew et al., 2007; Von Ruette et al., 2011; Yilmaz, 2010b). This method uses the cell value underlying the centroid of the polygon in GIS environment or the field verified center of a landslide (Dai and Lee, 2002). However, Simon et al. (2013) suggest that a point based measure is unrepresentative of the landslide body and is likely to mislead the results of subsequent analysis, and therefore recommend an area-based approach. The popularity of center-cell is followed by the use of cells within the landslide boundary (Ayalew and Yamagishi, 2005; Pradhan and Lee, 2010). The latest among the representation strategies is the use of seed cells (Süzen and Doyuran, 2004), which considers an area around the crown and flanks of the landslide as the best undisturbed morphological zone representing conditions before landslide occurrence. There have been several modifications to the original idea concerning the landslide crown and lateral planks. Bai et al. (2010) and Yesilnacar and Topal (2005) used a buffer around landslide boundaries; Che et al. (2012) used a buffer around the landslide center; and Wang et al. (2013) used buffers around the crown of landslides. In the use of landslide bodies as well as seed cells, some studies use all cells within the domain (Guzzetti et al., 2006; Trigila et al., 2015; Yilmaz, 2009), while others sample cells at random (Kawabata and Bandibas, 2009; Nefeslioglu et al., 2008a, 2008b; Yesilnacar and Topal, 2005). Van Den Eeckhaut et al. (2006) recommends the use of a single cell approach for representation to avoid spatial autocorrelation between sampled cells.

Abovementioned strategies use either the center-cell, a randomly sampled cell or all cells within the landslide boundary for LS analysis. However, there remains a possibility that these methods may not be representative of landslide initiation conditions as they could erroneously include cells from the accumulation zone (Clerici et al., 2006).

The proportional area-based approach of selecting the dominant representative cell value of a parameter within an area provides better representation of landslides (Simon et al., 2013). Though such a technique is new in LS research, a similar approach to representation were considered by Oguchi (1997) and Saito et al. (2009) for geological data. Oguchi (1997) defined the geology of each cell with the bedrock type occupying the largest area and similarly Saito et al. (2009) used the dominant geology in each watershed as the representative. However, for landslides with initiation zones generally located at the top and traveling downslope under the influence of gravity (Cruden, 1991; Pavel et al., 2011; Varnes, 1978) it is only relevant to select cells from the depletion zone (Atkinson and Massari, 1998). Depletion zone (or detachment zone, or rupture zone) is genetically and morphologically distinct zone on the upper part of a landslide where failure is effectively generated (Clerici et al., 2006).

1.4. Research objectives

Mountainous areas are susceptible to slope failures, which may be disastrous. Studies to mitigate landslides involve identification of susceptible areas, and as noted in the previous sections, we have amassed a vast understanding in this field. However, some issues still exist; three such issues investigated during this dissertation work constitute the research objectives.

– Landslide research commonly use various thematic information associated with slope conditions such as geology, soil depth, soil type, and landuse. However, such information generally have limitations of availability and scale that restrict their replicability in data-limiting areas. Therefore, the first objective of this dissertation is to develop a DEM-based LS model.

– The selection of an appropriate DEM scale is necessary to achieve high precision in LS research. However, single scale of topographic parameters, used in contemporary studies, might be restraining because the scaling of parameters is not uniform. Therefore, the second objective of this dissertation is to analyze how regional differences affect the scaling of topographic parameters and develop a methodology to incorporate such variations in LS mapping.

– The third issue relates to the topographic representation of landslides in LS modeling. Currently, different sampling and representation techniques are utilized in LS research. In order to identify the most suitable, the final objective of this dissertation is to analyze and propose the best method for the representation of landslides.

1.5. Outline of the thesis

This research is carried out to improve the contemporary practices in LS analysis and is essentially concentrated on the three issues mentioned in the previous section. The chapters are organized in the best possible manner to present an integrated idea.

Chapter 1 presents a general introduction to landslides. It includes literature reviews on causes of landslides, their impact to the society, methods of mitigation, current practices, and existing research problems. Subsequently, the research objectives and questions are introduced.

Chapter 2 introduces the study areas selected for this research and details their geo-environmental conditions.

Chapter 3 describes the data and the topographic parameters used in this research.

Chapter 4 details the methods adopted in this study, including newly proposed ones.

Chapter 5 includes the results obtained in the three study areas using the methods described in Chapter 4.

Chapter 6 discusses the results in order to associate the observations with their causes. It also discusses the results in comparison to previous studies.

Chapter 7 concludes the study and suggests directions for future research.

Chapter 2.   Study area

The analysis was first carried out for two equal-sized areas in Japan that differ in geological and environmental settings and landslide density (**Figure 2-1** and **Figure 2-2**). The study area in Niigata Prefecture, Honshu, was selected as a representative of an area with frequent landslides. It has an area of 625 km^2 (25 $\times$ 25 km), and the location has the highest landslide density in the region. An equal-sized area in Ehime Prefecture, Shikoku, was selected as a representative of an area with fewer landslides. These areas with different geo-environmental settings and landslide density allow us to examine how regional differences affect the scale and importance of each parameter.

Melamchi valley in Nepal was selected as a study area outside of Japan to assess the applicability of the research (**Figure 2-3**). The area experiences numerous landslides each year and more importantly, the landslides and devastations caused by the Gorkha-earthquake in 2015 were outstanding.

2.1.   Niigata, Japan

Niigata Prefecture in Japan is located on the island of Honshu along the coast of the Sea of Japan. The study area in Niigata lies between 138° 22' E and 138° 39' E and between 36° 58' and 37° 12.5' N. This area has very high density of landslides and landslides cover about 29% of the study area. The landslides in Niigata reflect specific geotectonic and climatic settings (Yamagishi et al., 2004; Yoshimatsu and Abe, 2006). The area lies in a large graben called North Fossa Magna (**Figure 2-2**) (Geological Survey of Japan, 1995) with active neotectonics (Inoue et al., 2012; Takeda et al., 2004; Takeuchi, 2008). According to a 10 m DEM (see **Section 3.2** for details), elevations in the area range from 5 to 1,284 m, with a mean of 369 m.

The mean slope is 17.4°. The mountains in the area mainly consist of Tertiary to Quaternary sedimentary rocks, including so-called "Green Tuff", and Quaternary volcanic rocks and their deeply weathered materials (Has et al., 2012; Matsuda et al., 1967; Takeuchi, 2004). The study area and its surroundings are characterized by heavy snowfall, whose meltwater contributes to rich groundwater and frequent landslides (Nakazato et al., 2013). The closest meteorological station (Tsunan, Japan Meteorological Agency) receives an average annual precipitation of about 1900 mm (1981–2010) with an average annual snowfall of about 1.35 m (1989–2010).

2.2. Ehime, Japan

The study area in Ehime is located in central Shikoku between 133° 11.7' E and 133° 29.7' E and between 33° 39' N and 33° 52.9' N. Its elevation ranges from 4 to 1895 m with a mean of 825 m. Mean slope is 31.7°, much larger than in Niigata. Located south of the Median Tectonic Line (**Figure 2-2**), most of the area is underlain by crystalline schist of the Jurassic complex from the Sanbagawa belt. Low-grade metamorphic greenstones from the Chichibu belt, a Jurassic accretionary complex zone, also dominate the southern section of the study area (Banno and Sakai, 1989; Suzuki and Ishizuka, 1998). A meteorological station in the southern part of the study area (Hongawa, Japan Meteorological Agency) and one in the northern part (Niihama, Japan Meteorological Agency) receive average annual precipitation of 3077 and 1305 mm (1981–2010), respectively. The study area is steep and affected by major tectonic lines (Hong et al., 2005), favoring landslides, but their density is lower than that in Niigata. Landslides in Ehime cover about 16% of the study area.
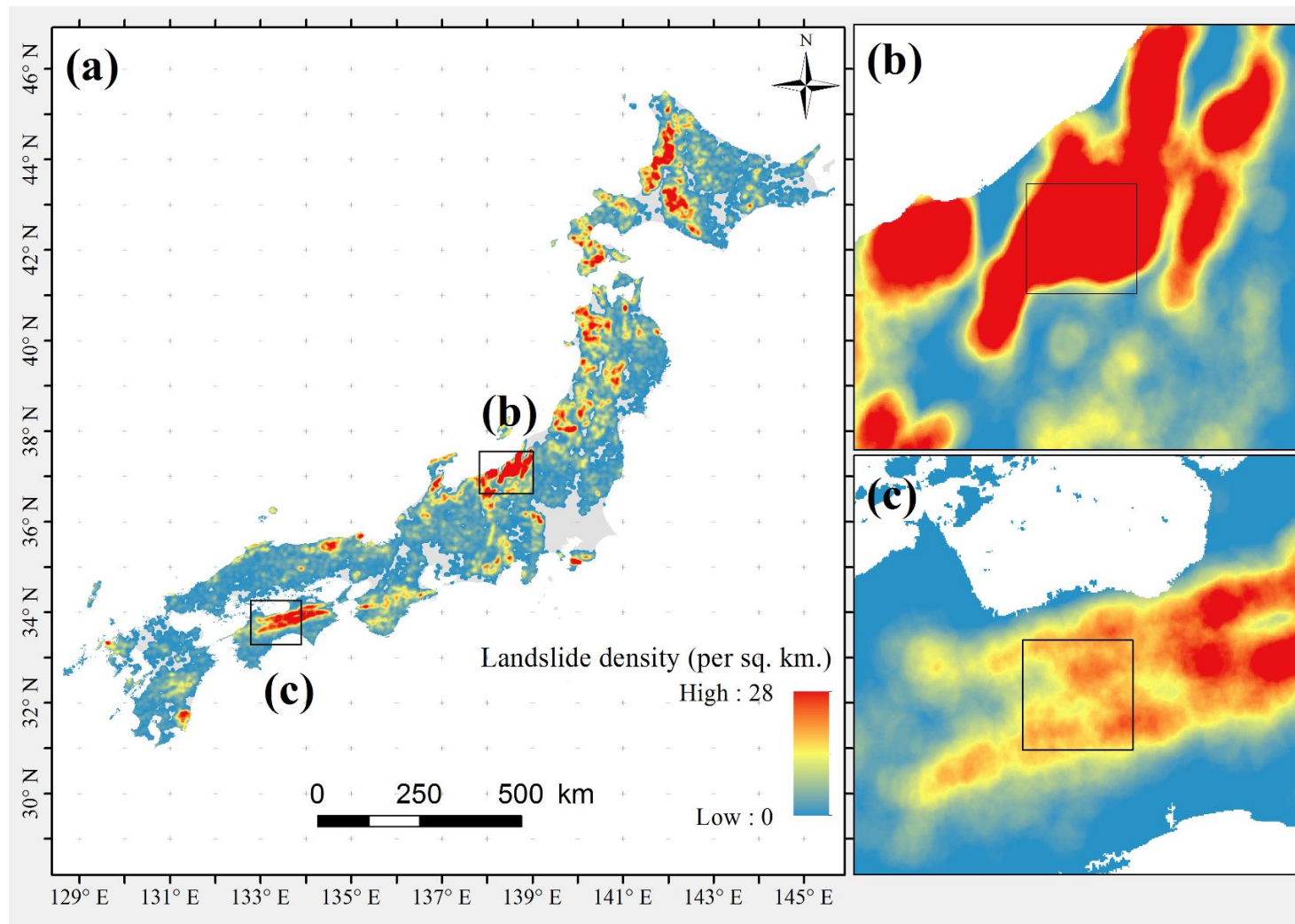
Figure 2-1. Map showing density of landslides in (a) Japan, (b) Niigata, and (c) Ehime, prepared using the landslides data from the NIED (National Research Institute for Earth Science and Disaster Prevention, Japan).

Study areas in Niigata (b) and Ehime (c) are indicated by square boxes. The color ramp corresponds to landslide density.
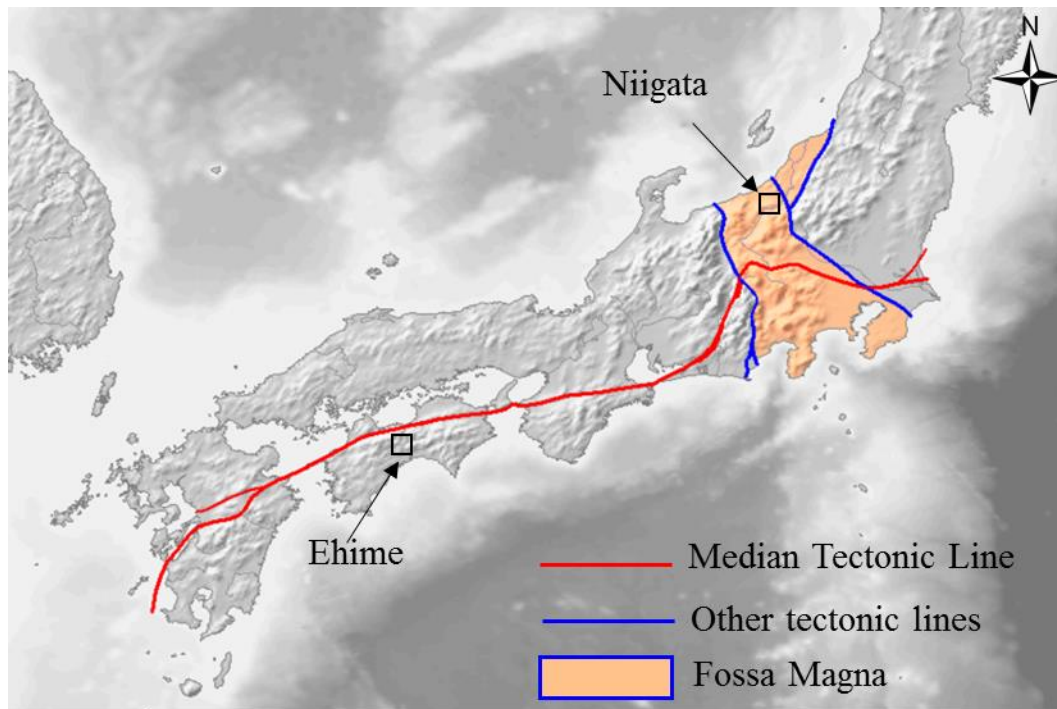
Figure 2-2. Map showing the location of the Median Tectonic Line and Fossa Magna.
(after Qrsk075, CC by SA 3.0)

2.3. Melamchi, Nepal

The study area in Nepal is the Melamchi valley and its surrounding areas located in Sindhupalchok district in the Bagmati zone of central Nepal (**Figure 2-3**). It has an area of 1023 km^2 and envelops the Melamchi valley with an area of 325 km^2. The valley is a typical narrow and steep Himalayan river-valley, 41 km in length with steep, rocky, and V-shaped lower and upper valley slopes. The mid-valley slopes are dissected to form alluvial terraces and are exploited for mountain agriculture and settlement (ADB, 2000). The lower part of the Melamchi valley has a sub-tropical climate, while the upper part has a cool temperate climate. Rainfall varies in the region depending on the elevation; higher elevation areas receive more rainfall than the lower areas. The annual average rainfall in the area is about 2800 mm, which is concentrated mostly during four months of the monsoon from June to September. Geologically, metamorphic quartzite rocks with soils of colluvial nature dominate the area. The combination of geological and environmental factors under the increasing anthropogenic pressures such as intense agriculture, grazing, deforestation and unplanned construction of rural roads accentuate the processes of mass wasting in the region (Tarolli and Sofia, 2016).

Mass movements can affect sediment discharge in rivers and hence water resources (Claessens et al., 2007; Hovius et al., 1997). The Melamchi watershed is the source water zone for the Melamchi water supply project under construction by the government of Nepal to bring drinking water to the Kathmandu valley (**Figure 2-3**) with the objective of diverting 170 MLD (million liters daily) of water from the Melamchi river through a 25.83 km long tunnel. A study of LS, therefore, provides information on the probable areas contributing to the sediment discharge so that

mitigation measures of slope protection could be targeted to those highly susceptible areas.
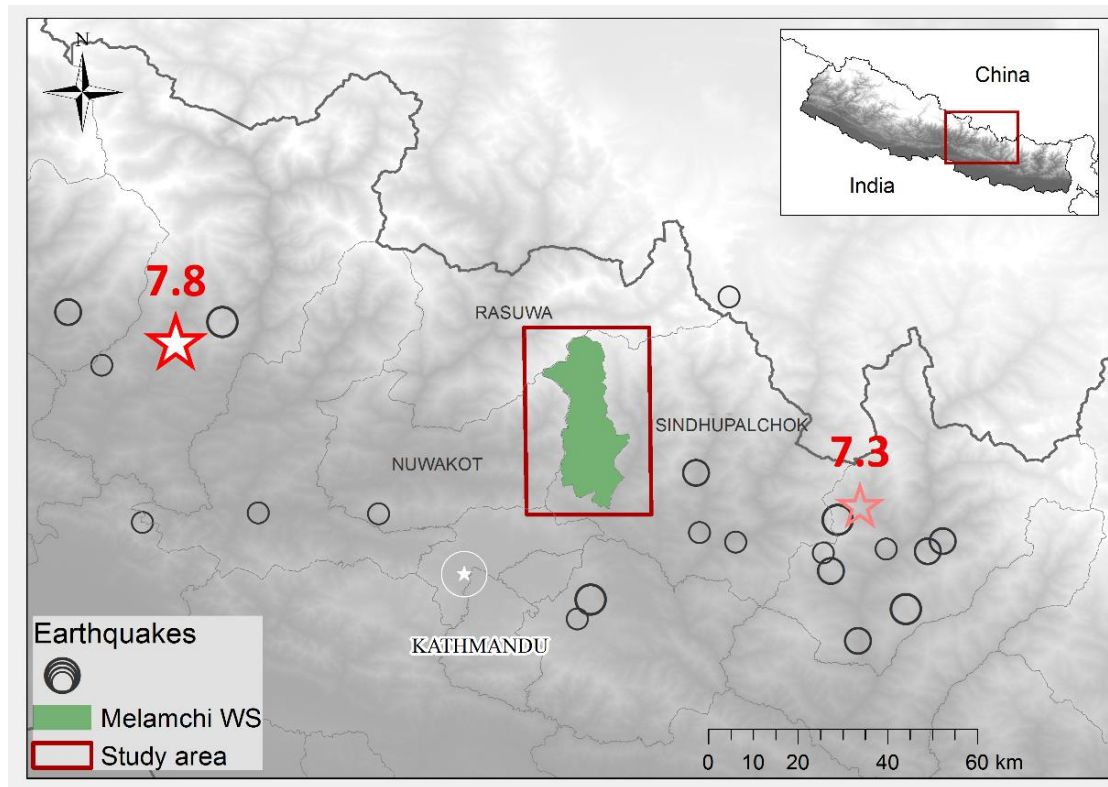
Figure 2-3. The Melamchi watershed and the study area in Nepal.
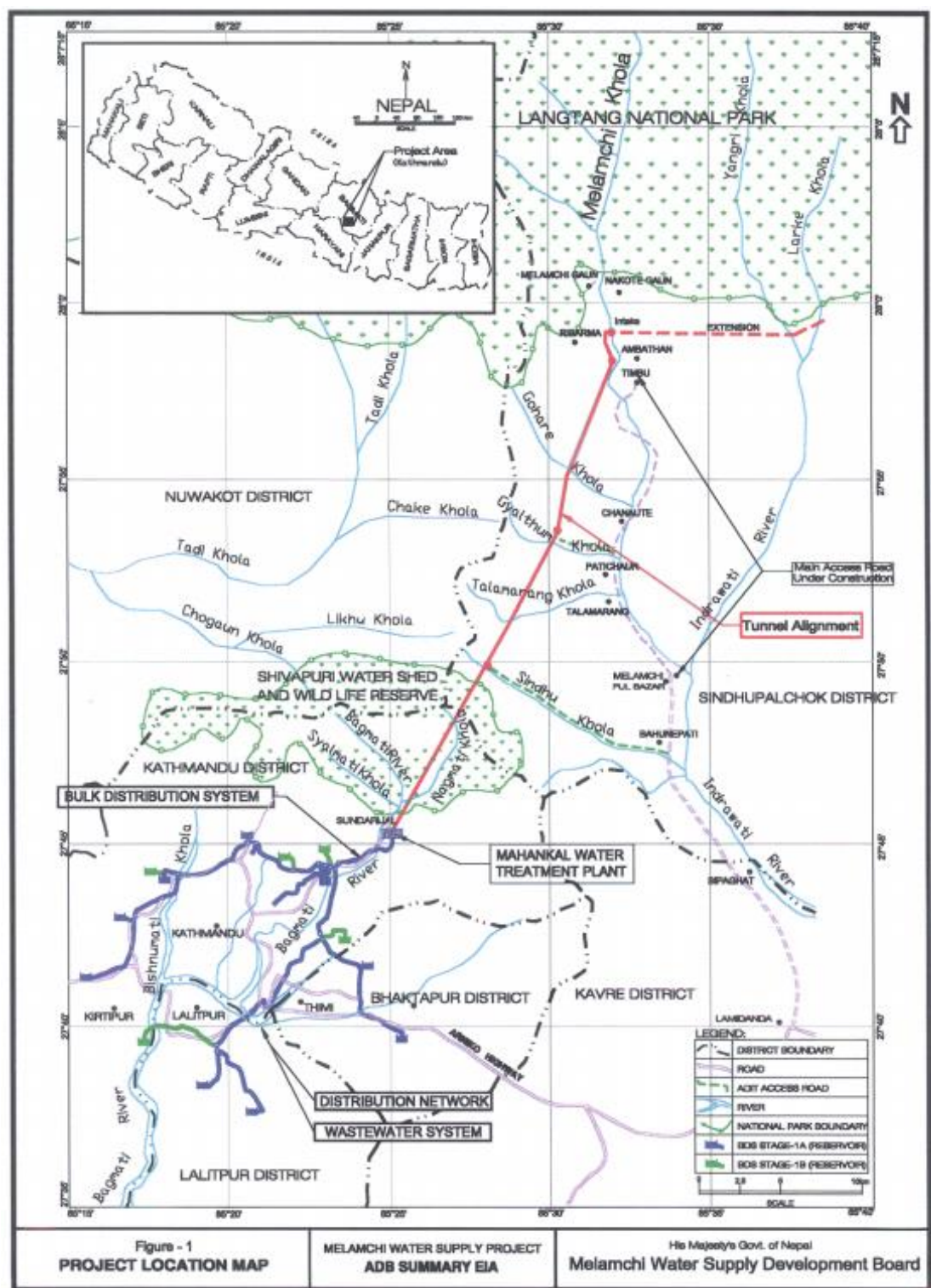
Figure 2-4. The Melamchi water supply project map (ADB, 2000).

On 25 April 2015, the Gorkha-earthquake [moment magnitude (M_w) 7.8] struck approximately 80 km to the northwest of the Nepalese capital of Kathmandu. It was followed by five aftershocks of ≥M 6.0 until 10 June 2015. The biggest aftershock on May 12, 2015 (M_w 7.3) occurred 80 km to the east-northeast of the Kathmandu. The earthquakes killed ~9000 people and severely damaged a 550×200 km region in Nepal and neighboring countries. Some mountain villages were completely destroyed, and the remote locations, blocked roads, and landslide-dammed rivers prevented ground access to many areas (Kargel et al., 2016)

Seismicity in the Himalaya dominantly results from the continental collision of the India and Eurasia plates, which are converging at a relative rate of 40–50 mm/yr – a fraction of which (~18 mm/yr) is driving the uplift of the Himalayan mountain range (USGS) (**Figure 2-4**). Northward underthrusting of India beneath Eurasia generates numerous earthquakes and consequently makes this area one of the most seismically hazardous regions on Earth. Gorkha-earthquake and subsequent aftershocks were also a consequence of this thrust faulting and were located on or near the main thrust interface.

The Gorkha-earthquake took a tremendous, tragic toll on human lives and culture. However, fortunately no damaging earthquake-caused glacier lake out- burst floods occurred. The total number of landslides was far fewer than those generated by comparable earthquakes elsewhere, probably because of a lack of surface ruptures, the concentration of deformation along the subsurface thrust fault at 10 to 15 km depth, and the regional dominance of competent high-grade metamorphic and intrusive igneous rock types. However, because immediate slow failures were less,

future rainfall events can result trigger greater number of landslides, including slopes in the Melamchi valley.
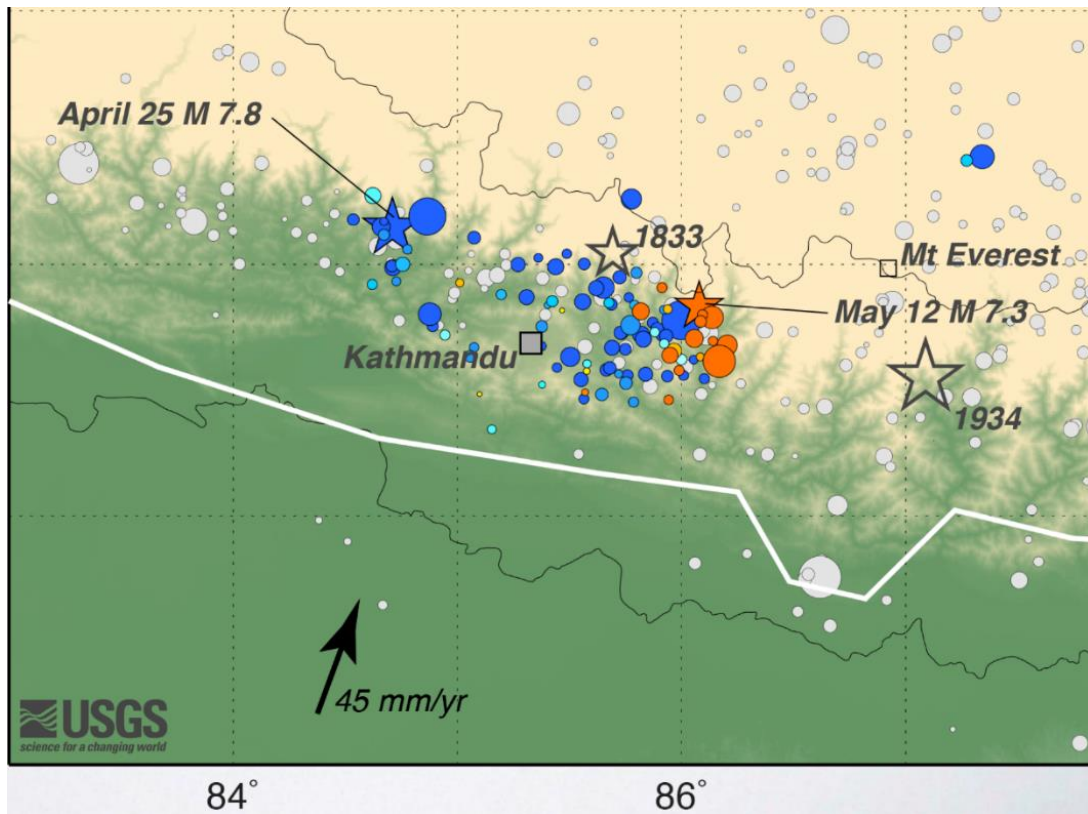
Figure 2-5. 2015 Nepal Earthquake and aftershocks (USGS).

Chapter 3.   Data

3.1. Landslide inventory

The landslide inventory for Japan was obtained from the NIED (National Research Institute for Earth Science and Disaster Prevention, Japan; available online at http://lsweb1.ess.bosai. go.jp/gis-data/index.html). The inventory covers the whole of Japan and was prepared by interpreting 1:40,000 aerial photographs. While the inventory includes many historical landslides identified from topographical discontinuities, small slope disturbances were not included. The use of historical (geomorphological) landslide inventories, which summarize past multiple landslide events (Malamud et al., 2004a), may enable a robust LSM because it reflects various environmental conditions, and the number of available data tends to be large. Landslides in Niigata cover about 29% of the study area while in Ehime they occupy about 16%. However, because of the uncertainties associated with the identification of large failures as unique (single) events (Guzzetti et al., 2002), and lack of differentiation among landslide types (inclusion of slow-moving earth flows and large fast-moving landslides), it seems better to remove exceptionally large landslides. The landslide data for the two study areas were hence extracted. The frequency distribution of the landslide area (**Figure 3-1**) shows that landslides greater than the 95th percentile in terms of the landslide area are out of the general trend. Therefore, these landslides larger than 95th percentile were not used in this study (**Figure 3-2**). The remaining 10662 landslides in Niigata and 2543 landslides in Ehime were investigated (**Table 3-1**).

Table 3-1. Statistical properties of the landslide inventory used in the study for Japan.

Properties	Area m^2	
	Niigata	**Ehime**
Min	212.0	1630.5
Max	47916.0	129713.9
Mean	9023.6	24196.3
Standard Deviation	8328.5	23824.4
Number	10662	2543
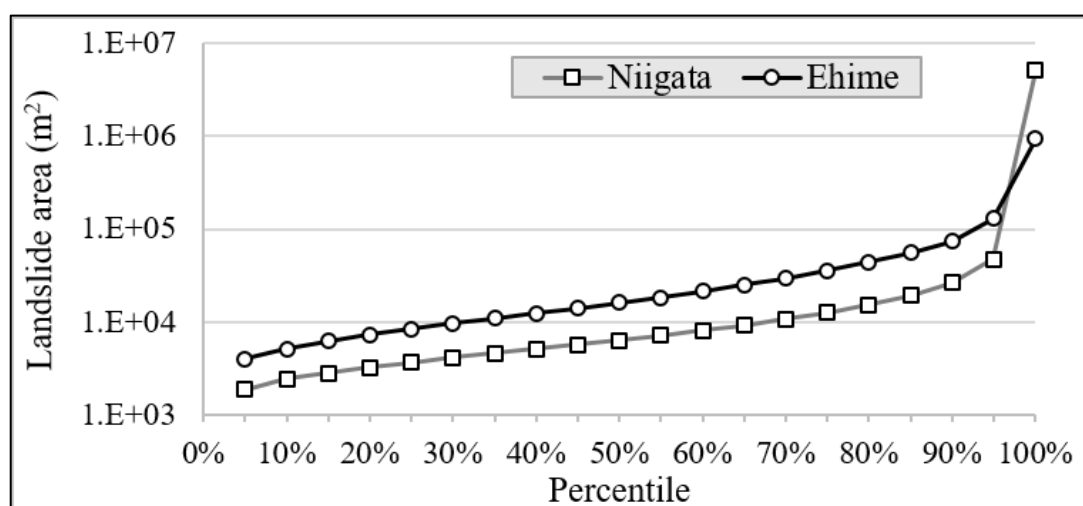
Figure 3-1. Percentile distribution of landslide area (Japan).
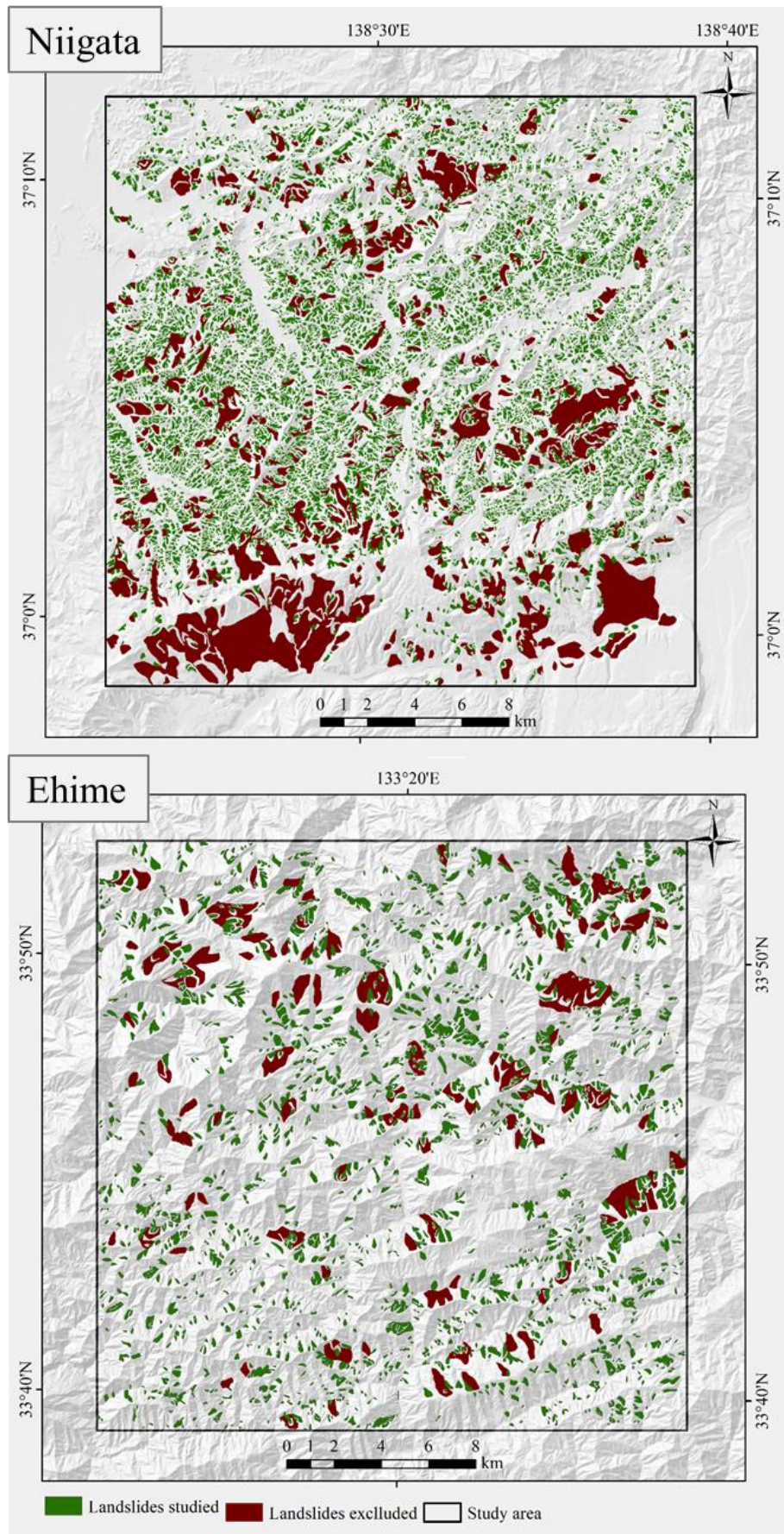
Figure 3-2. Landslide distribution in the Niigata and Ehime study areas.

Figure 3-3 and **Figure 3-4** show the frequency distribution of their areas. Medium to large landslides in both the areas show that the distribution can be described by a power-law with good fit over three order of area magnitude. Similar to previous observations for historical landslides in mountainous areas (Van Den Eeckhaut et al., 2007), the frequency distribution, exhibits negative power-law for larger landslides together with a significant rollover for smaller landslides. Rollover location for Niigata was around $5 \times 10^3 \text{ m}^2$ while the same for Ehime was around $10 \times 10^3 \text{ m}^2$. The slope of the area–frequency relationship (γ) in both the areas (Niigata $\gamma = -2.132$ and Ehime $\gamma = -2.037$) suggest the dominance of larger landslides (Malamud and Turcotte, 1999; Ohmori and Sugai, 1995).

A comparison of exponents of the power-law for all landslides in the area with the landslides that were included in this study (below 95th percentile) hints a low probability of occurrence of extremely large landslides in both the study areas (**Figure 3-3** and **Figure 3-4**). This observation in addition to the uncertainties associated with extremely large events explained earlier suggests that inclusion of such events might negatively influence the usability of LSM.
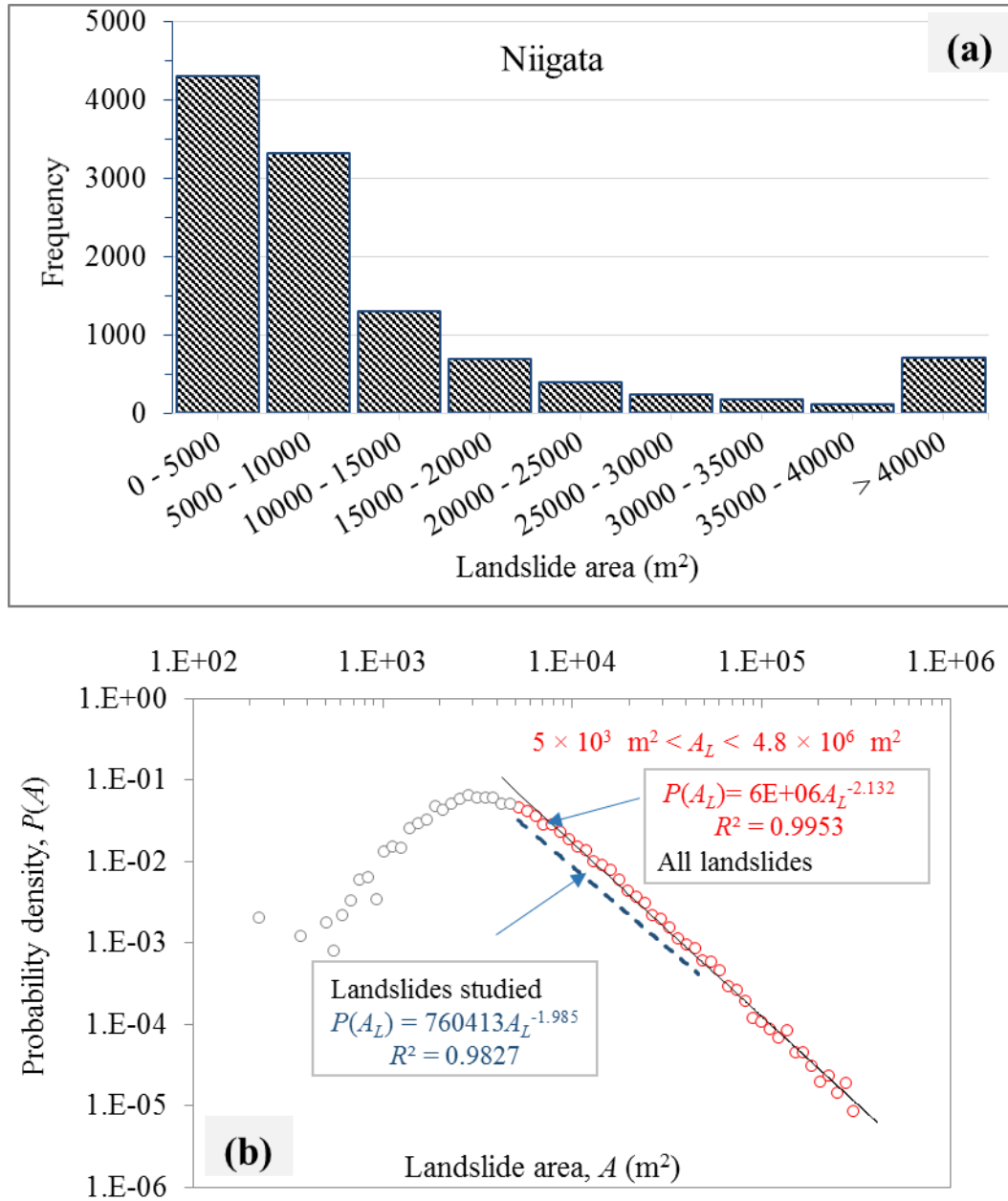
Figure 3-3. Landslide characteristics (Niigata). (a) Histogram showing the distribution of landslide areas. (b) Probability distribution of landslide areas
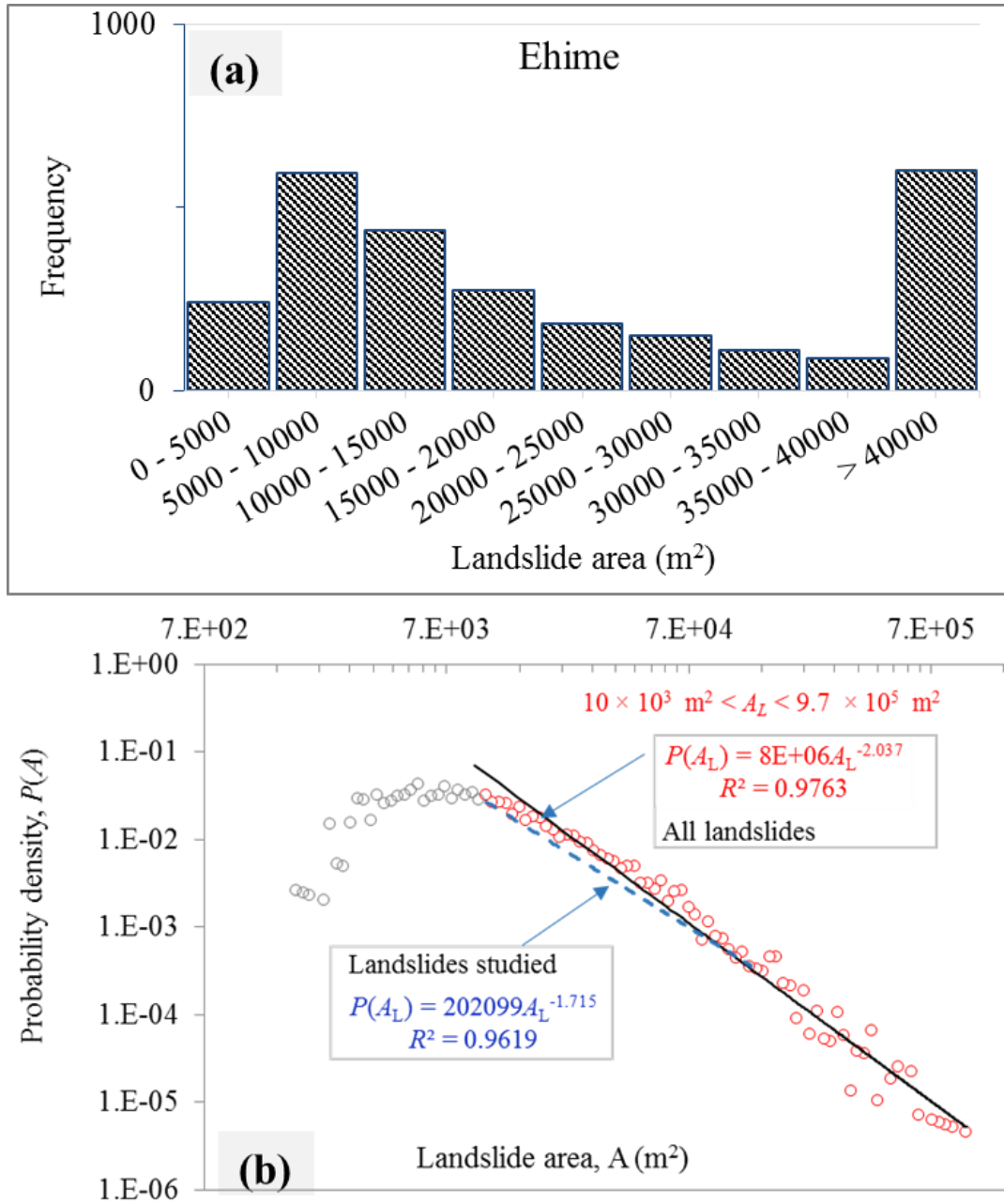
Figure 3-4. Landslide characteristics (Ehime). (a) Histogram showing the distribution of landslide areas. (b) Probability distribution of landslide areas.

