AN ABSTRACT OF THE THESIS OF

Yulin Sun for the degree of Master of Science in Soil Science presented on March 17, 2017.

Title: Artificial Means of Distinguishing Smooth-line Map Unit Boundaries: An Examination of Oregon Soil Survey Maps.

Abstract approved:

_____

Jay S. Noller

Soil is a complex living system with high heterogeneity, which makes locating soil map boundaries a challenge. In traditional soil survey, the placement of soil map boundaries relies largely (at least initially) on identifying the soil-forming factors of biota and topographic relief through stereo aerial photo pairs. Future soil survey is being automated with predictive digital soil mapping techniques. The blocky, raster edition of a soil map aesthetically contrasts with the smooth, visually more accurate lines of a traditional soil map. To set up rules for artificial intelligence to convert grids into smooth lines, it is essential to understand the association between the shapes of traditionally delineated soil map units and the properties used to define the map units.

In this study, taxonomically different map units with varied soil temperature regimes, soil moisture regimes, and parent materials were selected from Oregon for shape comparison. An array of shape descriptors, including Miller's circularity ratio, Schumm's elongation ratio, area deficiency ratio, full Procrustes distance, and elliptical irregularity index, were computed from soil geospatial (GIS) data to get quantitative 2-D shape descriptions for the delineations of each studied soil map unit. Kernel density estimation and two-sample Kolmogorov-Smirnov test were conducted

to evaluate the differences in shapes. Scatterplot matrix was used to display the separation of map units by linear combination of selected shape descriptors.

The purposes of this study were to i) analyze delineated 2-D shapes between taxonomically different soil map units using non-parametric statistical methods; ii) Understand if shape differences can be related to difference in soil climate regime or parent material; and iii) Find out if the linear combination of selected 2-D shape descriptors can be used to separate any pair of soil map units.

The results of this study indicate that i) delineated soil shapes appear random, but the frequency of shape archetypes is unique for each soil map unit; ii) Soil temperature and moisture regimes appear to influence such frequency, whereas parent material does not; and iii) Scatterplot matrices display separations between map units but not for all cases.

Through this study, a database containing the frequency of shape archetypes for each soil map unit can be built. This database can provide i) certain basic guidelines for machine learning, so that computer software can learn how to smooth raster grids; ii) Quality assurance and quality control (QA/QC) for traditional soil survey, so that the mapping and drawing techniques of a professional can be analyzed and learned by a novice soil mapper.

This study assessed an old question and delivered an answer foreseen but unsupported by prior work. The results of this study are more robust than earlier work, however, because this study used a more reliable database, involving a great diversity of soil orders and state factors, and advanced computer techniques and statistical analyses.

Artificial means of distinguishing smooth-line map unit boundaries: An examination of Oregon soil survey maps

by
Yulin Sun

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Master of Science

Presented March 17, 2017
Commencement June 2017

Master of Science thesis of Yulin Sun presented on March 17, 2017

APPROVED:

Major Professor, representing Soil Science

Head of the Department of Crop and Soil Science

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Yulin Sun, Author

ACKNOWLEDGEMENTS

I want to thank everyone who helped me during my study in Oregon State University. First and foremost, I want to thank my major professor, Dr. Jay Noller, for guiding me through this challenging project, and for patiently reviewing the drafts. Special thanks to my committee members, Dr. Kate Lajtha, Dr. Markus Kleber, and Dr. Jonathan Maynard, for providing valuable suggestions on my study and career. Thank you also to Dr. Duo Jiang, for serving as my Graduate Council Representative.

I thank Chris Ringo, for offering his expertise in GIS data analysis. My thanks to soil graduate students, Aimee Clark, Vance Almquist, Kris Osterlo, Tom Wanzek, Stephany Chacon, Trang Nguyen, and Pedro Rodrigues de Moraes Martinez, for their help in writing and nice conversation about soils. I also would like to thank fellow students from Statistics Department, Andrew Bernath, Kai Li, Kyle Dettloff, and Lu Wang, for their assistance in statistical analysis.

Last but not the least, I want to thank my family and my husband's family, for always being supportive. I especially want to thank my husband Mingyang Liu, who brought me to the adventure in soil science. Finally, I want to thank my daughter Lisa Liu, and my son George Liu, for their love along the way.

CONTRIBUTION OF AUTHORS

Dr. Jay S. Noller proposed and advised all the aspects of this study. Dr. Jay S. Noller and Aimee Clark were involved with the improvement of writing.

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

Page

LIST OF FIGURES

LIST OF FIGURES (Continued)

Figure Page

LIST OF FIGURES (Continued)

Figure Page

LIST OF TABLES

1. Introduction

To achieve suitable and sustainable land use and land management practices, it is essential that the information on the soil map is as accurate as possible at the scale of mapping. However, soil is a complex living system with high heterogeneity, and it is unrealistic to show every single detail of soil properties on map. Therefore, the purposes of mapping are to provide a general guideline about how detailed or how simplified the information shall be shown at the mapping scale, and which information has priority to be illustrated (Figure 1 and 2). Even with the guideline in mind, it is still a challenge for mappers to delineate areas without argument.

In traditional soil survey, aerial photographs give mappers the very first clue to where soil boundaries would occur (Figure 3). "Aerial photographs provide important clues about kinds of soil from the shape and color of the surface and the vegetation" (Soil Survey Staff, 1993). Places where vegetation types and/or topographic properties change abruptly would be considered to be places where soil properties are changing abruptly as well. Next, mappers go to the field, take soil samples around zones of change, use geomorphic characteristics to identify likely soil map units, and correlate soil to a known series. After that, mappers extrapolate the extent of delineation of each soil map unit based on their knowledge so that certainty and conformity are balanced (Schaetzl and Anderson, 2005).

The landscape cone (Figure 4) is a soil surveyors' mental model for soil climate regimes in a mountainous or hilly area. As elevation increases, temperature decreases but

precipitation rate increases. Therefore, soil climate regime at the summit of a mountain or the top of a hill is cool and wet, while that at the foot of a mountain/hill is relatively warm and dry. The boundary between two soil climate regimes can be estimated by elevation. However, the boundary declines towards north rather than being parallel to the sea level, because the rate of temperature decrease is greatest for slopes with poleward aspect.

Besides this soil cone model, each map unit also has its primary (objective) and non-primary (subjective) soil delineations. The delineations defined by geographical features such as rivers or hilltops are considered as primary delineations and usually delineated first by soil mappers, because the changes among soil delineations are abrupt and can be recognized easily. In contrast, the soil delineations along a hillslope with only gradual changes in topographic and other state factors and properties are defined more subjectively with soil mappers' best judgement (NRCS, 2003).

A soil map unit is a collection of delineated areas with similar composition of soil components (Figure 5). Because it is unrealistic to draw boundaries for every single pure soil component, one delineated area contains more than one soil component, and one soil component can be associated with other similar and/or dissimilar soil components within such an area. Consociation, association, complex, and undifferentiated group are the four types of map units used in current soil survey. The first one is relevant to this study, because the variation among consociations with the same map unit name is considered to be relatively low. In a consociation, more than half of the delineated area is occupied by one soil component and may or may not count its similar soil components; whereas in an

association, a complex, or an undifferentiated group, the area consists of two or more major soil components that are sufficiently different in behavior or morphology. The name of a consociation map unit usually contains the name of its dominant soil series, texture of the surface layer, and slope range (Soil Survey Staff, 1993).

The spatial distribution of a consociation map unit can be viewed as the spatial distribution of soils with similar compositions. In addition, Jenny's (1941) state equation, which is $S = f(cl, o, r, p, t)$, indicates that soil (S) is a function (f) of the five state factors: climate (cl), organic activity (o), relief (r), parent material (p), and time (t). Therefore, the spatial distribution of a map unit can also be viewed as the spatial distribution of area with similar state factors. In fact, digital soil mapping has been using the state factors along with other factors to predict the occurrence of map units (Minasny and McBratney, 2016), and traditional soil survey has been using vegetation, topographic properties, and geological information to locate the soil map boundaries (Soil Survey Staff, 1993).

To maintain consistent and accurate information on a soil map, it is critical to understand the correlation between the shape of map units and the factors that define the map units. Bilton (1982) investigated the correlation between vegetation delineation and soil delineation, and found that the correlation is weak. The overarching goal of this study was to investigate the correlation between delineated soil shapes and two other state factors, which are climate regimes and parent material in particular. The specific questions this study asked were: 1) Do taxonomically different map units have quantitatively different shapes? 2) Are the differences in delineated shapes considered to be statistically significant between paired map units? 3) Do the differences in shapes

correlate with state factors? 4) Can linear combination of selected shape descriptors differentiate paired soil map units? The results of this study would reveal the impact of climate regimes and parent material on the occurrence of soil map boundaries, and could be used to smoothing soil map boundaries in digital soil mapping.

2. Literature Review

Studies on the shape of soil map units date back at least to the 1950s, and Francis Hole is considered a pioneer of this type of study. Hole (1953) proposed that the pattern of soil delineation could be grouped into six classes varying from very simple to very complex, based on the ratio of the perimeter of the soil delineation over the perimeter of the circle with the same area (Figure 6). Hole (1978) then adopted Fridland's (1976) idea and divided soil shapes into three categories, disk, spot, and stripe, according to the ratio between the longest axis and the shortest axis (Figure 7 and 8). In Hole's later book (Hole and Campbell, 1985), he combined the disk shape into the spots category.

Later on in the 1980s, two graduate students at Oregon State University worked on shape analysis of soil map units (Bilton, 1982; Poore, 1986). Bilton (1982) studied the correlation between soil and vegetation delineations in the Oregon Coastal Range using a non-parametric statistical method. The results indicated that the shape of delineated soil was barely correlated with the shape of delineated vegetation using the dataset, and vegetation delineations appeared to be more rounded compared to soil delineations. Therefore, Bilton suggested further investigation was needed to determine if the non-statistically significant difference was merely coincidence. However, the study also found positive correlation between soil and vegetation delineations in uplands when using Schumm's elongation ratio to describe their shapes. Bilton (1982) posits that the similarity of soil and vegetation delineations in uplands could result from the similar response of soil and vegetation to topographic properties.

Using 452 delineations from five map units in Benton County of Oregon, Poore (1986) evaluated 43 shape indexes including primary measurements of soil delineation and its convex hull (Figure 9), as well as secondary measurements calculated from the primary measurements. Factor analysis showed that area of delineation, complexity of boundary, and Schumm's elongation ratio of convex hull were the three best quantitative descriptors for the spatial properties of sampled soil delineations. In addition, Poore used Mann-Whitney U test, a non-parametric statistical method, to compare shapes among the five map units, and the results showed significant differences between some but not all samples. Poore then suggested that classification of soils, erodibility of soils, intensity of soil survey, slopes, and landforms could lead to shape differences and similarities.

At the time of these theses, the application of computer technique was not common, and only hand-drawn delineations were available back then. With improvements in soil survey geodatabases, it is worthwhile to verify the results of former studies using a new dataset. Therefore, the purpose of the current study was to re-evaluate old questions by using the dataset from Soil Survey Geographic Database (SSURGO). With the application of advanced geospatial processing program (ArcMap), the current study was expected to provide robust outcomes.

3. Materials and Methods

This study is divided into two parts based on study areas and the criteria for grouping soil map units. In Part I, all samples were taken from Willamette Basin of Oregon and two (soil) map units were grouped solely based on the difference in family level of the taxonomic classifications of their dominant soil series (Figure 10). In Part II, the sampling area expanded to the entire state of Oregon and two map units were grouped based on the difference in soil climate regimes or types and origins of parent material (Figure 11). Yet the taxonomic classifications of the dominant soil series for paired map units were different at family level as well. In this section, the author first described the sampling process, then discussed the selected shape descriptors, and last talked about the statistical methods.

3.1. Sampling Process

For both parts of this study, the Soil Survey Geographic database (SSURGO) was used to delineate the survey area, and ArcMap (version 10.1) was used to display soil delineations. To minimize variations within a group and maximize the possibility that delineations represent the spatial distribution of a natural soil body, some delineations were excluded from this study. The exclusions included the delineations with holes inside, those not belonging to the map unit type of consociations, or those which were interrupted by county or landownership boundaries.

Besides the difference in sampling area as mentioned above, the slope range was different from Part I to Part II such that no restriction on slope range in Part I but slope

was restricted to less than 10% in Part II. Moreover, the sampling size of each map unit varied from Part I to Part II. R programming software (r-project.org) was used to randomly select 50 delineations for each map unit in Part I, and all the analysis were performed on these delineations. Delineations were selected from many soil map areas. This means that different mappers are included representing three or more decades of evolution in soil mapping practices.

In contrast, all delineations in each map unit of Part II were used for the analysis of shape difference, but only 30 delineations were selected to display the cluster pattern for each map unit. Delineations from each soil map unit were selected from a single soil survey area. This means that two to three authors performed linework within two decades.

In part I, five groups were developed based on the differentiation of taxonomic classes between the dominant soil series of each paired map units. The ones in Group A were considered to be similar, in Group B were very different, in Group C differed in organic matter content, in Group D differed in soil temperature regime, and in Group E differed in soil moisture regime (Table 1).

In part II, eight groups were generated based on the differentiation of climate regimes or parent material. The ones in Group 1, 2, and 3 had similar landform and parent material but different climate regime within each group, while those in group 4 through 8 had the same climate regime and landform but different parent material within each group (Table 2 and 3).

Also in Part II, one map unit from four different counties was selected to evaluate the shape differences among survey areas. The selected map unit is "Abiqua silty clay

loam, 0 to 3 percent slopes". The four counties selected for comparison are OR637 (Lane county), OR639 (Linn county), OR643 (Marion county), and OR053 (Polk county). Only two sample Kolmogorov-Smirnov test was conducted for this set of samples.

3.2. Point-based Shape Analysis

Shape analysis methods usually are grouped into point-based, contour-based, or region-based according to the information used (Zhang and Lu, 2004). The point-based methods involve the use of a landmark and a pseudo-landmark. A landmark is a point that carries fundamental morphological properties (Bookstein, 1986; Small, 1988) or certain mathematical meanings (Bookstein, 1991) to make shapes comparable, while a pseudo-landmark is more related to the construction of shape, such as the points regularly sampled along a contour (Dryden and Mardia, 1998; Claude, 2008). For soil delineations, no biological homological method defines their landmarks. However, soil delineation is a closed curve, and can be depicted by a decent number of points.

Full Procrustes distance is “the square root of the sum of squared distance between homologous coordinates of superimposed configuration” (Claude, 2008). In Part I, the full Procrustes distance between soil delineation and the mean shape of a map unit was used as a 2-D shape descriptor for each delineation.

To get this shape descriptor, soil delineations were exported one by one into Tagged Image File Format (Tiff) using ArcMap. The coding methods developed by Claude (2008) were used to read Tiff files, locate the starting point from digitalized images, and generate coordinates of pseudo-landmarks along the outline of soil

delineation. In this study, one of the two tangencies of long axis was used as the starting point, which was also considered as a landmark. The pseudo-landmarks were used as an approximation of the outline of delineated area, and better estimation of soil delineation would be achieved by using more pseudo-landmarks. However, the more pseudo-landmarks used, the slower the computing process would be for the latter steps. In this study, one hundred pseudo-landmarks were determined to be appropriate.

Then, the centroids of each delineation were computed based on the coordinates of pseudo-landmarks, the soil delineations were superimposed by their centroids, and a mean shape was generated for each map unit. After adjustment of size and rotation among soil delineations, the distances of similar pseudo-landmarks between delineations and their mean shape were squared and summed, which yield the value of full Procrustes distances for each delineation.

In part II, the full Procrustes distance was replaced by the elliptical irregularity index, because the Procrustes method required multiple steps to obtain values and was time consuming when dealing with a large number of samples.

3.3. Quantitative Shape Descriptors for Soil Map Delineations

Four quantitative 2-D shape descriptors used in this study were adopted from the Poore's study (Poore, 1986), namely Miller's circularity ratio, Schumm's elongation ratio, elliptical irregularity index, and area deficiency ratio. The first three shape descriptors can be viewed as comparisons between soil delineations to a standard archetype, such as

a circle or an ellipse; while area deficiency ratio compares a soil delineation to its convex hull.

The definition and equation of the shape descriptors are listed below. ArcMap (version 10.1) was used to build convex hulls for soil delineations, and parameters in the equations were directly taken from the attribute tables using ArcMap. Microsoft Excel 2010 was used to calculate the values for each shape descriptor.

Miller's circularity ratio compares a delineation to a circle, and the circle has the same perimeter as the delineation. The ratio equals to the area of delineation over the area of that circle (Miller, 1953), and the value ranges from 0 to 1. A value close to 1 implies that the soil delineation is more disk-like, while a value close to 0 indicates that the soil delineation is very complex. In this study, Miller's circularity ratio was calculated for each soil delineation itself.

Miller's Circularity Ratio = 4 * pi * Area / (Perimeter^2)

Schumm's elongation ratio also compares a delineation to a circle, but the circle has the same area as the delineation. The ratio equals to the diameter of that circle over the long-axis of the delineation (Schumm, 1956), and the value ranges from 0 to 1. A value close to 1 also implies that the soil delineation is more disk-like, while a value close to 0 implies that the soil delineation is more stripe-like. Poore (1986) evaluated Schumm's elongation ratio for soil delineation itself and that for its convex hull, and the ratio built on the convex hull was considered to be a better shape descriptor. In this study, Schumm's elongation ratio was calculated for the convex hull of each soil delineation, instead of soil delineation itself.

Schumm's Elongation Ratio = [(Area / pi) ^ 0.5] * 2 / long-axis

An area deficiency ratio compares a delineation to its convex hull (Poore, 1986). The ratio equals to perimeter of soil delineation (P1) over perimeter of the convex hull (P2). Values vary from 0 to ∞. A large value implies that the soil delineation is very complex, and a value close to 0 means the soil delineation is very simple.

Area Deficiency Ratio = (P1 – P2) / P2*100

Miller's circularity ratio, Schumm's elongation ratio, and the area deficiency ratio were used in both Part I and Part II. However, the full Procrustes distance in Part I was replaced by elliptical irregularity index in Part II.

Poore (1986) developed the elliptical irregularity index, which compares the soil delineation with an ellipse. The ellipse has the same area and long axis as the delineation, and the ratio equals to perimeter of this ellipse over perimeter of soil delineation. The perimeter of ellipse is calculated from its area and long axis using the equation developed by Weast et al. (1964). The index value ranges from 0 to 1. A delineation has a value close to 1 is more ellipse-like, and a value close to 0 is very complex.

Elliptical Irregularity Index = 2*pi*[(((long axis/2)^2 + (2*area/pi/long axis)^2)/2)^0.5]/perimeter

3.4. Statistical Analysis

Non-parametric statistics is a way to provide statistical reference when the data does not meet the assumption of parametric statistics (Daniel, 1990). Kernel density estimation is a very commonly used type of non-parametric density estimation (Thas,

2010). Unlike a histogram, which was used in Poore's study (Poore, 1986), a smooth weight function rather than rectangular weight function is used in kernel density estimation. Therefore, the shape of a kernel density estimation curve is not dependent on bin width, and the stepwise feature of a histogram can be avoided (Efromovich, 1999).

Two sample Kolmogorov-Smirnov test, which was also used in Bilton's study (Bilton, 1982), is another non-parametric statistical method frequently used to compare the empirical distribution functions of two distributions (Gibbons, 1985; Thas, 2010). In both Part I and Part II, a kernel density estimation curve was built to graphically compare the distributions of shape descriptors for the paired soil map units. A two sample Kolmogorov-Smirnov test was conducted to evaluate the difference between the distributions of each shape descriptor for the paired map units, and a small p value indicated that two distributions were from different populations. In addition, a scatterplot matrix was produced for each group to find the linear combination of measurements to best separate the paired map units.

4. Results

In this section, main findings from kernel density estimations, the two-sample Kolmogorov-Smirnov tests, and scatterplot matrices were described group by group. No comparison was made between groups, although some contradictory results appeared between Part I and Part II, explanation was made in the discussion section rather than here. The major task of this section was to provide a clear statement on what the results were, and whether the results supported or did not support each other within each group.

4.1. Results of Part I

In part I, nine soil map units in the Willamette Valley of Oregon were selected based on their taxonomic classifications, and five pairs were generated from these soil map units based on their similarities (Group A), dissimilarities (Group B), organic material content (Group C), soil temperature regime (Group D), or soil moisture regime (Group E) in the dominant soil series. Fifty polygons from each soil map unit were randomly selected for comparison. The delineated shapes of each polygon were described by four 2-D shape descriptors, and the distributions of those shape descriptors were compared for the paired soil map units within each group. Non-parametric statistical methods, e.g. kernel density estimation and the two-sample Kolmogorov-Smirnov test, were used to illustrate the significant level of the differences. The results showed that the differences between paired soil map units could be statistically significant, or not. As expected, the paired soil map units that were considered to be similar (Group A) based on the taxonomic classifications of the dominant soil series turned out to have similar

distributions for the selected shape descriptors. Surprisingly, the pair with different soil moisture regimes seemed to have similar distributions of shape descriptors as well. Scatterplot matrices were used to demonstrate the separation of the paired soil map units, and the results indicated that the method worked for some cases, but not all cases.

4.1.1. Group A Bashaw and Chehalem

The two soil map units in Group A are considered to be similar based on the taxonomy classifications of the dominant soil series. Bashaw is classified as very-fine, smectitic, mesic Xeric Endoaquerts, and Chehalem as fine, smectitic, mesic Cumulic Vertic Endoaquolls (Table 1). Therefore, both soil series have smectite as dominant clay minerals, are located in mesic soil temperature regime, and experience aquic soil moisture conditions. Although the distributions of selected shape descriptors displayed by kernel density estimation curves were not identical for Bashaw and Chehalem (Figure 12), the differences were not considered to be statistically significant according to the results of two–sample Kolmogorov-Smirnov test except for Miller's circularity ratio (Table 4). In addition, the scatterplot matrix indicated that combination of any two shape descriptors were not able to show clear separation between Bashaw and Chehalem, even the ones with Miller's circularity ratio (Figure 13).

4.1.2. Group B Salkum and Sifton

The two soil map units in Group B are considered to be very different based on the taxonomy classifications of the dominant soil series. Salkum is classified as fine,

kaolinitic, mesic Xeric Palehumults, and Sifton as medial over sandy or sandy-skeletal, mixed, mesic Typic Melanoxerands (Table 1). Therefore, Salkum series is a highly weathered soil with high clay content in the fine-earth fraction, and with kaolinite as dominant clay minerals. In contrast, Sifton series is a weakly weathered soil with more coarse fragments, and with a mixture of clay minerals. Yet, both are rich in organic matter, and locate in mesic soil temperature regime and xeric soil moisture regime. The distributions of selected shape descriptors displyed by kernel density estimation curves were quite different for Salkum and Sifton (Figure 14), and the differences were proved to be statistically significant according to the results of two-sample Kolmogorov-Smirnov test except for Schumm's elongation ratio (Table 4). However, scatterplot matrix barely showed clear separation between Salkum and Sifton except for the combination of the area deficiency ratio and full Procrustes distances (Figure 15).

4.1.3. Group C Cazadero and Willamette

The two soil map units in Group C are considered to have different organic matter content based on the taxonomy classifications of the dominant soil series. Cazadero is classified as fine, mixed, active, mesic Rhodic Paleudults, and Willamette as fine-silty, mixed, superactive, mesic Pachic Ultic Argixerolls (Table 1). Therefore, Cazedero is considered to have less organic matter than Willamette. The distributions of Miller's circularity ratio as well as Schumm's elongation ratio displyed by kernel density estimation curves were very different for Cazadero and Willamette (Figure 16), and the differences were considered to be statistically significant according to the results of two–

sample Kolmogorov-Smirnov test (Table 4). However, the distributions of the area deficiency ratio as well as full Procrustes distance for Cazadero and Willamette had lots of overlaps (Figure 16), and the differences were not considered to be statistically significant according to the results of two-sample Kolmogorov-Smirnov test (Table 4). However, the scatterplot matrix showed that the combination of area deficiency ratio with either Schumm's elongation ratio or full Procrustes distances can separate Cazadero from Willamette (Figure 17).

