Fuzzy Logic-based Digital Soil Mapping in the Laurel Creek Conservation Area, Waterloo, Ontario

by

Que Ren

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of

Master of Science
in
Geography

Waterloo, Ontario, Canada, 2012

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

ABSTRACT

The aim of this thesis was to examine environmental covariate-related issues, the resolution dependency, the contribution of vegetation covariates, and the use of LiDAR data, in the purposive sampling design for fuzzy logic-based digital soil mapping. In this design fuzzy c-means (FCM) clustering of environmental covariates was employed to determine proper sampling sites and assist soil survey and inference. Two subsets of the Laurel Creek Conservation area were examined for the purposes of exploring the resolution and vegetation issues, respectively. Both conventional and LiDAR-derived digital elevation models (DEMs) were used to derive terrain covariates, and a vegetation index calculated from remotely sensed data was employed as a vegetation covariate. A basic field survey was conducted in the study area. A validation experiment was performed in another area.

The results show that the choices of optimal numbers of clusters shift with resolution aggregated, which leads to the variations in the optimal partition of environmental covariates space and the purposive sampling design. Combining vegetation covariates with terrain covariates produces different results from the use of only terrain covariates. The level of resolution dependency and the influence of adding vegetation covariates vary with DEM source. This study suggests that DEM resolution, vegetation, and DEM source bear significance to the purposive sampling design for fuzzy logic-based digital soil mapping. The interpretation of fuzzy membership values at sampled sites also indicates the associations between fuzzy clusters and soil series, which lends promise to the applicability of fuzzy logic-based digital soil mapping in areas where fieldwork and data are limited.

ACKNOWLEDGEMENTS

I would first like to thank Dr. Jonathan Li for his willingness to serve as my supervisor. He has been a tremendous advisor during my master study, and I wish him all the best for the future. I would like to thank my committee members, Dr. Su-Yin Tan, Dr. Peter Deadman, and Dr. James Craig at the Department of Civil and Environmental Engineering, who found time in their busy schedules to provide much-appreciated input to this thesis. I owe much gratitude to Dr. Xiaoyuan Geng and Dr. David Kroetsch at Agriculture and Agri-Food Canada for their enduring the rain while helping me with fieldwork. Special thanks go to Dr. Xiaoyuan Geng for providing me with software, published papers, and a huge amount of datasets. I would like to thank Dr. A-Xing Zhu and Dr. Lin Yang at the University of Wisconsin-Madison for their kindness in answering my long questions through emails, and sharing the useful SoLIM software with public. I would also like to thank Dr. Merrin Macrae, Dr. Claude Duguay, and Dr. Su-Yin Tan again for their kind help and tips on my assignments and term papers. Many thanks also go to Eva Dodsworth, Jonathan Morgan, and Steve Xu at the University Map Library, and Mike Lackner at Mapping, Analysis and Design, for providing me with helpful datasets.

Many thanks go to Haiyan Guan for her tireless assistance, and also go to Sabrina Li who taught me a lot on writing in English. There is no possible way to express adequately my gratitude to my family: my husband, my dear father and mother, my parents-in-law, for their support through the years.

献给我亲爱的爸爸妈妈

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1 INTRODUCTION

1.1 Motivation

Activities in pedology, agriculture, natural resource management, environmental monitoring and modeling frequently necessitate detailed soil information (Moore et al., 1993; McBratney et al., 2003). Conventional soil surveys provide soil data in the form of vector-based polygon maps and are usually based on mental models. However these maps have been criticized for neither representing the variability of soil in a continuum nor quantitatively modeling soil (Burrough et al., 1997; McBratney et al., 2003; Scull et al., 2003; Zhu et al., 2008a). In *Factors of Soil Formation*, Jenny (1941) elaborated five soil forming factors: climate (*c*), organisms (*o*), relief (*r*), parent material (*p*), and time (*t*) in order to describe the quantifiable relationship between environment and soil. Jenny's theory was updated by McBratney et al. (2003) by adding two more factors: soil (*s*) as known soil property that can be used as an input, and space (*n*) representing spatial correlation. Recently this relationship between environmental factors and soil variation can be quantified on the use of remotely sensed data and geographic information system (GIS) tools which can extrapolate conventional soil maps and developing soil inference models (McBratney et al., 2003; Nauman, 2009; Geng et al., 2010). This field of work is broadly termed digital soil mapping (Geng et al., 2010), as well as predictive soil mapping, soil inference, or quantitative soil mapping (Scull et al., 2003).

Digital soil mapping can be defined as "the creation and population of spatial soil

information systems by numerical models inferring the spatial and temporal variations of soil types and soil properties from soil observation and knowledge and from related environmental variables" (Lagacherie and McBratney, 2007). A range of environmental covariates have been employed to predict soil types and properties, of which land-surface parameters are key components (McBratney et al., 2003). Land-surface parameters are often derived from digital elevation model (DEM) without further knowledge of the area represented (Pike et al., 2009), appearing in the same structure as their source DEM while presenting values of descriptive measures of the surface at each point. Depending on the derivation process, land-surface parameters are distinguished as primary or secondary (Pike et al., 2009; Wilson, 2011). Primary land-surface parameters (e.g., slope, aspect, and curvature) can be calculated directly from DEMs, while calculating secondary land-surface parameters (e.g., topographic wetness index (TWI), stream power index (SPI), and length-slope (LS) factor) requires additional processes based on primary parameters. Both primary and secondary parameters are useful for characterizing hydrological and geomorphological processes, and may capture the association between relief (*r*) and soil (Moore et al, 1991; Wilson and Gallant, 2000; McBratney et al., 2003).

A number of inferential tools for digital soil mapping are available, such as regression models (e.g., Moore et al., 1993; Odeh et al., 1994; Hengl et al., 2004), tree models (e.g., Bui et al., 1999; Bui and Moran, 2001), fuzzy logic (e.g., Zhu, 1994; Burrough et al., 1997). Under fuzzy logic, soil is considered as a spatial continuum and can be labelled as more than one soil series at a given point (Zhu, 1994). This method has great potential in soil

survey and mapping in terms of the reality that soil often varies continuously (McBratney and Odeh, 1997). It can also respond to the inadequacies surrounding conventional vector-based polygon soil maps by producing raster-based continuous soil maps (English, 2001). For example, since the mapping units are internally uniform with respect to specific soil properties or types in conventional soil maps, many studies have shown that within-unit variance is often unacceptably high; fuzzy soil mapping can overcome this difficulty by producing continuous soil maps (Burrough, 1987; Odeh et al., 1992; Zhu et al., 1997; Zhu et al., 2008a). Among the methods under fuzzy logic, Zhu et al. (2008a) proposed a purposive sampling design for digital soil mapping based on fuzzy clustering. This method can be useful for soil mapping in Canada in that it allows the reduction of the amount of fieldwork where very limited efforts exist (Geng et al., 2010). As well, this method can bypass the acquisition of expert knowledge to some extent (English, 2001). These advantages of this method lend great potential to its application in Canada. Most of the activities in soil survey and mapping in Canada are driven by private sector companies, and government led surveys are very limited due to the lack of funding and surveyors (Anderson and Smith, 2011). However private sector soil survey data does not go into the public realm, so that soil information is still strongly demanded, especially for agricultural lands where soils are managed intensively and evolve over time (Anderson and Smith, 2011). To overcome the difficulties in government led surveys, new techniques and methods in digital soil mapping are needed, especially, the purposive sampling design method by Zhu et al. (2008a) which is expected to greatly reduce the labour efforts.

The method by Zhu et al. (2008a) relies on environmental covariates, such as primary and secondary land-surface parameters, to design sampling sites and to infer soil variation. The source DEM for deriving land-surface parameters naturally plays a vital role in this method. Many subjects of concern in using DEM to derive land-surface parameters have been studied in the field of terrain analysis (Moore et al., 1991; Wilson and Gallant, 2000; Wilson, 2011). One of the key subjects is DEM resolution. In terrain analysis, the effects of DEM resolution on land-surface parameters have been investigated by a number of researchers (e.g., Quinn et al., 1991; Thompson et al., 2001; Wilson, 2011). Some of them claim that high-resolution DEM produces more accurate land-surface parameters while others argue that it is not necessary. Nevertheless, few available in the literature have investigated the effects of DEM resolution on fuzzy soil mapping, particularly, on the purposive sampling design. Since the resolution could have great impact on the quality of fuzzy soil mapping products, studies on examining the resolution issue are desirable (Behrens et al., 2010).

Another interesting issue in the purposive sampling design for fuzzy soil mapping relates to the contribution of environmental covariates other than land-surface parameters, such as vegetation covariates. DEM is still the primary source of environmental covariates in digital soil mapping (Lagacherie 2008). It is often to employ only the relief (r) factor to predict soil under fuzzy logic (Zhu et al., 2008a), while the contribution of vegetation covariates has not been widely explored. Vegetation is a fundamental factor contributing to soil formation and vegetation types are sometimes able to reflect soil texture, e.g., sand and

clay content, pore, and organic matter content (Jenny, 1941). It is not common to find the exactly same soil type both at the bottom of tall trees and low grasses in nature. Adding vegetation covariates in the purposive sampling design is possible to improve the quality of the design and fuzzy soil inference as long as the area under study could show an association between soil and vegetation. One of the most promising vegetation covariates is the Normalized Difference Vegetation Index (NDVI) that has shown some correlation to soil water content, sand and clay content (Sumfleth and Duttmann, 2008). Examining the contribution of vegetation covariates such as NDVI in naturally vegetated areas bears importance to the improvement of fuzzy soil inference.

In addition, a gap in the purposive sampling design for fuzzy soil mapping is that few studies available in the literature have employed LiDAR (light detection and ranging)-derived DEMs. The advantages of using LiDAR data in terrain analysis lie in their high-density sampling, high accuracy, and the possibility of generating a set of surface models from them (Wilson, 2011). At the much smaller resolution of LiDAR-derived DEM (e.g., <1m), the detailed terrain characteristics would be detected (Vaze et al., 2010). In Nelson et al. (2009), the accuracy of LiDAR data is approximately described as 0.15-1 m vertical accuracy and 1 m horizontal accuracy, higher than most of the other DEM sources. Some laser scanning systems also have the ability to provide the vegetation canopy and ground surface respectively, which may help with environmental modeling of heavily vegetated areas (Wilson, 2011). However, few in digital soil mapping have used LiDAR-derived DEMs, partly due to the difficulties in access to LiDAR data and in

detecting and correcting errors in them (Wilson, 2011). Involving LiDAR data in fuzzy soil inference remains to be a challenge. Particularly, it is necessary to understand the dependency of the purposive sampling design to the resolution of LiDAR-derived DEMs. The introduction of this new DEM source may have impact on terrain-based soil prediction, as well may challenge the available algorithms and experiences in digital soil mapping. With respect to the emerging popularity of LiDAR data (Nelson et al., 2009), addressing the topic of using LiDAR data in fuzzy soil inference will benefit the development of digital soil mapping.

1.2 Research Objectives and Questions

The aim of this study was to examine the environmental covariate-related issues in the purposive sampling design for fuzzy logic-based digital soil mapping, and to investigate if and to what degree the fuzzy clustering of environmental covariates can be employed to assist soil mapping. The objectives of this study can be summarized as follows:

1) To examine the sensitivity of the purposive sampling design for fuzzy logic-based digital soil mapping to the resolution of DEM used in constructing environmental covariates database for fuzzy clustering;
2) To explore the effect of adding a vegetation index, NDVI, into environmental covariates database on the purposive sampling design for fuzzy logic-based digital soil mapping;
3) To analyze if the fuzzy clustering of environmental covariates can support digital soil

mapping by examining the associations between soil series and fuzzy clusters.

The Laurel Creek Conservation area, located in the northwest corner of Waterloo, Ontario, was chosen as the study area. Fuzzy c-means (FCM) clustering of environmental covariates was performed at two subset units of the study area. The aim of FCM clustering is to find the optimal partitions of environmental covariate space. The clustering provides the optimal numbers of clusters and fuzzy membership value maps which could be used to determine optimal and limited field sampling sites. One subset area, Unit 1, was used to examine the effects of DEM resolution by resampling a 10 m conventional DEM into 20 m, 30 m, and 50 m, and a 1 m LiDAR-derived DEM into 5 m, 10 m, 20 m, 30 m, and 50 m. Another subset area, Unit 2, was applied to examine the contribution of vegetation covariates through the use of a vegetation index, NDVI, based on both conventional and LiDAR-derived DEMs at 10 m resolution. A basic field sampling survey was conducted in the study area assisted by soil scientists in order to analyze the associations between fuzzy clusters and soil series. FCM clustering based on terrain covariates at different resolutions was also performed in a subset area of the Waterloo Aquifer area for the purpose of validation.

The methodological framework introduced in this study may help to answer the following questions:

1) How do the purposive sampling design results vary with the resolution of conventional DEM?
2) How does the incorporation of NDVI into an environmental covariate database

influence the purposive sampling design results based on conventional DEM?

3) Does an association between fuzzy clusters and soil series exist?

4) What are the answers of above three questions when using LiDAR-derived DEM?

1.3 Thesis Structure

This thesis contains five chapters with this being the first.

1) Chapter 1 briefly introduces the research motivation followed by addressing the research objectives and questions. Finally the architecture of the thesis is given.

2) Chapter 2 reviews related literature on soil surveys in Canada, FCM clustering in digital soil mapping, and the environmental covariates-related issues: DEM resolution, vegetation covariates, and using LiDAR data.

3) Chapter 3 details the specific study area, a basic field sampling, environmental covariate data collection, and the methodology employed to design sampling sites based on FCM clustering.

4) Chapter 4 presents and discusses the FCM clustering results in two subset units addressing on issues of resolution, vegetation, and use of LiDAR data. Following is an interpretation of fuzzy membership values at sampling sites. Finally a validation experiment and its results are provided.

5) Chapter 5 summarizes the conclusions, limitations, and recommendations for future studies.

CHAPTER 2 LITERATURE REVIEW

This chapter first reviews a history of soil surveys in Ontario. Second is the state-of-the-art of soil survey in Canada, focusing on the fuzzy logic-based digital soil mapping method. Third, it outlines the theoretical basis of FCM clustering as applied in digital soil mapping. The major theme of the fourth section is to provide a critical review on environment covariate-related issues, resolution, vegetation, and the use of LiDAR data, in the purposive sampling design for fuzzy logic-based digital soil mapping.

2.1 A History of Soil Survey in Ontario

Soil surveying in Canada has a long history. The first soil survey was done by A.L. Galbraith in Ontario in 1914, with the expert advice provided by A. N. Coffey of the U.S. Bureau of Soils (Anderson and Smith, 2011). Coffey treated soils as geographic bodies that could be recognized as natural units (Anderson and Smith, 2011). Many efforts were made to identify these soil bodies (Coen, 1987; Geng et al., 2010).

The period from 1920 to early 1930s was the beginning of systematic soil surveys in Canada. The initial soil survey in Ontario during this period aimed to identify and map soil type for agricultural development use (McKeague and Stobbe, 1978). The classification system of the U.S. Bureau of soil was used, including 3 levels: soil province, soil series, and soil type (McKeague and Stobbe, 1978). Soil province was based on general surficial geology features; soil series was analogous to a geological formation involving color, origin of material, and weathering; soil type was then based on textural divisions within each

series. The scope of soil series was broader than this term used now, so only nine series were identified in the study area in this initial soil survey (McKeague and Stobbe, 1978). In this period soil maps without reports in many counties in Ontario were published which provided information about soil series and type, area, color, texture, relief, drainage, reaction, land use, and fertility needs (McKeague and Stobbe, 1978).

The following period from 1934 to 1944 was a cooperative federal-provincial soil survey period. In Ontario soil survey was based on a cooperative program conducted by federal and provincial Departments of Agriculture, and also Ontario Agricultural College at Guelph (McKeague and Stobbe, 1978). During this period, soil surveys were completed or stated in about 10 counties in southern Ontario. A numerical system for mapping soils was developed by G. A. Hills and symbols indicating texture were also added (McKeague and Stobbe, 1978).

Soil survey during the period from 1945 to 1970 was influenced by the National Soil Survey Committee (NSSC). A series of meetings of NSSC aimed to define a consistent, national classification system for soil survey and mapping in Canada. The first Canadian taxonomic system of soil classification was outlined during these meetings, which is the basis of the Canadian soil system used today. It had six levels: order, great group, subgroup, family, series, and type (McKeague and Stobbe, 1978). The soil survey and mapping had become more standardized since then. The Canada Land Inventory (CLI) program stated in 1963 resulted in a set of interpretive soil maps of capability for agriculture. In Ontario, many counties' soil surveys were completed during this period and used soil series as

mapping units, for the completion of the CLI mapping. A detailed soil survey of Waterloo County with reports was also finished during this period, which is still being used now.

During the period from 1970 to 1979, the re-named Canada Soil Survey Committee (CSSC) met several times, promoting the developments of the first edition of the Canadian System for Soil Classification and the Canada Soil Information System (CanSIS). Beginning in 1972, the first GIS in the world was developed in Canada for the purpose of storing soil maps and data electronically (Schut et al., 2011).

In the glory years of soil survey in Canada from 1950s to 1990s, activities in Ontario was among the most important surveys, where different physiographic and climatic regions provided a good opportunity for Canadian soil surveyors to investigate the soil types (McKeague and Stobbe, 1978). A series of national meetings promoted the establishment of the Canadian Soil Classification System which was widely used in soil surveys in Ontario. Soil surveyors started by preparing their own base maps and then moved to learn the skills of photo interpretation to predict soil. The legacy of soil survey for Ontario included a series of soil maps and reports in many counties, among which Waterloo County was mapped in detail.

2.2 A State-of-the-Art of Soil Survey in Canada

Field surveys led by government agencies reached their peak around 1970s and 1980s (Schut et al., 2011). However, a decline in field surveys and experienced field surveyors in government agencies has been seen between 1990s and 2010s (Anderson and Smith, 2011),

which has risen the challenge to satisfy the public need of soil information. During the recent era since 1995 till 2010, the work of pedologists has moved to the continuing upgrading and refinement of existing soil data, such as the development of the Soil Landscapes of Canada (SLC) framework, and the product of new “seamless” provincial coverages from existing information (Anderson and Smith, 2011). Few efforts exist in soil survey operations in provincial level, some of which are also sporadic and site-specific, such as the ongoing efforts to update existing soil information in Ontario (Anderson and Smith, 2011). On the other hand, the rise of private sector soil survey is evident, mainly driven by regulatory environment surrounding industrial developments, such as oil and gas pipelines and mine sites. However most of the private sector survey work does not go into public use and not well match the structure of government-developed soil information databases, such as CanSIS. Thus, national soil information databases are still challenged by the aging change of soil.

The emerging digital soil mapping provides an innovative and proactive way for soil mapping in Canada (Anderson and Smith, 2011). With assistance from the legacy of conventional soil maps and the advent of large amount of remotely sensed data and GIS tools, digital soil mapping may yield promising results for the public use of soil information in Canada (Geng et al., 2010). Geng et al. (2010) assessed the adequacy and usability of the legacy soil information for the applications of digital soil mapping in the near future in Canada. Their study revealed many challenges for incorporating soil legacy data into new digital soil mapping approaches, such as the various formats and scales. The methods of

extracting expert knowledge embedded in the existing soil data were also explored in their study. For example, in the National Soil Database (NSDB) within CanSIS, there could be a way to recover the geospatial linkage of soil classes and properties on the use of GIS to extract soil-landscape models for the purpose of digital soil mapping in Canada.

Digital soil mapping methods have not been widely used in soil survey in Canada. Grunwald (2009) summarized recent related studies in Canada. For example, MacMillan et al. (2007) proposed an automated approach to predictive ecological mapping that combined terrain derivatives, fuzzy and hard logic in a forest area of British Columbia. Liu et al. (2008) evaluated soil drainage mapping using remotely sensed data, a DEM, and apparent soil electrical conductivity for fields in a farm south of Ottawa, Ontario. Mabit et al. (2008) examined soil erosion and soil organic matter content using geostatistics tools for a location at Boyer River watershed, Quebec. Reynolds et al. (2008) found that structural regression was useful for determining soil physical quality indicators in Essex County, Ontario. These studies show that the application of digital soil mapping in Canada is encouraging. A range of soil inference tools and predictors have been examined in these studies. However, among many inference models, fuzzy logic has not been fully explored in these studies. Fuzzy logic-based digital soil mapping method can be very effective when dealing with the uncertainty in soil map and can reduce fieldwork that costs a large amount of money and labour efforts. Thus soil survey in Canada may take advantage of this method. Due to the limited experiences in using this method in Canada, studies on its related issues are meaningful and strongly needed.

2.3 Fuzzy C-Means Clustering in Digital Soil Mapping

In soil science, fuzzy logic is mainly used for soil classification in two perspectives: first, directly FCM clustering of observations in multivariate space into natural groups without an *a priori* number of clusters, and second, applying Semantic Import (SI) model based on imposed definitions and expert knowledge (McBratney and Odeh, 1997). In the first realm, the multivariate space usually refers to the multi-dimensional environmental covariate space, such as land-surface parameters. Examples in this realm include Odeh et al. (1992), Powell et al. (1991), McBratney and De Gruijter (1992), McBratney et al. (1992), Triantafilis and McBratney (1993), Zhu (1994), Burrough et al. (1997), De Gruijter et al. (1997), Lagacherie et al. (1997), Irvin et al. (1997), Burrough et al. (2000), Zhu et al. (2008a), Smith et al. (2010), Liu et al. (2011), and Yang et al. (2011). Burrough et al. (2000) conducted landform classification using FCM clustering in a farmland in Alberta. Smith et al. (2010) reported digital soil mapping at multiple scales using soil inference model based on FCM clustering of terrain covariates in British Columbia. Liu et al. (2011) used FCM clustering of terrain covariates along with existing soil maps to produce more accurate soil attribute information in a low relief area in south-central Manitoba. Yang et al. (2011) updated conventional soil maps based on FCM clustering of terrain covariates for an area in Wakefield, New Brunswick. These case studies have shown a positive sign for soil information users in Canada that the direct use of FCM clustering for digital soil mapping enables the production of continuous soil spatial information maps and overcomes the deficiency in fieldwork efforts and expert knowledge in soil survey and mapping. Although

many studies have applied fuzzy logic in the second realm that uses Semantic Import model, its application in Canada would be limited by its demand for the imposed definitions and expert knowledge. Therefore, directly using FCM clustering of environmental covariates in Canada is easier.

The commonly used FCM clustering method in those studies was introduced by Dunn (1974) and Bezdek (1974), and originates from the theory of *fuzzy sets* by Zadeh (1965). Fuzzy logic has been applied in various fields such as image processing, control theory, and artificial intelligence to complement traditional crisp logic. In real world, soil is physically continuous (Fridland, 1974) and therefore does not typically have sharp boundaries like those defined in conventional soil survey based on crisp logic. In other words, soil classification has to deal with uncertainty. Fuzzy logic is certainly effective when dealing with such uncertainty. In FCM clustering, observations in multivariate space are partitioned into relatively stable naturally occurring but continuous classes, and these observations are assigned continuous class membership values ranging from 0 to 1 (McBratney and Odeh, 1997). A membership value of 1 denotes observations that exactly match the class centroid, while observations that do not match the centroid receive membership values dependent on their degree of closeness to the centroid. FCM clustering allows an observation to belong to two or more clusters.

FCM clustering is based on minimization of the following objective function (Bezdek, 1981):

$$J_m\ (\ U\ ,v\ ) = \sum_{k=1}^{n}\ \sum_{i=1}^{c}\ \left(u_{ik}\right)^m \parallel\ y_k - \ v_i\ \parallel^2 \qquad (2.1)$$

where,

$Y = \{ y_1, y_2, \ldots\ldots, y_n \} \subset \mathbf{R}^n$ = the data,

c = number of clusters in Y; $2 \leq c \leq n$,

m = weighting exponent; $1 \leq m < \infty$,

U = fuzzy c-partition of Y,

$v = (v_1, v_2, \ldots\ldots, v_c)$ = vector of cluster centroids,

u_{ik} = membership of the kth object (y_k) belonging to the ith cluster, [0, 1],

J_m = squared error clustering criterion with fuzzy partition U of y at the centroids v.

J_m is a weighted measure of the squared distance between observations and cluster centroids. As J_m decreases, the clustering improves as observations tend to be overall closer to their representative centroids. The squared distance between kth object (y_k) to the centroid of cluster v_i is any inner product induced norm:

$$(d_{ik})^2 = \| y_k - v_i \|^2 \qquad (2.2)$$

There are infinitely many norms available, among which the Euclidean, Diagonal, and Mahalanobis norms enjoy most widespread use (Bezdek et al., 1984; English, 2001). For most geographic and geologic applications, the Euclidean norm is considered the most appropriate due to its identification of hyperspherical clusters (Bezdek et al., 1984; English, 2001). A detailed discussion of the different norms can be found in Bezdek (1981). In fuzzy logic-based digital soil mapping, the Euclidean norm is recommended (English, 2001).

The fuzzy partitioning is carried out through an iterative minimization of the objective function (2.1) with the update of membership u_{ik} and the cluster centroids v_i by (Bezdek

et al., 1984):

$$u_{ik} = 1 \Big/ \left[\sum_{j=1}^{C} \left(\frac{d_{ik}}{d_{jk}} \right)^{2/m-1} \right] \tag{2.3}$$

$$v_i = \sum_{k=1}^{n} (u_{ik})^m \, y_k \Big/ \sum_{k=1}^{n} (u_{ik})^m \tag{2.4}$$

which are based on the solutions of the minimization of J_m. This iteration will stop when changes between successive iterations are below a termination criterion ε, the maximum membership error, ranging from 0 and 1, which is expressed as:

$$max_{ik}\{|u_{ik}^{j+1} - u_{ik}^{j}|\} < \varepsilon \tag{2.5}$$

where, *j* is the iteration number.

This procedure converges to a local minimum or a saddle point of J_m (Bezdek et al., 1984).

In FCM clustering, the weighting exponent *m* controls the fuzziness degree. When *m* is 1, FCM clustering in fact becomes traditional hard clustering. Increasing *m* allows the partition blurring to the fuzziest state. There is no theoretical or computational approach to distinguish an optimal *m*, and the only way is to determine it experimentally (English, 2001). Another key issue of using FCM clustering is the number of clusters *c* is not known *a priori*. The determination of *c* is termed "cluster validity" (Bezdek, 1981). The most popular cluster validity indices are the partition coefficient (*F*) and entropy (*H*) which are defined as (Bezdek, 1981):

$$F(u) = \sum_{k=1}^{n} \sum_{i=1}^{c} \left(u_{ik} \right)^2 / n \tag{2.6}$$

$$H(u) = -\sum_{k=1}^{n}\sum_{i=1}^{c}\left(u_{ik}\ \log_a\left(u_{ik}\right)\right)/n \qquad (2.7)$$

where, logarithm base $a \in (1,\infty)$,

$$\frac{1}{c} \leq F \leq 1,$$

$$0 \leq H \leq \log_a(c).$$

F measures the average relative amount of overlap between clusters. *H* is a scalar measure of the amount of fuzziness (Bezdek, 1981). The optimal partition realizes the highest *F* and the lowest *H*. Note that *F* and *H* can reach their maximum and minimum values at the same number of clusters, respectively. The limitation of using *F* and *H* as cluster validity can be attributed to their monotonicity with the number of clusters *c* (Wang and Zhang, 2007). *F* decreases and *H* increases with the increase of the number of clusters monotonously, which leads to the selection of the optimal number of clusters is always at *c* = 2, the smallest *c*. Thus, an effective way is to examine the improvement in *F* or *H* over adjacent clusters (English, 2001; Zhu et al., 2008a; Yang et al., 2011). An optimal number of clusters *c* will be found where a significant improvement in *F* or *H* exists.

The FCM clustering technique provides an opportunity for soil classification in a continuous form that allows partial overlap of classes in attribute space and description of the gradual change of soil properties (Burrough et al., 1997; Zhu, 2006). To illustrate the use of FCM clustering, an example is given in Figure 2.1. The data *y* in Equation (2.1) is the environmental covariates vector (e1, e2,…, ep) at Point (x, y), and FCM clustering of environmental covariates is conducted in p-dimensional space. The algorithm is composed of the following steps:

First, define ε, fix m and c, initialize partition matrix $U^{(0)}$;

Second, calculate the centroids V by Equation (2.4);

Third, update membership matrix U by Equation (2.3);

Forth, stop when Equation (2.5) is met, otherwise return to the second step.

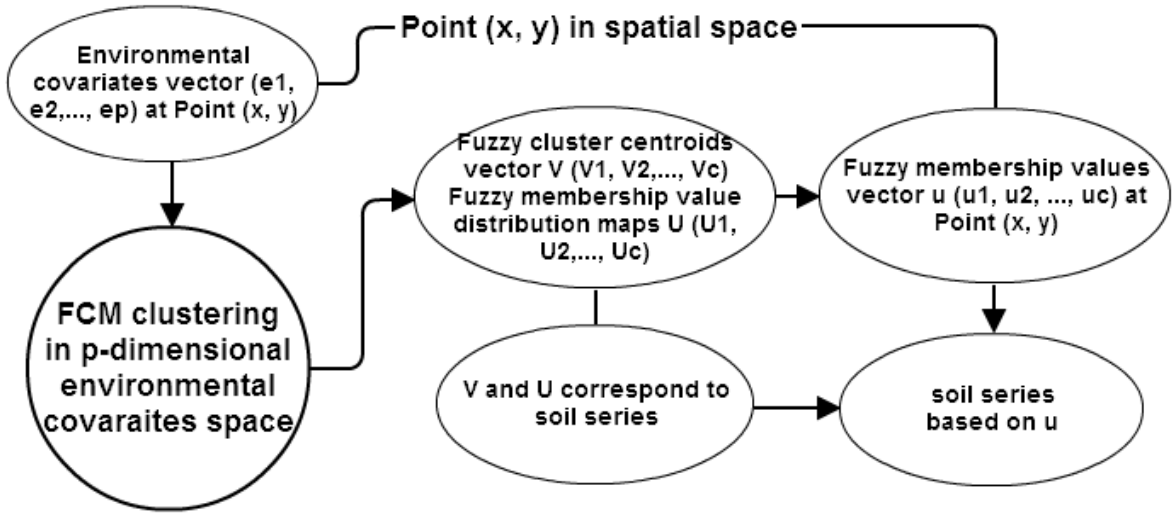

Figure 2.1 An example of the use of FCM clustering in fuzzy soil inference.

An optimal partition can be determined by examining the improvements in partition coefficient and entropy, in which the number of clusters is denoted as c in this example. Centroids of all the c clusters are recorded in a vector V (V1, V2, …, Vc) in Figure 2.1. Fuzzy membership values to a cluster are represented in a raster map in which each cell assigned with its membership value to this cluster. All the c clusters have such raster maps which are represented as a vector U (U1. U2, …, Uc) in Figure 2.1. The soil series associates with the centroids and membership values. Then, the vector u (u1, u2, …, uc) recording membership value to each cluster at Point (x, y) is available to predict the soil series at this point. Since the membership values to clusters are continuous, the assigned

soil series at Point (x, y) will also be continuous and overlapped. For example, soil at this point can be soil series A with membership value of 0.2, soil series B with membership value of 0.7, and soil series C with membership value of 0.1. Note that the sum of membership values to all clusters at a given point is always 1. Thus, the soil at this point is in a continuous form and overlapped among different types.

Among many applications of FCM clustering for digital soil mapping, the promising purposive sampling design for fuzzy soil mapping method (Zhu et al., 2008a) uses the results of FCM clustering of environmental covariates to design sampling sites and perform fuzzy soil inference. This method groups the observations in environmental covariates space into classes on the use of FCM clustering. Continuous fuzzy membership values to these classes are assigned to the observations. Each class could represent a unique environmental configuration. Many attempts suggest unique environmental configurations should beget distinctive soil types or properties (Jenny, 1941; Hudson, 1992; McSweeney et al., 1994; Moran and Bui, 2002; McBratney et al., 2003; Qi and Zhu, 2003; Zhu et al., 2008a), which is the underlying hypothesis of the purposive sampling design method. Based on the association between environmental configurations and soil, the fuzzy clusters can be deemed as corresponding to soil types or properties. Observations with highest fuzzy membership values to a specific class are regarded as typical instances which can represent the corresponding unique environmental configuration of this class. These observations are designed as the purposive sampling sites in field. The identified soil types or properties at these sites then are assigned to this class. Soil variation is inferred from these sites to the

whole area under study by the continuous fuzzy membership values and expert knowledge if available.

This purposive sampling design is a preliminary result of the digital soil mapping but bears great importance due to its guidance of fieldwork. An adequate soil sampling strategy can improve the quality of soil survey and mapping (Brady and Weil, 2000). Among many soil sampling strategies, such as random sampling, grid sampling, and zone sampling (Dinkins and Jones, 2008), the purposive sampling design could be outstanding because of its efficiency (Zhu et al., 2008a). This design aims at reducing field sampling sites and therefore reducing fieldwork to save money and time. Only the typical instances for each class are sampled, which may be less than those in other sampling strategies (Zhu et al., 2008a). In addition, the sampled sites are quantitative-based and more reasonable than those determined by the mental models in conventional soil survey (Zhu et al., 2008a). The determination of sampling sites is purposive and efficient. Moreover, to some extent soil expert knowledge can be quantified by fuzzy membership values and thus is not needed as strongly as that in conventional soil survey. These characteristics of this method allow its application in areas with limited resources. Since soil survey in Canada is lacking experienced soil surveyors and funding (Geng et al., 2010), use of this purposive sampling method is appealing. More studies on its related issues could greatly contribute to soil survey and mapping in Canada.

2.4 Environmental Covariate Issues

Understanding issues related to environmental covariates used in the purposive sampling design for fuzzy soil inference is capable of improving soil surveying and mapping. Issues such as the effects of DEM resolution, the contribution of vegetation covariates, and the use of LiDAR data are all worthy of investigation. A deep understanding of these issues will allow a good construction of environmental covariates database and a good design of sampling sites, and eventually should improve soil survey and mapping products.

2.4.1 DEM Resolution

The most commonly used environmental covariates are land-surface parameters. A range of factors have influences on land-surface parameters, such as the deriving algorithms and resolutions of their source DEMs (Moore et al., 1991; Schmidt and Dikau, 1999; Wilson et al., 2000; Wilson, 2011). Land-surface parameters are heavily dependent on resolution and scale at which they are calculated (Gallant et al., 2000). A variety of DEM products are available for deriving land-surface parameters ranging from low resolution to high resolution (Maune, 2007). Many have examined the effects of DEM resolution on land-surface parameters (e.g., Chang and Tsai, 1991; Wolock and Price, 1994; Gao, 1997; Chaplot et al., 2000; Florinsky and Kuryakova, 2000; Schoorl et al., 2000; Wilson et al., 2000; Thompson et al., 2001; Deng et al., 2007; Wu et al., 2008; Behrens et al., 2010; Vaze et al., 2010; Wilson, 2011). Some argue that a high resolution DEM is seen to produce more

detailed land-surface parameters that can better represent landform characteristics and assist in terrain analysis, while this argument is not necessary in other studies.

Although many efforts have been made on resolution issue with respect to land-surface parameters, few have examined the effects of resolution on further deliverables in digital soil mapping, e.g., the purposive sampling design results. Previous studies employing fuzzy logic in soil mapping mainly worked only at the resolution of their original environmental covariates, ignoring the uncertainty associated with the change of resolution (Arrell et al., 2007). Few attempts have been made to bridge this gap in the literature. Smith et al. (2006) and Zhu et al. (2008b) examined the combined effect of DEM resolution and neighborhood size on soil prediction under fuzzy logic. They concluded that DEM resolution may not be as important as neighborhood size at least for knowledge-based soil mapping, and that a high resolution DEM may not always produce high-quality digital soil mapping products. Nevertheless, with respect to the purposive sampling design for fuzzy soil mapping, previous studies have barely addressed the resolution issue. The successful application of this design in soil survey and mapping needs a solid understanding of its resolution dependency (Zhu et al., 2008b). The spatial resolution at which environmental covariates are derived may deviate from the spatial scale at which a set of soil forming processes occur in the specific area (Geng et al., 2012). How the design varies with changing in resolution remains to be answered, which will benefit a wise determination of the resolution to derive environmental covariates for designing sampling sites and also for inferring soil (Smith et al., 2006). Recently, more DEM products at high resolution are becoming

available for areas in Canada, such as the LiDAR-derived DEM in Prince Edward Island (Geng et al., 2012) and the South Western Ontario Orthophotography Project (SWOOP) 2010 (http://www.swoop2010.ca/). However, whether higher resolution necessarily promises better sampling design and soil inference products demands careful examinations. The resolution dependency issue is worthy of exploration.

2.4.2 Vegetation Covariates

The soil forming factors presented in Jenny (1941) contain two main altering organisms (*o*): vegetation and humans (McBratney et al., 2003). Vegetation information (e.g., indices, biomass) can be effectively obtained from remotely sensed data which are widely available (McBratney et al., 2003). Combining vegetation information with land-surface parameters has the potential to enhance soil prediction (Campling et al., 2002; Sommer et al., 2003; Sumfleth and Duttmann, 2008). For example, Dobos et al. (2000) found using indices calculated by satellite radiometric data, namely, NDVI, combined with DEM derivatives was among the most promising tools for soil survey. In Liu et al. (2011), NDVI was used to stratify the study area into different units to control influencing factors on soil formation.

However, in most previous studies on fuzzy soil inference, only land-surface parameters were employed as environmental covariates (Zhu et al., 1997; Burough et al., 2000; English, 2001; Zhu et al., 2008a; Yang et al., 2011). For example, Zhu et al. (2008a) used elevation, slope, plan curvature, profile curvature, and topographic wetness index. The

missing of vegetation as an environmental covariate may be explained as the studied areas are always controlled mainly by human cultivation where vegetation enjoys similarity. The purposive sampling design for areas occupied by natural vegetation, however, has not been well investigated. In such areas, vegetation covariates have great potential to improve soil prediction due to the possible association between soil and vegetation (Dobos et al., 2000; McBratney et al., 2003). Soil survey and mapping in naturally vegetated areas can benefit from the examination on the influence of adding vegetation covariates such as NDVI into environmental covariate analysis. For example, a global interest in forest motivates the gathering of soil information in forest areas where vegetation canopy may be an effective indicator (Owens et al., 1999). It is much likely that vegetation covariates can improve the quality of the sampling design in terms of the association between soil and vegetation, while investigations are still needed to verify this improvement. Therefore, it is necessary to bridge the gap that vegetation covariates are missing in constructing environmental covariates database for the purposive sampling design.

One of the most promising vegetation covariates is NDVI which has been shown to correlate well with soil water content, sand, and clay (Sumfleth and Duttmann, 2008). It is true that many new vegetation indices based on remotely sensed data have been developed, e.g., soil-adjusted vegetation index (SAVI), transformed SAVI (TSAVI), modified SAVI (MSAVI), and global environment monitoring index (GEMI) (Mulder et al., 2011). However, NDVI still dominates the practical applications due to its reliability, and the complexity and the need of additional information in calculating other new indices

(Rondeaux et al., 1996). In addition, the remotely sensed data for calculating NDVI are available almost everywhere and free of charge (Mulder et al., 2011). Thus, it is meaningful to evaluate the involvement of the commonly available index, NDVI, in the purposive sampling design for fuzzy logic-based digital soil mapping, so that many real-world applications may benefit for this evaluation.

2.4.3 Use of LiDAR-derived DEM

DEM can be created from a number of data sources, e.g., topographic maps or aerial photographs, ground-based survey data, interferometric synthetic aperture radar (InSAR) data, and airborne LiDAR data (Maune, 2007). Among these data sources, LiDAR data has become a very attractive source of terrain data partly because of its high accuracy and high cost-effectiveness (Pfeifer and Mandlburger, 2009). Producing DEM from LiDAR data now dominates local and regional projects everywhere and is likely to be the method of the future (Nelson et al., 2009). For example, Belgium and the Netherlands now have national LiDAR-derived digital surface models (DSM) at 2-5 m resolution (Wilson, 2011). Nevertheless, most studies on LiDAR data used in terrain analysis have only focused on comparison between LiDAR-derived land-surface parameters and those derived from conventional DEM (e.g., MacMillan et al, 2003; Murphy et al., 2009; Vaze et al., 2010). Experiences of using LiDAR data in digital soil mapping are very limited (Wilson, 2011). Lagacherie (2008) claimed that high-resolution DEM, such as LiDAR-derived DEM, can boost digital soil mapping in the near future, and thus studies on examining LiDAR data

used in digital soil mapping, particularly in fuzzy soil mapping, are immediately needed.

Using LiDAR-derived DEM in digital soil mapping is not without problems. LiDAR data can produce very dense and detailed DEM that may result in difficulties in handling with a large data volume (Nelson et al., 2009). Although the level of detail in LiDAR-derived DEM is much higher than that in conventional DEM, the cost of improved accuracy is the introduction of excessive details which are difficult to characterize by conventional methods (Roecher and Thompson, 2010). Previous studies have also argued that the accuracy of LiDAR-derived DEM varies with many factors, such as land cover and landform characteristics (Hodgson et al., 2005). High resolution and accuracy LiDAR-derived DEM may not necessarily promise high quality soil inference products. Nevertheless, to use LiDAR data in terrain analysis and digital soil mapping is definitely a future trend due to the more widely development in collecting LiDAR data (Nelson et al., 2009). Thus, attempts on employing LiDAR-derived DEM in soil inference are necessary. Given the potential of widely using LiDAR data in Canada in the near future (Geng et al., 2012), it is of great meaning to examine issues in its use in digital soil mapping. Fuzzy soil mapping based on LiDAR data can greatly benefit from the examination on the dependency of the purposive sampling design to the resolution of LiDAR-derived DEM. (Zhu et al., 2008b).

2.5 Chapter Summary

This chapter reviews soil surveying in Ontario, the state-of-the-art of soil survey in

Canada, and the principles of FCM clustering with an emphasis on the purposive sampling design method for fuzzy soil inference. Several issues in this method are reviewed: resolution, vegetation covariates, and the use of LiDAR data. Although many have investigated the influence of DEM resolution in terrain analysis, few of them have indicated its influence on the purposive sampling design results. Studies on the resolution dependency of the purposive sampling design are needed. In addition, previous studies on this method have utilized only terrain covariates, while vegetation covariates have not been well investigated. It is necessary to examine the contribution of vegetation covariates, such as NDVI, to this design. Moreover, with the popularity of LiDAR data, experiences in using LiDAR-derived DEM in the purposive sampling design for fuzzy soil inference are also needed. The challenges remaining among these issues trigger the motivation and objectives of this study.

CHAPTER 3 METHODOLOGY

This chapter describes the study area in the first section and introduces a field sampling survey in the following section. The last section details the environmental covariate data collection and the methodology employed to determine the optimal numbers of clusters and purposive sampling sites.