Following the Gorkha-earthquake, a large group of scientists and organizations, motivated by humanitarian needs, focused on satellite-based mapping of earthquake-induced landslides. The British Geological Society and partners at the Durham University prepared the coseismic and postseismic landslide inventory used in this study (British Geological Survey, 2015; Durham University, 2015). About 5500 landslides in the disaster-hit areas were identified by the interpretation of various types of satellite data (obtained via the International Charter Space and Major Disasters and directly from data suppliers) including WorldView, UK-DMC2, SPOT, Pleiades, and RADARSAT-2. Landslides in the inventory (**Figure 3-5**) are represented as polylines that extend from the head scarp to the toe and it includes 576 landslides within the study area. Their length ranges from 19 to 1350 m with a mean of 232 m. **Figure 3-6** presents their length-frequency distribution.
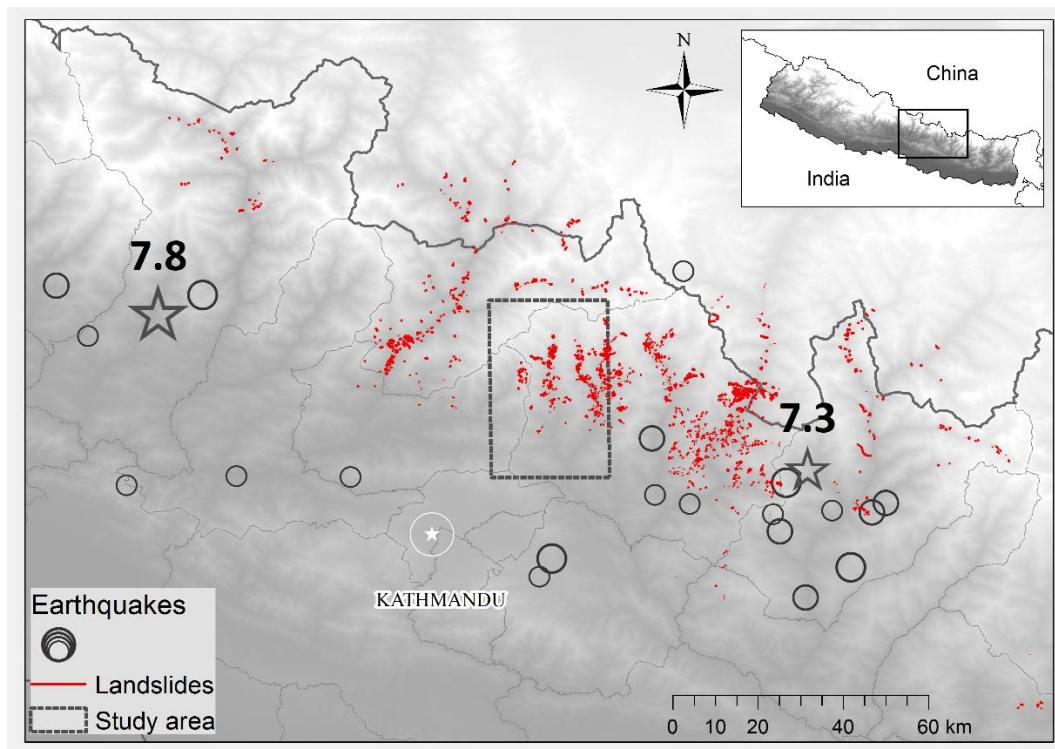
Figure 3-5. Landslides in Nepal after the 2015 Gorkha-earthquake.
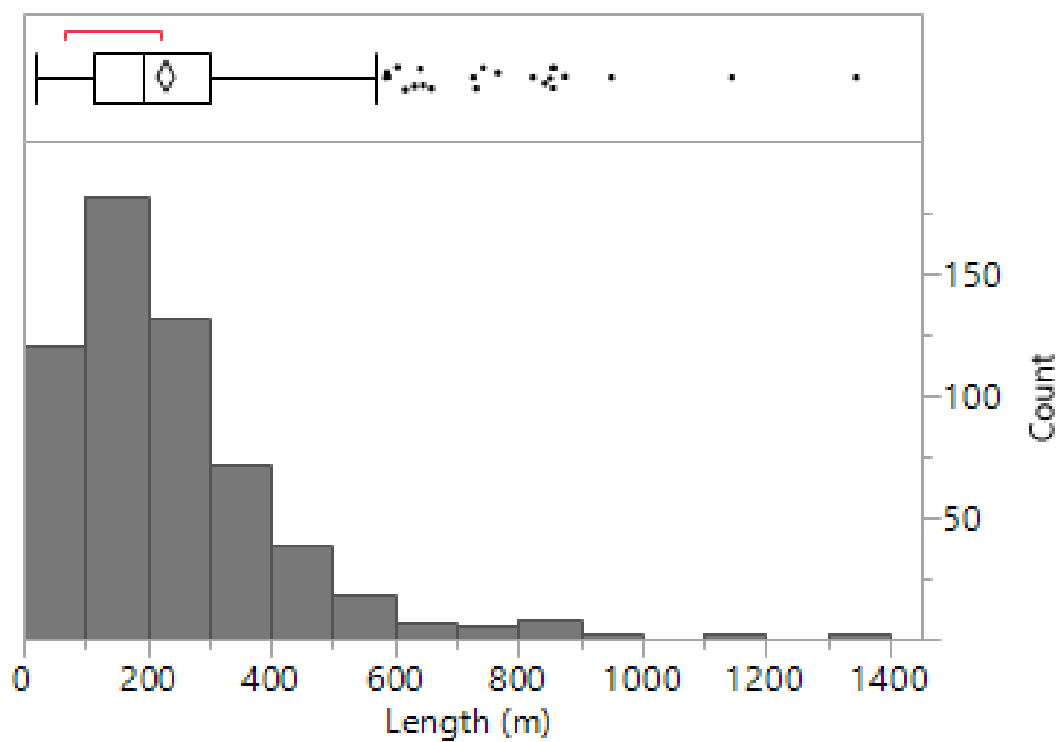
Figure 3-6. Length-frequency histogram of landslides in the study area.

3.2. Topographic and geological data

A 10-m DEM obtained from the GSI (Geospatial Information Authority of Japan) was used for topographic analysis in Japan. The minimum elevation for both study areas is about 5 m, whereas the mean and maximum elevations are, respectively, 369 and 1284 m for Niigata and 825 and 1895 m for Ehime.

A 30-m topographic data collected by the Shuttle Radar Topography Mission (SRTM) was used for the LS study in Nepal. The freely downloadable SRTM 1 Arc-Second Global elevation data (http://earthexplorer.usgs.gov/) offer worldwide coverage at a resolution of 1 arc-second (ca. 30 m). This global dataset was recently made public (starting 8th October 2014) and is the most reliable data source of global elevation in terms of spatial accuracy and coverage (Mondal et al., 2016). The elevation in the study area ranges from 769 to 5899 m with a mean of 2815 m.

The geological information of the study areas in Japan (**Figure 3-7**) is based on the seamless digital geological maps (scale: 1:200,000) provided by the Geological Survey of Japan (GSJ; Takeuchi and Yanagisawa, 2004). The data consisted of lithological information and location of known fault lines. The study area in Niigata consists of two major geological groups: sedimentary rocks and volcanic rocks further divided into 19 substrata according to the major lithological types and geological age (**Table 3-2**). Similarly, the study area in Ehime comprises of four major geological groups: sedimentary rocks, accretionary complex, metamorphic rocks, and plutonic rocks further divided into 18 substrata (**Table 3-2**). **Figure 3-8** shows the lithological composition of the study areas. Non-marine sediments form the dominant lithological type in Niigata and peltic schist followed by mafic schist dominate in Ehime.
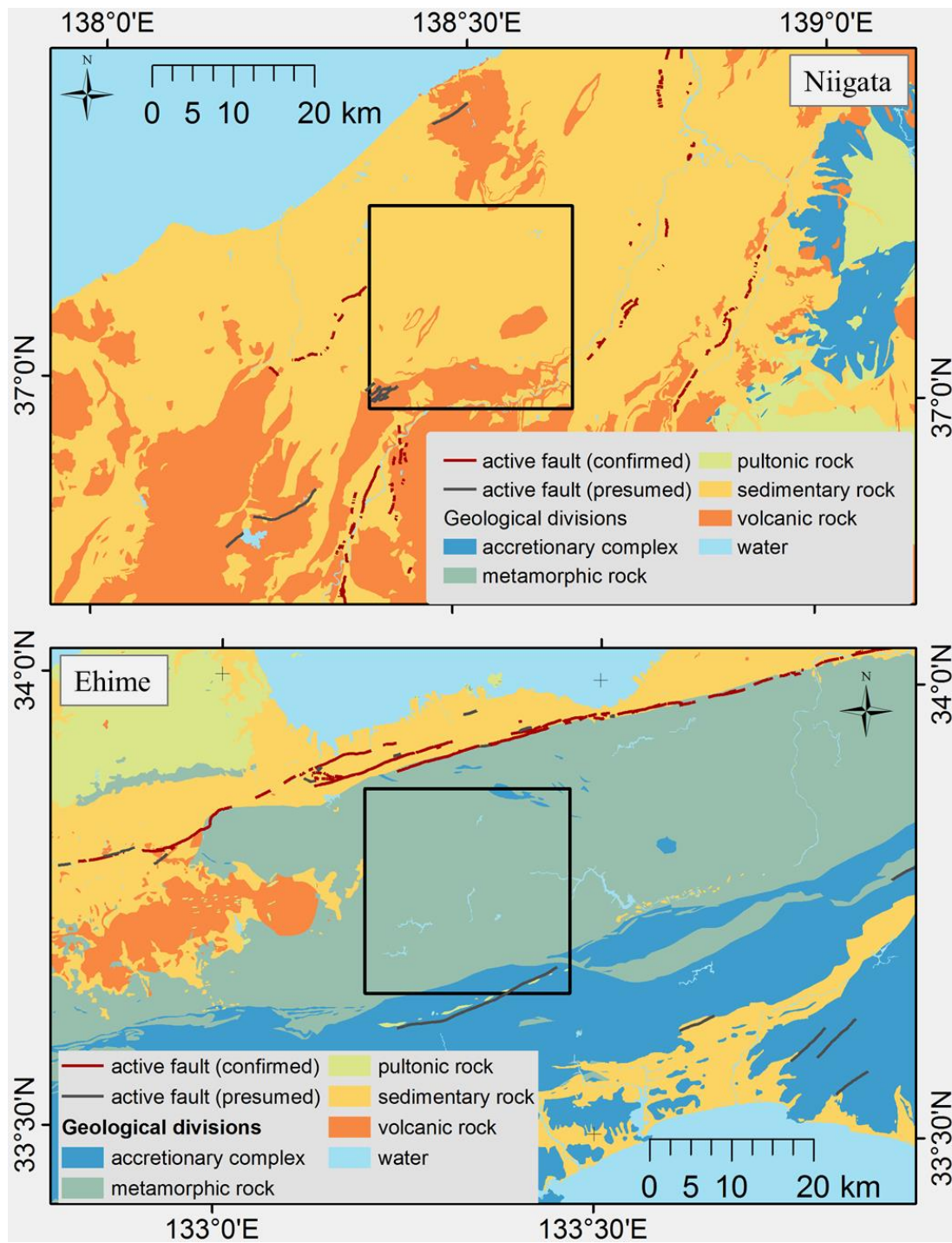
Figure 3-7. Geological maps of the study areas in Japan.

Table 3-2. Classification of geological substrata for the study areas in Japan.

	Code	Geological divisions	Lithological types	Age	% Area
Niigata	1	sedimentary rock	reclaimed land	Holocene	0.0001
	3	sedimentary rock	fan deposits	Late Pleistocene to Holocene	0.0048
	10	sedimentary rock	marine and non-marine sediments	Late Pleistocene to Holocene	0.0160
	20	sedimentary rock	marine and non-marine sediments	Late Pleistocene	0.0189
	22	sedimentary rock	lower terrace	Late Pleistocene	0.0224
	23	sedimentary rock	middle terrace	Late Pleistocene	0.0171
	24	sedimentary rock	higher terrace	Middle Pleistocene	0.0014
	30	sedimentary rock	marine and non-marine sediments	Middle Pleistocene	0.0054
	40	sedimentary rock	marine and non-marine sediments	Early Pleistocene	0.0629
	50	sedimentary rock	non-marine sediments	Late Miocene to Pliocene	0.3927
	60	sedimentary rock	non-marine sediments	Middle to Late Miocene	0.0867
	70	sedimentary rock	marine sedimentary rocks	Middle Miocene to Pliocene	0.1935
	711	volcanic rock	volcanic debris	Holocene	0.0213
	804	volcanic rock	non-alkaline felsic volcanic intrusive rocks	Late Miocene to Pliocene	0.0044
	826	volcanic rock	non-alkaline felsic volcanic rocks	Middle to Late Miocene	0.0305
	1010	volcanic rock	non-alkaline mafic volcanic rocks	Middle Pleistocene	0.0016
	1020	volcanic rock	non-alkaline mafic volcanic rocks	Early Pleistocene	0.1151
	1030	volcanic rock	non-alkaline mafic volcanic rocks	Late Miocene to Pliocene	0.0013
	2000	water	water	water	0.0040
Ehime	1	sedimentary rock	reclaimed land	Holocene	0.0001
	430	accretionary complex	melange matrix of J1-3 accretionary complex	Early to Late Jurassic	0.0417
	431	accretionary complex	sandstone of J1-3 accretionary complex	Early to Late Jurassic	0.0087
	437	accretionary complex	basalt block of J1-3 accretionary complex	Carboniferous to Permian	0.0083
	438	accretionary complex	limestone block of J1-3 accretionary complex	Carboniferous to Permian	0.0014
	439	accretionary complex	chert block of J1-3 accretionary complex	Carboniferous to Middle Jurassic	0.0138
	555	accretionary complex	ultramafic rocks	unknown age	0.0085
	1270	pultonic rock	felsic plutonic rocks	Middle to Late Miocene	0.0013
	1599	metamorphic rock	pelitic schist	Cretaceous	0.5058
	1600	metamorphic rock	psammitic schist	Cretaceous	0.0111

1620	metamorphic rock	mafic schist	Cretaceous	0.2323
1631	metamorphic rock	siliceous schist	Cretaceous	0.0346
1632	metamorphic rock	schist	Cretaceous	0.0099
1633	metamorphic rock	pelitic schist	Cretaceous	0.0713
1636	metamorphic rock	mafic schist	Cretaceous	0.0309
1638	metamorphic rock	siliceous schist	Cretaceous	0.0128
1640	metamorphic rock	mafic schist	Cretaceous	0.0020
2000	water	water	water	0.0055
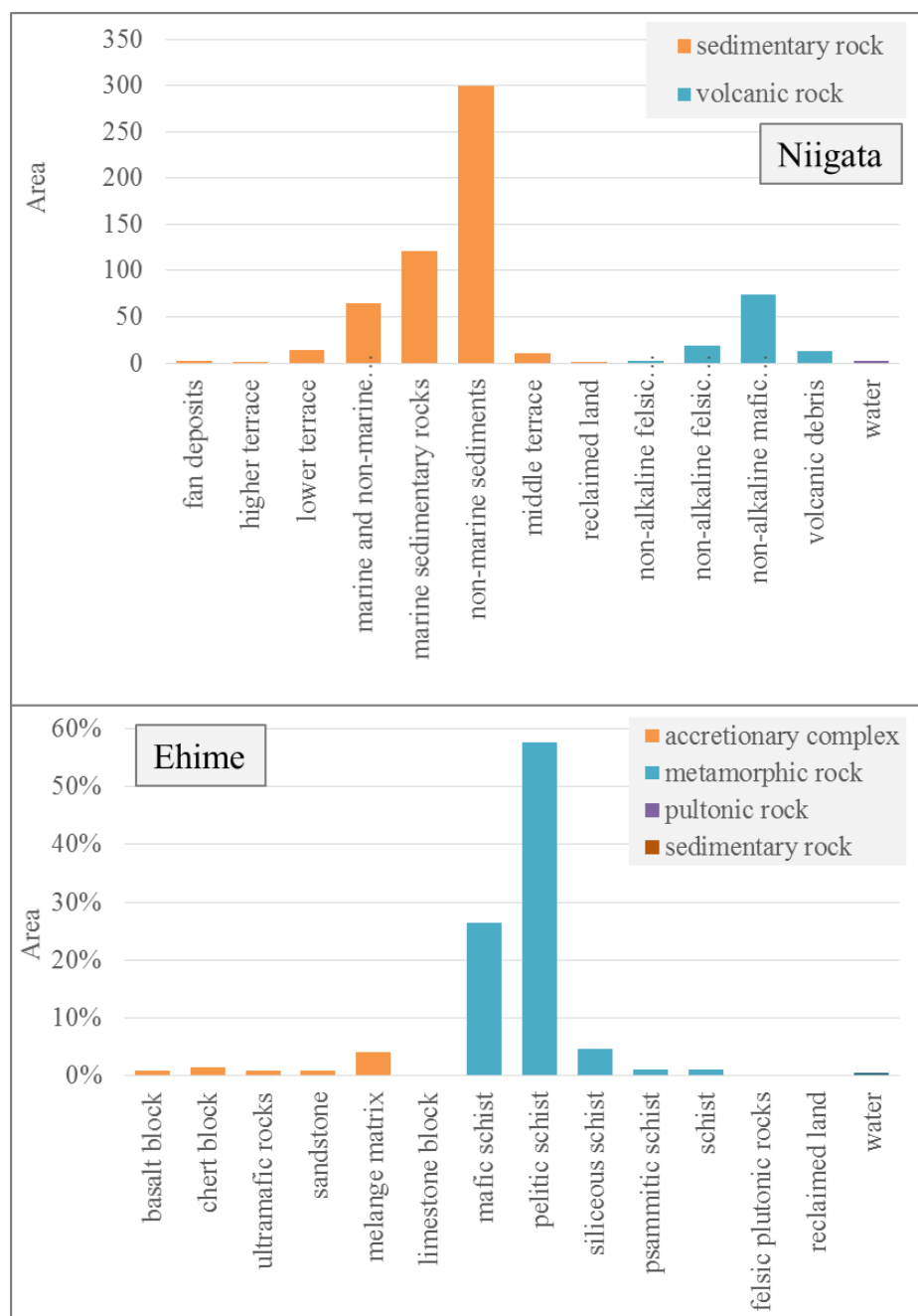
Figure 3-8. Lithological composition of the study areas.

3.3. Topographical parameters

The 10 m DEM was aggregated to create DEMs with 30, 60, 90, 120, 150, and 300 m cell sizes. The resulting seven DEM-scales were used to investigate the optimal DEM resolution for each landslide causative factor.

Various primary and secondary topographic attributes have been proposed and analyzed to characterize geomorphological characteristics of an area. This study employs 16 DEM-derived topographic parameters **(Table 3-3)** previously used in landslide research (Castellanos Abella and Van Westen, 2008; Chen and Yu, 2011; Guzzetti et al., 1999; Nefeslioglu et al., 2008a).

3.3.1. Elevation

Elevation (*El*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is a measure of the height of a surface above mean sea level and is considered an important causative factor, which influence slope stability. *El* is usually associated with landslides by the virtue of secondary factors detailed below.

3.3.2. Slope

Slope (*Sl*) (**Figures 3-9a**, **3-10a**, and **3-11a**) indicates the degree of inclination of the surface and shows the rate of elevation change. Slope gradient has a great influence on the susceptibility of a slope to landsliding and is frequently used in LS research (Yalcin, 2008). Gravity is the primary driving force for a landslide to occur. As slope gradient increases, the level of gravitation-induced shear stress in the colluviums or residual soils increases thereby reducing slope stability. Gentle slopes are expected to have a low frequency of landslides because of generally lower shear stresses associated with low gradients (Dai et al., 2001).

Table 3-3. Topographic parameters used for susceptibility modeling.

S.N.	Parameters	Abbreviation	Significance
1	Slope aspect	*Asp*	Solar insolation, evapotranspiration, species distribution and abundance
2	Total curvature	*Cr*	Total measure of surface curvature
3	Distance to drainage network	*Dtd*	Influence of fluvial processes
4	Distance to ridge	*Dtr*	Tectonics especially amplification of seismic shaking, accumulation of flow
5	Drainage density	*Dd*	Intensity of fluvial processes, stages of channelization
6	Drop	*Dr*	Strict hydrological slope and geomorphological slope
7	Elevation	*El*	Climate, vegetation, potential energy
8	Elevation-relief ratio	*Er*	Stages of landscape development; characterization of general topography
9	Internal relief	*Ir*	Characteristic of terrain roughness
10	Profile curvature	*Pfc*	Flow acceleration, erosion, deposition rate
11	Plan curvature	*Plc*	Converging/diverging flow, soil water content, soil characteristics
12	Slope	*Sl*	Velocity of surface and subsurface flow, soil water content
13	Stream power index	*SPI*	Measures erosive power of flowing water
14	Sediment transport capacity index	*STCI*	Net erosion and deposition rates; transportation capacity and erosion.
15	Terrain characterization index	*TCI*	Descriptor of terrain shapes and spatial variability of soil depths
16	Topographic wetness index	*TWI*	Soil moisture conditions and variability of soil types
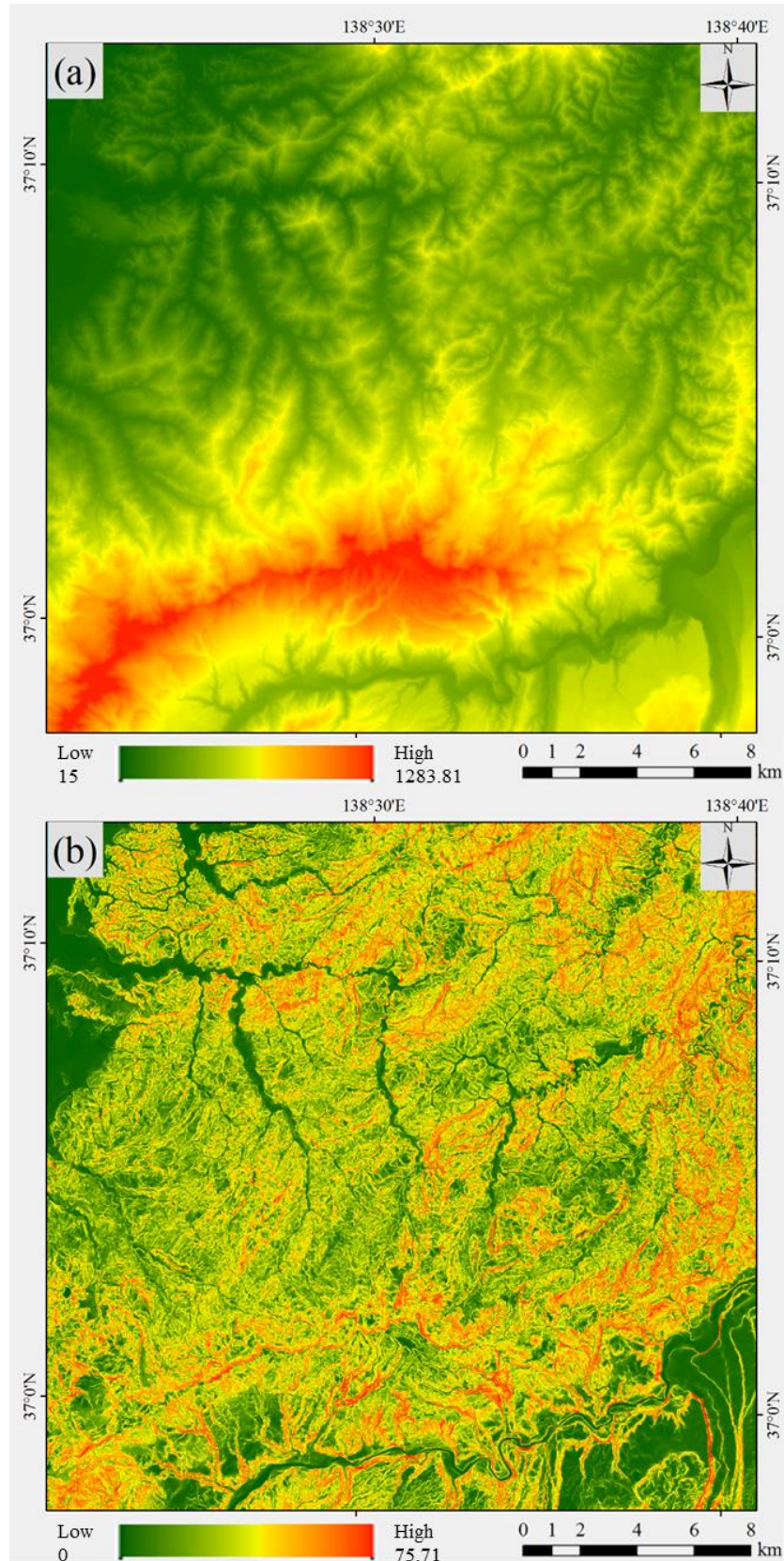
Figure 3-9. Maps of topographic factors for Niigata.
(a) *El*, (b) *Sl*, (c) *Dr*, (d) *Asp*, (e) *Pfc*, (f) *Plc*, (g) *Cr*, (h) *Dd*, (i) *Dtd*, (j) *Dtr*, (k) *Ir*,
(l) *Er*, (m) *STCI*, (n) *SPI*, (o) *TCI*, and (p) *TWI*.
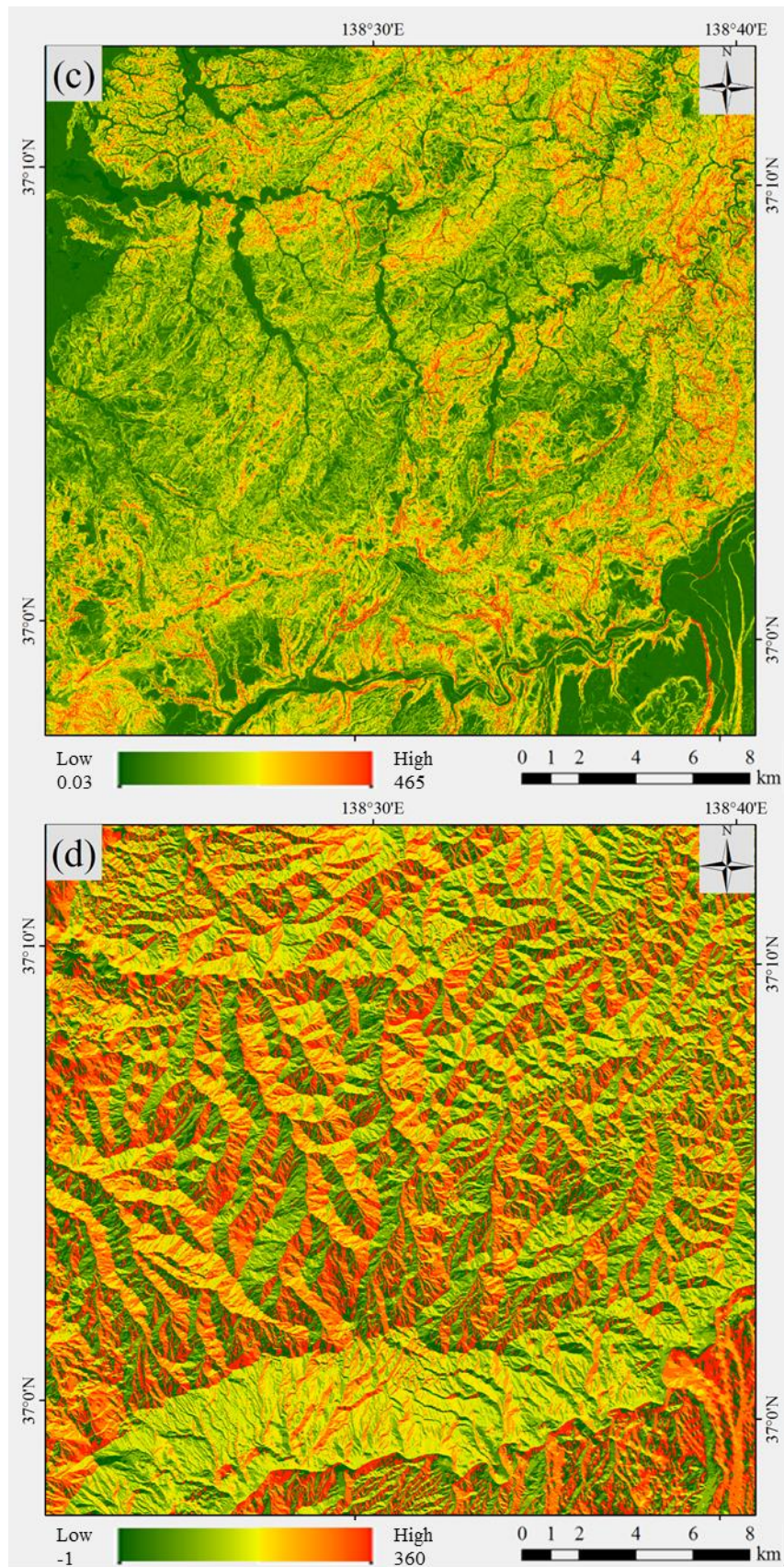
Figure 3-9. Continued.
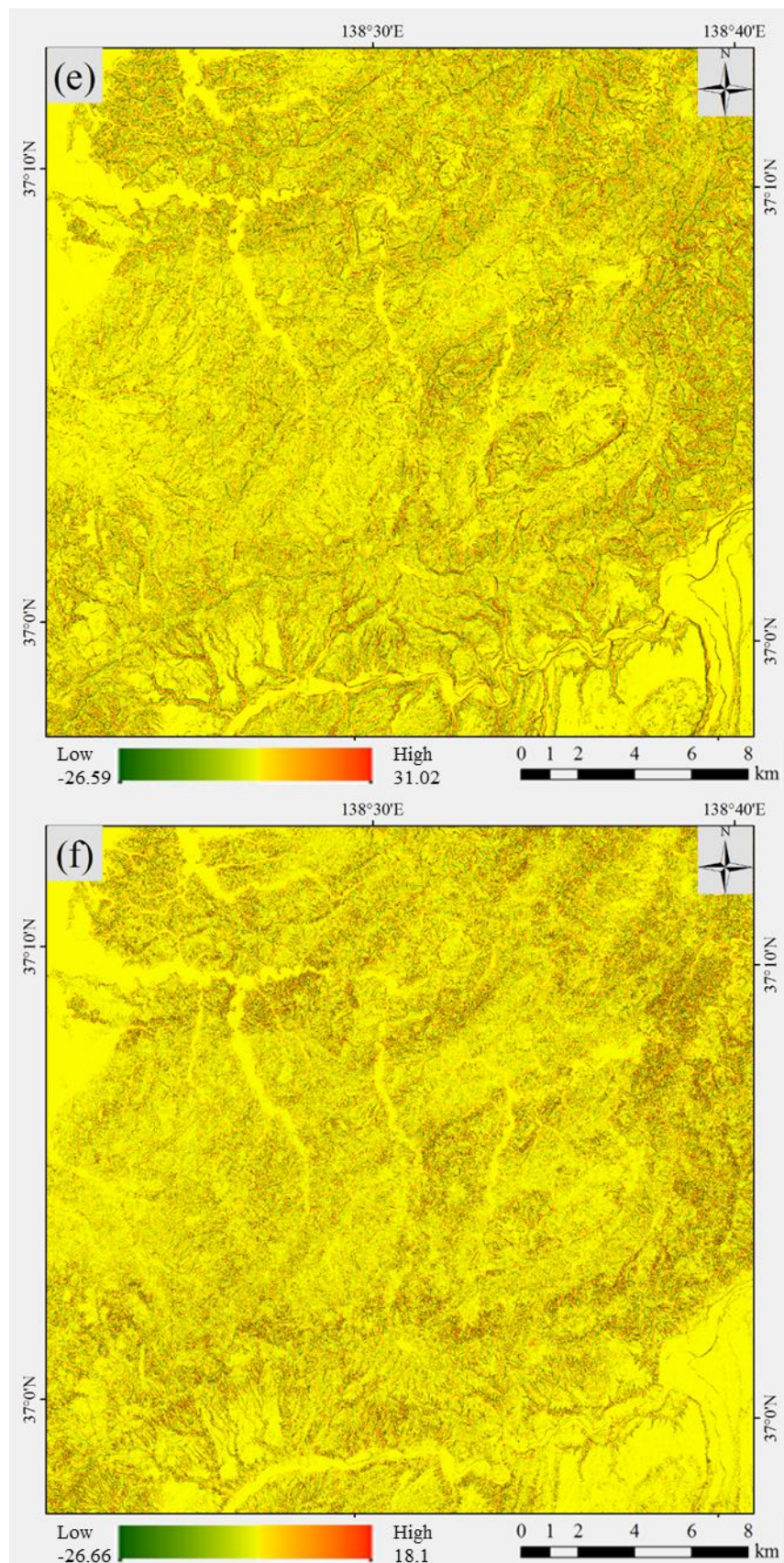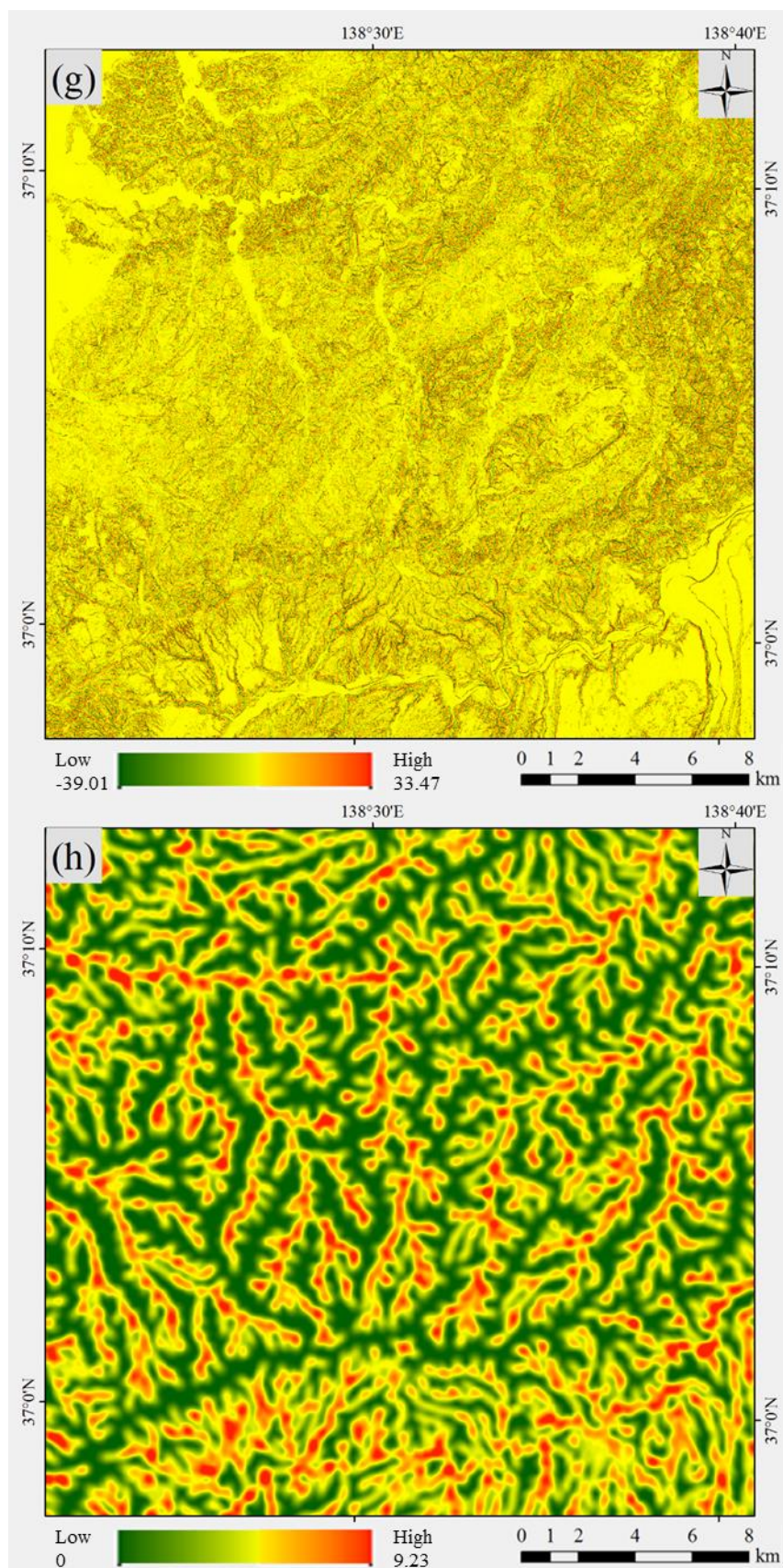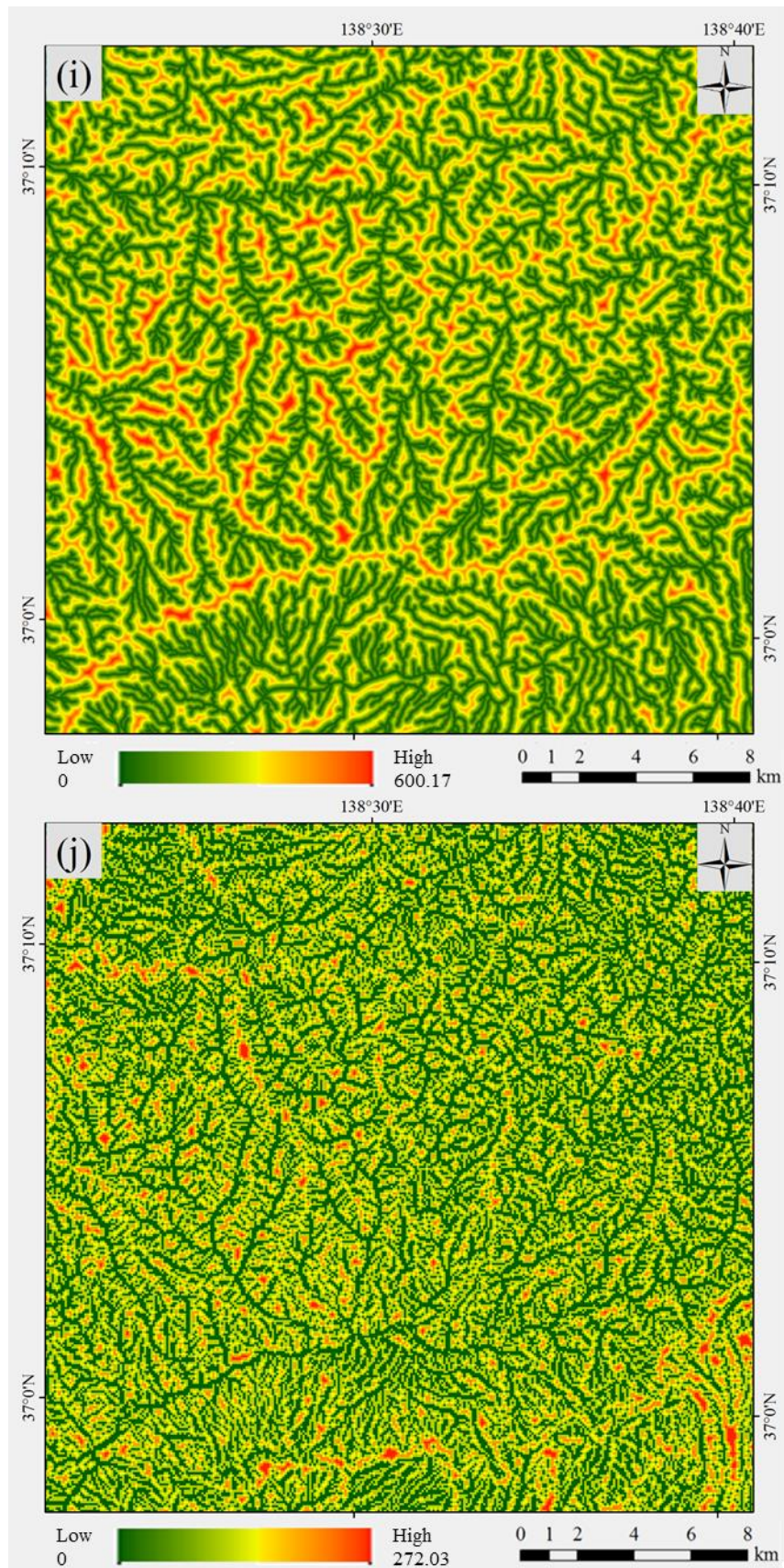
Figure 3-9. Continued.