4.1.4. Group D Keel and Sifton

The two soil map units in Group D are considered to have different soil temperature regime based on the taxonomy classifications of the dominant soil series. Both soil series are in the soil order of Andisols, but Keel is classified as Cryands whereas Sifton as Xerands (Table 1). The differences in the distributions of selected shape descriptors were not only displayed by kernel density estimation curves (Figure 18), but also indicated by the small p values of two–sample Kolmogorov-Smirnov test (Table 4). Moreover, the scatterplot matrix clearly showed separation of Keel from Sifton, and the separation was more obviously achieved by the combinations that used area deficiency ratio as one variable (Figure 19).

4.1.5. Group E Powell and Delena

The two soil map units in Group E are considered to have different soil moisture regime based on the taxonomy classifications of the dominant soil series (Table 1). Both

soil series have the same particle size class, minerology class, cation exchange activity class, and soil temperature regime. But Delena is classified as Humic Fragiaquepts, and Powell as Humic Fragixerepts. The distributions of selected shape descriptors were not identical but had lots of overlaps (Figure 20), and the differences were not statistically significant except for the distribution of full Procrustes distance (Table 4). In addition, the scatterplot matrix showed that none of the combinations of selected shape descriptors can separate Delena from Powell (Figure 21).

4.2. Results of Part II

In part II, the study area was expanded to the entire state of Oregon, and twelve soil map units with a slope of less than 10% were selected. Eight pairs were generated from these soil map units based on their differences in soil temperature and moisture regimes (Groups 1, 2, and 3) or parent material (Groups 4, 5, 6, 7, and 8). Three of the four shape descriptors from part I were continually used in part II, including Miller's circularity ratio, Schumm's elongation ratio, and the area deficiency ratio. The full Procrustes distance was replaced by the elliptical irregularity index, because the former shape descriptor requires several more steps for calculation, and is not considered a time efficient method if working with a large quantity of samples. The number of polygons in each soil map unit was not fixed at fifty for kernel density estimation and the two-sample Kolmogorov-Smirnov test in part II, and varied from around thirty to over two hundred. The unbalanced sample number between the paired soil map units was not a concern since non-parametric statistic methods were used in this part as well. The results showed

that soil moisture regime had an impact on the delineated shapes whereas soil temperature regime and parent material did not. This is contradictory with the results in Part I, in which soil temperature regime was found to have impact on delineated soil shapes whereas soil moisture regime did not. Therefore, the author proposed two possible explanations. 1) Although the map units were grouped based on the differences in soil temperature/moisture regimes, the differences in delineated soil shapes probably were not caused by the differences in soil temperature/moisture regimes. 2) The differences in soil temperature/moisture regimes were not large enough to lead to the differences in delineated soil shapes. The results of scatterplot matrices, with thirty delineations in each map unit, barely showed separation of the paired soil map units in Part II.

4.2.1. Group 1 through 3

According to the attribute table in SSURGO database, the soil map units within Group 1, 2, and 3 are considered to have different vegetation-climate regime (Table 2). The dominant soil series in the two soil map units of Group 1 are Lookingglass and Cowsly, respectively. The two soil series have similarities such as drainage class, land use, and morphology. However, Lookingglass is found at a lower elevation with warmer soil temperature regime compared with Cowsly. The kernel density estimation for the distributions of selected shape descriptors showed lots of overlaps (Figure 22), and the p values of two-sample Kolmogorov-Smirnov test indicated that the difference between the distributions was not statistically significant (Table 5). In addition, the scatterplot matrix

showed that the combination of shape descriptors can hardly separate Cowsly from Lookingglass (Figure 23).

The dominant soil series in the two soil map units of Group 2 are Clackamas and Kimberly, respectively. The two soil series have the same landform, parent material, and soil temperature regime, but Clackamas is poorly drained whereas Kimberly is well drained. The kernel density estimation curves indicated that the distriutions of Miller's circularity ratio and Schumm's elongation ratio for these two soil series had different peak values. Besides that, the distributions of the area deficiency ratio and the elliptical irregularity index had different peak heights (Figure 24). In addition, the results of the two-sample Kolmogorov-Smirnov test indicated that the differences between corresponding distributions were considered to be statistically significant (Table 5). However, the scatterplot matrix still showed that the combination of shape descriptors can hardly separate Clackamas from Kimberly (Figure 25).

The dominant soil series in the two soil map units of Group 3 are Bakeoven and Bocker. These two soil series have the same landform and parent material, but different soil temperature and moisture regimes. The distributions of selected shape descriptors displayed by kernel density estimation curves were quite different (Figure 26), and the differences were considered to be statistically significant (Table 5). Besides that, the scatterplot matrix showed separation of Bakeoven from Bocker (Figure 27).

4.2.2. Group 4 through 8

According to the attribute table in SSURGO database, the soil map units within Group 4 through 8 are considered to have different relief-parent material (Table 2). All of the dominant soil series in each soil map units have mesic soil temperature regime and xeric soil moisture regime, yet the parent material varies from sandstone and shale to tuff and basalt to sedimentary rock. However, the differences in the distributions of selected shape descriptors were not considered to be statistically significant (Table 5), even the kernel density estimation curves for each shape descriptor were unique for each soil series (Figure 28-32). In addition, the scatterplot matrices showed that the paired soil series were barely separated from each other (Figure 33-37).

4.2.3. Comparison of the Same Map Unit from Different Counties

The results of this section showed that different mappers tend to draw lines differently, as mentioned in Figure 2. Map unit of "Abiqua silty clay loam, 0 to 3 percent slopes" was selected from 4 counties, the frequency of shape archetypes was computed for each county (but results are not shown), and two-sample Kolmogorov-Smirnov tests were conducted to evaluate the shape differences among counties (Table 6).

According to Table 6, a certain level of difference was evident in line-drawings among different survey areas. The difference between county 639 (Linn county) and county 053 (Polk county) is the most obvious because all four shape descriptors showed differeces in the frequency of shape archetypes. In constrast, the difference between

county 053 (Polk county) and county 643 (Marion county) is the least obvious as only Miller's circularity ratio showed very significant statistical difference.

The differences in line-drawing among mappers, mentioned in Figure 2, is verified through this experiment. This type of difference in delineation shape of a single map unit can contribute to the positive results (significant difference in the frequency of shape archetypes) observed in Part II, in which most of the paired map units are selected from different survey areas, such as Group 2 and Group 3 (Table 3 and 5). One the other hand, this type of difference may contribute to the negative results (insignificant difference in the frequency of shape archetypes) of Part I as well, in which some of the map units contain soil delineations from varied survey areas, and the overlaps between paired map units might be larger than one expected. The bottom line is that different authors do lead to dispersion in study results.

5. Discussion

Four questions were raised in the introduction and they are: 1) Do taxonomically different map units have quantitatively different shapes? 2) Are the differences in delineated shapes considered to be statistically significant between paired map units? 3) Do the differences in shapes correlate with state factors? 4) Can linear combination of selected shape descriptors differentiate paired soil map units? The structure of following sections loosely builds upon these questions with similar order.

5.1. Shape Descriptors

Miller's circularity ratio compares the area of a soil delineation with the area of a circle which has the same perimeter of the delineation, and a ratio close to 1 indicates the delineation is more disk-like. Schumm's elongation ratio compares the long axis of a soil delineation with the diameter of a circle which has the same area of the convex hull of the soil delineation, and a ratio close to 0 indicates the delineation is more stripe-like. Area deficiency ratio compares the perimeter of a delineation to the perimeter of its convex hull, and a large value implies that the soil delineation is very complex. Elliptical irregularity index compares the perimeter of a soil delineation with the perimeter of an ellipse that has the same area and long axis of the delineation, and a value close to 1 indicates the delineation is more ellipse like. Refer to section 3.3 for the detailed definition and calculation of the above four quantitative shape descriptors.

Table 7 graphically displayed eight soil delineations from the same map unit of Jory series and listed the value of selected quantitative shape descriptors of these

delineations. The two smallest values of Miller's circularity ratio (C = 0.123 and C = 0.072) come from the two most complex soil delineations, and the two largest values (C = 0.785 and C = 0.907) come from the two most disk-like soil delineations. Therefore, Miller's circularity ratio is proved to be useful to distinguish disk-like soil delineations from complex soil delineations. In addition, elliptical irregularity index can also distinguish complex soil delineations from simple soil delineations, with smaller values correlating to more complex soil delineation, and values close to 1 indicating ellipse-like shapes. Besides these two, area deficiency ratio gave distinct quantitative descriptions for simple and complex soil delineations as well. However, Schumm's elongation ratio of the convex hull did not provide such distinct differentiation between simple soil delineations and complex soil delineations, probably because this ratio is based on the area of convex hull instead of the perimeter of soil delineation, and both complex and simple soil delineations can have similar convex hulls. From the data set in Table 7, as well as other data set (not shown), the soil delineations within one map unit appeared to have random shapes varying from disk-like to stripe-like, and from simple to complex.

5.2. Non-parametric Statistical Analysis

Poore (1986) used histogram to display the distribution of shape descriptors, but the selection of bin width influences the shape of histogram, and this limitation makes representation of continuous variable become discontinued. However, kernel density estimation curve conquers such a drawback, and draws a smooth line to show the frequency of variables.

Although the shape of delineation in each soil map unit appeared to be random (from the discussion in section 5.1), the frequency of each shape archetype is recognizable for each soil map unit. For instance, the distributions of Miller's circularity ratio, Schumm's elongation ratio, area deficiency ratio, and full Procrustes distance for Keel are different from that of Sifton (Figure 18). Compared to Keel, Sifton has a larger Miller's circularity ratio for the peak value of the kernel density curve, which indicates it has more disk-like delineations than Keel has. Moreover, Sifton has a higher spike for the peak value and a narrower width for area deficiency ratio, which indicates Sifton has more simple delineations than Keel does.

The difference between Keel and Sifton is also showed in the kernel density curves of full Procrustes distance. Full Procrustes distance used in this study is a minimized summed squared distance of similar pseudo-landmarks between a soil delineation and the mean shape of a soil map unit that the soil delineation belongs to, and a small value indicates the soil delineation is fairly similar to the mean shape. Therefore, the delineations in the map unit of Keel have less variation between each other than the delineations in the map unit of Sifton do (Figure 18).

Although the shape of kernel density curve is unique for each soil map unit, some curves have more overlaps within the paired map units than others do. The more overlaps the curves share, the less likely the shape difference between the paired map units is considered to be statistically significant. As an example, the distributions of Miller's circularity ratio, area deficiency ratio, and full Procrustes distance between Salkum and Sifton have less overlaps, and the two-sample Kolmogorov-Smirnov tests give small p

values when evaluating significant levels of differences between the paired distributions. In contrast, the distributions of Schumm's elongation ratio have more overlaps than the distributions of other three shape descriptors do, and the two-sample Kolmogorov-Smirnov tests give a p value of 0.549 (Figure 14 and Table 4).

5.3. State Factors

A consociation consists of one dominant soil component and one or more similar soil component(s) within one delineated area. The name of a consociation is the same as the dominant soil component, and usually is at series level in Soil Survey Geographic Database. Therefore, a consociation map unit named “Salkum silt clay loam, 8-16 percent slopes” has Salkum soil series as its dominant soil component in each delineated area.

On the other hand, soil is the result of continuous work of climate and vegetation on parent material at certain relief over a time, which is summarized in this equation: $S = f(cl, o, r, p, t)$. Therefore, spatial distribution of soil delineations from the same map unit can be viewed as spatial distribution of similar state factors. Bilton (1982) pointed out that the delineated shape of vegetation is barely correlated to the delineated shape of soil, whereas this study looked at correlation between delineated shapes of soils to soil climate regimes and parent material. The results indicated that the correlation between delineated shape of soil and parent material is weak, and the correlation between soil shapes and soil climate regimes appeared to be contradictory when comparing the results of Part I to the results of Part II.

The dominant soil series in the soil map units of Group D in Part I and Group 1 in Part II, for example, are not of the same soil temperature regime as their pair. The differences in the distributions of selected shape descriptors for Group D were proved to be statistically different, whereas that for Group 1 were not.

Another example of such contradictory findings in terms of the correlation between delineated soil shape to soil climate regime is Group E in Part I and Group 2 in Part II. The dominant soil series in each soil map unit within each paired group are considered to be of different in soil moisture regimes, but the results of two-sample Kolmogorov-Smirnov test did not support each other, either. The differences in the distributions of Miller's circularity ratio, Schumm's elongation ratio, and area deficiency ratio for Group E were not proved to be statistically different, whereas those for Group 2 were proved to be statistically different.

Inconsistent results may be caused by the mismatching of factors arising from focus on soil climate regime to the exclusion of differences or similarities in the other state factors. Such mismatches would contribute to the inconsistent results. The differences are arising from mismatch of classes that have discreet ranges for continuous variables such as temperature and slope. For example, compare soil map units Keel and Sifton in Group D of Part I, which have cryic and mesic soil temperature regimes, respectively, with soil map units Lookingglass and Cowsly in Group 1 of Part II, which have mesic and frigid soil temperature regimes, respectively. The minimum temperature difference between the upper bound for cryic and lower bound for mesic is greater than that between mesic and frigid, which is nil. We should expect similar non-linear

responses for the pseudo-continuous soil moisture regimes. Powell and Delena in Group E of Part I have xeric and aquic soil moisture regimes respectively, while Clackamas and Kimberly in Group 2 of Part II have aquic and aridic soil moisture regimes respectively. The moisture (both absolute and seasonal) difference between xeric and aquic is smaller than that between aquic and aridic.

However, no quantitative statement for the significant level of difference in soil temperature and moisture regime was established, and the above explanation was draw from minor examples. To confirm these possible explanations, a quantitative evaluation of the difference in soil temperature and moisture regime should be established, and more data need to be evaluated.

5.4. Scatterplot Matrix

Scatterplot matrix is used to display the separation of paired map units by the combination of selected shape descriptors, and it worked for some cases but not all. The separations are obvious in groups B, C, and D of Part I, especially the combination of area deficiency ratio and full Procrustes distance. This makes sense because the two-sample Kolmogorov-Smirnov tests also indicate significant differences between paired soil map units. However, the results of the two-sample Kolmogorov-Smirnov test and scattterplot matrices do not support each other for the data of groups 2 and 3 in Part II.

Poore (1986) pointed out that the shape descriptors have strong correlation, because the calculations of Miller's circularity ratio, Schumm's elongation ratio, elliptical irregularity index, and area deficiency ratio are based on the same primary measurements,

such as perimeter and area, etc. Therefore, the ability to separate different map units may be weakened by the autocorrelation among shape descriptors.

5.5. Implications

Through this type of study a database containing the frequency of shape archetypes for each soil map unit can be built. Through such a database, the frequency of shape archetypes for unknown-soil map units can be computed and then matched to known-soil map units. This is a many-to-many relationship. Conversely, the shape of one single unknown-soil delineation can be computed and then matched to its known-soil map unit. Therefore, the family level of taxonomic classification of an unknown-soil map unit or an unknown-soil delineation (individual) could be revealed.

Moreover, the linework of a novice soil mapper can be distinguished from that of an expert. As a learning tool through this type of study, the drawing skills of an expert can be analyzed and copied by a novice soil mapper. Therefore, the analysis of shape archetypes distribution can provide quality assurance and quality control for traditional soil survey.

Along with the development of computer science, digital soil mapping becomes more and more popular nowadays. Instead of drawing smooth vector delineations, at present a computer can only generate raster grids to represent soil boundaries. Using the approach advanced in this work, it may be possible to teach artificial intelligence to convert raster grids to vector delineations and thus make resulting maps more aesthetic.

This study used a dataset built from traditional soil survey, which itself has developed and been modified over decades. For instance, some soil series switched from one family to another family, and/or the physical, chemical, and morphological classification of some soil series might have changed as well. Therefore, the complications presented to this study are attributed to not only the complexity of soils, but also the complexity of a human-drawn dataset. To confirm (or challenge) the results of this study, and/or to apply the outcomes of this study, additional investigation is warranted.

6. Conclusion

All soil map unit delineations observed in this study appear random, drawing on a number of shape archetypes, and the delineated soil shapes within one soil map unit vary from simple to complex (Table 7). However, statistical analyses of morphometric shape on digital soil map data reveal that each soil map unit has a recognizable frequency of each of the shape archetypes, as displayed by the kernel density estimation curves.

This study verified prior work, based on rudimentary shape analysis, concluding that the distributions of selected shape descriptors are different among some soil map units but not all. Using non-parametric statistic methods, this study further confirmed that the difference between the distributions of selected shape descriptors is not always considered to be statistically significant. The two-sample Kolmogorov-Smirnov test was found to provide more statistical power when used to evaluate the significance level of the shape difference. Based on the p-values of these tests not all of the differences between paired soil map units were considered to be statistically significant.

Shapes of soil delineations seem to correlate with soil climate regimes, but the correlation is inconsistent. What variables lead to these inconsistent results is not fully understood. Perhaps, the difference in soil temperature and moisture regimes between map units was not significant enough to lead to a significant difference in the distributions of selected shape descriptors. Perhaps, the difference in distribution of selected shape descriptors was related with variables other than soil temperature and moisture regimes.

Shapes of soil delineations seem not to correlate with parent material. Analysis results are unclear as slope may have confounded analysis, or parent material does not influence the delineation of the examined soil map units. It would be valuable to investigate and compare the delineated soil shapes using series that have slopes steeper than 10% in future.

Scatterplot matrix did not show distinct separation between paired soil map units, probably it is caused by the correlation of shape descriptors themselves. Other cluster analysis such as Principle Component Analysis can be tested in future.

This type of study could be applied to provide quality assurance and quality control for both traditional soil survey and predictive (digital) soil mapping. However, this is the first step of the journey of thousands. To provide a fully developed system that can work routinely, more investigation is needed.

Bibliography

Bilton, J.L., 1982. Comparison of soil and vegetation map delineation shapes and areal correspondence, Oregon. Master thesis of Geography in Oregon State University.

Bookstein, F.L., 1986. Size and shape spaces for landmark data in two dimensions. Statistical Science 1: 181-242.

Bookstein, F.L., 1991. Morphometric tools for landmark data: geometry and biology. Cambridge University Press, NY.

Buckley, A., and Field, K., 2011. Making a meaningful map: A checklist for compiling more effective maps. Esri. http://www.esri.com/news/arcuser/0911/making-a-map-meaningful.html

Claude, J., 2008. Morphometrics with R. Springer NY.

Daniel, W.M., 1990. Applied nonparametric statistics. PWS-KENT Pub., Boston

Dryden, I.L., and Mardia, K.V., 1998. Statistical shape analysis. John Wiley and Sons, Inc., UK.

Efromovich, S., 1999. Nonparametric curve estimation: methods, theory, and applications. Springer NY.

Fridland, V.M., 1976. Pattern of the soil cover. Dokuchaev Soil Instityte, Academy of Sciences of the USSR, Israel Program for Scientific Translations, Jerusalem. Translated from Russian.

Gibbons, J.D., 1985. Nonparametric statistical inference. Marcel Dekker. NY.Cline, M.G., 1963. Logic of the new system of soil classification. Soil Science 96: 17-22.

Hole, F.D., 1953. Suggested terminology for describing soils as three-dimensional bodies. Soil Science Society of America Journal. 17: 131-135.

Hole, F.D., 1978. Approach to landscape analysis with emphasis on soils. Geoderma 21: 1-23.

Hole, F.D., and Cambell, J.B., 1985. Soil landscape analysis. Rowman and Allanheld, Totowa, N.J.

Jenny, H., 1941. Factors of soil formation: a system of quantitative pedology. McGraw-Hill, NY.

Miller, V.C., 1953. A quantitative geomorphic study of drainage basin characteristics in the Clinch Mountain area, Virginia and Tennessee. Dept. of Geology, Columbia University Technical Report Number 3, pp1-30.

Minasny, B., and McBratney, A.B., 2016. Digital soil mapping: A brief history and some lessons. Geoderma. 264: 301-311.

NRCS, 2003. Pacific Northwest MLRA Soil Survey Region Mo-1 Technical Note No.15, 18pp.

Poore, J.K., 1986. Quantitative shape analysis of soil map delineations in Benton county, Oregon. Master thesis of Soil Science in Oregon State University.

Schaetzl, R.J., and Anderson, S., 2005. Soils: genesis and geomorphology. Cambridge University Press, NY.

Schumm, S.A., 1956. Evaluation of drainage systems and slopes in Badlands at Perth Amboy, New Hersey. Bulletin of the Geological Society of America. 67: 597-646.

Small, C.G., 1988. Techniques of shape analysis on sets of points. International Statistical Review 56: 243-257.

Soil Survey Staff. 1993. Soil survey manual. Soil Conservation Service. U.S. Department of Agriculture Handbook 18.

Thas, O., 2010. Comparing distributions. Springer NY.

Weast, R.G., 1964. Mathematical tables from the handbook of chemistry and physics. 12th Edition. The Chemical Rubber Co., Cleveland, Ohio.

Zhang, D., and Lu, G., 2004. Review of shape representation and description techniques. Pattern Recognition 37: 1-9.

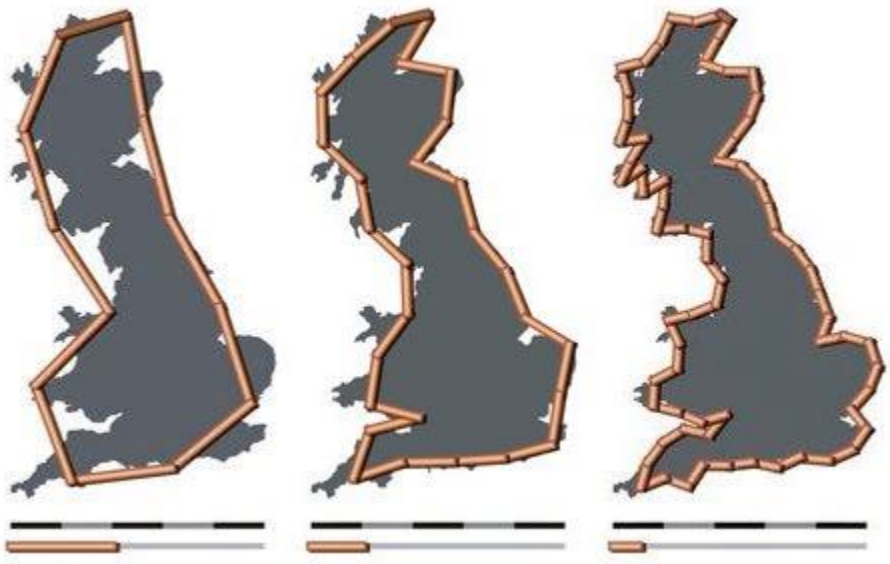


Figure 1. The coastal line of Britain. This figure is often used to illustrate coastal line paradox, but here it is used to explain the importance of mapping scale. The mapping scale increases from left to right, and more details can be shown at a larger scale (the image on the right). The perimeter and area defined by the orange line are different among the three images. The Soil Survey Geographic Database (SSURGO) used in this study has a mapping scale ranging from 1:12,000 to 1:63,360 (Soil Survey Staff, 1993). The link to the original image:
https://commons.wikimedia.org/wiki/File:Britain-fractal-coastline-combined.jpg
Another similar one can be found here:
https://www.computationallegalstudies.com/tag/law-as-a-complex-system/page/2/

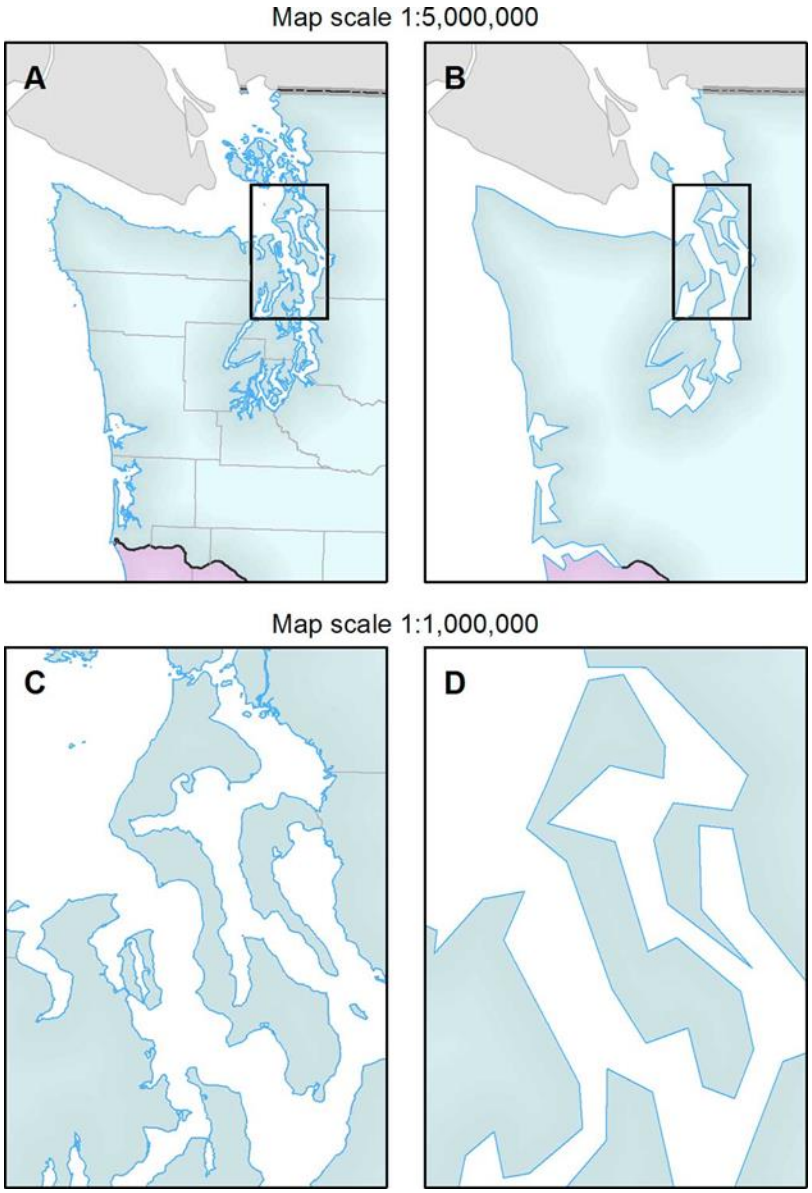


Figure 2. Map generalization. This image is taken from Buckley and Field (2011). Image A shows more details than image B does even both are shown at the same scale. This figure illustrates that different mappers can draw things differently, which will influence the parameters of a delineation, such as area and perimeter, which in turn will influence the values of shape descriptors. Link to the original dataset: http://www.esri.com/news/arcuser/0911/making-a-map-meaningful.html

Figure 3. The use of an aerial photo base in traditional soil survey (Soil Survey Staff, 1993).