3.1 Study Area

The study area was the Laurel Creek Conservation area, located in the northwest corner of the City of Waterloo, Ontario. This area was chosen because, first, it is an accessible rural area mainly covered by natural vegetation and water, which is proper for soil studies. Second, I was authorized to conduct a soil survey by Grand River Conservation Authority through the assist from two soil scientists at Agriculture and Agri-Food Canada before December 2011. Third, it is expected soil studies in the Laurel Creek Conservation area could contribute to the soil and hydrological studies on the Laurel Creek watershed, the Waterloo Aquifer area, and the Grand River watershed.

The Laurel Creek Conservation Area covers an area of approximately 293.3 hectares in the Regional Municipality of Waterloo (Figure 3.1). Although this area is completely encircled by urban development, the context of it remains in nature, which provides opportunities for environmental studies (Laurel Creek Conservation Area Master Plan, 2004). The main part of it is occupied by the Laurel Creek Reservoir constructed in 1960s for the purpose of adjusting flows of the Laurel Creek, which is a tributary of the Grand

River and enters the reservoir from the west and exits it from the east. The topography in the Laurel Creek Conservation area is characterized by the Waterloo Moraine that is a distinctive relief feature and an important groundwater recharge area. The elevation ranges from ~334 m to ~364 m. Annual average temperature in Waterloo Region is ~6.7℃ and total annual precipitation is ~900 mm (Laurel Creek Conservation Area Master Plan, 2004). Surficial geologic features in the Laurel Creek Conservation area include such as Tavistock Till and Ice-contact deposits (Figure 3.3), soils influenced by which are generally deep and permeable and have good water retention (Laurel Creek Conservation Area Master Plan, 2004).

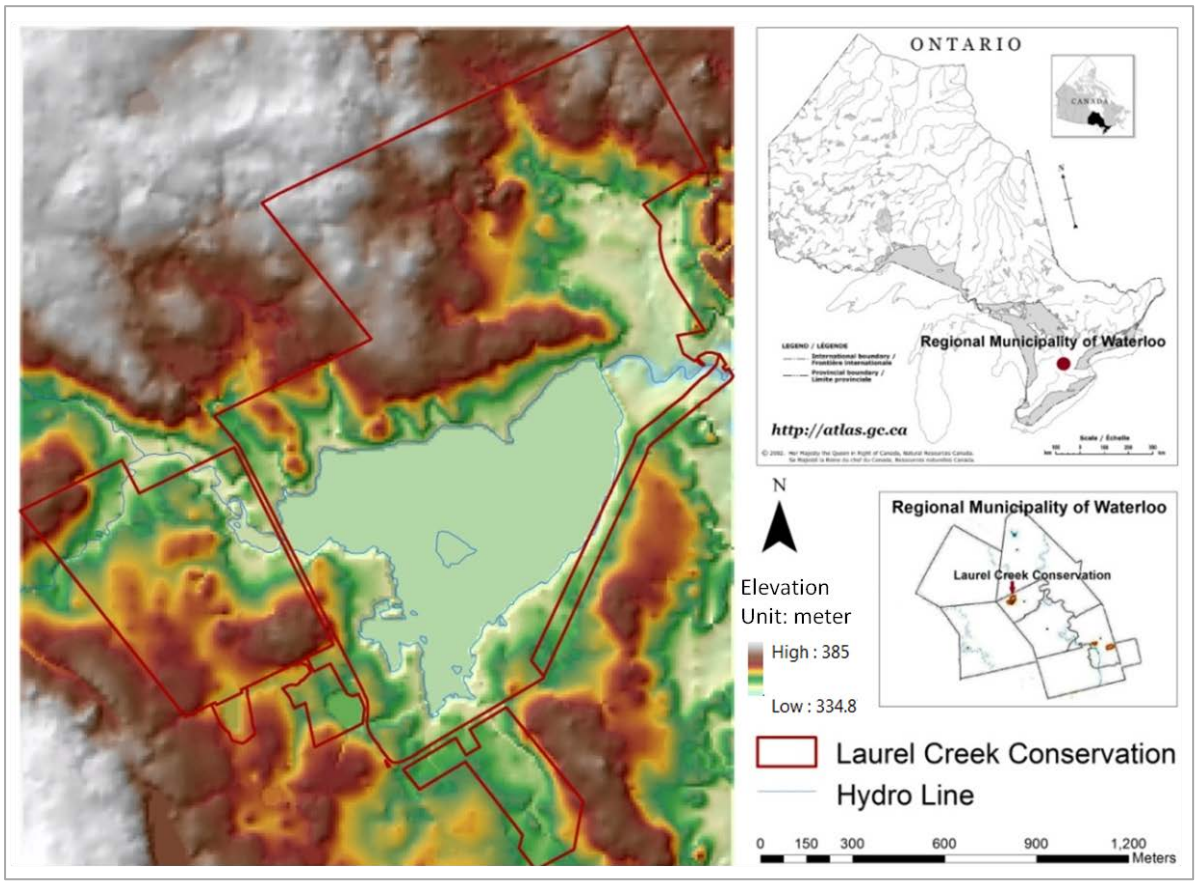

Figure 3.1 Location of the Laurel Creek Conservation area.

3.2 Field Survey and Sampling

A field observation and sampling in the Laurel Creek Conservation area was conducted on November, 8, 2011 (from ~8 am to ~3 pm), with air temperature ranging from 10℃ to 15℃ and high relative humidity (University of Waterloo Weather Station). The weather was proper for performing soil survey in field. The survey was supervised by soil scientists at Agriculture and Agri-Food Canada, Dr. Xiaoyuan Geng and Dr. David Kroetsch. Dr. Xiaoyuan Geng is a soil scientist experiencing in integration of geospatial science, soil science and hydrology on ecosystem modeling and decision making; Dr. David Kroetsch is a senior soil resource specialist in soil survey upgrades and soil re-survey techniques using digital information (Schut et al., 2011).

The sampling sites were determined by soil scientists based on the soil-landscape relationship and also the distribution of natural vegetation. These sites were mainly selected along transects and possible typical soils. Soil series were identified using the three letter CANSIS code (http://sis.agr.gc.ca/cansis/nsdb/slc/v2.2/domsub/name.html). *Soils of Waterloo County* (Presant and Wicklund, 1971) provided detailed description of the expected soil series in the study area, which were used as prototypes for comparisons. Finger assessments were used to determine soil texture, including feel test, moist cast test, ribbon test, taste test and others. Feel test included graininess test, dry feel test, and stickiness test. *Field Manual for Describing Soils in Ontario* was used as reference for finger assessments. The field records included site location and soil code. The survey tools were a Dutch auger and a profile sample box (Figure 3.2 (a)). Soil was pulled out using the

auger (Figure 3.2 (b)) and put into the profile sample box following the layer sequence (Figure 3.2 (c)). A soil profile could be observed in the profile sample box. A scale on the two sides of the profile sample box was used to measure the horizon depth. The use of the Dutch auger and profile sample box allowed a quick and easy sampling procedure in field.

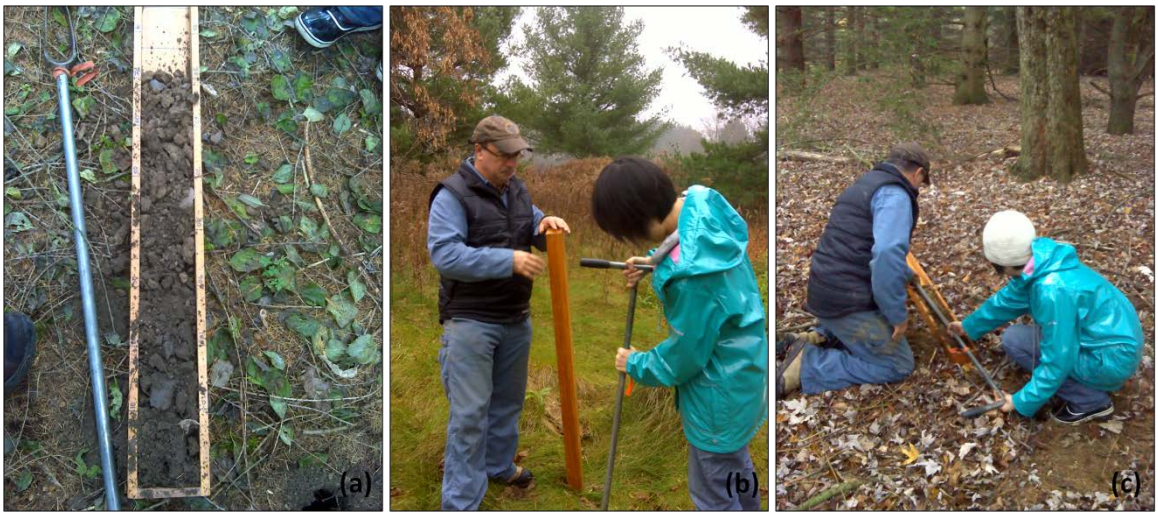

Figure 3.2 Field survey tools and method: (a) Dutch auger (left) and profile sample box (right); (b) using the Dutch auger to collect soil; and (c) using profile sample box to see soil profile.

It should be noted that the survey was limited by the accessibility in field and time. It was mainly developed in the northeast land area of the conservation, where total 19 sites were sampled. The northeast land area occupies relatively larger area than the west land area does, and the middle area is covered mostly by the reservoir. Thus the field survey was conducted mainly in the northeast area. The location and soil series code of the field sampling points are displayed on top of a surficial geology layer in Figure 3.3 (a). Their site number can be seen from Figure 3.3 (b) on top of contours with interval of 1 m based on 10 m conventional DEM. Table 3.1 records detailed information on the sampling points.

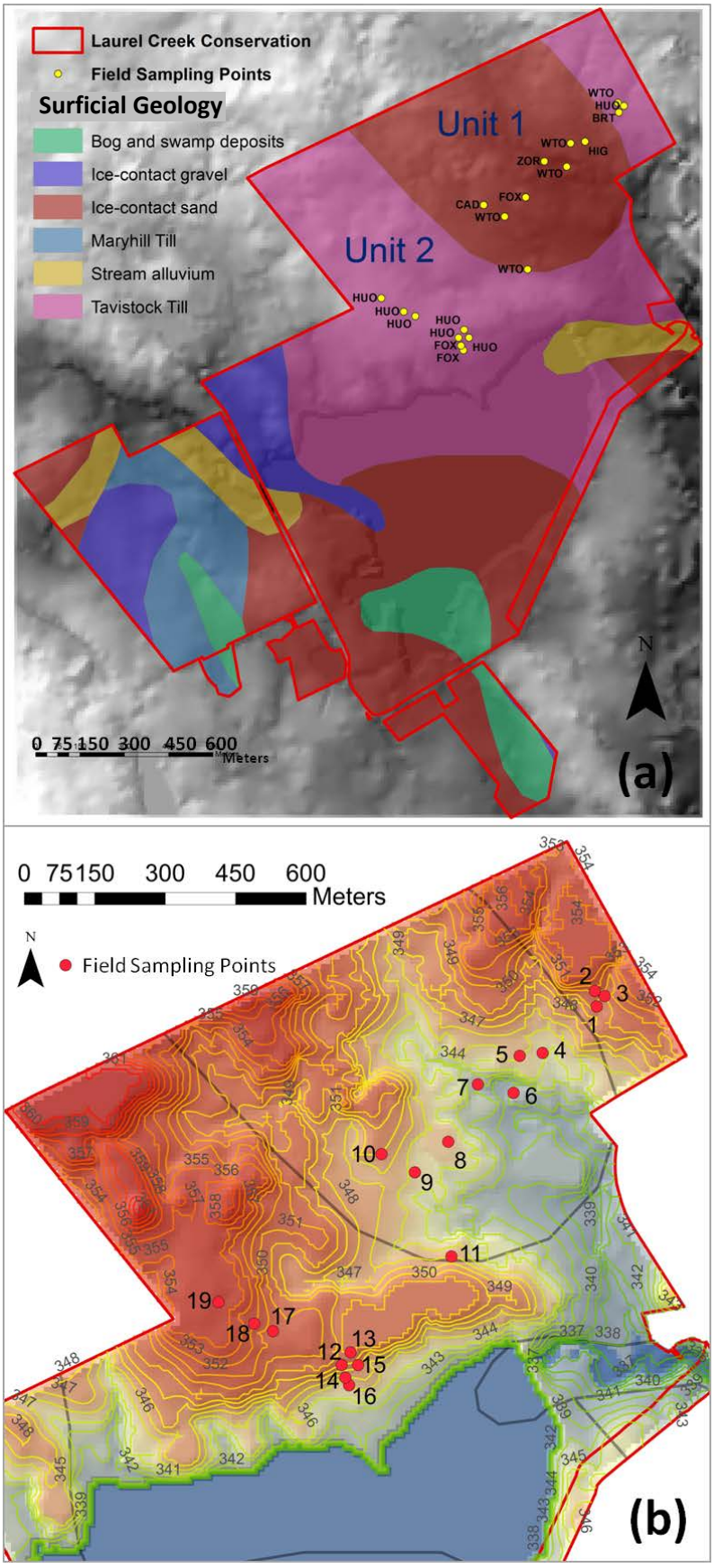

Figure 3.3 (a) Location and soil series code of field sampling points on surficial geology layer, and (b) location and site number of field sampling points on contours with interval of 1 m.

Table 3.1 Summarized information of field sampling points: sites number, identified soil series in three letter code, specific name of soil code, coordinates of sites, and remarks.

Site No.	Soil_Code	Name	Easting (m)	Northing (m)	Remarks
1	BRT	Brant loam	534835	4815412	Toe slope
2	WTO	Waterloo fine sandy loam	534831	4815445	Mid-slope
3	HUO	Huron loam	534852	4815434	Top slope
4	HIG	Heidelberg fine sandy loam	534720	4815313	Waterloo catena
5	WTO	Waterloo fine sandy loam	534671	4815307	
6	WTO	Waterloo fine sandy loam	534658	4815228	
7	ZOR	Organic soils	534582	4815246	Wetland soil
8	FOX	Fox sandy loam	534519	4815124	
9	WTO	Waterloo fine sandy loam	534448	4815059	
10	CAD	Caledon sandy loam	534377	4815098	
11	WTO	Waterloo fine sandy loam	534526	4814880	
12	HUO	Huron loam	534327	4814649	
13	HUO	Huron loam	534311	4814676	
14	HUO	Huron loam	534292	4814649	
15	FOX	Fox sandy loam	534308	4814606	
16	FOX	Fox sandy loam	534300	4814622	FOX over HUO
17	HUO	Huron loam	534146	4814721	
18	HUO	Huron loam	534106	4814737	
19	HUO	Huron loam	534030	4814783	PERTH

3.3 Data and Methodology

The intent of this study was to examine the purposive sampling design for digital soil mapping based on FCM clustering. The initial data include a 10 m conventional DEM, raw LiDAR points, and SPOT images. Figure 3.4 demonstrates the framework of the data collection and methodology. The first step was to assemble environmental covariate databases, including terrain covariates and vegetation covariates. The second step was a purposive sampling design based on FCM clustering of the databases. The cluster information extracted from FCM clustering results were used to analyze the effects of the resolution and the vegetation covariate, NDVI, on optimal partitions. The fuzzy

membership value maps then were employed in the interpretation of fuzzy membership values at sampled points in order to investigate the associations between soil series and fuzzy clusters. Specifically, the data collection and the purposive sampling design method are detailed in following sections.

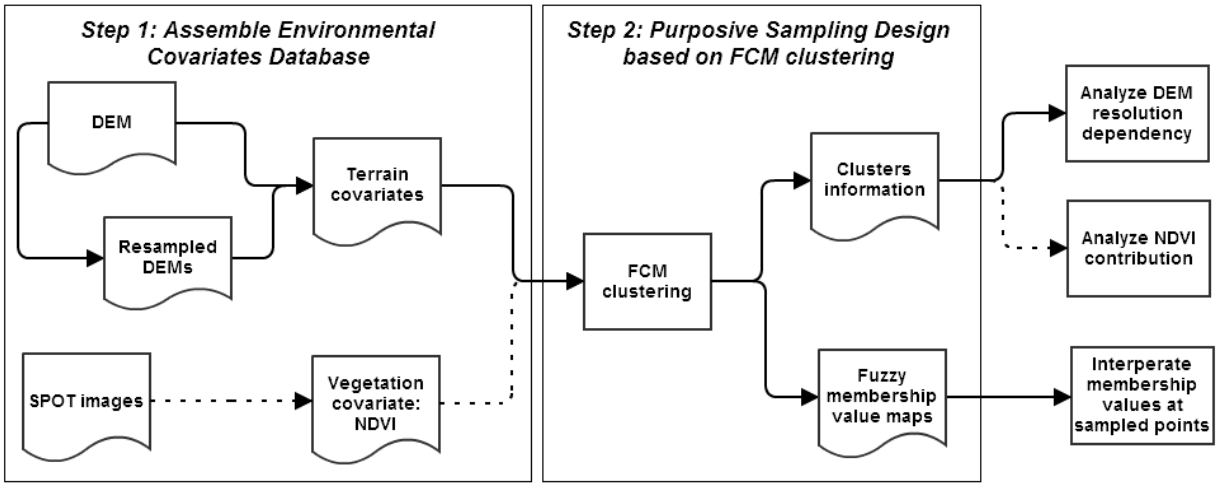

Figure 3.4 Framework of data collection and the purposive sampling design.

3.3.1 Environmental Covariates Data

The field observation and sampling led to some preliminary findings. By overlapping with surficial geologic feature layer (Figure 3.3), sampling sites could be roughly classified into two units, Ice-contact sand unit (Unit 1) containing more sand deposits and Tavistock Till unit (Unit 2) containing more clay deposits. In Unit 1, four of the eight sampling points were identified as WTO series consisting of well drained soils developed on fine and very fine sandy loams, while in Unit 2, seven of the eleven sampling points belonged to HUO series developed on moderately well drained clayey parent materials (Presant and Wicklund, 1971). It is likely that soils vary with geologic features in this area. I contend to control the

geologic factor in this study in order to examine only the interested issues. Therefore, the study area was stratified into those two units, and these two units were employed to examine the resolution and vegetation issues, respectively. This stratification strategy is also commonly used in terrain-based hydrological modeling, for example, channel cells and hillslope cells which are treated differently (Quinn et al., 1995; Gallant and Dowling, 2003; Deng et al., 2007).

3.3.1.1 Generating DEM from Raw LiDAR Points

A 1 m resolution LiDAR DEM was generated from raw LiDAR points in the Laurel Creek Conservation area. The raw LiDAR point data was provided by Mapping, Analysis and Design (MAD) in the University of Waterloo. This dataset covering most areas of the City of Waterloo was collected by Optech® with flying height of ~1300 m, pulse repetition frequency of 70 Hz and point density of 1.5 pt/m^2, on March 11, 2006. The generation of DEM from raw LiDAR points involved two steps: filtering and interpolating as shown in Figure 3.5.

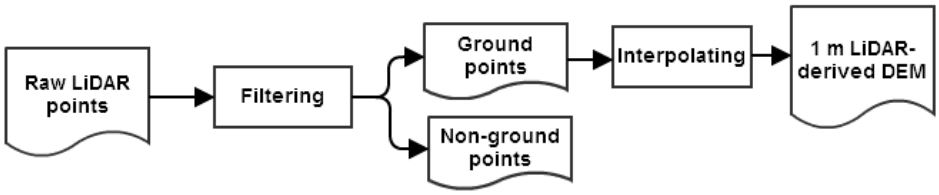

Figure 3.5 Generating DEM from raw LiDAR data.

Filtering is to remove non-ground points from raw LiDAR points. Two categories of filtering methods exist (Liu, 2008). Methods in the first category filter grid images that are interpolated from raw LiDAR points, such as wavelet-based filtering (Vu and Tokunaga,

2001), and segmentation-based filtering (Nardinocchi et al., 2003). Many of these methods stem from digital images processing methods. While it is evident that these methods might lose important information and introduce errors during transferring raw LiDAR points into grid images (Axelsson, 2000; Pfeifer and Mandlburger, 2009). It is therefore suggested to utilize methods in the other category which directly use raw LiDAR points in order to overcome those limitations (Axelsson, 2000; Liu, 2008). One of the popular methods in this category is Axelsson (2000): using adaptive TIN (Triangulated Irregular Networks) models, where a sparse TIN created from seed points is progressively densified by adding ground points which meet certain criteria. This method has been employed in many practical applications (Liu, 2008), and is effective in dense areas. Thus, this method was used in filtering the raw LiDAR points in the study area. Through testing and observation, it was decided to first segment the raw LiDAR point cloud data into subset datasets and then apply the filtering method for each subset individually. The reason for this segmentation could be explained as the LiDAR points belonging to trees and grasses were somewhat difficult to entirely clean up by conducting the filtering algorithm globally.

The filtering results, ground points, were then interpolated into a 1 m resolution DEM. Interpolation is to determine the elevation value at unsampled points using the sampled elevation values at neighbouring points. Commonly used interpolation methods include the inverse distance weighted (IDW), Spline, and Kriging methods. The IDW method assumes the sample points that are closer to the predicted point have more influence on the predicted value. Spline-based methods estimate values using a function to minimize the overall

surface curvature. Kriging is a geostatistical method based on both the distance and the degree of autocorrelation. Previous studies have shown that Kriging can yield the best interpolation results (Zimmerman et al, 1999). However, when dealing with high density points, such as LiDAR points, there are no significant differences between the performances of the IDW and Kriging methods (Liu, 2008; Pfeifer and Mandlburger, 2009). Additionally, the IDW method has a better computational performance than other methods (Pfeifer and Mandlburger, 2009). Therefore, this study employed the IDW method to interpolate the ground LiDAR points into a 1 m DEM in ArcGIS®. The generated LiDAR-derived DEM was filled in ArcGIS®.

3.3.1.2 Resampling

A conventional DEM at 10 m resolution in the Laurel Creek Conservation area also provided the initial data for this study and was used to extract land-surface parameters. This DEM was provided by the Map Library at the University of Waterloo. It was extracted from a DEM covering the whole Grand River watershed produced by Grand River Conservation Authority. This DEM was hydrologically-conditioned and originally constructed based on contour data and digital terrain data from 1:10,000 Ontario Base Maps in the fall of 2000. The horizontal precision is +/- 10 m, and the vertical precision is +/- 5 m. Since the dataset was unfilled, the first step was to fill it in order to avoid errors caused by artefacts.

To examine the effects of resolution on the purposive sampling design, different resolution DEMs were resampled from the basic 1 m LiDAR-derived DEM and the 10 m

conventional DEM. NEAREST, BILINEAR, and CUBIC options are among the commonly used resampling methods in ArcGIS® (Lo and Yeung, 2007). The NEARST option assigns the value of the nearest original cell to the corresponding resampled cell. The BILINEAR option takes a proximity-weighted average of the four original nearest cells to determine the resampled cell. In CUBIC option, the number of the determining original cells becomes16. Among them, the NEAREST option is the preferred method to use when applied to map layers in GIS because it does not change the attribute values of the original grid cells assigned to the reoriented grid cells and it is easier to compute than the other two options (Lo and Yeung, 2007). Previous studies also indicated that no significant differences exist among the three resampling options when computing land-surface parameters (Wu et al., 2008). Therefore, the 1 m LiDAR-derived DEM was resampled into 5 m, 10 m, 20 m, 30 m, and 50 m, and the 10 m conventional DEM was resampled into 20 m, 30 m, and 50 m (Figure 3.6), using the NEAREST option in ArcGIS®.

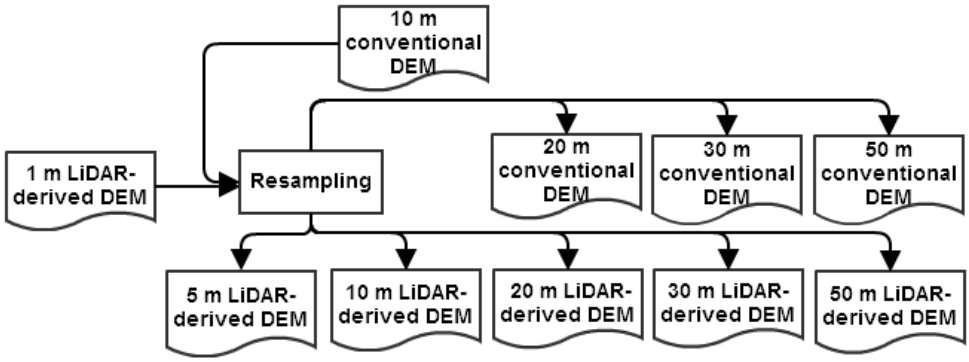

Figure 3.6 Resampling conventional DEM and LiDAR-derived DEM.

3.3.1.3 Deriving Land-Surface Parameters

Five land-surface parameters were decided to be primary importance to soil formation

in the study area: elevation, slope, plan curvature, profile curvature and topographic wetness index (TWI). Many previous studies deemed these land-surface parameters are essential in representing landform (English, 2001; Zhu et al., 2008b; Yang et al., 2011). The typical characteristics of the terrain can be well described using slope and curvatures, and in terms of the impact of soil moisture on soil formation, topographic wetness index can assist in classifying the soil. Moore et al. (1993) indicated that among a number of primary and secondary land-surface parameters, slope, curvatures, and wetness index have higher correlations with the soil properties. McSweeney et al. (1994) and English (2001) also contended that elevation, slope, plan curvature, profile curvature, and topographic wetness index exhibit the majority of relief influence on soil formation. Thus, these five land-surface parameters were input into FCM clustering as terrain covariates (Figure 3.7).

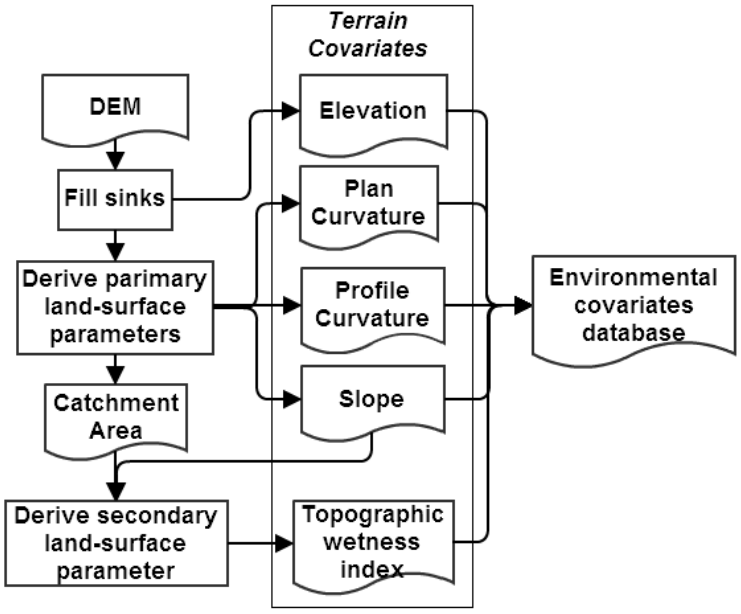

Figure 3.7 Deriving land-surface parameters as terrain covariates.

Primary land-surface parameters, slope, plan curvature, and profile curvature were directly derived from DEM as shown in Figure 3.7. Slope reflects the rate of change of

elevation in the steepest descent, and plan and profile curvatures are two parameters describing the concavity and convexity of the land surface (Olaya, 2009). The calculation of slope, plan curvature and profile curvature was based on the Evans-Young method (Evans, 1972; Young, 1978; Evans, 1979) in which the land surface was modeled by a quadratic polynomial. Another primary land-surface parameter, catchment area, was also derived from DEM directly for the purpose of calculating secondary land-surface parameters (Figure 3.7). Catchment area of a cell is the area of a watershed where water flows to this cell. This parameter was calculated based on a multiple flow direction method (Quinn et al., 1991) in which water in a cell flows into all lower neighboring cells. A secondary land-surface parameter, topographic wetness index, was calculated based on slope and catchment area. Topographic wetness index has the potential to reflect the tendency of a cell to accumulate water. It is defined as:

$$TWI = \ln\left(\frac{A}{\tan\beta}\right) \tag{3.1}$$

where A is catchment area and β is slope angle in degrees.

The elevation, slope, plan curvature, profile curvature, and topographic wetness index layers were all utilized as terrain covariates in FCM clustering. These primary and secondary land-surface parameters were calculated using terrain analysis algorithms embedded within a software package called “SoLIM Solutions 2010” developed by Zhu et al. (http://solim.geography.wisc.edu/software/index.htm). This useful software permitted the derivation of land-surface parameters as well as the purposive sampling design based on FCM clustering of environmental covariates in this study.

3.3.1.4 Deriving NDVI

To examine the contribution of vegetation covariate to the purposive sampling design based on FCM clustering, a vegetation index, NDVI, was employed in the environmental covariates database construction. NDVI is commonly calculated from satellite images using the formula proposed by Kriegler et al. (1969):

$$NDVI = \frac{\lambda_{NIR} - \lambda_{RED}}{\lambda_{NIR} + \lambda_{RED}} \quad (3.2)$$

where λ_{NIR} and λ_{RED} represent the reflectances in the near-infrared (NIR) and red bands, respectively. NDVI responds to the variations in green biomass, chlorophyll content, and canopy water stress (Liang, 2004). It ranges from -1.0 to 1.0, with negative values denoting non-vegetation and positive values indicating vegetation. The higher the value is, the denser the vegetation is. NDVI is regarded as one of the most effective indicators of vegetation in various applications (Ünsalan and Boyer, 2011).

As shown in Figure 3.8, NDVI was calculated using SPOT images downloaded from GeoBase Orthoimage 2005-2010 database (http://www.geobase.ca). This database consists of SPOT 4/5 images during the period 2005-2010 at 20 m resolution. In SPOT 4/5 images, Bands 2 and 3 are located in red and near-infrared, respectively. Thus, NDVI was calculated in ArcGIS® using:

$$NDVI = \frac{B_3 - B_2}{B_3 + B_2} \quad (3.3)$$

where B_2 and B_3 denoted the SPOT Bands 2 and 3, respectively. The NDVI layer was input into the FCM clustering as a vegetation covariate.

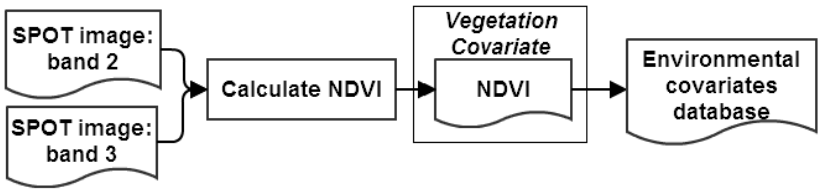

Figure 3.8 Calculating NDVI from SPOT images.

3.3.2 Purposive Sampling Design Method

The purposive sampling design based on FCM clustering of environmental covariates was conducted in Units 1 and 2 areas, respectively. In particular, the experiments were grouped into: Groups A and B on the use of conventional DEMs, and Groups C and D based on LiDAR-derived DEMs. Group A consisted of four experiments conducted in Unit 1 area based on four conventional DEMs at 10 m, 20 m, 30 m, and 50 m resolutions. Group B included two experiments conducted in Unit 2 area based on 10 m conventional DEM with and without NDVI, respectively. Experiments in Group C included six in Unit 1 area based on the LiDAR-derived DEMs with resolutions of 1 m, 5 m, 10 m, 20 m, 30 m, and 50 m. Group D involved two experiments in Unit 2 area using 10 m LiDAR-derived DEM with and without NDVI, respectively. Hence, the effects of resolution on the purposive sampling design could be analyzed based on the results of the experiments in Groups A and C, and the contribution of NDVI could be seen through the experiments in Groups B and D. The use of LiDAR-derived DEM could be analyzed through experiments in Groups C and D. In each experiment in these four groups, the purposive sampling design in general embodied two steps as shown in Figure 3.4.

1) The first step was to construct the environmental covariate database. The five

experiments in Group A constructed five databases embedded with five land-surface parameters (slope, plan curvature, profile curvature, topographic wetness index, and elevation) derived from the 10 m, 20 m, 30 m, and 50 m resolution conventional DEMs, respectively. In the first experiment of Group B, the database contained the five land-surface parameters derived from the 10 m conventional DEM and the resampled 10 m NDVI, while in the second experiment of Group B, the NDVI dataset was deleted from the database. The six experiments in Group C used the five land-surface parameters derived from the 1 m, 5 m, 10 m, 20 m, 30 m, and 50 m LiDAR-derived DEMs, respectively. In the first experiment in Group D, the database consisted of both the five land-surface parameters obtained from LiDAR-derived DEM and NDVI at 10 m resolution. The second experiment in Group D only contained the five LiDAR-derived land-surface parameters.

Although the initial NDVI was derived from the SPOT images at 20 m resolution, it was resampled into 10 m resolution to match other datasets in the experiments of Groups B and D. It should be noted that the process was only for the compatibility purpose, and the resolution of NDVI was only 20 m even though it was resampled into higher resolution.

The datasets in environmental covariates databases carried different numerical ranges, for example, elevation ranging around 330~360 m while slope ranging around 0~10 degree. By assuming that these covariates have the same weight in predicting soil variation, it was suggested to standardize these covariates to the same numerical range (Yang, 2006; Zhu et al., 2008a). Since cells in slope, topographic wetness index, and elevation layers carried only positive values, these layers were stretched into 0~100; while, plan curvature, profile

curvature, and NDVI layers were stretched to -50~50 because they originally ranged from negative values to positive values.

2) The second step was to conduct FCM clustering of environmental covariates. The aim of using FCM clustering was to identify the patterns of the environmental configuration. For each experiment in Groups A, B, C, and D, FCM clustering was performed five times across algorithms with fuzziness exponent m of 1.25, 1.5, 1.75, 2, and 2.25, respectively. For each run, the number of clusters c ranged from 2 to 15, the iteration number was 50 and the maximum membership error ε was 0.01. The FCM clustering results in each run included a set of fuzzy membership value maps in raster format corresponding to clusters, respectively, and a text file recording information on fuzziness exponent m, iteration number, cluster number c, partition coefficient F, entropy H, and fuzzy partition error (payoff) J_m. By examining the improvements in partition coefficient (F) and entropy (H) across the adjacent clustering, possible optimal numbers of clusters could be found in each run. When the improvement in H with cluster number changing from c to $c + 1$ was larger than the improvement with cluster number changing from $c - 1$ to c and from $c + 1$ to $c + 2$, c was regarded as an optimal cluster number in this run (Yang et al., 2011).

After FCM clustering of environmental covariates, an optimal partition of the study domain was able to be found. Each cluster in this partition would be presented in a fuzzy membership value map. Locations with highest membership values would be deemed as the fittest representations of the unique environmental configuration corresponded by this cluster because they were the closest to the cluster centroids. Field sampling could be

conducted only at such locations to interpret the environmental configuration as representative of soil series.

3.4 Chapter Summary

This chapter first presents the information of the Laurel Creek Conservation area and field sampling. The field survey indicated a strategy for the study of soil in this area, which was to stratify the area according to the suficial geology layer in order to control influential factors. Two units of geology features were selected as subset study areas. Following is the detailed framework of data collection and methodology in this study. Raw LiDAR point cloud data were utilized to generate a LiDAR-derived DEM. Both conventional and LiDAR-derived DEMs were resampled into coarser resolutions, based on which primary and secondary land-surface parameters were derived as terrain covariates input into environmental covariates database. NDVI calculated from SPOT images was used as vegetation covariate in the database. The purposive sampling design was based on FCM clustering of environmental covariates databases, and organized into four groups of experiments in which DEM source and NDVI were controlled.

CHAPTER 4 RESULTS AND DISCUSSION

This chapter first presents FCM clustering results of experiments in Groups A and C for Unit 1 area, based on which the effects of resolution on the purposive sampling design are discussed. Then it shows the results of experiments in Groups B and D in Unit 2 area, based on which the contribution of vegetation covariate is discussed. The fuzzy membership values at the sampled sites are interpreted in the third section. The demonstration of results in Groups C and D also implies an opportunity to discuss the use of LiDAR data in digital soil mapping. In the last section, a validation experiment in a subset area of the Waterloo Aquifer area is provided.

4.1 FCM Clustering Results in Unit 1 Area

4.1.1 Use of Conventional DEM

F and H

Graphical displays of the partition coefficient (*F*) and entropy (*H*) with number of clusters across five algorithms of weighting exponent of m = 1.25, 1.5, 1.75, 2, and 2.25 based on 10 m, 20 m, 30 m, and 50 m conventional DEMs in experiments of Group A are illustrated in Figure 4.1.

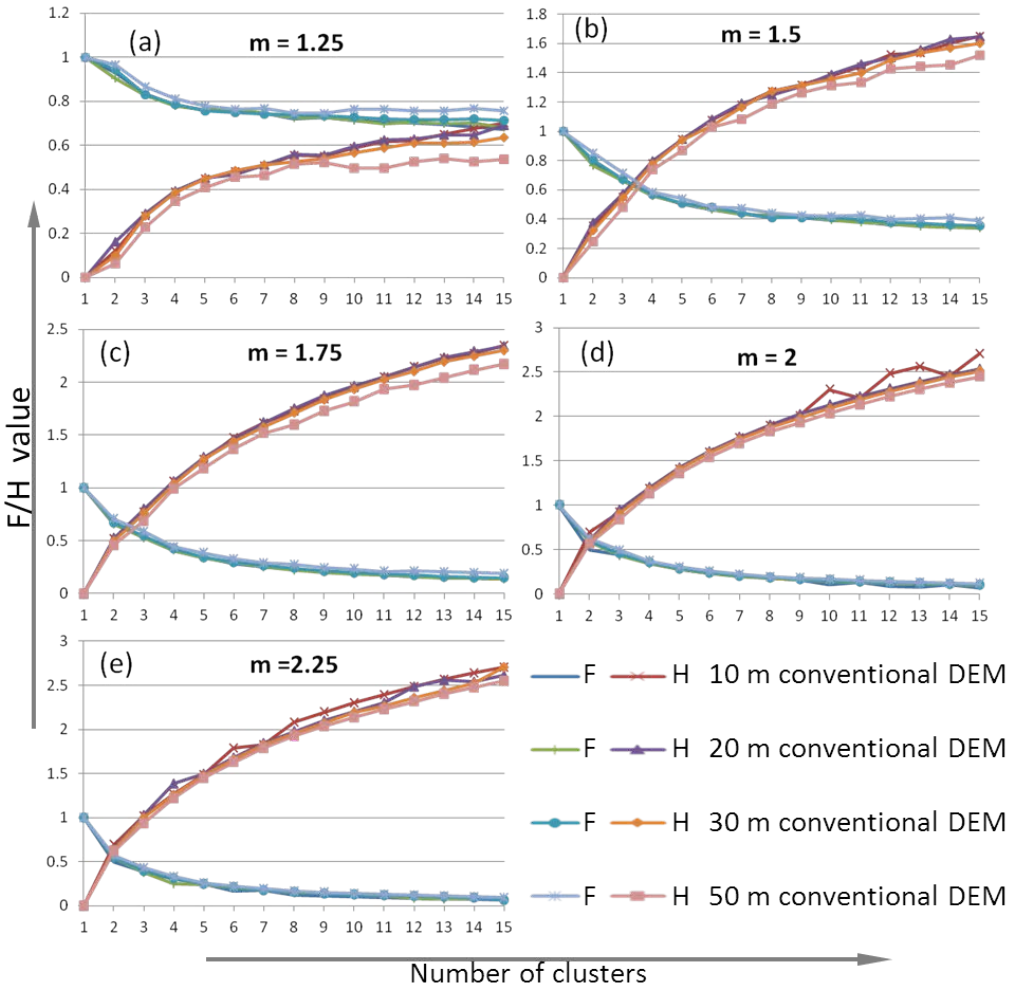

Figure 4.1 Partition coefficient (*F*) and entropy (*H*) plotted against the number of clusters based on 10 m, 20 m, 30 m, and 50 m conventional DEMs across five algorithms with *m* of (a) 1.25, (b) 1.5, (c) 1.75, (d) 2, and (e) 2.25 in Unit 1 area.

As the number of clusters increases, *F* decreases and *H* increases gradually. According to theory, *F* is inversely proportional to the overall average overlap between pairs of fuzzy clusters, and *H* is a scalar measure of the amount of fuzziness in a given fuzzy partition. Thus, the result presented in Figure 4.1 confirms that the overlap between clusters increases as the number of clusters increases. It also can be seen that increasing *m* leads to lower *F* and higher *H* and therefore reduces the estimated classification certainty. It is desirable to achieve a balance between fuzziness and certainty by choosing a proper weighting exponent *m*.

Slight changes of *F* and *H* with resolution exist. First, *F* tends to be larger and *H* tends to be smaller at each number of clusters when the resolution becomes lower. This trend is evident when *m* is small (e.g., *m* = 1.25). Second, when *m* is 1.25, 1.5, and 1.75, *H* at each number of clusters at 50 m resolution is obviously smaller than those at the other three higher resolutions. Third, when *m* is 2 and 2.25, the trend of *H* at 10 m resolution does not follow the others at lower resolutions. It experiences some small fluctuations. When *m* is 2.25, *H* experiences such fluctuations not only at 10 m resolution but also at 20 m resolution, while *F* keeps stable. Although a shift of the statistical population induced by the change in the number of evaluated cells might lead to these changes, it is still possible that they could imply the resolution dependency. For example, the different changes of *H* when the resolution is aggregated to 50 m might indicate a threshold of resolution for fuzzy clustering in Unit 1.

Improvements in F and H

It is useful to examine the improvements in *F* and *H* with numbers of clusters graphically in order to determine the optimal numbers of clusters. The improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and the entropy ($\Delta H = H(c) - H(c+1)$) over adjacent clusters across *m* of 1.25, 1.5, 1.75, 2, and 2.25 are demonstrated in Figure 4.2, 4.3, 4.4, 4.5, and 4.6, respectively. Each figure has four graphs showing ΔF and ΔH in four grid sizes. ΔH reaches its local maximum or minimum values and ΔF reaches its local minimum or maximum values at the same number of clusters, respectively.