Figure 3-9. Continued.

Figure 3-9. Continued.
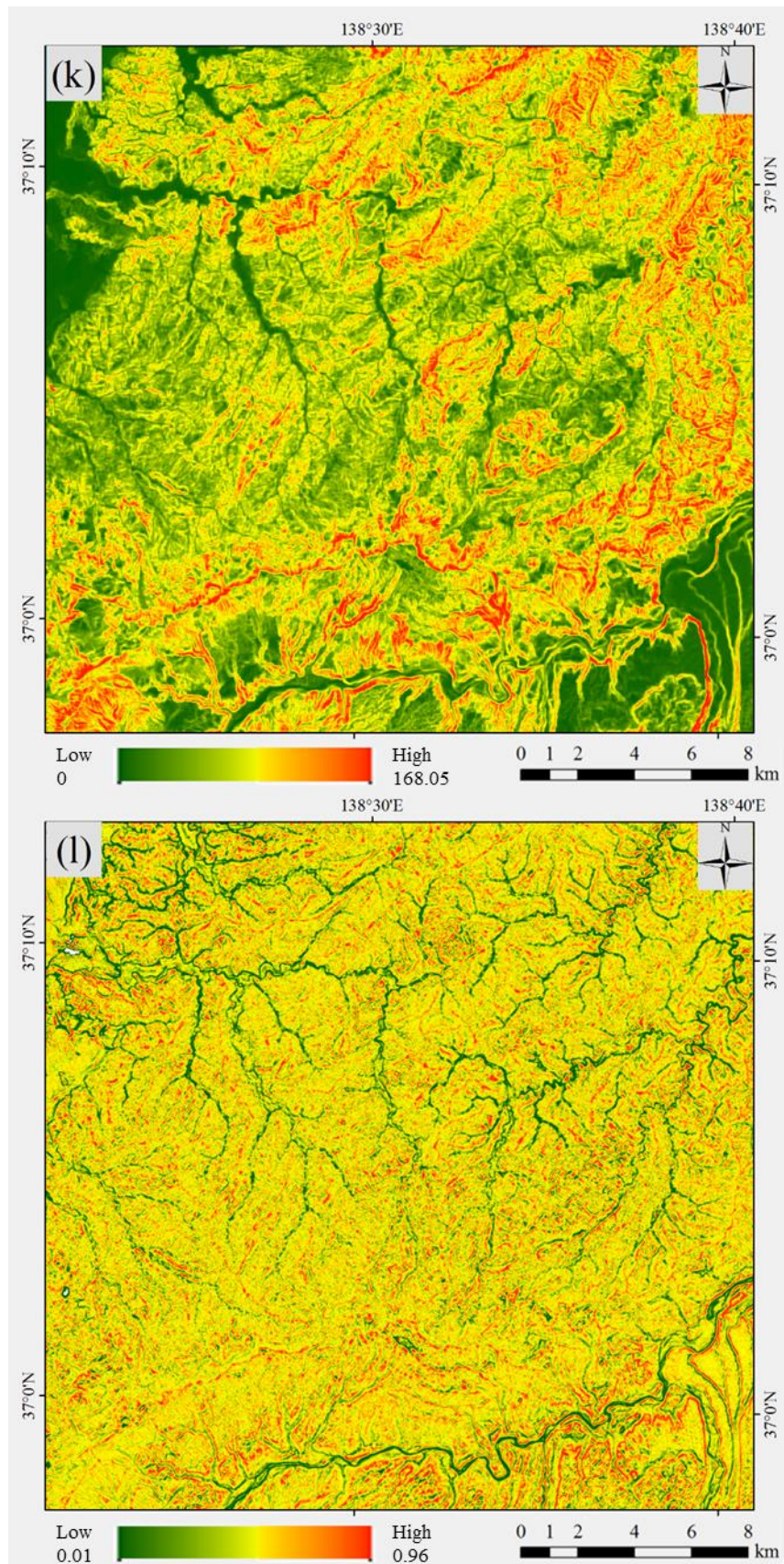
Figure 3-9. Continued.
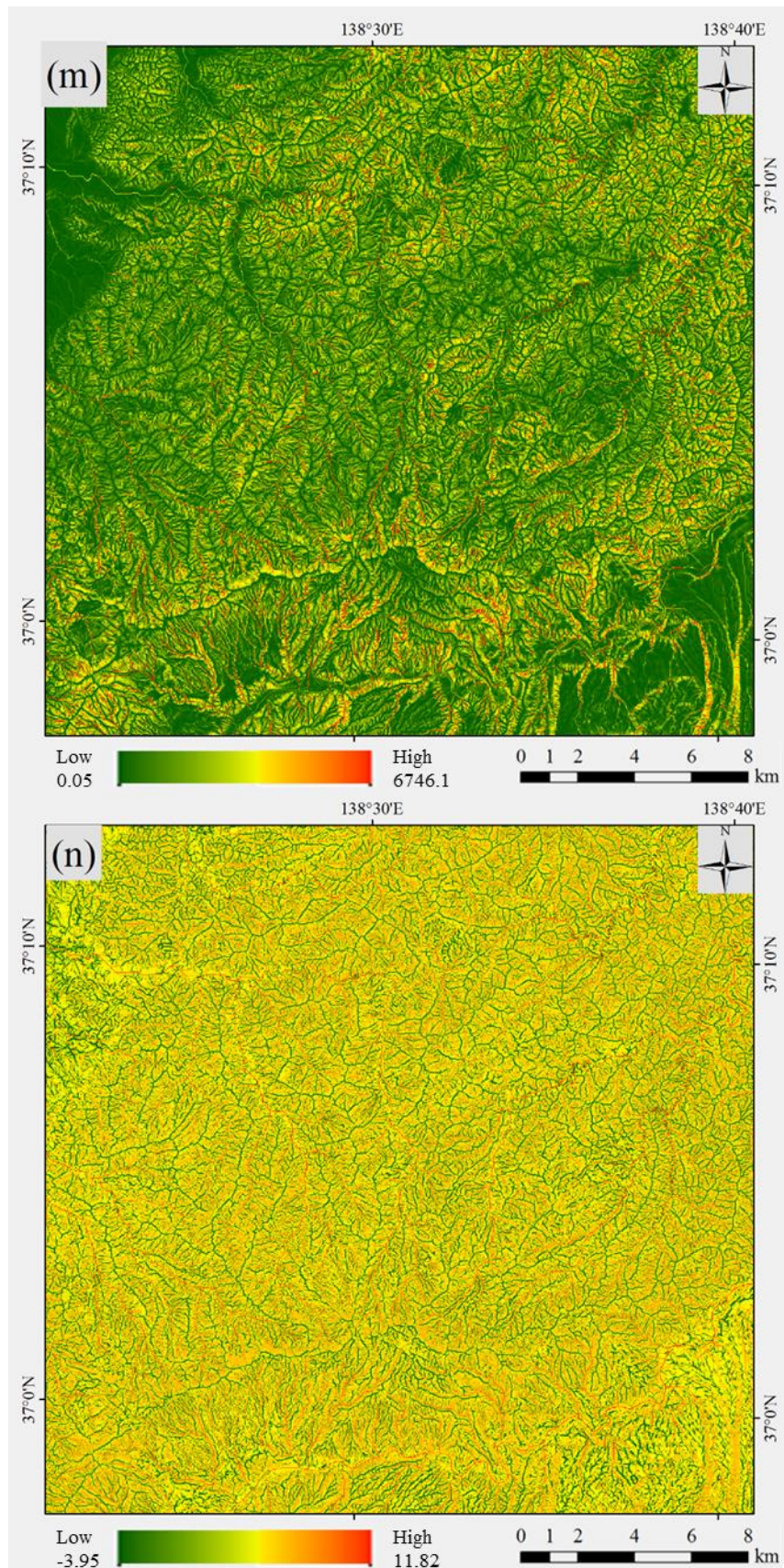
Figure 3-9. Continued.
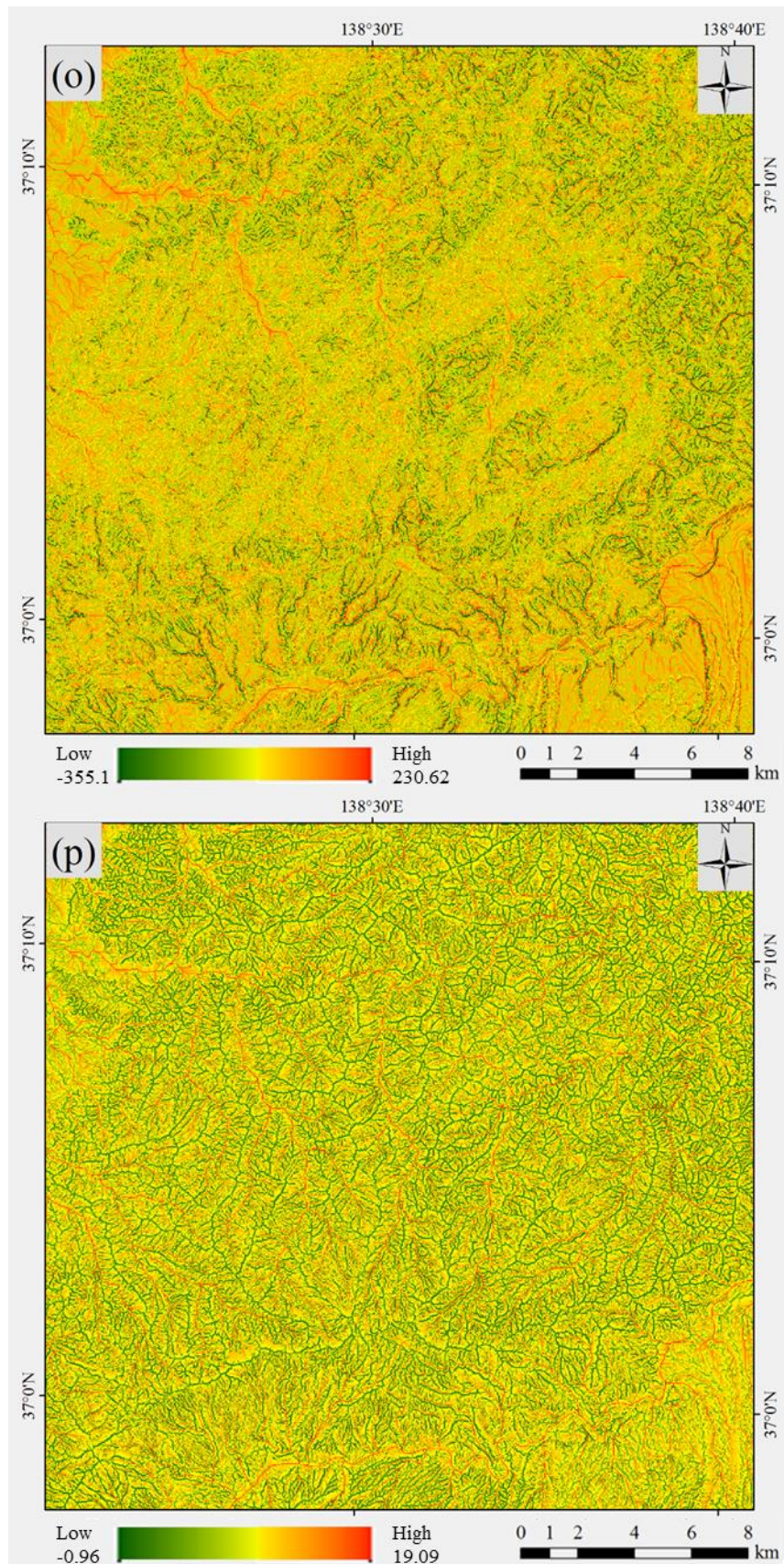
Figure 3-9. Continued.

.
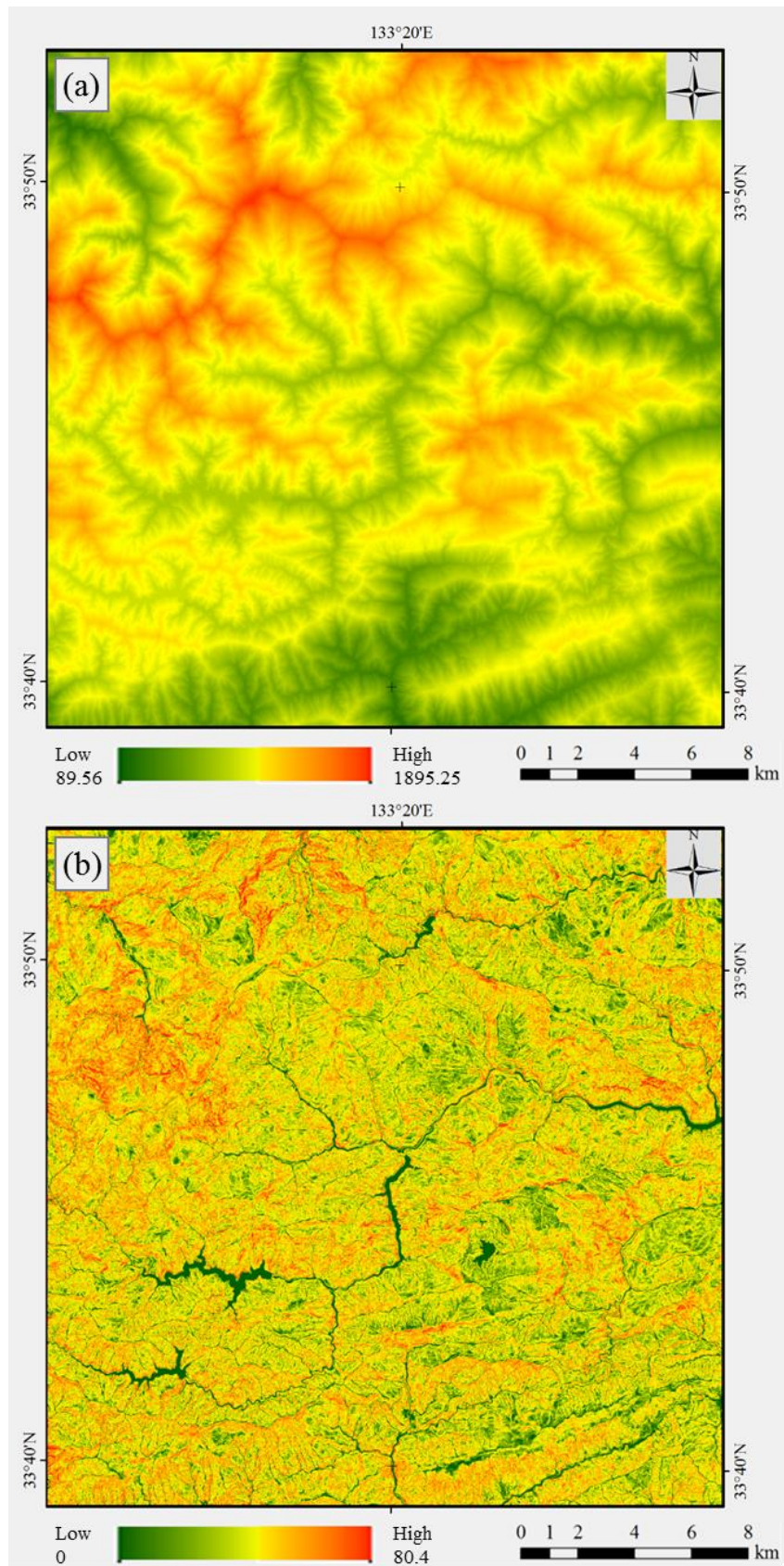
Figure 3-10. Maps of topographic factors for Ehime.
(a) *El*, (b) *Sl*, (c) *Dr*, (d) *Asp*, (e) *Pfc*, (f) *Plc*, (g) *Cr*, (h) *Dd*, (i) *Dtd*, (j) *Dtr*, (k) *Ir*,
(l) *Er*, (m) *STCI*, (n) *SPI*, (o) *TCI*, and (p) *TWI*.
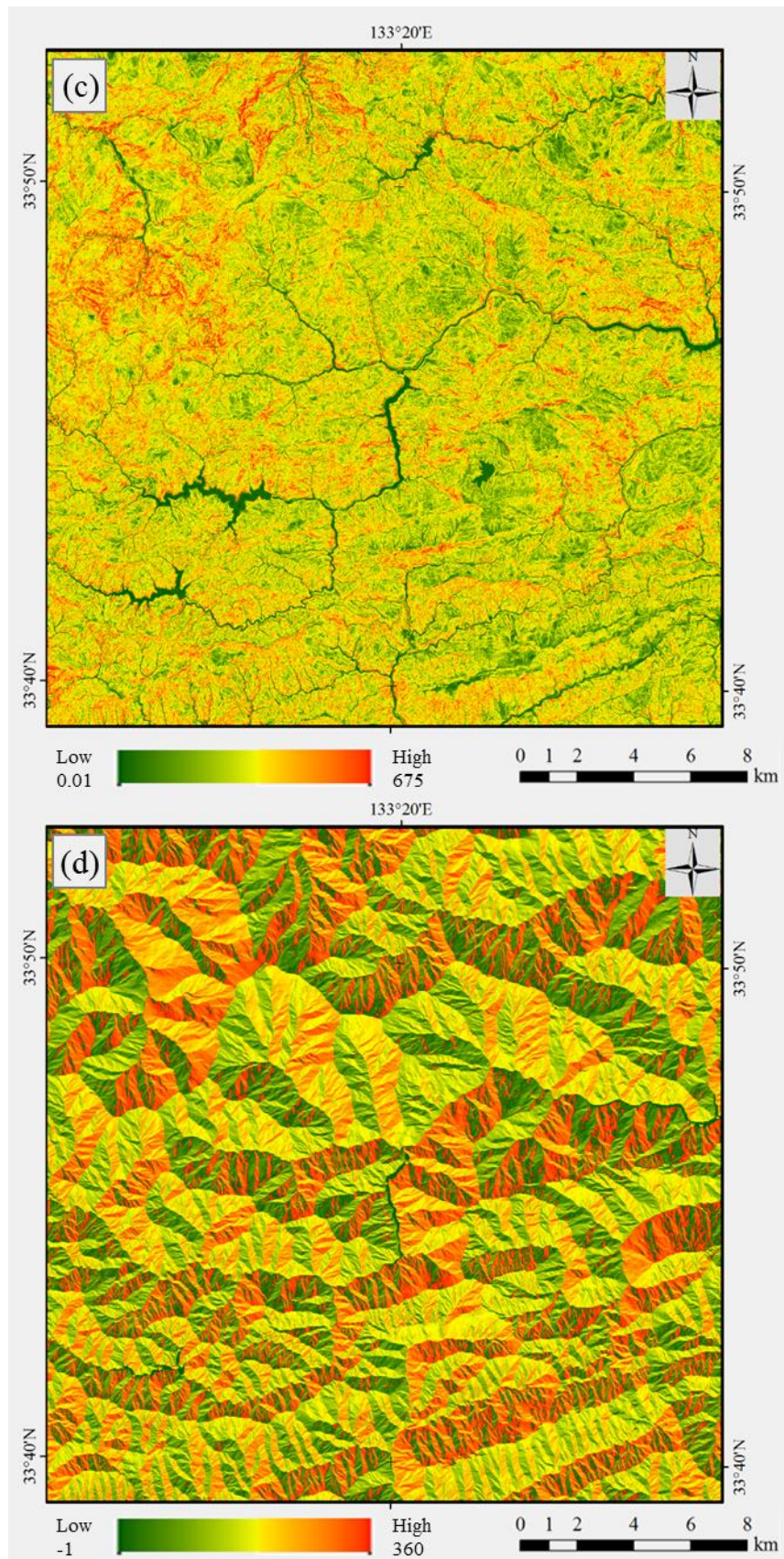
Figure 3-10. Continued.
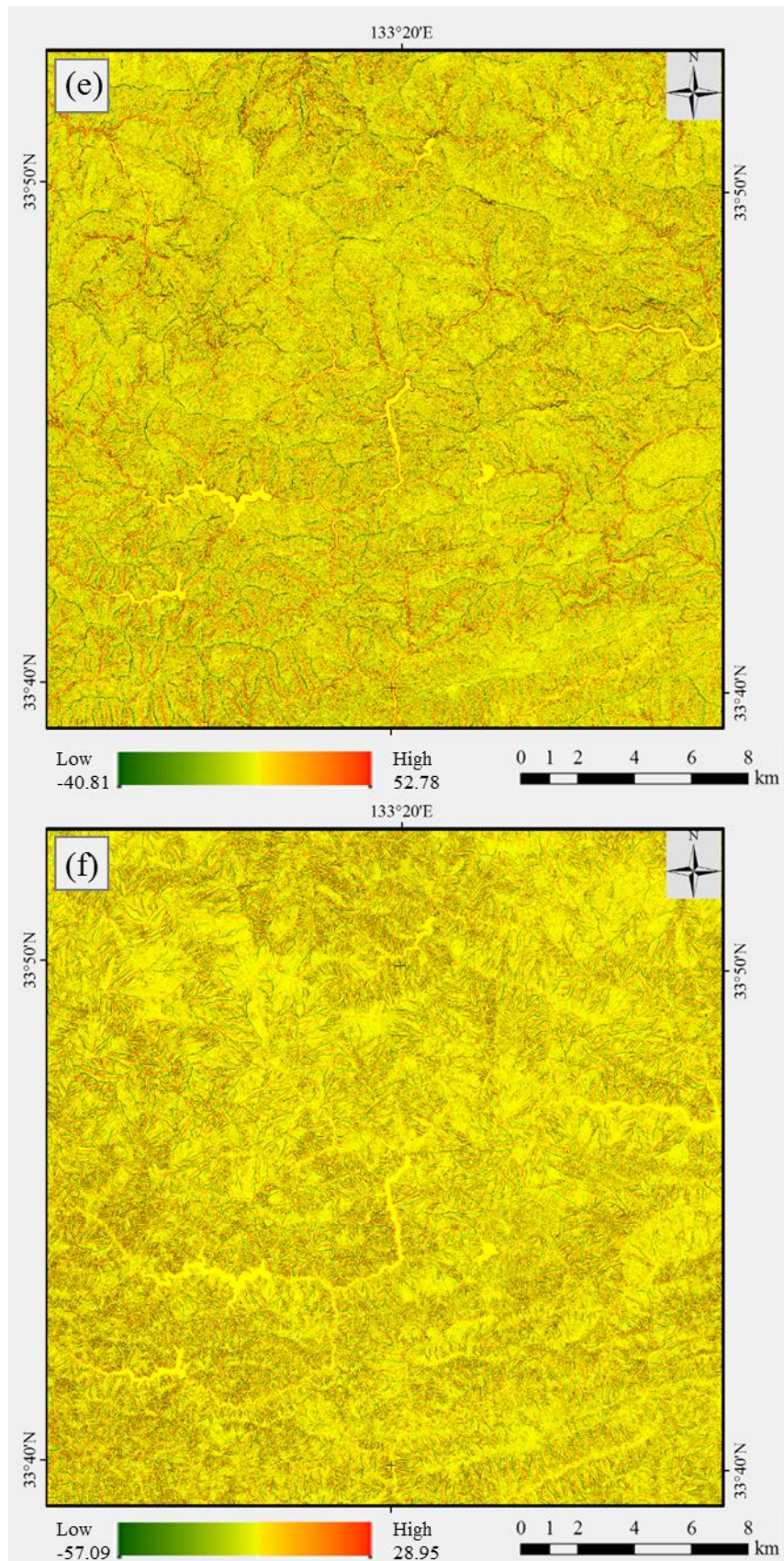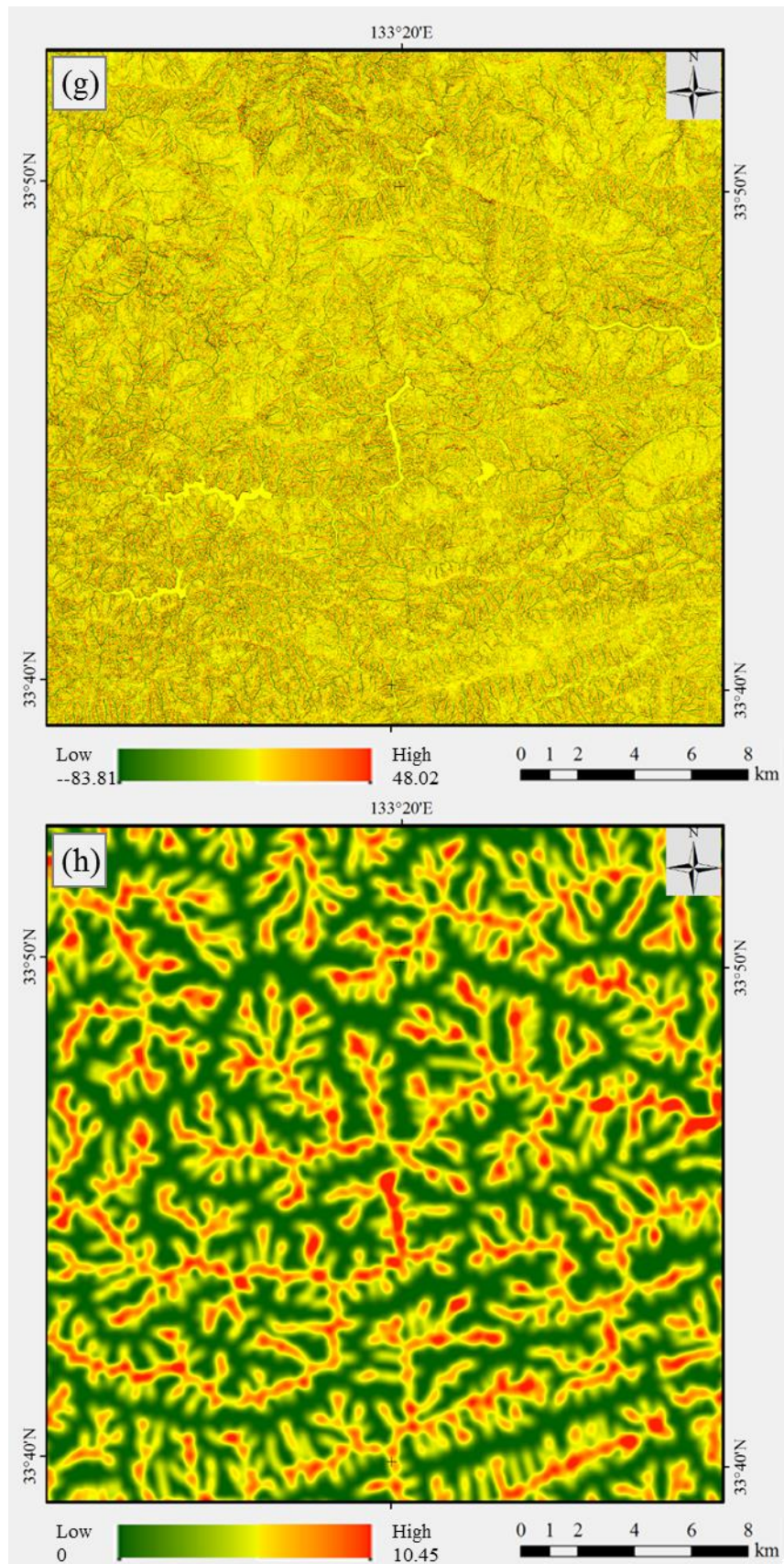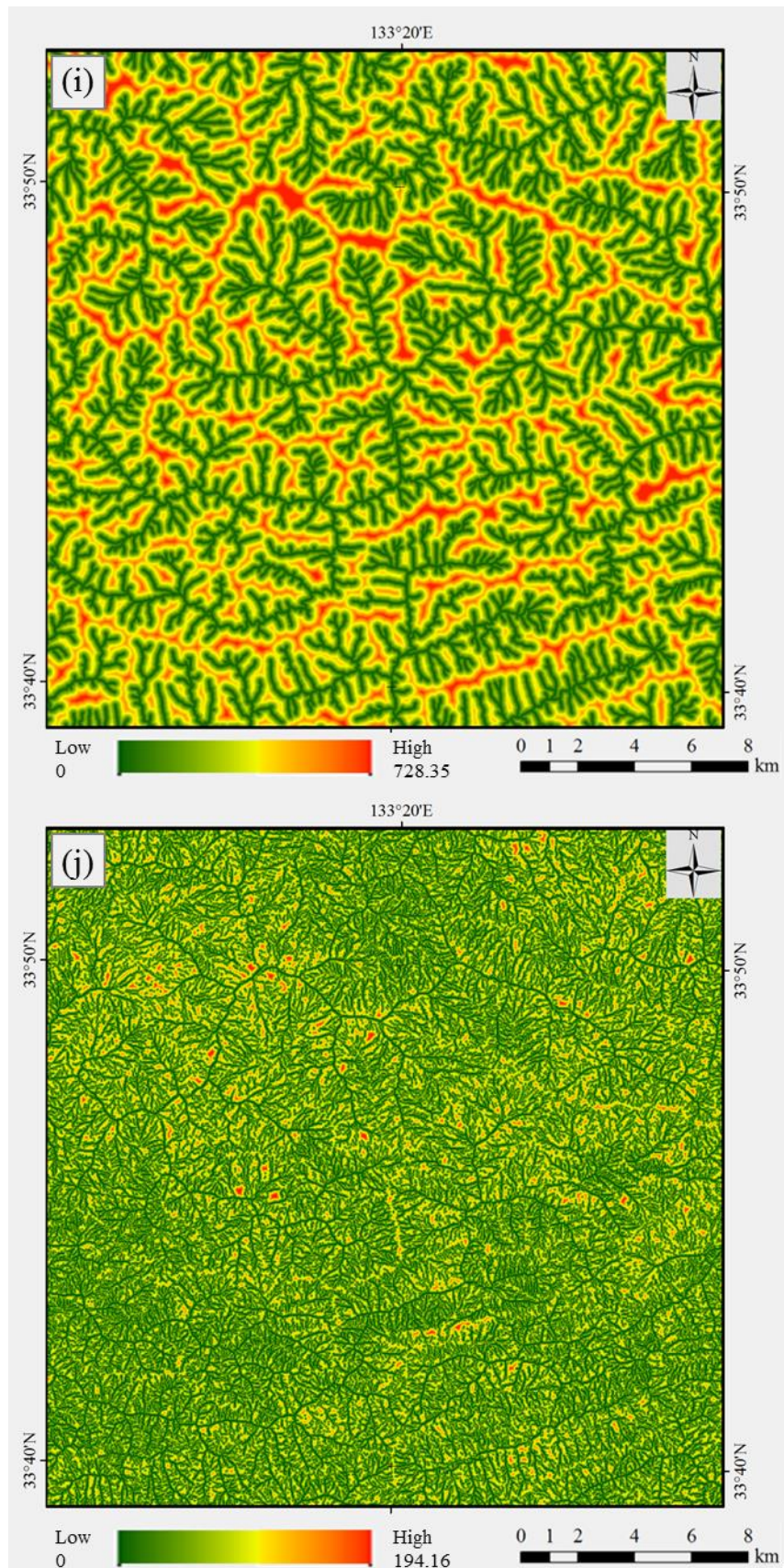
Figure 3-10. Continued.

Figure 3-10. Continued.

Figure 3-10. Continued.

Figure 3-10. Continued.

Figure 3-10. Continued.

Figure 3-10. Continued.