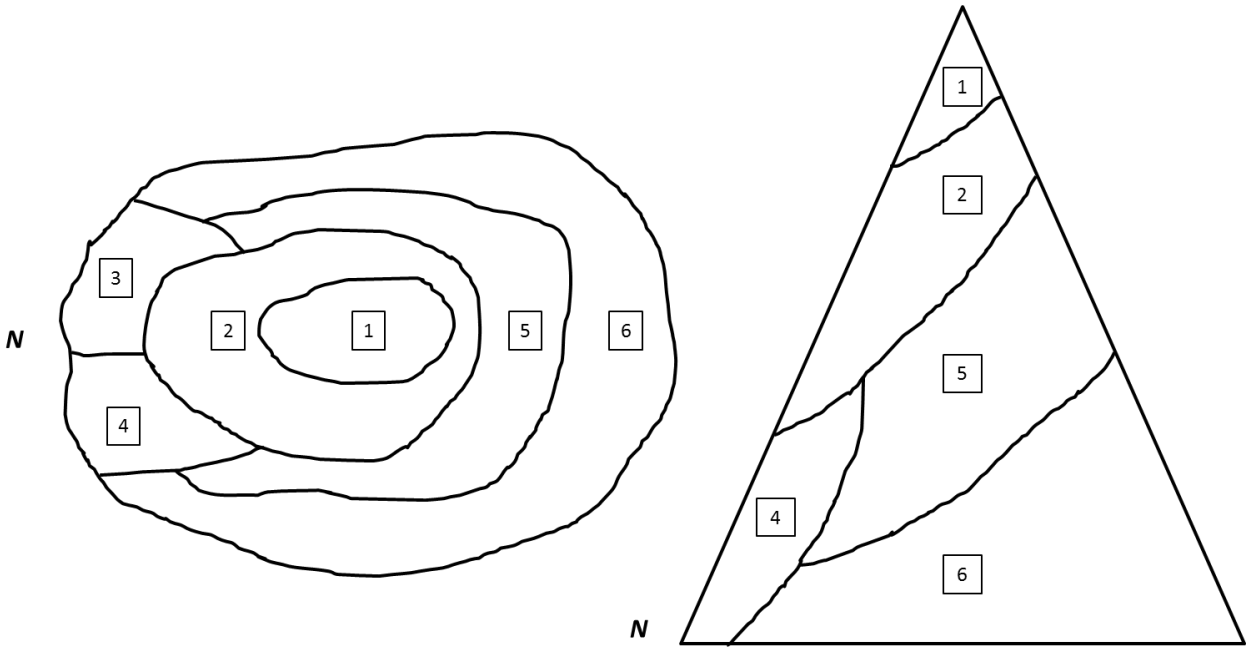


Figure 4. Illustration of a soil surveyor's mental landscape cone. The left and right images are map view and profile view of the same mountain, respectively. N represents for northern aspect. Number 1 through 6 represent different soil climate regime zones. Soil climate regime zone 1 locates at the highest elevation, and it has the highest precipitation and lowest soil temperature. In contrast, soil climate regime zone 6 locates at the lowest elevation, and it has the least precipitation and highest soil temperature. Soil climate regime zones 2 through 5 are somewhat in-between. Soil climate regime zone 4 has a northern aspect, and it has lower soil temperature than zone 6 has, as this is an illustration for northern hemisphere.

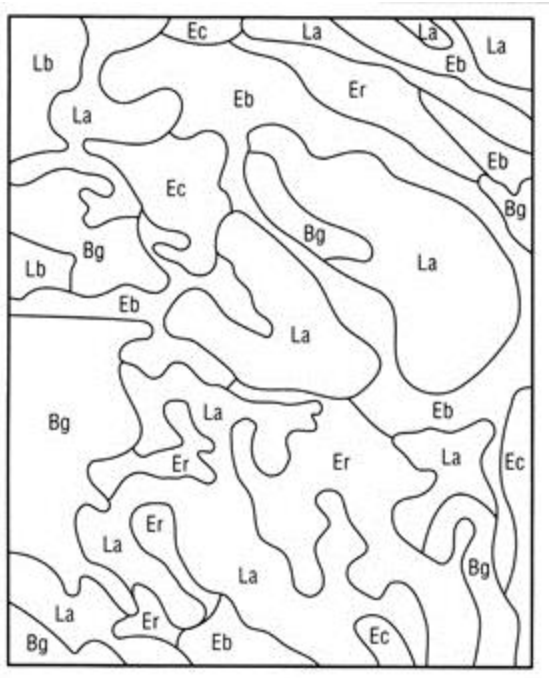


Figure 5. An example of soil delineations and the corresponding map units. In the above map, each enclosed area is called a soil delineation. The letters within each soil delineation are the map unit symbols. The soil delineations, which share the same map unit symbol, belong to the same map unit. A map unit is the collection of such delineations (Soil Survey Staff, 1993).

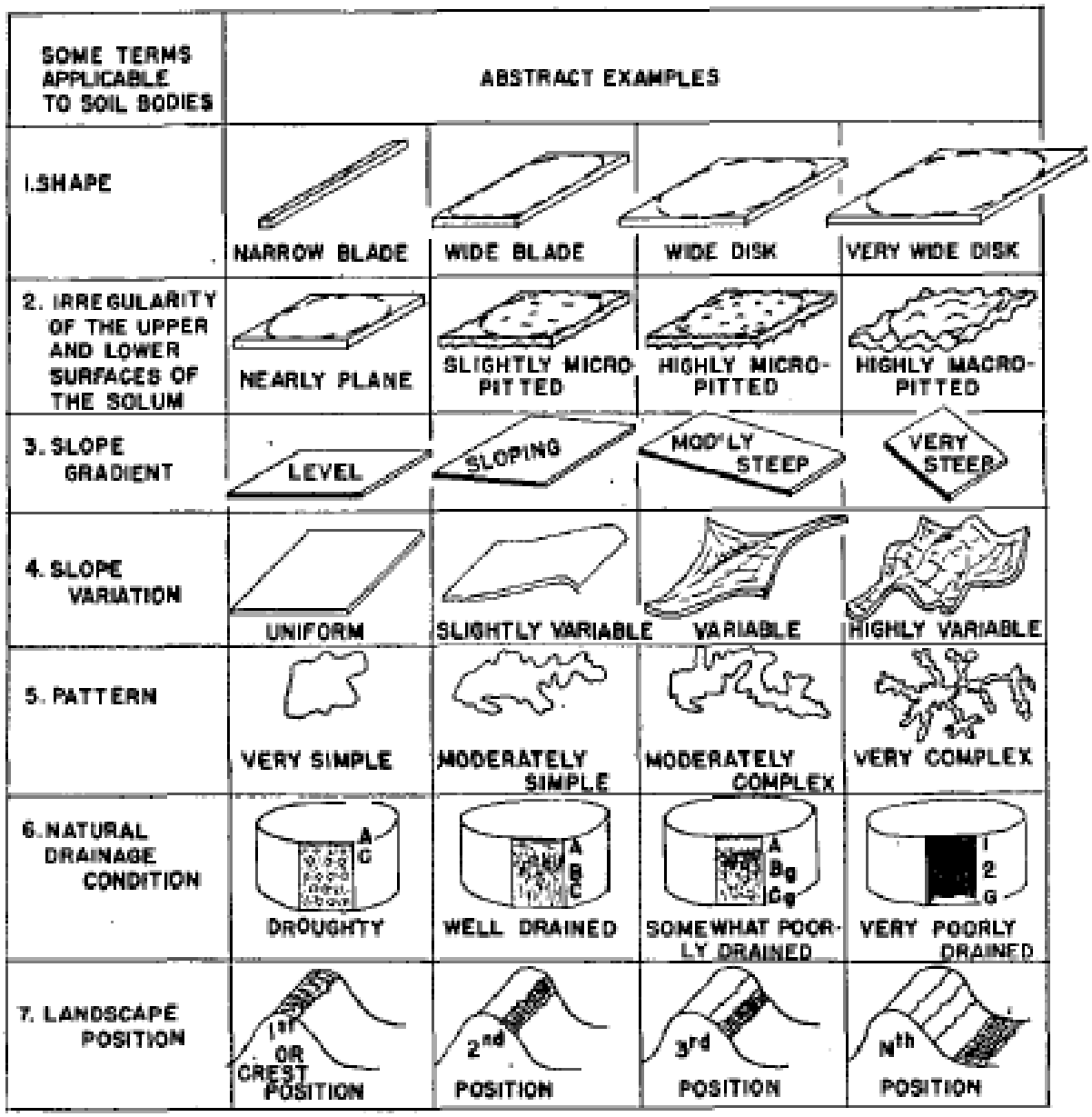


Figure 6. Six classes of soil shapes proposed by Hole. In his 1953 paper, Hole used the ratio of the perimeter of soil delineation over the perimeter of circle with the same area as the soil delineation to get a quantitative description for the shape of soil delineation. Based on the ratio, Hole ranked the delineated soil shapes into six classes varying from very simple to very complex. This figure is taken from Hole (1953).

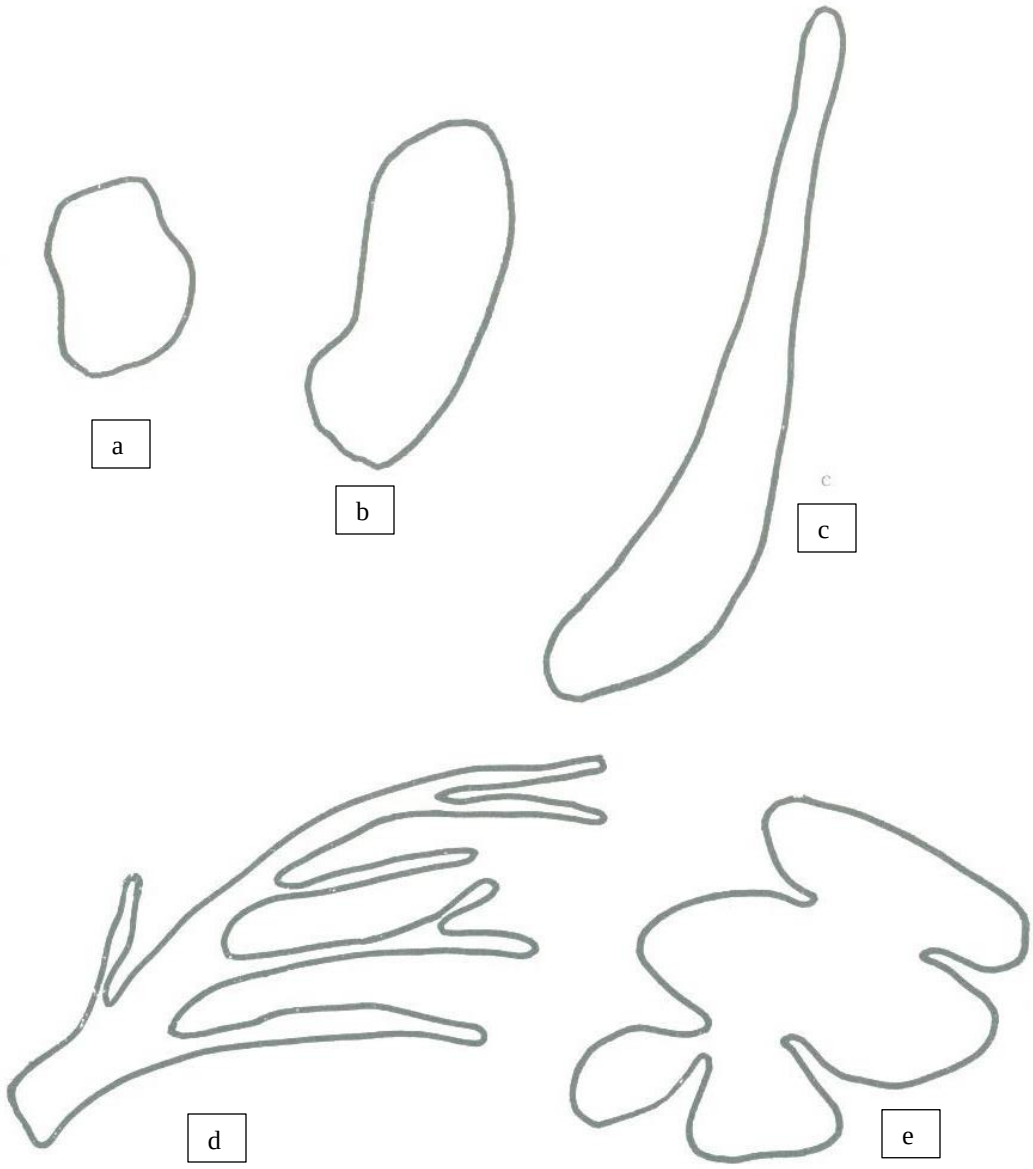


Figure 7. Five classes of soil shapes proposed by Fridland. In his 1976 book, Fridland used the ratio of the longest axis over the shortest axis to get a quantitative description for the shape of soil delineation. Based on the ratio, he grouped the delineated soil shapes into five categories: equilateral (a), elongated (b), linear (c), ramified asymmetric (d), and lobate symmetric (e). This figure is taken from Fridland (1976).

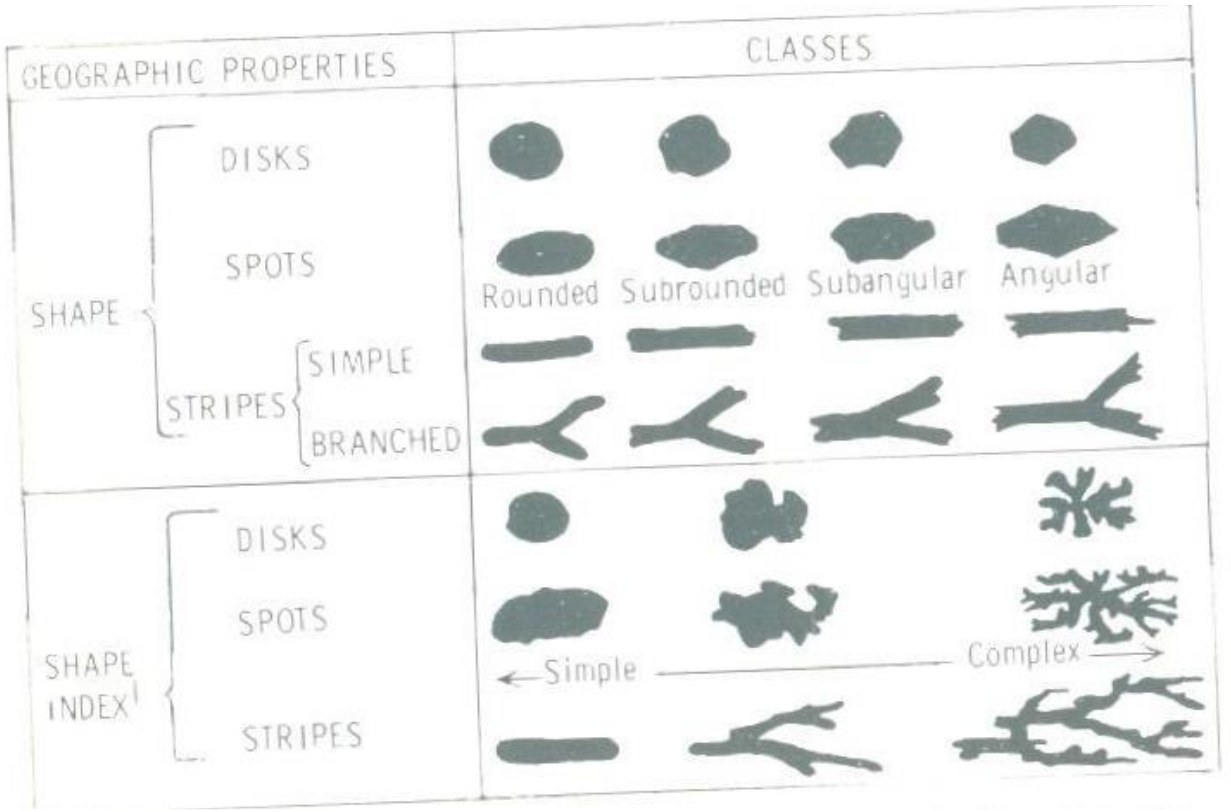


Figure 8. Three classes of soil shapes revised by Hole. In his 1978 paper, Hole adopted Fridland's idea, and used the ratio of the longest axis over the shortest axis to get a quantitative description for the shape of soil delineation. Based on the ratio, Hole grouped the delineated soil shapes into three categories: disk, spot, and stripe. In his 1985 book with Campbell, they further combined disk category into spot category. This figure is taken from Hole (1978).

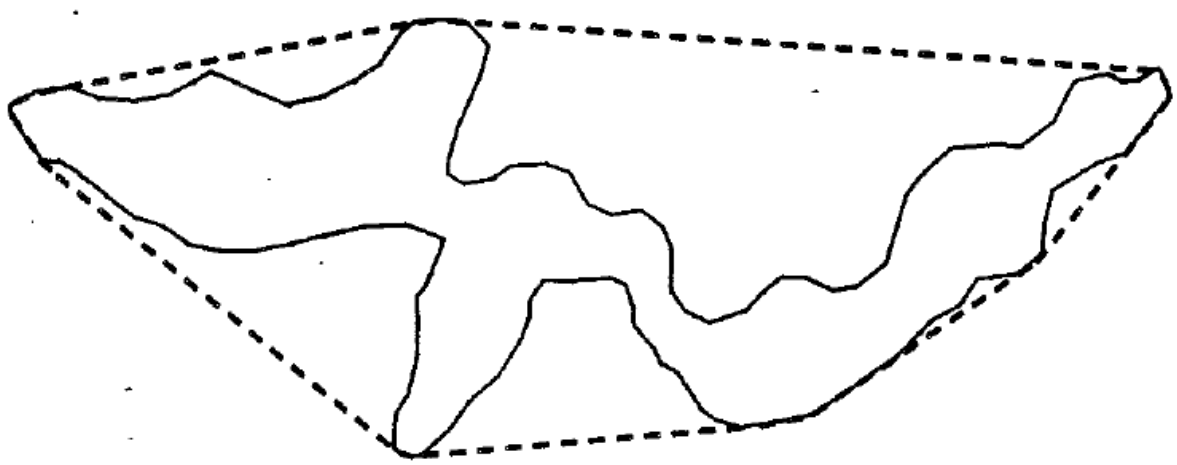


Figure 9. The concept of convex hull. A convex hull can be visualized as a polygon being wrapped by a rubber band. The solid line represents a soil delineation, and the dotted line represents its convex hull. This figure is taken from Poore (1986).

Figure 10. Study area of Part I. This graph shows the hydrologic units in Oregon at basin level, and the shadowed area is Willamette Basin. The zoomed in map shows a set of soil delineations from Cazadero soil series.

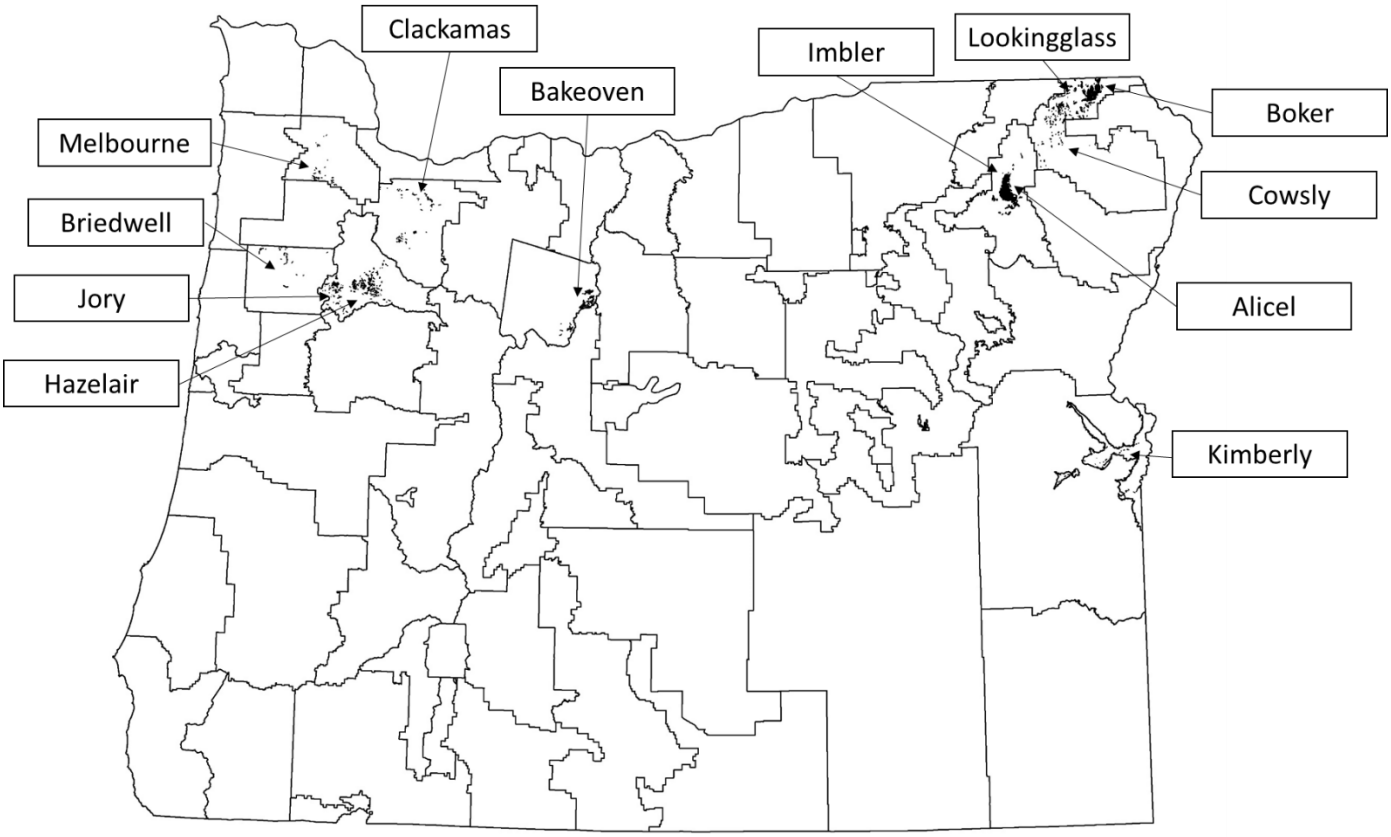


Figure 11. Study area of Part II. The selected soil map units were labeled by the dominant soil series and pointed to with arrows. Lookingglass was compared to Cowsly because of the difference in soil temperature regime. Clackamas was paired with Kimberly because of the difference in soil moisture regime. Bakeoven was paired with Bocker because of the differences in both soil temperature and moisture regimes. Melbourne was compared to Hazelair, Jory was compared to Hazelair and Melbourne separately, and Briedwell was compared to Imbler and Alicel separately, which are all based on difference in parent material. See Table 2 and 3 for more information.