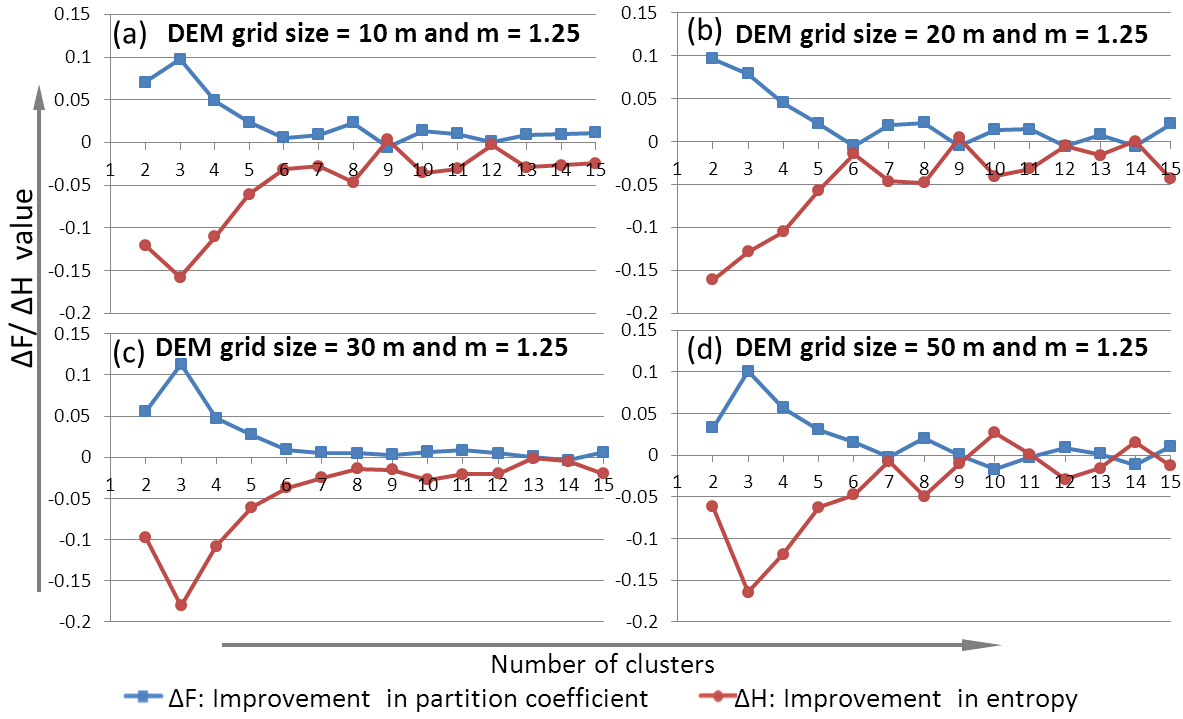

Figure 4.2 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, and (d) 50 m conventional DEMs, using the algorithm with *m* of 1.25 in Unit 1 area.

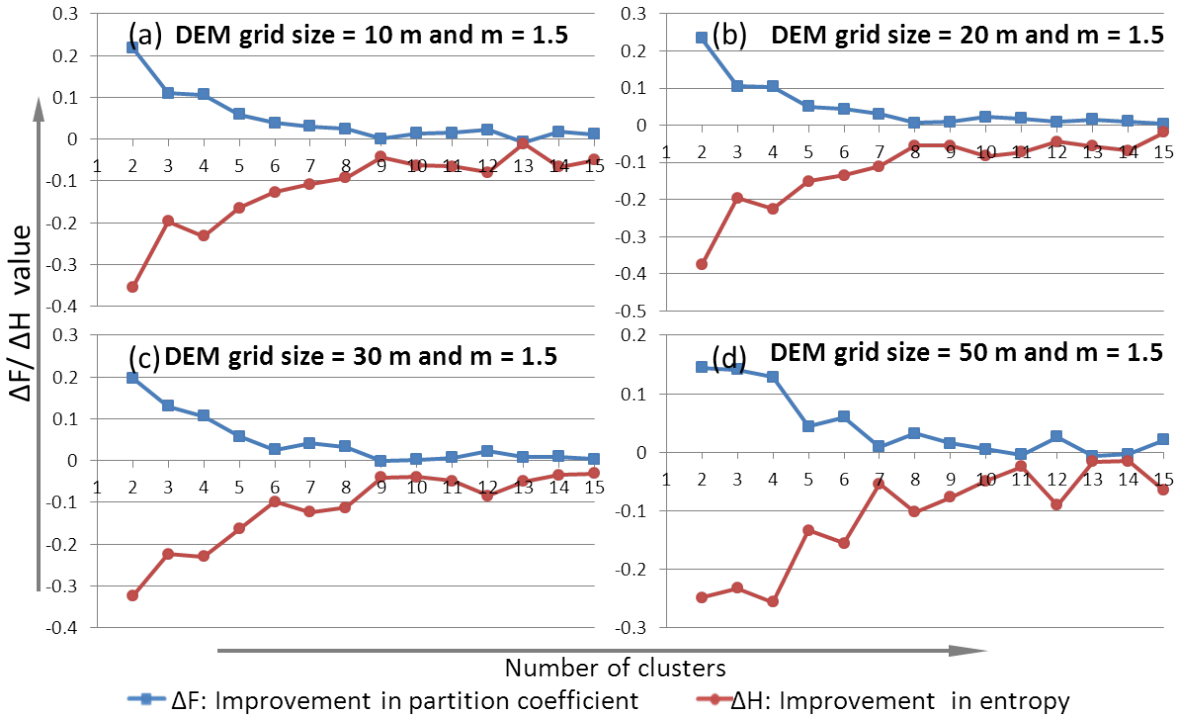

Figure 4.3 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, and (d)50 m conventional DEMs, using the algorithm with *m* of 1.5 in Unit 1 area.

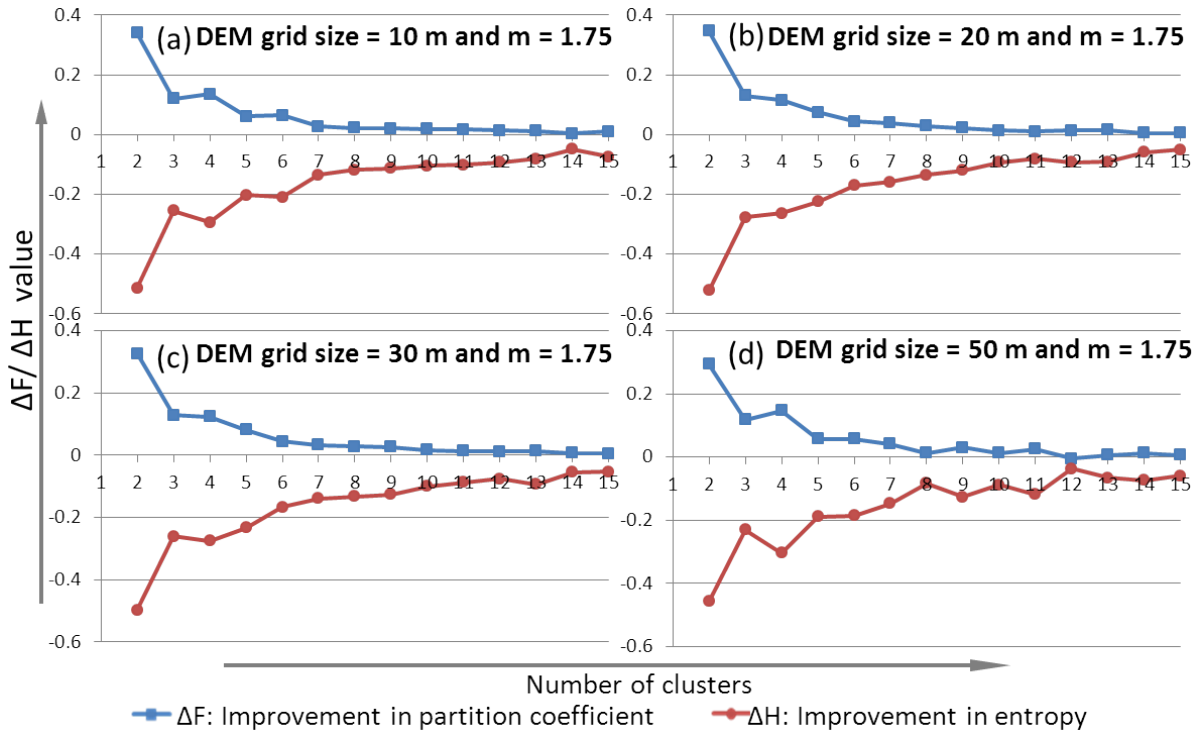

Figure 4.4 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, and (d)50 m conventional DEMs, using the algorithm with *m* of 1.75 in Unit 1 area.

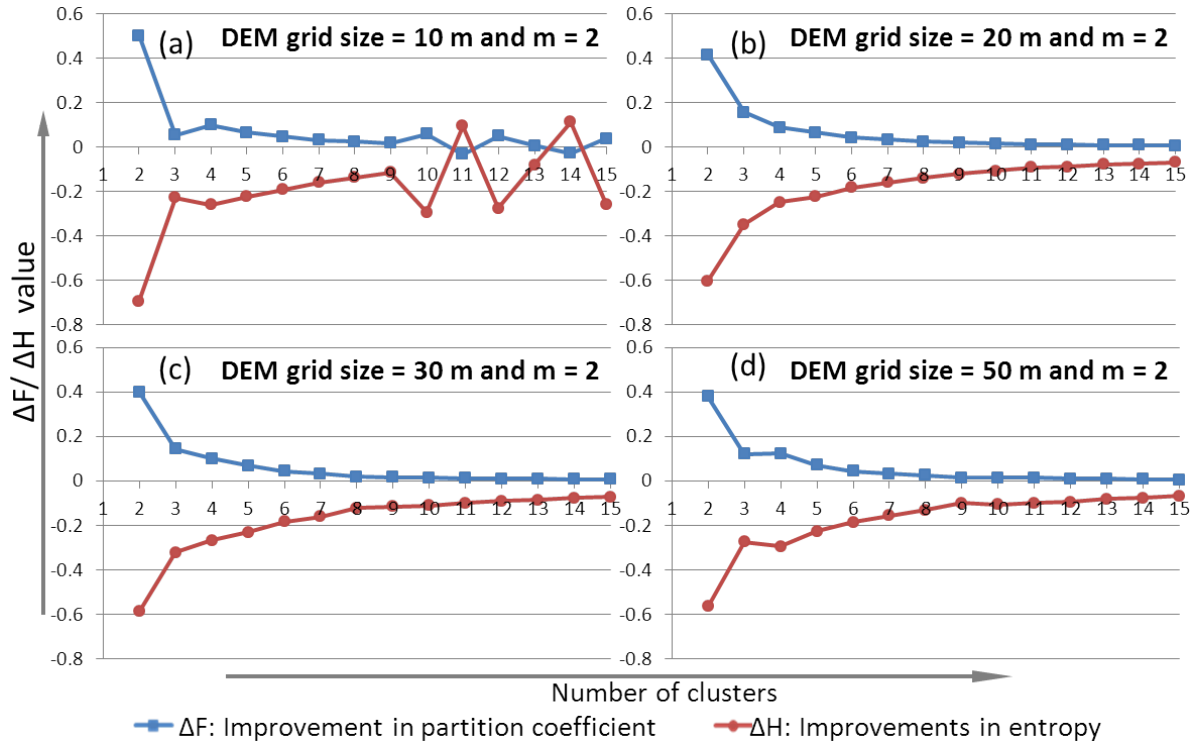

Figure 4.5 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, and (d)50 m conventional DEMs, using the algorithm with *m* of 2 in Unit 1 area.

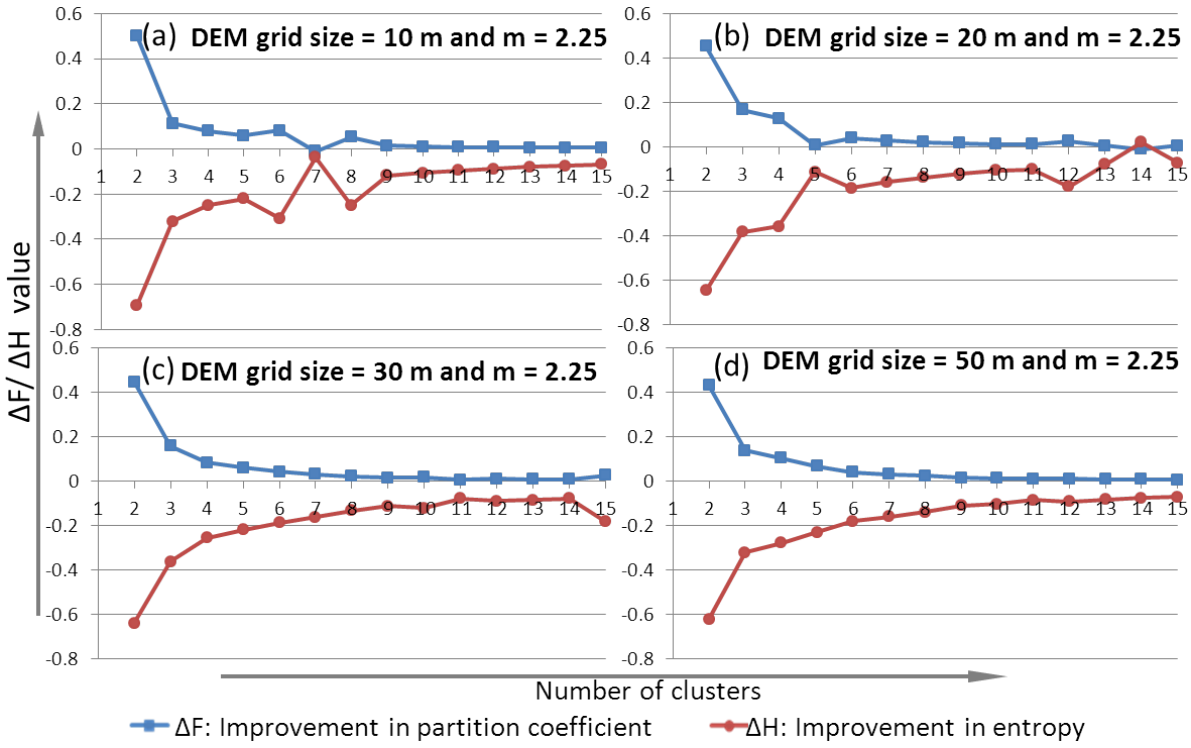

Figure 4.6 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, and (d)50 m conventional DEMs, using the algorithm with *m* of 2.25 in Unit 1 area.

Figures 4.2-4.6 show that increasing *m* from 1.25 to 2.25 generally weakens the fluctuations in ΔF and ΔH at each of the four resolutions. In Figure 4.2, when the grid size changes from 10 m to 20 m, the improvements have similar variations with the number of cluster increasing. When the grid size becomes 30 m, the improvements tend to be stable. However, when the grid size reaches 50 m, ΔF and ΔH become to fluctuate again. In Figure 4.3, ΔF and ΔH at 50 m are the most unstable. The fluctuation becomes larger when the grid size increases. In Figure 4.4, the improvements are roughly stable when the grid size increases from 10 m to 30 m. When the grid size becomes 50 m, the improvements show some variations with the increasing of clusters. In Figure 4.5, ΔF and ΔH are smooth when the grid size grows from 20 m to 50 m. When the grid size is 10 m, the improvements

undergo slight fluctuations with larger numbers of clusters. In Figure 4.6, when *m* comes to a larger value, 2.25, *ΔF* and *ΔH* become relatively smooth. However, when the grid size is smaller, sparse variations exist in the curves. When *ΔF* reaches its local minimum value and *ΔH* reaches its local maximum value, the cluster number is regarded as one optimal number of clusters. Table 4.1 summaries the possible optimal numbers of clusters with different *m* and at different grid sizes.

Table 4.1 Possible optimal numbers of clusters based on 10 m, 20 m, 30 m, and 50 m conventional DEMs across five algorithms with *m* of 1.25, 1.5, 1.75, 2, and 2.25 in Unit 1 area.

grid size / *m*	10 m		20 m				30 m	50 m		
1.25	9	12	6	9	12	14		7	10	14
1.5	9	13	12				6	5	7	11
1.75								8	10	12
2	11	14								
2.25	7		5	14						

It can be seen from Figures 4.2-4.6 and Table 4.1 that *ΔF* and *ΔH* vary with the grid size. Increasing the grid size from 10 m to 30 m generally mutes the fluctuations in *ΔF* and *ΔH* and therefore reduces choices of optimal numbers of clusters. However, when the grid size is 50 m, the fluctuations in *ΔF* and *ΔH* are enhanced and there are more choices of optimal numbers of clusters. These findings indicate that the clustering results can be different when the resolution shifts and therefore various environmental configurations may be identified. For example, a different environmental configuration containing fewer landform details may be established at a coarser resolution. This new configuration may correspond to a soil series that is not related to any configuration obtained at higher

resolutions. In other words, as the resolution changes different environmental configurations may become evident and significant in the prediction of soils.

These findings can be explained as environmental modeling (e.g., soil modeling) is restricted to spatial scales and resolutions on which the studied biophysical processes depend or the modeling *per se* depends (Zhuang and Montgomery, 1994; Band and Moore, 1995; Florinsky and Kuryakova, 2000; Deng et al., 2007). For example, land-surface parameters are dependent on the resolution at which they are calculated (Moore et al., 1993). In Deng et al. (2007), plan curvature and profile curvature derived from coarser resolutions are found to induce an unrealistic topography that is much smoother and gently rounded. As well, topographic wetness index is very sensitive to DEM resolution, while it may strongly relate to the landscape type (Zhang and Montgomery, 1994; Florinsky and Kuryakova, 2000; Deng et al., 2007). Thus, it is possible the variation of the five land-surface parameters used in this study with DEM resolution that has significant impact on the FCM clustering and the purposive sampling design.

A related study is done by Arrell et al. (2007) focussing on fuzzy identification of morphometric landform classes from DEMs. They argued that as the resolution increases, the number of fuzzy classes which can capture meaningful landform information will decrease and therefore the optimal number of clusters will decrease. It seems this conclusion is against the FCM clustering results based on conventional DEM in Unit 1 area that increasing resolution (from 30 m to 10 m) leads to more optimal numbers of clusters. However, it should be noted that the two studies are interested in different scales which can

influence the conclusions. In addition, it has not been examined whether the number of meaningful partitions increases with the increase of resolution in the study in Unit 1 area.

It can be seen through the findings that resolution has an influence on the recognition of the dominant environmental configuration. The spatial resolution of environmental covariates used in soil inference may not always correspond to the spatial scale of soil forming processes (Geng et al., 2012). Identifying the dominant environmental configurations at different resolutions is therefore important for understanding the scale of soil forming processes in the specific area.

An *m* equal to 1.5 represents a good compromise with respect to the improvements and stability of the classification. This result corresponds to previous studies which also used 1.5 as a proper value of *m* (e.g., Odeh et al., 1992; MacMillan et al., 2000; Burrough et al., 2001; Arrell et al., 2007). Thus, the clustering results with *m* of 1.5 are explored in demonstrating optimal partitions.

Hardened Class Maps

Fuzzy membership value maps in FCM clustering results indicate the spatial distribution of fuzzy membership values to clusters. A technique in fuzzy logic-based digital soil mapping is to harden these fuzzy membership value maps into conventional crisp form for the purpose of comparisons or interpretations (Burrough et al., 2000; English, 2001; Zhu et al., 2008a). The hardening present in this study is done by classifying each cell into the class to which the cell carries the highest membership value. This hardening can be further employed to create raster soil categorical maps if the association between

clusters and soils are obtained through soil inference (Zhu et al., 2008a). Four hardened class maps for the Unit 1 area are shown in Figure 4.7. These maps were drawn from the results of the FCM clustering in Group A experiments at four resolutions with *m* of 1.5. The numbers of clusters at the four resolutions were selected from Table 4.1. They are appropriate for partition because they appear multiple times (e.g., 9 at 10 m, 12 at 20 m, 7 at 50 m) or are the only choice (e.g., 6 at 30 m) among the possible optimal choices. A conventional 1: 50,000 soil map obtained from National Soil DataBase (NSDB), Agriculture Agri-Food Canada, is also presented for comparisons. This soil map was produced based on soil survey in 1970s (Presant and Wickland, 1971).

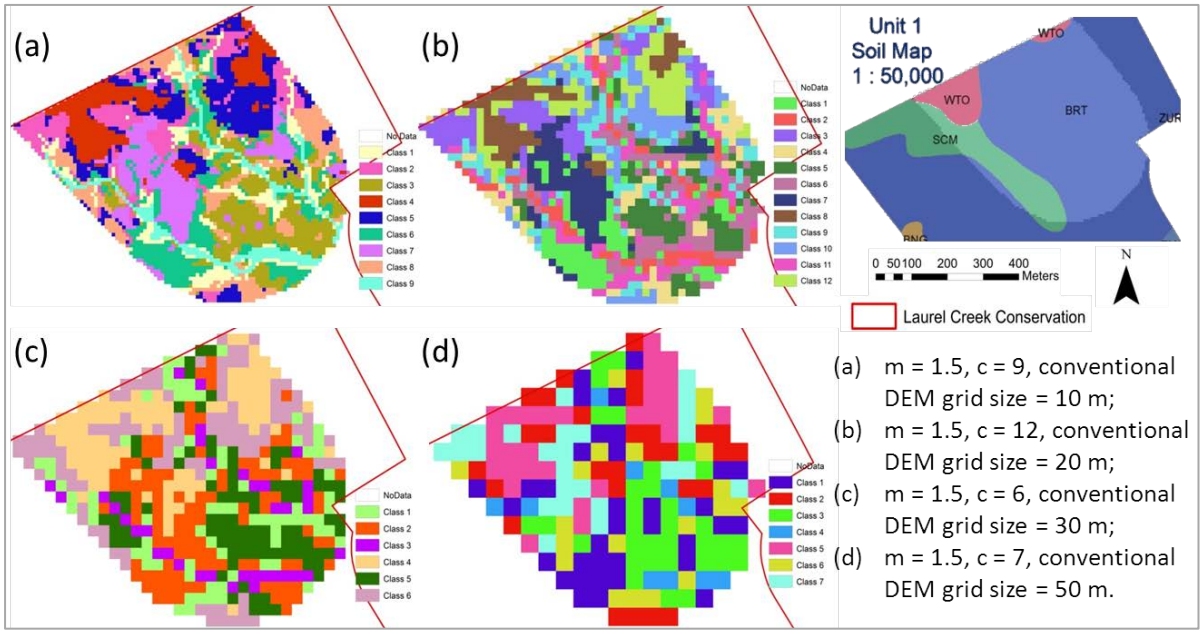

Figure 4.7 Hardened class maps based on conventional DEM with resolution of and number of clusters *c* of (a) 10 m and 9, (b) 20 m and 12, (c) 30 m and 6, and (d) 50 m and 7, respectively, using the algorithm with *m* of 1.5 in Unit 1 area.

In hardened class maps, each cell represents the most possibility of belonging to a cluster. It is likely that hardened class maps can capture both the local and overall variation

of soil information if the association between soil and the clusters exists (English, 2001). Although the area is classified into different numbers of clusters at four resolutions, a general consistent pattern can still be seen through the four maps. For example, the classes around the northwest area occupy almost the same area and have the same shape across four resolutions, namely, Class 4 in (a), Class 8 in (b), Class 4 in (c), and Class 5 in (d). These classes generally correspond to the distribution of soil series WTO in pink colour in the 1:50,000 conventional soil map. Although, it might be explained as both conventional soil survey and fuzzy clustering utilized terrain parameters, so that this corresponding is only a representation of terrain feature, this finding may still indicate the existing of the association between fuzzy clusters and soil series. In addition, all the four hardened class maps show more detailed variations than the 1:50,000 conventional soil map. Thus more detailed variations of soil are possible to be revealed in fuzzy logic-based digital soil mapping.

Through the fuzzy membership maps and hardened class maps, it is possible to identify the optimal viewing resolution of the clusters as 10 m for Unit 1 area. More details are present at this resolution and the surface generalisation is increased at other resolutions. In addition, classes at 10 m resolution are generally spatial adjacent, while classes at lower resolutions have considerable scatters. However, for soil classification purpose, it is not necessary that the highest resolution performs the best. A coarser resolution may provide a clearer and more meaningful classification result by reducing the classification noise (Arrell et al., 2007; MacMillan et al., 2010). It is likely the hardened class maps at lower

resolutions in this study can reveal much clearer patterns of soils than those at higher resolutions, which needs further validations. For example, the hardened map at 10 m resolution may reveal variations at a level of detail that does not match with the one reflecting soil forming processes. In other words, too much noise may exist at this scale for soil forming processes. In previous studies, Smith et al. (2006) argued that while very fine resolution DEMs contain much more detailed terrain information, they may not lead to more accurate soil resource inventories because fine resolution does not necessarily contribute to the differentiation of soil at the scale of interest. Zhu et al. (2008b) suggested that DEM resolution does not have significant impact on the accuracy of soil maps in that it does not have impact on derived slope values. They indicated that coarse DEM soothes out the fine details which are not important to soil classification at the studied scale. Similar findings can also be found in Howell et al. (2008). It is necessary to identify an optimal resolution at which fuzzy clustering and soil inference perform best.

Catenary Sequence

Images representing the spatial distribution of membership values to each class at three resolutions are shown in Figure 4.8 by arranging them as to appear in a catenary sequence, from the flat bottoms in southeast part to the tops in north part in Unit 1 area. Only the images at 10 m, 20 m, and 30 m resolutions are shown in this figure because when the grid size is 50 m it is unclear to observe such a catenary sequence from membership value maps. While the particular arrangement of numbers of clusters at each resolution differed numerically, each captured similar landscape positions in Unit 1 area.

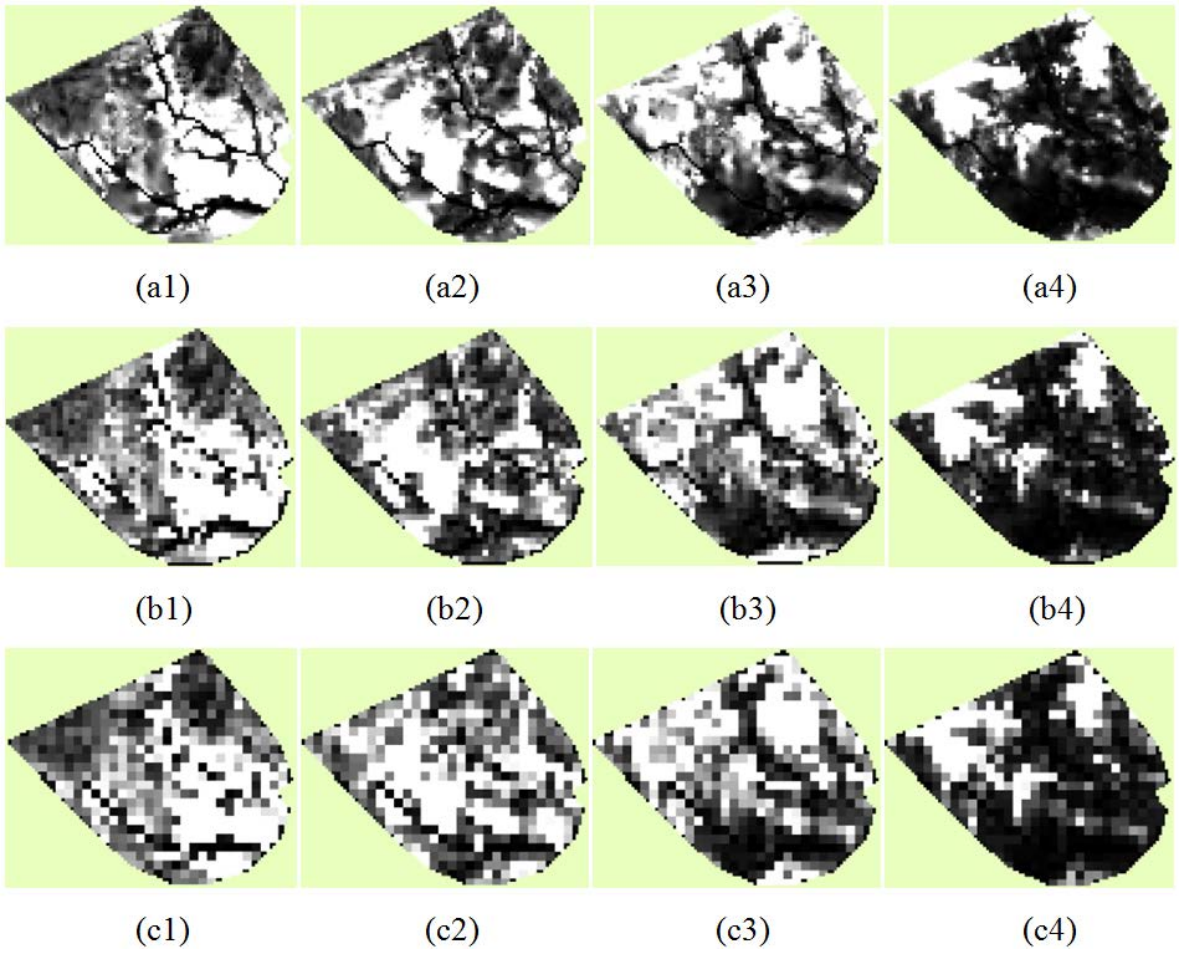

Figure 4.8 FCM membership value maps showing a catenary sequence using the algorithm with *m* of 1.5 and number of clusters *c* of 9 in Unit 1 area: (a1) Class 3, (a2) Class 7, (a3) Class 5, (a4) Class 4, based on 10 m conventional DEM; (b1) Class 8, (b2) Class 5, (b3) Class 3, (b4) Class 4, based on 20 m conventional DEM; (c1) Class 1, (c2) Class9, (c3) Class 2, (c4) Class 6, based on 30 m conventional DEM (light tone indicates high membership value).

In Figure 4.8, from left to right, cells with highest membership values correspond to landscape positions from flat and low bottoms to high and sloping tops in a catenary sequence. These classes are considered to be associated directly with obvious landscape positions (Odeh et al., 1992; Ahn et al., 1999; English, 2001). This result confirms the observations of Irvin et al. (1997), Burrough et al. (2000), and Arrell et al. (2007) that FCM clustering allows the extraction of geomorphologically significant classes. These classes may correspond to distinct environmental configurations which can be used to identify

soil-landscape associations. Other fuzzy classes not shown in Figure 4.8 then can be interpreted as transition zones between distinct landscape positions (Odeh et al., 1992; Ahn et al., 1999; English, 2001). Soil series related to these transition classes may also relate to transitive types. Another explanation could be that the other fuzzy classes are difficult to interpret and do not contain any geomorphologically significant meanings (Arrell et al., 2007).

Figure 4.8 also identifies the persistence of a catenary sequence at different resolutions. This catenary sequence is captured obviously and similarly at each of the three resolutions. If a landscape position is present in the same area at different spatial resolutions without changes in its shape or form, then it can be recognised as persistent and isometric (Arrell et al., 2007). The classes exhibiting the catenary sequence can be viewed as resolution independent and therefore are of great importance in soil inference. In particular, figures (a1), (b1), and (c1) show flat and low areas persistent at all three resolutions. These areas may imply soil series containing more water content because the water tends to flow and accumulate at these areas. For example, one of the sampled points in such areas, namely Point 7 (Table 3.1), was identified as ZOR series which is organic soils on wetlands. However, this persistence seems present only when the resolution is aggregated from 10 m to 30 m. When the resolution is 50 m, it is unlikely to identify such a persistent catenary sequence. The explanation may stem from that the resolutions lower than 50 m dramatically dilute the variation of environmental covariates, or that the landscape characteristics only dominate when resolution is higher than 50 m at the study scale. Therefore, it is suggested

that for Unit 1 area, it is better to use resolution higher than 50 m to classify landform or soil.

Although the purpose of examining FCM clustering results in Unit 1 area is to gain a fuller understanding of the sensibility of resolution to the purposive sampling design, it also poses the question of which is the best resolution to predict the soils. Knowledge within a soil context is needed for the determination. For example, examining the conventional soil maps and detailed soil survey reports in the study area can aid in the determination of the optimal resolution and number of clusters. The desired resolution depends on a number of issues, such as the environment of study area, available datasets, and also the computation ability. Another challenge is the identification of the best number of clusters. It is recommended that the number of clusters is better larger than the types of soils on the conventional soil map (Yang, 2006). It should be expected that each of all the soil types occurring on soil maps can be associated with at least one cluster unless this soil type does not exist anymore. It is also expected that new soil types never reported in conventional soil survey may appear to correspond to some clusters. Therefore, it is better to use more clusters than the known types of soils occurring in the study area to design the field survey and conduct soil inference.

Consequently, the FCM clustering results in Unit 1 area based on conventional DEM reveal that the resolution plays a vital role in the determination of optimal numbers of clusters and therefore in the partitioning of the environmental covariate space and identifying the environmental configurations. The choice of the optimal numbers of clusters

shifts with changing resolution: it decreases with the resolution aggregated from 10 m to 30 m, and suddenly increases when the resolution is 50 m. Hardened class maps imply the clusters associated with soils and it is not necessary the finer resolution can lead to better clustering results. The observation of a persistent catenary sequence through different resolutions indicates the association between landform positions and clusters, and also suggests fuzzy clustering at resolutions higher than 50 m in Unit 1 area.

4.1.2 Use of LiDAR-derived DEM

LiDAR-derived DEM

The 1 m DEM generated from raw LiDAR data is demonstrated in Figure 4.9, along with the raw LiDAR points and filtered ground points. Compared with conventional DEM in the study area (Figure 3.1), this LiDAR-derived DEM carries more details in relief variation but also some man-made features. For example, it is clear to see a dam on the reservoir from this DEM, which may be mis-interpolated in terrain analysis. The footprint of roads could also be captured in this DEM. Although such features could cause errors in deriving land-surface parameters, the impact of them is not deemed as significant in this study. The LiDAR-derived DEM does not cover a small part in the west corner of the Laurel Creek Conservation area due to the range of the raw LiDAR point data, while this missing does not influence the experiments in Units 1 and 2 areas located in the northeast part.

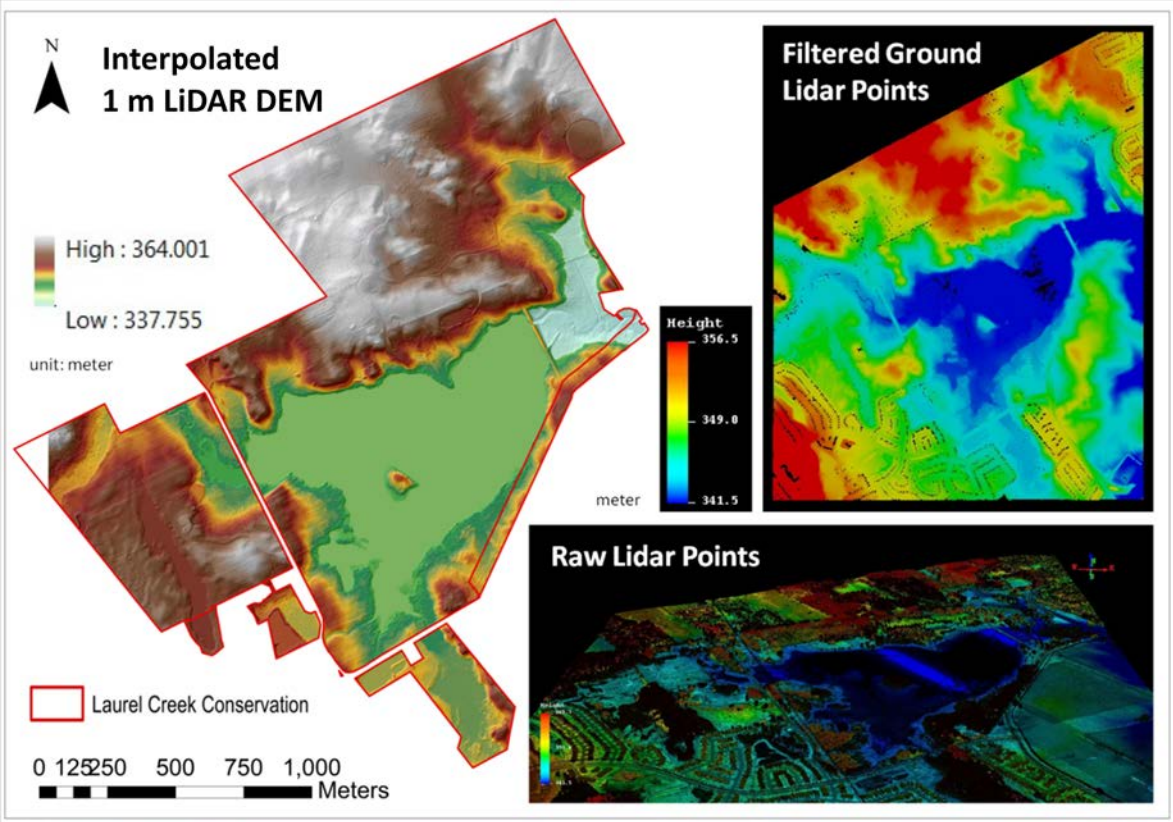

Figure 4.9 1 m LiDAR-derived DEM in the Laurel Creek Conservation area, with raw LiDAR points in right-bottom, and ground LiDAR points in right-top.

In an attempt to validate the LiDAR-derived DEM, the conventional 10 m DEM was compared with the resampled 10 m LiDAR-derived DEM. Total 100 points were randomly selected on the land of the Laurel Creek Conservation area. The root-mean-squre deviation (RMSD) of the elveation values extracted from the conventional 10 m DEM and the LiDAR-derived 10 m DEM at these points was 0.96487 m. Given the accuracy of the conventional 10 m DEM (vertical reliable: +/- 5 m), the value of the RMSD can be deemed as small enough to support the claim that the LiDAR-derived DEM has a high accuracy that could satisfy the need of this study.

Improvements in F and H

The FCM clustering results of experiments in Group C in Unit 1 area based on

LiDAR-derived DEM include information on the partition coefficient (F) and entropy (H) and fuzzy membership value maps. The results confirm the observation in Section 4.1.1 that with the increase of the cluster number, F decreases and H increases gradually. The improvements in partition coefficient ($\Delta F = F(c) - F(c + 1)$) and the entropy ($\Delta H = H(c) - H(c + 1)$) over adjacent clusters across m of 1.25, 1.5, 1.75, 2, and 2.25 are demonstrated in Figures 4.10, 4.11, 4.12, 4.13, and 4.14, respectively. Each figure has six graphs showing ΔF and ΔH at six grid sizes. ΔH reaches its local maximum or minimum values and ΔF reaches its local minimum or maximum values at the same number of clusters, respectively.

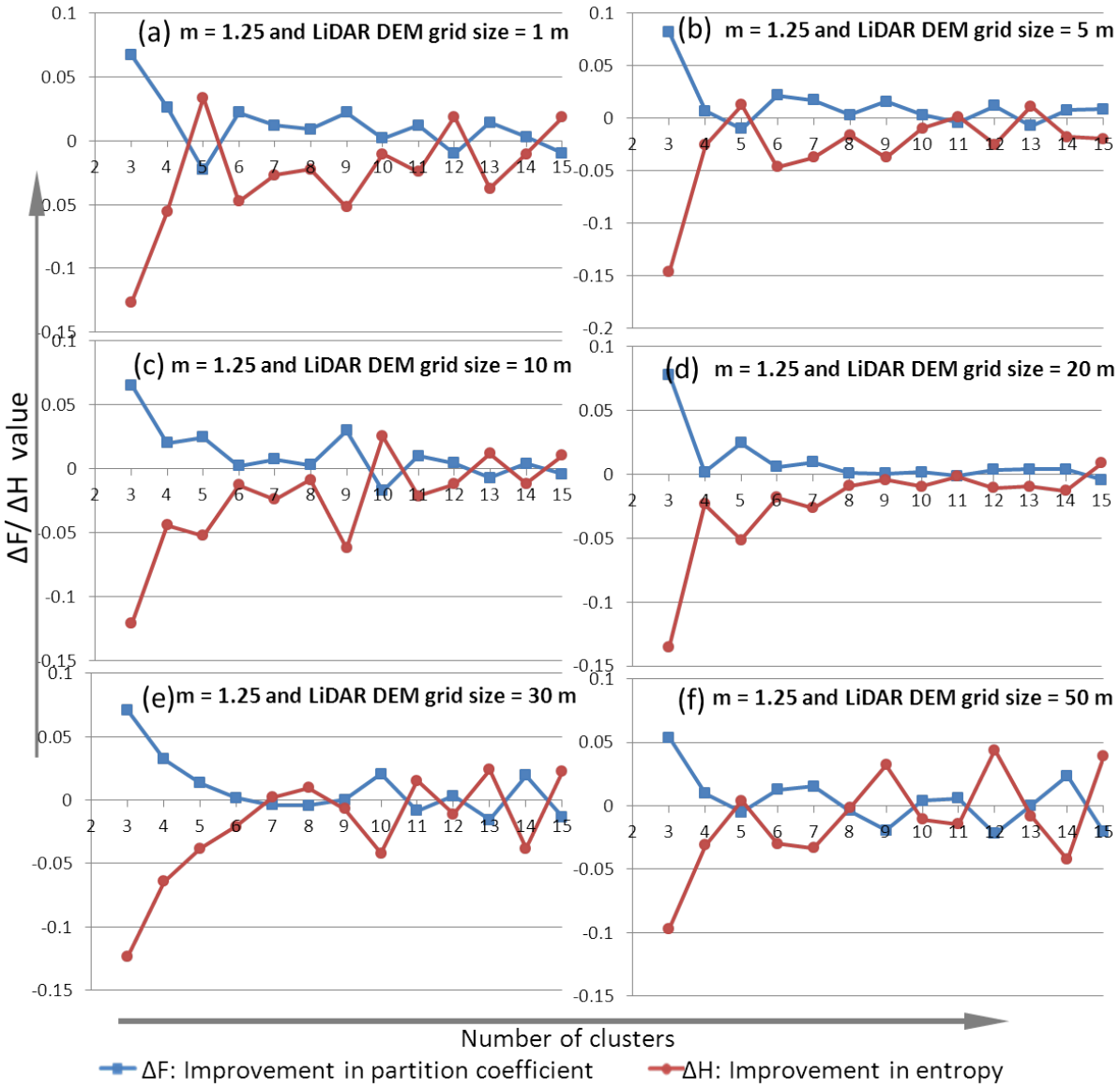

Figure 4.10 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 1 m, (b) 5 m, (c) 10 m, (d) 20 m, (e) 30 m, and (f) 50 m LiDAR-derived DEMs, using the algorithm with *m* of 1.25 in Unit 1 area.