49
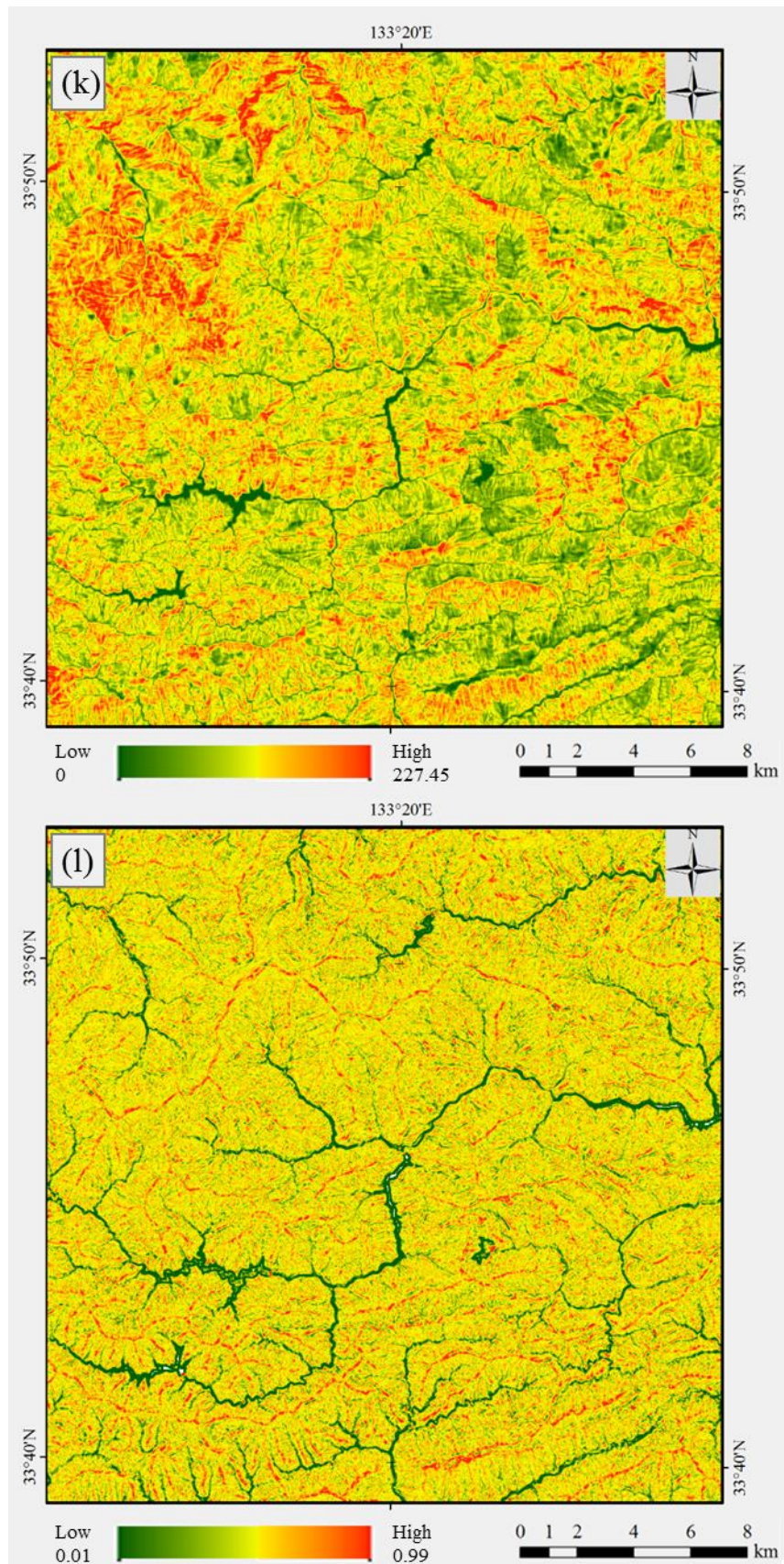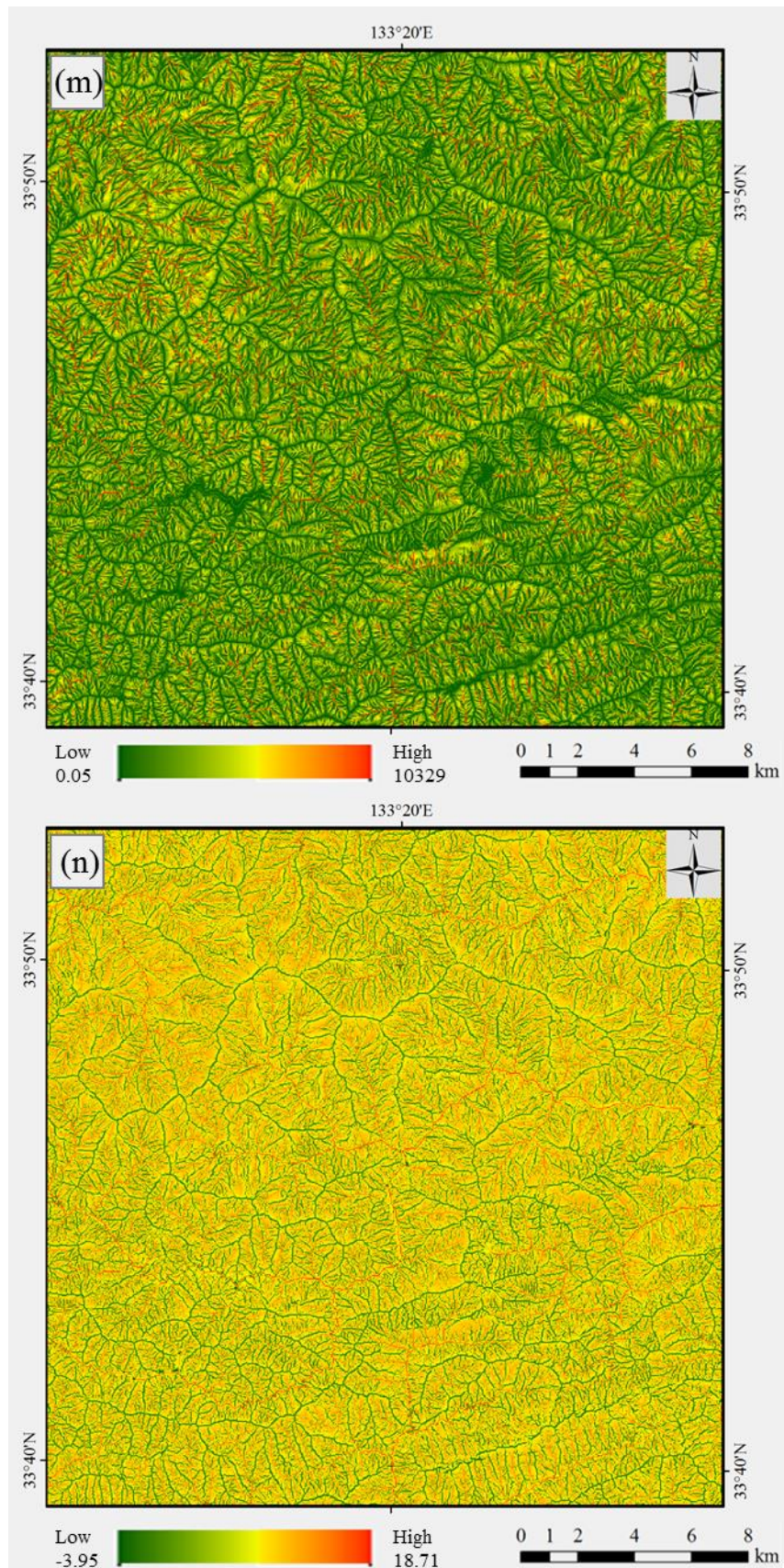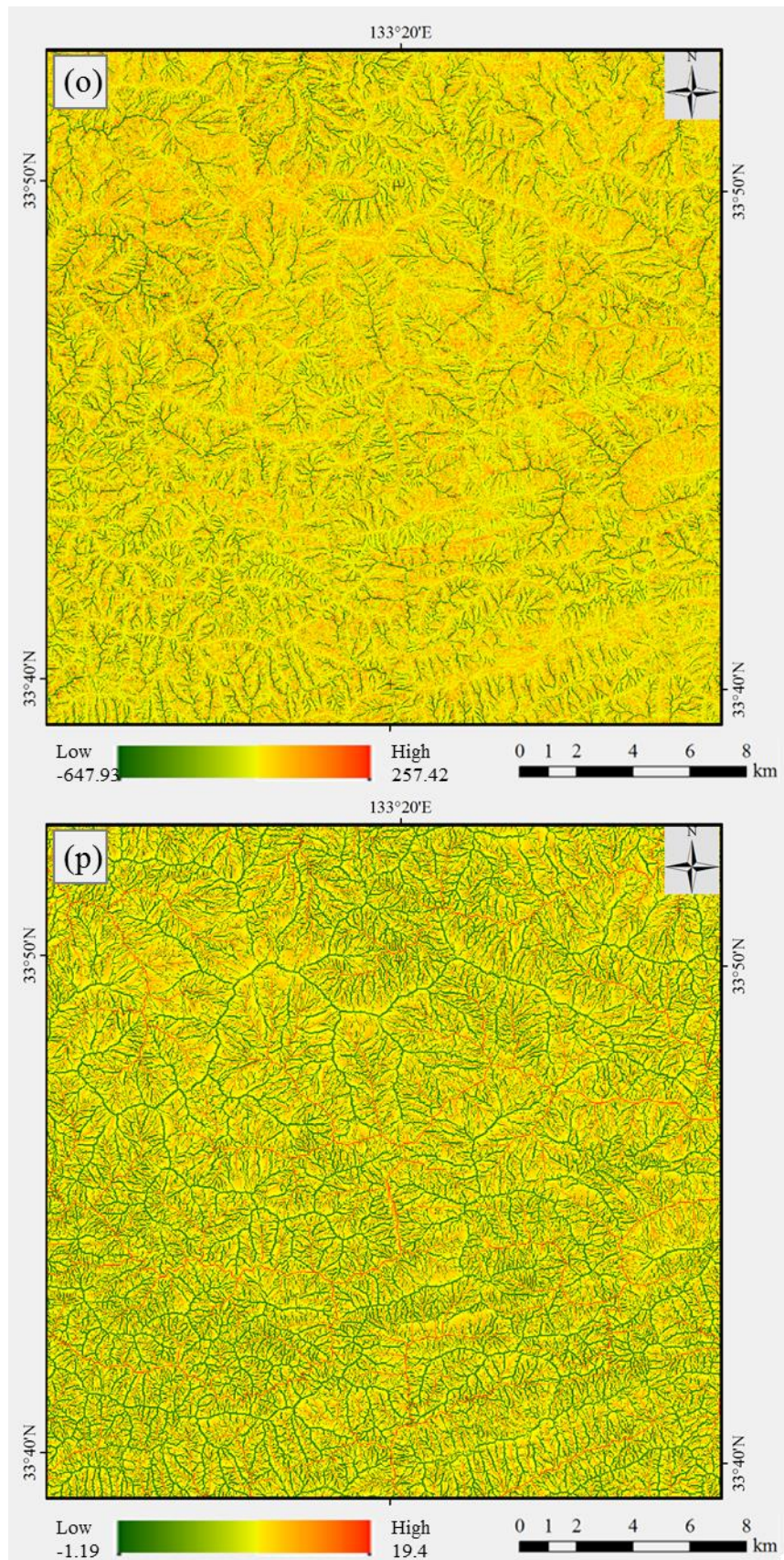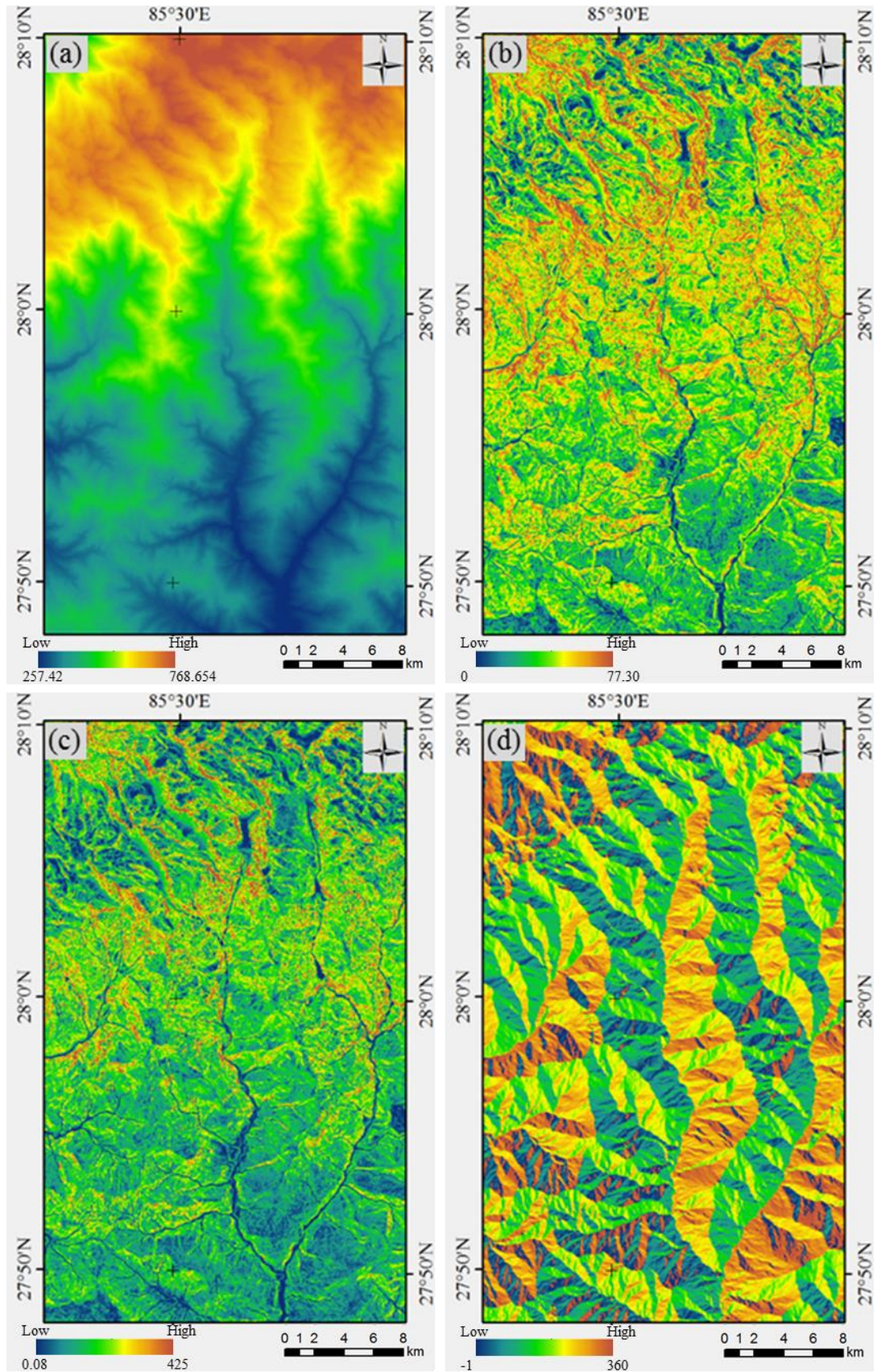
Figure 3-11. Maps of topographic factors for Melamchi, Nepal.
(a) *El*, (b) *Sl*, (c) *Dr*, (d) *Asp*, (e) *Pfc*, (f) *Plc*, (g) *Cr*, (h) *Dd*, (i) *Dtd*, (j) *Dtr*, (k) *Ir*,
(l) *Er*, (m) *STCI*, (n) *SPI*, (o) *TCI*, and (p) *TWI*.

50
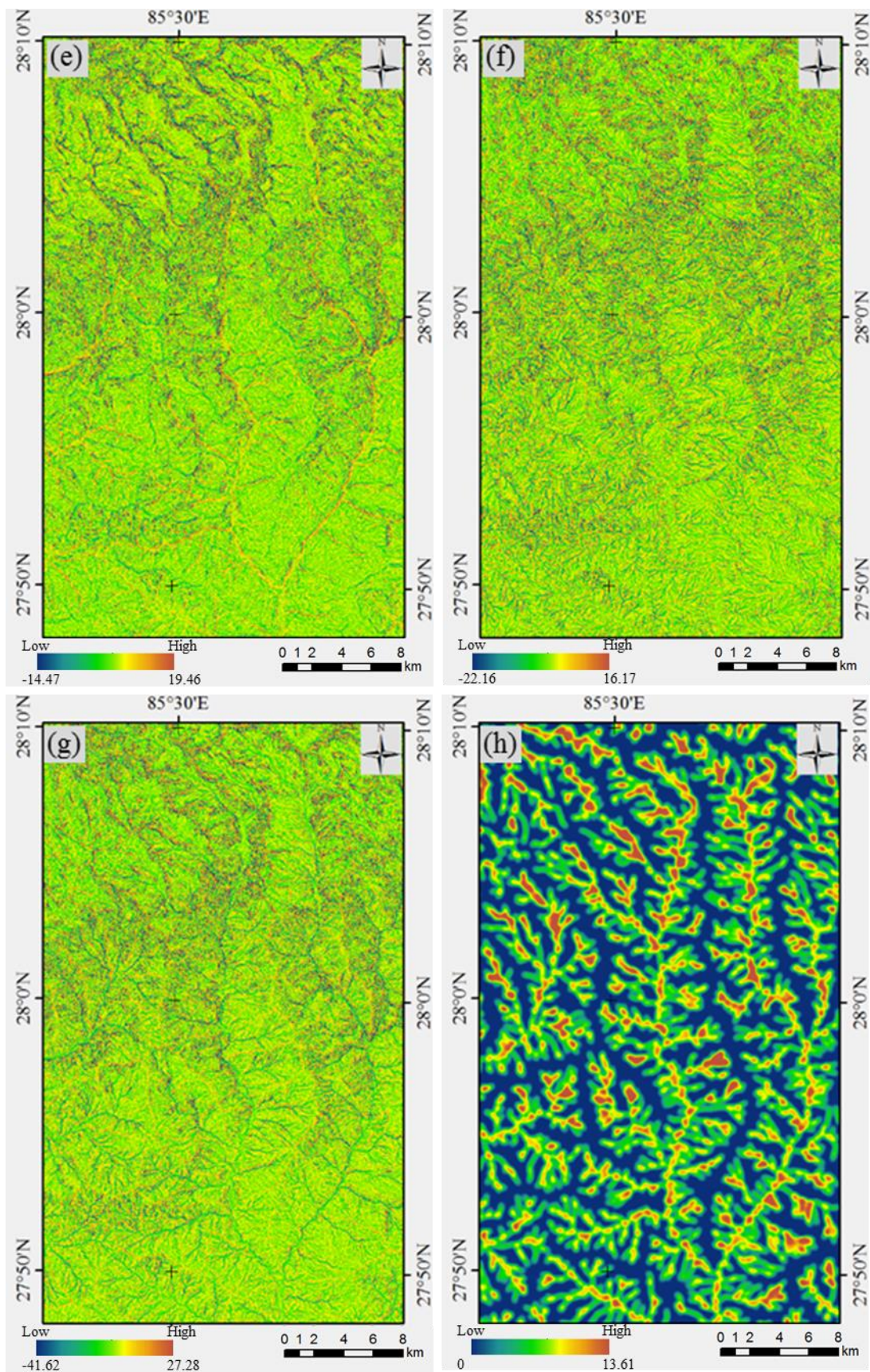
Figure 3-11. Continued.
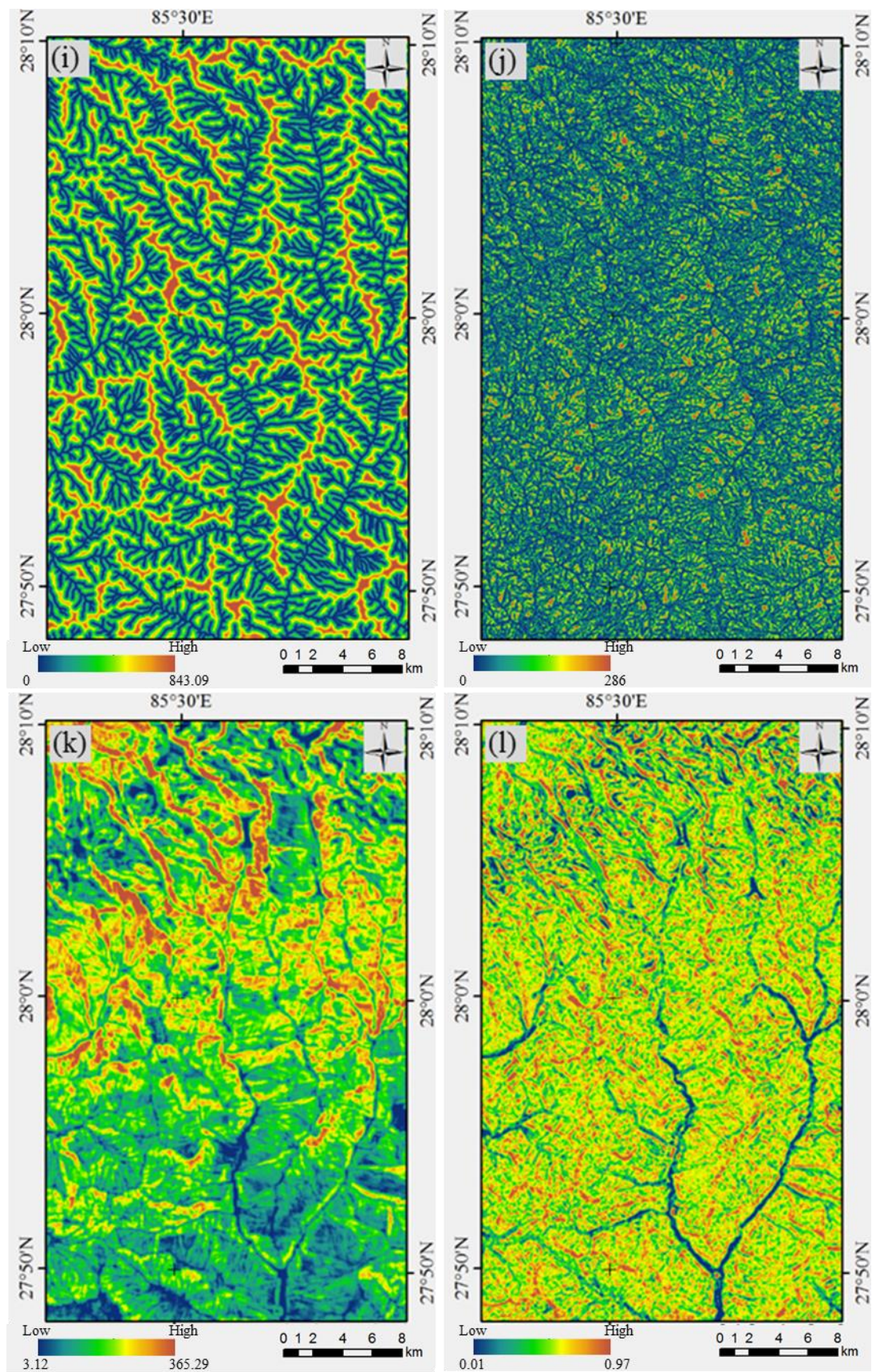
Figure 3-11. Continued.
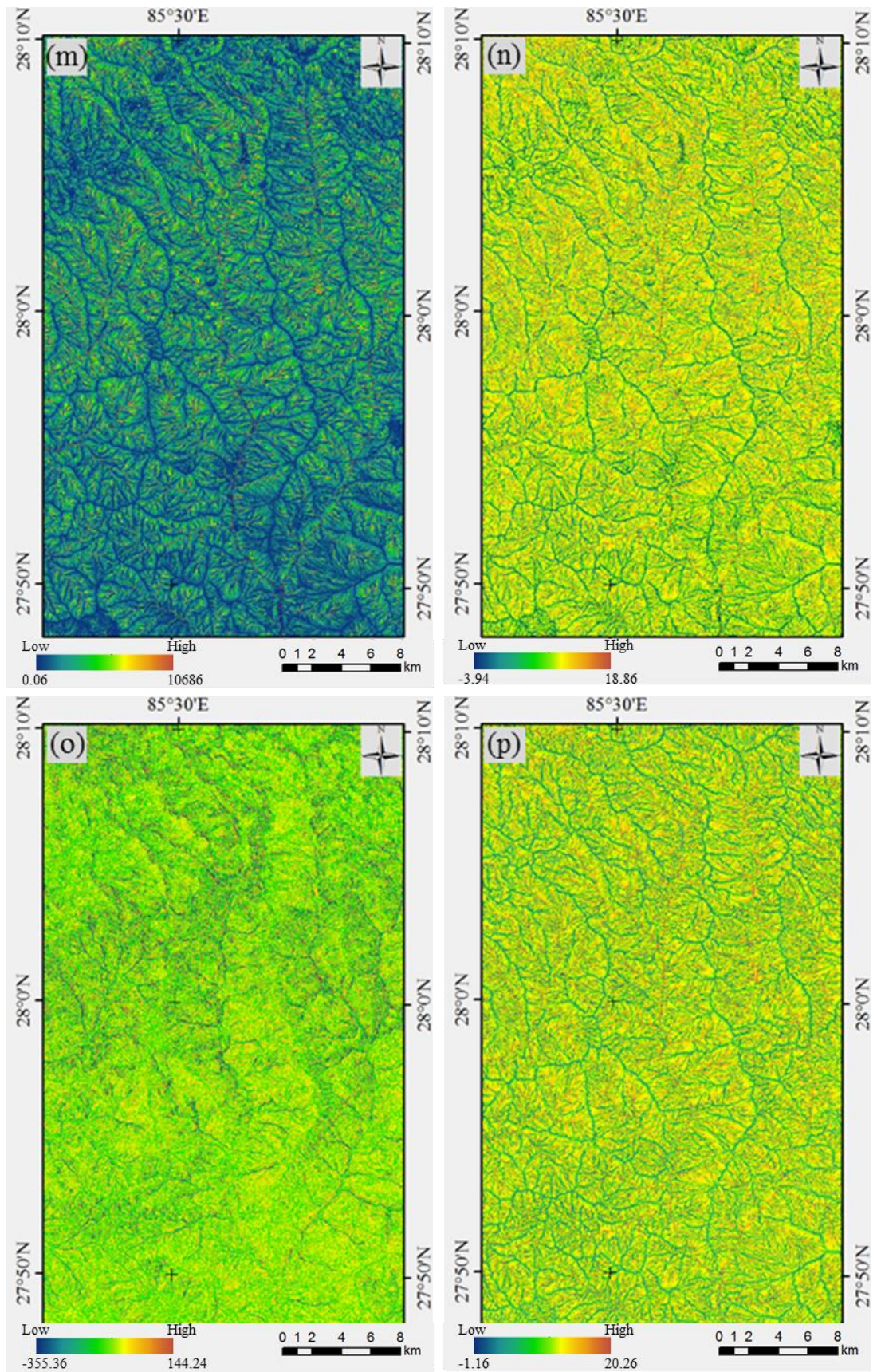
Figure 3-11. Continued.

3.3.3. Drop

Drop (*Dr*) (**Figures 3-9a**, **3-10a**, and **3-11a**), equivalent to the hydrologic slope (Claessens et al., 2005; Tarboton, 1997), shows the ratio of the maximum change in elevation along the direction of flow between cell-centers. It provides an exact measure of surface inclination in relation to flow (Prisley, 2011).

3.3.4. Slope aspect

Slope aspect (*Asp*) (**Figures 3-9a**, **3-10a**, and **3-11a**) represents the direction of the maximum slope. It can influence landslide initiation as it controls moisture retention and vegetation, which in turn may affect soil strength and susceptibility to landslides (Wieczorek et al., 1997). In many mountain ranges, *Asp* reflects the disproportionate distribution of precipitation due to the pronounced directional influence of prevailing winds creating distinct windward and leeward sides. A variation in the degree of weathering and basal erosion due to slope aspect was also reported by Ayalew and Yamagishi (2002) and Rech et al. (2001).

3.3.5. Profile curvature

Profile curvature (*Pfc*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is the surface curvature in the downslope direction (aspect) along a line formed by the intersection of an imaginary vertical plane with the ground surface. It directly controls the velocity of water flow, and therefore erosion (Duman et al., 2006).

3.3.6. Plan curvature

Plan curvature (*Plc*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is surface curvature perpendicular to slope direction or the curvature of the contours on a topographic

map (Ohlmacher, 2007). The influence of plan curvature on land degradation processes is the convergence or divergence of eroded material and water during downhill flow (Nefeslioglu et al., 2008b; Ohlmacher, 2007). In addition, this parameter constitutes one of the main factors controlling the geometry of the terrain surface where landslides occur (Evans et al., 1998).

3.3.7. Total curvature

Total curvature (*Cr*) (**Figures 3-9a**, **3-10a**, and **3-11a**) reflects both the plan and profile curvatures and represents the overall surface curvature. The *Cr* value intrinsically reflects two characteristics of surface morphology: magnitude of slope gradient and curvature of three-dimensional surface forms.

All curvature measures (*Pfc*, *Plc*, and *Cr*) take negative, zero, and positive values. The positive curvature value indicates a predominantly convex slope, and the convexity increases when the value becomes greater. On the other hand, the negative values indicate a predominantly concave slope. Where surface relief is minimal or in the straight sloping area, curvature value will tend to approach zero (Park et al., 2001). For most slopes, the convex part of the landscape may have been experiencing continuous denudational processes by surface erosion, soil creep, and other geomorphological and biological processes (Arnett and Conacher, 2007). In contrast, the concave parts of the slope receive materials from upslope areas and, consequently, surface aggradation occurs.

3.3.8. Drainage density

Ridges and channels are fundamental features of terrain morphology, and therefore are used in various terrain analyses (Band, 1986). In this study, drainage density (*Dd*) (**Figures 3-9a**, **3-10a**, and **3-11a**), the total length per unit area, was computed

using a circular moving window of radius equaling the length of 10 cells. This unit area changing with DEM resolution allowed the mapping of drainage texture ranging from purely local to the regional average drainage density (Tucker et al., 2001).

3.3.9. Distance to drainage network

The under-cutting action of a river and terrain modifications caused by gully erosion may trigger instability of slopes (Lee and Talib, 2005; Xu et al., 2012). Distance to drainages is therefore considered a controlling factor in LS. The shortest Euclidean distance to a drainage line (*Dtd*) (**Figures 3-9a**, **3-10a**, and **3-11a**) was estimated as a parameter in this study. Cells with flow accumulation higher than a threshold value were identified as drainage networks.

3.3.10. Distance to a ridge

In areas of high seismicity, the shortest distance to a ridge line (*Dtr*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is a significant LS parameter reflecting the amplified motion observed at mountain tops (Chang et al., 2007). *Dtr* was calculated from the DEM as Euclidean distance form ridges defined as lines of cells with zero flow accumulation.

3.3.11. Internal relief

Relative relief or internal relief (*Ir*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is the maximum elevation difference in a unit area (Castellanos Abella and Van Westen, 2008). *Ir* values were calculated locally for every cell using a moving window of 10×10 cells such that the unit of measurement represents the features relative to

the scale of analysis, e.g., fine-scale features with the 10 m DEM and coarser-scale form of hillslope with the coarser DEMs (Staley and Waskelwicz, 2013).

3.3.12. Elevation-relief ratio

Elevation-relief ratio (*Er*) (**Figures 3-9a**, **3-10a**, and **3-11a**) describes the area distribution at different elevations. Introduced by Wood and Snell (1960), *Er* is a substitute for the hypsometric integral designed to abstract the salient geometric characteristics of the topography at any scale (Pike and Wilson, 1971; Strahler, 1952). It reflects the stage of geomorphic development and lithological differences (Pérez-Peña et al., 2009; Schumm, 1956). It is defined as:

$$Er = (\text{mean elevation} - \text{min elevation}) / (\text{max elevation} - \text{min elevation})$$

(3-1)

Er values were also calculated locally for every cell using a moving window of 10 $\times$ 10 cells

3.3.13. Sediment transport capacity index

The sediment transport capacity index (*STCI*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is equivalent to the length–slope factor of the Revised Universal Soil Loss Equation (Chen and Yu, 2011). Therefore, it accounts for the effects of topography on both sediment transport and erosion (Moore et al., 1991). *STCI* is calculated as:

$$STCI = (m + 1) \, (A / 22.13)^m \, (\sin\beta / 0.0896)^n \qquad (3-2)$$

where *A* is the upslope contributing area (m^2), β is the local slope gradient (degrees), and *m* and *n* are constants. Because the sensitivity of erosion predictions is not strongly affected by the values of the constants (Kitahara et al., 2000), the values *m*

= 0.4 and n = 1.3 suggested by Chen and Yu (2011) for Taiwan, an area with landslide activity comparable to the study areas in this research, were used.

3.3.14. Stream power index

The stream power index (*SPI*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is a measure of the erosive power of water flow based on the assumption that discharge is proportional to the specific catchment area. Higher *SPI* values lead to an increased risk of slope erosion (Moore et al., 1991). It is defined as:

$$SPI = \ln (As \times \tan\beta) \tag{3-3}$$

where *As* is the specific catchment area (upslope contributing area per unit contour length).

3.3.15. Terrain characterization index

The terrain characterization index (*TCI*) (**Figures 3-9a**, **3-10a**, and **3-11a**) is related to the spatial variability of soil depth and sediment transportation capacity (Catani et al., 2010; Park et al., 2001) which is defined as:

$$TCI = Cr \times \ln As \tag{3-4}$$

TCI is expected to characterize the spatial distribution of soil properties at the continuous functions of terrain attributes (Park et al., 2001). The higher positive or negative *TCI* values may reflect the higher aggradation or degradation potential of soil materials, while smaller values more likely indicate either stability of material movement over the landscape.

3.3.16. Topographic wetness index

The topographic wetness index (*TWI*) (**Figures 3-9a**, **3-10a**, and **3-11a**) has been used to describe soil moisture distribution and found useful for LS studies (Beven, 1997; Moore et al., 1993); higher *TWI* values often occur in landslide bodies (Chen and Yu, 2011). Water infiltration to slope material increases pore water pressure and decreases the soil strength. High *TWI* values indicate zones that saturate first, and the slopes that are more likely to fail. *TWI* is defined as:

$$TWI = \ln (As \,/\, \tan\beta) \tag{3-5}$$

3.3.17. Random integer

In addition to the 16 topographic parameters, the random integer *rand* was also used as a model parameter to assess the performance of other parameters according to the parameter ranking provided by the RF model (Catani et al., 2013).

Although the DEM-derived parameters represent distinct terrain properties and processes (**Table 3-3**), their interrelationship (**Figure 3-12**) may lead to multicollinearity. However, for LSM, Harrell (2001) suggests that multicollinearity does not influence the predictions from training and testing datasets if both have the same type of collinearities. This applies to this study because all parameters used with the training and testing datasets are mathematical derivatives of the same 10 m DEM.
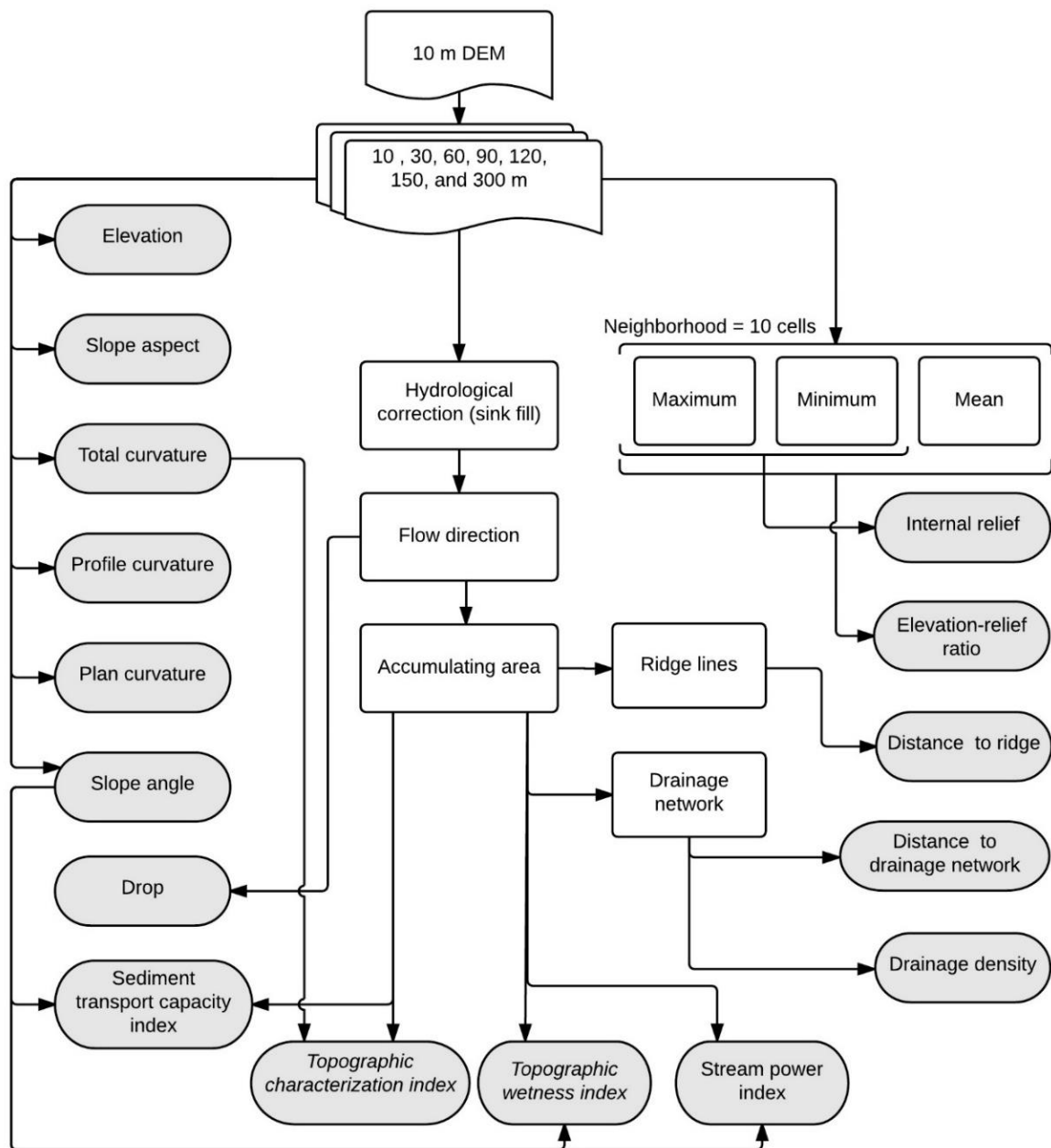
Figure 3-12. Interrelationship of the topographic parameters.

Chapter 4.   Methodology

RF was used to construct LS models using the topographic parameters and the landslide inventories. It was implemented in JMP Pro 11.2 (SAS Institute Inc. Cary, NC), and GIS-based calculations were performed using ArcGIS 10.3 (ESRI, Redlands, CA) and Python 2.7.

4.1.   Random Forest

4.1.1.   Concept

RF is an ensemble learning method of classification using regression trees which combines the idea of bagging with random feature selection (Breiman, 1996, 2001; Cutler et al., 2012; Ho, 1998). RF utilizes bootstrap and random techniques to select the subsample of data and predictor parameters while growing an ensemble of trees (hence called "forest"). In addition to constructing each tree using a different bootstrap sample of the data, RF changes how the classification or regression trees are constructed. In classical decision trees, each node is split using the best split among all parameters. In RF, by contrast, each node is split using the best among a subset of predictors randomly chosen at that node. This strategy leads to higher performance than many other classifiers such as discriminant analysis, SVM, and ANN; it also makes RF robust against overfitting (Breiman, 2001; Liaw and Wiener, 2002). RF has several other advantages: 1) it does not require assumptions on the distribution of explanatory parameters; 2) it allows for the mixed use of categorical and numerical parameters without using dummy parameters; and 3) it can account for interactions and nonlinearities between parameters (Catani et al., 2013).

Archer and Kimes (2008) simplify the RF algorithm as follows:

Each tree is grown using a bootstrapped sample from the original learning sample L. However, at each node of the tree, m of the p independent variables are randomly selected, the best split on these m is used to split the node. This random selection of features at each node decreases the correlation between the trees in the forest thus decreasing the forest error rate.

For $b = 1,\ldots, B$, sample n observations with replacement from L. This is referred to as the bootstrap sample, L_b. Use L_b to build a classification tree, with the following modifications:

1. At node t, randomly sample m of the p independent variables.

2. For each of the $k = 1,\ldots, m$ sampled variables, find the best split (s_k) among all possible splits for the k-th variable.

3. Choose the best split s^* from among the $k = 1,\ldots, m$ best splits s_k on which to split the node t. This j-th variable at its identified cut point c_{s^*} is used to split node t.

4. Split the data at this node by sending the $i = 1,\ldots, n$ observations with $x_{ij} < c_{s^*}$ to the left descendant node and all observations with $x_{ij} \geqslant c_{s^*}$ to the right descendant.

5. Repeat steps $1-4$ on all descendant nodes to grow a maximally sized tree, T_b.

The Gini criterion, how often a randomly chosen element from the set would be incorrectly labeled if it were randomly labeled according to the distribution of labels in the subset, is used to select the split with the lowest impurity at each node. For each tree in the forest, the predicted class for each

observation is obtained. The class with the maximum number of votes among the *B* trees in the forest is the predicted class of an observation.

RF produces multiple outputs to aid the interpretation of results, including out-of-bag (OOB) accuracy estimates and parameter importance measures (Bricher et al., 2013). OOB errors from RF classifications provide an alternative to cross-validation. For each tree in the forest, a random third of all observations are held out from the training set, and are referred to as OOB. The OOB error is, thus, the proportion of misclassified observations. The other crucial output is the measure of parameter importance, i.e., the statistical weight of each predictor variable. This study employs this measure to analyze the influence of scale on landslide causative parameters. OOB accuracy estimates provide the predictive efficacy of RF models. The change in generalized *R*-square (R^2), a measure of variance in the dependent variable explained by the independent variables, was analyzed to identify the required number of trees (*T#*). *T#* at which R^2 stops increasing and starts oscillating around a stabilized value indicates the optimal complexity required for the model.

4.1.2. Implementation

RF can be implemented in a variety of analytical environments. It is available in R (http://www.R-project.org) as 'randomForest', in Python (http://www.python.org) as 'RandomForestClassifier', in Matlab (Mathworks, U.S.A.) as 'Treebagger', and in JMP (SAS Institute Inc., Cary, NC) as 'Bootstrap Forest'. This study utilized JMP Pro 11.2 for the implementation of RF. The JMP-*statistical discovery* provides an intuitive interface to the RF modeling and several other statistical requirements.

Training and testing datasets

Classification data used in an RF model for LSM should contain information about both landslides and no-landslide areas. No-landslide points, whose number is equal to the number of landslides, were randomly created in no-landslide areas. Values of parameters for the landslide and no-landslide points were extracted. The data (50% landslide and 50% no-landslide) for Niigata and Ehime consist of 21,324 and 5,086 entries, respectively. The data were randomly divided into training (50%) and testing (50%) datasets.