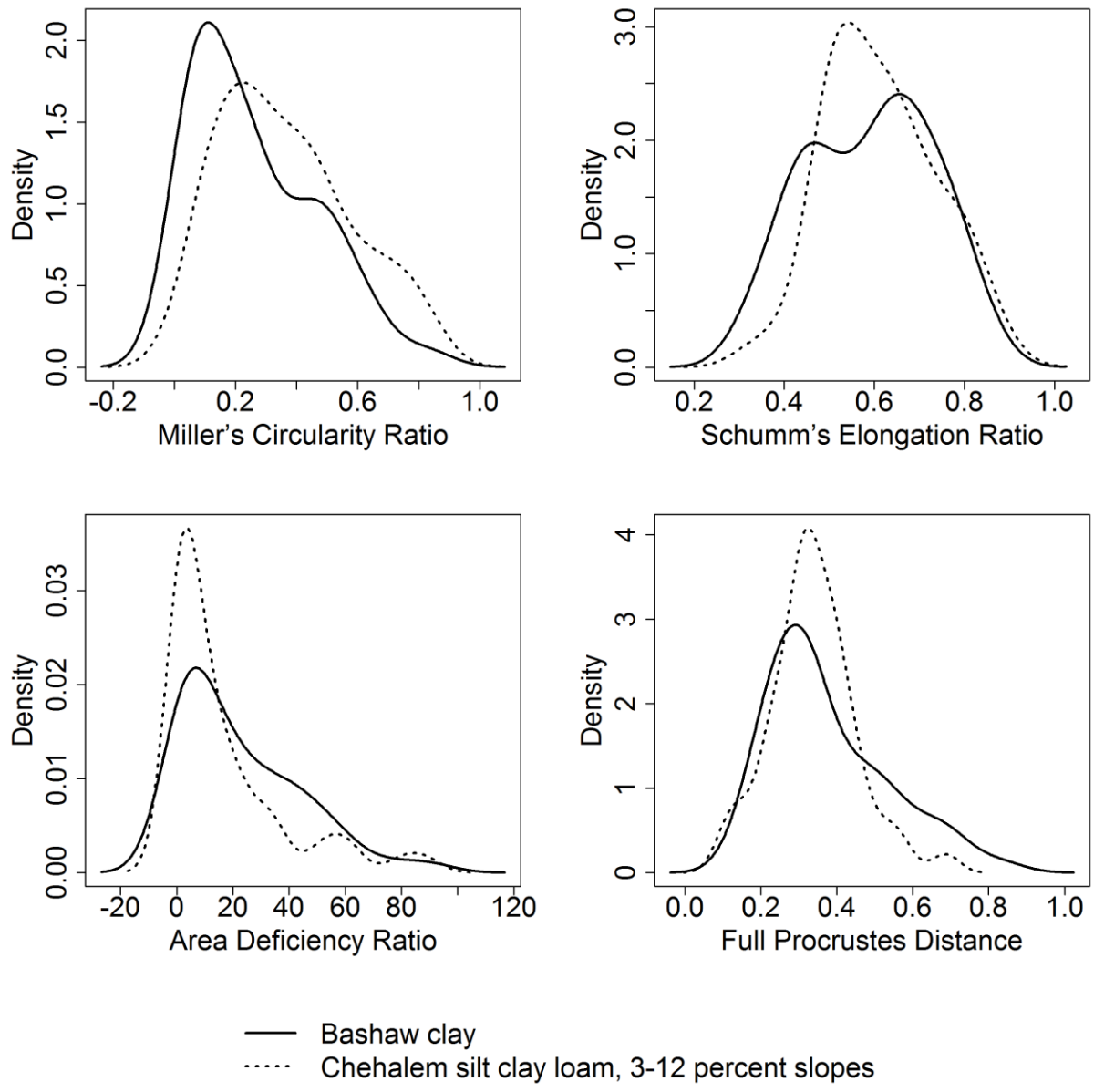


Figure 12. Kernel density estimation of selected shape descriptors for the two map units in Group A of Part I. The dominant soil series in each map unit are Bashaw (solid line) and Chehalem (dotted line) respectively. These two soil series are considered to be similar based on the taxonomic classification. For more information, see Table 1. Fifty delineations of each soil map unit were used for analysis. The slope range was not available for Bashaw in the SSURGO database, but it ranges from 0 to 12 percent based on the Official Soil Series Description online.

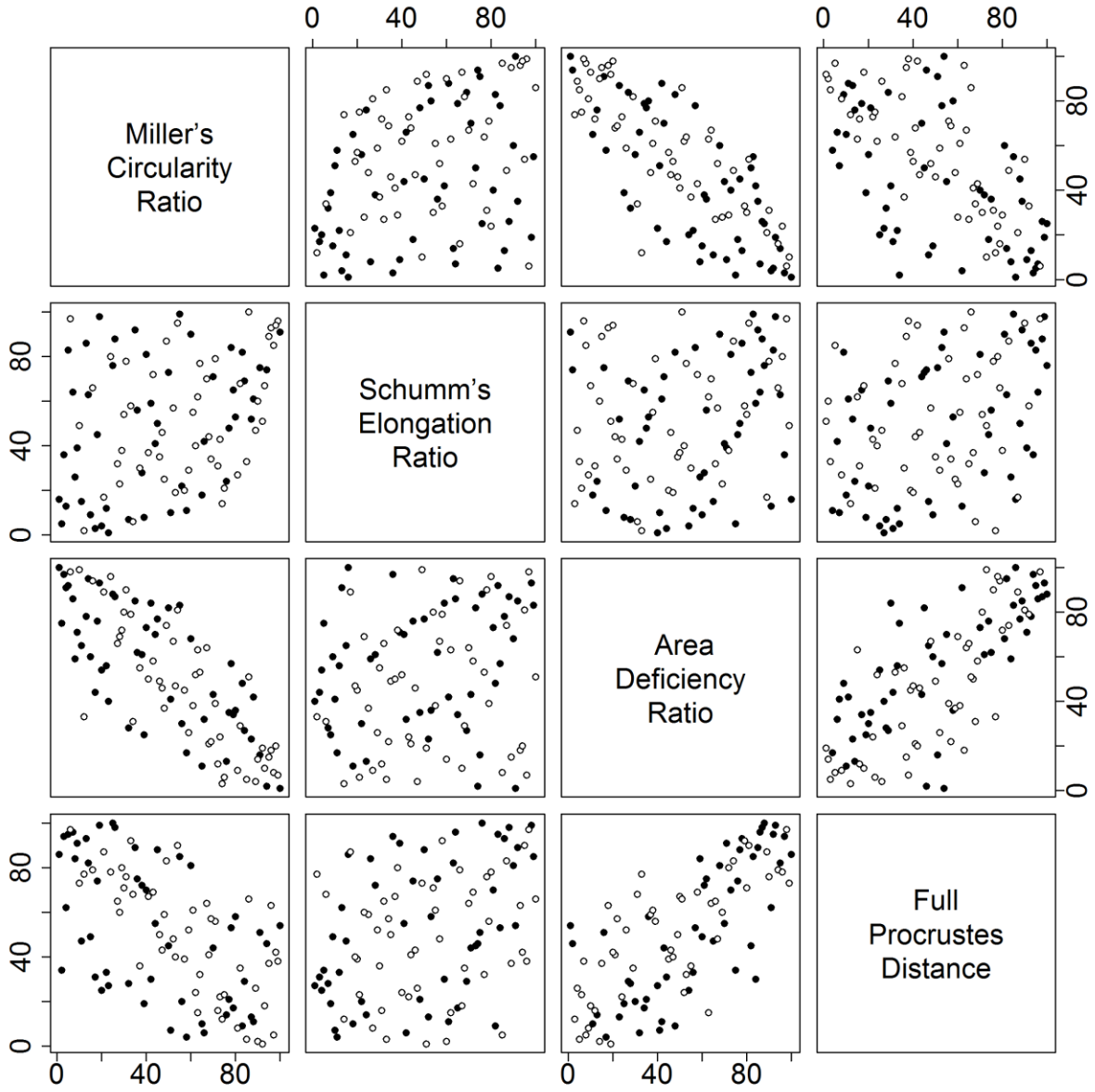


Figure 13. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group A of Part I. The dominant soil series in each map unit are Bashaw (black dots) and Chehalem (white dots) respectively. Fifty delineations of each soil map unit were used for analysis. The upper panel and the lower panel are mirror images along the diagonal. The separation of these two soil map units was not achieved by using any combination of these selected shape descriptors. This can be explained by the similarity of Bashaw and Chehalem based on their taxonomy classification.

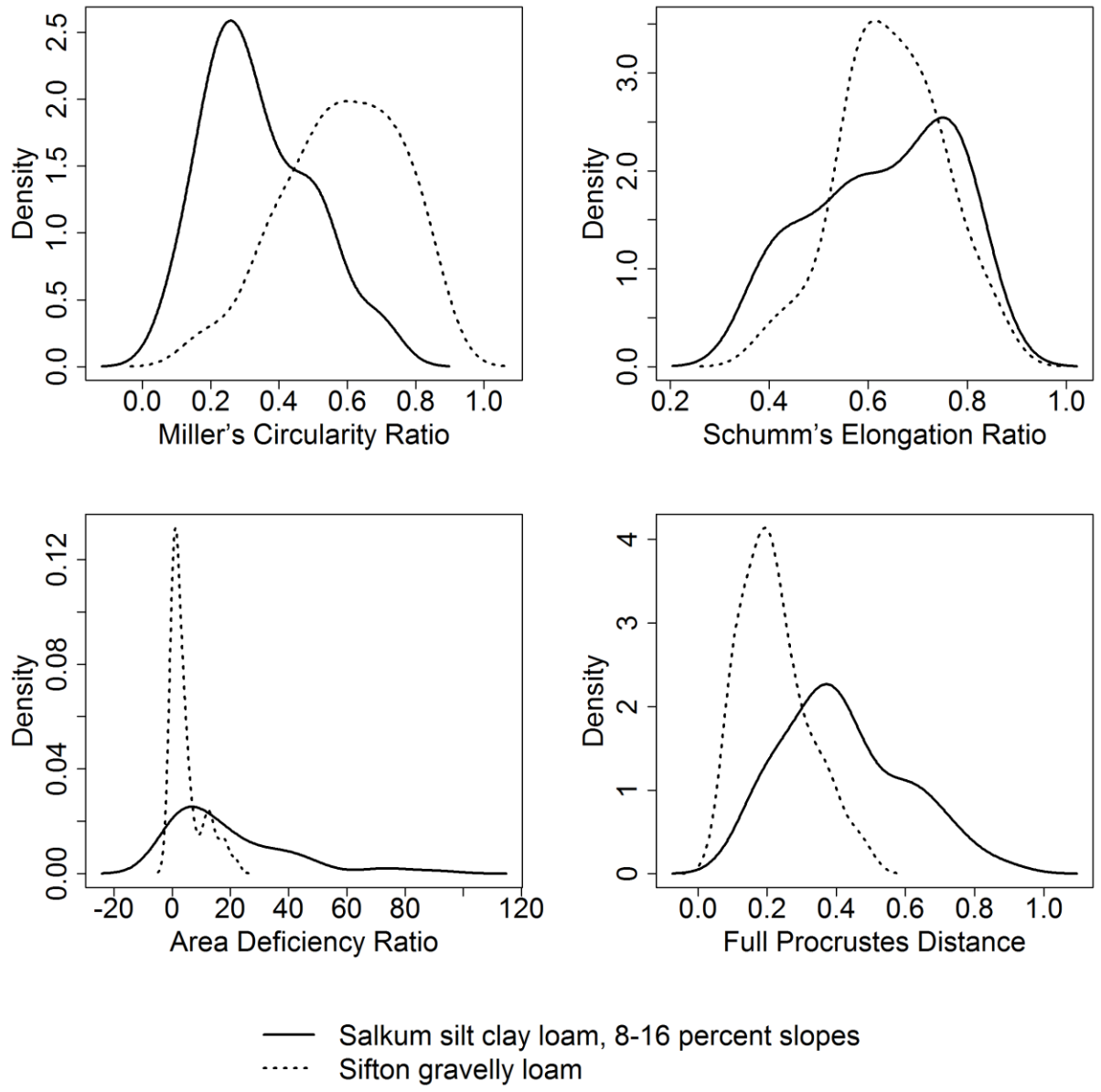


Figure 14. Kernel density estimation of selected shape descriptors for the two map units in Group B of Part I. The dominant soil series in each map unit are Salkum (solid line) and Sifton (dotted line) respectively. These two soil series are considered to be very different based on the taxonomic classification. For more information, see Table 1. Fifty delineations of each soil map unit were used for analysis. The slope range was not available for Sifton in the SSURGO database, but it ranges from 0 to 8 percent based on the Official Soil Series Description online.

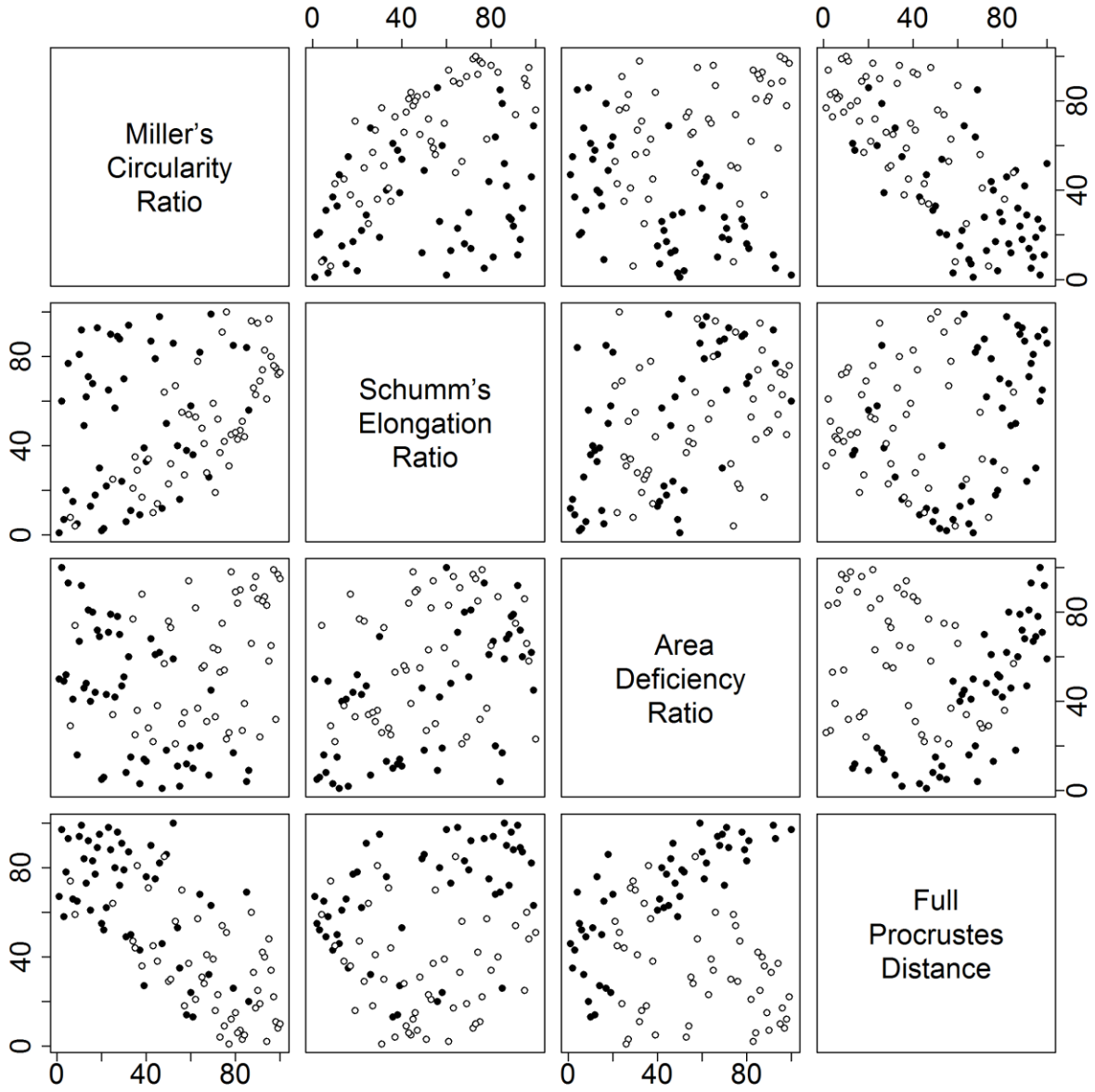


Figure 15. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group B of Part I. The dominant soil series in each map unit are Salkum (black dots) and Sifton (white dots) respectively. These two soil series are considered to be very different based on the taxonomy classification. Fifty delineations of each soil map unit were used for analysis. The upper panel and the lower panel are mirror images along the diagonal. The separation of these two soil map units was clearly achieved by using the combination of Miller's circularity ratio with any of the other three shape descriptors, and the combination of the area deficiency ratio with full Procrustes distance.

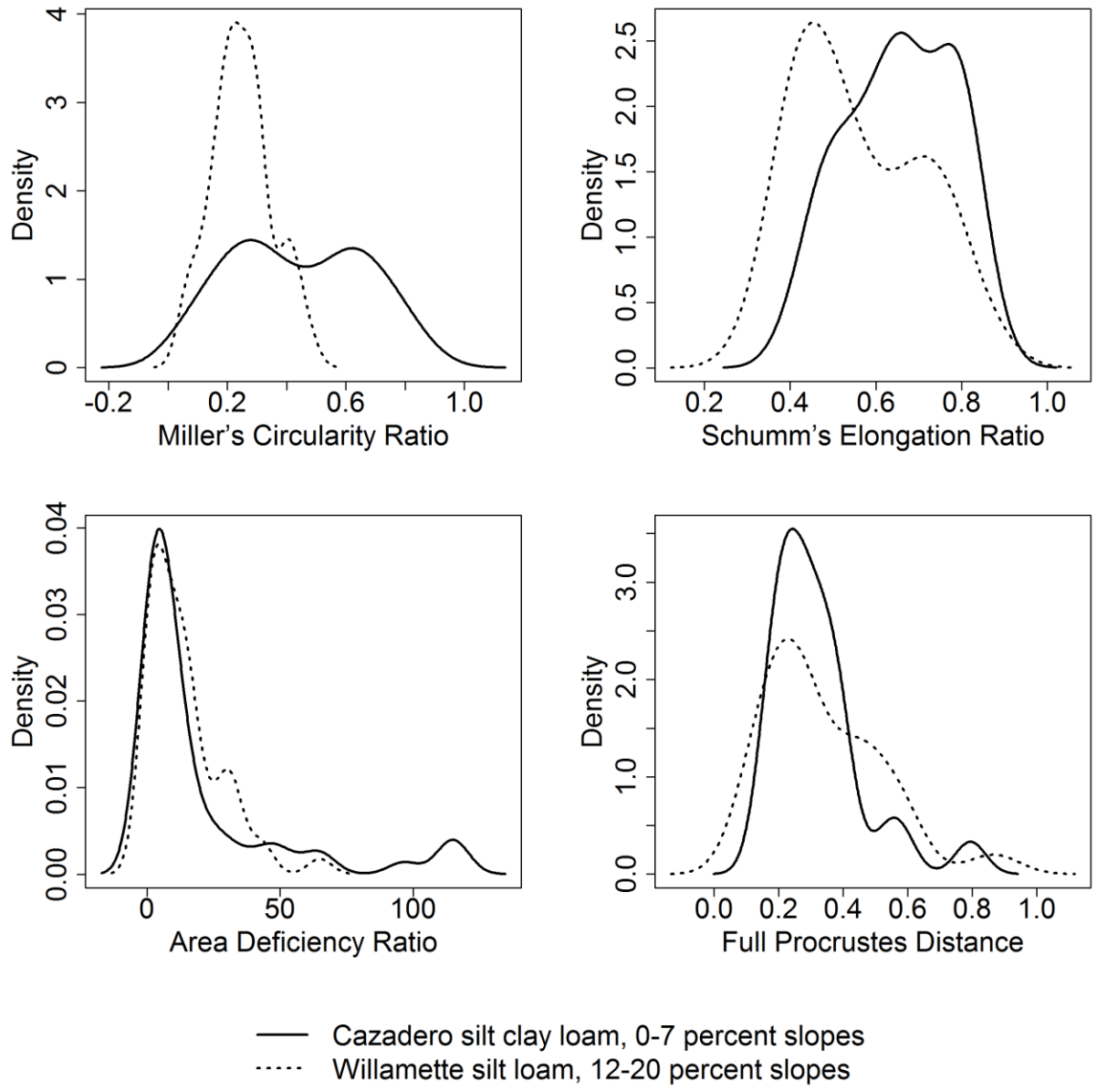


Figure 16. Kernel density estimation of selected shape descriptors for the two map units in Group C of Part I. The dominant soil series in each map unit are Cazadero (solid line) and Willamette (dotted line) respectively. These two soil series were considered to have different organic matter content based on the taxonomic classification. For more information, see Table 1. Fifty delineations of each soil map unit were used for analysis.

