

Classification of Geologic Units on Ganiki Planitia Quadrangle
(V14), Venus using Statistical Clustering Methods

Joey Richards

April 2005

Contents

1	Introduction to Problem	4
1.1	Venus Geology	4
1.1.1	Magellan Spacecraft	4
1.1.2	Ganiki Planitia Quadrangle (V14)	4
1.1.3	Geologic Goals	5
1.2	Mathematical Problem	5
1.2.1	Formulation of Problem	5
1.2.2	Clustering	7
2	Methods	9
2.1	Mixture Models	9
2.1.1	Multivariate Normal Distribution	10
2.1.2	Complete Data Likelihood	11
2.2	The EM Algorithm	13
2.2.1	EM Initialization	13
2.2.2	M-step (Maximization)	13
2.2.3	E-step (Expectation)	16
2.3	Properties of the EM Algorithm	18
2.3.1	General form of EM algorithm	18
2.3.2	$L(\Phi)$ is Non-decreasing for each Iteration	18
2.3.3	Convergence of the EM Algorithm	20
3	Results	22
3.1	Data Collection	22
3.2	Clustering Results	25
3.2.1	Relocation of Single Units	25
3.2.2	Removal of Single Units	27
3.3	More Clustering Results	29
3.3.1	Background Plains	29
3.3.2	A Less Constrained Model	30

4	Weighting	34
4.1	Weighting by Unit Area	34
4.1.1	Motivation	34
4.1.2	Methods	35
4.1.3	Choosing the Weighting Vector	38
4.1.4	Results	39
4.2	Weighting by Variable	44
4.2.1	Motivation	44
4.2.2	Methods	44
4.2.3	Results	45
5	Conclusions	46

Abstract

In this project, we analyzed data collected by NASA's Venus orbiter Magellan from the Ganiki Planitia Quadrangle (V14) to determine an optimal way to allocate geologic units into groups corresponding to unit type. Previously, a team led by Dr. Eric Grosfils created a geologic map of the region using Magellan radar images. However, they performed their analysis without analyzing the numerical data encoded in these images and without considering the physical property data sets of slope, surface emissivity, and surface reflectivity. We extracted the quantitative data corresponding to radar backscatter, elevation, slope, surface emissivity, and surface reflectivity from Magellan images. Then, using the existing geologic map as a baseline, we employed mixture models and the Expectation-Maximization (EM) algorithm to devise an optimal geologic map based on the data and to identify units whose data fits more closely with a different unit type than what was assigned by the mapping. We also created modified versions of the EM algorithm, considering issues such as the area of each geologic unit and the importance of different variables. Results showed that most units were classified the same way as specified by the original geologic map, while a handful of units were consistently assigned to different groups. We conclude that these areas should be more thoroughly examined geologically.

Chapter 1

Introduction to Problem

1.1 Venus Geology

1.1.1 Magellan Spacecraft

On May 4, 1989 the National Aeronautics and Space Administration (NASA) launched the Magellan Spacecraft to study the planet Venus. From August 10, 1990 until September 14, 1992 Magellan orbited Earth's sister planet, radar-mapping 98% of its surface at resolutions that were ten times better than any previous mapping of the planet [6]. Additionally, Magellan took detailed measurements of surface properties including elevation, slope, and reflectivity.

The influx of data from the Magellan mission has allowed planetary geologists to study surface-shaping processes on Venus at a remarkable depth. Scientists have discovered that geology on Venus is dominated by tectonics, volcanism, and impact cratering with little wind erosion. One important finding from the Magellan data is that the surface of the planet is only 500 million years old [6]. Venus formed around 4.6 billion years ago with the rest of the planets, so some sort of volcanic resurfacing event or events must have occurred approximately 500 million years ago. Subsequently, competing theories have arisen: (1) that the resurfacing of Venus occurred quickly in one planet-wide event, and (2) that the resurfacing happened in several smaller, localized events.

1.1.2 Ganiki Planitia Quadrangle (V14)

Ganiki Planitia quadrangle V14 (180°-210° E, 25°-50° N) is a section of Venus that was radar-mapped by the Magellan Spacecraft. It is a third the size of the United States. Situated between regions where extensive tectonic and volcanic activity has occurred in the past, Ganiki Planitia consists of volcanically-formed plains partially covered with tectonic, impact, and young volcanic features. Because such varying types of geology exist on Ganiki Planitia, studying the region may clarify the ways in which different geologic processes on Venus interact with each other and might hold the key to resolving the problem of the resurfacing of Venus 500 million years ago [5].

Over the past three years, Dr. Eric Grosfils, Assistant Professor in the Department of Geology at Pomona College has headed a team of undergraduate students to investigate the Magellan data from Ganiki Planitia. They have undertaken the project in attempts to understand the geologic processes on Venus more completely. The research is sponsored by a grant from NASA's Planetary Geology and Geophysics Program.