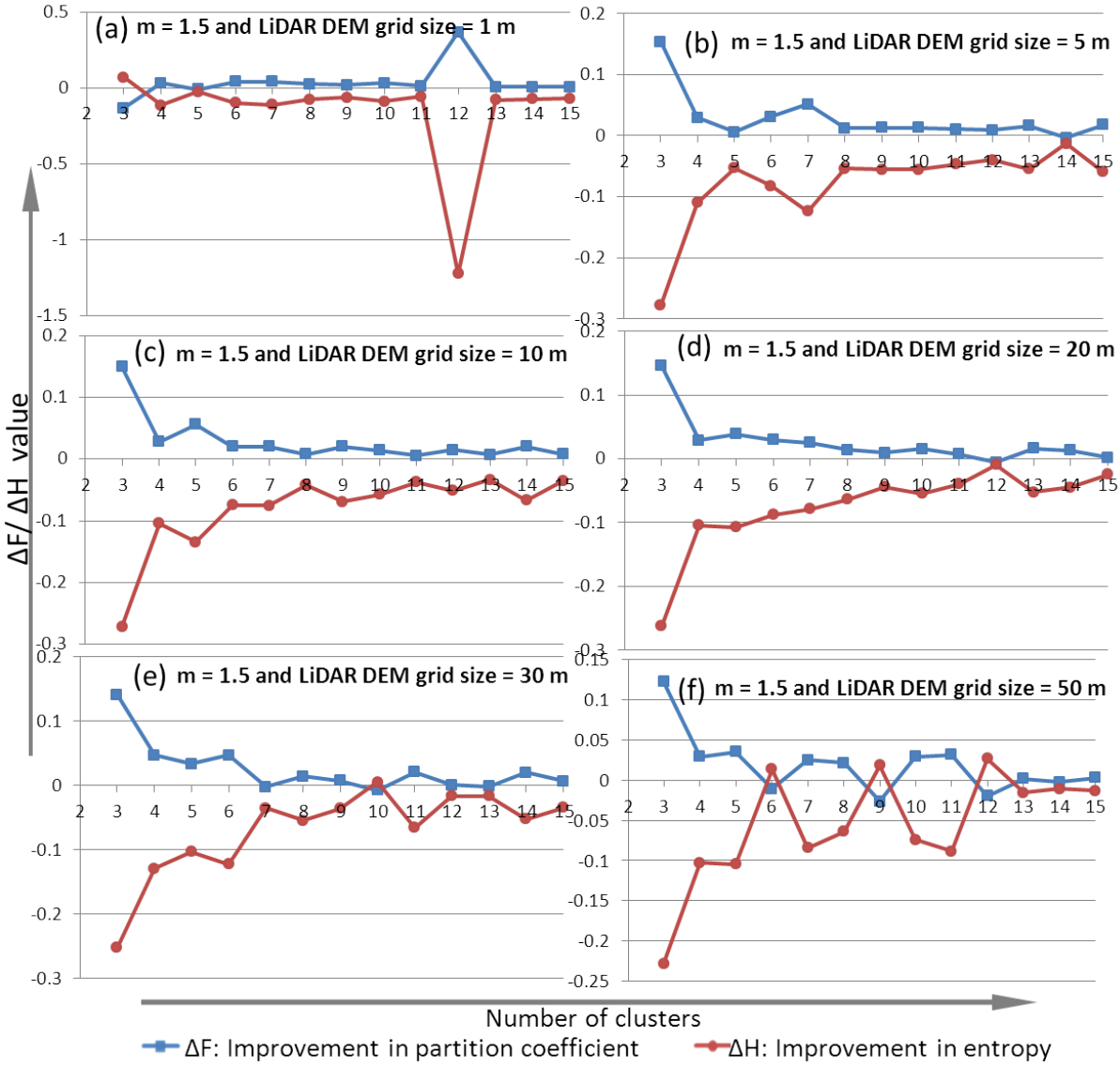

Figure 4.11 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 1 m, (b) 5 m, (c) 10 m, (d) 20 m, (e) 30 m, and (f) 50 m LiDAR-derived DEMs, using the algorithm with *m* of 1.5 in Unit 1 area.

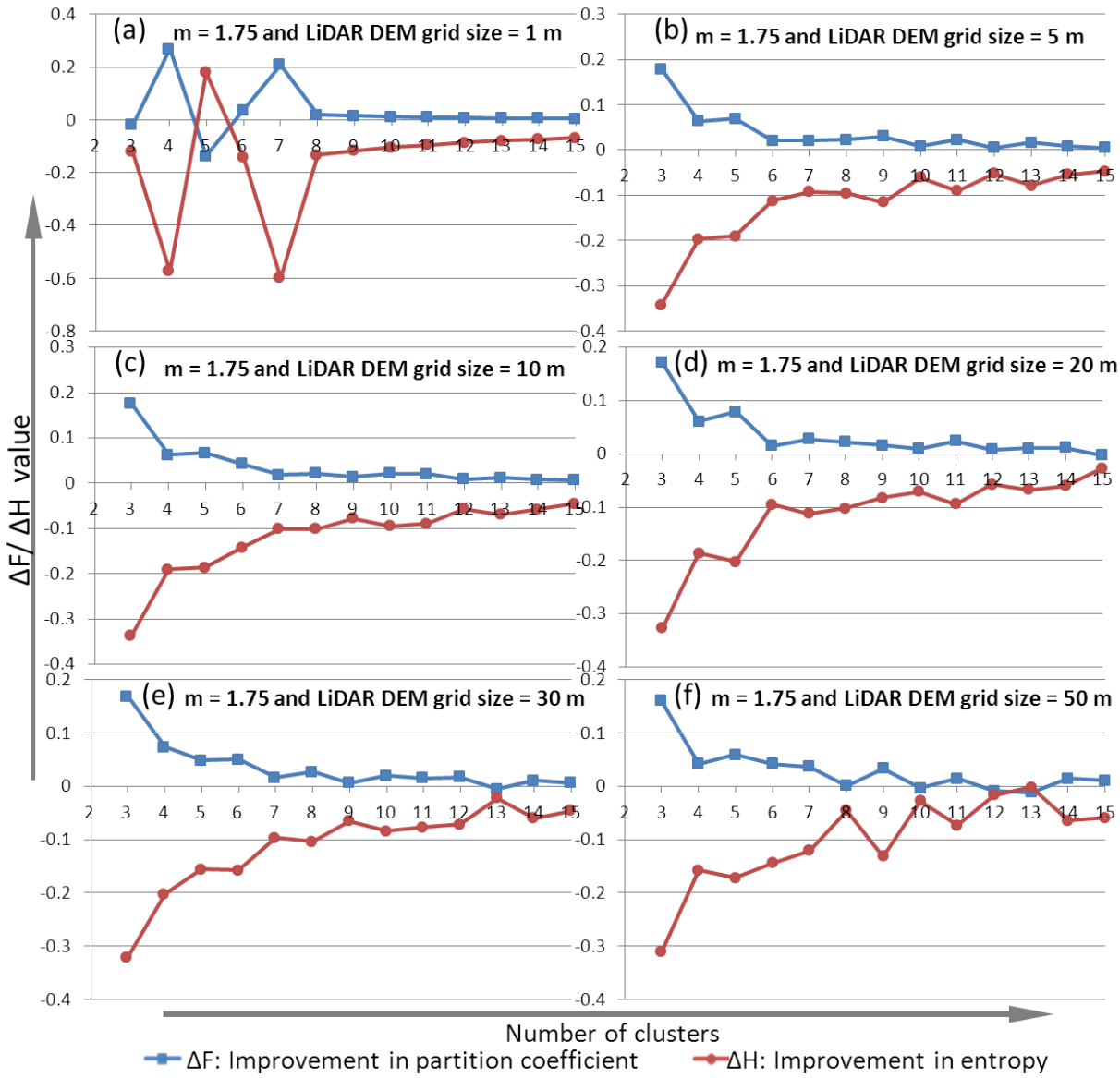

Figure 4.12 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 1 m, (b) 5 m, (c) 10 m, (d) 20 m, (e) 30 m, and (f) 50 m LiDAR-derived DEMs, using the algorithm with *m* of 1.75 in Unit 1 area.

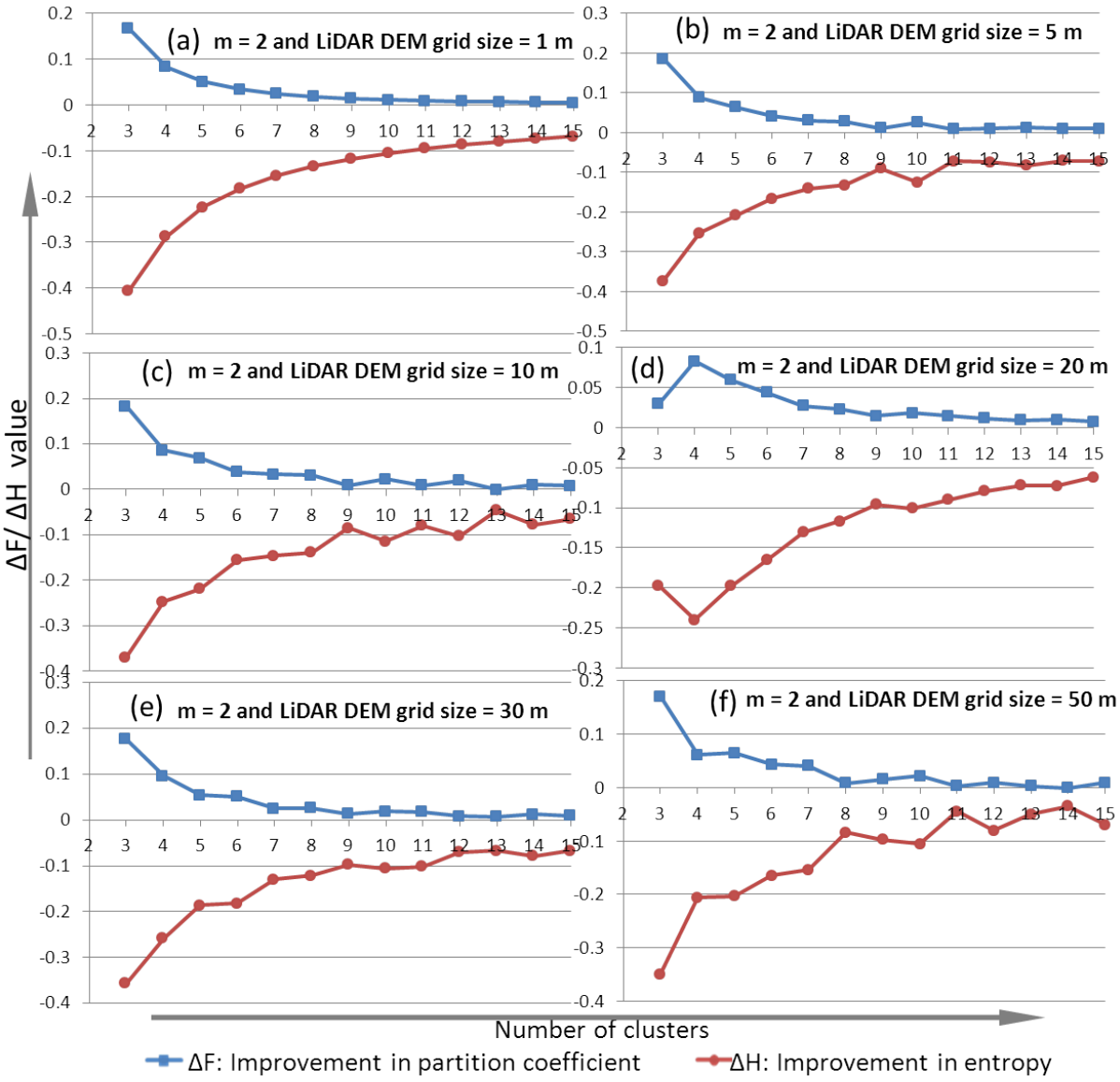

Figure 4.13 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 1 m, (b) 5 m, (c) 10 m, (d) 20 m, (e) 30 m, and (f) 50 m LiDAR-derived DEMs, using the algorithm with *m* of 2 in Unit 1 area.

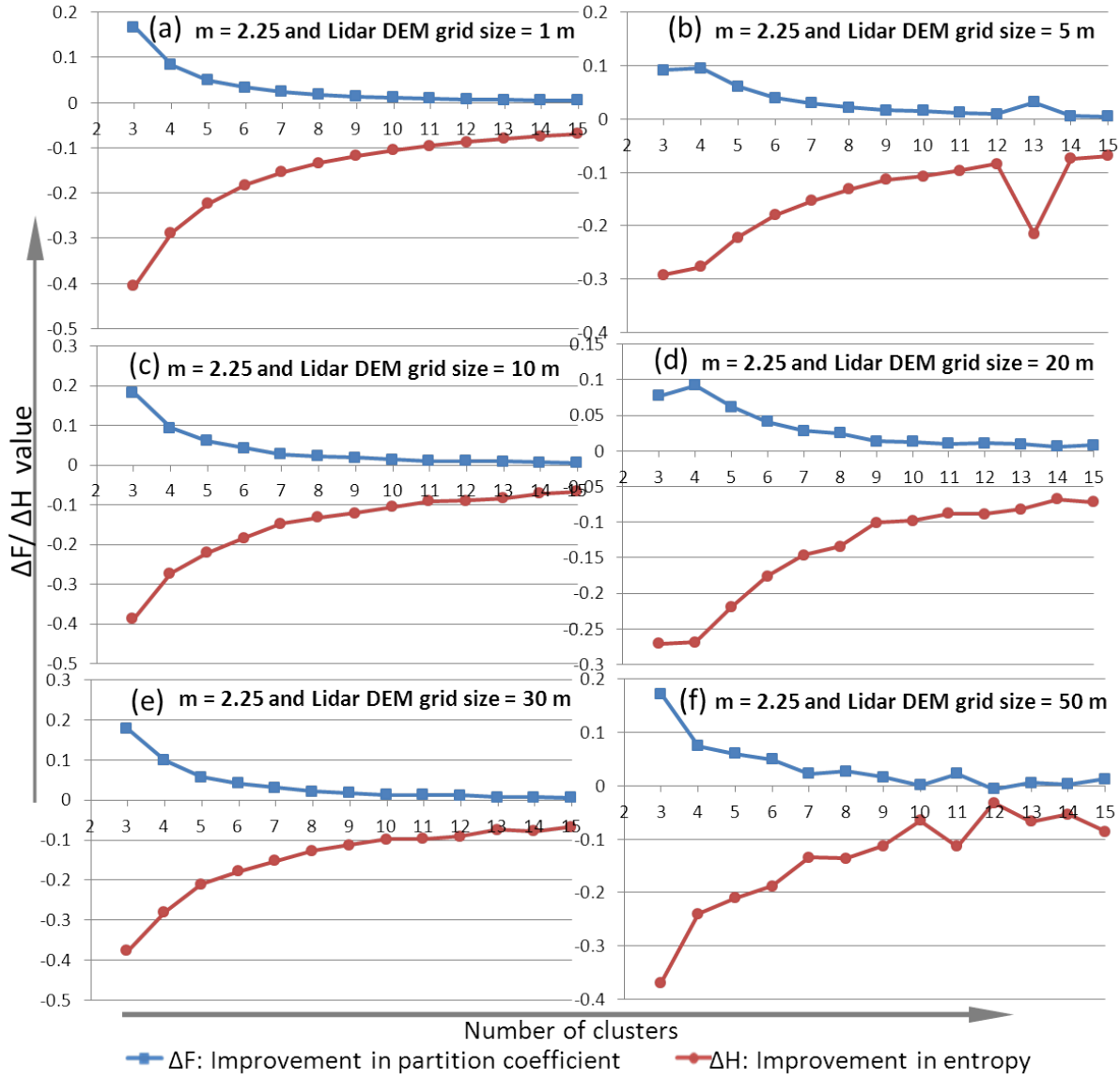

Figure 4.14 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 1 m, (b) 5 m, (c) 10 m, (d) 20 m, (e) 30 m, and (f) 50 m LiDAR-derived DEMs, using the algorithm with *m* of 2.25 in Unit 1 area.

Figures 4.10-4.14 confirm the finding that increasing *m* from 1.25 to 2.25 generally weakens the fluctuations in ΔF and ΔH at each resolution. The variation of ΔF and ΔH with number of clusters are different at each resolution. Increasing grid size from 1 m to 20 m generally mutes the fluctuations in ΔF and ΔH. When the grid size is 30 m, the fluctuations in ΔF and ΔH become bigger with smaller values of *m*, e.g., 1.25, 1.5, and 1.75, but generally follow the trend from 1 m to 20 m with *m* of 2 and 2.25. When the grid size is 50

m, the fluctuations in ΔF and ΔH are obviously enhanced.

The finding at 30 m resolution using LiDAR-derived DEM is different from that observed using conventional DEM shown in Section 4.1.1, while the trend from 1 m to 20 m using LiDAR-derived DEM agrees with the trend from 10 m to 30 m using conventional DEM. The finding at 50 m resolution of LiDAR-derived DEM corresponds to the clear change observed in 50 m conventional DEM.

Table 4.2 summaries the possible optimal numbers of clusters with different *m* and at different grid sizes, when ΔF reaches its local minimum value and ΔH reaches its local maximum value.

Table 4.2 Possible optimal numbers of clusters based on 1 m, 5 m, 10 m, 20 m, 30 m, and 50 m LiDAR-derived DEMs across five algorithms with *m* of 1.25, 1.5, 1.75, 2, and 2.25 in Unit 1 area.

m \ si-ze	1 m			5 m			10 m					20 m			30 m			50 m		
1.25	5	10	12	5	8	13	4	6	8	10	13	4	6	11	11	13		5	9	12
1.5	5			5	14		4	8	11	13		9	12		5	7	10	6	9	12
1.75	5			10	12		9	12				4	6	10	7	9	13	4	8	10
2				9			9	11	13						9			8	11	
2.25																		10	12	

It is likely that the choices of optimal number of clusters increase when resolution of LiDAR-derived DEM is aggregated from 1 m to 10 m and then decrease from 10 m to 20 m. When using conventional DEM, the choices for optimal number of clusters decrease with the grid size increasing from 10 m to 30 m, while this trend is true to LiDAR-derived DEM only from 10 m to 20 m. When using LiDAR-derived DEM at 30 m, there is no significant decline in the choices of optimal number of clusters as the one observed using conventional DEM. However, a sudden increase at 50 m is observed in both conventional and

LiDAR-derived DEM. In addition, the choices of optimal number of clusters in LiDAR-derived DEM are more than those in conventional DEM at 10 m, 20 m, 30 m, and 50 m resolutions, respectively.

The differences between the results using conventional and LiDAR-derived DEM may be explained by the differences between the environmental covariates. At the same resolution, LiDAR-derived DEM may delineate land-surface parameters in more detail than conventional DEM does, which stems from the different degrees of relief variation in conventional and LiDAR-derived DEMs at the same resolution (Remmel et al., 2008; Vaze et al., 2010). In Remmel et al. (2008), LiDAR-derived DEM provides a much more complete and accurate representation of the hydrological processes than the conventional DEM does. Vaze et al. (2010) also showed that a more detailed delineation of watersheds is given by the 25 m LiDAR-derived DEM but not the 25 m conventional DEM. The difference may then propagate to the FCM clustering of environmental covariates, resulting in more variations of environmental covariates generated from LiDAR-derived DEM, and more choices of optimal number of clusters found. Moreover, both results in conventional and LiDAR-derived DEM have shown a transition at 50 m that could be deemed as a threshold for the fuzzy soil inference in Unit 1 area.

Hardened Class Maps

Hardened class maps with *m* of 1.5 based on LiDAR-derived DEM at six resolutions are demonstrated in Figure 4.15. As aforementioned, an *m* equal to 1.5 was used due to its representing a good compromise with respect to the improvements and stability of the

partition. The numbers of clusters (*c*) are 5, 5, 13, 12, 10, and 12 with resolutions of 1 m, 5 m, 10 m, 20 m, 30 m, and 50 m, respectively. These numbers were selected from Table 4.2 based on which the partitions could be deemed as optimal.

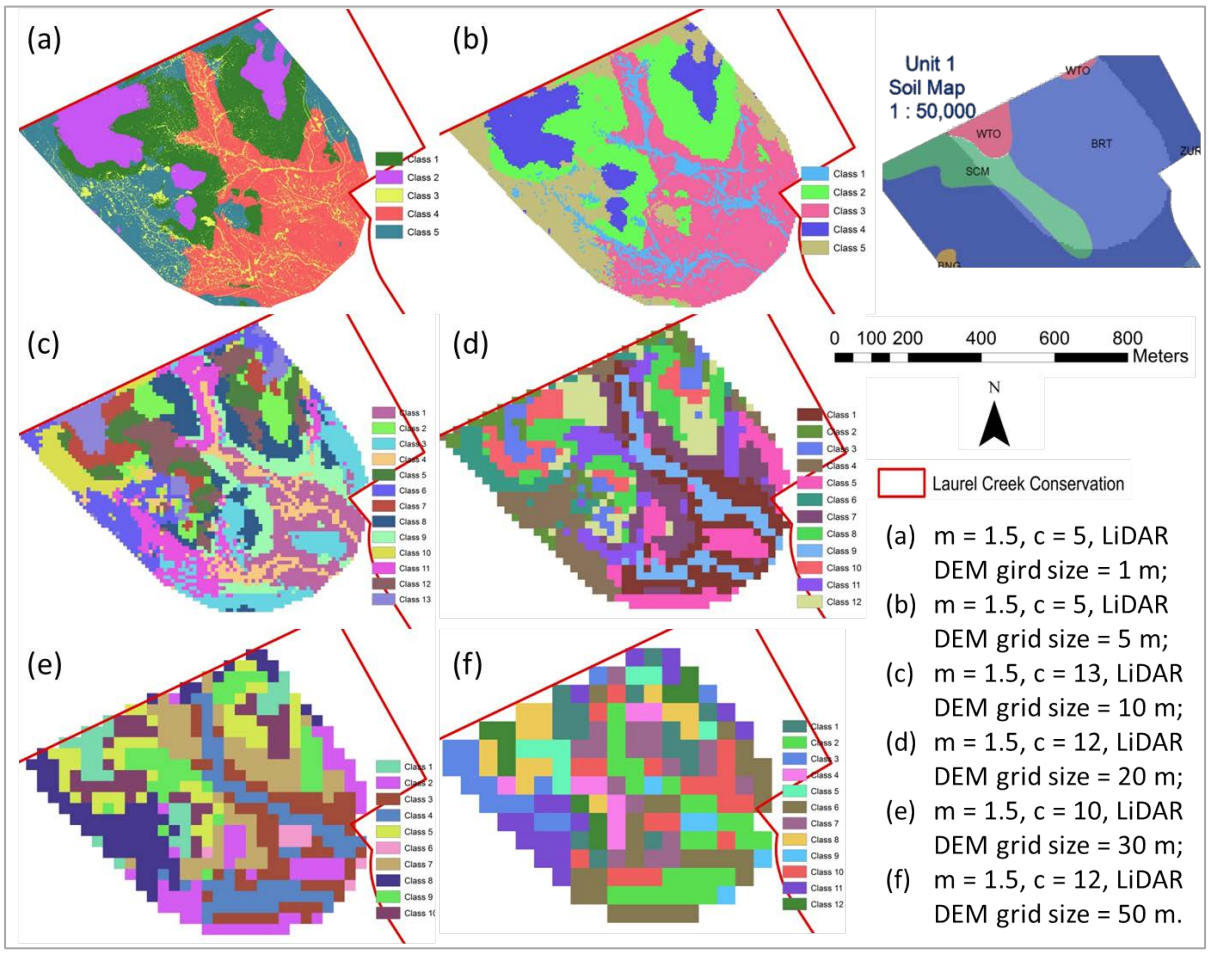

Figure 4.15 Hardened class maps based on LiDAR-derived DEM with resolution of and number of clusters *c* of (a) 1 m and 5, (b) 5 m and 5, (c) 10 m and 13, and (d) 20 m and 12, (e) 30 m and 10, (f) 50 m and 12, respectively, using the algorithm with *m* of 1.5 in Unit 1 area.

Although each of the six hardened class maps presents different partitions of Unit 1 area, a consistent pattern can still be seen through them. This consistency corresponds to the hardened class maps using conventional DEM shown in Section 4.1.1, which could confirm the domination of some landforms in this area. The graphs (a) and (b) at 1 m and 5 m resolution in Figure 4.15 carry very similar patterns based on the use of the same

number of clusters. Class 3 in graph (a) shown in yellow colour occupies thinner areas at the bottom than the corresponding Class 1 in graph (b) in blue. These areas may be interpolated into a channel that could develop soil with more water content. To determine which of the two graphs is more reasonable in representing such terrain feature, a good way may be to examine it in field by measuring the real scale over which it develops. Thus, more information is needed when selecting resolution of LiDAR-derived DEM to perform landform and soil classification.

At 20 m resolution, the hardened class map (d) in Figure 4.15 using LiDAR-derived DEM and the map (b) in Figure 4.7 using conventional DEM both have 12 classes, and thus it is of meaning to compare these two maps in terms of clustering patterns. The former one shows less scattering than the latter one and therefore is likely to provide clearer and continuous boundaries of clusters. Then, a better soil classification result could be produced from LiDAR-derived DEM with resolution of 20 m. This may correspond to many previous studies indicating that LiDAR-derived DEM could lead to improvements in the accuracy of geomorphic and hydrologic classification and prediction. (e.g., MacMillan et al., 2003; Murphy et al., 2008; Vaze et al., 2010). However, it is still impossible to claim that LiDAR-derived DEM at higher resolution and accuracy could perform better in soil classification than that at lower resolution and accuracy. For example, being both classified into 12 clusters, graph (d) in Figure 4.15 at 20 m presents more detailed variations of clusters than graph (f) at 50 m does, while a clear and meaningful classification may be seen from the coarser graph (f). A related observation in MacMillan et al. (2010) indicated

that finer resolution DEM with 5 m and 10 m resolution presents challenges for distinguishing short range from long range information, but lower resolution may perform better.

Consequently, it can be seen that the level of dependency of fuzzy clustering to resolution varies with DEM sources, conventional and LiDAR-derived DEM. When using LiDAR-derived DEM, the choices of optimal numbers of clusters increase with the resolution aggregated from 1 m to 10 m, decrease from 10 m to 20 m, generally keep stable when it comes to 30 m, and suddenly increase at 50 m. While using conventional DEM, the choices decrease from 10 m to 30 m, and increase at 50 m. At each resolution, the choices using LiDAR-derived DEM are more than those using conventional DEM. These variations may stem from the differences of the two sources of DEM in representing landscape. It has to be admitted that no perfect DEM resolution exists (Claessens et al., 2005) either in conventional DEM nor LiDAR-derived DEM. However fuzzy soil mapping could still profit from an explicit procedure on analyzing a fit resolution at which the most possible detailed and meaningful information on landscape could be gained.

4.2 FCM Clustering Results in Unit 2 Area

4.2.1 Use of Conventional DEM

F and H

In Group B, two experiments across five algorithms of weighting exponent of m = 1.25, 1.5, 1.75, 2, and 2.25 at 10 m resolution were conducted in Unit 2 area. One experiment

employed both the five terrain covariates and the vegetation covariate, NDVI, while the other one employed only the five terrain covariates. Graphical displays of the partition coefficient (*F*) and entropy (*H*) against the number of clusters are illustrated in Figure 4.16.

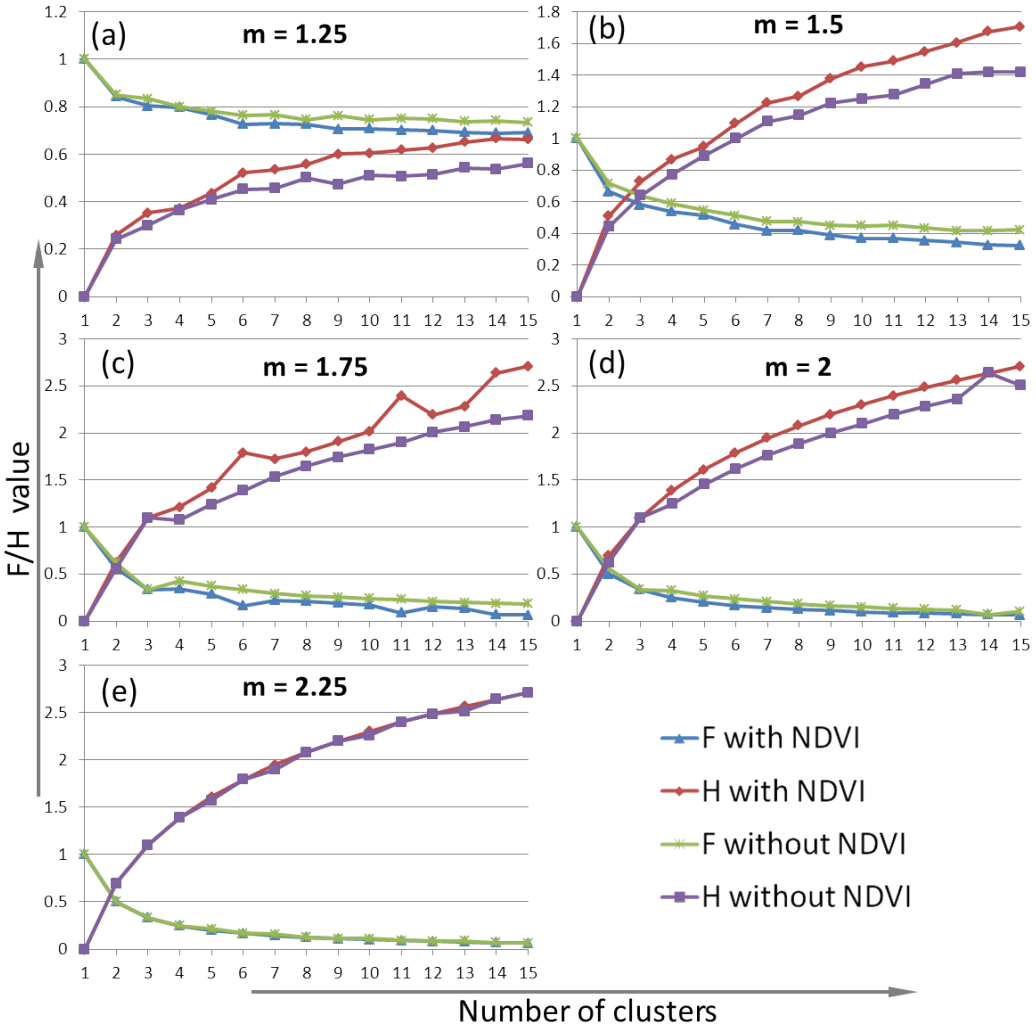

Figure 4.16 Partition coefficient (*F*) and entropy (*H*) plotted against the number of clusters based on 10 m conventional DEM across five algorithms with *m* of (a) 1.25, (b) 1.5, (c) 1.75, (d) 2, and (e) 2.25, where environmental covariates databases were constructed with NDVI and without NDVI in Unit 2 area.

Both of the two experiments (with NDVI and without NDVI) show that with the increase of the cluster number, *F* decreases and *H* increases generally, as well, increasing *m* leads to lower *F* and higher *H*. These results confirm the observations in Section 4.1 that the overlap between clusters increases as the number of clusters increases and the

classification certainty reduces as the value of *m* increases.

Some differences between the experiments with NDVI and without NDVI exist. First, when *m* grows from 1.25 to 2, the values of *F* with NDVI are smaller than those without NDVI at the same number of clusters, and the values of *H* with NDVI are larger than those without NDVI at the same number of clusters. However, when *m* comes to 2.25, there are no significant differences between the values of *F* or *H* with and without NDVI. Second, when *m* is 1.25, 1.5, 2, and 2.25, *F* with and without NDVI tend to decrease at the same speed, and *H* with and without NDVI tend to increase at the same speed as well. Both the increase and the decease are gradual. However, when *m* is 1.75, *F* and *H* with NDVI show some obvious fluctuations, while *F* and *H* without NDVI do not show such fluctuations. These differences may result from the change of dimensions in the environmental covariates space. It is also evident that such differences have dependence on the value of *m*.

Improvements in F and H

It is also necessary to examine the improvements in *F* and *H* with numbers of clusters graphically in order to determine the optimal numbers of clusters for experiments in Unit 2 area. The improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and the entropy ($\Delta H = H(c) - H(c+1)$) over adjacent clusters across *m* of 1.25, 1.5, 1.75, 2, and 2.25 at 10 m resolution are demonstrated in Figure 4.17. The figures on the left side show the results obtained with NDVI, while results shown on the right side are without NDVI. ΔH reaches its local maximum or minimum values and ΔF reaches its local minimum or maximum values at the same number of clusters, respectively.

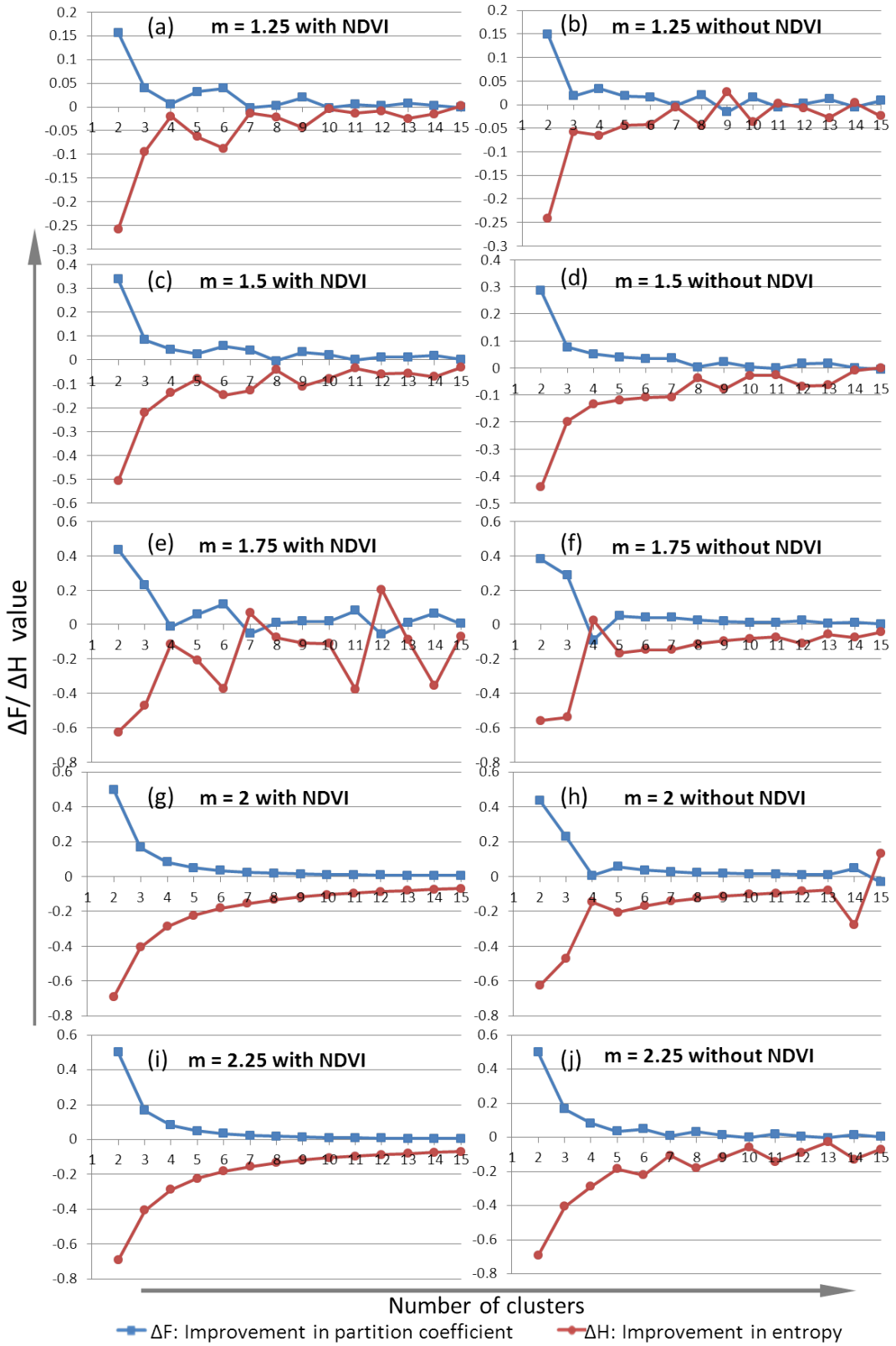

Figure 4.17 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on 10 m conventional DEM across the algorithms with *m* of and the construction of environmental covariates database (a) 1.25 with NDVI, (b) 1.25 without NDVI, (c) 1.5 with NDVI. (d) 1.5 without NDVI, (e) 1.75 with NDVI, (f) 1.75 without NDVI, (g) 2 with NDVI, (h) 2 without NDVI, (i) 2.25 with NDVI, and (j) 2.25 without NDVI, respectively, in Unit 2 area.

Figure 4.17 shows that the variations of ΔF and ΔH with the number of clusters are different when using NDVI and not using NDVI as a covariate. When *m* is 1.25 and 1.5, ΔF and ΔH with NDVI and without NDVI demonstrate similar fluctuations, while with *m* of 1.75, ΔF and ΔH with NDVI fluctuate strongly but those without NDVI tend to be stable. When *m* is 2 and 2.25, ΔF and ΔH with NDVI are very smooth, while those without NDVI undergo some fluctuations. Through these observations, it is necessary to examine the possible optimal numbers of clusters at which ΔF reaches its local minimum value and ΔH reaches its local maximum value. Table 4.3 summaries the possible optimal numbers of clusters when using NDVI and not using NDVI across *m* of five different values.

Table 4.3 Possible optimal numbers of clusters based on 10 m conventional DEM across five algorithms with *m* of 1.25, 1.5, 1.75, 2, 2.25, using environmental databases with and without NDVI in Unit 2 area.

m	1.25				1.5			1.75			2	2.25	
With NDVI	4	7	10		5	8	11	4	7	12			
Without NDVI	3	7	9	14	8			4			4	5	7

It can be seen that adding NDVI into the environmental covariates database generally reinforces the fluctuation of ΔF and ΔH across *m* of lower values, while this effect becomes negative with higher values of *m*. There are more choices of the optimal number of clusters when including NDVI than the situation when not including NDVI across *m* of lower values (< 2), while when *m* is higher, 2 and 2.25, the experiment using NDVI cannot reveal any optimal number of clusters and that without NDVI still shows some choices. As aforementioned, the larger *m*, 2 and 2.25, will result in more unstable classification. The results with *m* of lower values, such as 1.5, is more reliable. Thus, from this point of view, one then can argue that FCM clustering with NDVI in Unit 2 area based on conventional

DEM provides more optimal numbers of clusters than that without NDVI, and it is also possible that the introducing of more choices will lead to more effective classification results.

Many would contend that land-surface parameters are the most useful predictors for soil mapping (McKenzie et al., 2000; McBratney et al., 2003), while the best performance may take place where water and material flows are strongly governed by relief, e.g., hilly area, (Huggett, 1975; Park and Vlek, 2002). In other words, when the relief is relatively flat, other environmental covariates such as vegetation may take control in the prediction. That is much likely true for the Unit 2 area based on conventional DEM where the elevation range is only about 20 m (from ~340 m to ~360 m). Field investigation also indicates that the natural vegetation coverage in this area tends to follow the soil variation, which can be explained by the impact of the content of sand and clay on the vegetation growing. Hence, based on this study it is suggested to embody vegetation information when performing soil classification on the use of FCM clustering in this area.

Hardened Class Maps

Examining the hardened class maps for Unit 2 area can also provide useful information on the effectiveness of vegetation covariate. As shown in Table 4.3, using 8 as the number of clusters is one of the optimal options not only for the case with NDVI but also for that without NDVI. Thus, the hardening was done on the fuzzy membership value maps in Unit 2 area with *m* of 1.5 and *c* of 8 based on 10 m conventional DEM. The hardened class maps are shown in Figure 4.17, compared with the 1:50,000 conventional soil map in right top.

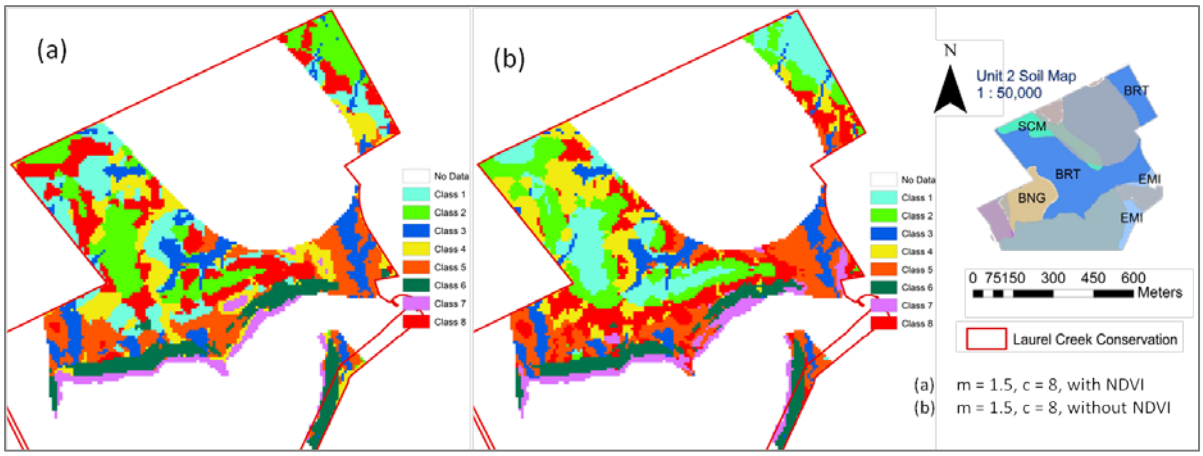

Figure 4.18 Hardened class maps based on 10 m conventional DEM and the construction of environmental covariates database (a) with NDVI and (b) without NDVI, using the algorithm with *m* of 1.5 and number of clusters *c* of 8 in Unit 2 area.

As shown in Figure 4.18, graphs (a) and (b) have the same number of classes and similar patterns, while the two graphs tend to differentiate each other by the degree of detail in the distribution of these classes. For example, in the central area, graph (a) contains more details than graph (b) does. This variance may be explained by the addition of NDVI. NDVI is able to reflect the two distinct vegetation communities, grassland and forest, found in Unit 2 area. Thus, graph (a) tends to reveal more information on vegetation communities than graph (b). In field investigation, in grassland area soils were easy to dig and contain more sand, while in forest area soils were always hard to dig due to more plant roots and clay content. This difference may reveal the association between vegetation covariate and soil series. However, answering the question of whether graph (a) is more reasonable and useful than graph (b) needs further studies on validation. In addition, both of the graphs show more detailed variations than the conventional soil map, suggesting that fuzzy logic-based digital soil mapping may produce more detailed soil maps.

The findings may indicate combining NDVI with land-surface parameters may

improve soil classification to some extent. This conclusion confirms the observations in McKenzie and Ryan (1999) and Park and Vlek (2002) that vegetation occurrence in addition to land-surface parameters improves the performance of soil prediction model. Although further studies on soil inference and validation in Unit 2 area are needed to better support this conclusion, it still bears significance to the initial work of the purposive sampling design for fuzzy logic-based digital soil mapping in this area. Zhu et al. (1997) implied that including vegetation information in an environmental covariate database may provide useful information for identifying soil attributes where vegetation can exhibit a great dependence on the soil conditions. However, in practice, many failed to include such information because the vegetation in the study area was always disturbed by human activities, for example, planted with single type of trees or crops. This is true to many soil surveys (e.g., Zhu et al., 2008a). In this case, other environmental covariates are expected to contribute more to gain better classification results.