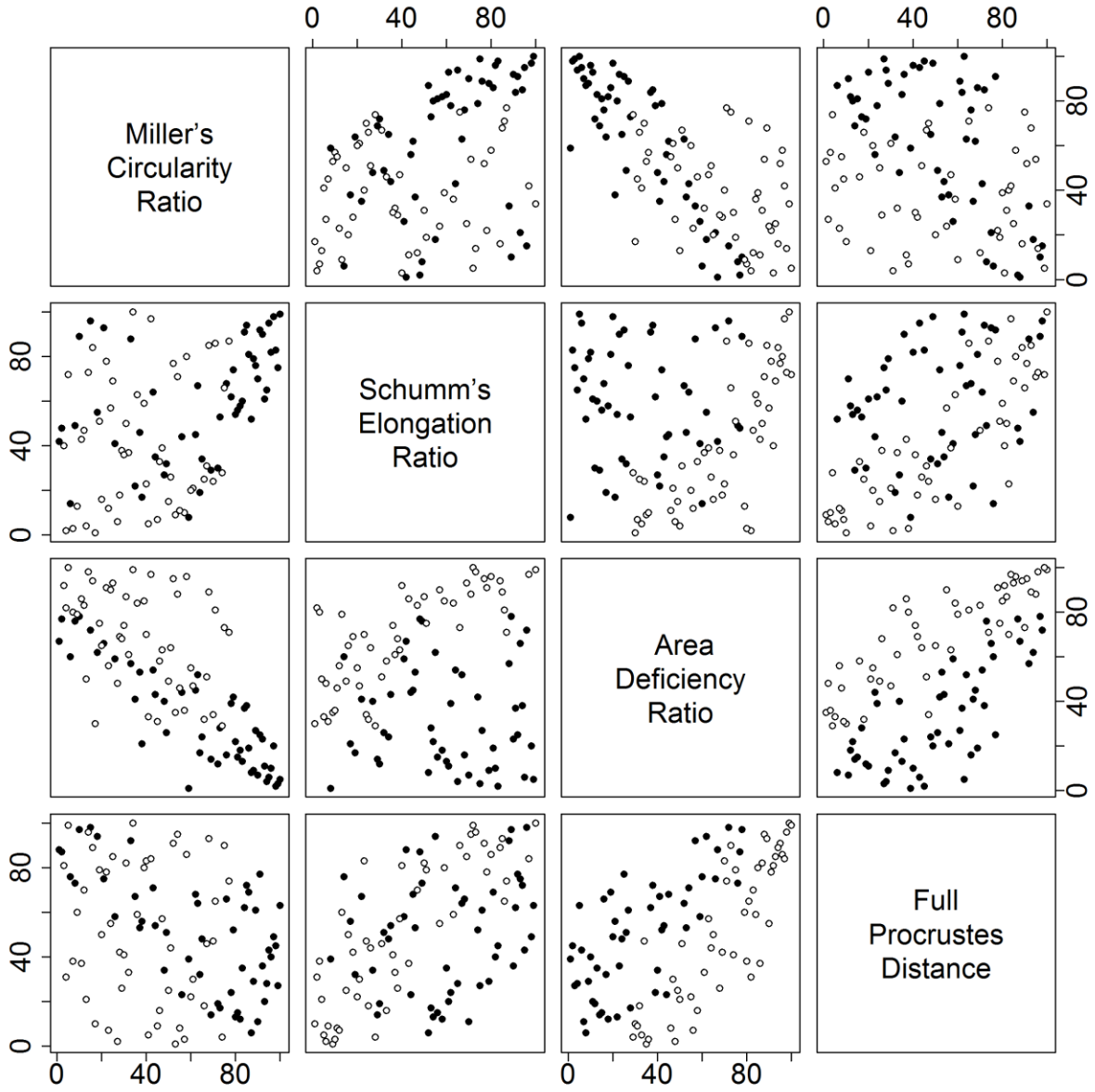


Figure 17. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group C of Part I. The dominant soil series in each map unit are Cazadero (black dots) and Willamette (white dots) respectively. These two soil series are considered to have different organic matter content based on the taxonomy classification. Fifty delineations of each soil map unit were used for analysis. The upper panel and the lower panel are mirror images along the diagonal. The separation of these two soil map units was clearly achieved by using the combination of the area deficiency ratio with either Schumm's elongation ratio or with full Procrustes distance.

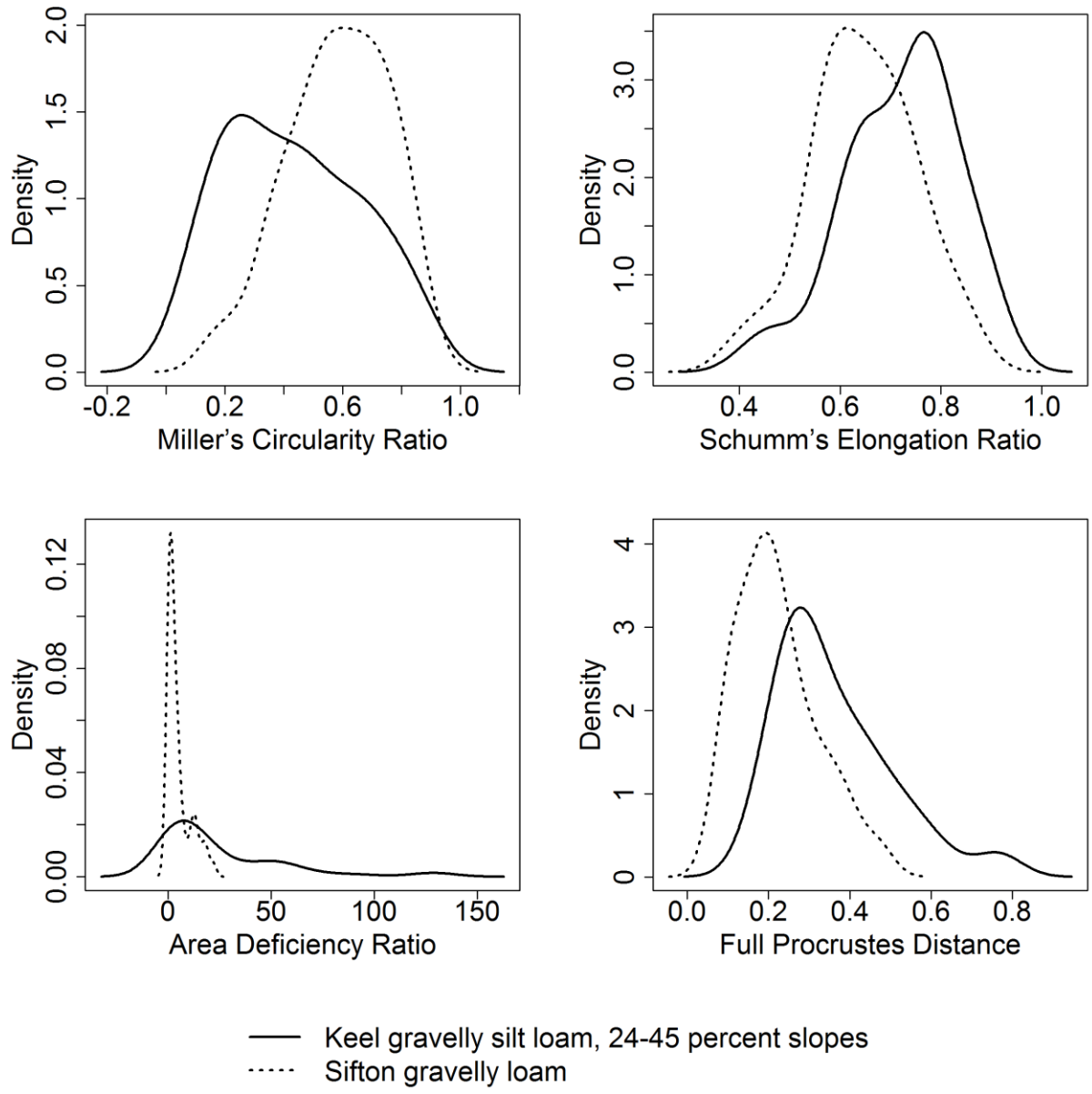


Figure 18. Kernel density estimation of selected shape descriptors for the two map units in Group D of Part I. The dominant soil series in each map unit are Keel (solid line) and Sifton (dotted line) respectively. These two soil series are considered to have different soil temperature regimes based on the taxonomic classification. For more information, see Table 1. Fifty delineations of each soil map unit were used for analysis. The slope range was not available for Sifton in the SSURGO database, but it ranges from 0 to 8 percent based on the Official Soil Series Description online.

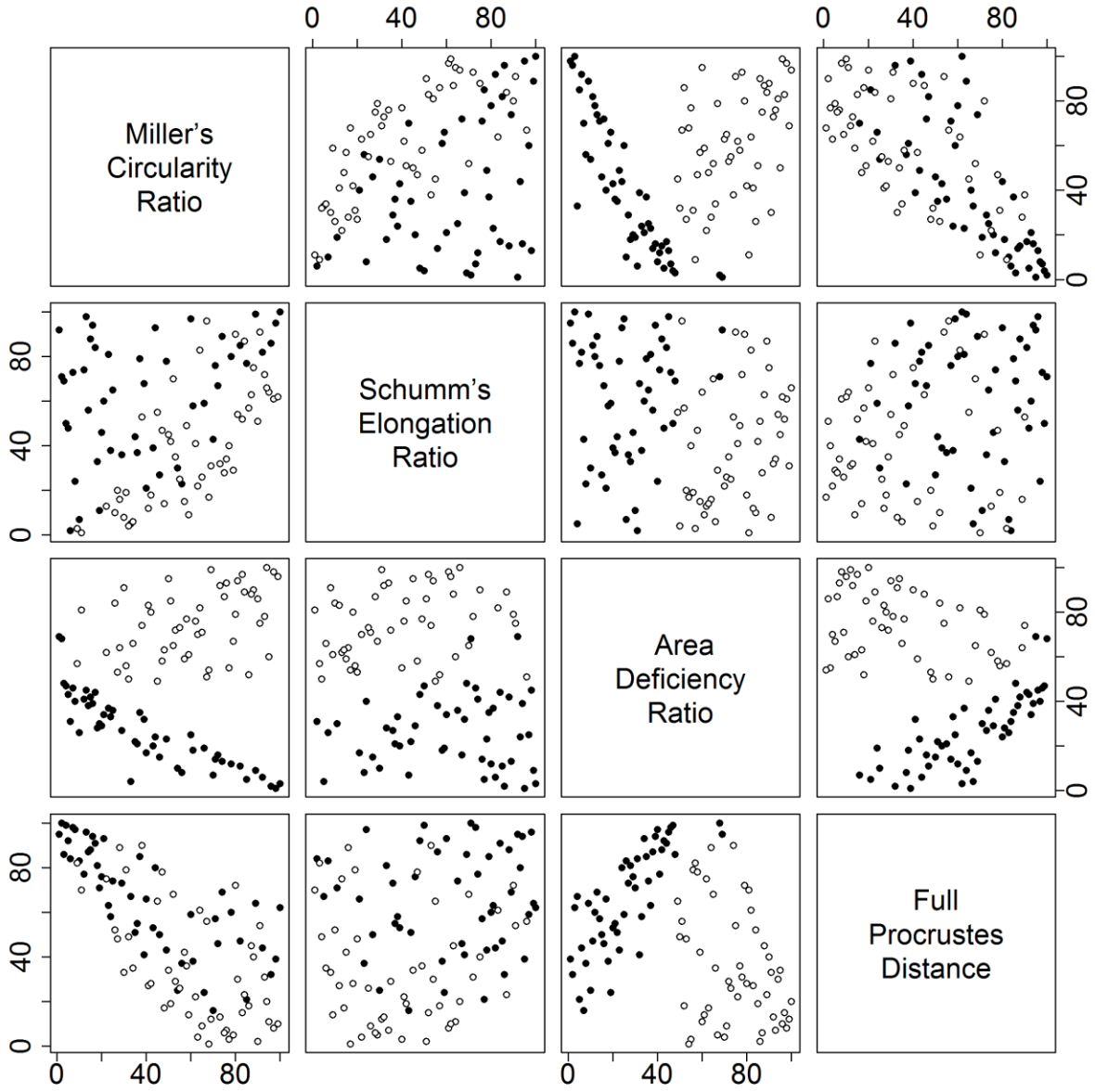


Figure 19. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group D of Part I. The dominant soil series in each map unit are Keel (black dots) and Sifton (white dots) respectively. These two soil series are considered to have different soil temperature regimes based on the taxonomy classification. The upper panel and the lower panel are mirror images along the diagonal. Fifty delineations of each soil map unit were used for analysis. The separation of these two soil map units was clearly achieved by using the combination of the area deficiency ratio with either Schumm's elongation ratio or with full Procrustes distance.

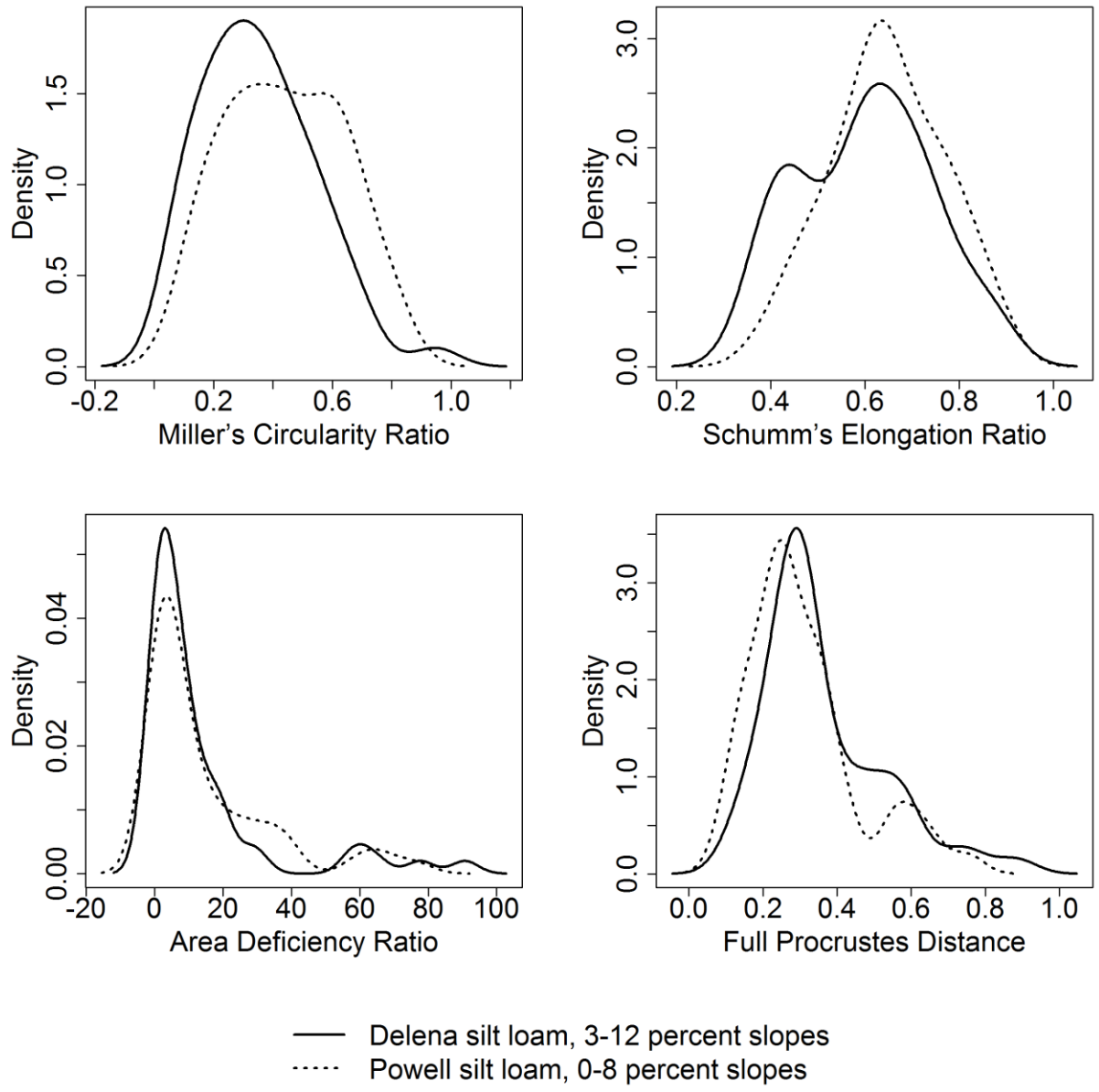


Figure 20. Kernel density estimation of selected shape descriptors for the two map units in Group E of Part I. The dominant soil series in each map unit are Delena (solid line) and Powell (dotted line) respectively. These two soil series are considered to have different soil moisture regimes based on the taxonomic classification. For more information, see Table 1. Fifty delineations of each soil map unit were used for analysis.

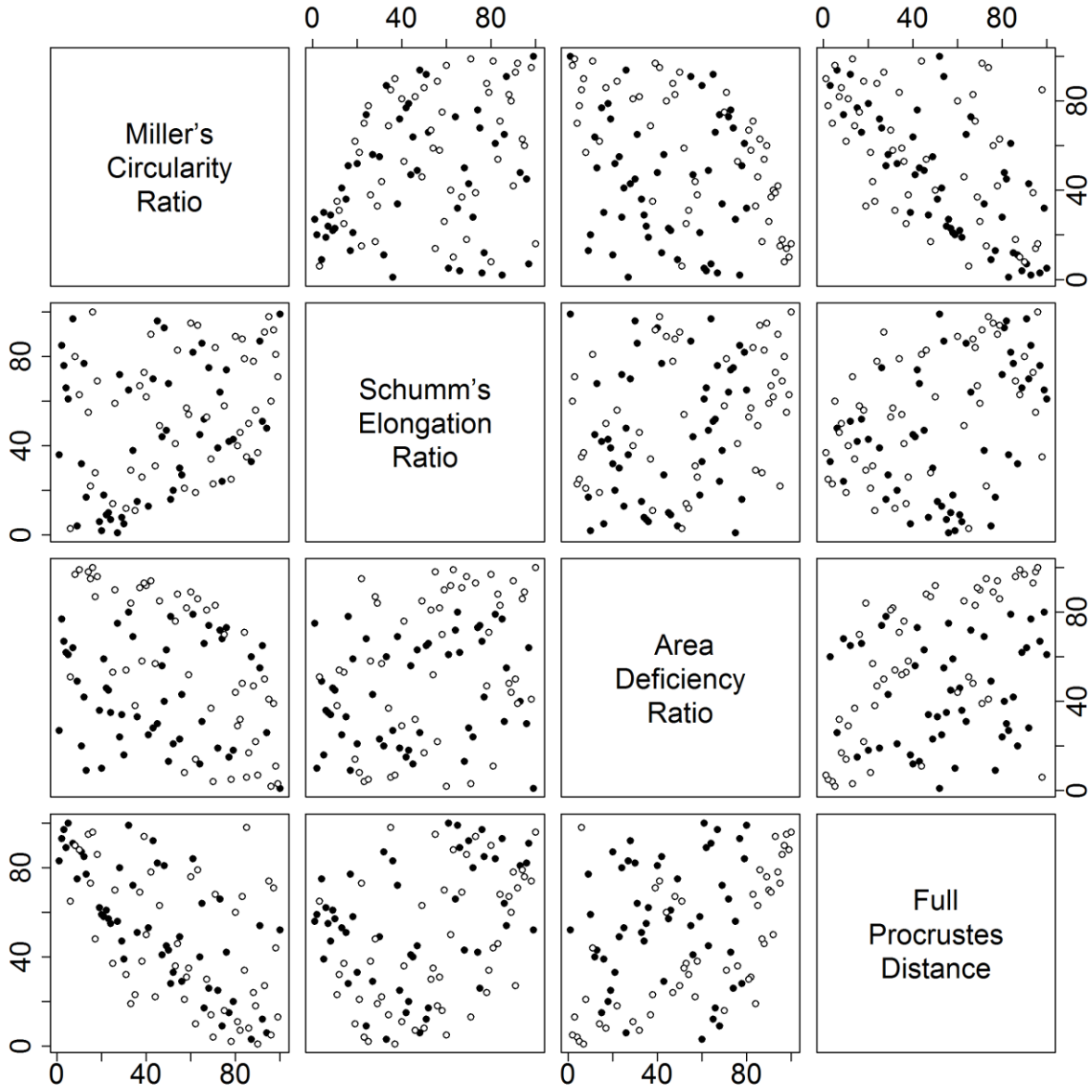


Figure 21. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group E of Part I. The dominant soil series in each map unit are Delena (black dots) and Powell (white dots) respectively. These two soil series are considered to have different soil moisture regimes based on the taxonomy classification. The upper panel and the lower panel are mirror images along the diagonal. Fifty delineations of each soil map unit were used for analysis. The separation of these two soil map units was not achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test also indicated the distributions of selected shape descriptors were not significantly different.

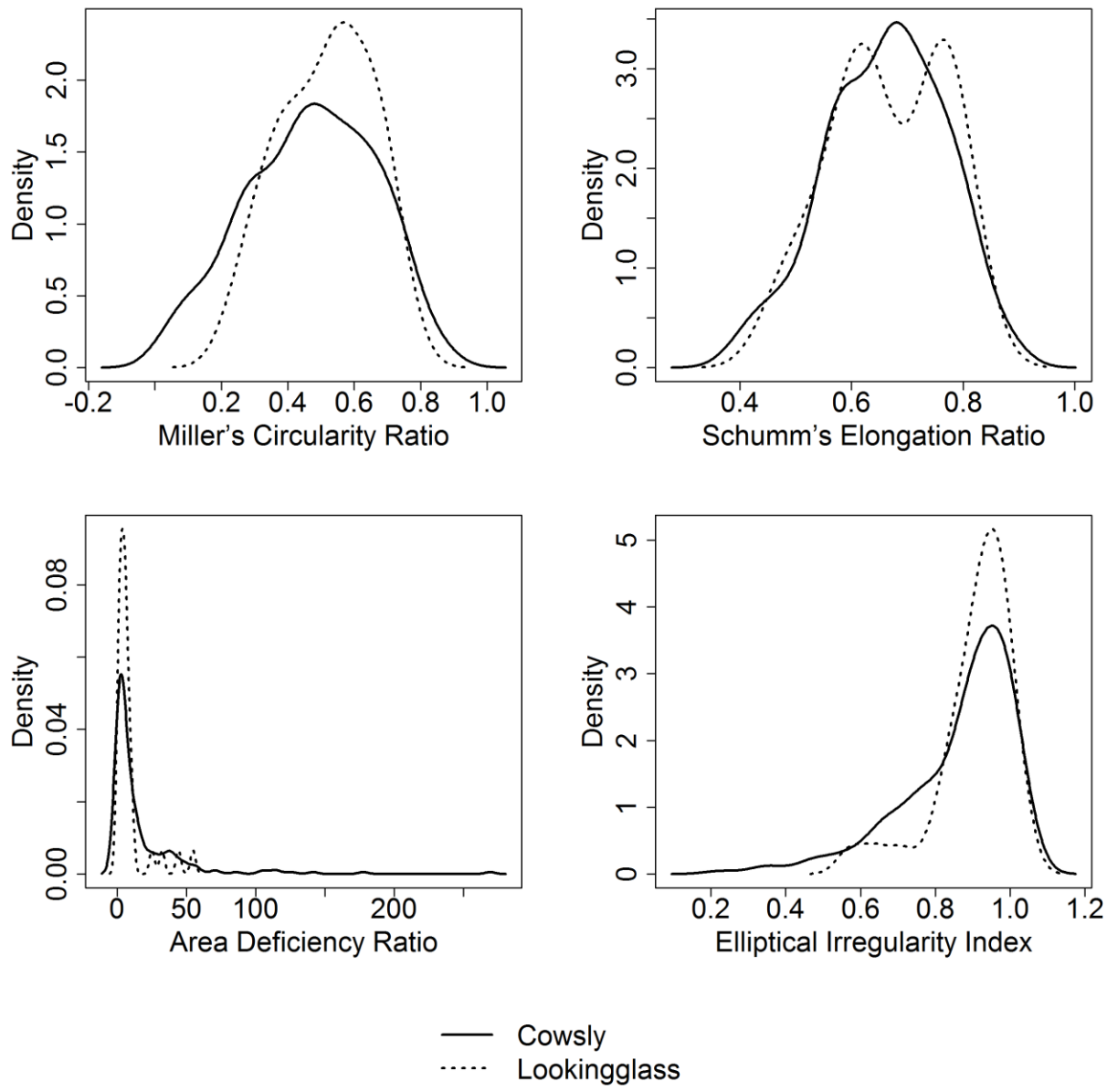


Figure 22. The kernel density estimation of selected shape descriptors for the two map units in Group 1 of Part II. The dominant soil series in each map unit are Cowsly (solid line) and Lookingglass (dotted line) respectively. These two soil series are considered to have different soil temperature regimes based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