Model evaluation

In statistical classification models, a receiver operating characteristic (ROC), or an ROC curve is used to evaluate their effectiveness and overall fit (Gorsevski et al., 2006). The curve is created by plotting the true positive rate (sensitivity) against the false positive rate (1-specificity). The area under the ROC curve (*AUC*) characterizes the quality of a prediction model (Yesilnacar and Topal, 2005). *AUC* varies from 0.5 (diagonal line) to 1, with higher values indicating a better predictive capability of the model. *AUC* values less than 0.7 correspond to poor predictive ability, between 0.7 and 0.8 to moderate, between 0.8 and 0.9 to good and >0.90 to excellent (Swets, 1988; Trigila et al., 2015) (**Figure 4-1**). RF models in this study were evaluated using their predictive accuracy and *AUC*.
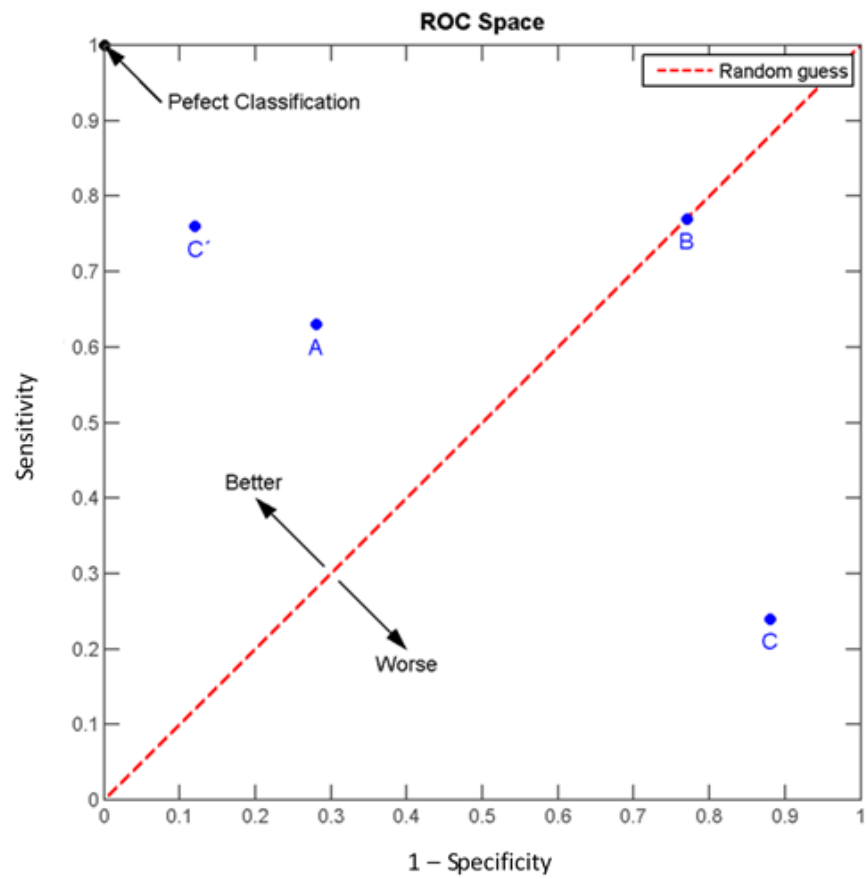
Figure 4-1. The ROC space.
(after Kai Walz, CC by SA 3.0)
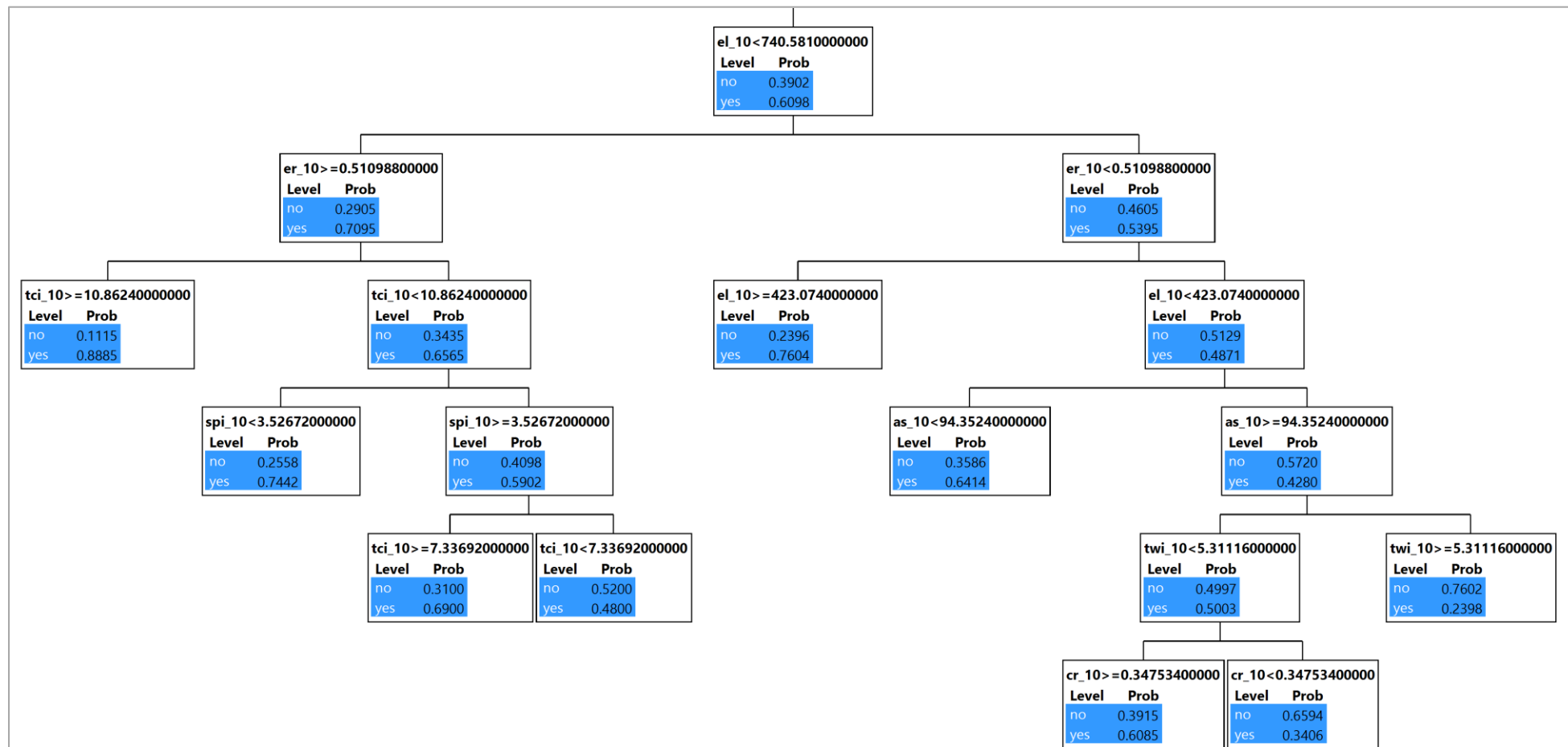
Figure 4-2. Section of a decision tree from a 'Random Forest' showing probabilities of landslide occurrence.

4.2. Influence of geology and lithology

It is expected that the properties of the slope-forming materials, such as strength and permeability that are involved in the failure, are related to lithology, which therefore should affect the likelihood of failure. (Dai and Lee, 2002). To analyze the degree of influence of lithological variations on the classification of landslides, a comparison was performed for LS models with and without the use of lithological information together with the topographic parameters discussed in previous section (**Section 3.3**). The geological parameters used are geological types, distance to geological boundary, and density of geological boundaries.

4.2.1. Geological types

Underlying geology governs the relative stability of hillslopes. Geological and structural settings of slopes are therefore studied as main predisposing factors controlling the development of mass movements (Grelle et al., 2011; Varnes, 1978). The geological substrata from the study areas were utilized to examine the influence of geology (*geo_code*; codes of geological substrata in **Table 3-2**).

4.2.2. Density of geological boundaries

The density of geological boundaries (*geo_den*; km/km^{-2}) was also calculated from the geological boundaries using the Spatial Analyst (**Figure 4-3**), setting the radius parameter at 1000 m which corresponds to the distance to geological boundary that includes most of the landslides (**Figure 4-4**) (Kawabata and Bandibas, 2009). Areas with higher density of geological boundaries are expected to be more susceptible to landslides.
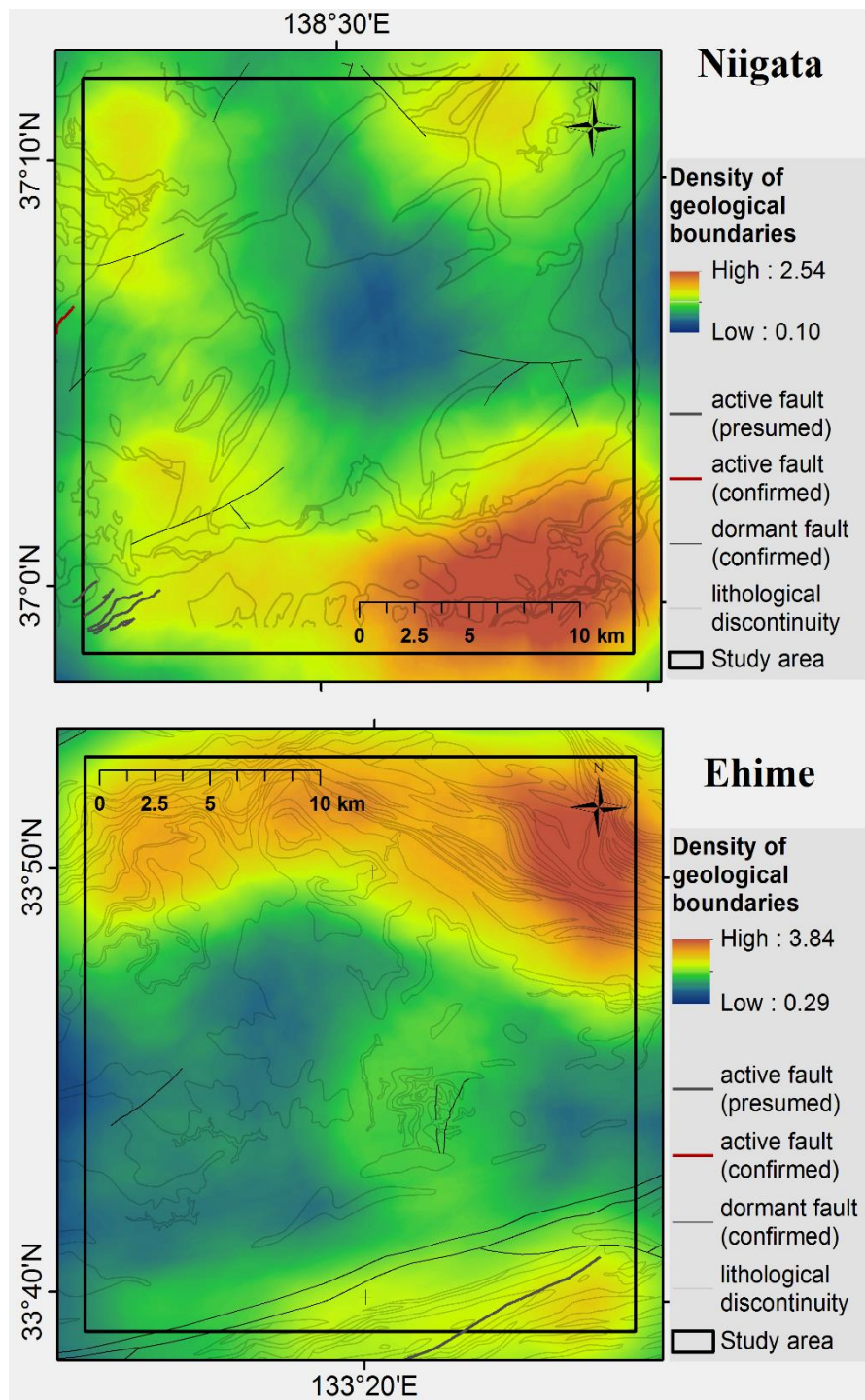
Figure 4-3. Map showing the density of geological boundaries for Niigata (top) and Ehime (bottom).
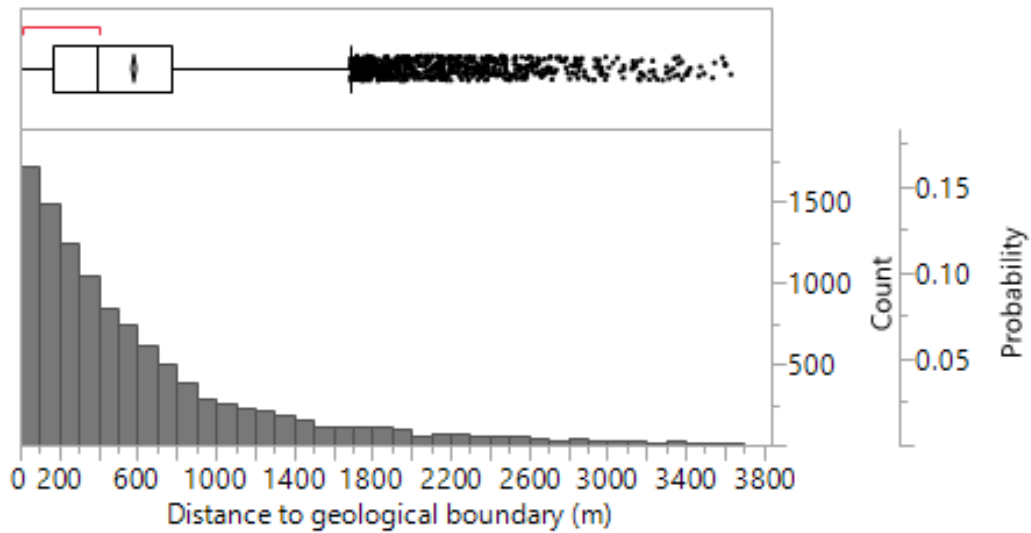
Figure 4-4. Landslide occurrence and distance to geological boundary.

4.2.3. Distance to geological boundary

Lithological boundaries as well as known fault lines were considered in this study as geological boundaries. These areas indicate lithological discontinuities and are known to be areas of weakness prone to slope failures (Kawabata and Bandibas, 2009). The shortest distance to the geologic boundary (*geo_dis*; m) was calculated from polyline data of geological boundaries using the Spatial Analyst extension of ArcGIS (**Figure 4-5**).
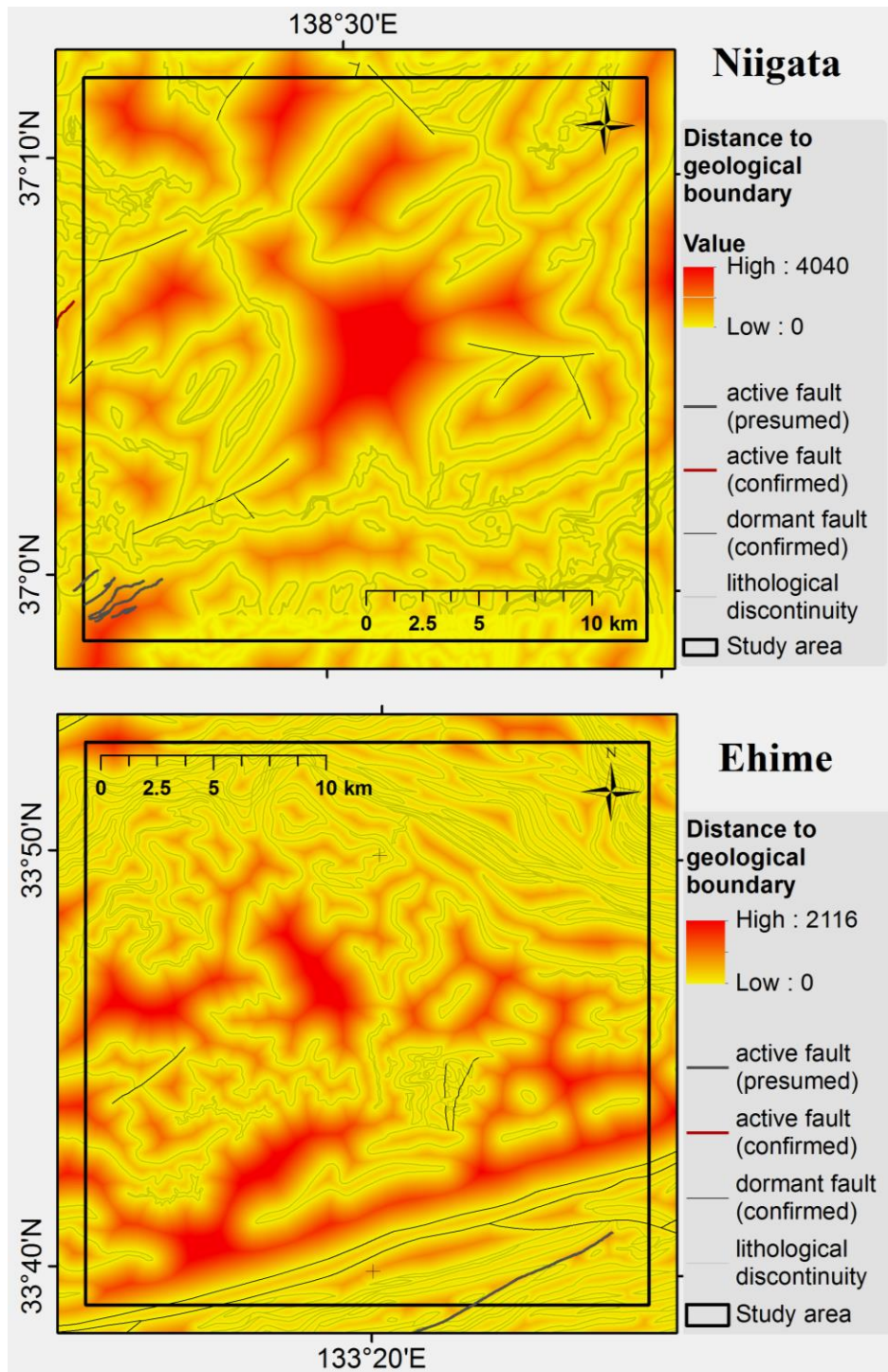
Figure 4-5. Map showing the distance to the geological boundary for Niigata (top) and Ehime (bottom).

4.3. Multi-resolution landslide susceptibility analysis

The seven DEM-scales (10 to 300 m) were applied to construct RF models classifying landslide presence and absence to identify the optimal resolution for each parameter. **Figure 4-6a** outlines the process. First, the values of the 16 parameters are computed for all the scales, and the values are used in an RF model to classify the landslide data. The scale of each parameter with the highest importance in the classification, determined as an average of 10 iterations, is considered optimal. This process is repeated for all the parameters, and finally, a combination of all parameters at their optimal scales are used to create a multi-resolution LS model and an LS map. The finest grid size among the parameters at their optimal scales is selected as the mapping unit. **Figure 4-6b** outlines the determination of parameter importance in a multi-resolution LS model. In the hypothetical example shown in **Figure 4-6**, *Sl* at 30 m contributes most to the classification, and therefore is the most important parameter for the LS study.
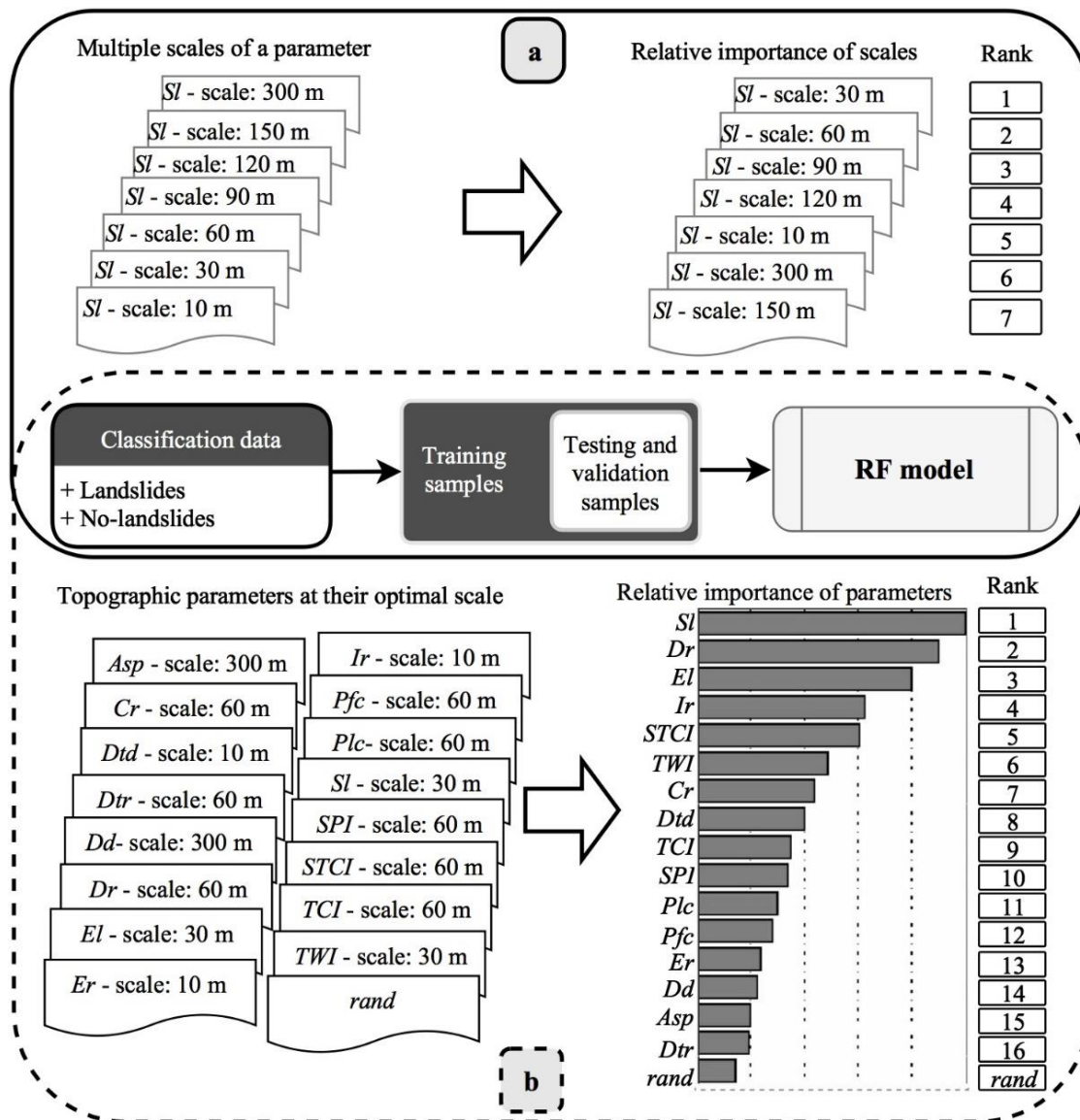
Figure 4-6. Outline of multi-resolution technique for: (a) selection of optimal parameter-scale and (b) determination of parameter importance (Paudel et al., 2016).

4.4. Topographic representation of landslides

4.4.1. Applied methods

Topographic representation of landslides in a grid-based analysis deals with which cells of landslides mapped as polygons are used subsequently in LS modeling. The methods of topographic representation were applicable only to landslides mapped as polygons and therefore limited to the study areas in Japan. Topographic representation techniques based on single cell approach suggested by Van Den Eeckhaut et al. (2006) was analyzed for five different strategies: center-cell, cells within landslide boundary, seed cells, cells within the depletion zone, and dominant within the depletion zone (DCD) (**Section 1.3.3**, **Figure 4-7**). This section improves the methodology proposed in **Section 4.3** and therefore deals with topographic parameters at their optimal scales.
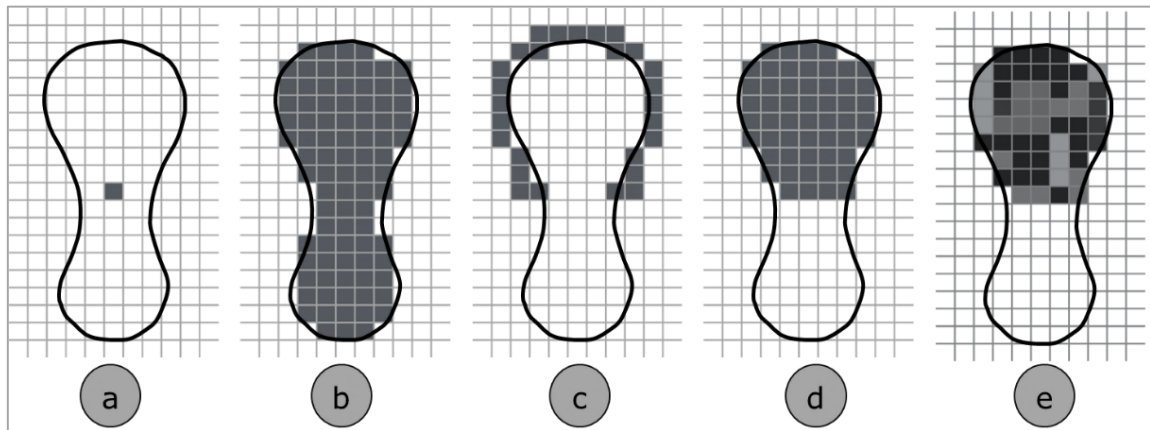
Figure 4-7. Illustration of the methods analyzed for the topographic representation of landslides. (a) Center-cell, (b) cells within landslide boundary, (c) seed cells, (d) cells within the depletion zone, and (e) dominant within the depletion zone.

The center-cell (**Figure 4-7a**) represents the cell underneath the centroid of the landslide polygon. For cells within the landslide boundary (**Figure 4-7b**), a representative cell was randomly selected for each landslide from cells within its boundary. For seed cells, the commonly used size of the buffer zone is 100 to 150 m (Süzen and Doyuran, 2004), but the buffer zone size may be different depending on experts' judgments. In this study, a 50 m buffer zone (**Figure 4-7c**) was selected as seed cells around the depletion zone, following the strategy of Wang et al. (2013). The representative cell was randomly selected from the seed cells of each landslide. Similarly, a cell was randomly selected from the depletion zone (**Figure 4-7d**) for the evaluation of method concerning the cells within the depletion zone. **Figure 4-7e** represents the "dominant within the depletion zone" (DCD) technique of representation newly proposed in this study. It is an assimilation of existing ideas of representative area-based approach (Simon et al., 2013) and established importance of depletion zone for the representation of landslide initiation conditions (Clerici et al., 2006; Trigila et al., 2015; Van Den Eeckhaut et al., 2006).

To identify the representative cell using the DCD, the dominant cell value within the depletion zone was estimated using the Zonal Statistics function of the Spatial Analyst extension of ArcGIS 10.3. The process requires categorical values. The continuous data sets were therefore classified into categories based on percentile divisions of cells within the depletion zone following the data driven strategy proposed by Süzen and Doyuran (2004).

4.4.2. Separation of depletion and deposition zones

A landslide consists two morphologically distinct zones: 1) the depletion zone (detachment zone or rupture zone), i.e. the upper part of a landslide where the failure

is effectively generated, and the accumulation zone, i.e. the lower part which is affected by the arrival of the depleted material (Clerici et al., 2006). Demarcation between these two zones is generally difficult and usually most landslide inventory maps show the depletion and accumulation zones of failures together as a single landslide (Dagdelenler et al., 2015). The landslide inventory for Japan also includes landslides as single polygons (**Section 3.1**) without differentiating depletion and deposition zones.

Landslides are elongated features with distinct heads and toes. This study hypothesizes that a transect passing through its center and perpendicular to the mean (angular) aspect (Guarneri, 2013) of the landslide body divides a landslide into two distinct depletion and accumulation zones (**Figure 4-8**). This was accomplished in a GIS toolset originally developed in Python 2.7. **Figure 4-9** represents the schematics of the GIS tool.
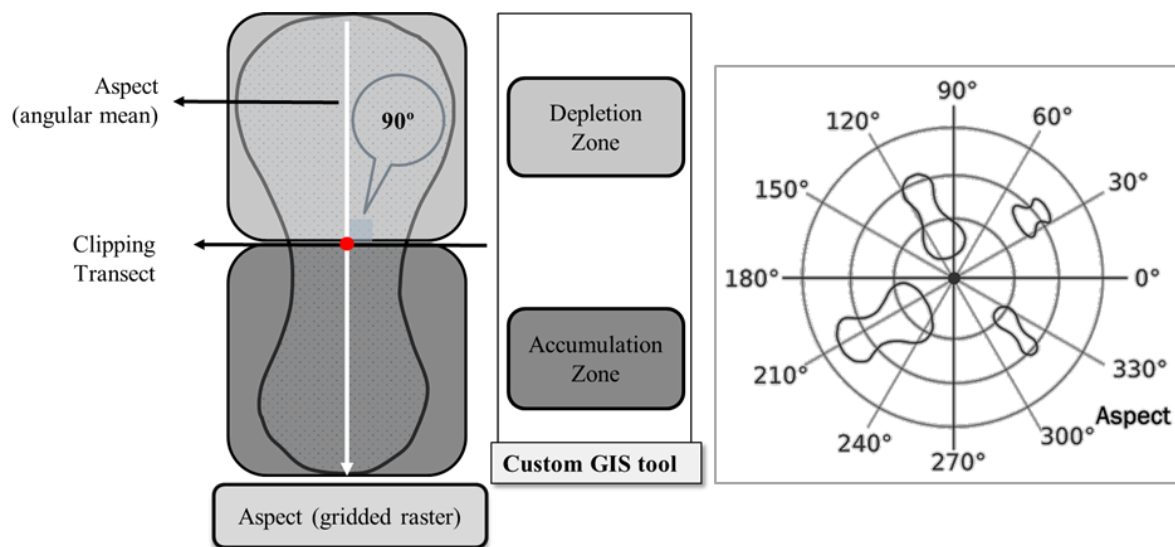
Figure 4-8. Outline for the separation of depletion and accumulation zones.
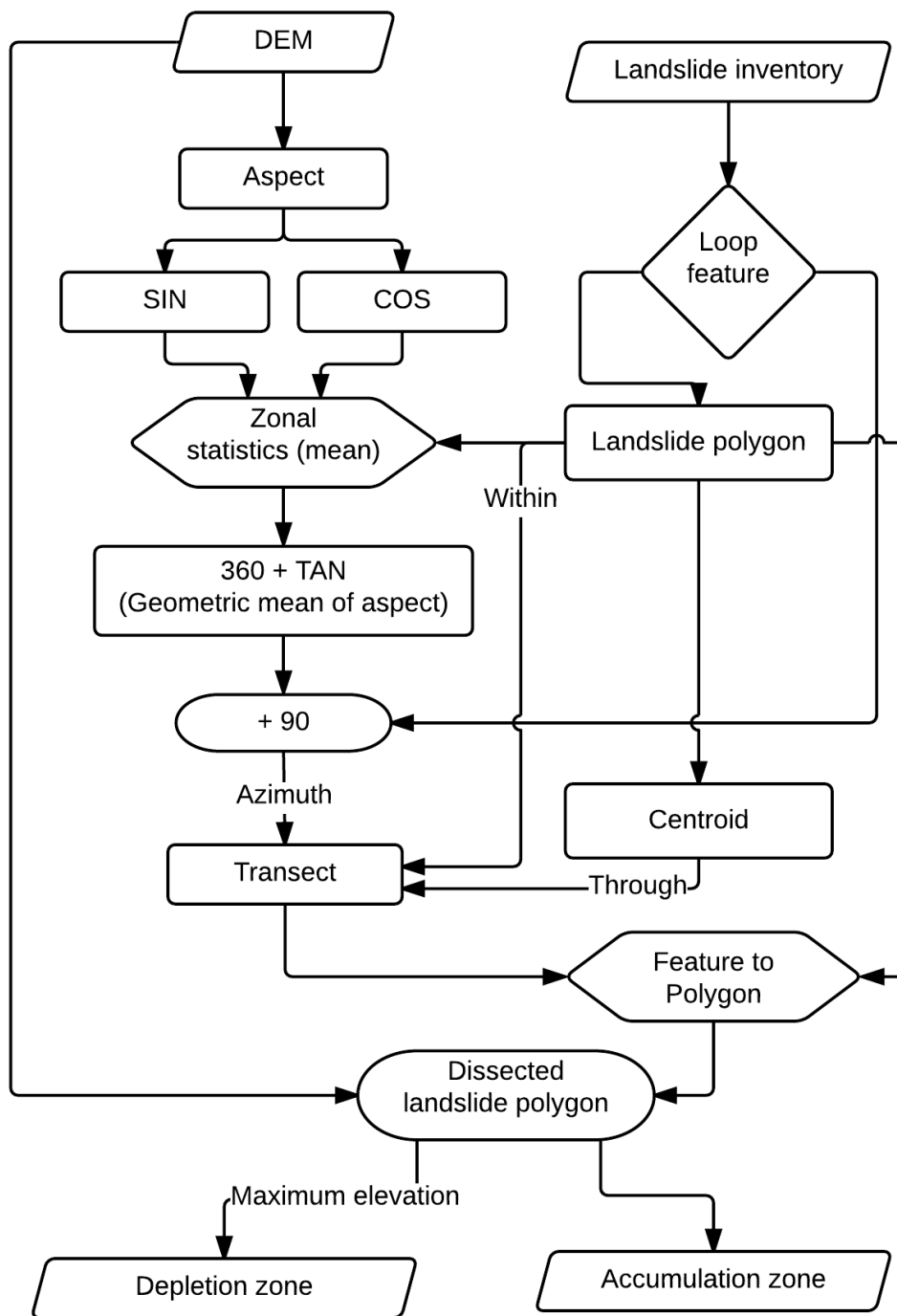
Figure 4-9. Flowchart for the separation of depletion and accumulation zones.

Chapter 5. Results

To identify the minimum $T\#$ required for a stable RF, $T\#$ was gradually increased to 5000. **Figure 5-1** shows that R^2 plateaued with $T\#$ greater than 100. This study however uses $T\# = 500$ to accommodate unseen inconsistencies, following the strategy of Díaz-Uriarte and Alvarez de Andrés (2006). All model results show values averaged over 10 iterations.
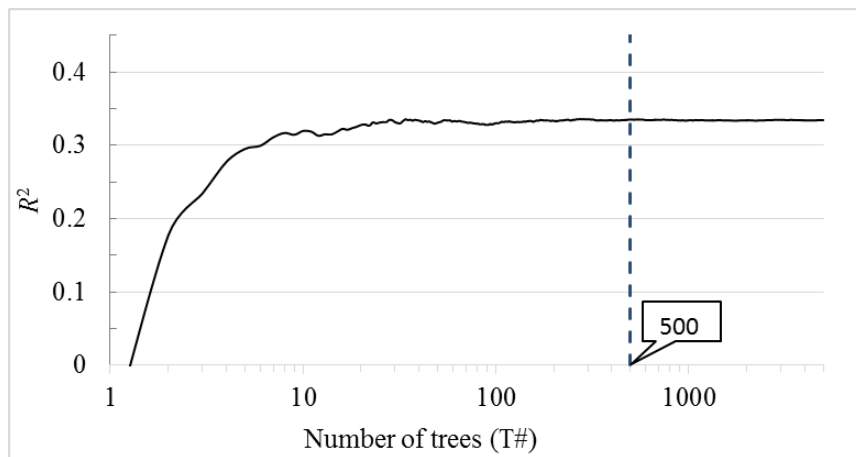
Figure 5-1. Generalized R-squared and the number of trees in an RF model.

At first, the LS analyses were conducted with landslide representation using the commonly employed center-cell method. Then different topographic representation methods were compared. The results presented in **Sections 5.1, 5.2,** and **5.3**; therefore utilize the center-cell method. **Section 5.4** includes a comparison of different topographic representation methods. **Section 5.5** includes results from LS models with different combination of representation methods and parameter scales. Center-cell method was again utilized in **Section 5.6** due to the type of landslide inventory.

5.1. Influence of geology and lithology

In comparison between the RF models with and without the use of geological information, it was observed that the inclusion of geological information resulted in very little improvement in the testing (prediction) accuracy of the LS models (**Figure 5-2**). Moreover, the influence of geology in both the study areas (Niigata and Ehime) was not uniform. In Niigata, addition of geological information was followed by an overall improvement of testing accuracy (about 2 %) of the LS model while in Ehime the same resulted in very little improvement in the predictive performance. This was well correlated with the rank of parameters for the two areas (**Table 5-1**). While two of the geological parameters (*geo_code* and *geo_den*) in Niigata contributed significantly to the classification, none for Ehime contributed significantly. In fact, *geo_code* ranked below *rand* in Ehime.
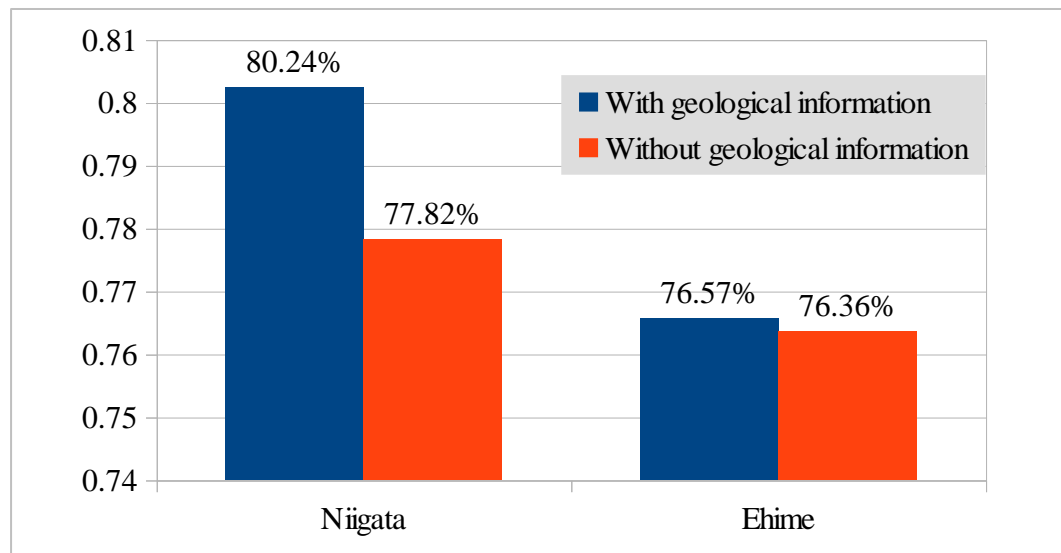
Figure 5-2. Influence of geological information on the accuracy of LS models.

Table 5-1. Importance of geological parameters in Niigata and Ehime.

Parameters	Rank of parameters	
	Niigata	Ehime
Geological types (*geo_code*)	3/20	20/20
Distance from geological boundary (*geo_dis*)	16/20	17/20
Density of geological boundary (*geo_den*)	8/20	16/20

5.2. Scale-sensitivity of topographical parameters

All 16 parameters used in this study were individually examined for scale-sensitivity. RF models for scale-sensitivity were trained at 50% and tested with the remaining 50% of the dataset. **Table 5-2** and **Table 5-3** show portion of classification contributed by each scale of parameters in Niigata and Ehime respectively. The scale corresponding to the highest portion was selected optimal. The results show that optimal scales differ among parameters.

For Niigata, the finest resolution (10 m) is optimal for parameters *Dtd*, *Er*, *Ir,* and *Sl*, while most of the other parameters exhibit optima at higher resolutions, preferably at 30, 60, and 300 m. Two parameters, *Asp* and *Dd,* contribute the most at 300 m, the coarsest resolution. The results from Ehime tend to be similar. For Ehime, parameters *Asp*, *Dtd*, *Dtr*, *El*, *Er*, *Ir*, and *Sl* are optimal at the finest scale, whereas the other parameters exhibit optima at coarser scales. Similar to Niigata, *Dd* in Ehime was found to be optimum at the coarsest scale. **Figure 5-3** compares the optimal parameter-scale between the two study areas. The optimal scales for most parameters show similarity in both areas except *Asp,* for which the coarsest resolution in Niigata and the finest resolution in Ehime are optimal.

Table 5-4 and Table 5-5 include correlation between parameter scales. The results suggest that for most parameters there is a decline in correlation away from the scale. *El* was however found to be almost perfectly correlated in both the study areas across all the scales. After the evaluation of optimal scale for each parameter, the correlation between them was assessed. Results in Table 5-6 and Table 5-7 show that some parameter pairs; *Ir* and *Dr*, *Plc* and *Cr*, *Dtr* and *SPI*, and *Ir* and *Sl* in Niigata and *Ir* and *Sl* in Ehime; show high correlation between them. Simple statistical summary of parameters at their finest, coarsest, and optimal scales is included in Table 5-8 and Table 5-9.
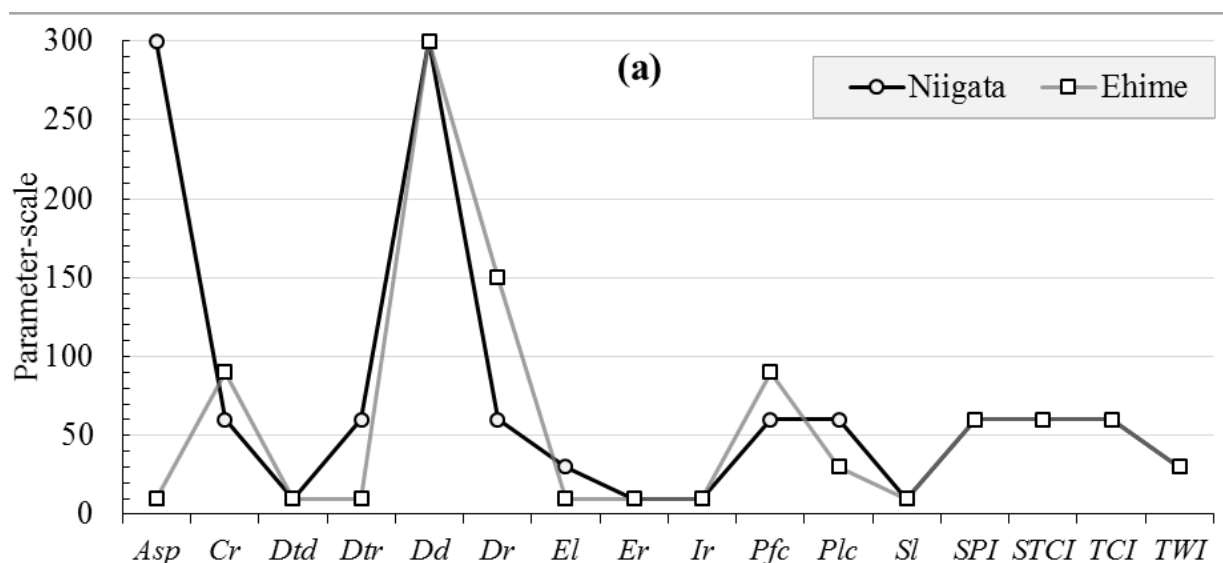
Figure 5-3. Optimal scales of parameters for Niigata and Ehime.