4.2.2 Use of LiDAR-derived DEM

Improvements in F and H

The improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and the entropy ($\Delta H = H(c) - H(c+1)$) over adjacent clusters across m of 1.25, 1.5, 1.75, 2, and 2.25 based on 10 m LiDAR-derived DEM in experiments of Group D are demonstrated in Figure 4.19. The figures on the left side show the results of using NDVI in the environmental covariates database, while results on the right side are without NDVI.

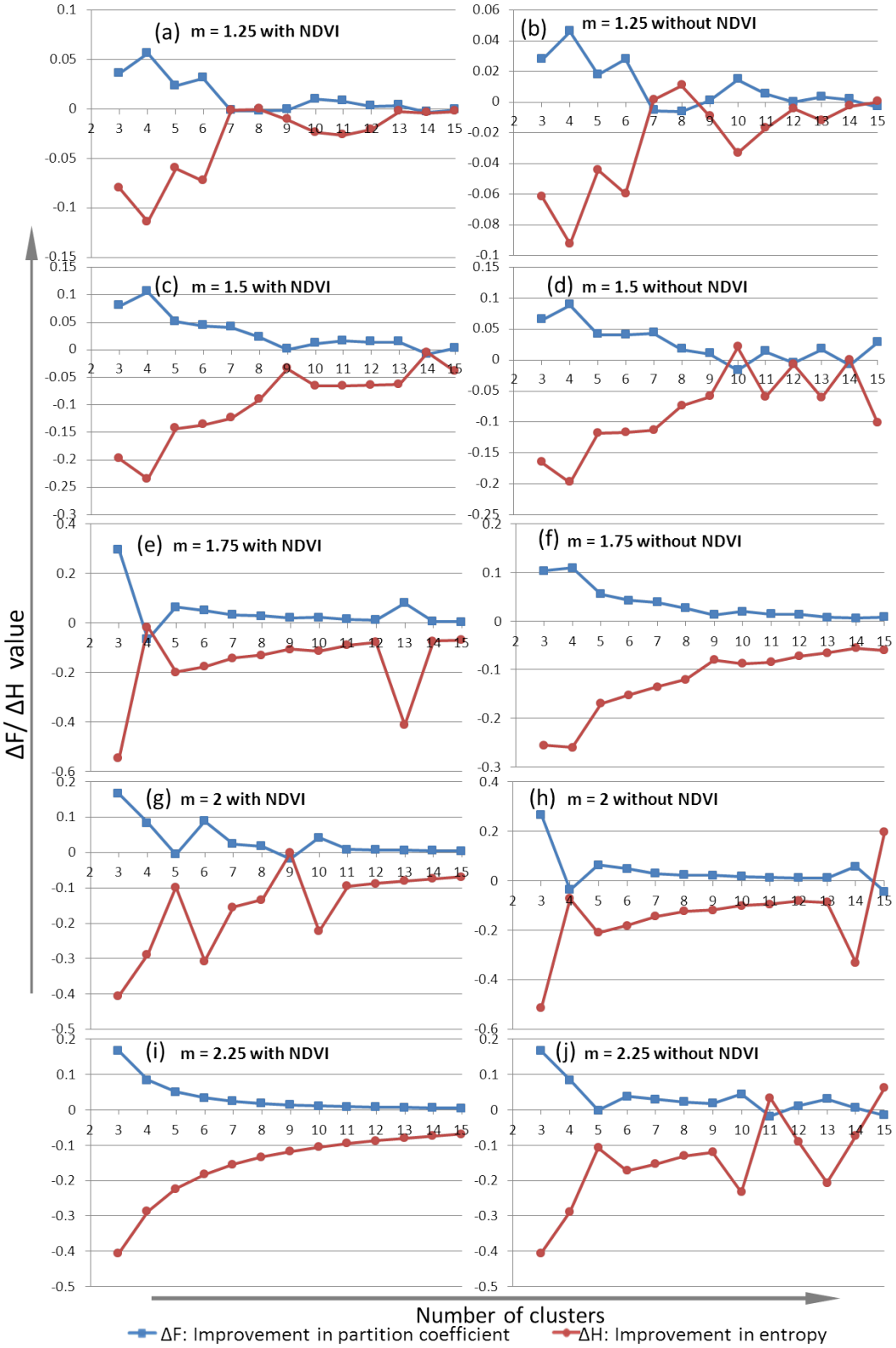

Figure 4.19 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on 10 m LiDAR-derived DEM across the algorithms with *m* of and the construction of environmental covariates database (a) 1.25 with NDVI, (b) 1.25 without NDVI, (c) 1.5 with NDVI. (d) 1. 5 without NDVI, (e) 1.75 with NDVI, (f) 1.75 without NDVI, (g) 2 with NDVI, (h) 2 without NDVI, (i) 2.25 with NDVI, and (j) 2.25 without NDVI, respectively, in Unit 2 area.

Figure 4.19 shows that the variations of ΔF and ΔH with the number of clusters when using NDVI are different from not using NDVI as a covariate. When *m* is 1.25, 1.5 or 2.25, the variations of ΔF and ΔH with NDVI are steadier than those without NDVI, while with *m* of 1.75 or 2 the variations of ΔF and ΔH without NDVI become steadier. This finding disagrees with the results using conventional DEM in Section 4.2.1 that adding NDVI into the environmental covariates database generally reinforces the fluctuation of ΔF and ΔH across *m* of lower values and mutes it with higher values of *m*. Possible optimal numbers of clusters at which ΔF reaches its local minimum value and ΔH reaches its local maximum value are summarized in Table 4.4.

Table 4.4 Possible optimal numbers of clusters based on 10 m LiDAR-derived DEM across five algorithms with *m* of 1.25, 1.5, 1.75, 2, and 2.25, using environmental databases with and without NDVI in Unit 2 area.

m	1.25		1.5			1.75	2		2.25	
With NDVI	5		9	14		4	5	9		
Without NDVI	5	12	10	12	14	9	4		5	11

It is likely to find more choices of optimal numbers of clusters when not using NDVI as an environmental covariate than those found in the situation with NDVI across *m* of 1.25, 1.5, and 2.25, while with *m* of 1.75 and 2 the choices using NDVI are not less than those without NDVI. This finding does not tie in with the observation in conventional DEM in Group B where involving NDVI leads to more optimal numbers of clusters with lower values of *m*, and less with larger values of *m*. This discrepancy may arise from the differences in the level of detail of environmental covariates. Terrain covariates derived from LiDAR data would contain more detailed variances than those derived from conventional DEM (MacMillan et al., 2010). However, NDVI datasets employed to

construct the environmental covariates databases in Groups B and D are the same. It is possible that the difference of the level of detail between LiDAR-derived terrain covariates and NDVI does not match with the one between conventional DEM-derived terrain covariates and NDVI. It may be this difference that leads to the disagreement in the findings in Groups B and D. In addition, this discrepancy may be attributed to the differences of the dominant terrain features identified by LiDAR and conventional DEM-derived environmental configurations (Vaze et al., 2010). These differences then could have impacts on the contribution of NDVI to the FCM clustering.

Hardened Class Maps

It is also helpful to examine hardened class maps for recognizing the contribution of NDVI on the use of LiDAR-derived DEM. Hardened class maps with NDVI and without NDVI based on LiDAR-derived DEM and *m* of 1.5 in Unit 2 area are illustrated in Figure 4.20.

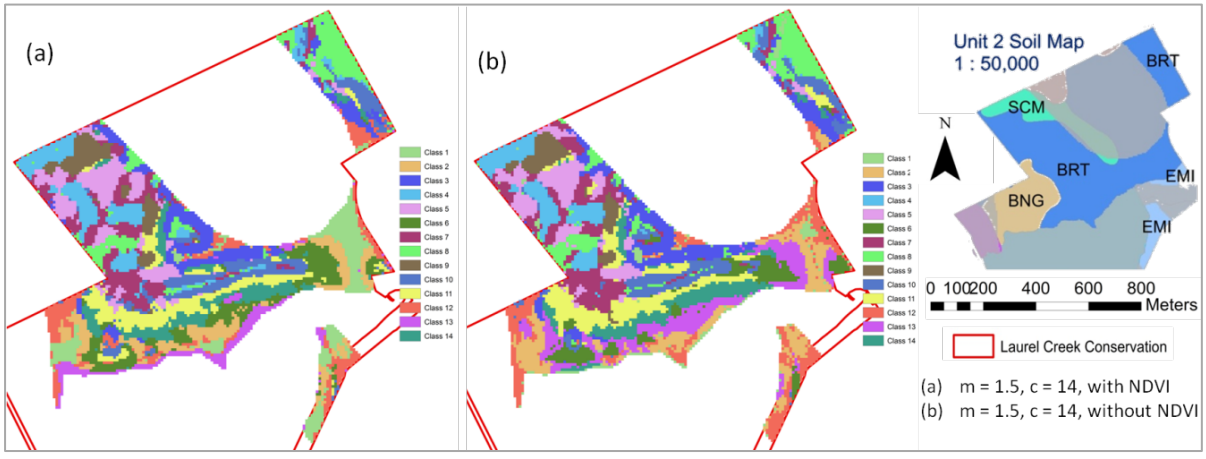

Figure 4.20 Hardened class maps based on 10 m LiDAR-derived DEM and the construction of environmental covariate database (a) with NDVI and (b) without NDVI, using the algorithm with *m* of 1.5 and number of clusters *c* of 14 in Unit 2 area.

Both of the two hardened class maps have 14 classes exhibiting similar patterns and show more detailed variances than the 1:50,000 conventional soil map, which confirm the observation in conventional DEM. Some slight differences of the patterns include the southeast area where the level of detail with NDVI is higher than the one without NDVI, which corresponds to the conclusion in Group B that adding NDVI leads to more detailed variances in hardened class maps. While in the middle-east area, the pattern without NDVI shows more variances than the one with NDVI, which is obviously different from the observation in conventional DEM (Figure 4.18). In this particular middle-east area, the 10 m LiDAR-derived and conventional DEMs carry very dissimilar information on elevation that could lead to such an obvious discrepancy in experiments of Groups B and D. However, as in Group B that more examinations and validations are needed to identify the map with more meaningful clustering information, the same applies to Group D.

Consequently, the contribution of NDVI to the purposive sampling design varies with DEM sources, which could still be ascribed to the difference between the levels of detail in DEM. Although findings from experiments based on LiDAR-derived and conventional DEMs are slightly different, it is still suggested to involve NDVI as a vegetation covariate in the fuzzy soil inference as long as the study area could demonstrate an association between soil and vegetation. The areas altered by human activities such as agriculture lands with uniform vegetation type, notwithstanding, might not carry such an association, where other environmental covariates may assist the fuzzy soil inference better.

4.3 Interpretation of Membership Values

4.3.1 Use of Conventional DEM

In fuzzy logic-based soil inference, it is suggested that if fuzzy clusters can be correctly interpreted as soil series, then soil maps can be inferred using the information of the cluster centroids (English, 2001). Cluster centroids are in the environmental covariates space with the membership values of 1, while in spatial space a cell with the membership value of 1 does not always exist. Cells with membership values very near 1 (e.g., > 0.9) are thus utilized to represent the most “typical” individuals of this cluster because they are the closest to the centroids. In this study, the results of the field sampling (Table 3.1) were employed to interpret the fuzzy membership values. An interpretation of fuzzy membership values at these points is possible to promote the understanding of the associations between fuzzy clusters and soil series in the study area.

Although fuzzy logic-based digital soil mapping avoids sharp boundaries in crisp form soil mapping, there could be confusion if the assigned membership values to two or more clusters at a cell are very similar (Burrough et al., 1997). Burrough et al. (1997) introduced a confusion index (CI) that is the ratio of the second to the first highest membership values at a cell to evaluate how well each individual observation in FCM clustering has been classified. As CI→0, the observation strongly associates with the first dominant class and there is little confusion; while as CI→1, there are small differences between the membership values to the two dominant classes, implying more confusion. When more confusion exists, the observation can be regarded as an intergrade between the two

dominant classes (in-between types). This index was calculated for sampled sites.

Table 4.5 presents the fuzzy membership values at the field sampling points in Unit 1 area with *m* of 1.5 and *c* of 9 based on 10 m conventional DEM in Group A. The calculated confusion index (CI) results are also recorded in Table 4.5.

Table 4.5 Fuzzy membership values and confusion index (CI) at sampled points based on 10 m conventional DEM using the algorithm with *m* of 1.5 and *c* of 9 in Unit 1 area (the first and second highest membership values in bold).

Membership value	Field Point ID and Soil_Code							
	4	5	6	7	8	9	10	11
	HIG	WTO	WTO	ZOR	FOX	WTO	CAD	WTO
Class 1	**0.581**	0.006	0.029	**0.015**	0.009	0.025	0.020	0.014
Class 2	0.039	0.030	**0.169**	0.001	0.025	0.027	0.033	**0.140**
Class 3	0.033	**0.127**	**0.283**	0.001	**0.530**	0.054	0.105	0.063
Class 4	0.007	0.025	0.023	0.000	0.037	0.004	0.047	0.109
Class 5	0.015	0.077	0.114	0.001	0.073	0.012	0.046	**0.483**
Class 6	**0.157**	0.057	0.148	0.002	0.059	**0.805**	**0.146**	0.034
Class 7	0.029	**0.660**	0.149	0.001	**0.201**	**0.064**	**0.580**	0.069
Class 8	0.017	0.017	0.080	0.001	0.065	0.008	0.020	0.085
Class 9	0.123	0.001	0.003	**0.977**	0.001	0.002	0.004	0.002
Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CI	0.270	0.192	0.597	0.015	0.379	0.080	0.252	0.290

In Table 4.5, Points 5, 6, 9, and 11 were all classified into WTO series by field survey, while their membership values to classes and their dominant classes vary largely. At Point 6, the confusion index (CI) is high (nearly 0.6) enough to be regarded as an intergrade between Classes 3 and 2 (Burrough et al., 1997). At Point 9, on the contrary, the CI value is close to 0 and thus there is little confusion about the dominant class, namely Class 6. For a site with high value of CI (e.g., Point 6), there will be more confusion about which class this site “really belongs”. This question may seem meaningless considering the “fuzziness”,

while such sites can be treated as intergrades lying between two classes in environmental covariates space and used to locate boundaries (Burrough et al., 1997; Arrell et al., 2007). In a further soil inference study, it is not recommended to use Point 6 to represent WTO series based on this partition because the high CI value implies this point may not be a "typical" WTO over environmental covariates space, while Point 9 with very low CI value is a better choice. Moreover, the very small membership value to the first dominant class (0.283) at Point 6 also hinders the use of this point to infer that class.

Among the CI values in Table 4.5, the lowest one is at Point 7, indicating that the dominant class (Class 9) at Point 7 can represent ZOR series (organic soils) well in soil inference. Point 7 was sampled at a low and flat site where the environmental configuration may be typical for organic soil forming. It is therefore suggested to map the distribution of organic soil in Unit 1 area according to the distribution of fuzzy membership values to Class 9.

Table 4.6 shows the fuzzy membership values at the field sampling points in Unit 2 area with *m* of 1.5 and *c* of 8 based on 10 m conventional DEM and NDVI in the first experiment of Group B. The calculated confusion index (CI) results are also recorded in Table 4.6.

Table 4.6 Fuzzy membership values and confusion index (CI) at sampled points based on 10 m conventional DEM and environmental covariates database with NDVI, using the algorithm with *m* of 1.5 and *c* of 8 and in Unit 2 area (the first and second highest membership values in bold).

Mem-bership value	Field Point ID and Soil_Code										
	1	2	3	12	13	14	15	16	17	18	19
	BRT	WTO	HUO	HUO	HUO	HUO	FOX	FOX	HUO	HUO	HUO
Class 1	**0.486**	**0.804**	**0.097**	**0.438**	**0.198**	**0.713**	**0.088**	**0.128**	0.166	0.104	0.026
Class 2	0.044	0.016	0.038	0.018	0.080	0.013	0.004	0.005	**0.227**	**0.159**	**0.871**
Class 3	0.016	0.009	**0.729**	0.012	0.017	0.017	0.009	0.009	**0.292**	0.051	0.008
Class 4	0.084	0.020	0.035	0.041	0.040	0.025	0.032	0.030	0.084	**0.534**	0.009
Class 5	0.176	**0.081**	0.063	**0.355**	0.114	**0.176**	**0.851**	**0.808**	0.139	0.058	0.010
Class 6	0.012	0.006	0.010	0.022	0.016	0.015	0.005	0.006	0.009	0.010	0.001
Class 7	0.002	0.001	0.004	0.001	0.001	0.001	0.001	0.001	0.004	0.005	0.000
Class 8	**0.180**	0.064	0.024	0.113	**0.533**	0.042	0.010	0.013	0.079	0.078	**0.074**
Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CI	0.370	0.101	0.133	0.811	0.371	0.247	0.103	0.158	0.777	0.298	0.085

As shown in Table 4.6, Points 3, 12, 13, 14, 17, 18, and 19 were identified as HUO series and Points 15, and 16 were within FOX series. For HUO, the CI values at Points 12 and 17 are very high (>0.6, a threshold defined by Burrough et al. in 1997) and should be considered as intergrades that cannot be used as “typical” instances in soil inference based on this partition. Among these seven points identified as HUO, Point 19 with the lowest CI value would be a good instance for inferring HUO in Unit 2 area. For FOX, Points 15 and 16 with low CI values and identified both as FOX series have a same dominant class, namely, Class 5. Thus, these two points are adjacent not only in spatial space (Figure 3.3) but also in environmental covariates space. They both can be regarded as “typical” FOX.

In fuzzy soil inference, a point with a very low CI value can be used to infer the soil series identified at this point to the entire area under study based on the distribution of fuzzy membership values and expert knowledge. As aforementioned, Point 7 in Table 4.5

and Point 19 in Table 4.6 are good and typical instances for inferring ZOR and HUO series, respectively. The class dominating the typical point is believed to correspond to a unique environmental configuration related to the soil series. However, two major concerns following this may be, first, other soil series may happen to correspond to the environmental configuration defined by the class. Second, other different environmental configurations may also represent the same series. In fact, they are much likely to be true in real world. As in the first situation, Zhu et al. (2008a) suggested to survey two or three sites with the highest membership values to a class in field to determine the associated soil series. In terms of the second concern, the combination of more than one environmental configuration is indeed analogous to the soil expert defining a soil series happening in different landform types (English, 2001). Thus, expert knowledge can be captured with the use of fuzzy logic in digital soil mapping to some extent.

4.3.2 Use of LiDAR-derived DEM

The fuzzy membership values at sampled points using LiDAR-derived DEM were also employed to discover the associations between fuzzy clusters and soil series. Take the experiment based on the 1 m LiDAR-derived DEM in Unit 1 area for example. Fuzzy membership values and the CI values at sampled points are shown in Table 4.7, with *m* of 1.5 and *c* of 5 that could be deemed as an optimal partition. The CI values at Points 6, 7 and 10 are near 0, and thus these three points are optimal to be used as “typical” instances of the identified soil series in Unit 1 area, namely, WTO, ZOR, and CAD, respectively. It is evident to use Points 6, 7, and 10 to represent Classes 4, 3, and 2 to which these points have

the highest fuzzy membership values, respectively. Then the fuzzy membership distribution of these classes accompanied by expert knowledge would be used to infer soil series from these points to areas without sampled points.

Table 4.7 Fuzzy membership values and confusion index (CI) at sampled points based on 1m LiDAR-derived DEM using the algorithm *m* of 1.5 and *c* of 5 in Unit 1 area (the first and second highest membership values in bold).

Membership value	Field Point ID and Soil_Code							
	4	5	6	7	8	9	10	11
	HIG	WTO	WTO	ZOR	FOX	WTO	CAD	WTO
Class 1	0.152	**0.699**	0.002	0.031	0.144	**0.622**	**0.059**	**0.450**
Class 2	0.020	0.030	0.000	0.012	0.148	0.012	**0.893**	0.177
Class 3	0.053	0.018	**0.004**	**0.856**	0.042	0.024	0.005	0.026
Class 4	**0.583**	**0.144**	**0.989**	**0.063**	**0.180**	**0.246**	0.009	0.093
Class 5	**0.192**	0.109	0.004	0.039	**0.486**	0.096	0.034	**0.254**
Total	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
CI	0.329	0.207	0.004	0.073	0.370	0.395	0.066	0.565

Through the interpretation of fuzzy membership values at 19 sampled points in Units 1 and 2 areas based on LiDAR-derived and conventional DEMs, it is indeed possible to interpret fuzzy clusters as soil series. As well, the nature of in-between types in soils and expert knowledge can also be quantitatively captured by fuzzy membership values. Hence, it is much likely to conclude that the associations between soil series and fuzzy clusters exist and fuzzy logic can assist soil inference well by providing quantitative measure of in-between types and expert knowledge to some extent.

4.4 Validation of Results

A subset area located in the southwest corner of the Waterloo Aquifer area (Figure 4.21) was utilized to validate the results in this study. This validation area covering an area of 56

km^2 is much larger than the Laurel Creek Conservation area and carries different landforms and vegetation types (mainly crops). The elevation ranges from the lowest in southeast part (~310 m) to the highest in the north part (~420 m). This area mainly consists of agricultural lands which are not likely to exhibit associations between soil and vegetation, and thus the contribution of vegetation covariate was not validated in this area. Only the resolution dependency and associations between soil series and clusters were examined and presented in this section.

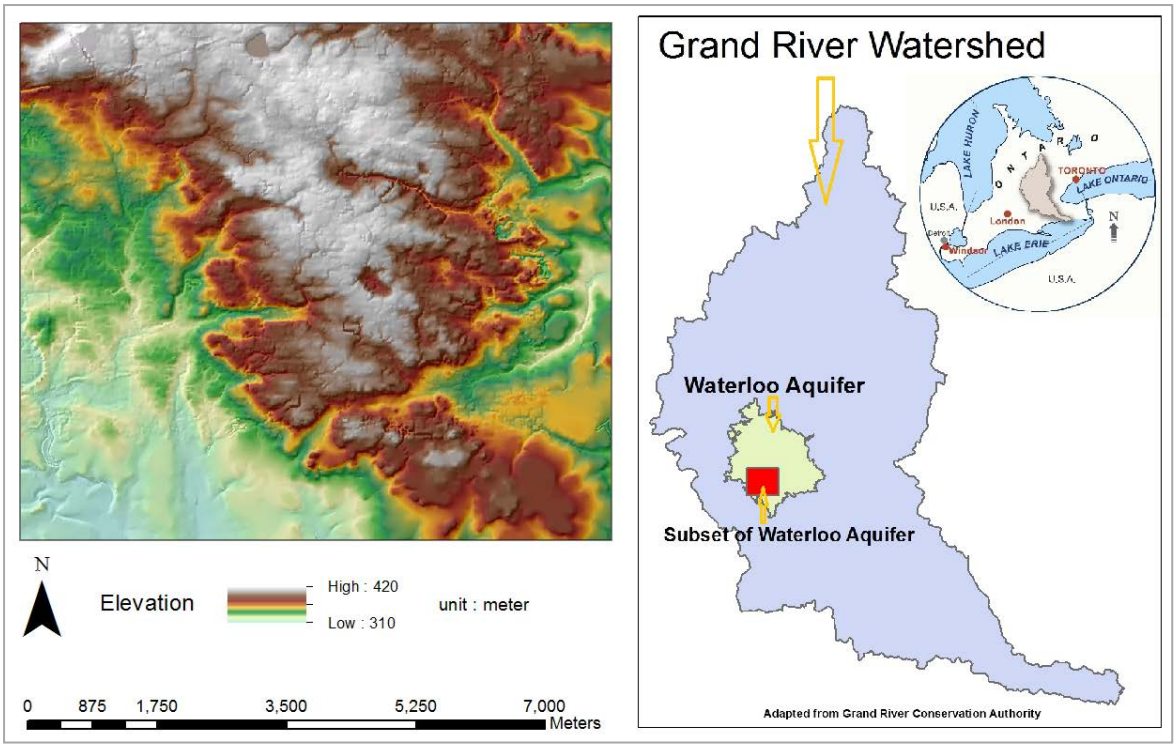

Figure 4.21 Location of the subset area of the Waterloo Aquifer area.

The 10 m conventional DEM covering this area was first filled and then resampled into 20 m, 30 m, 50 m, and 100 m. Land-surface parameters, namely, slope, plan curvature, profile curvature, topographic wetness index, and elevation, were derived from these DEMs and input into FCM clustering across five algorithms of m = 1.25, 1.5, 1.75, 2, and 2.25,

respectively. For each run, the number of clusters c ranged from 2 to 30, the iteration number was 50, and the maximum membership error was 0.01. The reason for using the range from 2 to 30 as the number of clusters, which was twice larger than the one used in the Laurel Creek Conservation area, was that the soils in the conventional soil map at a scale of 1:50,000 in the subset area include more than 15 types, but not more than 30 types. As aforementioned, it is better to use the number of clusters more than the number of existing soil types in soil maps.

The improvements in partition coefficient ($\Delta F = F(c) - F(c + 1)$) and the entropy ($\Delta H = H(c) - H(c + 1)$) over adjacent clusters based on conventional DEM at 10 m, 20 m, 30 m, 50 m, and 100 m with m of 1.25, 1.5, 1.75, 2, and 2.25 are shown in Figures 4.22-4.26. Table 4.8 summarises the optimal numbers of clusters in each run.

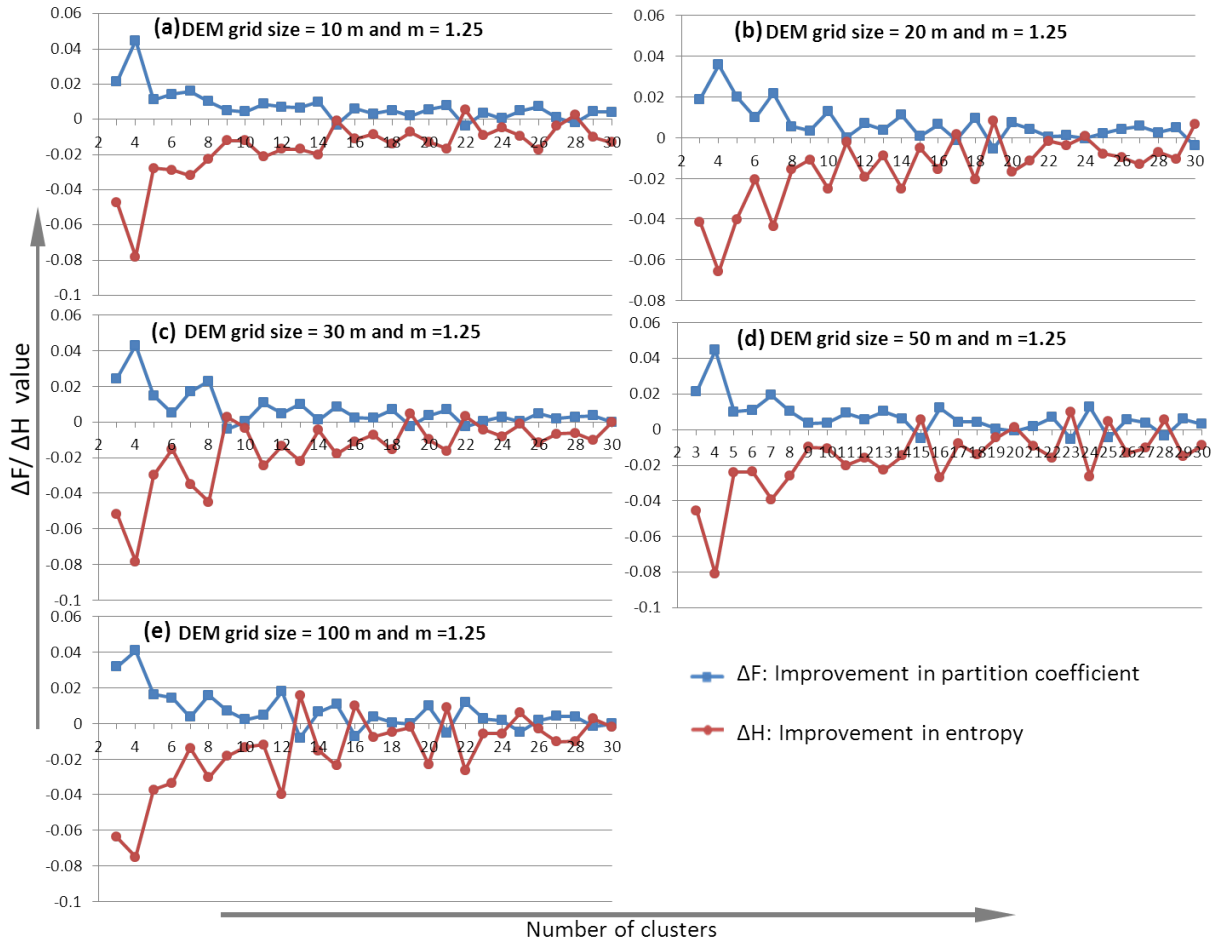

Figure 4.22 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, (d) 50 m, and (e) 100 m conventional DEMs using the algorithm with *m* of 1.25, in the subset area of the Waterloo Aquifer area.

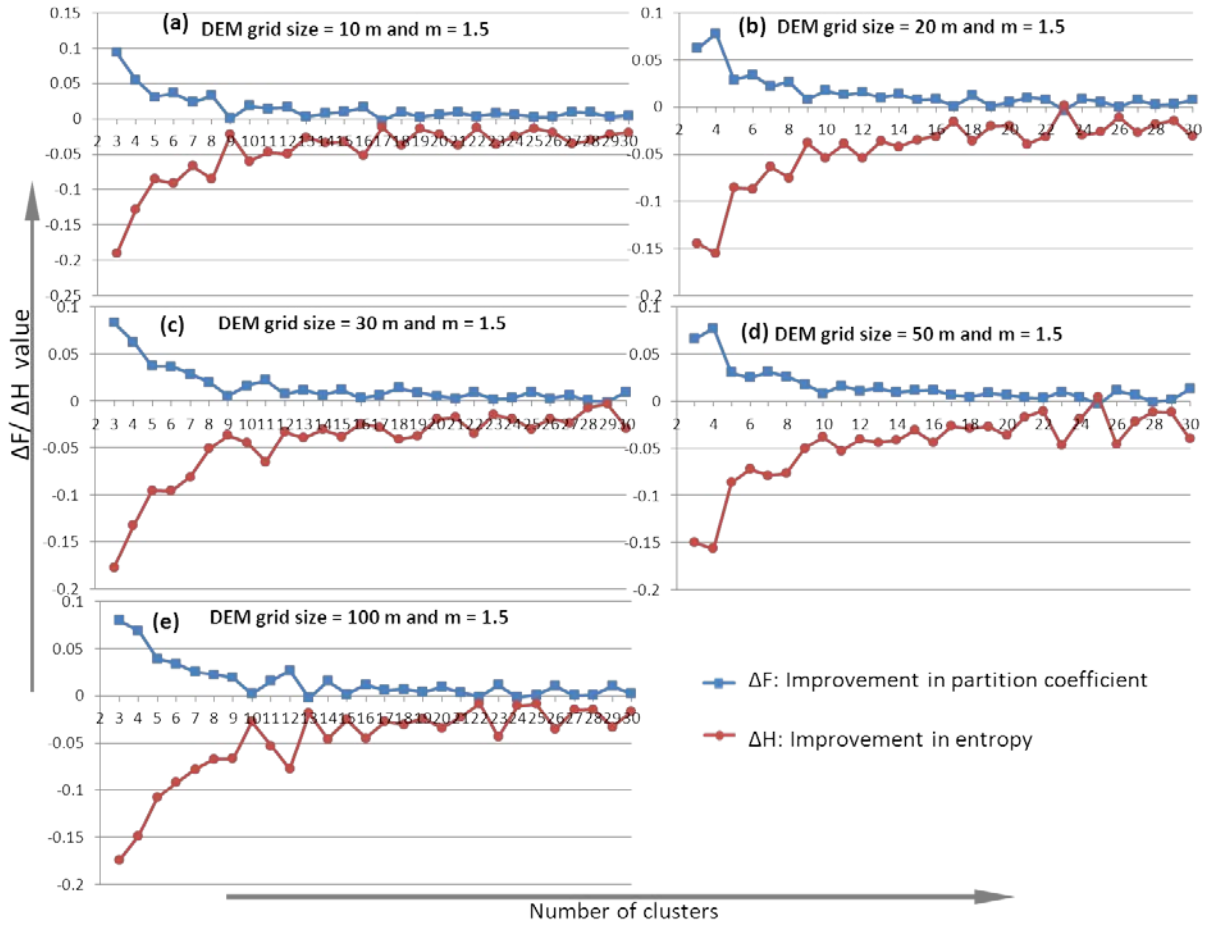

Figure 4.23 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, (d) 50 m, and (e) 100 m conventional DEMs using the algorithm with m of 1.5, in the subset area of the Waterloo Aquifer area.

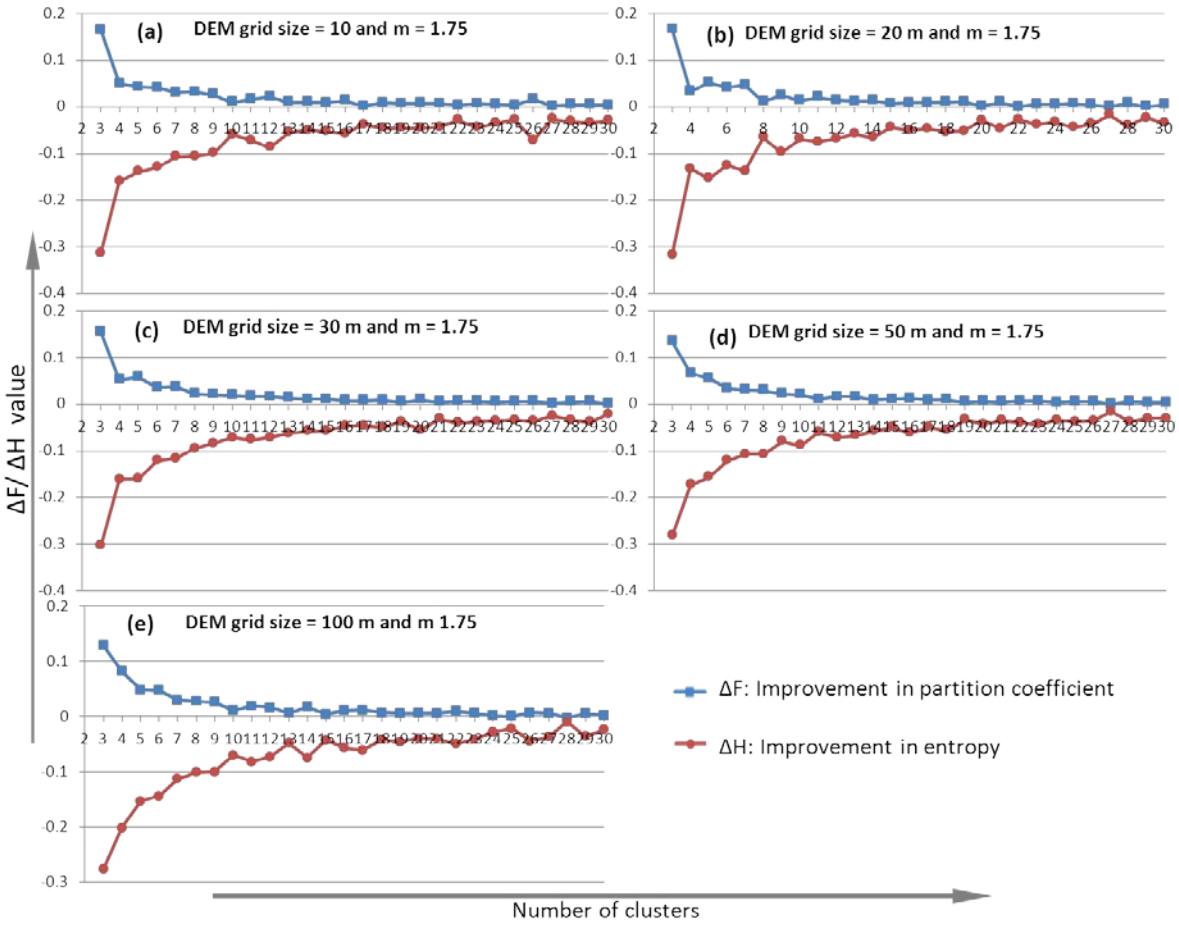

Figure 4.24 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, (d) 50 m, and (e) 100 m conventional DEMs using the algorithm with *m* of 1.75, in the subset area of the Waterloo Aquifer area.

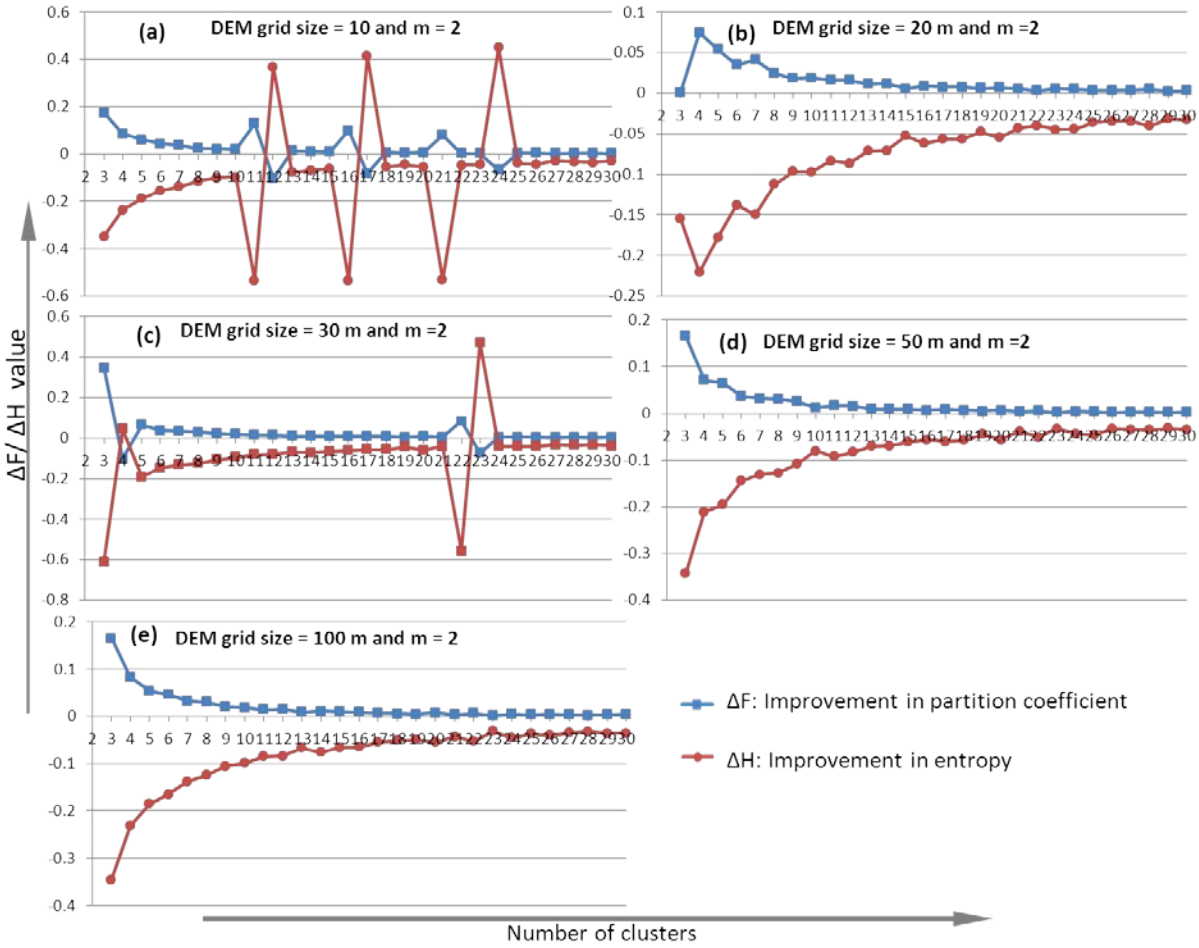

Figure 4.25 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, (d) 50 m, and (e) 100 m conventional DEMs using the algorithm with *m* of 2, in the subset area of the Waterloo Aquifer area.

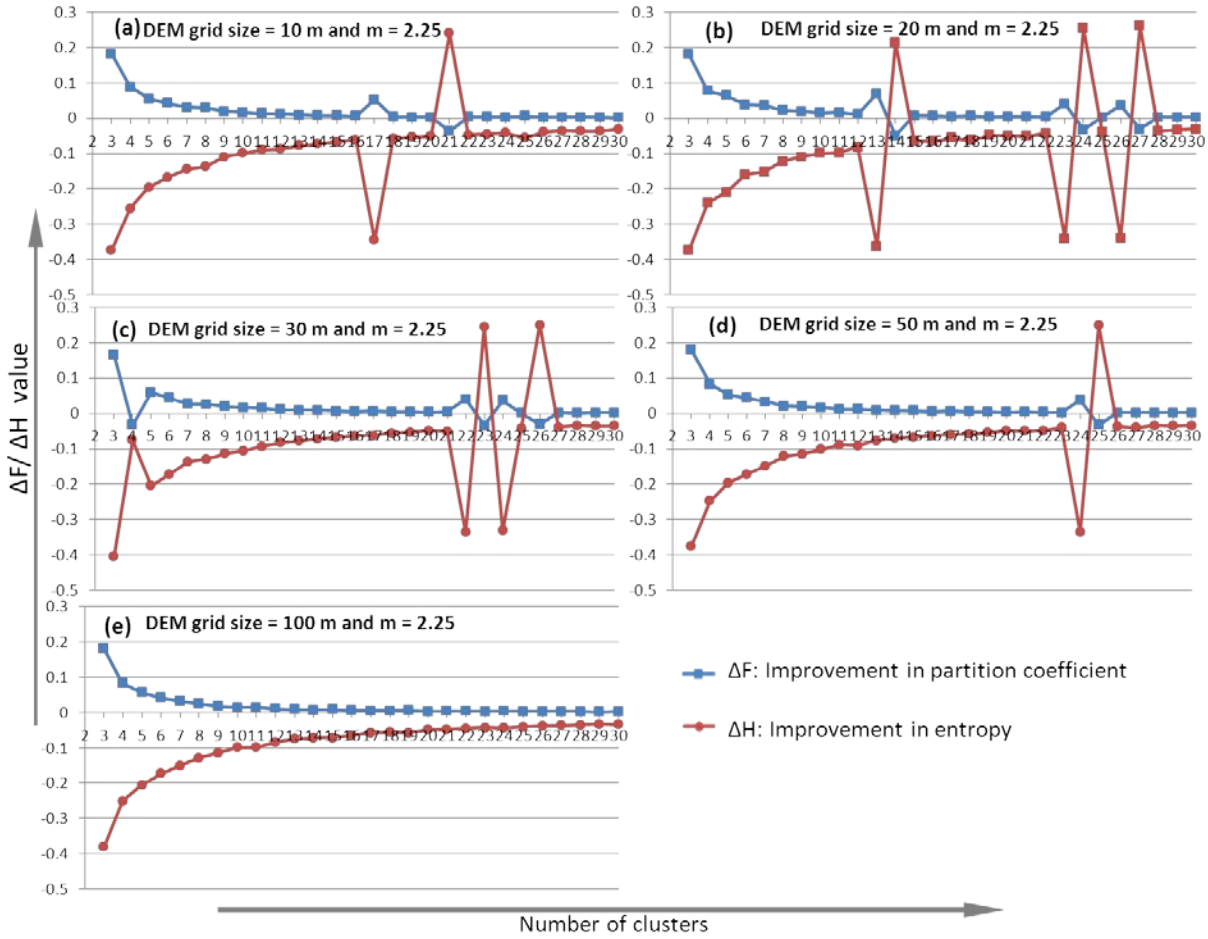

Figure 4.26 Improvements in partition coefficient ($\Delta F = F(c) - F(c+1)$) and entropy ($\Delta H = H(c) - H(c+1)$) plotted against the number of clusters based on (a) 10 m, (b) 20 m, (c) 30 m, (d) 50 m, and (e) 100 m conventional DEMs using the algorithm with *m* of 2.25, in the subset area of the Waterloo Aquifer area.