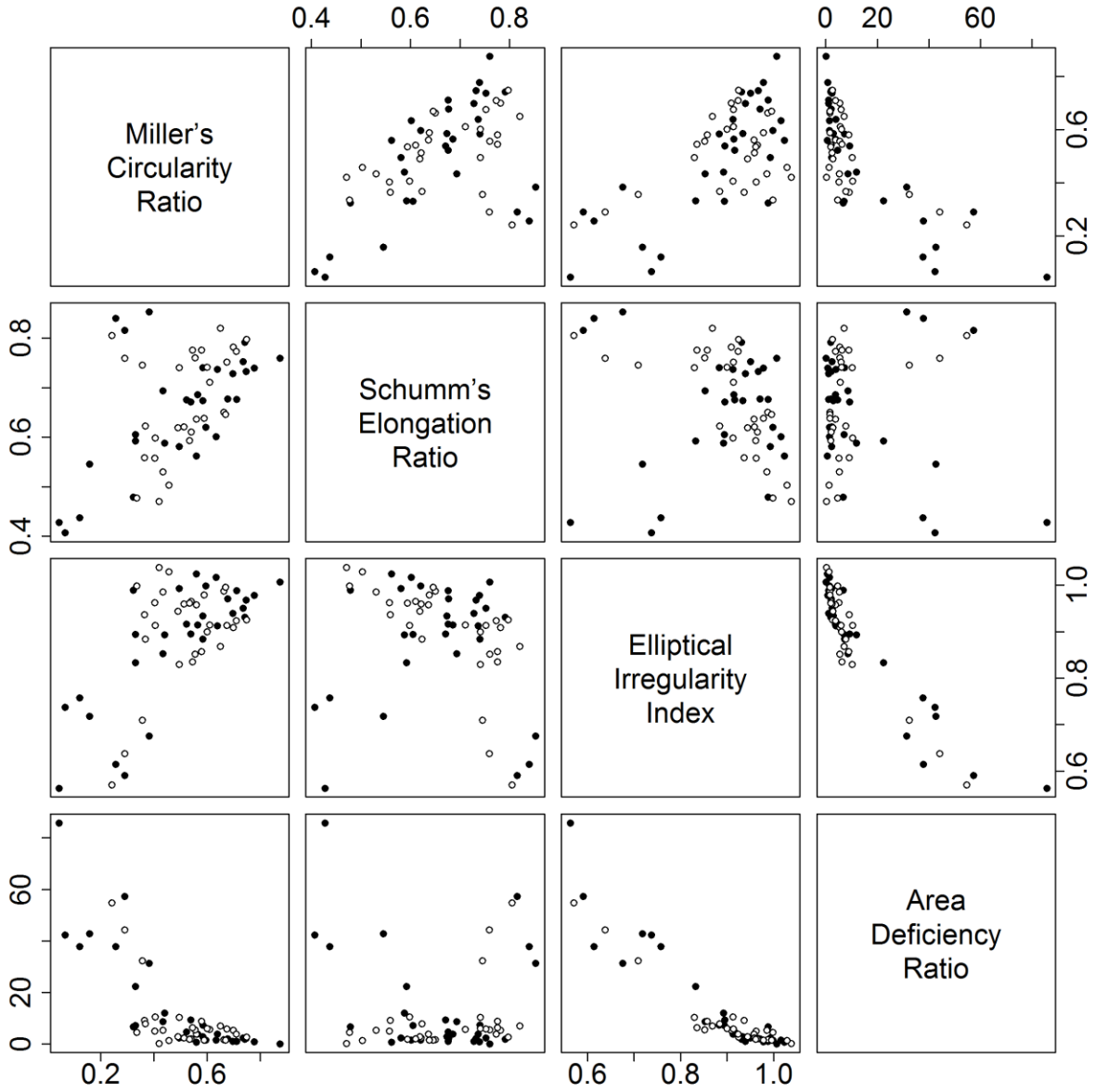


Figure 23. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 1 of Part II. The dominant soil series in each map unit are Cowsly (black dots) and Lookingglass (white dots) respectively. These two soil series are considered to have different soil temperature regimes based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test also indicated the distributions of selected shape descriptors were not significantly different.

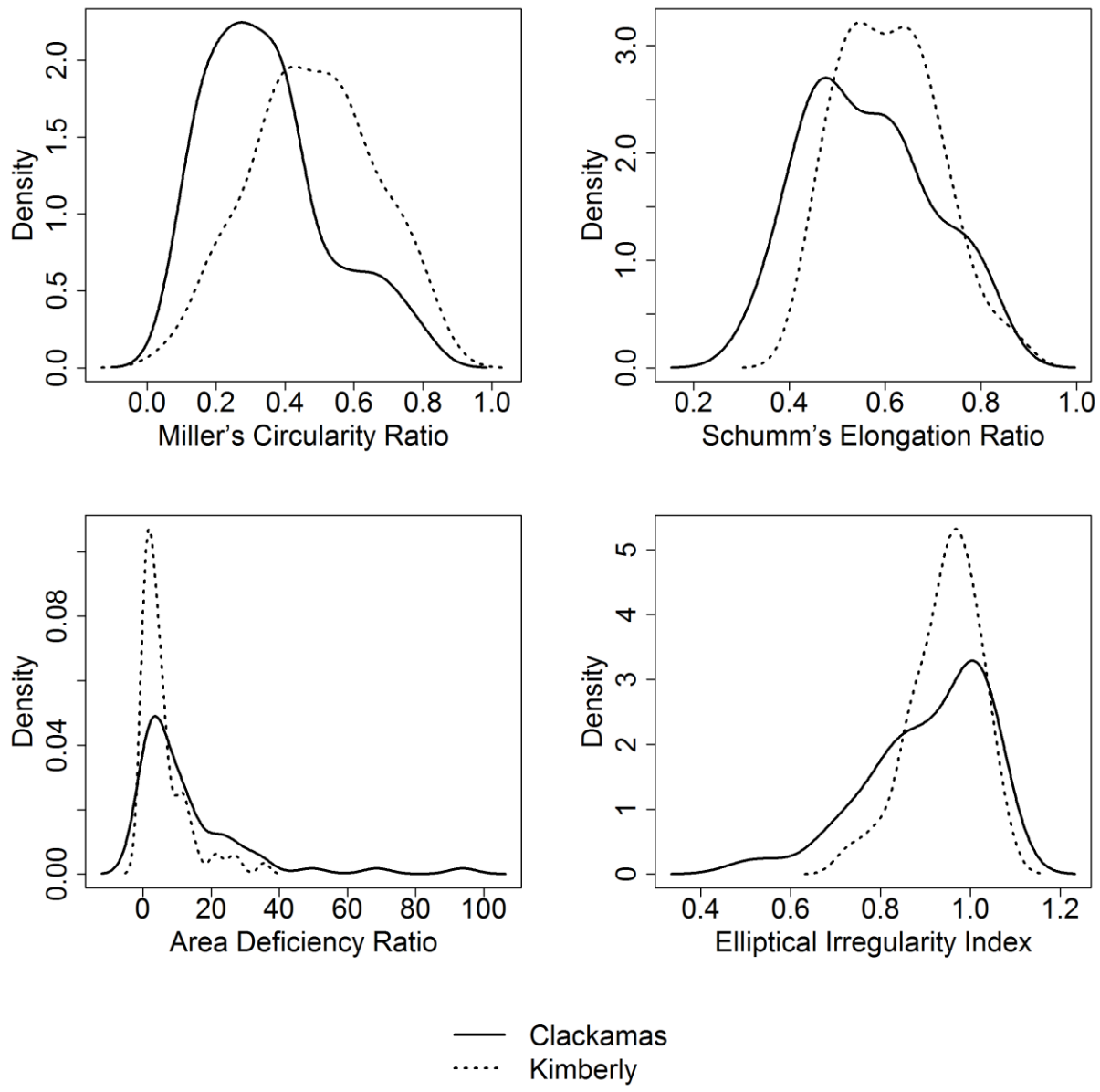


Figure 24. The kernel density estimation of selected shape descriptors for the two map units in Group 2 of Part II. The dominant soil series in each map unit are Clackamas (solid line) and Kimberly (dotted line) respectively. These two soil series are considered to have different soil moisture regimes based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

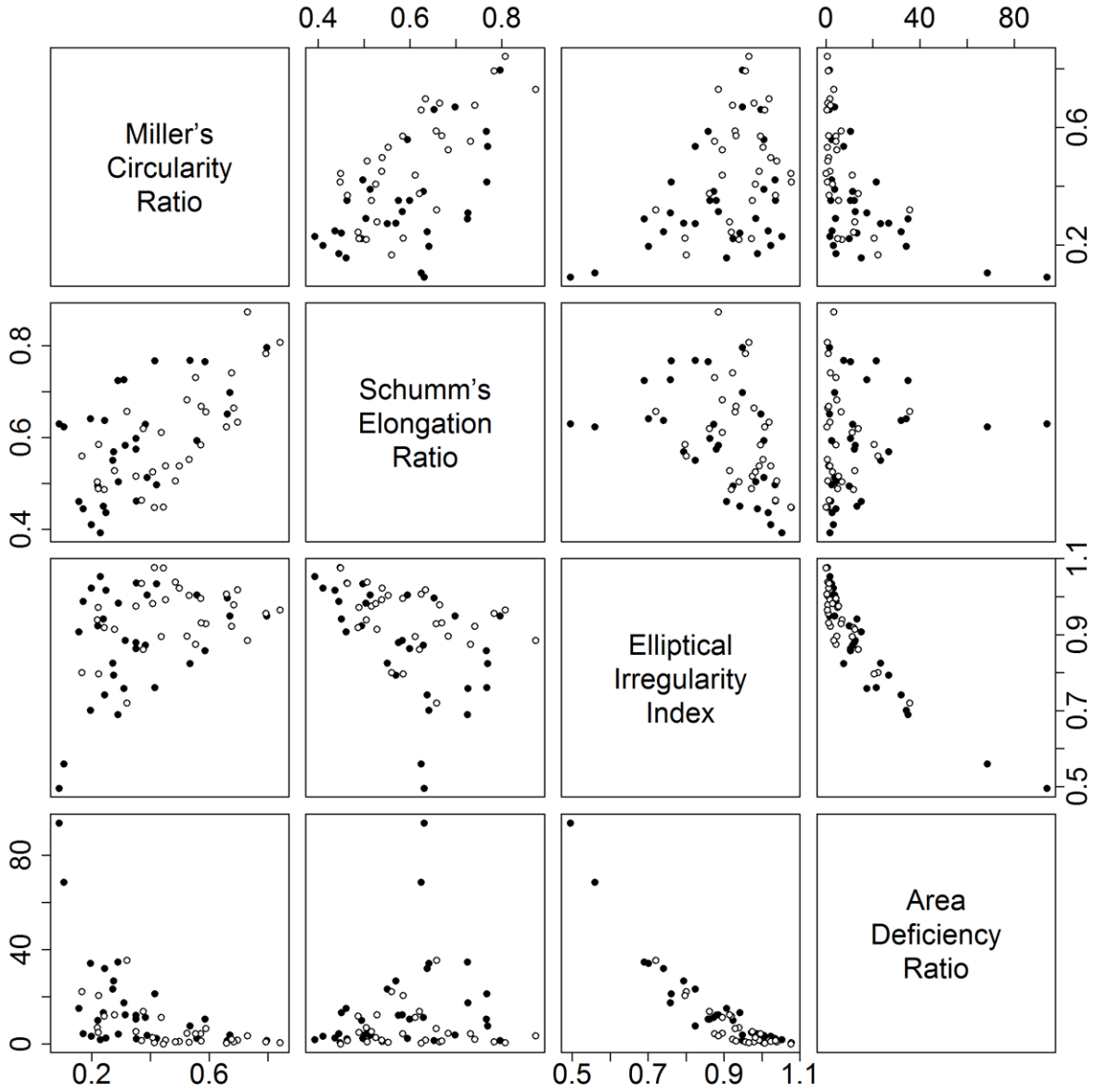


Figure 25. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 2 of Part II. The dominant soil series in each map unit are Clackamas (black dots) and Kimberly (white dots) respectively. These two soil series are considered to have different soil moisture regimes based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. Even the results of the two-sample Kolmogorov-Smirnov test indicated the distributions of selected shape descriptors were significantly different; the separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors.

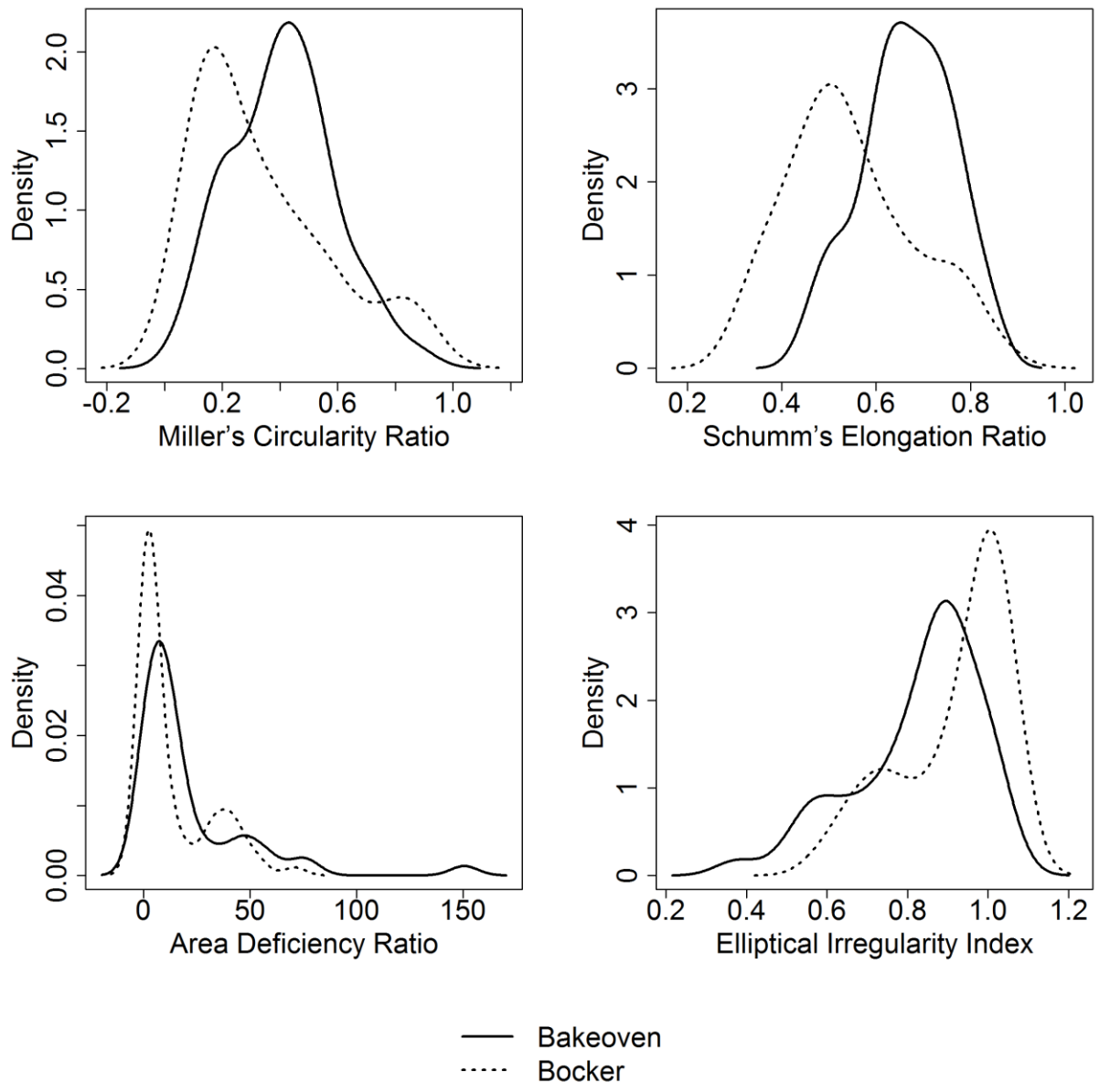


Figure 26. The kernel density estimation of selected shape descriptors for the two map units in Group 3 of Part II. The dominant soil series in each map unit are Bakeoven (solid line) and Bocker (dotted line) respectively. These two soil series were considered to have different soil temperature and moisture regimes based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

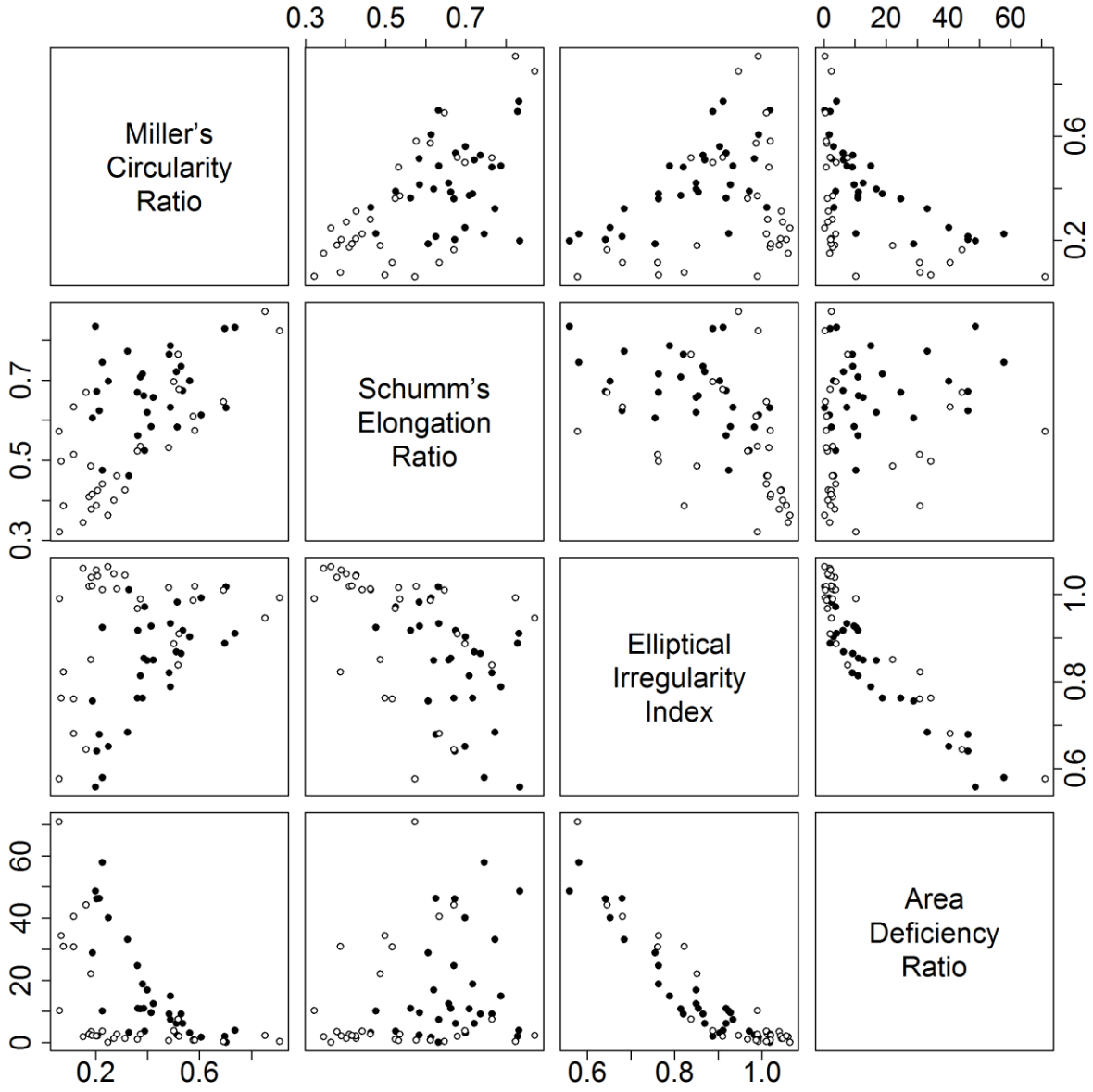


Figure 27. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 3 of Part II. The dominant soil series in each map unit are Bakeoven (black dots) and Bocker (white dots) respectively. These two soil series are considered to have different soil temperature and moisture regimes based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was achieved by using the combination of shape descriptors.

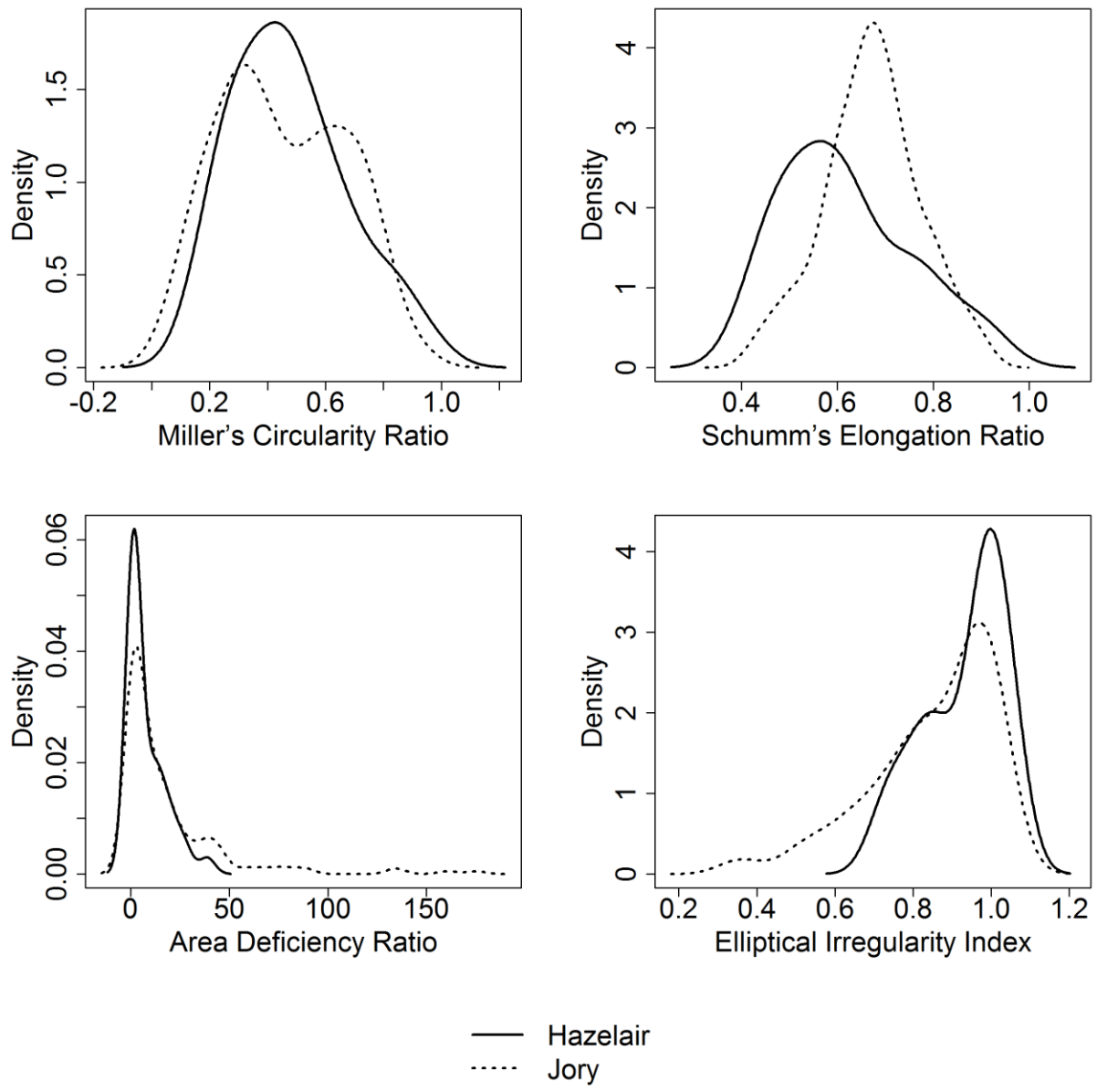


Figure 28. The kernel density estimation of selected shape descriptors for the two map units in Group 4 of Part II. The dominant soil series in each map unit are Hazelair (solid line) and Jory (dotted line) respectively. These two soil series were considered to have different parent material based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