Table 5-2. Scale-sensitivity of parameters in Niigata. Values indicate the portion contributed by each parameter-scale in the classification. Highest contributor is selected as the optimal scale for the parameter.

Parameters	Scales							
	Optimal	10	30	60	90	120	150	300
Asp	300	0.113	0.108	0.112	0.131	0.152	0.146	**0.238**
Cr	60	0.085	0.171	**0.28**	0.192	0.121	0.099	0.052
Dtd	10	**0.347**	0.226	0.126	0.093	0.075	0.072	0.061
Dtr	60	0.151	0.256	**0.265**	0.139	0.074	0.048	0.067
Dd	300	0.168	0.166	0.128	0.113	0.125	0.115	**0.185**
Dr	60	0.153	0.178	**0.217**	0.164	0.123	0.102	0.061
El	30	0.176	**0.189**	0.152	0.124	0.111	0.107	0.141
Er	10	**0.198**	0.17	0.116	0.144	0.114	0.117	0.14
Ir	10	**0.356**	0.211	0.112	0.071	0.057	0.08	0.112
Pfc	60	0.084	0.152	**0.286**	0.206	0.126	0.094	0.053
Plc	60	0.129	0.198	**0.237**	0.156	0.117	0.09	0.074
Sl	10	**0.232**	0.214	0.21	0.123	0.087	0.062	0.072
SPI	60	0.101	0.214	**0.259**	0.153	0.104	0.081	0.088
STCI	60	0.089	0.232	**0.258**	0.158	0.113	0.077	0.073
TCI	60	0.045	0.139	**0.282**	0.187	0.137	0.108	0.102
TWI	30	0.12	**0.224**	0.221	0.155	0.105	0.079	0.095

Table 5-3. Scale-sensitivity of parameters in Ehime. Values indicate the portion contributed by each parameter-scale in the classification. Highest contributor is selected as the optimal scale for the parameter.

Parameters	Scales							
	Optimal	10	30	60	90	120	150	300
Asp	10	**0.192**	0.172	0.153	0.128	0.124	0.11	0.121
Cr	90	0.141	0.153	0.162	**0.169**	0.132	0.141	0.103
Dtd	10	**0.338**	0.209	0.125	0.116	0.08	0.079	0.052
Dtr	10	**0.205**	0.191	0.157	0.122	0.088	0.147	0.089
Dd	300	0.14	0.141	0.136	0.15	0.128	0.151	**0.154**
Dr	150	0.127	0.139	0.157	0.164	0.138	**0.18**	0.095
El	10	**0.156**	0.155	0.138	0.125	0.138	0.134	0.155
Er	10	**0.161**	0.147	0.134	0.139	0.129	0.138	0.152
Ir	10	**0.264**	0.165	0.119	0.112	0.107	0.106	0.127
Pfc	90	0.128	0.129	0.15	**0.176**	0.138	0.159	0.12
Plc	30	0.168	**0.176**	0.165	0.163	0.114	0.113	0.101
Sl	10	**0.21**	0.175	0.162	0.137	0.116	0.105	0.094
SPI	60	0.142	0.141	**0.181**	0.14	0.113	0.161	0.122
STCI	60	0.11	0.131	**0.194**	0.158	0.124	0.164	0.119
TCI	60	0.122	0.153	**0.198**	0.155	0.117	0.143	0.111
TWI	30	0.16	**0.169**	0.169	0.144	0.113	0.141	0.105

Table 5-4. Correlation between parameter scales (Niigata).

Parameters	Optimal	Scales						
		10	30	60	90	120	150	300
Asp	300	0.28	0.31	0.36	0.41	0.47	0.54	1.00
Cr	60	0.37	0.68	1.00	0.67	0.61	0.45	0.21
Dtd	10	1.00	0.98	0.93	0.88	0.83	0.77	0.56
Dtr	60	0.29	0.53	1.00	0.55	0.46	0.34	0.16
Dd	300	0.70	0.70	0.70	0.70	0.70	0.73	1.00
Dr	60	0.65	0.84	1.00	0.85	0.80	0.70	0.51
El	30	1.00	1.00	1.00	1.00	1.00	1.00	0.99
Er	10	1.00	0.52	0.33	0.26	0.22	0.19	0.12
Ir	10	1.00	0.81	0.63	0.52	0.44	0.38	0.20
Pfc	60	0.29	0.64	1.00	0.66	0.60	0.47	0.27
Plc	60	0.35	0.61	1.00	0.60	0.50	0.34	0.09
Sl	10	1.00	0.89	0.72	0.60	0.52	0.46	0.29
SPI	60	0.34	0.54	1.00	0.54	0.48	0.36	0.18
STCI	60	0.28	0.43	1.00	0.43	0.43	0.35	0.20
TCI	60	0.14	0.38	1.00	0.42	0.30	0.22	0.13
TWI	30	0.55	1.00	0.54	0.39	0.30	0.24	0.12

Table 5-5. Correlation between parameter scales (Ehime).

Parameters	Optimal	Scales						
		10	30	60	90	120	150	300
Asp	10	1.00	0.83	0.69	0.62	0.54	0.49	0.33
Cr	90	0.26	0.49	0.67	1.00	0.63	0.56	0.27
Dtd	10	1.00	0.98	0.94	0.89	0.83	0.79	0.59
Dtr	10	1.00	0.63	0.35	0.20	0.12	0.09	0.04
Dd	300	0.71	0.71	0.72	0.71	0.71	0.75	1.00
Dr	150	0.30	0.43	0.62	0.72	0.79	1.00	0.66
El	10	1.00	1.00	1.00	1.00	1.00	1.00	0.99
Er	10	1.00	0.48	0.27	0.20	0.18	0.15	0.11
Ir	10	1.00	0.72	0.50	0.41	0.35	0.32	0.24
Pfc	90	0.19	0.44	0.67	1.00	0.64	0.57	0.35
Plc	30	0.69	1.00	0.65	0.41	0.25	0.16	0.02
Sl	10	1.00	0.84	0.63	0.48	0.38	0.33	0.21
SPI	60	0.37	0.54	1.00	0.52	0.44	0.34	0.17
STCI	60	0.28	0.38	1.00	0.49	0.43	0.34	0.18
TCI	60	0.24	0.46	1.00	0.37	0.21	0.10	0.02
TWI	30	0.54	1.00	0.54	0.36	0.26	0.21	0.10

Table 5-6. Correlation between the parameters at optimal scale (Niigata).

Parameters	Asp_300	Cr_60	Dtd_10	Dtr_60	Dd_300	Dr_60	El_30	Er_10	Ir_10	Pfc_60	Plc_60	Sl_10	SPI_60	STCI_60	TCI_60	TWI_30
Asp_300	1.00															
Cr_60	0.00	1.00														
Dtd_10	0.00	0.23	1.00													
Dtr_60	-0.02	-0.56	-0.24	1.00												
Dd_300	0.00	-0.08	-0.53	0.12	1.00											
Dr_60	0.01	0.51	0.31	-0.32	-0.25	1.00										
El_30	-0.04	0.12	0.17	-0.11	-0.17	0.28	1.00									
Er_10	0.00	0.46	0.30	-0.32	-0.21	0.44	0.22	1.00								
Ir_10	0.02	0.03	0.13	-0.08	-0.18	**0.77**	0.21	0.12	1.00							
Pfc_60	0.00	-0.85	-0.31	0.47	0.17	-0.55	-0.16	-0.57	-0.03	1.00						
Plc_60	0.00	**0.87**	0.09	-0.50	0.02	0.34	0.05	0.23	0.03	-0.48	1.00					
Sl_10	0.01	0.03	0.08	-0.08	-0.14	0.64	0.16	0.11	**0.82**	-0.02	0.02	1.00				
SPI_60	-0.01	-0.68	-0.26	**0.86**	0.13	-0.30	-0.10	-0.37	0.00	0.55	-0.61	0.00	1.00			
STCI_60	0.00	-0.50	-0.23	0.47	0.11	-0.10	-0.01	-0.26	0.18	0.43	-0.43	0.14	0.66	1.00		
TCI_60	-0.02	0.01	-0.11	0.60	0.11	-0.08	-0.07	-0.08	-0.19	-0.03	-0.01	-0.17	0.59	0.16	1.00	
TWI_30	-0.01	-0.54	-0.18	0.48	0.09	-0.33	-0.11	-0.32	-0.13	0.42	-0.51	-0.13	0.54	0.33	0.24	1.00

Table 5-7. Correlation between the parameters at optimal scale (Ehime).

Parameters	Asp_10	Cr_90	Dtd_10	Dtr_10	Dd_300	Dr_150	El_10	Er_10	Ir_10	Pfc_90	Plc_30	Sl_10	SPI_60	STCI_60	TCI_60	TWI_30
Asp_10	1.00															
Cr_90	0.02	1.00														
Dtd_10	0.02	0.38	1.00													
Dtr_10	0.03	-0.25	-0.04	1.00												
Dd_300	-0.01	-0.16	-0.54	0.05	1.00											
Dr_150	0.04	0.34	0.50	-0.04	-0.43	1.00										
El_10	0.00	0.18	0.38	-0.06	-0.38	0.36	1.00									
Er_10	0.03	0.39	0.28	-0.14	-0.20	0.33	0.18	1.00								
Ir_10	0.06	0.03	0.06	0.04	-0.05	0.40	0.11	0.08	1.00							
Pfc_90	-0.01	-0.82	-0.42	0.20	0.28	-0.45	-0.27	-0.47	-0.04	1.00						
Plc_30	-0.03	0.39	0.07	-0.37	0.00	0.05	0.06	0.10	-0.03	-0.24	1.00					
Sl_10	0.06	0.02	0.02	0.06	-0.02	0.28	0.08	0.07	**0.75**	-0.02	-0.04	1.00				
SPI_60	0.00	-0.55	-0.27	0.34	0.13	-0.17	-0.14	-0.32	0.01	0.44	-0.42	0.02	1.00			
STCI_60	0.01	-0.42	-0.26	0.19	0.15	-0.13	-0.13	-0.29	0.04	0.33	-0.31	0.03	0.68	1.00		
TCI_60	-0.02	0.10	0.05	0.09	-0.01	0.08	0.01	0.10	-0.04	-0.09	0.17	-0.04	0.33	-0.03	1.00	
TWI_30	0.03	-0.42	-0.17	0.47	0.08	-0.12	-0.11	-0.25	-0.02	0.32	-0.63	-0.01	0.54	0.37	0.05	1.00

Table 5-8. Statistical details of parameters for Niigata at 10 m, 300 m, and the optimal scale.

Parameters	Optimal scale	Minimum			Maximum			Mean			Std. Dev		
		10 m	300 m	Optimal	10 m	300 m	Optimal	10 m	300 m	Optimal	10 m	300 m	Optimal
Asp	300	-1.00	0.02	0.02	360.00	359.94	359.94	180.38	186.92	186.92	105.80	111.13	111.13
Cr	60	-39.01	-0.32	-3.48	33.47	0.37	3.06	0.00	0.00	0.00	2.11	0.09	0.60
Dtd	10	0.00	0.00	0.00	600.17	848.53	600.17	154.19	223.64	154.19	108.80	184.40	108.80
Dtr	60	0.00	0.00	0.00	272.03	670.82	305.94	18.96	164.89	45.73	14.54	167.24	40.26
Dd	300	0.00	0.00	0.00	9.23	6.96	6.96	2.09	1.53	1.53	1.66	1.46	1.46
Dr	60	0.03	0.02	0.03	465.00	42.50	166.67	32.47	15.07	27.79	23.41	9.21	17.26
El	30	15.00	15.13	15.00	1283.81	1223.81	1283.26	385.58	386.65	385.69	237.18	236.33	237.19
Er	10	0.01	0.18	0.01	0.96	0.70	0.96	0.49	0.45	0.49	0.11	0.07	0.11
Ir	10	0.00	97.14	0.00	168.05	755.66	168.05	36.61	323.67	36.61	18.58	128.47	18.58
Pfc	60	-26.59	-0.23	-2.50	31.02	0.20	2.40	0.03	0.00	0.02	1.48	0.05	0.34
Plc	60	-26.66	-0.24	-2.31	18.10	0.22	2.49	0.03	0.00	0.02	1.05	0.05	0.36
Sl	10	0.00	0.09	0.00	75.71	22.84	75.71	18.26	7.47	18.26	10.84	3.87	10.84
SPI	60	-3.95	-3.75	-3.92	17.82	16.92	17.84	4.34	4.81	4.78	2.83	6.72	4.99
STCI	60	0.05	0.07	0.05	6746.10	1728.15	4331.62	30.68	87.60	63.11	41.40	127.45	101.25
TCI	60	-355.11	0.08	-35.02	230.62	19.02	38.58	3.16	6.37	4.15	12.41	6.25	4.65
TWI	30	-0.96	0.95	-0.52	19.09	19.08	19.09	5.81	7.79	6.39	2.41	5.80	3.64

Table 5-9. Statistical details of parameters for Ehime at 10 m, 300 m, and the optimal scale.

Parameters	Optimal scale	Minimum			Maximum			Mean			Std. Dev		
		10 m	300 m	Optimal	10 m	300 m	Optimal	10 m	300 m	Optimal	10 m	300 m	Optimal
Asp	10	-1.00	0.03	-1.00	360.00	359.97	360.00	176.90	177.46	176.90	105.70	102.46	105.70
Cr	90	-83.81	-0.51	-2.64	48.02	0.52	2.49	0.00	0.00	0.00	2.72	0.17	0.61
Dtd	10	0.00	0.00	0.00	728.35	848.53	728.35	168.20	236.11	168.20	118.94	194.17	118.94
Dtr	10	0.00	0.00	0.00	194.16	670.82	194.16	22.56	172.35	22.56	17.53	169.24	17.53
Dd	300	0.00	0.00	0.00	10.45	7.64	7.64	1.93	1.51	1.51	1.68	1.52	1.52
Dr	150	0.01	0.11	0.02	675.00	42.50	85.00	61.66	28.50	42.92	26.44	11.70	17.50
El	10	89.56	147.75	89.56	1895.25	1802.10	1895.25	891.49	892.03	891.49	291.62	287.57	291.62
Er	10	0.01	0.26	0.01	0.99	0.70	0.99	0.50	0.47	0.50	0.08	0.07	0.08
Ir	10	0.00	284.58	0.00	227.45	1234.76	227.45	69.43	706.74	69.43	21.39	174.00	21.39
Pfc	90	-40.81	-0.42	-1.71	52.78	0.39	1.91	0.02	0.00	0.01	1.76	0.09	0.30
Plc	30	-57.09	-0.37	-10.09	28.95	0.35	7.07	0.02	0.00	0.02	1.56	0.10	1.01
Sl	10	0.00	0.23	0.00	80.40	38.41	80.40	31.68	16.80	31.68	10.13	6.85	10.13
SPI	60	-3.95	-3.43	-3.95	18.71	18.06	18.49	5.29	5.72	5.75	2.67	6.59	4.88
STCI	60	0.05	0.10	0.05	10329.02	3726.50	6334.71	62.28	206.43	136.34	72.82	278.56	187.04
TCI	60	-647.93	0.07	-48.75	257.42	17.97	34.07	2.17	6.08	3.16	17.12	5.71	5.79
TWI	30	-1.19	0.34	-0.74	19.40	19.41	19.40	5.88	7.51	6.38	2.47	6.02	3.77

5.3. Relative importance of landslide causative parameters

The relative importance of each optimal parameter is shown in **Figure 5-4.** Parameters *Dr*, *Sl*, *El*, *Ir*, *STCI,* and *TWI* for Niigata and *Dr*, *TCI*, *Plc*, *Cr*, *Sl*, and *TWI* for Ehime constitute the top six parameters explaining the distribution of landslides in each area. Parameters with higher relative importance (≥8; **Figure 5-4**) in both areas are *Cr*, *Dtd*, *Dr*, *Ir*, *Sl*, *TCI*, and *TWI*. Some parameters were found to be of greater importance in one area but not in the other. The parameters with large differences in relative importance between the two areas (≥5; **Figure 5-4**) are *Asp*, *El*, *Ir Plc*, *SPI*, *STCI*, and *TCI*; Parameters *El*, *Ir*, *SPI*, and *STCI* have higher importance in Niigata while *Asp*, *Plc*, and *TCI* have higher importance in Ehime. The random variable "*rand*" was correctly identified as the least important parameter in the analysis (**Table 5-10**).
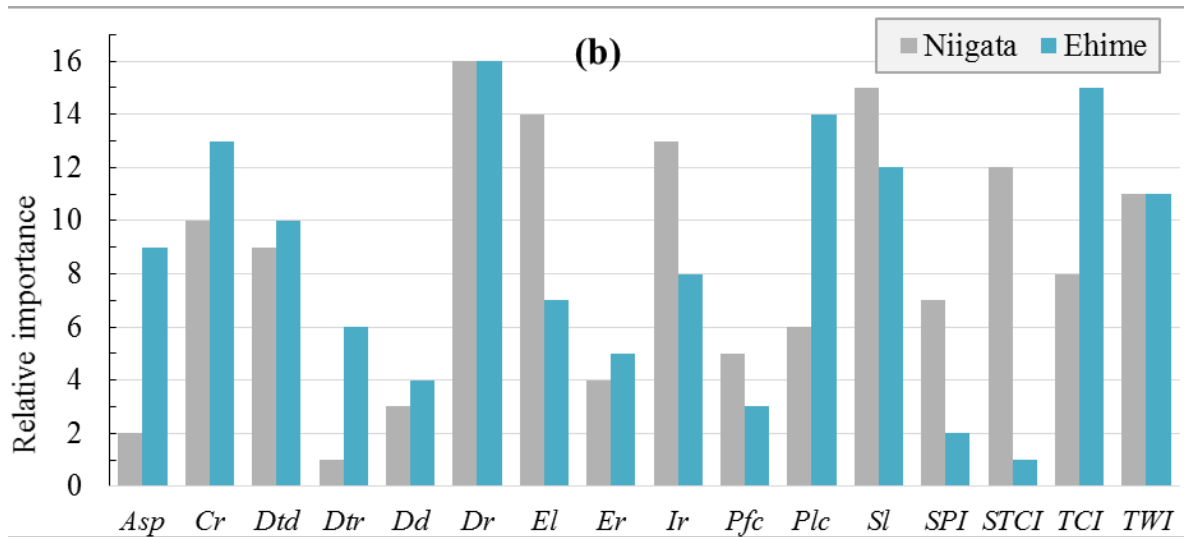
Figure 5-4. Relative importance of parameters for Niigata and Ehime.

Table 5-10. Rank of importance for parameters at their optimal scale (Niigata and Shikoku).

Rank	Niigata			Ehime		
	Parameter	Portion contributed	Likelihood ratio chi-square (G^2)	Parameter	Portion contributed	Likelihood ratio chi-square (G^2)
1	Dr_60	0.1711	392429	Dr_150	0.1067	88023
2	Sl_10	0.1101	252625	TCI_60	0.0945	77974
3	El_30	0.1100	252218	Plc_30	0.0911	75154
4	Ir_10	0.1081	247949	Cr_90	0.0803	66303
5	STCI_60	0.1022	234473	Sl_10	0.0745	61485
6	TWI_30	0.0809	185533	TWI_30	0.0641	52909
7	Cr_60	0.0570	130701	Dtd_10	0.0638	52672
8	Dtd_10	0.0495	113647	Asp_10	0.0612	50470
9	TCI_60	0.0474	108675	Ir_10	0.0541	44660
10	SPI_60	0.0336	76989	El_10	0.0514	42406
11	Plc_60	0.0316	72377	Dtr_10	0.0477	39328
12	Pfc_60	0.0240	55074	Er_10	0.0436	35965
13	Er_10	0.0224	51487	Dd_300	0.0428	35335
14	Dd_300	0.0212	48609	Pfc_90	0.0388	31994
15	Asp_300	0.0202	46257	SPI_60	0.0357	29456
16	Dtr_60	0.0108	24680	STCI_60	0.0262	21628

5.4. Comparison of topographic representation techniques

The results of comparative analysis of topographic representation techniques in **Figure 5-5** and **Table 5-11** indicate that representation methods perform the best with parameters at optimal scale. Additionally, the newly proposed DCD, with a testing accuracy of 81.11% in Niigata and 83.28% in Ehime, outperforms other methods. It is followed by the center-cell approach, while the seed cells approach of representation lagged behind.
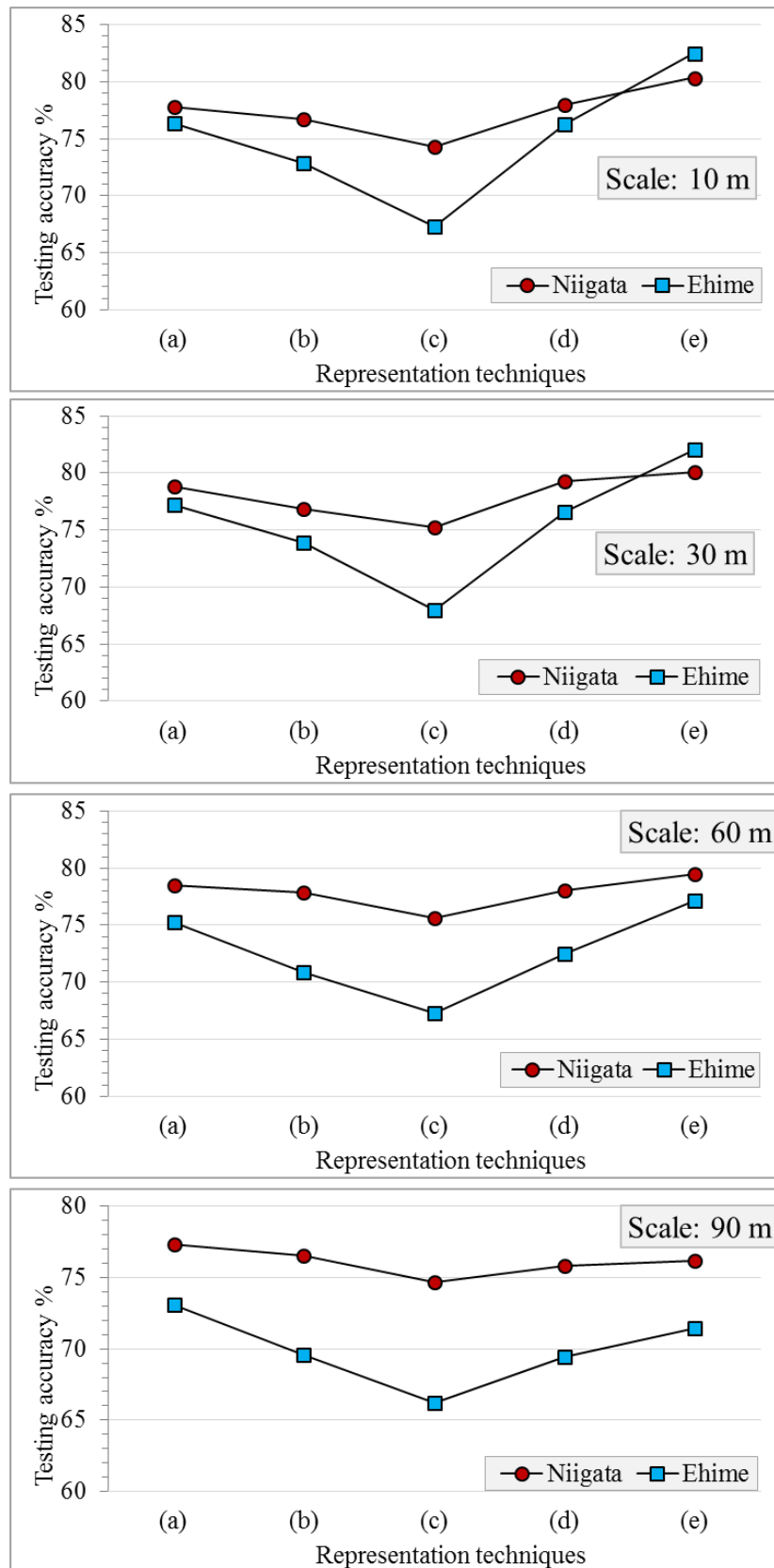
Figure 5-5. Accuracy assessment of the representation techniques at various scales. (a) center-cell, (b) cells within landslide polygon, (c) seed cells, (d) cells within the depletion zone, and (e) dominant within the depletion zone (DCD).
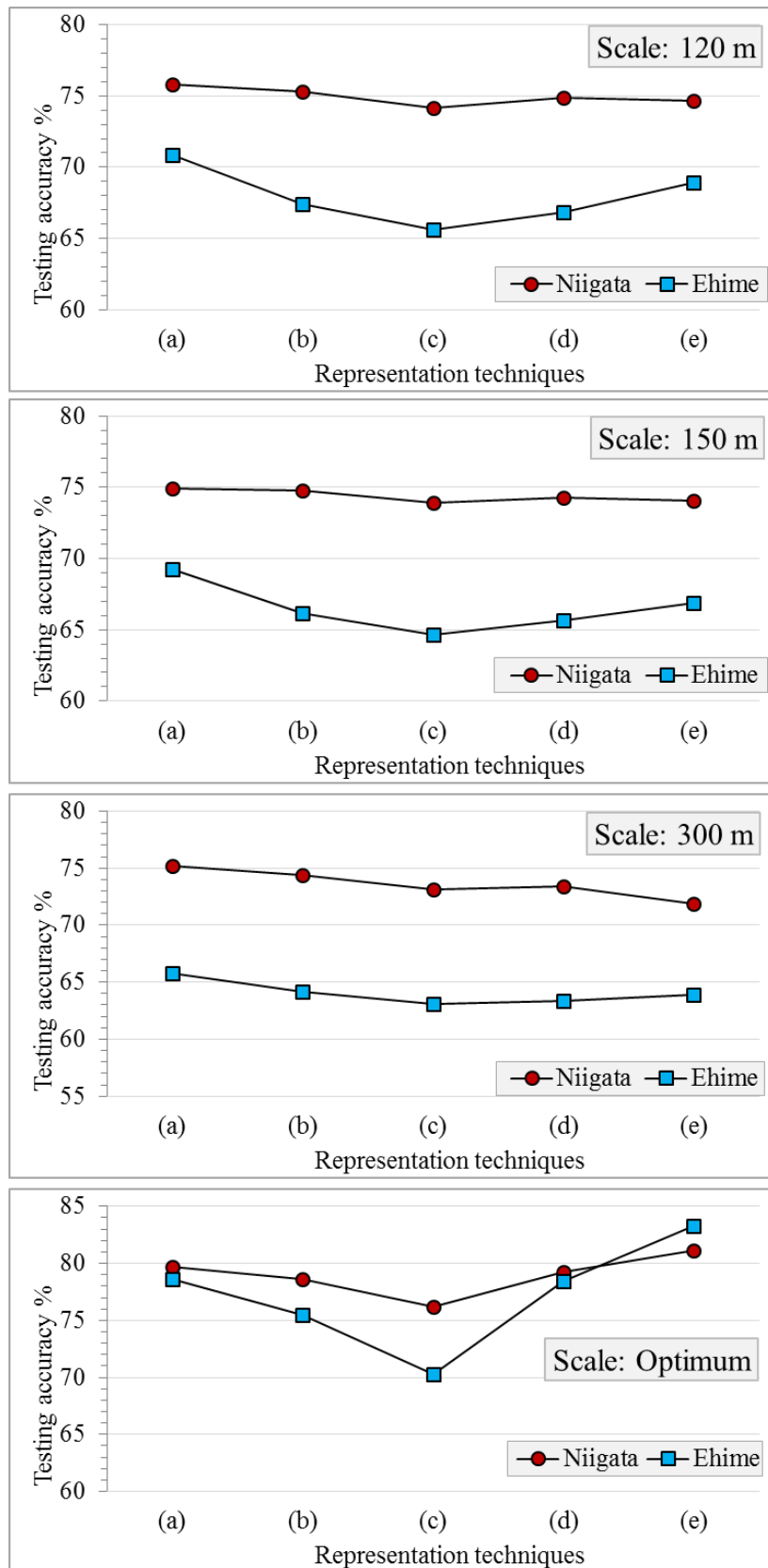
Figure 5-5. Continued.

Table 5-11. Accuracy assessment of the representation techniques at various scales.

Symbols	Representation	Scale	Testing accuracy	
			Niigata	Ehime
(a)	Center-cell	10	77.80	76.32
(b)	Cells within landslide polygon	10	76.69	72.86
(c)	Seed cells	10	74.28	67.30
(d)	Cells within the depletion zone	10	77.94	76.28
(e)	DCD	10	80.37	82.55
(a)	Center-cell	30	78.85	77.22
(b)	Cells within landslide polygon	30	76.84	73.90
(c)	Seed cells	30	75.28	67.97
(d)	Cells within the depletion zone	30	79.24	76.57
(e)	DCD	30	80.07	82.04
(a)	Center-cell	60	78.47	75.23
(b)	Cells within landslide polygon	60	77.83	70.87
(c)	Seed cells	60	75.64	67.27
(d)	Cells within the depletion zone	60	78.03	72.51
(e)	DCD	60	79.47	77.16
(a)	Center-cell	90	77.32	73.11
(b)	Cells within landslide polygon	90	76.54	69.60
(c)	Seed cells	90	74.71	66.24
(d)	Cells within the depletion zone	90	75.85	69.45
(e)	DCD	90	76.20	71.44
(a)	Center-cell	120	75.82	70.88
(b)	Cells within landslide polygon	120	75.30	67.43
(c)	Seed cells	120	74.13	65.61
(d)	Cells within the depletion zone	120	74.89	66.82
(e)	DCD	120	74.67	68.91
(a)	Center-cell	150	74.89	69.26
(b)	Cells within landslide polygon	150	74.74	66.14
(c)	Seed cells	150	73.94	64.67
(d)	Cells within the depletion zone	150	74.27	65.63
(e)	DCD	150	74.03	66.90
(a)	Center-cell	300	75.21	65.74
(b)	Cells within landslide polygon	300	74.41	64.14
(c)	Seed cells	300	73.13	63.08
(d)	Cells within the depletion zone	300	73.38	63.37
(e)	DCD	300	71.92	63.90
(a)	Center-cell	Optimum	79.70	78.62
(b)	Cells within landslide polygon	Optimum	78.63	75.51
(c)	Seed cells	Optimum	76.20	70.27
(d)	Cells within the depletion zone	Optimum	79.28	78.45
(e)	DCD	Optimum	81.11	83.28

5.5. Landslide susceptibility models

Figure 5-6 presents the testing and training accuracies (averaged over 10 iterations) of the RF models constructed with parameters at various scales using center-cell method. In both areas, the training accuracies (85.08% for Niigata and 95.44% for Ehime) and testing accuracies (79.70% and 78.62%, respectively) were found to be highest for the models with parameters at the optimal scales. Among the model iterations, the models closest to the mean testing accuracy were used to produce LS maps. **Figure 5-7** shows 25 km^2 parts of the LS maps for the combined optimal scale and the 10, 30, 90, and 150 m scales.

The effectiveness of the model with parameters at the optimal scales was evaluated using *AUC* (**Section 4.1**). **Figure 5-8** illustrates *AUC* values on test data for RF models at different scales. The model with the parameters at optimal scales shows an *AUC* value of 0.877 for Niigata and 0.870 for Ehime. The highest *AUC* values show that multi-resolution LS modeling outperforms the conventional single-scale modeling.

	optimum	10	30	60	90	120	150	300
Nii-train	85.08	83.87	84.69	83.92	83.05	82.11	81.45	81.47
Nii-test	79.70	77.80	78.85	78.47	77.32	75.82	74.89	75.21
Ehi-test	78.62	76.32	77.22	75.23	73.11	70.88	69.26	65.74
Ehi-train	95.44	94.43	94.15	93.76	95.04	94.40	95.24	95.10
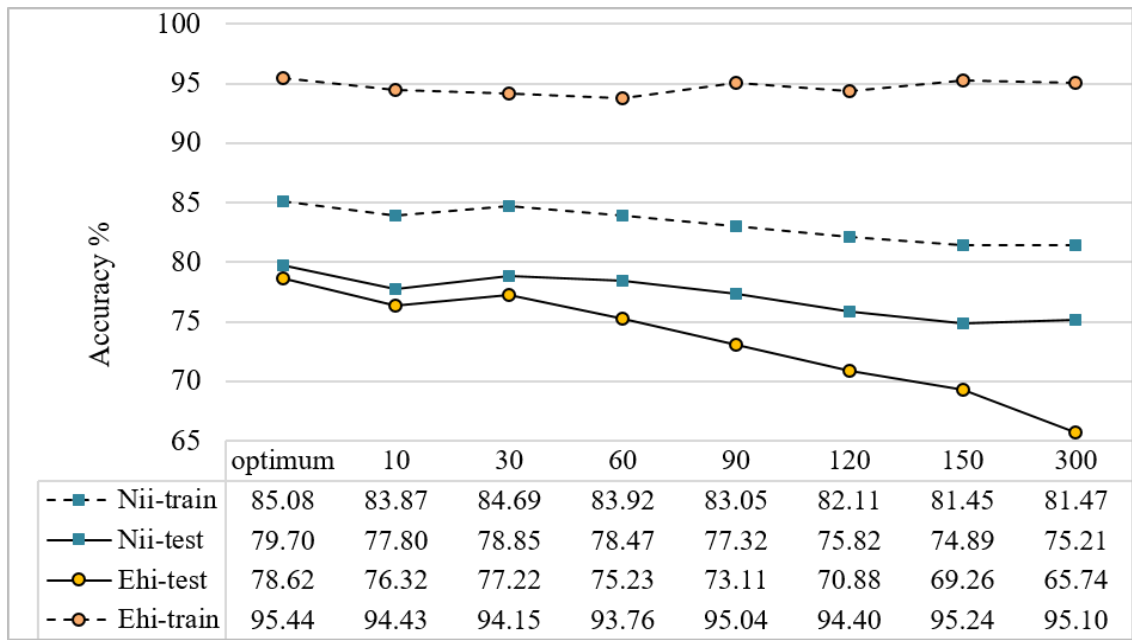
Figure 5-6. Training and testing accuracies of LS models for different parameter scales for Niigata and Ehime.
(Nii-train = training samples at Niigata; Nii-test = testing samples at Niigata; Ehi-train = training samples at Ehime; and Ehi-test = training samples at Ehime)
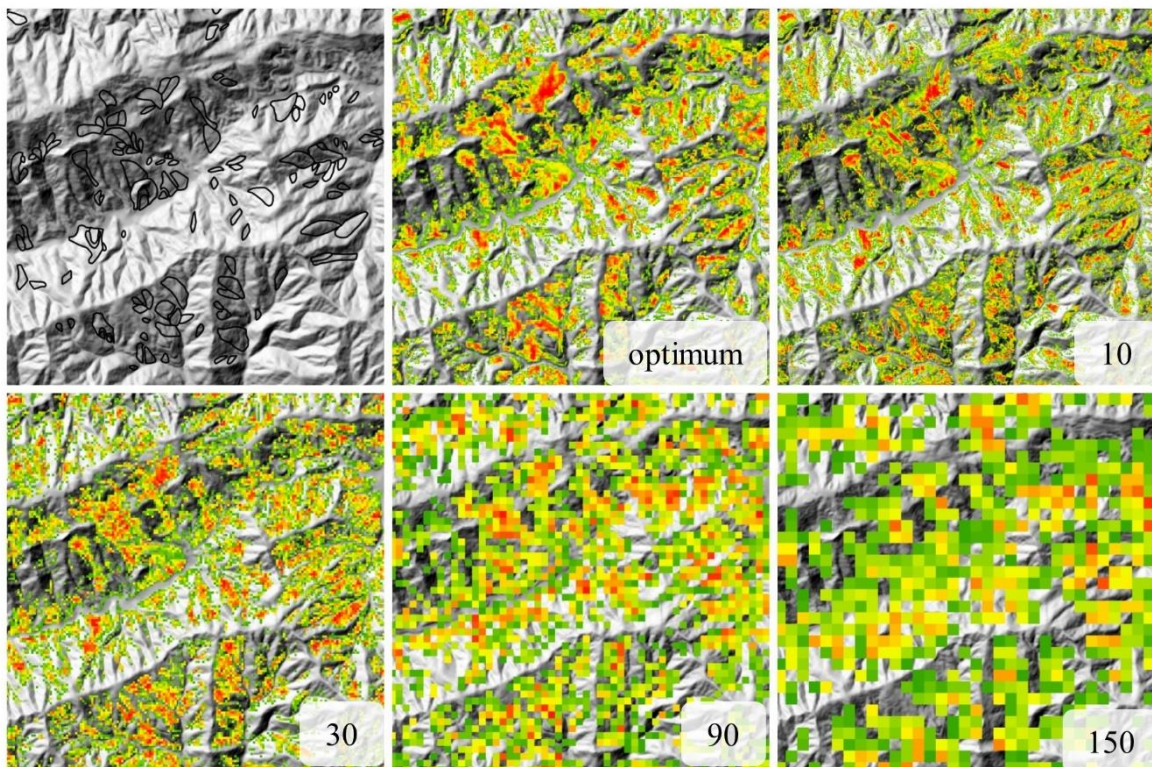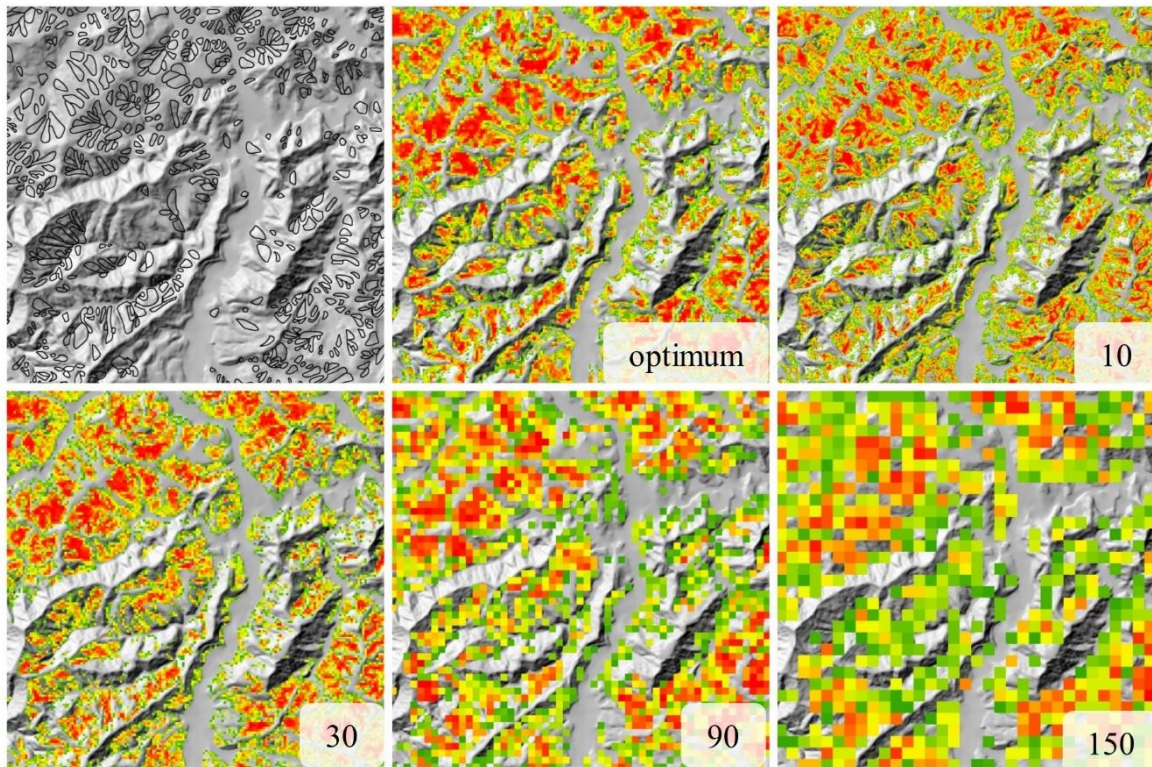
Figure 5-7. Landslide distribution map (gray) and part of LS maps (25 km^2 and probability >= 0.5) from RF models with different parameter scales for Niigata (upper) and Ehime (lower).
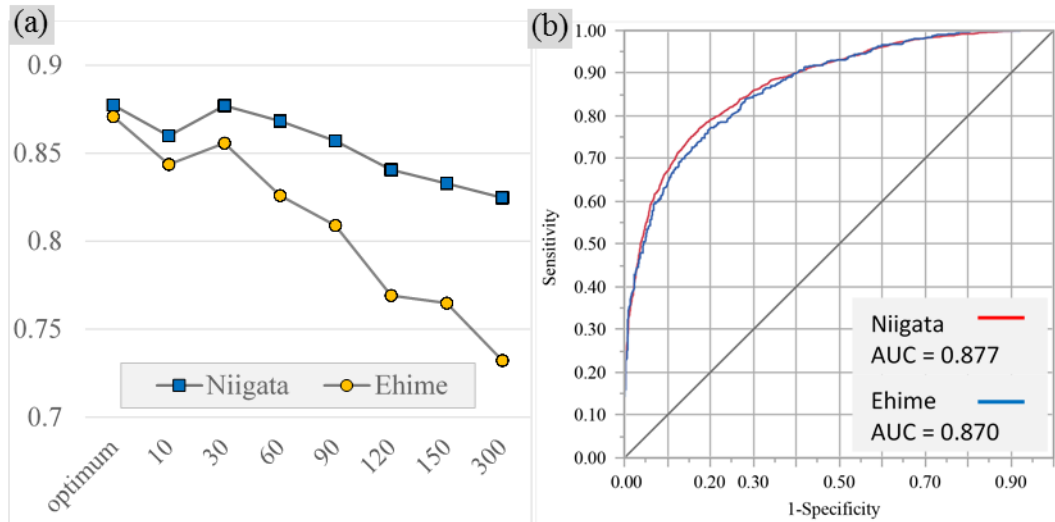
Figure 5-8. Results of ROC analysis. (a) *AUC* values of ROC on test data for the models at various scales. (b) ROC curves for the multi-resolution LS model with parameters at optimal scale.