Table 4.8 Possible optimal numbers of clusters based on 10 m, 20 m, 30 m, and 50 m conventional DEMs across five algorithms with *m* of 1.25, 1.5, 1.75, 2, and 2.25, in the subset area of the Waterloo Aquifer area.

m \ grid size	10 m				20 m				30 m			50 m			100 m			
1.25	5	15	17	19	6	9	11	13	6	9	12	12	15	20	7	13	16	
	22	24	28		15	17	19	28	14	19	25	23	25	28	21	25		
1.5	5	7	9	13	7	9	11	13	9	12	14	6	10	12	10	13	15	
	17	19	22	25	17	23	26		16	26		25			22			
1.75	10	17	22		4	8	10	20	19			11	19	27	10	13	15	28
2	12	17	24		6	15			4	23		10						
2.25	21				14	24	27		4	23	26	25						

The FCM clustering results in the subset area of the Waterloo Aquifer area show that increasing *m* from 1.25 to 2.25 generally weakens the fluctuations in ΔF and ΔH at each of

the five resolutions, which confirms the observations in the study in the Laurel Creek Conservation area. It is also likely to find that increasing the grid size from 10 m to 100 m generally mutes the fluctuations in ΔF and ΔH and therefore reduces choices of optimal numbers of clusters with m of larger values such as 2 and 2.25. However, when m is smaller, the muting effect is not obvious. Table 4.8 shows that the choices of optimal numbers of clusters do not suffer obvious change when the resolution is aggregated from 10 m to 20 m, but enjoy slight decrease from 20 m to 100 m.

In Unit 1 area, when using LiDAR-derived DEM, the choices of optimal numbers of clusters increase with the resolution aggregated from 1 m to 10 m, decrease from 10 m to 20 m, generally keep stable when it comes to 30 m, and suddenly increase from 30 m to 50 m. While in using conventional DEM in Unit 1 area, the choices decrease from 10 m to 30 m, and increase from 30 m to 50 m. It can be seen the general trend of decreasing in the choices of optimal number of clusters in the subset area of the Waterloo Aquifer area from 20 m to 100 m confirms the decrease from 10 m to 30 m in Unit 1 area on the use of conventional DEM, and the decrease from 10 m to 20 m when using LiDAR-derived DEM. Although the corresponding resolutions are different, they show a consistent decreasing with resolution aggregated. The sudden change at 50 m resolution in Unit 1 area is not observed in the subset area, which may be explained as the scale and landform in the subset area are different from those in Unit 1 area. From the analysis of results in Unit 1 area, it is possible that 50 m resolution may be a threshold for fuzzy soil inference in this area, while for the subset area of the Waterloo Aquifer area, the corresponding threshold may be found

at lower resolution, lower than the tested 100 m. After all, the subset area is much larger than Unit 1 area. When study scales are different, the identified environmental configurations may be very different. Nevertheless, in general, the results in the subset area could confirm the conclusions on resolution issue obtained in Unit 1 area that the choices of optimal number of clusters shift with resolution aggregated and the application of the purposive sampling design for fuzzy soil inference requires the consideration of the resolution dependency.

In the study in the Laurel Creek Conservation area, an *m* equal to 1.5 shows a balance between stability and fuzziness, while in the subset area of the Waterloo Aquifer area, it seems an *m* equal to 2 performs better in establishing optimal clustering. In particular, graphs in Figure 4.23 with *m* of 1.5 are fluctuating all the time, but as shown in Figure 4.25, when *m* is 2, the fluctuation tends to be stable with the increase of the number of clusters, which can be seen even through the limited range from 2 to 30. For example, when *m* is 2 and the gird size is 10 m, the last optimal number of clusters before 30 is 24, and after 24 the curve becomes very stable, indicating an optimal clustering at 24. This phenomenon was not obvious in the study in the Laurel Creek Conservation area, which might be due to the limited number of clusters or the influence of scale issue in these different study areas.

In terms of the purposive sampling design, take the partition with *m* of 2 and *c* of 24 at 10 m resolution in the subset area for example. For each cluster, three points with the highest membership values to this cluster were designed as purposive sampling points for this cluster, which are shown in Figure 4.27. There are total 72 points in the figure with

their cluster numbers as labels (from 1 to 24). By overlapping with 1:50,000 conventional soil map, it is found that in 13 classes (54.2%) at least two of the three designed sampling points belonged to same soil series. Table 4.9 records the 13 classes, their purposive sampling points, and the corresponding soil series. Similar work has also been done in two nearby subset areas which have the same areas with this one. In one of these two areas, a purposive sampling design based on the partition with *m* of 2 and *c* of 20 at 10 m resolution, which was an optimal partition, demonstrates 12 classes (60%) carry at least two points of the three designed points belonging to same soil series. In the other area, the result is 12 classes out of 18 classes (66.7%) based on one optimal partition with *m* of 2 and *c* of 18 at 10 m resolution.

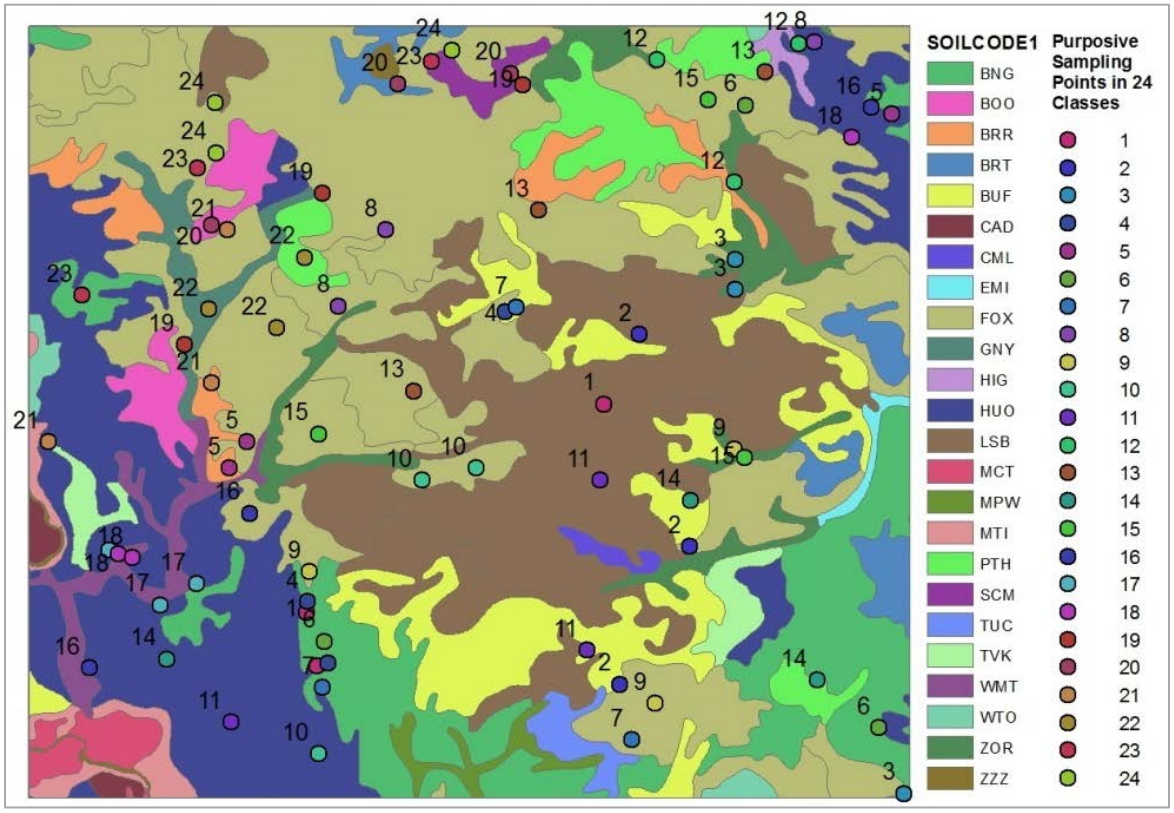

Figure 4.27 Purposive sampling points in the subset area of the Waterloo Aquifer area based on the partition with m of 2 and c of 24 and10 m conventional DEM.

Table 4.9 Classes with two or three purposive sampling points belonging to the same soil series in the subset area of the Waterloo Aquifer area based on the partition with *m* of 2 and *c* of 24 and 10 m conventional DEM.

Class No.	SOIL_CODE1	Point ID		Class No.	SOIL_CODE1	Point ID
1	BNG	2, 3		13	FOX	38, 39
3	ZOR	7,8		15	FOX	43, 44
4	BNG	10, 11		17	HUO	49, 50, 51
6	BNG	16, 17		18	HUO	52, 54
8	FOX	21, 22		19	FOX	55, 56, 57
9	FOX	26, 27		24	FOX	70, 71, 72
10	FOX	28, 30				

To some extent, this finding can confirm the conclusion in Section 4.3 that the associations between clusters and soil series exist. Although the conventional 1:50,000 soil map has been produced more than 3 decades ago, it still carries significant soil information that embodies with expert knowledge (Geng et al., 2010). Overlaying the soil map with the designed sampling points is a first approximation in evaluating the sampling design and also provides evidence of the associations between clusters and soils. According to the method by Zhu et al. (2008), sampling two to three points in each class is good enough for the establishment of the association between clusters and soil series. In the experiments in three subset areas of the Waterloo Aquifer area, more than half of the classes in each experiment involve two or three purposive sampling points belonging to same soil series. Therefore, it is much likely to verify the associations between fuzzy clusters and soils. This finding could confirm the conclusions drawn from the study in the Laurel Creek Conservation area. The real fieldwork also can be done based on this overlaying, in which the focus for soil surveyors is on validating the known soil types in soil maps at the designed points.

It should be noted that the experiments in the subset area of the Waterloo Aquifer area

was conducted at different landscape and scale from those in the experiments in the Laurel Creek Conservation area. Conclusions on resolution dependency drawn from the experiments in the subset area could only validate the trend in the experiments of Groups A and C generally and coarsely. It is less possible to obtain the exactly same conclusions from these experiments. In other words, it is likely to expect some discrepancies between the results in the subset area and Unit 1 area at the same resolution. However, this validation still bears meanings to this study and could provide valuable information on the understanding of the resolution dependency of the purposive sampling design. Additionally, the ultimate goal of this study is to provide preliminary suggestions for the future soil survey and mapping in the Waterloo Aquifer area. The future soil survey in this area bears significance to the protection of the groundwater resources and the precision agriculture in the Waterloo County. Thus, a validation experiment based on its subset area could also initiate future studies in the whole Waterloo Aquifer area.

Consequently, these findings in the subset areas of the Waterloo Aquifer area could confirm the conclusion in the study in the Laurel Creek Conservation area that the resolution of DEM has influence on the purposive sampling design, the choices of optimal number of clusters vary with resolution aggregated, and the associations between fuzzy clusters and soil series exist.

4.5 Chapter Summary

This chapter presents FCM clustering results in Units 1 and 2 areas in the Laurel Creek

Conservation area based on both conventional and LiDAR-derived DEMs. The dependency of the purposive sampling design in Unit 1 area varies with DEM resolution as well as DEM source. Although results of the contributions of vegetation covariate based on conventional and LiDAR-derived DEM in Unit 2 area are slightly different, it is still evident to employ vegetation covariate to infer soil as long as the association between soil and vegetation could be found. Interpretation of fuzzy membership values to clusters at sampled points indicates the existing of the associations between fuzzy cluster and soil series, and lends promise to quantifying expert knowledge by fuzzy membership values. The findings related to the issues of the resolution dependency and associations between soil series and clusters were generally confirmed by a validation experiment in a subset area of the Waterloo Aquifer area.

CHAPTER 5 CONCLUSIONS AND RECOMMENDATIONS

This study examines three issues in the purposive sampling design for fuzzy logic-based digital soil mapping: the resolution dependency, the contribution of vegetation covariates, and the use of LiDAR data. The results and discussion yield several findings that provide empirical answers to the research questions and suggest interesting avenues for future studies. This concluding chapter summarises the results of this study by answering the original research questions: (1) How do the purposive sampling design results vary with the resolution of conventional DEM? (2) How does the involvement of NDVI in environmental covariates database influence the purposive sampling design results? (3) Do the associations between fuzzy clusters and soil series exist? (4)What are the answers of above three questions when using LiDAR-derived DEM? The first section in this chapter presents conclusions on these research questions, and the following section discusses the limitations in this study and the direction for further research.

5.1 Conclusions

5.1.1 Resolution Dependency

The first research question this study attempted to answer was the resolution dependency of the purposive sampling design. It was examined by resampling basic DEM into coarser resolutions and performing FCM clustering of environmental covariates, including five land-surface parameters (elevation, slope, plan curvature, profile curvature, and topographic wetness index), derived from the basic and resampled DEMs. FCM

clustering results consisted of clusters information on partition coefficient and entropy, and fuzzy membership value maps. Optimal partitions of the environmental covariate space were distinguished by significant improvements in partition coefficient and entropy.

Through the experiments, the resolution of DEM has been recognized as an influential factor to the purposive sampling design for fuzzy soil mapping. The choices of optimal numbers of clusters shift with resolution aggregated. Different environmental configurations may be identified at different resolutions, which may result in diverse establishments of associations between fuzzy clusters and soil series. In other words, as the resolution changes different environmental configurations may become evident and significant in the prediction of soils. This shifting may arise from the effects of DEM resolution on land-surface parameters, as environmental modeling is restricted to spatial scales and resolutions on which the studied biophysical processes depend or the modeling *per se* depends. In addition, hardened class maps imply that finer resolution does not necessary lead to better classification results. By comparing with conventional soil map, more detailed variation may be obtained through fuzzy soil mapping method than conventional mapping method does, and there may be associations between patterns in fuzzy clusters and the distribution of soil series in the conventional soil map. Moreover, the observation of a persistent catenary sequence in fuzzy membership value maps across different resolutions indicates the association between landform positions and clusters, which may be independent of resolution and therefore be important in soil inference. The consistent catenary sequence may imply the dominate terrain features that have great

impact on soil development, while its disappearance at much lower resolution may suggest a resolution threshold for fuzzy soil inference in the study area.

Consequently, choosing a proper resolution of DEM to derive land-surface parameters for the specific study area is essential. Taking account of the sensitivity of FCM clustering to DEM resolution can not only improve the quality of the purposive sampling design, but also the understanding of landforms and soils in the area under study. Although it has to be admitted that no perfect DEM resolution exists, fuzzy soil mapping could still profit from explicit procedures on analyzing the influence brought by resolution change, and on attempting to locate a fit resolution at which the most possible detailed and meaningful information on landscape could be gained.

5.1.2 Vegetation Covariate

The second research question addressed the missing of vegetation covariates in previous attempts on the purposive sampling design for fuzzy soil inference. This study experimented with NDVI, a vegetation covariate derived from remotely sensed data, in the construction of the environmental covariates database. The control experiments based on NDVI and the five land-surface parameters (elevation, slope, plan curvature, profile curvature, and topographic wetness index) demonstrated the influence of NDVI on the purposive sampling design.

In FCM clustering, combing NDVI with land-surface parameters derived from conventional DEM is able to provide more optimal numbers of clusters for the sampling design. It is possible that the introducing of more choices will lead to more effective

classification results and thus improvements in the fuzzy soil inference, while more field studies are needed to support this argument. Additionally, hardened class maps demonstrate more details revealed through embodying NDVI. Although land-surface parameters are the most useful predictors for soil mapping, their best performance may take place where water and material flows are strongly governed by relief. For areas with relatively flat relief, the role played by other environmental covariates in the prediction may be strengthened, such as vegetation covariates. It should be noted that vegetation covariates are expected to complement terrain covariates but not to replace them because terrain covariates are still the most practical and influential predictors in digital soil mapping.

Consequently, it is recommended to add vegetation covariates in fuzzy logic-based digital soil mapping as long as the association between vegetation and soil can be found in the study area. In the state of art, more new environmental covariates other than parameters derived from elevation data are becoming available, arising from the boost in remotely sensed data. Vegetation covariates are typical examples of these data. Many studies have found combining new remotely sensed covariates such as NDVI with terrain covariates would produce the better environmental modeling and prediction results. How these remotely sensed covariates can assist the fuzzy soil inference is worthy of exploration.

5.1.3 Associations between Fuzzy Clusters and Soil Series

The third research question on the existence of the associations between fuzzy clusters and soil series was examined by interpreting fuzzy membership values at field sampling sites. Fuzzy membership values were obtained from optimal partitions of the environmental

covariates in the study area. The confusion degree of an observation belonging to more than one class was measured by a confusion index.

Associations between fuzzy clusters and soil series can be seen from the interpretations of fuzzy membership values at the sampling points. Points with more confusion carry the in-between characteristic and are not recommended for representing the corresponding class, while those with low confusion are deemed as typical instances representing the dominant class. Interpretations of fuzzy membership values are able to discover that more than one cluster correspond to the same soil series, which is analogous to the knowledge of soil experts when they define a soil series happening in different landform types. It is also possible that within a same cluster more than one typical soil series are found, in which situation effective strategies are needed to establish the corresponding soil series by the specific environmental configurations. In the purposive sampling design, the corresponding points to each cluster obtained from an optimal fuzzy partition of the environmental covariates space will be surveyed, and these points need to have low confusion and obvious dominant class.

As a result, the in-between nature of soil and expert knowledge can be represented through fuzzy membership values in a quantitative form, and the associations between fuzzy clusters and soil series can be discovered through interpreting fuzzy membership values. It is true that in practical applications of fuzzy soil mapping the establishment of the associations between environmental configurations and soil series or properties is assisted not only by the FCM clustering results but also expert knowledge. Soil experts design rules

to define those associations. Nevertheless, fuzzy membership values provide an opportunity to quantify such expert knowledge, and they are able to aid experts in defining rules. It can be seen that the associations between fuzzy clusters and soil series exist and thus FCM clustering of environmental covariates represents a useful avenue for soil survey and mapping with limited expert knowledge and resources.

5.1.4 Use of LiDAR Data

To address the last research question, LiDAR data were employed to examine the resolution dependency, the contribution of vegetation covariates, and the associations between fuzzy clusters and soil series. A high-resolution DEM was generated from raw LiDAR points and resampled into coarser resolutions to derive land-surface parameters. Control experiments on NDVI, the vegetation covariate, were performed based on LiDAR-derived land-surface parameters as well. The interpretation of fuzzy membership values obtained from an optimal partition based on LiDAR-derived environmental covariates was also conducted.

It can be seen that the level of dependency of the purposive sampling design to resolution varies with DEM sources, conventional and LiDAR-derived DEMs. The shifting pattern of the choices of optimal numbers of clusters using LiDAR-derived DEM is different from the one using conventional DEM. At each resolution, the choices using LiDAR-derived DEM are more than those using conventional DEM. These variances may arise from the discrepancy in the level of detail of relief represented by conventional and LiDAR-derived DEMs. LiDAR data have higher vertical and horizontal accuracy than most

of the other DEM sources, which could result in the discrepancy found in this study. It is evident that LiDAR-derived DEM can delineate land-surface parameters in more detail than conventional DEM does partly because the different degrees of relief variation and accuracies in the two DEM sources, which may correspond to the finding in many studies that LiDAR-derived DEM provides a much more complete and accurate representation of the hydrological processes than the conventional DEM does. However, it is not necessarily the case that high-resolution LiDAR-derived DEM performs better than low-resolution LiDAR-derived DEM in fuzzy soil inference. With the popularity of high resolution DEM data, investigations on distinguishing the optimal resolution for specific study area are important.

In terms of the contribution of vegetation covariates, slightly different results are found using LiDAR-derived DEM from those using conventional DEM. Combining NDVI with LiDAR-derived land-surface parameters could not introduce more choices of optimal number of clusters, which may be ascribed to the difference between the levels of detail in land-surface parameters derived from the two source DEM, and may also arise from the discrepancy in the level of detail of NDVI and land-surface parameters. Notwithstanding, it is still suggest to consider vegetation covariates in fuzzy soil inference based on LiDAR data, because the discrepancy in the level of detail of vegetation covariate data and LiDAR data will be diminished by the increasing accuracy of remotely sensed data.

The interpretation of fuzzy membership values obtained from the LiDAR-based fuzzy partition at sampled sites agrees with the one based on conventional DEM. Typical

instances for specific classes can be seen from the results. The associations between fuzzy clusters and soil series have the potential to assist fuzzy soil inference and the quantification of expert knowledge to some extent. Although there are some difficulties in using LiDAR data, such as the large volume of data and the challenges to conventional algorithms, LiDAR data has become an auspicious source to generate DEM for terrain analysis and also digital soil mapping. The results in this study thus provide useful experiences to the use of LiDAR data in fuzzy soil inference.

In conclusion, this study has accomplished its overall objectives by answering the four research questions. This initial work on the purposive sampling design will minimize the extent and the amount of field investigation efforts. The examination on related issues in this design also carries considerable potentials for assisting the determination of the optimal resolution and vegetation as environmental covariates, and relevance to the auspicious use of LiDAR data in digital soil mapping. The most telling contribution of this study is the encouragement for the success of extrapolating soil patterns from environmental covariates under fuzzy logic.

5.2 Limitations and Recommendations for Future Studies

This study bears significances to the initial work of fuzzy logic-based digital soil mapping in the Laurel Creek Conservation area, however, it also carries some limitations that can be improved in future studies.

First, the yielded results and conclusions were based on this case study in the Laurel

Creek Conservation area only. Similar studies in other areas with different landform and vegetation types should be carefully validated. Many previous studies have indicated that their conclusions may only hold true for their specific study areas or landform types (e.g., Arrell et al., 2007; Deng 2007; Deng et al., 2007; Wu et al., 2008; Vaze et al., 2010). Although a subset area of the Waterloo Aquifer area, carrying different landform and vegetation types was used as validation, it is still necessary to carefully consider bias and uncertainty involved in terrain analysis and fuzzy soil inference. In this study, bias and uncertainty may exist in the initial data, DEM. The conventional DEM carries a low accuracy because it was produced based on low resolution source data. Errors may exist in LiDAR-derived DEM too, suggesting a validation based on field survey of elevation in the future work. Moreover, the accuracy of the primary and secondary land-surface parameters relies on the performance of the deriving algorithms. The scale that these algorithms are developed at may influence the resolution dependency results to some extent. In addition, the procedure of FCM clustering carries some uncertainties, such as the determination of the fuzzy degree, the maximum membership error, and the iteration number. Hence, it is suggested to thoughtfully infer the obtained conclusions in this study to other study areas with different or even similar landform and vegetation types.

Second, it should be noted that the stratification according to the geology layer is not always obligatory in fuzzy logic-based digital soil mapping but is likely to be necessary for areas with soil displaying significant dependence on geology. Terrain factor is deemed as the basic influential factor in soil formation, using which is very practical because of the

widespread elevation data (MacBratney et al., 2003). Considering the geology layer as an environmental covariate in soil inference is also promising due to the role of parent materials in the formation of soil in Jenny (1941). However, in this study the influential factors other than terrain or vegetation were controlled in order to reach its objectives through a much clearer manner. In addition, the geology characteristics in the study area involve much more issues which need careful investigations, such as the Waterloo Moraine. It might be impossible to comprehensively capture these characteristics in this study with limited data and fieldwork. Thus, it is recommended to research more on geologic predictors for soil inference in the future work, particularly, with respect to the glacial deposits in southern Ontario.

Third, the basic field survey assisted by two soil scientists promoted the understanding of the soil in the study area, while it will be necessary to pursue a comprehensive field survey based on the results of the purposive sampling design in the future work. It has to be noted that this study did not conduct such a comprehensive survey because of the lack of resources in surveyors and tools, and also because of the limited time and accessibility. Although digital soil mapping involves the use of large remotely sensed data and GIS tools, fieldwork is still inevitable. All the new techniques are aiming at reducing it but hardly replacing it. Moreover, fieldwork for digital soil mapping may deviate from conventional soil survey because, first, preliminary results such as the purposive sampling design can provide quantitative guidelines and estimations for real fieldwork, and second, it can be supported by a legacy of soil maps, such as the overlapping strategy in the validation

experiment in Section 4.4. Thus, a field survey based on the purposive sampling design is needed in the future work to validate the design and to infer soil series or properties.

Fourth, this study assumed all the land-surface parameters have the same weight in soil prediction, while this may not always hold true. For example, in Behrens et al. (2010), it is reported that topographic wetness index and profile curvature have more weights than other parameters when predicting fluvial soils, while soil classes occurring on steeper slopes show stronger correlation with slope than other parameters do. Future studies need to evaluate the different weights of predictors based on an investigation of soil forming processes and landscape characteristics in field. Moreover, the determination of land-surface parameters used in fuzzy soil inference is still arbitrary itself, and strongly depends on knowledge in domains including not only soil science but also geomorphometry. Although the utilization of fuzzy clustering aims at minimizing requirements for domain knowledge in determining input parameters, such knowledge can still control and influence the output and the interpretation of the results. A future application of this fuzzy soil inference method in areas such as the Waterloo Aquifer area will definitely need experts with knowledge not only about the soils developed in these areas but also their specific terrain characteristics.

Fifth, one main objective of this study was to explore the resolution dependency of the purposive sampling design for fuzzy soil inference. However it has to be noted that the integration of multiple scales has become a new trend in terrain analysis (Deng, 2007). This integration is expected to improve the modeling of hydrological and geomorphological

processes which carry obvious multi-scale characteristics (Gessler et al., 2009), such as fractal (Deng, 2007). There have been many efforts on the integration of multiple scales in terrain analysis (e.g., Fisher et al., 2004; Schmidt and Andrew, 2005; Behrens et al., 2010), based on approaches of contextual spatial information (e.g., Behrens et al., 2010), wavelet analysis (e.g., Lark, 2007), incorporating scale into land-surface parameters (e.g., Schmidt and Hewitt, 2004), and so on. Fuzzy logic has also been incorporated to multi-scale landform classification. For example, in Deng and Wilson (2008), multiple spatial scales and multiple semantic meanings were combined to delineate mountain peaks as fuzzy entities, through considering the peak properties and spatial scale dependency. This initial work has shown a positive sign of using multiple spatial scales in fuzzy landform classification. Given the basic connections between geomorphometry and digital soil mapping, explorations on multi-scale fuzzy soil inference will be a future research priority. The findings in this study could also provide preliminary indications for the integration of multiple scales.

Sixth, it should be noted that soil is a very complex system that is always difficult to model accurately. Soil is often continuous, so that fuzzy logic used in digital soil mapping indeed can assist in representing this continuousness characteristic. However, since soil is complicate, it may not always be continuous. Sudden changes may be observed within very short spatial distance. This may always happen, particularly, when surveying soils developed based on glacier deposits, in which soil profiles might suffer abrupt disturbs during the movement of glaciers. In the field survey in this study, such a sudden change

was found among Points 12-16 , at which soil profiles changed quickly against the spatial distance. If there is a way to “see” the whole soils horizontally in that particular area, it might be found abrupt or crisp changes, not a gentle continuous change in soil profiles horizontally. Thus, the fuzzy clustering of soil in such situations needs more future efforts.

In conclusion, although some limitations exist in this study, it can still provide useful insights for future studies on fuzzy soil inference. However, it has to be admitted that finding a number of clusters that can completely satisfy a complex system such as soil is very difficult. As Odeh et al. (1992) pointed out, “the biggest problem in fuzzy c-means clustering is to determine how many classes exist in the data and how fuzzy they are”. Indeed, there is still much work need to be done in order to solve this problem and to take the most advantage of fuzzy logic in digital soil mapping.

REFERENCES

Ahn, C.W., Baumgardner, M.F., and Biehl, L.L., 1999. Delineation of soil variability using geostatistics and fuzzy clustering analyses of hyperspectral data. S*oil Science Society of America Journal*, 63, 142-150.

Anderson, D.W., and Smith, C.A.S., 2011. A history of soil classification and soil survey in Canada: Personal perspectives. *Canadian Journal of Soil Science*, 91(5), 675-694.

Arrell, K.E., Fisher, P.F., Tate, N.J., and Bastin, L., 2007. A fuzzy c-means classification of elevation derivatives to extract the morphometric classification of landforms in snowdonia, Wales. *Computers and Geosciences*, 33(10), 1366-1381.

Axelsson, P., 2000. DEM generation from laser scanner data using adaptive TIN models. *International Archive of Photogrammetry and Remote Sensing*, 33(B4), 110–117.

Band, L.E. and Moore, I.D., 1995, Scale: Landscape attributes and geographical information systems. *Hydrological Processes*, 9, 401–422.

Behrens, T., Zhu, A., Schmidt, K., and Scholten, T., 2010. Multi-scale digital terrain analysis and feature selection for digital soil mapping. *Geoderma* 155(3-4), 175-185.

Bezdek, J.C., 1974. Numerical taxonomy with fuzzy sets. *Journal of Mathematical Biology* 1, 57-71.

Bezdek, J.C., 1981. *Pattern Recognition with Fuzzy Objective Function Algorithms*. New York: Plenum Press.

Bezdek, J.C., Ehrlich, R., and Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. *Computers and Geosciences*, 10(2-3), 191-203.

Brady, N.C., and Weil, R.R., 2000. *Elements of the nature and properties of soils*. Upper Saddle River, N.J. : Prentice Hall.

Bui, E.N., Loughhead, A., and Corner, R., 1999. Extracting soil–landscape rules from previous soil surveys. *Australian Journal of Soil Research*, 37, 495– 508.

Bui, E.N., and Moran, C.J., 2001. Disaggregation of polygons of surficial geology and soil maps using spatial modelling and legacy data. *Geoderma*, 103, 79–94.

Burrough, P.A., 1987. Mapping and mapping analysis: New tools for land evaluation. *Soil Use Management*, 3, 20-25.

Burrough, P.A., Wilson, J., van Gaans, P.F.M., and Hansen, A., 2001. Fuzzy k-means classification of topo-climatic data as an aid to forest mapping in the Greater Yellowstone Area, USA. *Landscape Ecology*, 16, 523–546.

Burrough, P.A., van Gaans, P.F.M., and Hootsmans, R., 1997. Continuous classification in soil survey: Spatial correlation, confusion and boundaries. *Geoderma*, 77(2-4), 115-135.

Burrough, P.A., van Gaans, P.F.M., and MacMillan, R.A., 2000. High-resolution landform classification using fuzzy k-means. *Fuzzy Sets and Systems*, 113(1), 37-52.

Campling, P., Gobin, A., and Feyen, J., 2002. Logistic modeling to spatially predict the probability of soil drainage classes. *Soil Science Society of America Journal*, 66, 1390–1401.

Chaplot, V., Walter, C., and Curmi, P., 2000. Improving soil hydromorphy prediction according to DEM resolution and available pedological data. *Geoderma*, 97, 405– 422.

Chang, K. and Tsai, B., 1991. The effect of DEM resolution on slope and aspect mapping. *Cartography and Geographic Information Systems*, 18, 69–77.

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

Coen, J.M., 1987. *Soil Survey Handbook*. Vol. 1 Technical Bulletin 1987-9E, ISBN 0-662-15374-X, Agriculture Canada, Ottawa, ON.

De Gruijter, J.J., Walvooer, D.J.J, and van Gaans, P.F.M., 1997. Continuous soil maps: a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. In: De Gruijter, J.J., McBratney, A.B., McSweeney, K. (Eds.), *Fuzzy Sets in Soil Science*. *Geoderma*, 77, 169-195

Deng, Y., 2007. New trends in digital terrain analysis: landform definition, representation, and classification. *Progress in Physical Geography*, 31(4), 405-419.

Deng, Y., and Wilson, J.P., 2008. Multi-scale and multi-criteria mapping of mountain peaks as fuzzy entities. *International Journal of Geographical Information Science*, 22(2), 205-218.

Deng, Y., Wilson, J.P., and Bauer, B.O., 2007. DEM resolution dependencies of terrain attributes across a landscape. *International Journal of Geographical Information Science*, 21(2), 187-213.

Dinkins, C.P., and Jones, C., 2008. Soil sampling strategies. Adapted from http://msuextension.org/publications/AgandNaturalResources/MT200803AG.pdf.

Dobos, E., Micheli, E., Baumgardner, M.F., Biehl, L., and Helt, T., 2000. Use of combined digital elevation model and satellite radiometric data for regional soil mapping. *Geoderma, 97*(3-4), 367-391.

Dunn, J.C., 1974. A fuzzy relative of the isodata process and its use in detecting compact, well-separated clusters. *Journal of Cybernetics*. 3, 22-57.

English, E.M., 2001. Assisting Knowledge-Based Inference Soil Mapping: The Application of Fuzzy c-Means Clustering to Expose Environmental Niches. M.S. Thesis, University of Wisconsin-Madison.

Evans, I.S., 1972. General geomorphometry, derivatives of altitude, and descriptive statistics. In: Chorley, R.J. (Ed.), *Spatial Analysis in Geomorphology*. Harper & Row, 17-90.

Evans, I.S., 1979. An integrated system of terrain analysis and slope mapping. Final Report (Report 6) on Grant DA-ERO-591-73-G0040. *Statistical Characterization of Altitude Matrices by Computer*. Department of Geography, University of Durham, 192.

Fisher, P, Wood, J., and Cheng, T., 2004 Where is Helvellyn? Fuzziness of multi-scale landscape morphometry. *Transactions of the Institute of British Geographers*, 29,106–128.

Florinsky, I.V. and Kuryakova, G.A., 2000: Determination of grid size for digital terrain modelling in landscape investigations: exemplified by soil moisture distribution at a micro-scale. *International Journal of Geographical Information Science*,14, 815–32.

Fridland, V.M., 1974. Structure of the soil mantle. *Geoderma*, 12, 35-41.

Gallant, J.C. and Dowling, T.I., 2003. A multiresolution index of valley bottom flatness form mapping depositional areas. *Water Resources Research*, 39, 1347–1359.

Gallant, J.C., Hutchinson, M.F., and Wilson, J.P., 2000. Future directions for terrain analysis. In: Wilson, J.P. and Gallant, J.C. (eds.), *Terrain Analysis Principles and Applications.* New York: John Wiley, 423–427

Gao, J., 1997, Resolution and accuracy of terrain representation by grid DEMs at a microscale. *International Journal of Geographical Information Science*, 11, 199–212.

Geng, X., Burcher, R., Kroetsch, D., and Mitchell, S., 2012. Multi-scale feature data processing and landscape analysis toolkit for predictive soil mapping. Unpublished conference paper.

Geng, X., Fraser, W., VandenBygaart, B., Smith, S., Waddell, A., Jiao, Y., and Patterson, G., 2010. Toward digital soil mapping in canada: Existing soil survey data and related expert knowledge. In Boettinger J. L., Howell D. W., Moore A. C., Hartemink A. E. and Kienast-Brown S.(Eds.), *Digital Soil Mapping, Progress in Soil Science 2*, Springer Netherlands.

Gessler, P., Pike, R., MacMillan, R. A., Hengl, T., and Reuter, H. I., 2009. The future of geomorphometry. In: Hengl, T. and Reuter, H.I., (Eds), *Geomorphometry: Concepts, Software, Applications*, Elsevier, 637-652.

Grunwald, S. 2009. Multi-criteria characterization of recent digital soil mapping and modeling approaches. *Geoderma*, 152, 195–207.

Hengl, T., Heuvelink, G. B. M., and Stein, A., 2004. A generic framework for spatial prediction of soil variables based on regression-Kriging. *Geoderma*,120(1-2),75-93.

Hodgson, M.E., Jensen, J.R., Raber, G., Tullis, J., Davis, B., Schuckman, K., and Thompson, G., 2005. An evaluation of LiDAR-derived elevation and terrain slope in leaf-off conditions. *Photogrammetric Engineering and Remote Sensing* 71, 817–823.

Howell, D., Kim, Y.G., and Haydu-Houdeshell, C.A., 2008. Development and application of digital soil mapping within traditional soil survey: What will it grow into? In: Hartemink, A.E., McBratney, A. and Mendonça-Santos, M.d.L.(Eds.), *Digital Soil Mapping with Limited Data*, Springer Netherlands, 43-52.

Hudson, B.D., 1992. The soil survey as paradigm-based science. *Soil Science Society of America Journal*, 56, 836–841.

Huggett, R.J., 1975. Soil landscape systems: a model of soil genesis. *Geoderma,* 13, 1– 22.

Irvin, B.J., Ventura, S.J., and Slater, B.K., 1997. Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Wisconsin, *Geoderma* ,77,137-154.

Jenny, H., 1941. *Factors of Soil Formation: A System of Quantitative Pedology*. New York: McGraw-Hill.

Kriegler, F.J., Malila, W.A., Nalepka, R.F., and Richardson, W., 1969. Preprocessing transformations and their effects on multispectral recognition, in Proceedings of the Sixth International Symposium on Remote Sensing of Environment, University of Michigan, Ann Arbor, MI, 97-131.

Lagacherie, P., 2008. Digital soil mapping: A state of the art. In: Hartemink, A.E., McBratney, A.B., and Mendonca Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. New York: Springer, 3-14.

Lagacherie, P., Cazemier, D.R., Van Gaans, P., and Burrough, P.A., 1997. Fuzzy k-means clustering of fields in an elementary catchment and extrapolation to a larger area. In: De Gruijter, J.J., McBratney, A.B., Mc-Sweeney, K. (Eds.), Fuzzy Sets in Soil Science. *Geoderma*, 77, 197-216.

Lagacherie, P., and McBratney, A.B., 2007. Spatial soil information systems and spatial soil inference systems: perspectives for Digital Soil Mapping. In: Lagacherie P., McBratney, A.B., Voltz, M. (Eds.), *Digital Soil Mapping: An Introductory Perspective*. Amsterdam: Elsevier, 3–24.

Lark, R.M., 2007. Decomposing digital soil information by spatial scale. In: Lagacherie P., McBratney, A.B., and Voltz, M. (Eds.), *Digital Soil Mapping: An Introductory Perspective.* Amsterdam: Elsevier, 301-326.

Laurel Creek Conservation Area Master Plan, 2004. http://www.grandriver.ca/index/document.cfm?Sec=46&Sub1=7.

Liang, S., 2004. *Quantitative Remote Sensing of Land Surfaces*. Hoboken, N.J.: Wiley-Interscience.

Liu, X., 2008. Airborne lidar for DEM generation: some critical issues. *Progress in Physical Geography* 32(1), 31-49.

Liu, F., Geng, X., Zhu, A., Fraser, W., and Waddell, A., 2011. Soil texture mapping over low relief areas using land surface feedback dynamic patterns extracted from MODIS. *Geoderma,* 171-172, 44-52.

Liu, J., Pattey, E., Nolin, M.C., Miller, J.R., and Ka, O., 2008. Mapping within-field soil drainage using remote sensing, DEM and apparent soil electrical conductivity. *Geoderma*, 143, 261–272.

Lo, C.P., and Yeung, A.K.W., 2007. *Concepts and Techniques of Geographic Information Systems* (2nd ed.). Upper Saddle River, N.J. Pearson Prentice Hall.

Mabit, L., Bernard, C., Makhlouf, M., and Laverdière, M.R., 2008. Spatial variability of erosion and soil organic matter content estimated from 137Cs measurements and geostatistics. *Geoderma*, 146, 245–251.

MacMillan, R.A., Martin, T.C., Earle, T.J., and McNabb, D.H., 2003. Automated analysis and classification of landforms using high resolution digital elevation data: applications and issues. *Canadian Journal of Remote Sensing* 29, 592–606.

MacMillan, R.A., Moon, D.E., and Coupé, R.A., 2007. Automated predictive ecological mapping in a forest region of B.C., Canada, 2001–2005. *Geoderma*, 140, 353–373.

MacMillan, R., Pettapiece, W., Nolan, S., and Goddard, T., 2000. A generic procedure automatically segmenting landforms into landform elements using DEMs, heuristic rules and fuzzy logic. *Fuzzy sets and Systems*, 113, 81-109.

MacMillan, R.A., Moon, D.E., Coupé, R.A., and Phillips, N., 2010. Predictive ecosystem mapping (PEM) for 8.2 ha of forestland, British Columbia, Canada.

Maune, D.F., 2007. *Digital Elevation Model Technologies and Applications: The DEM Users Manual* (2nd ed.). Bethesda, Md.: American Society for Photogrammetry and Remote Sensing.

McBratney, A.B., and De Gruijter, J.J., 1992. A continuum approach to soil classification by modified fuzzy k-means with extragrades. *Journal of Soil Science.* 43, 159-175.

McBratney, A.B., De Gruijter, J.J., and Brus, D.J., 1992. Spacial prediction and mapping of continuous soil classes. *Geoderma*, 54, 39-64.

McBratney, A.B., Mendonça Santos, M.L., and Minasny, B., 2003. On digital soil mapping. *Geoderma,* 117(1-2), 3-52.

McBratney, A.B., and Odeh, I.O.A., 1997. Application of fuzzy sets in soil science: Fuzzy logic, fuzzy measurements and fuzzy decisions. *Geoderma,* 77(2-4), 85-113.

McKeague, J.A., and Stobbe, P.C., 1978. History of soil survey in Canada 1914-1975. *Historical Series No.11*. T & H Printers.

McKenzie, N.J., Gessler, P.E., Ryan, P.J., and O'Connel, D., 2000. The role of terrain analysis in soil mapping. In: Wilson, J.P., Gallant, J.C. (Eds.), *Terrain Analysis-Principles and Applications,* New York: Wiley, 245-265.

McKenize, N.J., and Ryan, P.J., 1999. Spatial prediction of soil properties using environmental correlation. *Geoderma*, 89, 67-94.

McSweeney, K., Slater, B.K., Hammer, R.D., Bell, J.C., Gessler, P.E., and Petersen, G.W., 1994. Towards a new framework for modeling the soil-landscape continuum. In: Amundson R. (Ed.), *Factors of Soil Formation: A Fiftieth Anniversary Publication*, Madison, WI: Soil Society of America, 127-154.