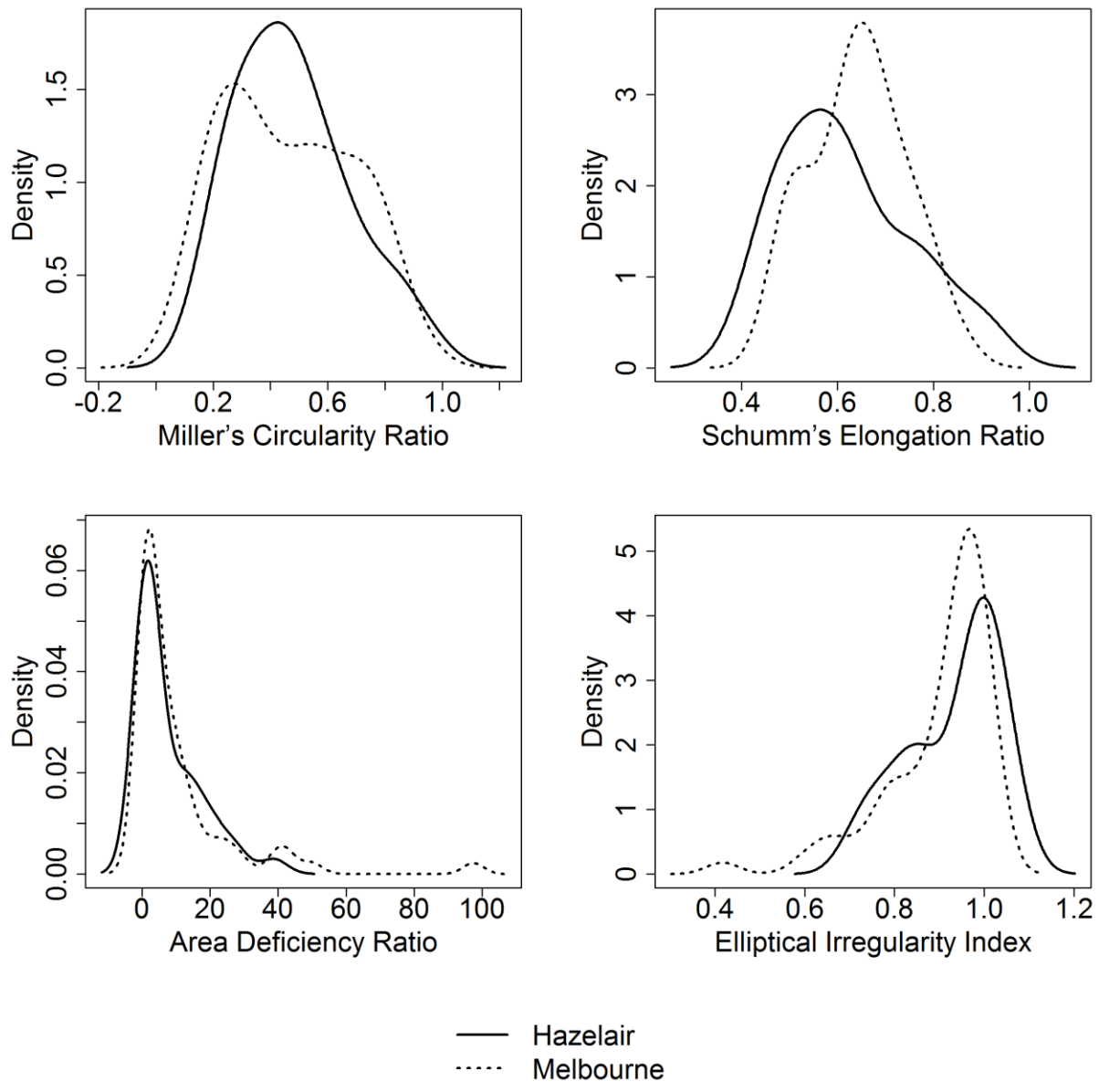


Figure 29. The kernel density estimation of selected shape descriptors for the two map units in Group 5 of Part II. The dominant soil series in each map unit are Hazelair (solid line) and Melbourne (dotted line) respectively. These two soil series were considered to have different parent material based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

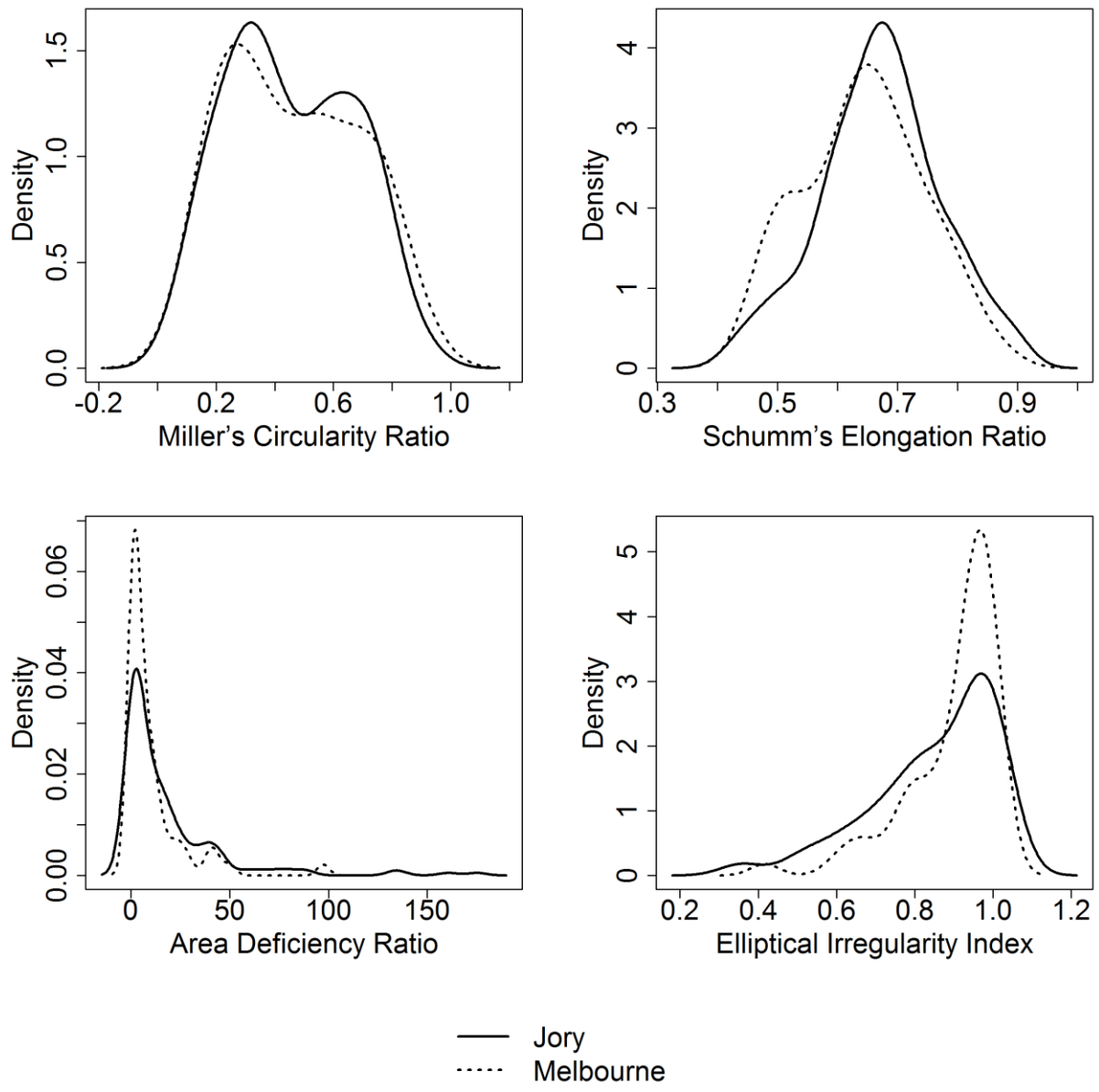


Figure 30. The kernel density estimation of selected shape descriptors for the two map units in Group 6 of Part II. The dominant soil series in each map unit are Jory (solid line) and Melbourne (dotted line) respectively. These two soil series were considered to have different parent material based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

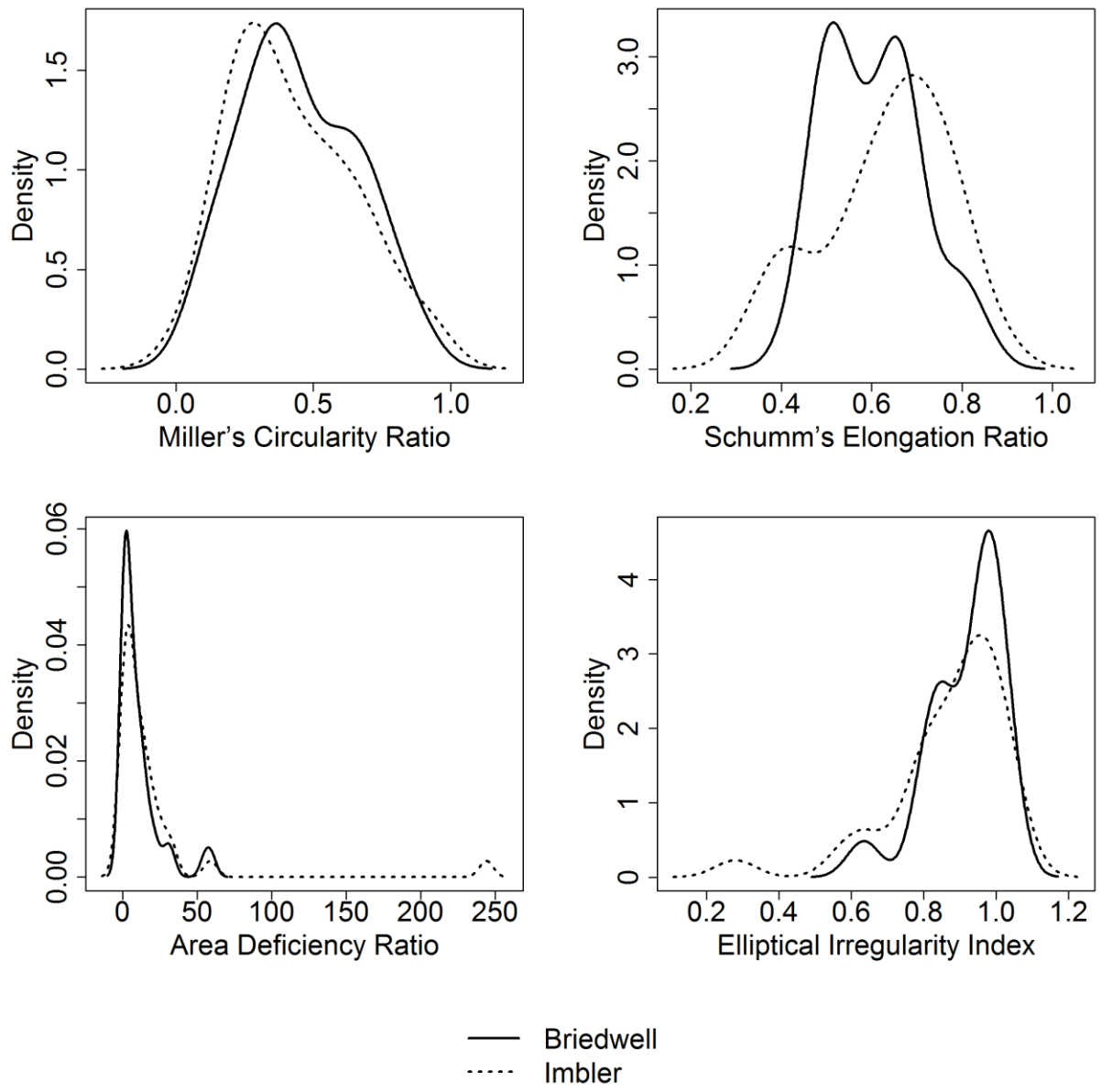


Figure 31. The kernel density estimation of selected shape descriptors for the two map units in Group 7 of Part II. The dominant soil series in each map unit are Briedwell (solid line) and Imbler (dotted line) respectively. These two soil series were considered to have different parent material based on the SSURGO database. For more information, see Tables 2 and 3. All delineations of each soil map unit were used for analysis.

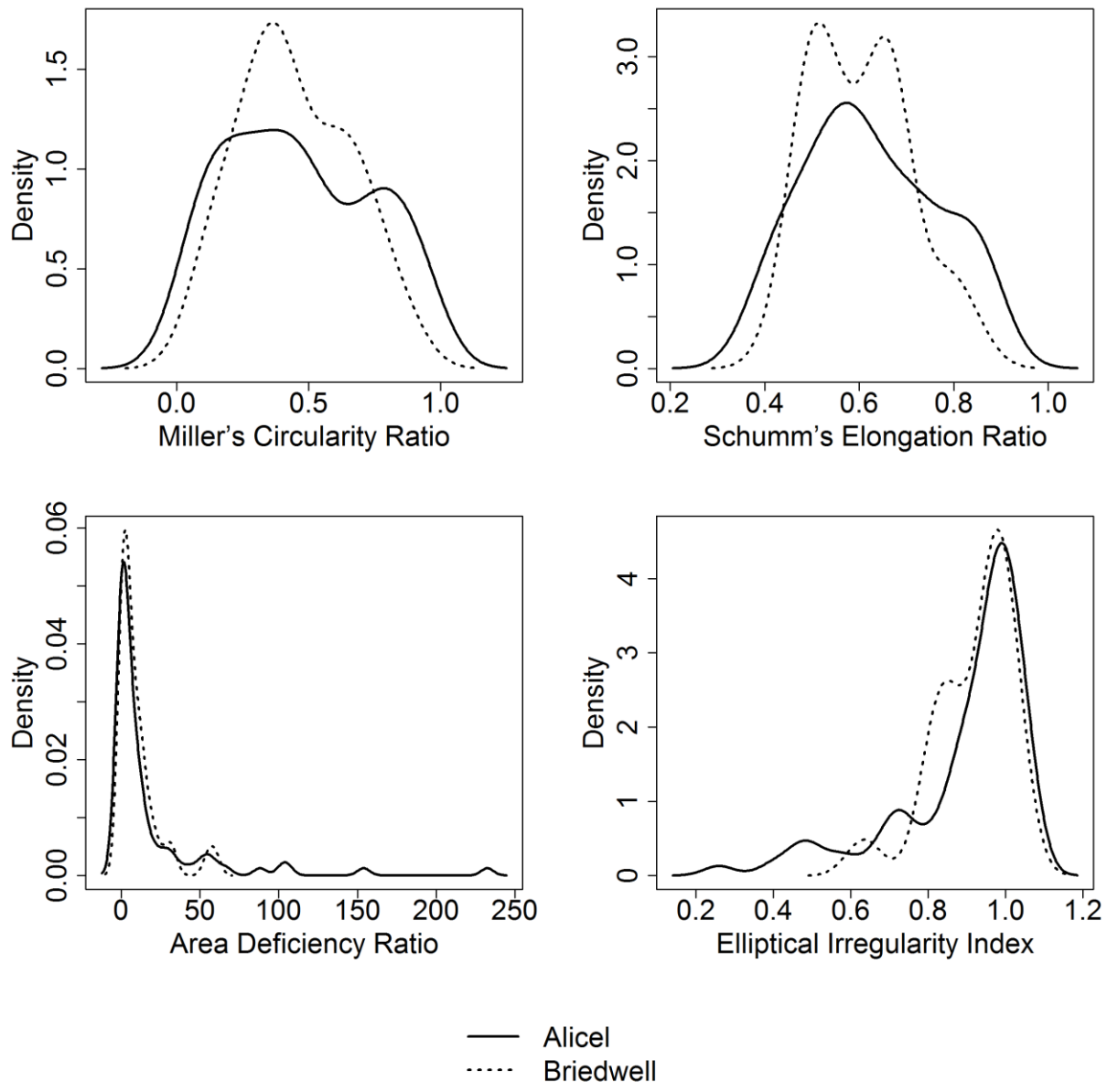


Figure 32. The kernel density estimation of selected shape descriptors for the two map units in Group 8 of Part II. The dominant soil series in each map unit are Alicel (solid line) and Briedwell (dotted line) respectively. These two soil series were considered to have different parent material based on the SSURGO database. For more information, see Table 2 and 3. All delineations of each soil map unit were used for analysis.

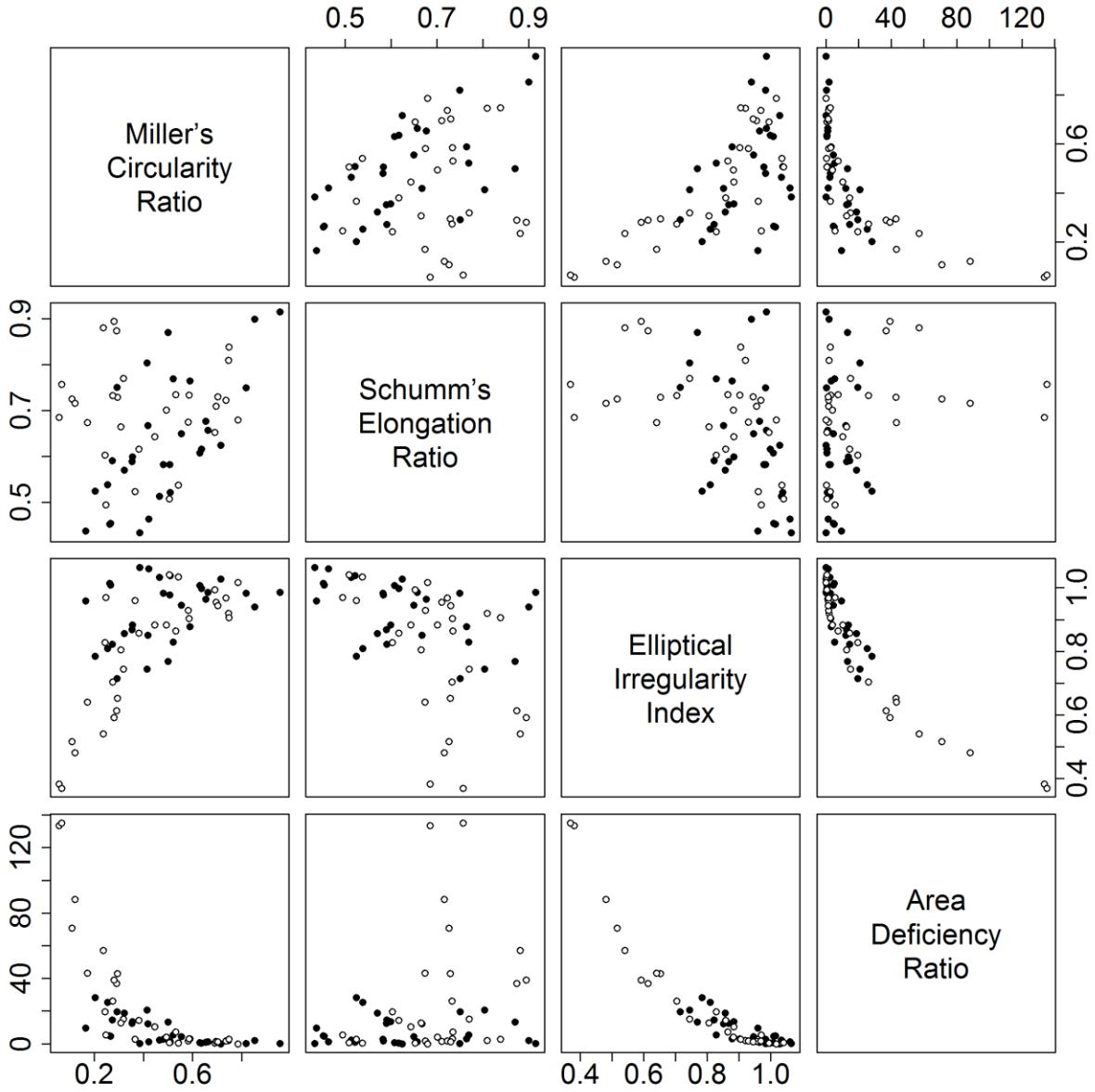


Figure 33. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 4 of Part II. The dominant soil series in each map unit are Hazelair (black dots) and Jory (white dots) respectively. These two soil series are considered to have different parent material based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test also indicated the distributions of selected shape descriptors were not significantly different.

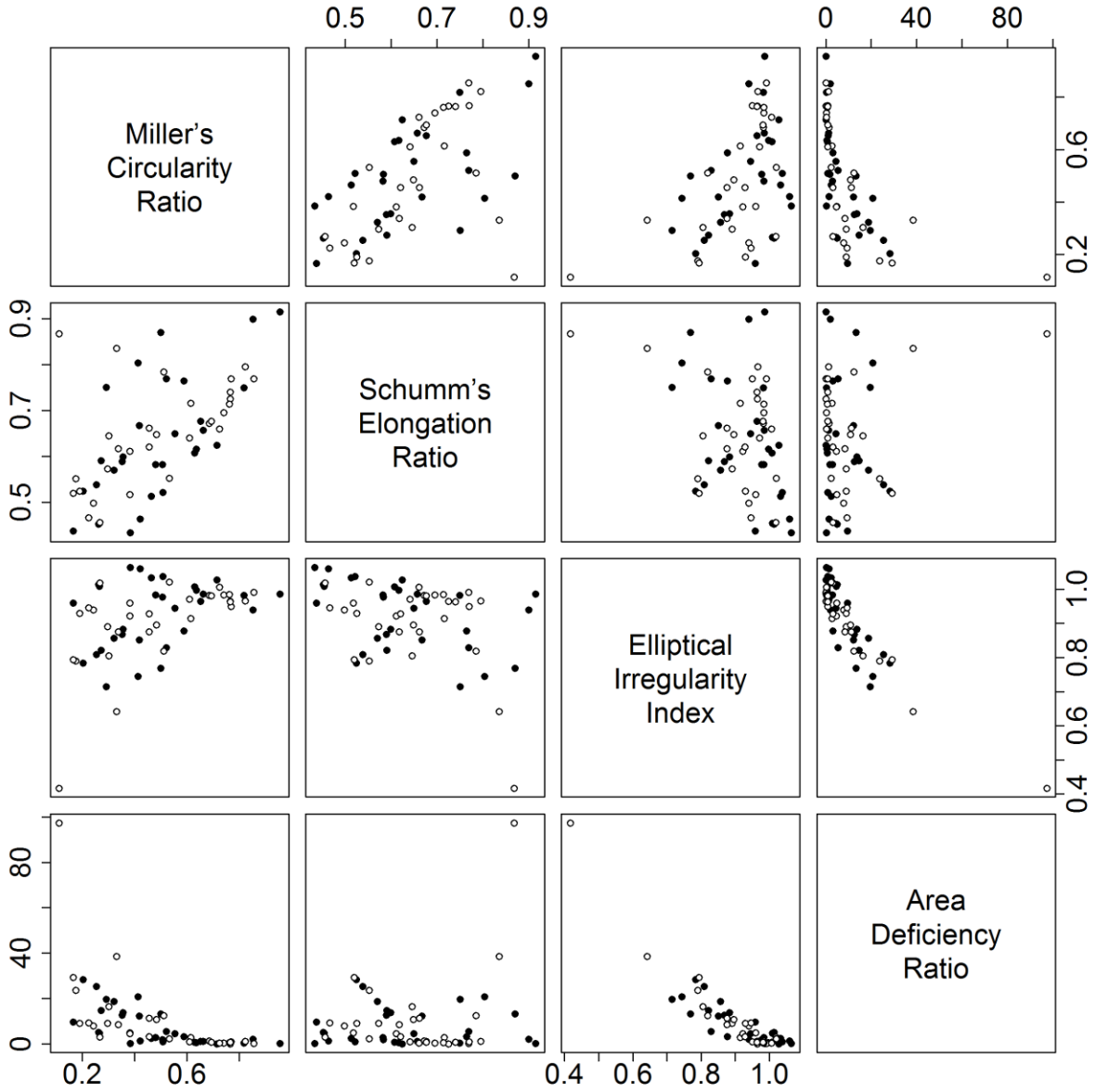


Figure 34. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 5 of Part II. The dominant soil series in each map unit are Hazelair (black dots) and Melbourne (white dots) respectively. These two soil series are considered to have different parent material based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test indicated the distributions of selected shape descriptors were not significantly different.