Table 5-12 presents a summary of the testing accuracies and *AUC* estimates for the LS models with various representation techniques with parameters at optimal scale. LS models with testing accuracy of 81.11% in Niigata and 83.28% in Ehime were obtained with a combined use of the multi-resolution method (**Section 4.3**) and DCD technique (**Section 4.4**). The *AUC* values of 0.89 for Niigata and 0.92 for Ehime support their excellent predictive capacity (**Figure 5-9**). **Figure 5-10** shows LS maps prepared using the same LS models.

Table 5-12. Testing accuracies and *AUC* values for representation techniques with parameters at optimal scale.

Representation technique	Niigata		Ehime	
	Testing accuracy %	*AUC*	Testing accuracy %	*AUC*
Center-cell	79.70	0.88	78.62	0.87
Cells within landslide polygon	78.63	0.87	75.51	0.84
Seed cells	76.20	0.84	70.27	0.77
Cells within the depletion zone	79.28	0.87	78.45	0.86
DCD	81.11	0.89	83.28	0.92
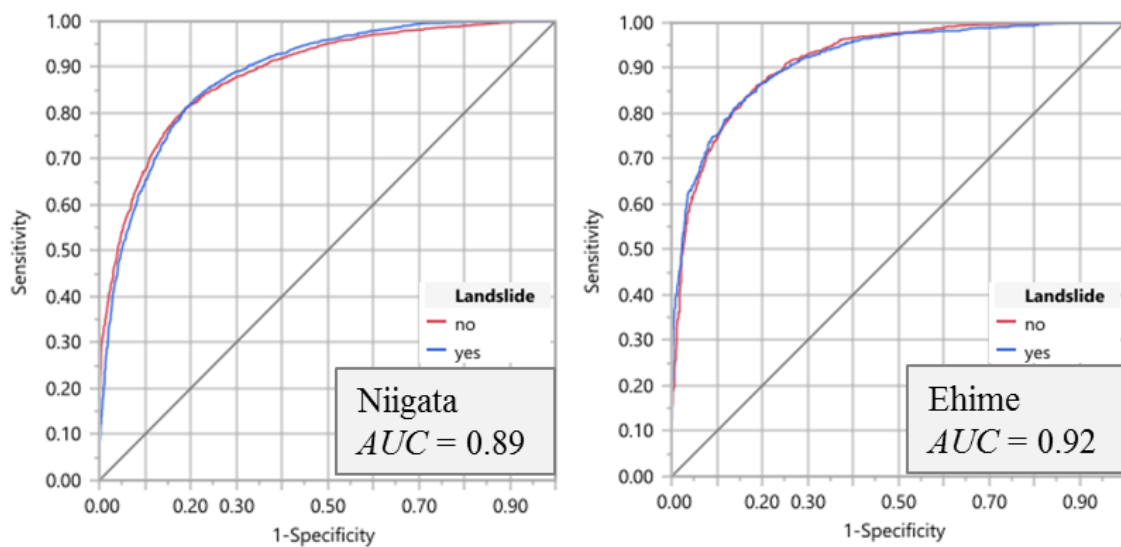
Figure 5-9. ROC plots for DCD with parameters at optimal scale.
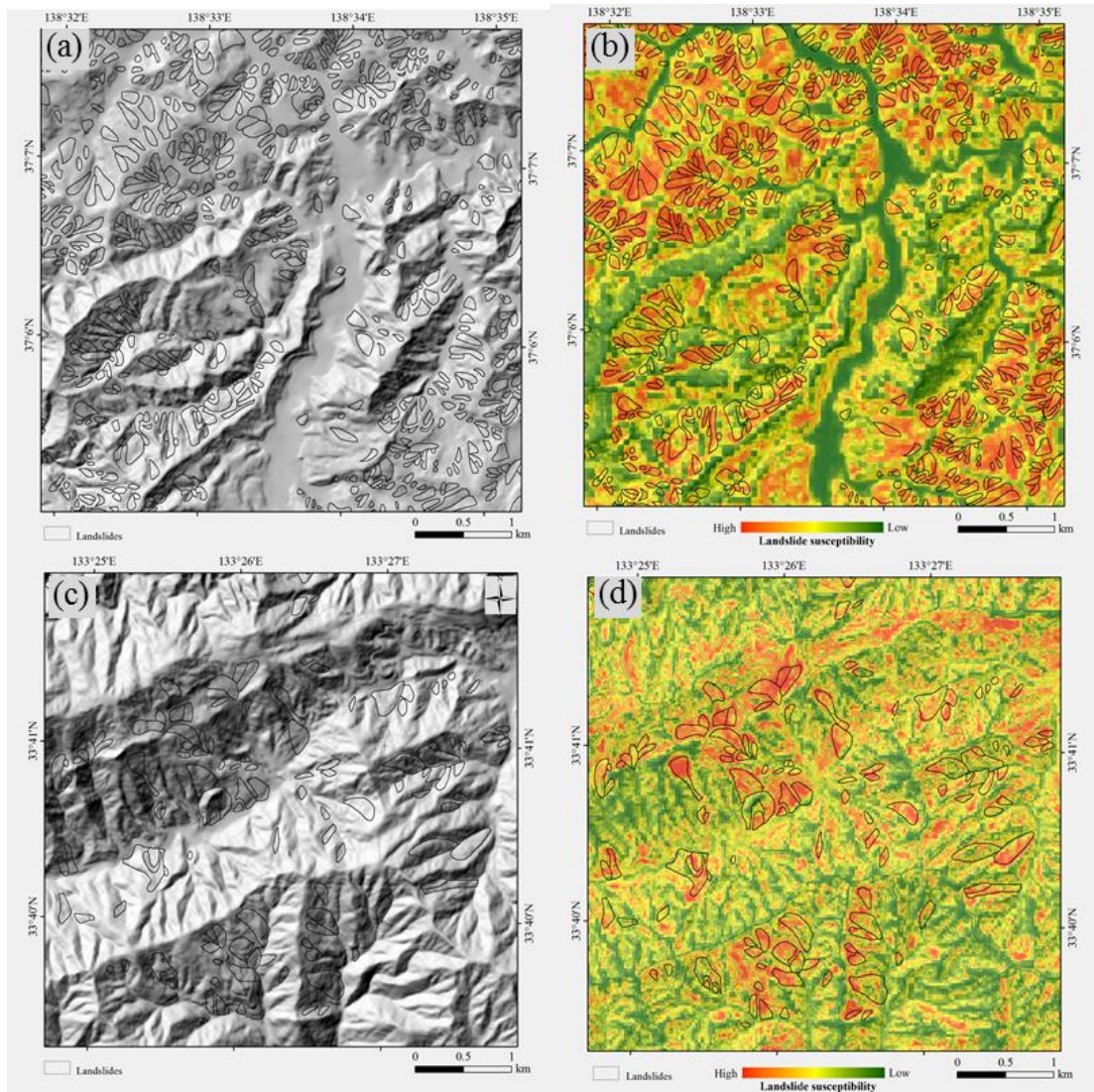
Figure 5-10. Landslide distribution map (gray) and parts of LS maps (25 km^2) for Niigata (a and b) and Ehime (c and d) with a combined use of optimal parameter-scale and DCD.

5.6. Landslide susceptibility mapping in Melamchi

The developed methodologies were applied to evaluate LS in Melamchi, Nepal. Only the multi-resolution LS analysis could be implemented because the landslides there were mapped as polylines without areal extent. Five DEM-scales (30 to 150 m) were used to identify the optimal scales of topographic parameters, which were then used to construct a multi-resolution LS model (**Section 4.3**). **Table 5-12** presents the optimal parameter-scale, their rank, and their contribution to the classification of the testing. *El*, *Er*, *Ir*, and *Sl* were four most influential parameters associated with landslide occurrences in the region. Topographic parameters, *Plc*, *Dtr*, and *TCI* were ranked below *rand* suggesting that they provided no information to the classification. **Figure 5-11** includes the training accuracies as well as testing accuracies of LS models with parameters at multiple scales. As in the Japanese study areas, the LS model with the optimal parameter-scale (testing accuracy = 84.5% and *AUC* = 0.92) was found to perform better than other single parameter-scale models. **Figure 5-12** shows the ROC plot for the multi-resolution model, and **Figure 5-13** shows the produced LS map based on the best model.

Table 5-13. Rank of importance for parameters at their optimal scale in Melamchi for LSM.

Rank	Optimal parameter-scale		Likelihood ratio chi-square (G^2)	G^2	Portion contributed
1	*El*	90	47300.2	+++++++++++++++++	0.1687
2	*Er*	150	33982.4	++++++++++++	0.1212
3	*Ir*	30	24719.9	++++++++	0.0882
4	*Sl*	120	23071.5	++++++++	0.0823
5	*Dr*	30	21346.3	+++++++	0.0761
6	*Dd*	90	16837.4	++++++	0.06
7	*STCI*	120	15602.1	+++++	0.0556
8	*SPI*	150	14142.7	+++++	0.0504
9	*Asp*	150	13129.6	++++	0.0468
10	*Dtd*	30	12105.6	++++	0.0432
11	*Pfc*	90	11974.2	++++	0.0427
12	*Cr*	90	11790.9	++++	0.042
13	*TWI*	120	7580.32	++	0.027
14	*rand*		7321.59	++	0.0261
15	*Plc*	120	6784.94	++	0.0242
16	*Dtr*	120	6518.76	++	0.0232
17	*TCI*	120	6216.95	++	0.0222

	Optimal	30	60	90	120	150
- ■ - Training	93.94	92.84	92.77	93.76	93.48	93.43
—○— Testing	84.50	81.59	81.70	82.25	81.54	81.95
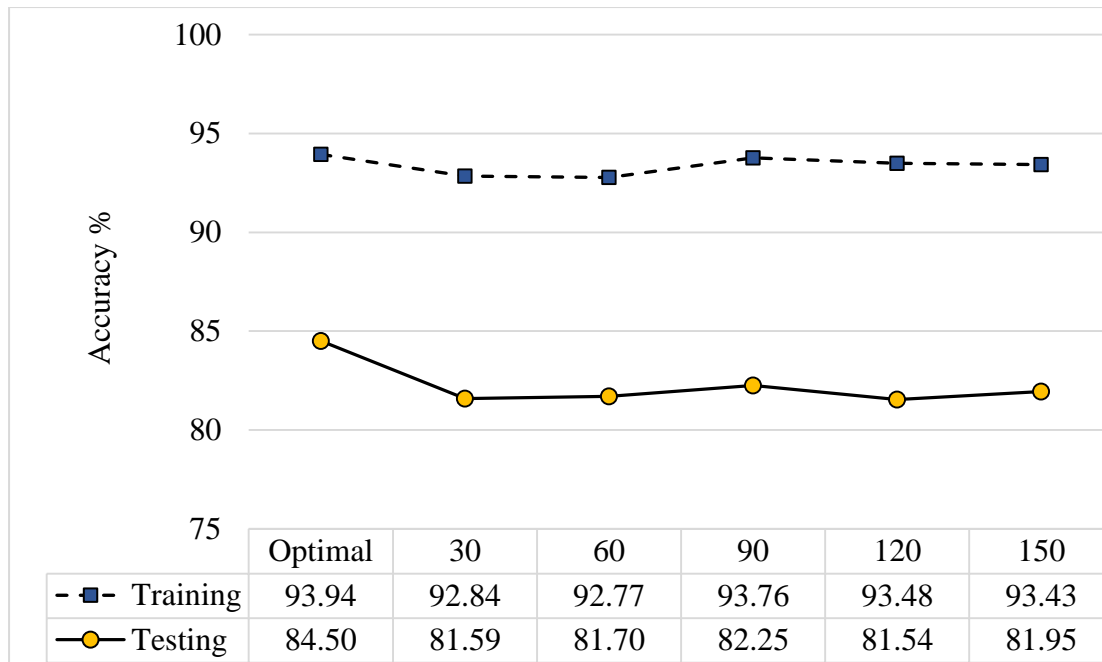
Figure 5-11. Testing accuracy of LS models for Melamchi at various scales.

Landslide	Area
no	0.9254
Yes	0.9254
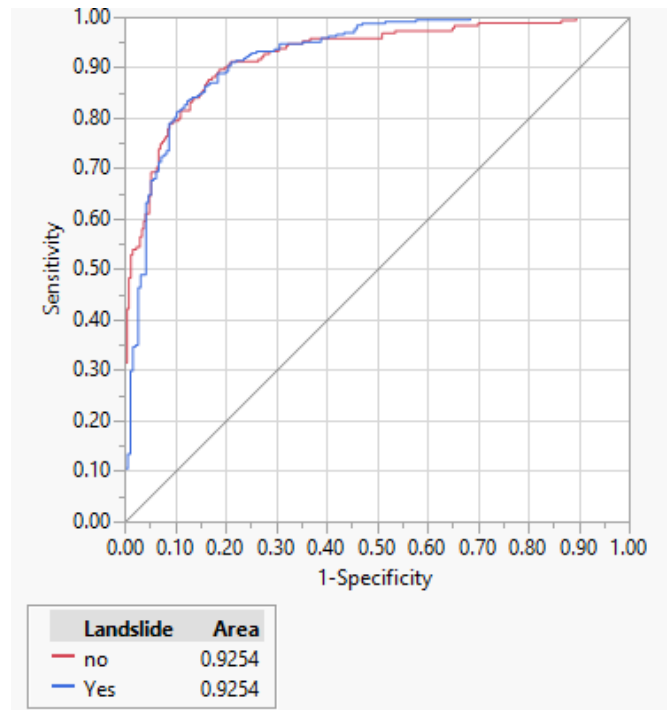
Figure 5-12. ROC plot for the LS model with the optimal parameter-scale in Melamchi.
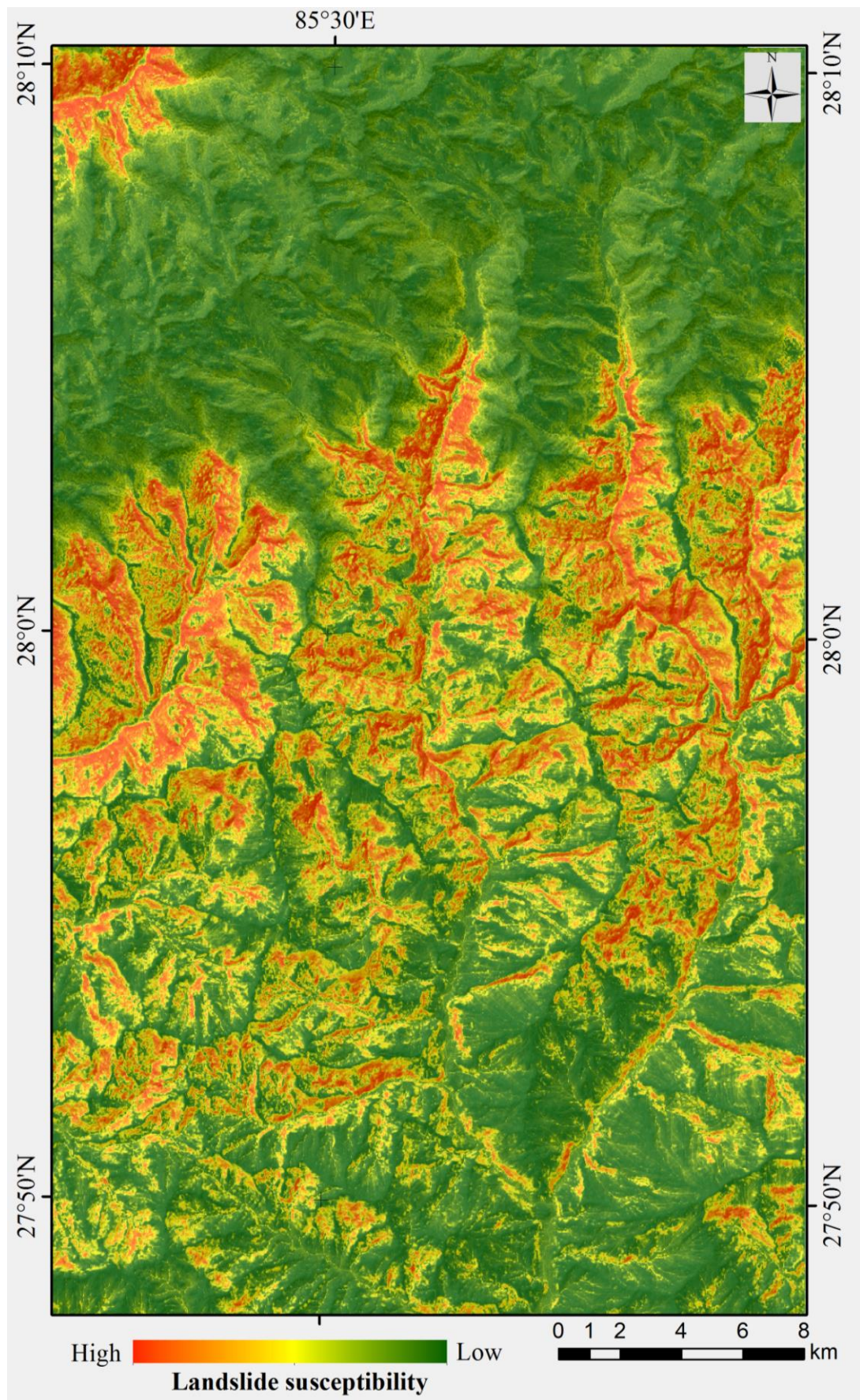
Figure 5-13. LS map of Melamchi with the use of the multi-resolution technique.

Chapter 6. Discussion

6.1. Random Forest for evaluating landslide susceptibility

The result for the number of trees for a robust LS modeling with RF (**Figure 5-1**) shows that RF models can achieve suitable complexity with forest structure easily manageable by most desktop computers ($T\# < 100$; Catani et al., 2013). Application in different fields often require much larger forest densities (Bachmair and Weiler, 2012). This enables the use of larger datasets in LS modeling (Archer and Kimes, 2008), otherwise restricted to the use of smaller samples due to computational requirements of other ML methods as in Kawabata and Bandibas (2009). RF models can handle categorical data, unbalanced data as well as data with missing values, not possible in other ML techniques such as SVM (Pal, 2005), which further encourages its practical implementation. Another aspect of the RF model, valuable in LS analysis, is the measure of variable importance. It was instrumental in this study for the identification of the optimal scale of topographic parameters and the proposed multi-resolution technique (**Section 4.3**).

6.2. Significance of DEM-based landslide susceptibility analysis

Numerous parameters from diverse themes are used in LSM. However, limitations in availability and scale of data may restrict the use of such studies. A DEM-based approach however permits quantitative, reproducible, and efficient LSM. The global availability of DEMs with resolutions suitable for LSM is also useful in this regard. This study investigated the usability of a DEM-based LSM using RF. The methodologies designed using high quality data in Japan were also successfully applied to a study area in Nepal, where high quality data lacks.

There is a general consensus that geological information is one of the most decisive parameters regarding landslide manifestation and therefore is included in most LS studies (Atkinson and Massari, 1998; Clerici et al., 2006; Nourani et al., 2013; Pourghasemi et al., 2012; Rossi et al., 2010). Picanço et al. (2014) also suggested the importance of geology in landslide studies due to its high correlation to occurrence and typology of landslides.

However, this study has revealed that the necessity of using geological information in addition to topographic parameters is not always high. While there was a little positive improvement on the overall accuracy of the LS model in Niigata when geological information was added, its influence in Ehime was insignificant. This observation for Niigata is consistent to Kawabata and Bandibas (2009) who found that geological information in Niigata was crucial in LS; model accuracy halved with the exclusion of geological information. Such dependence might be because the study analyzed landslides triggered by the 2004 Mid-Niigata Prefecture earthquake while the current study analyzes historical landslides irrespective of triggering mechanism. Nevertheless, the positive influence of geology altogether, and higher importance of *geo_code* and *geo_den* (**Table 5-1**) might be linked to the high seismicity of the region (Wang et al., 2007) because of the high correlation between seismic velocity and surface geology (Bard, 1995). The higher importance of *geo_den* might therefore be due to increased concentration of tectonic activity along the relatively weaker geological regions prone to landslides (Chuang et al., 2009). In Shikoku, Yamasaki and Chigira (2008) suggest that lithology exerts strong control on weathering and therefore the distribution of landslides. However, current study in Ehime found that the addition of geological information resulted in very little improvement of the LS model. This could be because while geology is

important, the scale of geological maps might not fully depict the conditions leading to slope failures. For example, the degree of weathering or the thickness of unstable regolith is usually not presented in geological maps. Dhakal et al. (2000) also pointed out that the spatial variation in geology relating to landslides is often not captured in geological maps. Instead, terrain variations at finer scales can better reflect terrain attributes, including soil characteristics (Iwahashi and Pike, 2007; Moore et al., 1993; Prima et al., 2006). Therefore, there are cases in which LSM based only on DEMs are meaningful enough as in the present study.

In a DEM-based LSM, however, it should be noted that post-failure DEMs include terrain modifications due to landslides. This makes the use of historical landslides (or those occurred prior to the DEM acquisition) in LS more challenging. Some techniques to mitigate this issue by reconstructing the pre-failure terrain have also been proposed (Gorum et al., 2008; Van Den Eeckhaut et al., 2006), but are not commonly employed. These imperfections of the DEM are therefore a target of future LS studies.

6.3. Factors influencing scale-sensitivity of topographic factors

Different terrain parameters vary in different ways when the DEM resolution changes (Zhou and Chen, 2011). Results from Ehime and Niigata also show that the optimum scales for LS modeling differ according to parameters (**Table 5-2** and **Table 5-3**), and tend to be common for the two study areas **(Figure 5-2)**. Although it is expected that the finest DEM can describe detailed topography and is hence suitable for LSM, several parameters are more significant at relatively coarse scales ($\geq$30 m). This suggests that the smallest-scale variabilities of these parameters do not well represent the physical processes of landslide triggering, as suggested by

some previous studies (Freer et al., 2002; Tarolli and Tarboton, 2006). For example, all curvature-related variables (*Cr*, *Pfc*, and *Plc*) and the composite topographic indices (*SPI*, *STCI*, *TCI*, and *TWI*) show meaningful influences on LS at scales equal to or larger than 30 m (**Figure 5-2**). This suggests that a 3×3 moving window for parameter computation properly encompasses a meaningful topographic unit, including both the detachment and deposition areas of a landslide, only at relatively coarse resolutions (Catani et al., 2013). The landslide size distribution of the two study areas (**Figure 3-1**, **Table 3-1**) also suggests that most landslides are larger than the terrain represented at 10 m resolution. Scaling of *Dr* with the optimal 60 or 150 m resolution also seems to correspond to the landslide size distribution. In contrast, for *Ir* and *Er*, the finest resolution (10 m) is the optimal scale for both areas, and it is ascribable to a larger 10×10 moving window used for their computation, which can represent a relatively large area even at the finest resolution. However, although computed using a 3×3 moving window, *Sl* is optimal at the finest scale. This seems to reflect the high sensitivity of slope calculation to DEM resolution. It is widely known that coarser resolution DEMs result in lower *Sl* values for the same terrain (**Figure 6-1**) (Zhang et al., 1999). Because *Sl* is directly related to gravitational force triggering landslides, its accurate computation using fine DEMs is important. A similar explanation can be given for *Dtd,* which is also optimal at the finest 10 m scale. Fluvial activity such as channel erosion tends to induce landslides along the river course. The accurate location of rivers is better represented if the finest DEM is used (Wang and Yin, 1998).
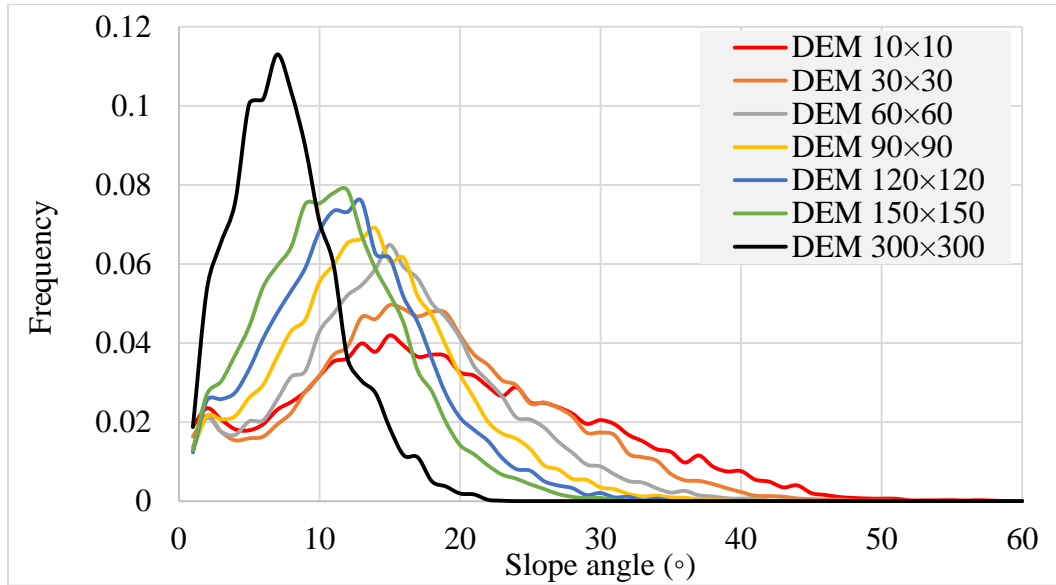
Figure 6-1. Relationship between DEM resolution and slope angle for Niigata.

For parameter *Dtr*, a coarser scale (60 m) for Niigata and the finest scale (10 m) for Ehime were found to be optimal. Ridges extracted at coarser scales usually correspond to major ridge lines, while at finer scales they include local topographic highs (ESRI, 2016). *Dtr* for Niigata at a coarser scale could therefore include the amplified motion observed along the major outstanding ridges during seismic events (Chang et al., 2007). Indeed, some landslides in Niigata were due to high seismicity, as was the case of the 2004 Chuestu earthquake (Wang et al., 2007).

The optimal scales of *Dd* and *Asp* differ significantly from those of the other parameters; the coarsest scale (300 m) is optimum for *Dd,* while *Asp* shows the largest deviation between the two areas, 10 and 300 m. *Dd* in this study is estimated over a unit area dependent on the scale of analysis, thus at finer scales it might be related only to the local presence/absence of drainage lines. However, at coarser scales, it can reflect the known relationship between general relief characteristics and landslide occurrences (Lin and Oguchi, 2004; Oguchi, 1997; Strahler, 1952). For *Asp*, the finest resolution (10 m) is optimal for Ehime, while the coarsest

resolution (300 m) is optimal for Niigata. Local meteorological conditions and their relationship with LS may explain this variation. Both the study areas receive large amounts of precipitation; however, Niigata receives a significant portion of its precipitation as snowfall (1.35 m per year). The increased overburden due to accumulated snow and the increased soil moisture from snowmelt are responsible for landslides there (Kawagoe et al., 2009). *Asp* at a coarser scale indicates the overall direction of a hillslope and suggests that the difference in the deposition thickness of snow on the windward and leeward sides is crucial for LS in Niigata. Dip-slope and dip-aspect of the predominantly sedimentary geology in Niigata might also have influenced this scale dependency. In contrast, *Asp* at a finer scale could depict local variations in micro-climate, such as insolation and related groundwater conditions, which is related to rock weathering (Rech et al., 2001). The close relationship between weathering and distribution of landslides has been reported in Shikoku Island including Ehime (Yamasaki and Chigira, 2008), indicating the effect of finer scale *Asp* on local climate, weathering, and landslides.

6.4. Geo-environmental influence on the importance of landslide causative factors

Among the values of relative parameter importance (**Figure 5-3**), higher values for *Cr*, *Dtd*, *Dr*, *Ir*, *Sl, TCI,* and *TWI* in both Niigata and Ehime suggest that these parameters are instrumental in landslide occurrences. Landslide probability generally increases with terrain slope because of increased shear stress, and slope is considered very important in LS studies (Lee and Talib, 2005; Nefeslioglu et al., 2008b). Therefore, the higher importance of *Dr* and *Sl* is reasonable. The higher importance of *Dr* compared to *Sl* confirms that *Dr* is a more direct representation of local maximum slope. Claessens et al. (2005) provided a similar observation on

the different effects of these slope parameters. The higher importance of *Ir* in both study areas also suggests the importance of topographic steepness.

The importance of *Dtd* is explained by the bidirectional relationship between fluvial processes and slope failures. While landslides contribute to channel initiation, stream incision also contributes to landslides (Gerrad and Gardner, 2000; Montgomery and Dietrich, 1989; Ng, 2006; Oguchi, 1997). The higher relative importance of *TCI* and *TWI* may reflect the significance of hydrological variations related to rock weathering and soil properties (Iwahashi and Pike, 2007; Moore et al., 1993; Prima et al., 2006). There is a general consensus regarding *Cr* that landslides are more likely to occur on concave slopes because of groundwater concentration (Ayalew et al., 2004). In contrast, earthquake-induced landslides may be more likely on convex slopes with higher ground acceleration. The importance of *Cr* in both study areas therefore hints to such mechanisms controlling LS.

Parameters *Asp*, *Plc,* and *TCI* have markedly higher importance in Ehime than in Niigata. A combination of geological and environmental variables may explain this observation. As noted, the importance of *Asp* in Ehime may be due to local micro-climatic differences that lead to differential weathering. By contrast, the higher importance of *Plc* in Ehime indicates the positive influence of horizontal flow movement in LS, as suggested by Nefeslioglu et al. (2008b) (**Table 3-3**). The area in Ehime receives a larger amount of rainfall than in Niigata; hence increased water concentration may contribute more to landslides. The higher importance of *TCI* in Ehime can be explained similarly because it is a parameter strongly related to terrain curvature.

6.5. Assessment of LS models

The performances of LS models at eight different scales for Niigata and Ehime and six different scales for Melamchi were compared based on their accuracy estimates (**Figure 5-6** and **Figure 5-11**) and *AUC* values (**Figure 5**-**8**). The scale dependency of input parameters was also observed in the accuracy estimates of the LS models (**Figure 5-6** and **Figure 5-11**). For the study areas in Japan, except for the training accuracies of LS models for Ehime and an LS model for Niigata at 300 m resolution, the accuracy estimates decrease with an increasing analytical scale beyond 30 m. The slight increase in the model accuracy for Niigata at 300 m is the effect of the two parameters that are optimal at that scale (**Figure 5-2**). The discrepancy observed with the training accuracies for Ehime might be due to the smaller number of training samples. Higher testing accuracies were obtained for models at coarser scales ($\geq$30 m) than at the finest scale (10 m). Similar results were obtained in Melamchi where the finest scale (30 m) was found to be the least suitable scale for LSM. For example, while parameter-scales show large differences, the influence of the most important parameter in Melamchi (*El*) is best explained at the scale of 90 m (**Figure 5-11**)**.** The proposed multi-resolution LS technique resulted in a boost of testing accuracies (**Figure 5-6** and **Figure 5-11)** in all the study areas. Obtained *AUC* values also suggest their excellent predictive abilities (**Section 4.1**, **Figure 5-8**).

Two examples of local LS maps at different scales for Niigata and Ehime (**Figure 5-7**) indicate that the usability of an LS map depends on the mapping scale as well as the model used. The LS maps in **Figure 5-10** and **Figure 5-13** suggest the broad applicability of LSM using the proposed methodologies.

6.6. Topographic representation of landslides

Topographic representation is a crucial, yet not commonly discussed aspect of LS studies. It greatly affects the outcome of LSM because selecting one over the others leads to a possibility of selecting dissimilar initiation conditions for the same landslide.

A comparison between the five representation techniques (**Section 5.4**) shows that the newly proposed DCD method of representation always has the highest performance (higher testing accuracies and *AUC*, **Figure 5-5**, **Table 5-11**, and **Table 5-12**). This is logical because the dominant cell within the depletion zone most likely represents the dominant processes governing landslide initiations (Clerici et al., 2006; Trigila et al., 2015). The results in **Figure 5-5** and **Table 5-11** also suggest that LS models irrespective of representation methods perform better with parameters at optimal scale. This might be because optimal scales differ among parameters and LS models perform better whenever optimum scale for parameters is evaluated and utilized in LSM. **Section 6.3** details the scale sensitivity of parameters and the importance of the multi-resolution LS analysis in this regard.

The representation techniques exhibit marked distinctiveness at finer scales but not in coarser scales (**Figure 5-5**). This might be due to the terrain smoothing at coarser scales as well as the scale of analysis itself. At finer scales, different representation techniques identify distinct/dissimilar terrain conditions while at coarser scales they might relate to same/similar conditions producing similar results. Although this scaling of the representation techniques was observed in both the study areas, it was more pronounced in Niigata than in Ehime. Size of landslides in the two study areas (smaller in Niigata and larger in Ehime) (**Table 3-1**) could explain this observation.

The performance of DCD was followed by the center-cell and cells-within-the-depletion-zone types of topographic representation. While conditions at the depletion zone are more closely related to the actual conditions of landslide initiation, all cells are not the same. Random sampling of representative cell within the depletion zone therefore causes a decrease in its representative efficacy, nearing the performance of center-cell representation technique in corresponding LS models. Regarding the center-cell technique, Atkinson and Massari (2011) suggested its use to ensure that no preference is given to large or small landslides. However, the results in this study suggest that such a strategy will is not always valid.

All of the abovementioned techniques are restricted to the depletion zone. The same explains the lower efficiency of the cells-within-landslide-polygon technique, which includes cells sampled from all of the landslide body (depletion and accumulation). The cells in the accumulation zone are rarely representative of the landslide initiation conditions required in LSM.

Seed cells are regarded as a practical representation of pre-failure conditions and therefore preferred in LSM (Nefeslioglu et al., 2008b; Suzen and Doyuran, 2004b; Yilmaz, 2010b), however in this study it was not suitable (**Figure 5-5**, **Table 5-11**, and **Table 5-12**). Suzen and Doyuran (2004b) and Che et al. (2012) suggest that the seed cell approach is mostly suitable for small shallow translational slides which does not alter slope form. The dominance of large landslides (**Figure 3-3** and **Figure 3-4)** might therefore be responsible for the lowest performance of the seed cells. The subjectivity associated with the buffer zone length for seed-cell extraction (Wang et al., 2013), use of historical landslides, and terrain modifications since landsliding could have also lowered the efficacy of the representation technique.

However, for fast moving landslides covering large distances, the seed cell approach might be more useful than the other representation techniques because the distinction between depletion and deposition zones employed for other techniques cannot be applied to such landslides.

Chapter 7.   Conclusions

LSM provides the relative likelihood of future landslides, conditional on local geomorphic and topographic characteristics (Chung and Fabbri, 2008). Results of this study suggest that a single parameter-scale analysis falls short in accommodating the heterogeneity of geomorphological characteristics of the landslides and their surrounding areas. This study proposes a multi-resolution LSM technique to incorporate such variabilities resulting in more accurate LSM than any single-scale analysis. The method requires an identification of optimum scales for all parameters to best represent the conditions of slope failure. The parameters at different optimum scales are then brought together for the final LSM.

The study also demonstrated the usefulness of a DEM-based LS analysis in areas without other sets of high-quality thematic data. The influence of geological information on LSM was also analyzed. The result suggest that unless, available at higher scales, inclusion of geological information in LS analysis results only in a little increase in predictive ability of an LS model. Therefore, a DEM-based LSM looks promising, at least for steep areas experiencing active landsliding such as Japanese and Nepalese mountains. In other areas where landslides are more limited, landslides might concentrate in zones with particular factors including geological and anthropogenic ones.

The analysis of scale and importance of the DEM-derived parameters reveal that while some parameters show similar importance and scale dependency for different regions, environmental differences result in variability between regions. The performance of LS models also suggests that the finest scale of analysis is not always the best. The proposed multi-resolution LS analysis permits higher accuracy LSM than any single-scale analysis.

The study also evaluated different topographic representation techniques that determine how the modeling datasets are prepared. A new area-based representation technique concerning the dominant cell within the depletion zone is also proposed in this study. The proposed method of representation was found to be better than other four contemporary methods: center-cell, cells within landslide boundary, cells within depletion zone, and seed cells. The application of the newly proposed representation technique on top of the multi-resolution technique resulted in a cumulative increase in the efficacy of the DEM-based LSM.

Most importantly, the proposed methodologies, developed using higher resolution datasets in Japan, were also successfully applied to Nepal where the data quality is relatively low. This indicates broad applicability of the methodologies and fulfills the broad objective for this study.

Further study is recommended to confirm the usefulness of the proposed methodologies. Especially research in areas whose environmental characteristic differ from those in the Japanese and Nepalese study areas is necessary.