Moore, I.D., Gessler, P.E., Nielsen, G.A., and Peterson, G.A., 1993. Soil attribute prediction using terrain analysis. *Soil Science Society of America Journal*, 57(2), 443–452.

Moore, I.D., Grayson, R.B., and Ladson, A.R., 1991. Digital terrain modeling - a review of hydrological, geomorphological, and biological applications. *Hydrological Processes*, 5(1), 3-30.

Moran, C.J., and Bui, E.N., 2002. Spatial data mining for enhanced soil map modeling**.** *International Journal of Geographical Information Science*, 16, 533–549.

Mulder, V.L., de Bruin, S., Schaepman, M.E., and Mayr, T.R., 2011. The use of remote sensing in soil and terrain mapping — A review. *Geoderma*, 162(1-2), 1-19.

Murphy, P.N.C., Ogilvie, J., Meng, F.-R., and Arp, P., 2008. Stream network modelling using LiDAR and photogrammetric digital elevation models: a comparison and field verification. *Hydrological Processes*, 22 (12), 1747-1754.

Murphy, P. N. C., Ogilvie, J., and Arp, P., 2009. Topographic modelling of soil moisture conditions: A comparison and verification of two models. *European Journal of Soil Science*, 60(1), 94-109.

Nardinocchi, C., Forlani, G., and Zingaretti, P., 2003. Classification and filtering of laser data. *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, 34(3W13), 8p.

Nauman, T., 2009. *Digital Soil-landscape Classification for Soil Survey Using ASTER Satellite and Digital Elevation Data in Organ Pipe Cactus National Monument*, Arizona. M.S. Thesis. University of Arizona.

Nelson, A., Reuter, H.I., and Gessler, P., 2009. DEM production methods and sources. In: Hengl, T. and Reuter, H.I., (Eds), *Geomorphometry: Concepts, Software, Applications*, Elsevier, 65-85.

Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J., 1992. Fuzzy-c-means and kriging for mapping soil as a continuous system. *Soil Science Society of America Jouranal.* 56, 1848-1854.

Odeh, I.O.A., McBratney, A.B., and Chittleborough, D.J., 1994. Spatial prediction of soil properties from landform attributes derived from a digital elevation model. *Geoderma*, 63, 197–214.

Olaya, V., 2009. Basic land-surface parameters. In: Hengl, T. and Reuter, H.I., (Eds), *Geomorphometry: Concepts, Software, Applications,* Elsevier, 141-169.

Owens, K.E., Reed, D.D., Londo, A.J., Maclean, A.L., and Mroz, G.D., 1999. A landscape level comparison of pre-European settlement and current soil carbon content of a forested landscape in upper Michigan. *Forest Ecology and Management,* 113, 179–189.

Park, S.J., and Vlek, P.L.G., 2002. Environmental correlation of three-dimensional soil spatial variability: A comparison of three adaptive techniques. *Geoderma,* 109(1-2), 117-140.

Pike, R.J., Evans, I.S., and Hengl, T., 2009. Geomorphometry: A brief guide. In: Hengl, T. and Reuter, H.I., (Eds), *Geomorphometry: Concepts, Software, Applications,* Elsevier, 3-30.

Powell, B., McBratney, A.B., and MacLeod, D.A., 1991. The application of fuzzy classification to soil pH profiles in the Lockyer Valley, Queensland, Australia. *Catena* , 18, 409-420.

Presant, E.W., and Wickland, R.E., 1971. *The Soils of Waterloo County, Ontario Soil Survey*, Report 44.

Pfeifer, N., and Mandlburger, G., 2009. Lidar data filtering and DTM generation. In: Shan, J., and Toth, K. (Eds.), *Topographic Laser Ranging and Scanning Principles and Processing*. Boca Raton: CRC Press/ Taylor & Francis Group, 307–331.

Qi, F., and Zhu, A.X., 2003. Knowledge discovery from soil maps using inductive learning. *International Journal of Geographical information Science,* 17,771–795.

Quinn, P., Beven, K., Chevallier, P., and Planchon, O., 1991. The prediction of hillslope flow paths for distributed hydrological modeling using digital terrain models. *Hydrological Processes*, 5, 59-79.

Quinn, P., Beven, K. and Lamb, R., 1995. The ln(a/tanb) index: How to calculate it and how to use it within the TOPMODEL framework. *Hydrological Processes*,9, 162–182.

Remmel, T.K., Todd, K.W., and Buttle, J., 2008. A comparison of existing surficial hydrological data layers in a low-relief forested Ontario landscape with those derived from a Lidar DEM. *Forestry Chronicle* 84(6):850-865.

Reynolds, W.D., Drury, C.F., Yang, X.M., and Tan, C.S., 2008. Optimal soil physical quality inferred through structured regression and parameter interactions. *Geoderma,* 146, 466–474.

Roecker, S.M., and Thompson, J.A., 2010. Scale effects on terrain attribute calculation and their use as environmental covariates for digital soil mapping. In Boettinger, J., Howell, D., Moore, A., Hartemink, A., and Kienast-Brown, S., (Eds), *Digital Soil Mapping, Bridging Research, Environmental Applications, and Operation*, London: Springer, 55-66.

Rondeaux, G., Steven, M., and Baret, F., 1996. Optimization of Soil-Adjusted Vegetation Indices. *Remote Sensing of Environment*, 55, 95-107.

Schmidt, J., and Andrew, R., 2005. Multi-scale landform characterization. *Area* 37(3), 341-350.

Schmidt, J. and Dikau, R., 1999. Extracting geomorphometric attributes and objects from digital elevation models – semantics, methods, and future needs. In: Dikau, R., and Sourer, H. (Eds), *GIS for Earth Surface Systems: Analysis and Modelling of the Natural Environment.* Gebrüder Borntraeger, Berlin, 153-226.

Schmidt, J., and Hewitt, A. E., 2004. Fuzzy land element classification from DTMs based on geometry and terrain position. *Geoderma* 121, 243–256.

Schoorl, J.M., Sonneveld, M.P.W., and Veldkamp, A., 2000. Three-dimensional landscape process modelling: The effect of DEM resolution. *Earth Surface Processes and Landforms,* 25(9), 1025-1034.

Schut, P., Smith, S., Fraser, W., Geng, X.Y., and Kroetsch, D., 2011. Soil landscapes of Canada: Building a national framework for environmental information. *Geomatica,* 65(3), 293-309.

Scull, P., Franklin, J., Chadwick, O., and McArthur, D., 2003. Predictive soil mapping: A review. *Progress in Physical Geography*, 27(2), 171-197.

Smith, S., Bulmer, C., Flager, E., Frank, G. and Filatow, D., 2010. Digital soil mapping at multiple scales in British Columbia, Canada. In: *4th Global Workshop on Digital Soil Mapping*, 24-26 May, Rome, Italy, 17.

Smith, M.P., Zhu, A.X., Burt, J.E., and Stiles, C., 2006. The effects of DEM resolution and neighborhood size on digital soil survey. *Geoderma,* 137(1-2), 58-69.

Sommer, M., Wehrhan, M., Zipprich, M., Castell, Z.W., Weller, U., Castell, W., Ehrich, S., Tandler, B., and Selige, T., 2003. Hierarchical data fusion for mapping soil units at field scale. *Geoderma* 112:179–196.

Sumfleth, K., and Duttmann, R., 2008. Prediction of soil property distribution in paddy soil landscapes using terrain data and satellite information as indicators. *Ecological Indicators*, 8(5), 485-501.

Thompson, J.A., Bell, J.C., and Butler, C.A., 2001. Digital elevation model resolution: Effects on terrain attribute calculation and quantitative soil-landscape modeling. *Geoderma* , 100(1-2), 67-89.

Triantafilis, J., and McBratney, A.B., 1993. Application of continuous methods of soil classification and land suitability assessment in the lower Namoi valley. *CSIRO Australia Division of Soils*, Divisional Rep. 121, CSIRO, Melbourne.

Ünsalan, C., and Boyer, K.L., 2011. *Multispectral Satellite Image Understanding: From Land Classification to Building and Road Detection*. London: Springer, 185pp.

Vaze, J., Teng, J., and Spencer, G., 2010. Impact of DEM accuracy and resolution on topographic indices. *Environmental Modelling and Software*, 25(10), 1086-1098.

Vu, T.T., and Tokunaga, M., 2001. Wavelet and scale-space theory in segmentation of airborne laser scanner data. *Proceedings of the 22nd Asian Conference on Remote Sensing*, Singapore, 5p.

Wang, W., and Zhang, Y., 2007. On fuzzy cluster validity indices. *Fuzzy Sets and Systems,* 158(19), 2095-2117.

Wilson, J.P., 2011. Digital terrain modeling. *Geomorphology,* 137(1), 107-121.

Wilson, J.P. and Gallant, J.C., 2000. Digital terrain analysis. In: Wilson, J.P. and Gallant, J.C. (eds.), *Terrain Analysis Principles and Applications*. New York: John Wiley, 1–27

Wilson, J.P., Repetto, P.L. and Snyder, R.D., 2000. Effect of data source, grid resolution, and flow-routing method on computed topographic attributes. In: Wilson, J.P. and Gallant, J.C. (eds.), *Terrain Analysis Principles and Applications*. New York: John Wiley, 133–161.

Wolock, D.M. and Price, C.B., 1994, Effects of digital elevation model and map scale and data resolution on a topography-based watershed model. *Water Resources Research*, 30, 3041–3052.

Wu, S., Li, J., and Huang, G.H., 2008. A study on DEM-derived primary topographic attributes for hydrologic applications: Sensitivity to elevation data resolution. *Applied Geography*, 28(3), 210-223.

Yang, L., 2006. *Extraction of Knowledge on Soil-environment Relationships Using Fuzzy c-means (FCM) Clustering*. M.S. Thesis. Beijing Normal University.

Yang, L., Jiao, Y., Fahmy, S., Zhu, A., Hann, S., Burt, J.E., and Qi, F., 2011. Updating conventional soil maps through digital soil mapping. *Soil Science Society of America,* 75, 1044-1053.

Young, M., 1978. Terrain analysis: program documentation. Report 6 on Grant DA-ERO-591-73-G0040. Statistical characterization of altitude matrices by computer. Department of Geography, University of Durham, Durham, UK, 27.

Zadeh, L.A., 1965. Fuzzy sets. *Information of Control 8*, 338-353.

Zhang, W. and Montgomery, D.R., 1994. Digital elevation model grid size, landscape representation, and hydrologic simulation. *Water Resources Research*, 30, 1019–1028.

Zhu, A.X., 1994. *Soil Pattern Inference Using GIS under Fuzzy Logic*. Ph.D. Thesis, University of Toronto.

Zhu, A.X., 2006. Fuzzy logic models. In: Grunwald, S. (Ed.), *Environmental Soil-Landscape Modeling: Geographic Information Technologies and Pedometrics*, Boca Raton: CRC/Taylor & Francis, 215-240.

Zhu, A X., Band, L., Vertessy, R., and Dutton, B., 1997. Derivation of soil properties using a soil land inference model (SoLIM). *Soil Science Society of America Journal*, 61(2), 523.

Zhu, A.X., Burt, J.E., Smith, M., Wang, R., and Gao, J., 2008b. The impact of neighbourhood size on terrain derivatives and digital soil mapping, In: Zhou Q., Lees B. and Tang G. (Eds.), *Advances in Digital Terrain Analysis*, Berlin: Springer, 333-348.

Zhu, A.X., Yang, L., Li, B.L., Qin, C.Z., English, E., Burt, J.E., and Zhou, C.H., 2008a. Purposive sampling for digital soil mapping for areas with limited data. In: Hartemink, A.E., McBratney, A.B., and Mendonca Santos, M.L. (Eds.), *Digital Soil Mapping with Limited Data*. New York: Springer, 233–245.

Zimmerman, D., Pavlik, C., Ruggles, A., and Armstrong, M.P., 1999. An experimental comparison of ordinary and universal Kriging and inverse distance weighting. *Mathematical Geology* 31, 375–389.

APPENDIX A

The FCM clustering results of the experiments in Group A.

Table A.1 FCM clustering results in Unit 1 area based on 10 m conventional DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	26	2	0.929398	0.070602	0.120283	-0.12028	1533041
	19	3	0.831701	0.097697	0.278084	-0.1578	1216207
	30	4	0.782653	0.049048	0.388312	-0.11023	1058811
	36	5	0.758931	0.023722	0.449154	-0.06084	953621
	36	6	0.753382	0.005549	0.480577	-0.03142	874692.6
	50	7	0.74428	0.009102	0.508232	-0.02766	816909.2
	50	8	0.721199	0.023081	0.555044	-0.04681	773439.4
	50	9	0.726765	-0.00557	0.551638	0.003406	723419.8
	50	10	0.713023	0.013742	0.586859	-0.03522	692992.5
	28	11	0.702939	0.010084	0.617896	-0.03104	665699.3
	50	12	0.701918	0.001021	0.620771	-0.00287	643174.1
	50	13	0.692577	0.009341	0.649801	-0.02903	622123
	50	14	0.683072	0.009505	0.676731	-0.02693	603739.5
	50	15	0.671609	0.011463	0.701284	-0.02455	588842.6
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	21	2	0.781231	0.218769	0.354508	-0.35451	1428935
	20	3	0.67179	0.109441	0.551007	-0.1965	1070556
	46	4	0.565003	0.106787	0.783295	-0.23229	890978.2
	50	5	0.505498	0.059505	0.947566	-0.16427	778491.7
	50	6	0.466053	0.039445	1.074836	-0.12727	700554.2
	50	7	0.43554	0.030513	1.182386	-0.10755	641425.3
	50	8	0.410653	0.024887	1.274344	-0.09196	595273.9
	50	9	0.409801	0.000852	1.316449	-0.04211	555134.5
	50	10	0.396011	0.01379	1.379552	-0.0631	523621.7
	50	11	0.381082	0.014929	1.444622	-0.06507	496975
	50	12	0.358854	0.022228	1.52452	-0.0799	475268.1
	50	13	0.365806	-0.00695	1.535383	-0.01086	452927.8
	50	14	0.348436	0.01737	1.601902	-0.06652	435242.7
	50	15	0.336828	0.011608	1.651369	-0.04947	419011.8
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	19	2	0.66094	0.33906	0.513691	-0.51369	1266796
	25	3	0.540783	0.120157	0.768321	-0.25463	890753.4
	22	4	0.405148	0.135635	1.061649	-0.29333	695894.4
	39	5	0.344865	0.060283	1.264898	-0.20325	578252.6
	20	6	0.280303	0.064562	1.47459	-0.20969	500821.9
	33	7	0.253307	0.026996	1.610316	-0.13573	443390.5
	39	8	0.231081	0.022226	1.729064	-0.11875	399815.4
	36	9	0.209948	0.021133	1.84264	-0.11358	365248.3
	50	10	0.191991	0.017957	1.948067	-0.10543	337057.2
	38	11	0.17457	0.017421	2.049629	-0.10156	313585
	30	12	0.160421	0.014149	2.14303	-0.0934	293632.3
	41	13	0.14901	0.011411	2.223863	-0.08083	276429.2
	49	14	0.144889	0.004121	2.27233	-0.04847	261109.5
	34	15	0.135034	0.009855	2.346547	-0.07422	247857.8
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	3	2	0.50003	0.49997	0.693117	-0.69312	1123909
	31	3	0.445347	0.054683	0.921911	-0.22879	713402.6

	28	4	0.346868	0.098479	1.182663	-0.26075	520943.1
	28	5	0.281282	0.065586	1.405835	-0.22317	410184.7
	26	6	0.233673	0.047609	1.597895	-0.19206	339008.4
	26	7	0.201779	0.031894	1.756556	-0.15866	289309.2
	36	8	0.177113	0.024666	1.894519	-0.13796	252652.1
	32	9	0.159122	0.017991	2.008894	-0.11438	224406.6
	2	10	0.100091	0.059031	2.302132	-0.29324	224759.8
	43	11	0.132973	-0.03288	2.208393	0.093739	183373
	2	12	0.083405	0.049568	2.484478	-0.27609	187306.2
	2	13	0.077004	0.006401	2.564421	-0.07994	172888.4
	29	14	0.105	-0.028	2.451172	0.113249	143938.6
	2	15	0.066729	0.038271	2.70758	-0.25641	149843.6
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	2	2	0.500013	0.499987	0.693135	-0.69314	0.693135
	37	3	0.386813	0.1132	1.015023	-0.32189	556937.5
	36	4	0.306866	0.079947	1.265163	-0.25014	381945.6
	36	5	0.247908	0.058958	1.484581	-0.21942	284811.5
	3	6	0.166818	0.08109	1.791306	-0.30673	239336.4
	34	7	0.177431	-0.01061	1.828688	-0.03738	184571.4
	2	8	0.125159	0.052272	2.078806	-0.25012	167042.5
	2	9	0.111218	0.013941	2.196743	-0.11794	144187.1
	2	10	0.100065	0.011153	2.302261	-0.10552	126397.5
	2	11	0.091071	0.008994	2.397004	-0.09474	112178
	2	12	0.083385	0.007686	2.484595	-0.08759	100641.3
	2	13	0.076981	0.006404	2.564573	-0.07998	91055.25
	2	14	0.071512	0.005469	2.638472	-0.0739	82994.06
	2	15	0.066711	0.004801	2.707719	-0.06925	76144.49

Table A.2 FCM clustering results in Unit 1 area based on 20 m conventional DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	22	2	0.903138	0.096862	0.161122	-0.16112	393129
	21	3	0.824249	0.078889	0.289142	-0.12802	315248.2
	33	4	0.779146	0.045103	0.394095	-0.10495	273986.8
	26	5	0.758314	0.020832	0.450431	-0.05634	246134
	50	6	0.763098	-0.00478	0.464621	-0.01419	224138.8
	50	7	0.743803	0.019295	0.510552	-0.04593	209767.3
	50	8	0.721661	0.022142	0.55829	-0.04774	199129
	36	9	0.726387	-0.00473	0.553581	0.004709	185883.7
	30	10	0.712449	0.013938	0.593682	-0.0401	178392.2
	50	11	0.698205	0.014244	0.625001	-0.03132	171781.5
	44	12	0.703769	-0.00556	0.629493	-0.00449	165323.6
	50	13	0.695684	0.008085	0.645255	-0.01576	159903.2
	50	14	0.701115	-0.00543	0.645021	0.000234	155498.9
	50	15	0.680372	0.020743	0.688408	-0.04339	151152.7
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	15	2	0.765619	0.234381	0.376237	-0.37624	363355.3
	20	3	0.660289	0.10533	0.571825	-0.19559	276332.1
	44	4	0.557299	0.10299	0.796965	-0.22514	230182
	27	5	0.506625	0.050674	0.947765	-0.1508	200575.5
	32	6	0.462821	0.043804	1.082854	-0.13509	180475.2
	50	7	0.431999	0.030822	1.193647	-0.11079	165566.9
	39	8	0.425163	0.006836	1.248906	-0.05526	152982.8
	49	9	0.416222	0.008941	1.304342	-0.05544	142751.3
	44	10	0.39398	0.022242	1.386987	-0.08265	134800.7
	50	11	0.376101	0.017879	1.458795	-0.07181	127977
	50	12	0.367347	0.008754	1.503485	-0.04469	121797.4
	50	13	0.352489	0.014858	1.559417	-0.05593	116833.9
	50	14	0.342492	0.009997	1.62733	-0.06791	112447.7
	50	15	0.338667	0.003825	1.645793	-0.01846	107655.9

m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	14	2	0.654135	0.345865	0.521862	-0.52186	320910.3
	24	3	0.523871	0.130264	0.798373	-0.27651	228596.6
	23	4	0.408527	0.115344	1.062783	-0.26441	179044.1
	19	5	0.333815	0.074712	1.28733	-0.22455	148975
	28	6	0.289836	0.043979	1.458105	-0.17078	128698.9
	24	7	0.250884	0.038952	1.61658	-0.15848	114093.3
	27	8	0.222594	0.02829	1.752548	-0.13597	102912.8
	24	9	0.200289	0.022305	1.873049	-0.1205	94034
	35	10	0.187155	0.013134	1.965195	-0.09215	86750.43
	48	11	0.17677	0.010385	2.046748	-0.08155	80680.87
	41	12	0.162521	0.014249	2.139579	-0.09283	75560.79
	30	13	0.147276	0.015245	2.230041	-0.09046	71179.18
	50	14	0.141648	0.005628	2.289926	-0.05989	67276.5
	50	15	0.136016	0.005632	2.340968	-0.05104	63779.65
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	15	2	0.584741	0.415259	0.603093	-0.60309	276691.7
	27	3	0.428514	0.156227	0.950312	-0.34722	181818.5
	32	4	0.340077	0.088437	1.198917	-0.24861	133668.8
	21	5	0.274774	0.065303	1.42176	-0.22284	105360.4
	26	6	0.23158	0.043194	1.604403	-0.18264	87064.87
	28	7	0.199414	0.032166	1.763245	-0.15884	74276.97
	26	8	0.174688	0.024726	1.901838	-0.13859	64839.73
	25	9	0.155668	0.01902	2.023625	-0.12179	57567.13
	19	10	0.140478	0.01519	2.130091	-0.10647	51784.73
	30	11	0.130009	0.010469	2.222455	-0.09236	47040.03
	25	12	0.119449	0.01056	2.311183	-0.08873	43120.29
	30	13	0.110494	0.008955	2.390341	-0.07916	39810.6
	40	14	0.102522	0.007972	2.46697	-0.07663	36967.96
	41	15	0.09542	0.007102	2.536552	-0.06958	34501.87
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	19	2	0.545595	0.454405	0.645965	-0.64597	235716.5
	31	3	0.378764	0.166831	1.028563	-0.3826	141204.7
	4	4	0.250187	0.128577	1.385922	-0.35736	100289
	28	5	0.242653	0.007534	1.498191	-0.11227	72941.79
	29	6	0.202506	0.040147	1.682507	-0.18432	57626.8
	29	7	0.173647	0.028859	1.839845	-0.15734	47298.64
	24	8	0.15176	0.021887	1.977126	-0.13728	39921.02
	26	9	0.13493	0.01683	2.09743	-0.1203	34406.92
	25	10	0.121752	0.013178	2.203815	-0.10639	30148.09
	23	11	0.1097	0.012052	2.305541	-0.10173	26763.1
	3	12	0.083747	0.025953	2.482425	-0.17688	25385.24
	3	13	0.077454	0.006293	2.5615	-0.07908	22959.24
	46	14	0.088418	-0.01096	2.540405	0.021095	19789.86
	38	15	0.08231	0.006108	2.611337	-0.07093	18160.14

Table A.3 FCM clustering results in Unit 1 area based on 30 m conventional DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	20	2	0.944204	0.055796	0.098092	-0.09809	191104.1
	20	3	0.831058	0.113146	0.278485	-0.18039	152042.9
	47	4	0.78371	0.047348	0.38694	-0.10846	132418
	50	5	0.756341	0.027369	0.447568	-0.06063	119224.7
	50	6	0.747245	0.009096	0.484494	-0.03693	109224.9
	50	7	0.741897	0.005348	0.509698	-0.0252	101047.2
	40	8	0.736993	0.004904	0.523599	-0.0139	94487.68
	50	9	0.734246	0.002747	0.538694	-0.0151	89246.59
	50	10	0.727403	0.006843	0.566374	-0.02768	85345.84

	36	11	0.719218	0.008185	0.587332	-0.02096	81617.12
	42	12	0.714538	0.00468	0.607691	-0.02036	79029.48
	50	13	0.7142	0.000338	0.608938	-0.00125	76284.56
	50	14	0.717684	-0.00348	0.613925	-0.00499	73788.74
	50	15	0.711863	0.005821	0.633867	-0.01994	71527.63
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	20	2	0.802652	0.197348	0.324161	-0.32416	179551.8
	17	3	0.672837	0.129815	0.54829	-0.22413	133958.4
	35	4	0.567004	0.105833	0.77807	-0.22978	111473.9
	23	5	0.50907	0.057934	0.940964	-0.16289	97329.72
	42	6	0.482858	0.026212	1.038767	-0.0978	87037.5
	41	7	0.442074	0.040784	1.162105	-0.12334	79578.63
	30	8	0.409031	0.033043	1.274028	-0.11192	73905.15
	50	9	0.409787	-0.00076	1.314447	-0.04042	68880.58
	50	10	0.407849	0.001938	1.353552	-0.03911	64728.08
	50	11	0.400966	0.006883	1.402132	-0.04858	61305.13
	50	12	0.378806	0.02216	1.485799	-0.08367	58665.02
	50	13	0.370109	0.008697	1.536117	-0.05032	56212.16
	50	14	0.361116	0.008993	1.570421	-0.0343	53821.44
	46	15	0.358128	0.002988	1.601436	-0.03102	51703.77
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	16	2	0.674001	0.325999	0.497564	-0.49756	160109.2
	19	3	0.545895	0.128106	0.759341	-0.26178	111695.5
	26	4	0.42294	0.122955	1.034562	-0.27522	87138.82
	22	5	0.341592	0.081348	1.268175	-0.23361	72479.54
	38	6	0.298274	0.043318	1.435903	-0.16773	62574.18
	26	7	0.266353	0.031921	1.575158	-0.13926	55277.71
	21	8	0.238537	0.027816	1.708135	-0.13298	49739.97
	20	9	0.213277	0.02526	1.835034	-0.1269	45396.67
	31	10	0.196895	0.016382	1.935848	-0.10081	41870.91
	44	11	0.183761	0.013134	2.023949	-0.0881	38944.86
	50	12	0.172625	0.011136	2.100469	-0.07652	36524.97
	50	13	0.159451	0.013174	2.193968	-0.0935	34376.04
	50	14	0.152695	0.006756	2.249124	-0.05516	32503.42
	50	15	0.147154	0.005541	2.30322	-0.0541	30815.49
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	16	2	0.599478	0.400522	0.586686	-0.58669	138754.9
	21	3	0.454283	0.145195	0.906631	-0.31995	89764.94
	21	4	0.351865	0.102418	1.172679	-0.26605	65477.05
	21	5	0.282228	0.069637	1.402845	-0.23017	51510.13
	31	6	0.237895	0.044333	1.586245	-0.1834	42528.2
	37	7	0.20541	0.032485	1.744945	-0.1587	36273.44
	26	8	0.185946	0.019464	1.866594	-0.12165	31615.66
	22	9	0.168712	0.017234	1.981143	-0.11455	28011.91
	34	10	0.152597	0.016115	2.09066	-0.10952	25169.37
	25	11	0.138993	0.013604	2.189843	-0.09918	22867.02
	34	12	0.12784	0.011153	2.279925	-0.09008	20965.56
	31	13	0.118	0.00984	2.364845	-0.08492	19361.12
	25	14	0.109615	0.008385	2.441136	-0.07629	17998.13
	26	15	0.101776	0.007839	2.512946	-0.07181	16816.06
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	17	2	0.553343	0.446657	0.637797	-0.6378	118606.9
	21	3	0.395537	0.157806	1.000034	-0.36224	70399.17
	25	4	0.311407	0.08413	1.255652	-0.25562	48147.08
	26	5	0.251873	0.059534	1.475053	-0.2194	35849.14
	24	6	0.20942	0.042453	1.662571	-0.18752	28283.54
	23	7	0.178603	0.030817	1.824501	-0.16193	23210.66
	23	8	0.157771	0.020832	1.957298	-0.1328	19589.71

	27	9	0.142693	0.015078	2.069792	-0.11249	16880.39
	21	10	0.125576	0.017117	2.19081	-0.12102	14800.47
	38	11	0.118471	0.007105	2.26936	-0.07855	13102.6
	27	12	0.10869	0.009781	2.358486	-0.08913	11750.89
	38	13	0.100238	0.008452	2.442371	-0.08389	10633.16
	31	14	0.092652	0.007586	2.520288	-0.07792	9700.546
	3	15	0.067312	0.02534	2.703205	-0.18292	9705.696

Table A.4 FCM clustering results in Unit 1 area based on 50 m conventional DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	18	2	0.966994	0.033006	0.061967	-0.06197	72682.74
	17	3	0.866114	0.10088	0.226352	-0.16439	55885.79
	24	4	0.809458	0.056656	0.34528	-0.11893	48993.36
	40	5	0.778825	0.030633	0.407764	-0.06248	43833.01
	50	6	0.763396	0.015429	0.455555	-0.04779	40503.55
	50	7	0.765788	-0.00239	0.462891	-0.00734	37580.83
	50	8	0.745762	0.020026	0.51235	-0.04946	35435.11
	50	9	0.745366	0.000396	0.522452	-0.0101	33542.62
	37	10	0.762506	-0.01714	0.49588	0.026572	31400.38
	50	11	0.765055	-0.00255	0.495013	0.000867	29936.82
	50	12	0.756023	0.009032	0.52398	-0.02897	28749.96
	39	13	0.754697	0.001326	0.539873	-0.01589	27645.08
	50	14	0.766338	-0.01164	0.524433	0.01544	26615.31
	50	15	0.7558	0.010538	0.537499	-0.01307	25807.71
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	23	2	0.855655	0.144345	0.248655	-0.24866	69506.98
	16	3	0.714593	0.141062	0.481178	-0.23252	50091.63
	38	4	0.586422	0.128171	0.737666	-0.25649	41769.31
	29	5	0.542546	0.043876	0.871522	-0.13386	36237.98
	50	6	0.482888	0.059658	1.027162	-0.15564	32707.67
	50	7	0.473932	0.008956	1.081364	-0.0542	29728.71
	40	8	0.441637	0.032295	1.183885	-0.10252	27446.07
	36	9	0.426744	0.014893	1.26079	-0.07691	25564.95
	33	10	0.421403	0.005341	1.310129	-0.04934	23961.61
	50	11	0.426034	-0.00463	1.334509	-0.02438	22586.02
	50	12	0.400232	0.025802	1.425083	-0.09057	21551.61
	39	13	0.406768	-0.00654	1.44075	-0.01567	20521.3
	50	14	0.410224	-0.00346	1.456152	-0.0154	19711.14
	50	15	0.389198	0.021026	1.521094	-0.06494	19084.01
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	17	2	0.706577	0.293423	0.458327	-0.45833	63018.61
	19	3	0.588214	0.118363	0.68938	-0.23105	42473.45
	17	4	0.442167	0.146047	0.993494	-0.30411	33033.29
	24	5	0.385436	0.056731	1.182445	-0.18895	27394.08
	26	6	0.327815	0.057621	1.368666	-0.18622	23666.06
	27	7	0.287278	0.040537	1.516311	-0.14765	20927.31
	44	8	0.274331	0.012947	1.599631	-0.08332	18735.25
	27	9	0.245123	0.029208	1.726359	-0.12673	17018.55
	45	10	0.232626	0.012497	1.815249	-0.08889	15631.57
	29	11	0.207966	0.02466	1.934678	-0.11943	14510.54
	50	12	0.213805	-0.00584	1.972212	-0.03753	13533.53
	50	13	0.20741	0.006395	2.038196	-0.06598	12719
	50	14	0.195511	0.011899	2.112738	-0.07454	12017.87
	45	15	0.189549	0.005962	2.171693	-0.05896	11408.32
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	15	2	0.619283	0.380717	0.564264	-0.56426	55042.18
	22	3	0.496433	0.12285	0.837954	-0.27369	34661.45

	19	4	0.371791	0.124642	1.131305	-0.29335	25089.76
	24	5	0.300642	0.071149	1.357729	-0.22642	19727.51
	19	6	0.255436	0.045206	1.541967	-0.18424	16267.54
	25	7	0.222806	0.03263	1.697456	-0.15549	13853.77
	26	8	0.196525	0.026281	1.828379	-0.13092	12058.19
	46	9	0.181209	0.015316	1.927754	-0.09938	10651.5
	31	10	0.166488	0.014721	2.032613	-0.10486	9535.556
	29	11	0.151993	0.014495	2.131793	-0.09918	8637.793
	19	12	0.140749	0.011244	2.225424	-0.09363	7904.329
	27	13	0.130963	0.009786	2.3056	-0.08018	7290.091
	22	14	0.122447	0.008516	2.380685	-0.07509	6767.329
	34	15	0.117102	0.005345	2.447961	-0.06728	6315.574
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	15	2	0.569555	0.430445	0.620243	-0.62024	47243.78
	26	3	0.432715	0.13684	0.940204	-0.31996	27562.83
	22	4	0.32927	0.103445	1.218018	-0.27781	18620.84
	21	5	0.261979	0.067291	1.44734	-0.22932	13850.91
	22	6	0.222007	0.039972	1.628181	-0.18084	10904
	23	7	0.191112	0.030895	1.787613	-0.15943	8936.112
	18	8	0.166889	0.024223	1.926398	-0.13879	7534.423
	27	9	0.152179	0.01471	2.035969	-0.10957	6479.779
	28	10	0.138803	0.013376	2.139567	-0.1036	5666.912
	44	11	0.128729	0.010074	2.225057	-0.08549	5003.071
	23	12	0.118919	0.00981	2.316217	-0.09116	4479.413
	29	13	0.110143	0.008776	2.399039	-0.08282	4049.541
	29	14	0.10204	0.008103	2.47578	-0.07674	3690.758
	28	15	0.096184	0.005856	2.547028	-0.07125	3387.22

APPENDIX B

The FCM clustering results of the experiments in Group B.

Table B.1 FCM clustering results in Unit 2 area based on 10 m conventional DEM with NDVI.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	24	2	0.843523	0.156477	0.258415	-0.25842	4499411
	35	3	0.804117	0.039406	0.352449	-0.09403	3746483
	28	4	0.797914	0.006203	0.372285	-0.01984	3226231
	45	5	0.765668	0.032246	0.434485	-0.0622	2891016
	36	6	0.726149	0.039519	0.522096	-0.08761	2681762
	50	7	0.72838	-0.00223	0.535622	-0.01353	2490018
	38	8	0.725669	0.002711	0.556997	-0.02137	2336227
	50	9	0.706138	0.019531	0.600332	-0.04334	2226737
	50	10	0.707772	-0.00163	0.604794	-0.00446	2123938
	50	11	0.702697	0.005075	0.618362	-0.01357	2046468
	50	12	0.700496	0.002201	0.626623	-0.00826	1965786
	49	13	0.692507	0.007989	0.650947	-0.02432	1895588
	50	14	0.689651	0.002856	0.66547	-0.01452	1834249
	50	15	0.690979	-0.00133	0.663075	0.002395	1762415
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	20	2	0.663206	0.336794	0.507858	-0.50786	3999978
	39	3	0.580117	0.083089	0.729874	-0.22202	3171527
	39	4	0.537334	0.042783	0.868217	-0.13834	2696243
	45	5	0.513792	0.023542	0.948017	-0.0798	2353132
	32	6	0.456075	0.057717	1.09592	-0.1479	2112066
	50	7	0.415042	0.041033	1.224825	-0.12891	1932679
	50	8	0.419933	-0.00489	1.266148	-0.04132	1780567
	50	9	0.387457	0.032476	1.376324	-0.11018	1673006
	50	10	0.367073	0.020384	1.454492	-0.07817	1574711
	50	11	0.36724	-0.00017	1.489596	-0.0351	1490309
	50	12	0.35521	0.01203	1.548373	-0.05878	1420775
	50	13	0.343488	0.011722	1.604471	-0.0561	1359488
	50	14	0.325358	0.01813	1.675077	-0.07061	1307219
	50	15	0.324078	0.00128	1.707403	-0.03233	1257140
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	21	2	0.563623	0.436377	0.625954	-0.62595	3430385
	2	3	0.333392	0.230231	1.098524	-0.47257	2548925
	45	4	0.345506	-0.01211	1.210531	-0.11201	2020907
	50	5	0.286315	0.059191	1.417804	-0.20727	1706922
	2	6	0.16678	0.119535	1.791421	-0.37362	1515562
	50	7	0.220005	-0.05323	1.722772	0.068649	1321828
	50	8	0.210343	0.009662	1.798373	-0.0756	1187828
	50	9	0.192293	0.01805	1.908023	-0.10965	1084076
	50	10	0.172606	0.019687	2.018207	-0.11018	1000610
	2	11	0.090959	0.081647	2.397622	-0.37942	961943.2
	50	12	0.148213	-0.05725	2.193751	0.203871	869150.2
	50	13	0.136446	0.011767	2.283128	-0.08938	821678.5
	2	14	0.0715	0.064946	2.63856	-0.35543	802764.5
	2	15	0.066732	0.004768	2.70756	-0.069	762289
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	2	2	0.50009	0.49991	0.693057	-0.69306	2905120
	3	3	0.333354	0.166736	1.098582	-0.40553	1936767

	3	4	0.25005	0.083304	1.386194	-0.28761	1452570
	3	5	0.200158	0.049892	1.609042	-0.22285	1162028
	2	6	0.166722	0.033436	1.791593	-0.18255	968390.7
	3	7	0.142931	0.023791	1.945651	-0.15406	830034.5
	3	8	0.12506	0.017871	2.079202	-0.13355	726291.3
	2	9	0.11117	0.01389	2.19696	-0.11776	645588.1
	2	10	0.10004	0.01113	2.302387	-0.10543	581038.8
	2	11	0.090937	0.009103	2.397743	-0.09536	528217.5
	3	12	0.083384	0.007553	2.484606	-0.08686	484192
	2	13	0.076963	0.006421	2.564688	-0.08008	446955.5
	2	14	0.071468	0.005495	2.638782	-0.07409	415029.1
	2	15	0.066705	0.004763	2.707766	-0.06898	387362.4
m = 2.25	Iteration Number	Cluster number	*F*	Δ*F* *F*(*c*)-*F*(*c*+1)	*H*	Δ*H* *H*(*c*)-*H*(*c*+1)	J_m
	1	1	1		0		
	2	2	0.500052	0.499948	0.693095	-0.6931	2442928
	3	3	0.333344	0.166708	1.098597	-0.4055	1471627
	3	4	0.250025	0.083319	1.386244	-0.28765	1.386244
	3	5	0.20008	0.049945	1.609238	-0.22299	777125.3
	2	6	0.1667	0.03338	1.791659	-0.18242	1.791659
	3	7	0.142896	0.023804	1.945776	-0.15412	510306.9
	3	8	0.125031	0.017865	2.079317	-0.13354	431863.9
	2	9	0.111149	0.013882	2.197052	-0.11774	372742.2
	2	10	0.100027	0.011122	2.30245	-0.1054	326748.9
	2	11	0.090928	0.009099	2.397791	-0.09534	290049.8
	3	12	0.083361	0.007567	2.484742	-0.08695	260156
	2	13	0.076949	0.006412	2.564778	-0.08004	235392
	2	14	0.071455	0.005494	2.638871	-0.07409	214566.3
	2	15	0.066694	0.004761	2.707846	-0.06898	196838.5

Table B.2 FCM clustering results in Unit 2 area based on conventional DEM without NDVI.

m = 1.25	Iteration Number	Cluster number	*F*	Δ*F* *F*(*c*)-*F*(*c*+1)	*H*	Δ*H* *H*(*c*)-*H*(*c*+1)	J_m
	1	1	1		0		
	28	2	0.851203	0.148797	0.241698	-0.2417	3468140
	30	3	0.833034	0.018169	0.299186	-0.05749	2797131
	42	4	0.799009	0.034025	0.364136	-0.06495	2456003
	50	5	0.78082	0.018189	0.408623	-0.04449	2195969
	32	6	0.764837	0.015983	0.451382	-0.04276	2012657
	46	7	0.76635	-0.00151	0.456421	-0.00504	1846045
	50	8	0.746117	0.020233	0.500376	-0.04396	1736236
	50	9	0.761514	-0.0154	0.472886	0.02749	1640311
	44	10	0.746154	0.01536	0.509719	-0.03683	1547401
	50	11	0.751497	-0.00534	0.507041	0.002678	1458992
	50	12	0.74906	0.002437	0.513956	-0.00692	1387381
	50	13	0.737118	0.011942	0.541468	-0.02751	1328720
	50	14	0.742259	-0.00514	0.537649	0.003819	1268257
	50	15	0.733411	0.008848	0.56156	-0.02391	1229411
m = 1.5	Iteration Number	Cluster number	*F*	*ΔF F(c)-F(c+1)*	*H*	*ΔH H(c)-H(c+1)*	J_m
	1	1	1		0		
	15	2	0.714616	0.285384	0.440514	-0.44051	3102352
	32	3	0.638374	0.076242	0.637908	-0.19739	2408627
	50	4	0.587316	0.051058	0.772687	-0.13478	2028765
	50	5	0.546749	0.040567	0.890694	-0.11801	0.890694
	31	6	0.511862	0.034887	1.000041	-0.10935	1.000041
	50	7	0.476459	0.035403	1.106669	-0.10663	1458022
	50	8	0.47272	0.003739	1.145009	-0.03834	1342665
	50	9	0.45127	0.02145	1.222301	-0.07729	1254957
	50	10	0.447944	0.003326	1.250065	-0.02776	1174543
	43	11	0.449116	-0.00117	1.276356	-0.02629	1104892
	50	12	0.432816	0.0163	1.34441	-0.06805	1049713
	50	13	0.415568	0.017248	1.408099	-0.06369	1005223
	50	14	0.41559	-2.2E-05	1.41819	-0.01009	953589.1
	50	15	0.420639	-0.00505	1.418423	-0.00023	904902.1

m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	14	2	0.620792	0.379208	0.560156	-0.56016	2687527
	2	3	0.333367	0.287425	1.098562	-0.53841	2035962
	37	4	0.423378	-0.09001	1.072867	0.025695	1550493
	37	5	0.374621	0.048757	1.238908	-0.16604	1300945
	29	6	0.334679	0.039942	1.386316	-0.14741	1127017
	50	7	0.294413	0.040266	1.532746	-0.14643	1000244
	40	8	0.270202	0.024211	1.646791	-0.11405	901351.2
	50	9	0.252556	0.017646	1.741308	-0.09452	0.252556
	50	10	0.240682	0.011874	1.823314	-0.08201	757631.8
	50	11	0.228621	0.012061	1.897235	-0.07392	1.897235
	50	12	0.205201	0.02342	2.007327	-0.11009	659701.9
	50	13	0.198431	0.00677	2.064487	-0.05716	619319.9
	50	14	0.187154	0.011277	2.139715	-0.07523	585762.8
	50	15	0.183093	0.004061	2.184186	-0.04447	554286.9
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	19	2	0.56341	0.43659	0.626238	-0.62624	2293540
	2	3	0.333351	0.230059	1.098586	-0.47235	1547002
	37	4	0.326681	0.00667	1.245565	-0.14698	1136157
	40	5	0.271014	0.055667	1.451586	-0.20602	907326.2
	43	6	0.233497	0.037517	1.618958	-0.16737	754771.4
	50	7	0.206231	0.027266	1.760101	-0.14114	646012.2
	50	8	0.184492	0.021739	1.884298	-0.1242	564286.9
	50	9	0.166342	0.01815	1.998024	-0.11373	501001
	49	10	0.15188	0.014462	2.099978	-0.10195	450305.9
	50	11	0.13812	0.01376	2.196218	-0.09624	409108.7
	50	12	0.128117	0.010003	2.281697	-0.08548	374630.9
	50	13	0.118456	0.009661	2.360513	-0.07882	345637.7
	2	14	0.071518	0.046938	2.638433	-0.27792	2.638433
	50	15	0.103438	-0.03192	2.506628	0.131805	2.506628
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	1	1	1		0		
	2	2	0.50011	0.49989	0.693037	-0.69304	1951258
	2	3	0.333344	0.166766	1.098596	-0.40556	1175470
	3	4	0.250084	0.08326	1.386126	-0.28753	820407.7
	27	5	0.215124	0.03496	1.571302	-0.18518	617514.3
	2	6	0.166734	0.04839	1.791558	-0.22026	494215.5
	34	7	0.158149	0.008585	1.897887	-0.10633	404801.8
	3	8	0.125087	0.033062	2.079096	-0.18121	344936.7
	2	9	0.111204	0.013883	2.196807	-0.11771	297708
	35	10	0.111059	0.000145	2.255727	-0.05892	259154.6
	2	11	0.090954	0.020105	2.39765	-0.14192	231671
	2	12	0.083399	0.007555	2.48451	-0.08686	207791.2
	50	13	0.086682	-0.00328	2.511334	-0.02682	186632
	2	14	0.071489	0.015193	2.638632	-0.1273	171374.4
	2	15	0.066727	0.004762	2.7076	-0.06897	157215.5

APPENDIX C

The FCM clustering results of the experiments in Group C.