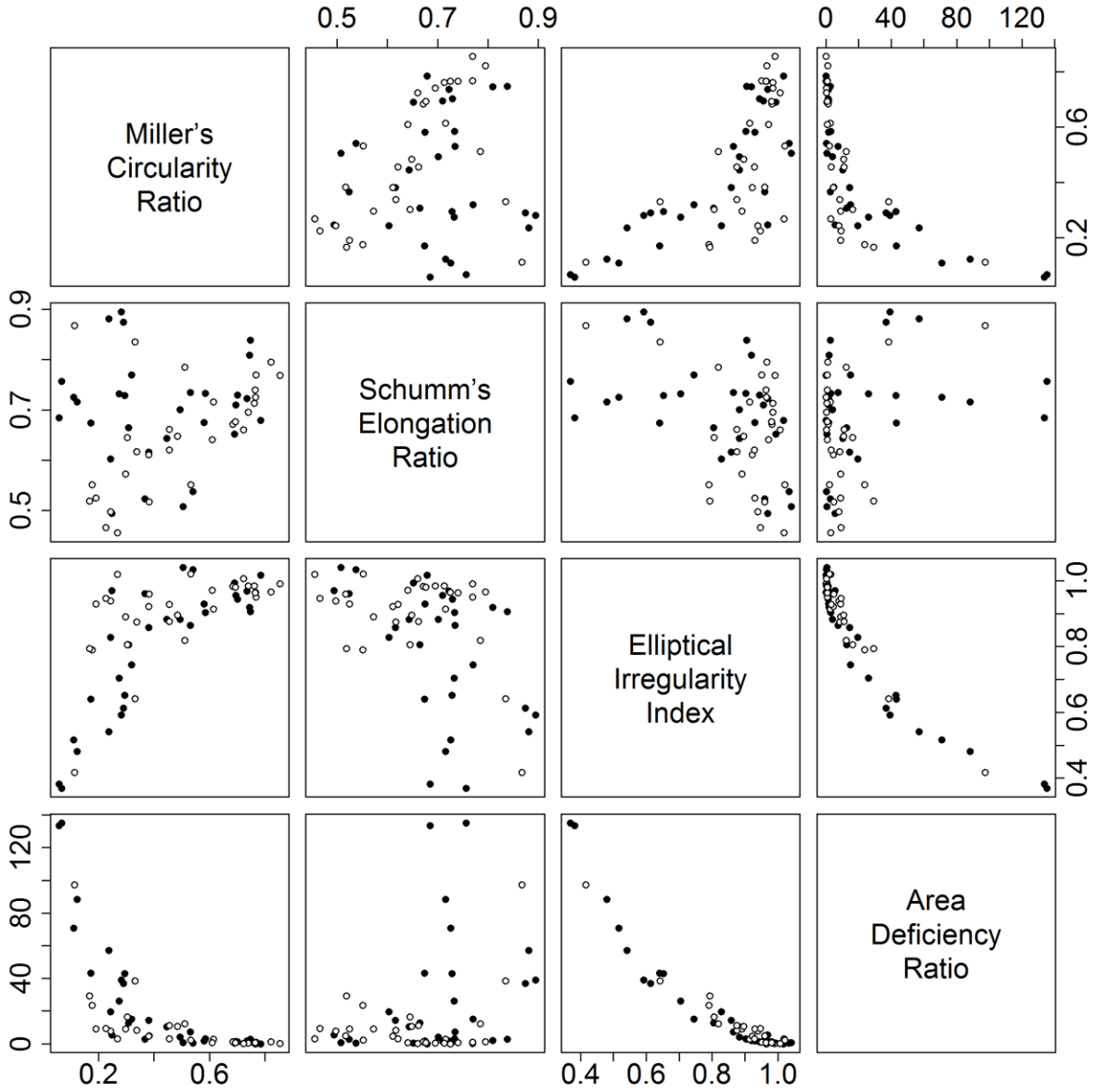


Figure 35. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 6 of Part II. The dominant soil series in each map unit are Jory (black dots) and Melbourne (white dots) respectively. These two soil series are considered to have different parent material based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test indicated the distributions of selected shape descriptors were not significantly different.

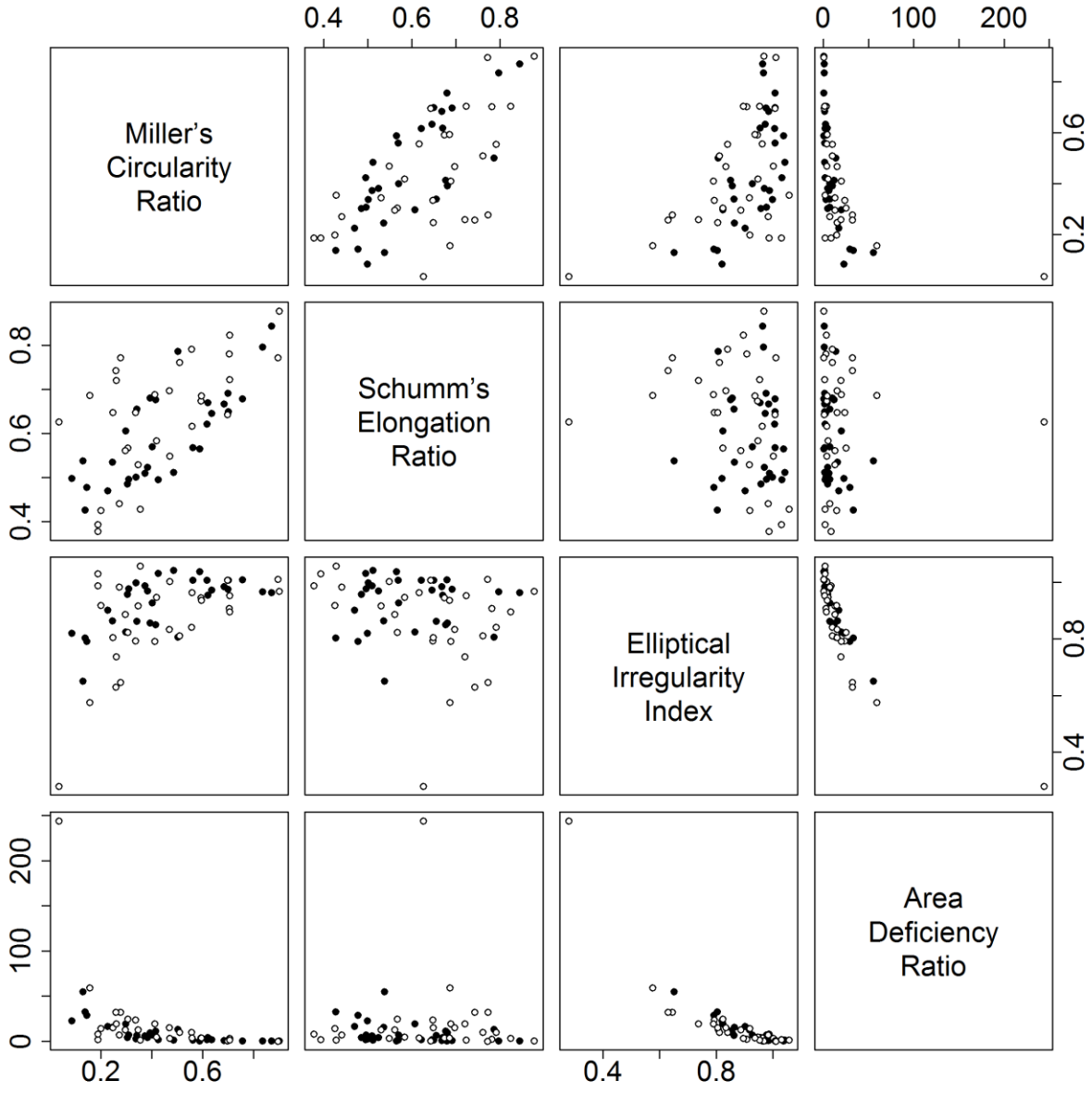


Figure 36. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 7 of Part II. The dominant soil series in each map unit are Briedwell (black dots) and Imbler (white dots) respectively. These two soil series are considered to have different parent material based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test indicated the distributions of selected shape descriptors were not significantly different.

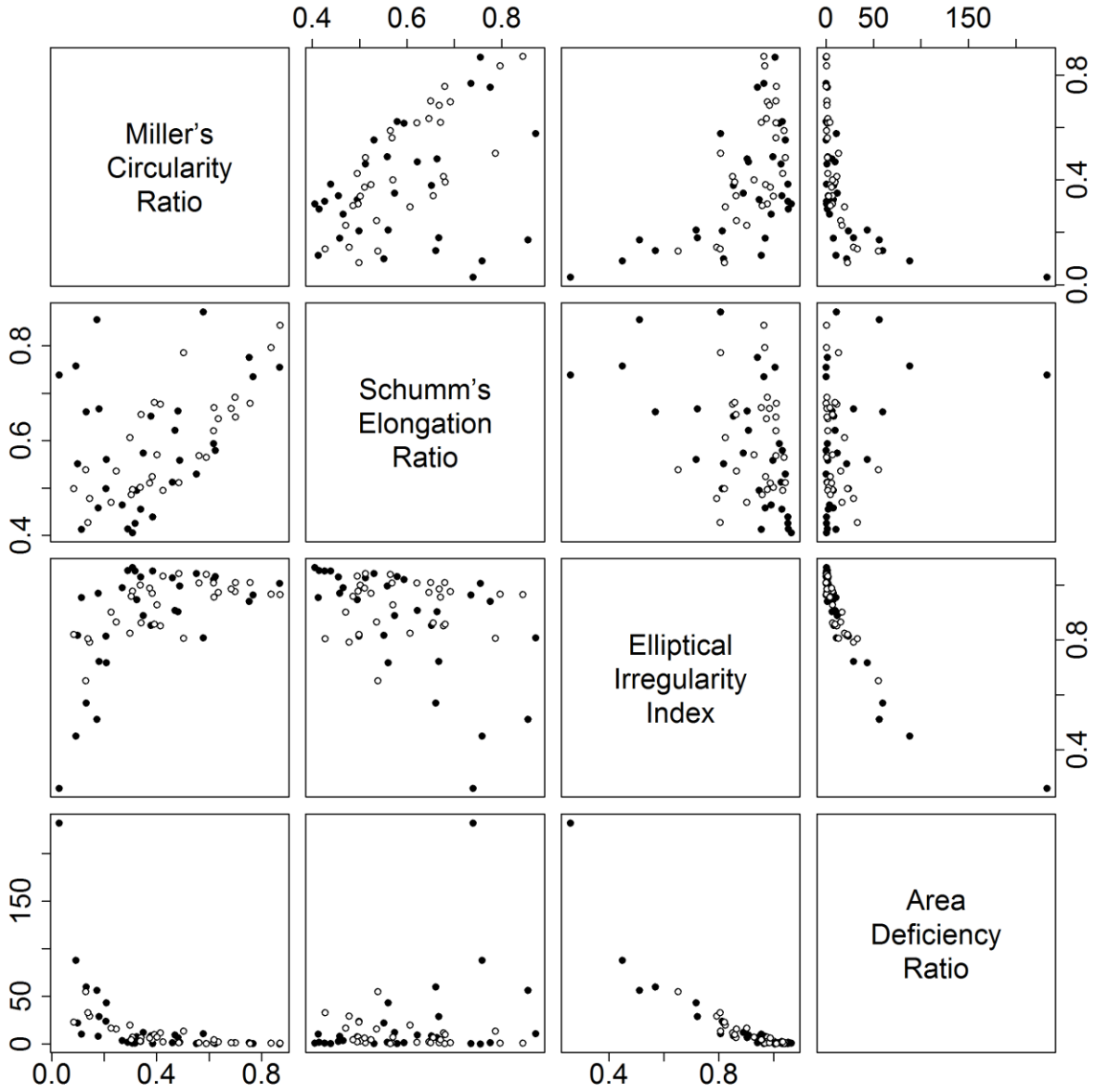


Figure 37. Scatterplot displays the linear combination of selected shape descriptors for the two map units in Group 8 of Part II. The dominant soil series in each map unit are Alicel (black dots) and Briedwell (white dots) respectively. These two soil series are considered to have different parent material based on the information in SSURGO database. The upper panel and the lower panel are mirror images along the diagonal. Thirty delineations of each soil map unit were used for analysis. The separation of these two soil map units was barely achieved by using any combination of these selected shape descriptors, and the results of the two-sample Kolmogorov-Smirnov test indicated the distributions of selected shape descriptors were not significantly different.

Table 1. Grouped soil series in Part I and their taxonomic classes. Each group contained two soil series. Based on the taxonomic classes, the soil series in group A were considered to be similar. Those in group B were considered to be very different. The soil series in group C contain different organic matter content. The soil series in group D have different soil temperature regimes, and those in group E possess different soil moisture regimes.

Group	Taxonomic class	Map unit name	Landform
A. Similar	Very-fine, smectitic, mesic Xeric Endoaquerts	Bashaw clay	alluvial fans, flood plains, terraces
	Fine, smectitic, mesic Cumulic Vertic Endoaquolls	Chehalem silt clay loam, 3-12 percent slopes	alluvial fans
B. Very different	Fine, kaolinitic, mesic Xeric Palehumults	Salkum silt clay loam, 8-16 percent slopes	terraces
	Medial over sandy or sandy-skeletal, mixed, mesic Typic Melanoxerands	Sifton gravelly loam	terraces
C. Different organic matter content	Fine, mixed, active, mesic Rhodic Paleudults	Cazadero silt clay loam, 0-7 percent slopes	terraces
	Fine-silty, mixed, superactive, mesic Pachic Ultic Argixerolls	Willamette silt loam, 12-20 percent slopes	valleys, terraces
D. Different soil temperature regimes	Medial, amorphic Typic Haplocryands	Keel gravelly silt loam, 25-45 percent slopes	mountains
	Medial over sandy or sandy-skeletal, mixed, mesic Typic Melanoxerands	Sifton gravelly loam	terraces
E. Different soil moisture regimes	Fine-silty, mixed, superactive, mesic Humic Fragiaquepts	Delena silt loam, 3-12 percent slopes	swales, terraces
	Fine-silty, mixed, superactive, mesic Humic Fragixerepts	Powell silt loam, 0-8 percent slopes	terraces

Table 2. Grouped soil series in Part II and their temperature regimes (T), moisture regimes (M), types of landforms, and kinds of parent material. Each group contained two soil series. Within groups 1 through 3, the landforms and parent material of grouped series were similar while T and/or M were different. Within groups 4 through 8, the landforms, T, and M of grouped series were similar while parent materials were different.

Group	Series	T	M	Landform	Parent Material
1. Different soil temperature regimes	Looking-glass	Mesic	Xeric	Structural Benches	Loess with minor amounts of volcanic ash over colluvium and residuum derived from basalt
	Cowsly	Frigid	Xeric	Plateaus	Loess with a minor influence of volcanic ash over colluvium and residuum derived from basalt
2. Different soil moisture regimes	Clackamas	Mesic	Aquic	Terraces	Mixed gravelly alluvium
	Kimberly	Mesic	Aridic	Terraces	Alluvium
3. Different soil temperature and moisture regimes	Bakeoven	Mesic	Aridic	Plateaus	Colluvium and residuum weathered from basalt
	Bocker	Frigid	Xeric	Plateaus	Loess and colluvium derived from basalt
4. Different parent material	Hazelair	Mesic	Xeric	Hills	Residuum weathered from sandstone and shale
	Jory	Mesic	Xeric	Hills	Colluvium derived from tuff and basalt
5. Different parent material	Melbourne	Mesic	Xeric	Hills	Residuum and colluvium derived from sedimentary rock
	Hazelair	Mesic	Xeric	Hills	Residuum weathered from sandstone and shale
6. Different parent material	Jory	Mesic	Xeric	Hills	Colluvium derived from tuff and basalt
	Melbourne	Mesic	Xeric	Hills	Residuum and colluvium derived from sedimentary rock
7. Different parent material	Imbler	Mesic	Xeric	Terraces	Mixed eolian deposits derived from basalt and andesite
	Briedwell	Mesic	Xeric	Terraces	Old, mixed, gravelly alluvium over irregular substratum of siltstone
8. Different parent material	Alicel	Mesic	Xeric	Terraces	Mixed eolian deposits derived from basalt and andesite
	Briedwell	Mesic	Xeric	Terraces	Old, mixed, gravelly alluvium over irregular substratum of siltstone

Table 3. Grouped soil series in Part II and their county code, total number of delineations, and taxonomic classes. Each group contained two soil series. Within groups 1 through 3, the landforms and parent material of grouped series were similar while T and/or M were different. Within groups 4 through 8, the landforms, T, and M of grouped series were similar while parent materials were different.

Group	Series	County Code	Total Number of Delineations	Taxonomic Class
1. Different soil temperature regimes	Looking-glass	OR670	32	Fine, smectitic, mesic Xerertic Argialbolls
	Cowsly	OR670	206	Fine, smectitic, frigid Xerertic Argialbolls
2. Different soil moisture regimes	Clackamas	OR610	55	Fine-loamy, mixed, superactive, mesic Typic Argiaquolls
	Kimberly	OR641	67	Coarse-loamy, mixed, superactive, mesic Torrifluventic Haploxerolls
3. Different soil temperature and moisture regimes	Bakeoven	OR674	45	Loamy-skeletal, mixed, superactive, mesic Aridic Lithic Haploxerolls
	Bocker	OR670	68	Loamy-skeletal, mixed, superactive, frigid Lithic Haploxerolls
4. Different parent material	Hazelair	OR643	35	Very-fine, smectitic, mesic Vertic Haploxerolls
	Jory	OR643	164	Fine, mixed, active, mesic Xeric Palehumults
5. Different parent material	Melbourne	OR067	59	Fine, mixed, superactive, mesic Ultic Palexeralfs
	Hazelair	OR643	35	Very-fine, smectitic, mesic Vertic Haploxerolls
6. Different parent material	Jory	OR643	164	Fine, mixed, active, mesic Xeric Palehumults
	Melbourne	OR067	59	Fine, mixed, superactive, mesic Ultic Palexeralfs
7. Different parent material	Imbler	OR625	31	Coarse-loamy, mixed, superactive, mesic Pachic Haploxerolls
	Briedwell	OR053	36	Loamy-skeletal, mixed, superactive, mesic Ultic Haploxerolls
8. Different parent material	Alicel	OR625	78	Fine-loamy, mixed, superactive, mesic Pachic Haploxerolls
	Briedwell	OR053	36	Loamy-skeletal, mixed, superactive, mesic Ultic Haploxerolls

Table 4. The p values of two-sample Kolmogorov-Smirnov tests for Part I. The tests were conducted between grouped soil series, using the distributions of each shape descriptor. A p value smaller than 0.05 indicates the difference between two distributions is significant and these values are noted with *; and a p value smaller than 0.01 indicates the difference between two distributions is very significant and these values are noted with **. The information of paired soil series in each group can be found in Table 1.

Group	p Values for Selected 2-D Shape Descriptors			
	Miller's Circularity Ratio	Schumm's Elongation Ratio	Area Deficiency Ratio	Full Procrustes Distance
A	0.039*	0.272	0.179	0.272
B	4.93e-07**	0.549	2.46e-04**	1.45e-07**
C	3.80e-05**	2.84e-03**	0.717	0.179
D	1.32e-03**	2.84e-03**	2.46e-04**	1.58e-06**
E	0.112	0.272	0.869	0.039*

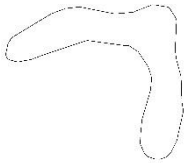
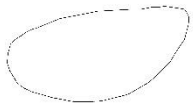
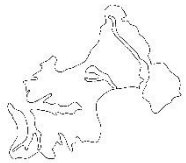
Table 5. The p values of two-sample Kolmogorov-Smirnov tests for Part II. The tests were conducted between grouped soil series, using the distributions of each shape descriptor. A p value smaller than 0.05 indicates the difference between two distributions is significant and these values are noted with *; and a p value smaller than 0.01 indicates the difference between two distributions is very significant and these values are noted with **. The information of paired soil series in each group can be found in Tables 2 and 3.

Group	p Values for Selected 2-D Shape Descriptors			
	Miller's Circularity Ratio	Schumm's Elongation Ratio	Elliptical Irregularity Index	Area Deficiency Ratio
1	0.247	0.439	0.286	0.046*
2	8.70e-05**	0.022*	0.090	0.017*
3	4.74e-03**	6.49e-07**	8.48e-05**	2.32e-03**
4	0.735	2.74e-03**	0.047*	0.195
5	0.494	0.122	0.075	0.853
6	0.980	0.225	0.058	0.097
7	0.677	0.104	0.415	0.644
8	0.402	0.149	0.583	0.058

Table 6. The p values of two-sample Kolmogorov-Smirnov tests for the same map unit from different counties. The tests were conducted between grouped counties, using the distributions of each shape descriptor. A p value smaller than 0.05 indicates the difference between two distributions is significant and these values are noted with *; and a p value smaller than 0.01 indicates the difference between two distributions is very significant and these values are noted with **. This table showed that the same map unit taken from different survey areas had differences in the frequency of shape archetypes.

Map Unit	County Code	p Values for Selected 2-D Shape Descriptors			
		Miller's Circularity Ratio	Schumm's Elongation Ratio	Area Deficiency Ratio	Elliptical Irregularity Index
Abiqua silty clay loam, 0 to 3 percent slopes	OR637vsOR639	0.032*	8.62e-03**	0.065	0.120
	OR637vsOR053	0.133	0.046*	0.039*	0.065
	OR637vsOR643	2.29e-03**	0.662	0.01*	2.06e-03**
	OR639vsOR053	3.77e-03**	1.48e-04**	0.011*	0.026*
	OR639vsOR643	0.499	0.142	9.67e-04**	2.72e-03**
	OR053vsOR643	4.162e-05**	0.121	0.131	0.136

Table 7. Selected delineations in Jory soil series. This table took some delineations of Jory soil series from the same map unit, and lists some of the quantitative properties of each delineation. ID is the FID number of the delineation in the digital shape file. Scale is the scale when the delineation was being exported as an image from ArcMap. C is Miller's circularity ratio which ranges from 0 to 1; a value close to 1 indicates a disk-like shape. Elo is Schumm's elongation ratio which also ranges from 0 to 1; a value close to 1 indicates a stripe-like shape. Ell is Elliptical irregularity index which normally ranges from 0 to 1; a value close to 1 indicates an ellipse-like shape. However, because the equation that is used to calculate the perimeter of ellipse is merely an estimation, the index value will be larger than 1. D is area deficiency ratio; a small value indicates a simple shape while a large value indicates a complex shape. From this table, one can see that within one soil map unit, the delineated shape of soils appeared to be random, as the shape varies from very simple to very complex. However, analysis on their morphometric shape descriptors indicated that the frequency of each shape varies from map unit to map unit.

Delineations	Quantitative Descriptions	Delineations	Quantitative Descriptions
	ID: 0		ID: 13
	Scale: 1:2,847		Scale: 1:1,538
	C: 0.366		C: 0.319
	Elo: 0.524		Elo: 0.770
	Ell: 0.960		Ell: 0.744
	D: 2.683		D: 15.076
	ID: 149		ID: 100
	Scale: 1:2,574		Scale: 1:1,200
	C: 0.375		C: 0.387
	Elo: 0.460		Elo: 0.439
	Ell: 1.031		Ell: 1.059
	D: 1.182		D: 0.223
	ID: 24		ID: 152
	Scale: 1:1,883		Scale: 1:769
	C: 0.785		C: 0.907
	Elo: 0.679		Elo: 0.798
	Ell: 1.017		Ell: 1.001
	D: 0.050		D: 0.029
	ID: 140		ID: 93
	Scale: 1:10,461		Scale: 1:14,775
	C: 0.123		C: 0.072
	Elo: 0.698		Elo: 0.769
	Ell: 0.494		Ell: 0.335
	D: 88.242		D: 160.630