References

ADB, 2000: Summary environmental impact assessment of Melamchi water supply project.

Akgun, A., 2011: A comparison of landslide susceptibility maps produced by logistic regression, multi-criteria decision, and likelihood ratio methods: a case study at İzmir, Turkey. Landslides 9, 93–106.

Archer, K.J., Kimes, R. V., 2008: Empirical characterization of random forest variable importance measures. Comput. Stat. Data Anal. 52, 2249–2260.

Arnett, R.R., Conacher, A.J., 2007: Drainage basin expansion and the nine unit landsurface model. Aust. Geogr. 12, 237–249.

Atkinson, P.M., Massari, R., 2011: Autologistic modelling of susceptibility to landsliding in the Central Apennines, Italy. Geomorphology 130, 55–64.

Atkinson, P.M., Massari, R., 1998: Mapping susceptibility to landsliding in the central Apennines, Italy 24, 373–385.

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

Ayalew, L., Yamagishi, H., 2002: Landslides in the Kakuda-Yahiko Mountains of Niigata, Their Analyses and Description Using GIS, in: National Congress of Engineering Geological Society of Japan. Takamatsu, Japan, pp. 131–134.

Ayalew, L., Yamagishi, H., Ugawa, N., 2004: Landslide susceptibility mapping using GIS-based weighted linear combination, the case in Tsugawa area of Agano River, Niigata Prefecture, Japan. Landslides 1, 73–81.

Bachmair, S., Weiler, M., 2012: Hillslope characteristics as controls of subsurface flow variability. Hydrol. Earth Syst. Sci. 16, 3699–3715.

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.

Bai, S.B., Wang, J., Pozdnoukhov, A., Kanevski, M., 2009: Validation of Logistic Regression Models for Landslide Susceptibility Maps, in: 2009 WRI World Congress on Computer Science and Information Engineering. IEEE, pp. 355–358.

Band, L.E., 1986: Topographic Partition of Watersheds with Digital Elevation Models. Water Resour. Res. 22, 15–24.

Banno, S., Sakai, C., 1989: Geology and metamorphic evolution of the Sanbagawa metamorphic belt, Japan. Geol. Soc. London, Spec. Publ. 43, 519–532.

Bard, P., 1995: Effects of surface geology on ground motion: Recent results and remaining issues. Proc. 10Th Eur. Conf. Earthquaker Eng. 10, 305–323.

Beven, K., 1997: TOPMODEL: A critique. Hydrol. Process. 11, 1069–1085.

Birkeland, P.W., 1984: Soils and geomorphology. Oxford University Press.

Brabb, E., 1984: Innovative approaches to landslide hazard mapping, in: 4th International Symposium on Landslides. Toronto, pp. 307–324.

Brabb, E.E., Pampeyan, E.H., Bonilla, M.G., 1972: Landslide susceptibility in San Mateo County, California. Misc. F. Stud. Map.

Brardinoni, F., Hassan, M.A., Slaymaker, H.O., 2003: Complex mass wasting response of drainage basins to forest management in coastal British Columbia. Geomorphology 49, 109–124.

Breiman, L., 2001: Random forests. Mach. Learn. 45, 5–32.

Breiman, L., 1996: Bagging predictors. Mach. Learn. 24, 123–140.

Brenning, A., 2005: Spatial prediction models for landslide hazards: review, comparison and evaluation. Nat. Hazards Earth Syst. Sci. 5, 853–862.

Bricher, P.K., Lucieer, A., Shaw, J., Terauds, A., Bergstrom, D.M., 2013: Mapping Sub-Antarctic Cushion Plants Using Random Forests to Combine Very High Resolution Satellite Imagery and Terrain Modelling. PLoS One 8, 1–15.

British Geological Survey, 2015: Nepal earthquake response 2015 [WWW Document]. Earth Planet. Obs. Mapp. URL http://www.bgs.ac.uk/research/earthHazards/epom/NepalEarthquakeRespons e.html (accessed 12.10.15).

Cascini, L., Fornaro, G., Peduto, D., 2010: Advanced low- and full-resolution DInSAR map generation for slow-moving landslide analysis at different scales. Eng. Geol. 112, 29–42.

Castellanos Abella, E. a., Van Westen, C.J., 2008: Qualitative landslide susceptibility assessment by multicriteria analysis: A case study from San Antonio del Sur, Guantánamo, Cuba. Geomorphology 94, 453–466.

Catani, F., Lagomarsino, D., Segoni, S., Tofani, V., 2013: Landslide susceptibility estimation by random forests technique: Sensitivity and scaling issues. Nat. Hazards Earth Syst. Sci. 13, 2815–2831.

Catani, F., Segoni, S., Falorni, G., 2010: An empirical geomorphology-based approach to the spatial prediction of soil thickness at catchment scale. Water Resour. Res. 46, n/a-n/a.

Chang, K.-T., Chiang, S.-H., Hsu, M.-L., 2007: Modeling typhoon- and earthquake-induced landslides in a mountainous watershed using logistic regression. Geomorphology 89, 335–347.

Che, V.B., Kervyn, M., Suh, C.E., Fontijn, K., Ernst, G.G.J., Del Marmol, M. a., Trefois, P., Jacobs, P., 2012: Landslide susceptibility assessment in Limbe (SW Cameroon): A field calibrated seed cell and information value method. Catena 92, 83–98.

Chen, C.-Y., Yu, F.-C., 2011: Morphometric analysis of debris flows and their source areas using GIS. Geomorphology 129, 387–397.

Chigira, M., Yagi, H., 2006: Geological and geomorphological characteristics of landslides triggered by the 2004 Mid Niigta prefecture earthquake in Japan. Eng. Geol. 82, 202–221.

Choi, J., Oh, H.-J., Lee, H.-J., Lee, C., Lee, S., 2012: Combining landslide susceptibility maps obtained from frequency ratio, logistic regression, and artificial neural network models using ASTER images and GIS. Eng. Geol. 124, 12–23.

Chuang, S.C., Chen, H., Lin, G.W., Lin, C.W., Chang, C.P., 2009: Increase in basin sediment yield from landslides in storms following major seismic disturbance. Eng. Geol. 103, 59–65.

Chung, C.-J., Fabbri, A.G., 2008: Predicting landslides for risk analysis — Spatial models tested by a cross-validation technique. Geomorphology 94, 438–452.

Claessens, L., Heuvelink, G.B.M., Schoorl, J.M., Veldkamp, A., 2005: DEM resolution effects on shallow landslide hazard and soil redistribution modelling. Earth Surf. Process. Landforms 30, 461–477.

Claessens, L., Knapen, A., Kitutu, M.G., Poesen, J., Deckers, J.A., 2007: Modelling landslide hazard, soil redistribution and sediment yield of landslides on the Ugandan footslopes of Mount Elgon. Geomorphology 90, 23–35.

Clerici, A., Perego, S., Tellini, C., Vescovi, P., 2006: A GIS-based automated procedure for landslide susceptibility mapping by the Conditional Analysis method: The Baganza valley case study (Italian Northern Apennines). Environ. Geol. 50, 941–961.

Coblentz, D., Pabian, F., Prasad, L., 2014: Quantitative Geomorphometrics for Terrain Characterization. Int. J. Geosci. 5, 247–266.

Coe, J.A., Reid, M.E., Brien, D.L., Michael, J.A., 2011: Assessment of topographic and drainage network controls on debris-flow travel distance along the west coast of the United States, in: Genevois, R., Hamilton, D.L., Prestininzi, A. (Eds.), Proceedings of the 5th International Conference on Debris Flow Hazards Mitigation, Mechanics, Prediction and Assessmemt. Italian Journal of Engineering Geology and Environment and Casa Editrice Universita La Sapienza, Rome, Italy, Paduna, Italy, pp. 199–209.

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.

Crozier, M.J., 1999: Prediction of rainfall-triggered landslides: A test of the antecedent water status model. Earth Surf. Process. Landforms 24, 825–833.

Cruden, D.M., 1991: A simple definition of a landslide. Bull. Int. Assoc. Eng. Geol. 43, 27–29.

Cutler, A., Cutler, D.R., Stevens, J.R., 2012: Random Forests, in: Ensemble Machine Learning. pp. 157–175.

Cutler, D.R., Edwards, T.C., Beard, K.H., Cutler, A., Hess, K.T., Gibson, J., Lawler, J.J., 2007: Random Forest for classification in Ecology. Ecology 88, 2783–2792.

Dagdelenler, G., Nefeslioglu, H., Gokceoglu, C., 2015: Modification of seed cell sampling strategy for landslide susceptibility mapping: an application from the Eastern part of the Gallipoli Peninsula (Canakkale, Turkey). Bull. Eng. Geol. Environ.

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

Dai, F.C., Lee, C.F., Li, J., Xu, Z.W., 2001: Assessment of landslide susceptibility on the natural terrain of Lantau Island, Hong Kong. Environ. Geol. 40, 381–391.

Dhakal, A.S., Amada, T., Aniya, M., 2000: Landslide Hazard Mapping and its Evaluation Using GIS : An Investigation of Sampling Schemes for a Grid-Cell Based Quantitative Method. Photogramm. Eng. Remote Sens. 66, 981–989.

Díaz-Uriarte, R., Alvarez de Andrés, S., 2006: Gene selection and classification of microarray data using random forest. BMC Bioinformatics 7, 3.

Dong, J.-J., Tung, Y.-H., Chen, C.-C., Liao, J.-J., Pan, Y.-W., 2009: Discriminant analysis of the geomorphic characteristics and stability of landslide dams. Geomorphology 110, 162–171.

Drăguţ, L., Eisank, C., 2011: Object representations at multiple scales from digital elevation models. Geomorphology 129, 183–189.

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

Durham University, 2015: Nepal earthquake landslide locations, 30 June 2015 [WWW Document]. Humanit. DATA Exch. URL https://data.humdata.org/dataset/nepal-earthquake-landslide-locations-30-june-2015

ESRI, 2016: Flow Accumulation (Spatial Analyst) [WWW Document]. ArcGIS Resour. Cent. URL http://help.arcgis.com/En/Arcgisdesktop/10.0/Help/index.html#//009z000000 51000000.htm (accessed 1.1.16).

Evans, I.S., Evans, I.S., Lane, S., Lane, S., Richards, K., Richards, K., Chandler, J., Chandler, J., 1998: What do terrain statistics really mean?, in: Landform Monitoring, Modelling and Analysis. pp. 119–138.

Fernández-Delgado, M., Cernadas, E., Barro, S., Amorim, D., Amorim Fernández-Delgado, D., 2014: Do we Need Hundreds of Classifiers to Solve Real World Classification Problems? J. Mach. Learn. Res. 15, 3133–3181.

Freer, J., McDonnell, J.J., Beven, K.J., Peters, N.E., Burns, D.A., Hooper, R.P., Aulenbach, B., Kendall, C., 2002: The role of bedrock topography on subsurface storm flow. Water Resour. Res. 38, 5-1-5–16.

Galli, M., Ardizzone, F., Cardinali, M., Guzzetti, F., Reichenbach, P., 2008: Comparing landslide inventory maps. Geomorphology 94, 268–289.

Geological Survey of Japan, 1995: Strip Map of the Itoigawa-Shizuoka Tectonic Line Active Fault System, 1:100,000 (in Japanese with English abstract).

Germanoski, D., 2001: Relationships between topography, geology, and differential erosion in eastern Pennsylvania, in: GSA Annual Meeting. Boston.

Gerrad, A.J., Gardner, R.A.M., 2000: The role of landsliding in shaping the landscape of the Middle Hills, Nepal. Zeitschrift fur Geomorphol. Suppl. 122, 47–62.

Glade, T., Crozier, M., Smith, P., 2000: Applying Probability Determination to Refine Landslide-triggering Rainfall Thresholds Using an Empirical "Antecedent Daily Rainfall Model." Pure Appl. Geophys. 157, 1059–1079.

Goetz, J.N., Brenning, A., Petschko, H., Leopold, P., 2015: Evaluating machine learning and statistical prediction techniques for landslide susceptibility modeling. Comput. Geosci. 81, 1–11.

Gorsevski, P. V., Gessler, P.E., Foltz, R.B., Elliot, W.J., 2006: Spatial Prediction of Landslide Hazard Using Logistic Regression and ROC Analysis. Trans. GIS 10, 395–415.

Gorum, T., Gonencgil, B., Gokceoglu, C., Nefeslioglu, H.A., 2008: Implementation of reconstructed geomorphologic units in landslide susceptibility mapping: the Melen Gorge (NW Turkey). Nat. Hazards 46, 323–351.

Grelle, G., Revellino, P., Donnarumma, A., Guadagno, F.M., 2011: Bedding control on landslides: A methodological approach for computer-aided mapping analysis. Nat. Hazards Earth Syst. Sci. 11, 1395–1409.

Guarneri, J., 2013: From a Different Angle: Averaging Upslope Aspect [WWW Document]. Trust Me, I'm a Geogr. URL https://gisjay.wordpress.com/2013/03/04/from-a-different-angle-averaging-upslope-aspect/ (accessed 11.17.15).

Günther, A., Reichenbach, P., Malet, J.-P., Van Den Eeckhaut, M., Hervás, J., Dashwood, C., Guzzetti, F., 2013: Tier-based approaches for landslide

susceptibility assessment in Europe. Landslides 10, 529–546.

Guthrie, R.H., 2002: The effects of logging on frequency and distribution of landslides in three watersheds on Vancouver Island, British Columbia. Geomorphology 43, 273–292.

Guzzetti, F., Carrara, A., Cardinali, M., Reichenbach, P., 1999: Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy. Geomorphology 31, 181–216.

Guzzetti, F., Malamud, B.D., Turcotte, D.L., Reichenbach, P., 2002: Power-law correlations of landslide areas in central Italy. Earth Planet. Sci. Lett. 195, 169–183.

Guzzetti, F., Peruccacci, S., Rossi, M., Stark, C.P., 2008: The rainfall intensity-duration control of shallow landslides and debris flows: An update. Landslides 5, 3–17.

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

Harrell, F.E., 2001: Multivariable Modeling Strategies, in: Regression Modeling Strategies, Springer Series in Statistics. Springer New York, pp. 53–85.

Has, B., Noro, T., Maruyama, K., Nakamura, A., Ogawa, K., Onoda, S., 2012: Characteristics of earthquake-induced landslides in a heavy snowfall region-landslides triggered by the northern Nagano prefecture earthquake, March 12, 2011, Japan. Landslides 9, 539–546.

Hasegawa, S., Dahal, R.K., Nishimura, T., Nonomura, A., Yamanaka, M., 2009: DEM-Based analysis of earthquake-induced shallow landslide susceptibility. Geotech. Geol. Eng. 27, 419–430.

Ho, T.K., 1998: The random subspace method for constructing decision forests. IEEE Trans. Pattern Anal. Mach. Intell. 20, 832–844.

Hong, Y., Hiura, H., Shino, K., Sassa, K., Suemine, A., Fukuoka, H., Wang, G., 2005: The influence of intense rainfall on the activity of large-scale crystalline schist landslides in Shikoku Island, Japan. Landslides 2, 97–105.

Hovius, N., Stark, C.P., Allen, P.A., 1997: Sediment flux from a mountain belt derived by landslide mapping. Geology 25, 231.

Huggel, C., Clague, J.J., Korup, O., 2012: Is climate change responsible for changing landslide activity in high mountains? Earth Surf. Process. Landforms 37, 77–91.

Hussin, H.Y., Zumpano, V., Reichenbach, P., Sterlacchini, S., Micu, M., van Westen, C., Bălteanu, D., 2015: Different landslide sampling strategies in a grid-based bi-variate statistical susceptibility model. Geomorphology.

Inoue, K., Mori, T., Mizuyama, T., 2012: Three Large Historical Landslide Dams and Outburst Disasters in the North Fossa Magna Area , Central Japan 5, 145–154.

Iwahashi, J., Pike, R.J., 2007: Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature. Geomorphology 86, 409–440.

Iwahashi, J., Watanabe, S., Furuya, T., 2001: Landform analysis of slope movements using DEM in Higashikubiki area, Japan. Comput. Geosci. 27, 851–865.

Kanungo, D.P., Sarkar, S., Sharma, S., 2011: Combining neural network with fuzzy, certainty factor and likelihood ratio concepts for spatial prediction of landslides. Nat. Hazards 59, 1491–1512.

Kargel, J.S., Leonard, G.J., Shugar, D.H., Haritashya, U.K., Bevington, A., Fielding, E.J., Fujita, K., Geertsema, M., Miles, E.S., Steiner, J., Anderson, E., Bajracharya, S., Bawden, G.W., Breashears, D.F., Byers, A., Collins, B., Dhital, M.R., Donnellan, A., Evans, T.L., Geai, M.L., Glasscoe, M.T., Green, D., Gurung, D.R., Heijenk, R., Hilborn, A., Hudnut, K., Huyck, C., Immerzeel, W.W., Liming, J., Jibson, R., Kaab, A., Khanal, N.R., Kirschbaum, D., Kraaijenbrink, P.D.A., Lamsal, D., Shiyin, L., Mingyang, L., McKinney, D., Nahirnick, N.K., Zhuotong, N., Ojha, S., Olsenholler, J., Painter, T.H., Pleasants, M., Pratima, K.C., Yuan, Q.I., Raup, B.H., Regmi, D., Rounce, D.R., Sakai, A., Donghui, S., Shea, J.M., Shrestha, A.B., Shukla, A., Stumm, D., van der Kooij, M., Voss, K., Xin, W., Weihs, B., Wolfe, D., Lizong, W., Xiaojun, Y., Yoder, M.R., Young, N., 2016: Geomorphic and geologic controls of geohazards induced by Nepals 2015 Gorkha earthquake. Science (80-.). 351, aac8353-aac8353.

Kawabata, D., Bandibas, J., 2009: Landslide susceptibility mapping using geological data, a DEM from ASTER images and an Artificial Neural Network (ANN). Geomorphology 113, 97–109.

Kawagoe, S., Kazama, S., Ranjan Sarukkalige, P., 2009: Assessment of snowmelt triggered landslide hazard and risk in Japan. Cold Reg. Sci. Technol. 58, 120–129.

Keefer, D.K., 1984: Landslides caused by earthquakes. Geol. Soc. Am. Bull.

Keijsers, J.G.S., Schoorl, J.M., Chang, K.T., Chiang, S.H., Claessens, L., Veldkamp, a., 2011: Calibration and resolution effects on model performance for predicting shallow landslide locations in Taiwan. Geomorphology 133, 168–177.

Kitahara, H., Okura, Y., Sammori, T., Kawanami, A., 2000: Application of Universal Soil Loss Equation (USLE) to mountainous forests in Japan. J. For. Res. 5, 231–236.

Lee, J.W., Lee, J.B., Park, M., Song, S.H., 2005: An extensive comparison of recent

classification tools applied to microarray data. Comput. Stat. Data Anal. 48, 869–885.

Lee, S., Choi, J., 2004: Landslide susceptibility mapping using GIS and the weight-of-evidence model. Int. J. Geogr. Inf. Sci. 18, 789–814.

Lee, S., Talib, J.A., 2005: Probabilistic landslide susceptibility and factor effect analysis. Environ. Geol. 47, 982–990.

Liaw, a, Wiener, M., 2002: Classification and Regression by randomForest. R news 2, 18–22.

Lin, Z., Oguchi, T., 2004: Drainage density, slope angle, and relative basin position in Japanese bare lands from high-resolution DEMs. Geomorphology 63, 159–173.

Malamud, B.D., Turcotte, D.L., 1999: Self-organized criticality applied to natural hazards. Nat. Hazards 20, 93–116.

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004a: Landslide inventories and their statistical properties. Earth Surf. Process. Landforms 29, 687–711.

Malamud, B.D., Turcotte, D.L., Guzzetti, F., Reichenbach, P., 2004b: Landslides, earthquakes, and erosion. Earth Planet. Sci. Lett. 229, 45–59.

Mathew, J., Jha, V.K., Rawat, G.S., 2007: Application of binary logistic regression analysis and its validation for landslide susceptibility mapping in part of Garhwal Himalaya, India. Int. J. Remote Sens. 28, 2257–2275.

Matsuda, T., Nakamura, K., Sugimura, A., 1967: Late Cenozoic orogeny in Japan. Tectonophysics 4, 349–366.

Meyer, D., Leisch, F., Hornik, K., 2003: The support vector machine under test. Neurocomputing 55, 169–186.

Mondal, A., Khare, D., Kundu, S., 2016: Uncertainty analysis of soil erosion modelling using different resolution of open-source DEMs. Geocarto Int. 6049, 1–16.

Montgomery, D.R., Dietrich, W.E., 1989: Source areas, drainage density, and channel initiation. Water Resour. Res. 25, 1907–1918.

Moore, I., Gessler, P., Nielsen, G.A., Peterson, G.A., 1993: Soil attribute prediction using terrain analysis. Soil Sci. Soc. Am. J. 57, 443–452.

Moore, I.D., Grayson, R.B., Ladson, A.R., 1991: Digital terrain modelling: A review of hydrological, geomorphological, and biological applications. Hydrol. Process. 5, 3–30.

Nakazato, H., Shoda, D., Inoue, K., Suzuki, H., 2013: A Case Study of Behavior Observation of Landslide Induced by Snowmelt After an Earthquake, in: Ugai,

K., Yagi, H., Wakai, A. (Eds.), Earthquake-Induced Landslides: Proceedings of the International Symposium on Earthquake-Induced Landslides, Kiryu, Japan, 2012. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 341–345.

Nandi, a., Shakoor, a., 2010: A GIS-based landslide susceptibility evaluation using bivariate and multivariate statistical analyses. Eng. Geol. 110, 11–20.

Nefeslioglu, H., Duman, T.Y., Durmaz, S., 2008a: Landslide susceptibility mapping for a part of tectonic Kelkit Valley (Eastern Black Sea region of Turkey). Geomorphology 94, 401–418.

Nefeslioglu, H., Gokceoglu, C., Sonmez, H., 2008b: An assessment on the use of logistic regression and artificial neural networks with different sampling strategies for the preparation of landslide susceptibility maps. Eng. Geol. 97, 171–191.

Ng, K.Y., 2006: Landslide locations and drainage network development: A case study of Hong Kong. Geomorphology 76, 229–239.

Nourani, V., Pradhan, B., Ghaffari, H., Sharifi, S.S., 2013: Landslide susceptibility mapping at Zonouz Plain, Iran using genetic programming and comparison with frequency ratio, logistic regression, and artificial neural network models. Nat. Hazards 71, 523–547.

Oguchi, T., 1997: Drainage density and relative relief in humid steep mountains with frequent slope failure. Earth Surf. Process. landforms 22, 107–120.

Ohlmacher, G.C., 2007: Plan curvature and landslide probability in regions dominated by earth flows and earth slides. Eng. Geol. 91, 117–134.

Ohmori, H., Sugai, T., 1995: Toward geomorphometric models for estimating landslide dynamics and forecasting landslide occurrence in Japanese mountains. Zeitschrift für Geomorphol. 101, 149–164.

Pal, M., 2005: Random forest classifier for remote sensing classification. Int. J. Remote Sens. 26, 217–222.

Park, S., McSweeney, K., Lowery, B., 2001: Identification of the spatial distribution of soils using a process-based terrain characterization. Geoderma 103, 249–272.

Paudel, U., Oguchi, T., Hayakawa, Y., 2016: Multi-Resolution Landslide Susceptibility Analysis Using a DEM and Random Forest. Int. J. Geosci. 7, 726–743.

Pavel, M., Nelson, J.D., Jonathan Fannin, R., 2011: An analysis of landslide susceptibility zonation using a subjective geomorphic mapping and existing landslides. Comput. Geosci. 37, 554–566.

Pérez-Peña, J. V., Azañón, J.M., Booth-Rea, G., Azor, a., Delgado, J., 2009: Differentiating geology and tectonics using a spatial autocorrelation technique for the hypsometric integral. J. Geophys. Res. 114, F02018.

Petley, D., 2011: Global deaths from landslides in 2010 (updated to include a comparison with previous years) [WWW Document]. URL http://blogs.agu.org/landslideblog/2011/02/05/global-deaths-from-landslides-in-2010/ (accessed 5.24.16).

Picanço, J., Pinto, C.A., Mesquita, M.J., Moraes, M., Soares, L.F., Cardoso, F., 2014: Typology of Rainfall-Triggered Landslides in the Urban Area of Antonina, Southern Brazil, in: Sassa, K., Canuti, P., Yin, Y. (Eds.), Landslide Science for a Safer Geoenvironment. Springer International Publishing, Cham, pp. 379–383.

Pike, R.J., Wilson, S.E., 1971: Elevation-relief ratio, hypsometric integral, and geomorphic area-altitude analysis. Bull. Geol. Soc. Am. 82, 1079–1084.

Pourghasemi, H.R., Mohammady, M., Pradhan, B., 2012: Landslide susceptibility mapping using index of entropy and conditional probability models in GIS: Safarood Basin, Iran. Catena 97, 71–84.

Pradhan, B., 2013: A comparative study on the predictive ability of the decision tree, support vector machine and neuro-fuzzy models in landslide susceptibility mapping using GIS. Comput. Geosci. 51, 350–365.

Pradhan, B., Lee, S., 2010: 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.

Prima, O.D.A., Echigo, A., Yokoyama, R., Yoshida, T., 2006: Supervised landform classification of Northeast Honshu from DEM-derived thematic maps. Geomorphology 78, 373–386.

Prima, O.D.A., Yoshida, T., 2010: Characterization of volcanic geomorphology and geology by slope and topographic openness. Geomorphology 118, 22–32.

Prisley, S., 2011: When GIS slope isn't what you think. For. source 16, 11.

Rech, J.A., Reeves, R.W., Hendricks, D.M., 2001: The influence of slope aspect on soil weathering processes in the Springerville volcanic field, Arizona. Catena 43, 49–62.

Regmi, N.R., Giardino, J.R., Vitek, J.D., 2010: Modeling susceptibility to landslides using the weight of evidence approach: Western Colorado, USA. Geomorphology 115, 172–187.

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

Saito, H., Nakayama, D., Matsuyama, H., 2009: Comparison of landslide susceptibility based on a decision-tree model and actual landslide occurrence: The Akaishi Mountains, Japan. Geomorphology 109, 108–121.

Schumm, S.A., 1956: Evolution of drainage systems and slopes in Badlands at Perth Amboy, New Jersey. Geol. Soc. Am. Bull. 67, 597.

Simon, N., Crozier, M., Roiste, M. de, Rafek, A., 2013: Point Based Assessment: Selecting the Best Way to Represent Landslide Polygon as Point Frequency in Landslide Investigation. Electron. J. Geotech. Eng. 18, 775–784.

Staley, D.M., Waskelwicz, T.A., 2013: The Use of Airborne Laser Swath Mapping on Fans and Cones: An Example from the Colorado Front Range, in: Schneuwly-Bollschweiler, M., Stoffel, M., Rudolf-Miklau, F. (Eds.), Dating Torrential Processes on Fans and Cones, Advances in Global Change Research. Springer Netherlands, Dordrecht, pp. 147–164.

Strahler, A.N., 1952: Hypsometric (area-altitude) analysis of erosional topography. Geol. Soc. Am. Bull. 63, 1117.

Suzen, M., Doyuran, V., 2004a: A comparison of the GIS based landslide susceptibility assessment methods: multivariate versus bivariate. Environ. Geol. 45, 665–679.

Suzen, M., Doyuran, V., 2004b: Data driven bivariate landslide susceptibility assessment using geographical information systems: a method and application to Asarsuyu catchment, Turkey. Eng. Geol. 71, 303–321.

Suzuki, S., Ishizuka, H., 1998: Low-grade metamorphism of the Mikabu and northern Chichibu belts in central Shikoku, SW Japan: implications for the areal extent of the Sanbagawa low-grade metamorphism. J. Metamorph. Geol. 16, 107–116.

Svetnik, V., Liaw, A., Tong, C., Christopher Culberson, J., Sheridan, R.P., Feuston, B.P., 2003: Random Forest: A Classification and Regression Tool for Compound Classification and QSAR Modeling. J. Chem. Inf. Comput. Sci. 43, 1947–1958.

Swets, J., 1988: Measuring the accuracy of diagnostic systems. Science (80-.). 240, 1285–1293.

Takeda, T., Sato, H., Iwasaki, T., Matsuta, N., Sakai, S., Iidaka, T., Kato, A., 2004: Crustal structure in the northern Fossa Magna region, central Japan, modeled from refraction/wide-angle reflection data. Earth, Planets Sp. 56, 1293–1299.

Takeuchi, A., 2008: Duplex Stress Regime in the North Fossa Magna, Central Japan. Bull. Earthq. Res. Inst.

Takeuchi, A., 2004: Basement-involved tectonics in North Fossa Magna , central Japan : The significance of the northern Itoigawa-Shizuoka Tectonic Line. Earth Planets Sp. 56, 1261–1269.

Takeuchi, K., Yanagisawa, Y., 2004: 1:50,000 Digital Geological Map of the Uonuma Tegion, Niigata Prefecture (Ver. 1), GSJ Open-file Report 478, Geological Survey of Japan.

Tarboton, D., 1997: A new method for the determination of flow directions and upslope areas in grid digital elevation models. Water Resour. Res. 33, 309.

Tarolli, P., Sofia, G., 2016: Human topographic signatures and derived geomorphic processes across landscapes. Geomorphology 255, 140–161.

Tarolli, P., Tarboton, D., 2006: A new method for determination of most likely landslide initiation points and the evaluation of digital terrain model scale in terrain stability mapping. Hydrol. Earth Syst. Sci. 10, 663–677.

Tien Bui, D., Pradhan, B., Lofman, O., Revhaug, I., 2012: Landslide susceptibility assessment in vietnam using support vector machines, decision tree, and nave bayes models. Math. Probl. Eng. 2012, 1–26.

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

Tucker, G.E., Catani, F., Rinaldo, A., Bras, R.L., 2001: Statistical analysis of drainage density from digital terrain data. Geomorphology 36, 187–202.

Van Asch, T.W.J., Buma, J., Van Beek, L.P.H., 1999: A view on some hydrological triggering systems in landslides. Geomorphology 30, 25–32.

Van Den Eeckhaut, M., Poesen, J., Govers, G., Verstraeten, G., Demoulin, A., 2007: Characteristics of the size distribution of recent and historical landslides in a populated hilly region. Earth Planet. Sci. Lett. 256, 588–603.

Van Den Eeckhaut, M., Vanwalleghem, T., Poesen, J., Govers, G., Verstraeten, G., Vandekerckhove, L., 2006: Prediction of landslide susceptibility using rare events logistic regression: A case-study in the Flemish Ardennes (Belgium). Geomorphology 76, 392–410.

Varnes, D.J., 1984: Landslide hazard zonation: a review of principles and practice, Natural Hazards. UNESCO, Paris.

Varnes, D.J., 1978: Slope movement types and processes, in Schuster, R.L., and Krizek, R.J.,eds., Landslides—Analysis and control. Natl. Acad. Sci. Transp. Res. Board Spec. Rep. 176, 11–33.

Von Ruette, J., Papritz, A., Lehmann, P., Rickli, C., Or, D., 2011: Spatial statistical modeling of shallow landslides—Validating predictions for different landslide inventories and rainfall events. Geomorphology 133, 11–22.

Vorpahl, P., Elsenbeer, H., Märker, M., Schröder, B., 2012: How can statistical models help to determine driving factors of landslides? Ecol. Modell. 239, 27–39.

Wade, A., 1935: The Relationship between Topography and Geology. Aust. Surv. 5, 367–371.

Wang, H.B., Sassa, K., Xu, W.Y., 2007: Analysis of a spatial distribution of landslides triggered by the 2004 Chuetsu earthquakes of Niigata Prefecture, Japan. Nat. Hazards 41, 43–60.

Wang, L.-J., Sawada, K., Moriguchi, S., 2013: Landslide susceptibility analysis with logistic regression model based on FCM sampling strategy. Comput. Geosci. 57, 81–92.

Wang, X., Yin, Z.Y., 1998: A comparison of drainage networks derived from digital elevation models at two scales. J. Hydrol. 210, 221–241.

Wieczorek, G.F., Mandrone, G., DeCola, L., 1997: The Influence of Hillslope Shape on Debris-Flow Initiation, in: Debris-Flow Hazards Mitigation: Mechanics, Prediction, and Assessment. ASCE, pp. 21–31.

Wilson, J.P., Gallant, J.C., 2000: Terrain Analysis: Principles and Applications. John Wiley & Sons.

Wood, W.F., Snell, J.B., 1960: A quantitative system for classifying landforms. Defense Technical Information Center, Ft. Belvoir.

Wu, B., Abbott, T., Fishman, D., McMurray, W., Mor, G., Stone, K., Ward, D., Williams, K., Zhao, H., 2003: Comparison of statistical methods for classification of ovarian cancer using mass spectrometry data. Bioinformatics 19, 1636–1643.

Xu, C., Xu, X., Dai, F., Saraf, A.K., 2012: Comparison of different models for susceptibility mapping of earthquake triggered landslides related with the 2008 Wenchuan earthquake in China. Comput. Geosci. 46, 317–329.

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 of China. Nat. Hazards 68, 883–900.

Xu, C., Xu, X., Yao, X., Dai, F., 2014: Three (nearly) complete inventories of landslides triggered by the May 12, 2008 Wenchuan Mw 7.9 earthquake of China and their spatial distribution statistical analysis. Landslides 11, 441–461.

Yalcin, A., 2008: GIS-based landslide susceptibility mapping using analytical hierarchy process and bivariate statistics in Ardesen (Turkey): comparisons of results and confirmations. Catena 72, 1–12.

Yalcin, A., Reis, S., Aydinoglu, A.C., Yomralioglu, T., 2011: A GIS-based comparative study of frequency ratio, analytical hierarchy process, bivariate statistics and logistics regression methods for landslide susceptibility mapping in Trabzon, NE Turkey. CATENA 85, 274–287.

Yamagishi, H., Ayalew, L., Horimatsu, T., Kanno, T., Hatamoto, M., 2004: Recent landslides in Niigata region, Japan [WWW Document]. Japan Landslide Soc. URL http://www.landslide-soc.org/e-content/Recent_landslides_in_Niigata_region.pdf (accessed 7.8.14).

Yamasaki, S., Chigira, M., 2008: Weathering mechanism of pelitic schist and its relationships with landslides: analysis of undisturbed drilled cores from the Sambagawa Metamorphic belt in Shikoku. J. Geol. Soc. Japan 114, 109–126.

Yeon, Y., Han, J.-G., Ryu, K.H., 2010: Landslide susceptibility mapping in Injae, Korea, using a decision tree. Eng. Geol. 116, 274–283.

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.

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

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

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.

Yin, Y., Wang, F., Sun, P., 2009: Landslide hazards triggered by the 2008 Wenchuan earthquake, Sichuan, China. Landslides 6, 139–152.

Yoshimatsu, H., Abe, S., 2006: A review of landslide hazards in Japan and assessment of their susceptibility using an analytical hierarchic process (AHP) method. Landslides 3, 149–158.

Zare, M., Pourghasemi, H.R., Vafakhah, M., Pradhan, B., 2012: Landslide susceptibility mapping at Vaz Watershed (Iran) using an artificial neural network model: a comparison between multilayer perceptron (MLP) and radial basic function (RBF) algorithms. Arab. J. Geosci. 6, 2873–2888.

Zhang, X., Drake, N.A., Wainwright, J., Mulligan, M., 1999: Comparison of slope estimates from low resolution DEMs: scaling issues and a fractal method for their solution. Earth Surf. Process. Landforms 24, 763–779.

Zhilin, L., 2008: Multi-Scale Digital Terrain Modelling and Analysis, in: Advances in Digital Terrain Analysis. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 59–83.

Zhou, Q., Chen, Y., 2011: Generalization of DEM for terrain analysis using a compound method. ISPRS J. Photogramm. Remote Sens. 66, 38–45.

Appendix

1. Calculation of the density of geological boundaries using Spatial Analyst extension in ArcGIS.

Density of geological boundaries in this study was computed using the line-density function of the Spatial Analyst extension in ArcGIS. The function calculates the density of linear features in the neighborhood of each output raster cell. Density is calculated in units of length per unit area.

Conceptually, a circle is drawn around each raster cell center using the search radius. The length of the portion of each line that falls within the circle is summed and the total length is divided by the circle's area. The figure below illustrates this concept:
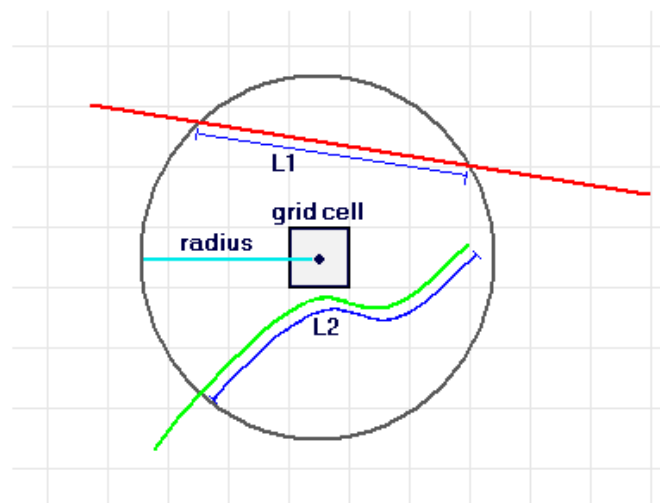

Figure A-1. Conceptual representation of line density (ArcGIS Desktop Help; http://webhelp.esri.com)

In Figure A-1, a raster cell is shown with its circular neighborhood. Lines L1 and L2 represent the length of the portion of each line that falls within the circle. The corresponding population field values are V1 and V2. Thus: (L1×V1 + L2×V2) / (area of circle) = Density. A search radius of 1000 m was used in this study (see **Section 4.2.2**).

2. Calculation of the distance to the geological boundary using the Spatial Analyst extension in ArcGIS.

Distance to the geological boundary in this study was computed using the Euclidean-distance function of the Spatial Analyst extension in ArcGIS. The function produces an output raster that records for each cell the shortest distance to the closest source (geological boundaries in this study).

Euclidean distance is calculated from the center of the source cell to the center of each of the surrounding cells. Conceptually, the Euclidean-distance algorithm works as follows: For each cell, the distance to each source cell is determined by calculating the hypotenuse with *x_max* and *y_max* as the other two legs of the triangle. This calculation derives the true Euclidean distance, rather than the cell distance.

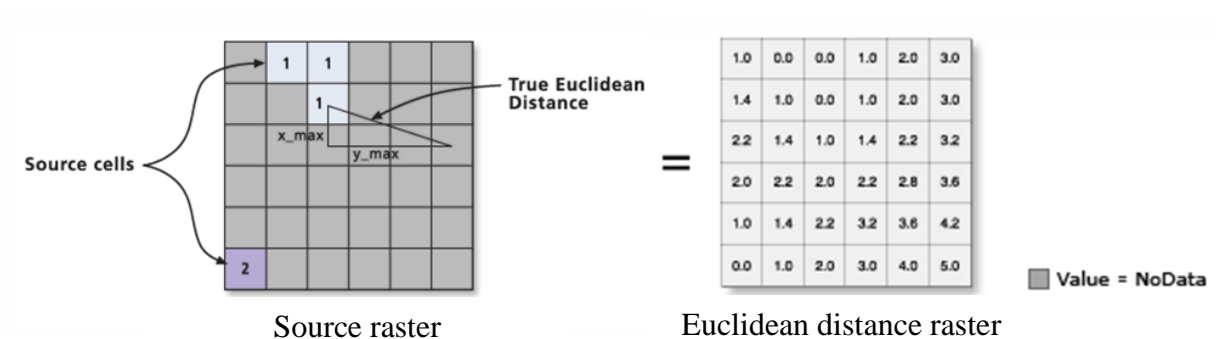Source raster   Euclidean distance raster

Figure A-2. Illustration of Euclidean distance (ArcGIS Desktop Help;
http://webhelp.esri.com)