Table C.1 FCM clustering results in Unit 1 area based on 1 m LiDAR-derived DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	14	2	0.903368		0.1628		393327442
	50	3	0.836297	0.067071	0.2893	-0.12646	334352858
	26	4	0.810182	0.026115	0.345	-0.05578	288180257
	31	5	0.833077	-0.022895	0.3112	0.03387	243899400
	50	6	0.811046	0.022031	0.3588	-0.0476	220481363
	50	7	0.799154	0.011892	0.3855	-0.02676	205318728
	36	8	0.790272	0.008882	0.4078	-0.02227	196877050
	50	9	0.768115	0.022157	0.4597	-0.0519	182348914
	38	10	0.766128	0.001987	0.47	-0.01029	173640791
	50	11	0.754152	0.011976	0.4944	-0.02444	167800493
	41	12	0.764191	-0.010039	0.4758	0.01861	159507855
	50	13	0.749784	0.014407	0.5133	-0.03747	154646613
	50	14	0.746877	0.002907	0.5239	-0.0106	149516234
	50	15	0.756381	-0.009504	0.5055	0.018354	144925103
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500004		0.6931		416720483
	21	3	0.634671	-0.134667	0.6234	0.069783	287545670
	29	4	0.60418	0.030491	0.7394	-0.11607	240509174
	39	5	0.614832	-0.010652	0.7648	-0.02534	205374086
	50	6	0.575676	0.039156	0.865	-0.10025	181902175
	50	7	0.535354	0.040322	0.975	-0.10993	166498177
	50	8	0.510988	0.024366	1.0509	-0.07595	153295137
	50	9	0.491614	0.019374	1.1139	-0.06301	142526406
	50	10	0.460648	0.030966	1.2042	-0.0903	134560163
	50	11	0.448841	0.011807	1.261	-0.05681	127666271
	2	12	0.083345	0.365496	2.4848	-1.22382	170122644
	2	13	0.076929	0.006416	2.5649	-0.08007	163449759
	2	14	0.071438	0.005491	2.639	-0.07409	157503407
	2	15	0.066681	0.004757	2.7079	-0.06895	152161042
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500001		0.6931		350418924
	23	3	0.518786	-0.018785	0.8134	-0.12023	233473866
	2	4	0.250005	0.268781	1.3863	-0.57291	208359728
	46	5	0.387038	-0.137033	1.2072	0.179111	156027517
	50	6	0.35027	0.036768	1.348	-0.14083	135188375
	2	7	0.142865	0.207405	1.9459	-0.59788	136940765
	2	8	0.125004	0.017861	2.0794	-0.13355	123891317
	2	9	0.111119	0.013885	2.1972	-0.11776	113415718
	2	10	0.100002	0.011117	2.3026	-0.10539	104799376
	2	11	0.090913	0.009089	2.3979	-0.0953	97569195.3
	2	12	0.083338	0.007575	2.4849	-0.08701	91405162.9
	2	13	0.076925	0.006413	2.5649	-0.08005	86079668.7
	2	14	0.071432	0.005493	2.639	-0.0741	81425604.9
	2	15	0.066672	0.00476	2.708	-0.06898	77318899.6
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500001		0.6931		294666063
	3	3	0.333341	0.16666	1.0986	-0.40545	196443166
	2	4	0.250003	0.083338	1.3863	-0.28769	147332745
	2	5	0.200005	0.049998	1.6094	-0.22314	117865931
	2	6	0.16667	0.033335	1.7918	-0.18232	98221690.4
	2	7	0.142861	0.023809	1.9459	-0.15415	84189853.9

	2	8	0.125002	0.017859	2.0794	-0.13354	73666317.7
	2	9	0.111116	0.013886	2.1972	-0.11777	65480870.9
	2	10	0.100001	0.011115	2.3026	-0.10537	58933090.3
	2	11	0.090911	0.00909	2.3979	-0.09531	53575424.7
	2	12	0.083336	0.007575	2.4849	-0.08701	49110774.9
	2	13	0.076924	0.006412	2.5649	-0.08005	45333101.4
	2	14	0.071431	0.005493	2.639	-0.0741	42094953.7
	2	15	0.06667	0.004761	2.708	-0.06898	39288415.7
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.5		0.6931		247783655
	2	3	0.333335	0.166665	1.0986	-0.40546	149265058
	2	4	0.250002	0.083333	1.3863	-0.28768	104180041
	2	5	0.200003	0.049999	1.6094	-0.22314	78821814.9
	2	6	0.166669	0.033334	1.7918	-0.18232	62758124.1
	2	7	0.14286	0.023809	1.9459	-0.15415	51759015
	2	8	0.125001	0.017859	2.0794	-0.13354	43802283.2
	2	9	0.111114	0.013887	2.1972	-0.11778	37805480.1
	2	10	0.100001	0.011113	2.3026	-0.10537	33140534.3
	2	11	0.090911	0.00909	2.3979	-0.09531	29418326.4
	2	12	0.083335	0.007576	2.4849	-0.08701	26386523.1
	2	13	0.076924	0.006411	2.5649	-0.08005	23874258.3
	2	14	0.07143	0.005494	2.639	-0.0741	21761986.6
	2	15	0.066669	0.004761	2.708	-0.06899	19963764.4

Table C.2 FCM clustering results in Unit 1 area based on 5 m LiDAR-derived DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	12	2	0.91004		0.152581		15613184
	50	3	0.828254	0.081786	0.298921	-0.14634	13239047
	30	4	0.821608	0.006646	0.324433	-0.02551	11304045
	31	5	0.831485	-0.00988	0.311569	0.012864	9716689
	42	6	0.809577	0.021908	0.35778	-0.04621	8774545
	50	7	0.792213	0.017364	0.394933	-0.03715	8153679
	31	8	0.789129	0.003084	0.410721	-0.01579	7580596
	50	9	0.773154	0.015975	0.448069	-0.03735	7161781
	47	10	0.770133	0.003021	0.457666	-0.0096	6789150
	50	11	0.774074	-0.00394	0.456325	0.001341	6430663
	50	12	0.76207	0.012004	0.481328	-0.025	6214690
	40	13	0.769402	-0.00733	0.470047	0.011281	5905351
	50	14	0.761716	0.007686	0.487873	-0.01783	5702968
	50	15	0.753183	0.008533	0.507191	-0.01932	5537988
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	13	2	0.800314		0.325988		14410456
	16	3	0.647124	0.15319	0.603648	-0.27766	11420588
	27	4	0.617555	0.029569	0.713669	-0.11002	9511319
	50	5	0.612038	0.005517	0.766775	-0.05311	8197186
	39	6	0.581259	0.030779	0.84921	-0.08244	7247452
	50	7	0.529495	0.051764	0.972903	-0.12369	6621757
	50	8	0.517046	0.012449	1.026594	-0.05369	6090320
	50	9	0.503953	0.013093	1.081791	-0.0552	5627535
	50	10	0.490779	0.013174	1.137321	-0.05553	5274321
	36	11	0.480533	0.010246	1.183979	-0.04666	4975898
	50	12	0.471717	0.008816	1.223991	-0.04001	4717564
	50	13	0.455663	0.016054	1.27844	-0.05445	4509059
	50	14	0.4592	-0.00354	1.291826	-0.01339	4291020
	50	15	0.441389	0.017811	1.350886	-0.05906	4133170
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	15	2	0.706099		0.456674		12810853
	16	3	0.527088	0.179011	0.7993	-0.34263	9311679
	32	4	0.463054	0.064034	0.995586	-0.19629	7389924
	46	5	0.394045	0.069009	1.186474	-0.19089	6204674
	50	6	0.37339	0.020655	1.298577	-0.1121	5351280

	50	7	0.35297	0.02042	1.391387	-0.09281	4720076
	50	8	0.330435	0.022535	1.486625	-0.09524	4244524
	50	9	0.300048	0.030387	1.601969	-0.11534	3867706
	50	10	0.291966	0.008082	1.663207	-0.06124	3545667
	50	11	0.269795	0.022171	1.753207	-0.09	3287343
	50	12	0.264839	0.004956	1.80551	-0.0523	3062177
	50	13	0.248149	0.01669	1.884616	-0.07911	2879780
	50	14	0.240042	0.008107	1.939016	-0.0544	2710297
	50	15	0.235071	0.004971	1.986211	-0.04719	2566196
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	17	2	0.637922		0.541736		11143121
	18	3	0.452995	0.184927	0.915917	-0.37418	7372980
	45	4	0.365205	0.08779	1.16864	-0.25272	5500449
	44	5	0.30192	0.063285	1.377878	-0.20924	4389640
	50	6	0.260843	0.041077	1.54444	-0.16656	3651344
	50	7	0.231306	0.029537	1.685693	-0.14125	3124940
	50	8	0.203578	0.027728	1.818222	-0.13253	2730954
	50	9	0.193078	0.0105	1.909	-0.09078	2422327
	50	10	0.167899	0.025179	2.035136	-0.12614	2183339
	50	11	0.159999	0.0079	2.107778	-0.07264	1980871
	50	12	0.150787	0.009212	2.182266	-0.07449	1812032
	50	13	0.138722	0.012065	2.266105	-0.08384	1673555
	50	14	0.12953	0.009192	2.337945	-0.07184	1551161
	50	15	0.12058	0.00895	2.410415	-0.07247	1448659
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500012		0.693135		9988386
	20	3	0.408165	0.091847	0.984872	-0.29174	5744124
	37	4	0.313504	0.094661	1.261871	-0.277	4001385
	39	5	0.252396	0.061108	1.48361	-0.22174	3025546
	50	6	0.212568	0.039828	1.66316	-0.17955	2408031
	34	7	0.183007	0.029561	1.816026	-0.15287	1985412
	43	8	0.161174	0.021833	1.947727	-0.1317	1679769
	50	9	0.144515	0.016659	2.061591	-0.11386	1449527
	50	10	0.129399	0.015116	2.168555	-0.10696	1270776
	48	11	0.117522	0.011877	2.265477	-0.09692	1127859
	50	12	0.108655	0.008867	2.349765	-0.08429	1011543
	3	13	0.076982	0.031673	2.564565	-0.2148	962229.3
	2	14	0.071462	0.00552	2.638827	-0.07426	877166.4
	2	15	0.066693	0.004769	2.70785	-0.06902	804712.9

Table C.3 FCM clustering results in Unit 1 area based on 10 m LiDAR-derived DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	12	2	0.912007		0.148731		3838980
	50	3	*0.8469*	0.06511	*0.26993*	-0.12119	*3238855*
	26	4	0.82689	0.020007	0.314028	-0.0441	2738223
	50	5	0.80217	0.02472	0.366334	-0.05231	2468987
	50	6	0.799943	0.002227	0.379119	-0.01279	2221415
	50	7	0.792526	0.007417	0.403259	-0.02414	2064308
	50	8	0.789739	0.002787	0.411765	-0.00851	1910341
	50	9	0.760073	0.029666	0.473733	-0.06197	1844362
	48	10	0.777267	-0.01719	0.448457	0.025276	1709991
	50	11	0.767297	0.00997	0.46993	-0.02147	1639330
	50	12	0.762839	0.004458	0.481784	-0.01185	1573809
	50	13	0.77013	-0.00729	0.470013	0.011771	1502174
	50	14	0.76614	0.00399	0.481648	-0.01164	1447069
	50	15	0.770281	-0.00414	0.47131	0.010338	1379567
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	12	2	0.806218		0.317126		3548574
	13	3	0.656787	0.149431	0.588389	-0.27126	2802196
	21	4	0.629038	0.027749	0.692486	-0.1041	2317853
	50	5	0.573338	0.0557	0.827194	-0.13471	2026947

	50	6	0.553366	0.019972	0.901887	-0.07469	1807405
	37	7	0.533455	0.019911	0.977727	-0.07584	1648913
	50	8	0.525675	0.00778	1.019472	-0.04174	1513278
	50	9	0.505377	0.020298	1.0887	-0.06923	1406654
	36	10	0.491656	0.013721	1.146382	-0.05768	1323263
	50	11	0.486016	0.00564	1.183694	-0.03731	1250668
	48	12	0.471358	0.014658	1.234816	-0.05112	1187180
	50	13	0.464792	0.006566	1.268455	-0.03364	1130912
	50	14	0.44433	0.020462	1.335062	-0.06661	1088381
	50	15	0.436301	0.008029	1.370266	-0.0352	1044752
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	14	2	0.713491		0.447008		3162200
	15	3	0.537656	0.175835	0.783258	-0.33625	2292356
	34	4	0.475496	0.06216	0.973951	-0.19069	1812845
	50	5	0.408364	0.067132	1.160075	-0.18612	1521042
	50	6	0.365866	0.042498	1.302184	-0.14211	1316784
	50	7	0.34719	0.018676	1.403359	-0.10118	1164976
	40	8	0.325283	0.021907	1.504322	-0.10096	1047321
	50	9	0.310684	0.014599	1.581639	-0.07732	953119.8
	50	10	0.289602	0.021082	1.676238	-0.0946	878180.5
	50	11	0.269343	0.020259	1.765325	-0.08909	816470.7
	50	12	0.26096	0.008383	1.821835	-0.05651	761293.5
	50	13	0.249459	0.011501	1.890919	-0.06908	715450.2
	50	14	0.241644	0.007815	1.94841	-0.05749	675080.1
	50	15	0.235097	0.006547	1.993422	-0.04501	640220.1
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	16	2	0.645286		0.532907		2756434
	18	3	0.462204	0.183082	0.902294	-0.36939	1820559
	30	4	0.376342	0.085862	1.149611	-0.24732	1355803
	39	5	0.307671	0.068671	1.368267	-0.21866	1084103
	50	6	0.269378	0.038293	1.523636	-0.15537	899413.9
	36	7	0.236678	0.0327	1.670163	-0.14653	769978.4
	50	8	0.205873	0.030805	1.809458	-0.1393	673818.1
	50	9	0.196877	0.008996	1.894851	-0.08539	597181.1
	45	10	0.173887	0.02299	2.009438	-0.11459	537602.2
	50	11	0.164976	0.008911	2.089258	-0.07982	488197.3
	50	12	0.145775	0.019201	2.191681	-0.10242	447834.8
	50	13	0.146112	-0.00034	2.237238	-0.04556	411764.2
	50	14	0.135728	0.010384	2.315289	-0.07805	382332.8
	50	15	0.128278	0.00745	2.380259	-0.06497	356598.7
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	18	2	0.598496		0.587709		2371589
	21	3	0.415281	0.183215	0.974314	-0.38661	1421734
	33	4	0.321924	0.093357	1.247375	-0.27306	989584.8
	44	5	0.260424	0.0615	1.46787	-0.2205	748181.6
	28	6	0.217135	0.043289	1.650803	-0.18293	595478.7
	50	7	0.189252	0.027883	1.798232	-0.14743	490875
	45	8	0.166834	0.022418	1.929258	-0.13103	415283.6
	42	9	0.14733	0.019504	2.050496	-0.12124	358453
	50	10	0.133106	0.014224	2.155238	-0.10474	314169.5
	50	11	0.12244	0.010666	2.245671	-0.09043	278824.1
	43	12	0.111514	0.010926	2.334764	-0.08909	250093
	38	13	0.102276	0.009238	2.418327	-0.08356	226308.8
	39	14	0.095785	0.006491	2.490231	-0.0719	206269.9
	41	15	0.089776	0.006009	2.556289	-0.06606	189189

Table C.4 FCM clustering results in Unit 1 area based on 20 m LiDAR-derived DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	13	2	0.913049		0.145682		1027312
	50	3	0.83502	0.078032	0.28043	-0.13475	856089
	18	4	0.833108	0.001909	0.303549	-0.02312	719344.3

	25	5	0.808547	0.024561	0.354596	-0.05105	640114.4
	26	6	0.802841	0.005706	0.372708	-0.01811	580385.9
	38	7	0.793216	0.009625	0.398859	-0.02615	537834
	50	8	0.792105	0.001111	0.408068	-0.00921	501276.1
	50	9	0.791506	0.000599	0.412338	-0.00427	467888
	50	10	0.789676	0.00183	0.421627	-0.00929	442575.3
	30	11	0.790858	-0.00118	0.423333	-0.00171	420328.9
	50	12	0.787286	0.003572	0.433814	-0.01048	402371.8
	50	13	0.783175	0.004111	0.443099	-0.00929	384086
	50	14	0.778979	0.004196	0.45593	-0.01283	371221.4
	50	15	0.783336	-0.00436	0.446989	0.008941	355889.3
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	14	2	0.810821		0.310291		951631.1
	12	3	0.66486	0.145961	0.573145	-0.26285	745651.7
	19	4	0.636121	0.028739	0.677875	-0.10473	613487.8
	30	5	0.59721	0.038911	0.785365	-0.10749	531808.9
	32	6	0.567192	0.030018	0.87329	-0.08793	474065
	50	7	0.542253	0.024939	0.951963	-0.07867	431876.8
	50	8	0.528651	0.013602	1.015402	-0.06344	400002.4
	41	9	0.519517	0.009134	1.059599	-0.0442	368231.3
	50	10	0.503883	0.015634	1.113896	-0.0543	346099.3
	50	11	0.497087	0.006796	1.153308	-0.03941	325622.8
	50	12	0.503065	-0.00598	1.163146	-0.00984	308415.6
	50	13	0.487289	0.015776	1.215779	-0.05263	294115.3
	50	14	0.474169	0.01312	1.260539	-0.04476	281058.7
	50	15	0.472335	0.001834	1.285077	-0.02454	269191.3
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	15	2	0.717847		0.441347		849998.6
	13	3	0.546821	0.171026	0.768106	-0.32676	612823.7
	23	4	0.485834	0.060987	0.954597	-0.18649	482611.7
	19	5	0.407302	0.078532	1.156956	-0.20236	406013.8
	36	6	0.392189	0.015113	1.252235	-0.09528	347739.2
	37	7	0.363919	0.02827	1.363977	-0.11174	307621.4
	50	8	0.341208	0.022711	1.465737	-0.10176	276223.5
	43	9	0.324635	0.016573	1.547848	-0.08211	250913.4
	50	10	0.314307	0.010328	1.618239	-0.07039	230625.6
	50	11	0.290364	0.023943	1.711873	-0.09363	214334.5
	50	12	0.282113	0.008251	1.768813	-0.05694	200122.3
	50	13	0.271003	0.01111	1.836272	-0.06746	187898.5
	50	14	0.25964	0.011363	1.896465	-0.06019	177258.5
	50	15	0.262323	-0.00268	1.923199	-0.02673	167370.7
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	3	2	0.500011		0.693136		802737.6
	14	3	0.470272	0.029739	0.889927	-0.19679	488308.7
	27	4	0.387724	0.082548	1.130047	-0.24012	362583.3
	50	5	0.328396	0.059328	1.327585	-0.19754	288764.1
	29	6	0.284502	0.043894	1.493012	-0.16543	239727.4
	34	7	0.257252	0.02725	1.623475	-0.13046	204797.9
	42	8	0.2343	0.022952	1.740207	-0.11673	178710.1
	48	9	0.2195	0.0148	1.835902	-0.09569	158280.3
	34	10	0.201453	0.018047	1.936837	-0.10094	142142
	50	11	0.186999	0.014454	2.026557	-0.08972	129035.5
	50	12	0.17573	0.011269	2.105556	-0.079	118095.5
	50	13	0.166303	0.009427	2.177838	-0.07228	108862.1
	50	14	0.156348	0.009955	2.250721	-0.07288	100996.4
	50	15	0.149013	0.007335	2.312773	-0.06205	94172.63
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	3	2	0.500003		0.693144		675020.8
	16	3	0.422463	0.07754	0.963998	-0.27085	382219.7
	29	4	0.330408	0.092055	1.232659	-0.26866	265576.6
	41	5	0.268424	0.061984	1.451645	-0.21899	200632.3
	32	6	0.227239	0.041185	1.627628	-0.17598	159585.2

	44	7	0.198628	0.028611	1.774212	-0.14658	131522.3
	33	8	0.173862	0.024766	1.908835	-0.13462	111271.7
	48	9	0.160664	0.013198	2.009749	-0.10091	95898.16
	44	10	0.147696	0.012968	2.107755	-0.09801	83973.65
	50	11	0.137468	0.010228	2.195622	-0.08787	74463.8
	50	12	0.126363	0.011105	2.284058	-0.08844	66735.63
	42	13	0.11659	0.009773	2.365942	-0.08188	60334.18
	49	14	0.110572	0.006018	2.43362	-0.06768	54958.25
	50	15	0.102645	0.007927	2.5053	-0.07168	50393.01

Table C.5 FCM clustering results in Unit 1 area based on 30 m LiDAR-derived DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	11	2	0.920252		0.134678		522513.1
	50	3	0.849566	0.070686	0.25796	-0.12328	431683.3
	32	4	0.817274	0.032292	0.322184	-0.06422	369221.8
	50	5	0.803781	0.013493	0.360536	-0.03835	329212.3
	50	6	0.802122	0.001659	0.381402	-0.02087	301453.7
	50	7	0.806272	-0.00415	0.379444	0.001958	276323.4
	48	8	0.810471	-0.0042	0.369907	0.009537	248003.1
	50	9	0.810189	0.000282	0.37694	-0.00703	231311
	50	10	0.789748	0.020441	0.419008	-0.04207	221076.4
	50	11	0.798249	-0.0085	0.403894	0.015114	206005
	50	12	0.79489	0.003359	0.415007	-0.01111	193927.1
	36	13	0.810712	-0.01582	0.390967	0.02404	185454.5
	50	14	0.79132	0.019392	0.429221	-0.03825	177080.6
	36	15	0.804697	-0.01338	0.406843	0.022378	170790.3
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	11	2	0.818264		0.299422		484456.5
	21	3	0.677158	0.141106	0.552137	-0.25272	377623.4
	32	4	0.63003	0.047128	0.681175	-0.12904	313384.8
	34	5	0.596696	0.033334	0.784163	-0.10299	271784.7
	50	6	0.549777	0.046919	0.906527	-0.12236	244884
	50	7	0.552345	-0.00257	0.941741	-0.03521	221005.4
	50	8	0.538548	0.013797	0.996117	-0.05438	203363
	50	9	0.531604	0.006944	1.03224	-0.03612	188618.8
	49	10	0.539194	-0.00759	1.027466	0.004774	171154.6
	50	11	0.518828	0.020366	1.09306	-0.06559	162032.7
	34	12	0.518026	0.000802	1.109895	-0.01684	151907
	46	13	0.519908	-0.00188	1.126202	-0.01631	143161.5
	50	14	0.499944	0.019964	1.178908	-0.05271	136785.9
	40	15	0.493749	0.006195	1.213063	-0.03415	130627.9
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	11	2	0.725008		0.431747		432785.6
	19	3	0.556111	0.168897	0.752664	-0.32092	310855.7
	41	4	0.482365	0.073746	0.955482	-0.20282	245960
	35	5	0.43381	0.048555	1.110861	-0.15538	204772.4
	44	6	0.383258	0.050552	1.268488	-0.15763	177454
	43	7	0.366644	0.016614	1.365462	-0.09697	156925.4
	50	8	0.33921	0.027434	1.469465	-0.104	141100.4
	43	9	0.332572	0.006638	1.535168	-0.0657	128210.9
	43	10	0.312947	0.019625	1.619245	-0.08408	118065.2
	50	11	0.297928	0.015019	1.696568	-0.07732	109656.5
	50	12	0.281086	0.016842	1.767874	-0.07131	102421.8
	48	13	0.286287	-0.0052	1.790422	-0.02255	95831.63
	50	14	0.275394	0.010893	1.850726	-0.0603	90304.54
	48	15	0.269042	0.006352	1.896441	-0.04572	85634.63
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	11	2	0.655073		0.521004		377988.2
	22	3	0.478192	0.176881	0.877054	-0.35605	247863
	32	4	0.382573	0.095619	1.135481	-0.25843	184619
	30	5	0.328326	0.054247	1.32207	-0.18659	146577.2

	28	6	0.276961	0.051365	1.50447	-0.1824	121780.6
	39	7	0.251451	0.02551	1.633785	-0.12932	104093.4
	42	8	0.225574	0.025877	1.755439	-0.12165	90899.16
	45	9	0.212397	0.013177	1.852415	-0.09698	80564.95
	50	10	0.193837	0.01856	1.958058	-0.10564	72432.39
	44	11	0.175875	0.017962	2.059257	-0.1012	65832.26
	50	12	0.167266	0.008609	2.129743	-0.07049	60295.5
	50	13	0.160291	0.006975	2.197142	-0.0674	55612.89
	50	14	0.148155	0.012136	2.275914	-0.07877	51670.16
	50	15	0.13916	0.008995	2.343278	-0.06736	48210.05
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	12	2	0.606901		0.578124		325726.2
	28	3	0.428633	0.178268	0.953841	-0.37572	194162.2
	27	4	0.329038	0.099595	1.233355	-0.27951	135091.5
	22	5	0.271171	0.057867	1.443171	-0.20982	101921.8
	31	6	0.229051	0.04212	1.62107	-0.1779	80988.95
	36	7	0.198224	0.030827	1.773269	-0.1522	66739.12
	35	8	0.176755	0.021469	1.899538	-0.12627	56427.75
	47	9	0.159017	0.017738	2.011805	-0.11227	48641.01
	44	10	0.146382	0.012635	2.109982	-0.09818	42591.46
	35	11	0.133592	0.01279	2.206341	-0.09636	37778.48
	32	12	0.121389	0.012203	2.297402	-0.09106	33878.3
	35	13	0.113995	0.007394	2.371775	-0.07437	30643.7
	29	14	0.106313	0.007682	2.449201	-0.07743	27926.04
	33	15	0.099874	0.006439	2.5157	-0.0665	25614.75

Table C.6 FCM clustering results in Unit 1 area based on 50 m LiDAR-derived DEM.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	10	2	0.910742		0.150141		201500
	50	3	0.857019	0.053723	0.247161	-0.09702	*165560*
	27	4	0.847435	0.009584	0.2784	-0.03124	138390.4
	48	5	0.852426	-0.00499	0.274513	0.003887	117105.7
	44	6	0.839959	0.012467	0.30458	-0.03007	105058.8
	50	7	0.824591	0.015368	0.337996	-0.03342	96190.99
	50	8	0.828546	-0.00396	0.339736	-0.00174	88078.98
	50	9	0.848077	-0.01953	0.30768	0.032056	80371.35
	37	10	0.844137	0.00394	0.318459	-0.01078	74000.21
	50	11	0.838198	0.005939	0.332877	-0.01442	72094.21
	25	12	0.860389	-0.02219	0.289155	0.043722	64067.73
	50	13	0.860352	3.7E-05	0.29723	-0.00808	61328.36
	50	14	0.836932	0.02342	0.339617	-0.04239	59164.32
	28	15	0.857412	-0.02048	0.30058	0.039037	56102.7
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	8	2	0.805551		0.316835		185725.7
	42	3	0.682976	0.122575	0.544783	-0.22795	144712.7
	14	4	0.653757	0.029219	0.647444	-0.10266	118304.6
	35	5	0.618433	0.035324	0.751857	-0.10441	103149.1
	50	6	0.628877	-0.01044	0.738097	0.01376	87177.87
	50	7	0.603586	0.025291	0.822452	-0.08436	78186.53
	26	8	0.582333	0.021253	0.886052	-0.0636	70682.36
	50	9	0.608615	-0.02628	0.867346	0.018706	64988.7
	50	10	0.579305	0.02931	0.94124	-0.07389	59967.73
	50	11	0.547668	0.031637	1.029603	-0.08836	57038.38
	44	12	0.567203	-0.01954	1.002736	0.026867	52436.96
	39	13	0.565403	0.0018	1.01841	-0.01567	49423.84
	50	14	0.567487	-0.00208	1.029316	-0.01091	46445.49
	50	15	0.564579	0.002908	1.04268	-0.01336	44089.39
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	8	2	0.715009		0.443638		164891.1
	20	3	0.554702	0.160307	0.75451	-0.31087	118602.1
	12	4	0.512527	0.042175	0.911678	-0.15717	92519.48

	29	5	0.45355	0.058977	1.08375	-0.17207	77472.98
	47	6	0.41136	0.04219	1.228059	-0.14431	67175.15
	32	7	0.374506	0.036854	1.348566	-0.12051	59673.04
	43	8	0.373517	0.000989	1.394236	-0.04567	52974.02
	28	9	0.340327	0.03319	1.525404	-0.13117	49009.31
	40	10	0.343628	-0.0033	1.553254	-0.02785	44238.67
	43	11	0.329126	0.014502	1.626599	-0.07335	41086.16
	50	12	0.338131	-0.00901	1.64344	-0.01684	38089.49
	40	13	0.349342	-0.01121	1.644748	-0.00131	35433.96
	50	14	0.33449	0.014852	1.709541	-0.06479	33366.17
	50	15	0.32358	0.01091	1.768929	-0.05939	31635.03
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	8	2	0.648233		0.528461		143385.9
	15	3	0.478089	0.170144	0.877797	-0.34934	94114.34
	14	4	0.416634	0.061455	1.084114	-0.20632	69310.35
	24	5	0.352058	0.064576	1.287025	-0.20291	55181.93
	40	6	0.308772	0.043286	1.451792	-0.16477	45860.35
	19	7	0.268539	0.040233	1.606023	-0.15423	39333.84
	40	8	0.260161	0.008378	1.690194	-0.08417	34254.28
	34	9	0.244922	0.015239	1.787509	-0.09732	30398.83
	34	10	0.223624	0.021298	1.893543	-0.10603	27355.68
	50	11	0.220432	0.003192	1.938015	-0.04447	24560.73
	32	12	0.2117	0.008732	2.018296	-0.08028	22473.17
	39	13	0.209079	0.002621	2.068802	-0.05051	20711.53
	41	14	0.210272	-0.00119	2.103812	-0.03501	19137.69
	50	15	0.201252	0.00902	2.173608	-0.0698	17870.48
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	8	2	0.601403		0.583951		123208
	17	3	0.429413	0.17199	0.953334	-0.36938	73492.73
	16	4	0.355024	0.074389	1.193247	-0.23991	50762.99
	24	5	0.294774	0.06025	1.403745	-0.2105	38257.77
	27	6	0.245191	0.049583	1.591702	-0.18796	30449.56
	21	7	0.222223	0.022968	1.725871	-0.13417	25064.93
	28	8	0.19492	0.027303	1.862207	-0.13634	21215.95
	23	9	0.17847	0.01645	1.975122	-0.11292	18301.44
	50	10	0.177484	0.000986	2.040508	-0.06539	16010.98
	31	11	0.154637	0.022847	2.15427	-0.11376	14224.36
	48	12	0.160454	-0.00582	2.186558	-0.03229	12681.01
	50	13	0.154752	0.005702	2.253617	-0.06706	11454.32
	50	14	0.152174	0.002578	2.307344	-0.05373	10421.47
	50	15	0.139701	0.012473	2.392944	-0.0856	9560.609

APPENDIX D

The FCM clustering results of the experiments in Group D.

Table D.1 FCM clustering results in Unit 2 area based on 10 m LiDAR-derived DEM with NDVI.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	16	2	0.865008		0.224155		9430906
	22	3	0.82877	0.036238	0.304063	-0.07991	*7678461*
	50	4	0.772601	0.056169	0.417888	-0.11383	6934210
	42	5	0.749121	0.02348	0.477659	-0.05977	6360727
	40	6	0.717778	0.031343	0.550011	-0.07235	5991687
	50	7	0.719513	-0.00173	0.551394	-0.00138	5534865
	50	8	0.72158	-0.00207	0.551521	-0.00013	5192402
	50	9	0.722385	-0.00081	0.562142	-0.01062	4938665
	50	10	0.712522	0.009863	0.585658	-0.02352	4684860
	50	11	0.704284	0.008238	0.611992	-0.02633	4496432
	50	12	0.701543	0.002741	0.632795	-0.0208	4402048
	50	13	0.698049	0.003494	0.635295	-0.0025	4158041
	50	14	0.701108	-0.00306	0.63919	-0.00389	4007472
	50	15	0.702052	-0.00094	0.641533	-0.00234	3859835
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	14	2	0.727008		0.426339		8496134
	23	3	0.646918	0.08009	0.624544	-0.19821	6588596
	27	4	0.541297	0.105621	0.859687	-0.23514	5679313
	48	5	0.490158	0.051139	1.002809	-0.14312	5018773
	45	6	0.446649	0.043509	1.138879	-0.13607	4562072
	50	7	0.40527	0.041379	1.263425	-0.12455	4209675
	50	8	0.382334	0.022936	1.35402	-0.0906	3921969
	49	9	0.381341	0.000993	1.388469	-0.03445	3665049
	50	10	0.369572	0.011769	1.454386	-0.06592	3459576
	50	11	0.353348	0.016224	1.519866	-0.06548	3288806
	50	12	0.339105	0.014243	1.583987	-0.06412	3138621
	50	13	0.325099	0.014006	1.646805	-0.06282	3007447
	50	14	0.333228	-0.00813	1.652232	-0.00543	2879956
	50	15	0.330548	0.00268	1.691638	-0.03941	2775617
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	15	2	0.628441		0.551828		7393730
	3	3	0.333598	0.294843	1.098214	-0.54639	5637736
	30	4	0.399259	-0.06566	1.114772	-0.01656	4291464
	30	5	0.335243	0.064016	1.314382	-0.19961	3620733
	50	6	0.284322	0.050921	1.491193	-0.17681	3157586
	50	7	0.252277	0.032045	1.633519	-0.14233	2810540
	50	8	0.224814	0.027463	1.763413	-0.12989	2542213
	43	9	0.204443	0.020371	1.869489	-0.10608	2325547
	46	10	0.181785	0.022658	1.983196	-0.11371	2149572
	43	11	0.167601	0.014184	2.072995	-0.0898	2000817
	40	12	0.156622	0.010979	2.151712	-0.07872	1873381
	2	13	0.076995	0.079627	2.564484	-0.41277	1877029
	2	14	0.07151	0.005485	2.63849	-0.07401	1775522
	2	15	0.066788	0.004722	2.707142	-0.06865	1685862
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500054		0.693094		6425881
	3	3	0.33343	0.166624	1.098467	-0.40537	4283965
	3	4	0.250039	0.083391	1.386216	-0.28775	3212962
	26	5	0.255479	-0.00544	1.483907	-0.09769	2509916
	3	6	0.166793	0.088686	1.79138	-0.30747	2141797
	3	7	0.143001	0.023792	1.945407	-0.15403	1835801

	2	8	0.125072	0.017929	2.079152	-0.13375	1606405
	30	9	0.142183	-0.01711	2.079766	-0.00061	1396096
	2	10	0.100069	0.042114	2.302238	-0.22247	1285130
	2	11	0.090993	0.009076	2.397433	-0.09519	1168269
	2	12	0.083402	0.007591	2.484494	-0.08706	1070915
	2	13	0.076963	0.006439	2.564692	-0.0802	988574.7
	2	14	0.071474	0.005489	2.638738	-0.07405	917957.9
	2	15	0.066738	0.004736	2.707518	-0.06878	856731.4
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500032		0.693115		5403539
	2	3	0.333391	0.166641	1.098526	-0.40541	3255183
	3	4	0.250019	0.083372	1.386257	-0.28773	2271929
	2	5	0.200089	0.04993	1.609215	-0.22296	1718879
	3	6	0.166729	0.03336	1.791572	-0.18236	1368564
	3	7	0.14293	0.023799	1.945654	-0.15408	1128710
	2	8	0.125047	0.017883	2.079255	-0.1336	955210.3
	2	9	0.111174	0.013873	2.196944	-0.11769	824420.4
	2	10	0.100046	0.011128	2.302357	-0.10541	722715.6
	2	11	0.090969	0.009077	2.397569	-0.09521	641533.2
	2	12	0.083384	0.007585	2.484605	-0.08704	575414.4
	2	13	0.076951	0.006433	2.564771	-0.08017	520640.2
	2	14	0.07146	0.005491	2.638837	-0.07407	474578.2
	2	15	0.066718	0.004742	2.707662	-0.06882	435361.4

Table D.2 FCM clustering results in Unit 2 area based on 10 m LiDAR-derived DEM without NDVI.

m = 1.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	15	2	0.880965		0.198191		8036673
	22	3	0.853376	0.027589	0.259823	-0.06163	*6317678*
	42	4	0.807402	0.045974	0.352389	-0.09257	5619001
	49	5	0.789723	0.017679	0.39667	-0.04428	5073561
	20	6	0.761873	0.02785	0.456395	-0.05973	4745396
	50	7	0.767547	-0.00567	0.454967	0.001428	4363066
	50	8	0.773816	-0.00627	0.444056	0.010911	3988072
	50	9	0.772933	0.000883	0.45317	-0.00911	3751905
	50	10	0.758199	0.014734	0.486498	-0.03333	3543421
	50	11	0.752875	0.005324	0.503586	-0.01709	3379222
	50	12	0.752968	-9.3E-05	0.508047	-0.00446	3243571
	50	13	0.749538	0.00343	0.520217	-0.01217	3096582
	50	14	0.747786	0.001752	0.522918	-0.0027	3003774
	50	15	0.75104	-0.00325	0.522819	9.9E-05	2856024
m = 1.5	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	13	2	0.756981		0.384804		7289732
	25	3	0.691804	0.065177	0.550444	-0.16564	5495251
	50	4	0.602813	0.088991	0.748128	-0.19768	4692323
	50	5	0.561465	0.041348	0.866259	-0.11813	4104538
	39	6	0.520801	0.040664	0.983279	-0.11702	3710737
	50	7	0.477126	0.043675	1.096603	-0.11332	3414766
	47	8	0.459553	0.017573	1.170369	-0.07377	3163540
	50	9	0.449676	0.009877	1.229723	-0.05935	2956161
	50	10	0.466148	-0.01647	1.208703	0.02102	2720260
	50	11	0.452377	0.013771	1.268207	-0.0595	2575561
	50	12	0.457245	-0.00487	1.275113	-0.00691	2425962
	50	13	0.43994	0.017305	1.336167	-0.06105	2314306
	50	14	0.447704	-0.00776	1.336373	-0.00021	2199447
	50	15	0.418987	0.028717	1.438779	-0.10241	2177604
m = 1.75	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	13	2	0.663184		0.509245		6388458
	30	3	0.559712	0.103472	0.765322	-0.25608	4511519
	22	4	0.450365	0.109347	1.025948	-0.26063	3623171
	31	5	0.395366	0.054999	1.19576	-0.16981	3034167
	50	6	0.352996	0.04237	1.348242	-0.15248	1.348242

	42	7	0.314485	0.038511	1.483717	-0.13548	2345719
	50	8	0.287648	0.026837	1.6038	-0.12008	2118952
	45	9	0.275172	0.012476	1.683806	-0.08001	1930030
	50	10	0.255815	0.019357	1.771234	-0.08743	1781391
	50	11	0.241233	0.014582	1.855795	-0.08456	1655605
	49	12	0.227735	0.013498	1.928215	-0.07242	1550763
	50	13	0.220171	0.007564	1.993864	-0.06565	1456333
	50	14	0.214212	0.005959	2.049377	-0.05551	1375587
	50	15	0.206233	0.007979	2.109316	-0.05994	1305517
m = 2	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	16	2	0.600307		0.584913		5500278
	3	3	0.333523	0.266784	1.098327	-0.51341	3790678
	31	4	0.369013	-0.03549	1.17077	-0.07244	2679962
	36	5	0.305701	0.063312	1.379672	-0.2089	2138902
	28	6	0.257402	0.048299	1.560021	-0.18035	1781965
	48	7	0.227478	0.029924	1.703878	-0.14386	1526362
	44	8	0.20408	0.023398	1.826863	-0.12299	1334054
	50	9	0.182424	0.021656	1.944833	-0.11797	1185891
	41	10	0.165288	0.017136	2.045842	-0.10101	1066788
	48	11	0.152128	0.01316	2.14046	-0.09462	969599.2
	50	12	0.141243	0.010885	2.220704	-0.08024	888487
	42	13	0.129	0.012243	2.307993	-0.08729	820645.7
	3	14	0.071555	0.057445	2.638172	-0.33018	812088.4
	50	15	0.115411	-0.04386	2.442067	0.196105	710556.5
m = 2.25	Iteration Number	Cluster number	F	ΔF $F(c)$-$F(c+1)$	H	ΔH $H(c)$-$H(c+1)$	J_m
	2	2	0.500049		0.693098		4781465
	3	3	0.333426	0.166623	1.098473	-0.40538	2880400
	3	4	0.250039	0.083387	1.386217	-0.28774	2010373
	27	5	0.251437	-0.0014	1.49384	-0.10762	1473683
	35	6	0.213986	0.037451	1.666139	-0.1723	1172269
	40	7	0.184774	0.029212	1.819706	-0.15357	966803.7
	48	8	0.162534	0.02224	1.949667	-0.12996	817926.8
	50	9	0.144206	0.018328	2.069295	-0.11963	706075.5
	2	10	0.100074	0.044132	2.302213	-0.23292	639470.6
	37	11	0.118533	-0.01846	2.26836	0.033853	549272.7
	38	12	0.10777	0.010763	2.357597	-0.08924	492781.8
	3	13	0.076987	0.030783	2.564531	-0.20693	460645.4
	3	14	0.0715	0.005487	2.63856	-0.07403	419882.4
	50	15	0.087413	-0.01591	2.577019	0.061541	372729.1