1.1.3 Geologic Goals

With the Ganiki Planitia data, Dr. Grosfils and his geology students have been investigating an array of geologic issues. First, they have analyzed the volcanic features on Ganiki Planitia in attempts to conclude whether or not there are systematic changes in the type of volcanism on the region over time. Secondly, they have looked at tectonic activity to determine how regional stresses are distributed and how they have changed. Lastly, they are in the process of constructing a region-wide chronology of geologic events in order to place time constraints on the resurfacing of the region to provide insight into the problem of global resurfacing [5].

However, none of these issues can be investigated without first creating a geologic map of Ganiki Planitia. A geologic map is a division of a region into many smaller units based on a mapper's interpretation of what the main features of the region are. Each unit is subsequently labeled as a certain type of geologic structure based on what the unit is. A completed geologic map looks like a colored jigsaw puzzle, with each piece representing a unit and each color representing a unit type. Researchers cannot begin to consider the relative timing of geologic events until a geologic map is constructed. Therefore, it is extremely important for Dr. Grosfils' research group that a complete and accurate geologic map of the Ganiki Planitia region be made. Additionally, one of the geologists' goals is to submit a final map of the region to the U.S. Geological Survey for publication.

Throughout the past three years, the group led by Dr. Grosfils has created a map of the Ganiki Planitia region (Figure 1.1). The mapping consists of 200 units divided into 18 groups (e.g. craters, volcanic flows, radar-dark plains, radar-light plains). To construct the map, the research team studied the Magellan radar image of the region via the computer imaging program ArcGis. The team also looked at stereo images, which combine elevation information and the radar image to create a three-dimensional overhead picture of the region. The researchers did not explicitly use any of the numerical information that was retrieved by Magellan to make their map.

1.2 Mathematical Problem

1.2.1 Formulation of Problem

Although Dr. Grosfils and his research team have spent three years studying the images of Ganiki Planitia in the program ArcGis, they have not analyzed the numerical data encoded in these images. Each of the more than 150,000,000 pixels that make up the radar image of the Ganiki Planitia region has associated with it a numerical value. Also, each of the 350,000 pixels of each of the four physical properties: elevation, slope, emissivity, and reflectivity,



Figure 1.1: Geologic map of V14 created by Dr. Grosfils' team. Each section represents a geologic unit and each color represents a type of unit. There are 200 units and 18 unit types.

has a numerical value. This information is readily available and may reveal much about the region. Yet, it has not been utilized by Dr. Grosfils' research team.

The goal of our project is to devise an allocation of the 200 units into a number of groups based on the numerical data of the radar backscatter and the four physical property variables. Performing this classification will allow us to see whether or not the numerical data of backscatter and elevation are consistent with the geologists' allocation of units based on the images produced by these two data sets in ArcGis. It will also provide insight into whether or not the existing geologic map is consistent with the other three physical properties that were not used by Grosfils' team.

However, it is important to keep in mind that many properties of these geologic units do not appear in the data and can only be seen by performing rigorous geologic analysis of the images on ArcGis. Much of the region is an intricate web of tectonically-formed

lineaments and volcanic flows. Failing to take into account the geologists' interpretation means depriving oneself of a massive amount of information that is not detectable in the data alone. Therefore, throughout this project we depended heavily on the original geologic mapping and kept close correspondence with Dr. Grosfils and his research team. Three years have been spent devising the original geologic map. Therefore the ultimate goal of this project is not to propose a massive overhaul of the geologists' work but rather to expose areas where the data sets are telling us that something is happening that does not coincide with the geologists' interpretation.

1.2.2 Clustering

We have data from 200 different geologic units and wish to classify each unit as belonging to one unit type. The information with which we will perform this classification is the mean, standard deviation, and median of pixel values within each unit boundary for each of the five variables. To perform this classification, we will use cluster analysis. Clustering is the allocation of n data points (in our case geologic units) into G groups (unit types) based on k variables. There are several types of clustering methods available, ranging from heuristic procedures to formal, model-based techniques.

Hierarchical clustering begins with a partitioning of the data points and either merges groups (agglomerative) or divides groups (divisive) based on some criterion. For example, one agglomerative clustering method is nearest neighbor (single link), where in each iteration the two groups that are closest to one another are merged. The stopping point of the algorithm is highly arbitrary, as the investigator must decide on an optimal number of clusters or proceed until either the data is in one cluster (agglomerative) or each data point is in its own cluster (divisive). Another problem with hierarchical clustering is that once a merging or splitting is made, it can never be undone, even if in later steps it appears that it should be. Hierarchical clustering is only stepwise optimal, and will not necessarily lead to an optimal final solution. Hierarchical techniques are also computationally intensive [3].

Relocation (hill-climbing) clustering methods begin with some initial partition of the data into groups. Then, the change in some clustering criterion is calculated for every possible move of a data point to a different group. The change which leads to the greatest improvement in the value of the clustering criterion is made, and then the process is repeated until an optimal value of the clustering criterion is reached. One widely-used relocation clustering method is k-means. In each iteration step in k-means, the centroid of every cluster is calculated. The clustering criterion to be minimized is the distance from the data points in each cluster to that cluster's centroid. Drawbacks to the k-means method is the large amount of computation necessary, especially when the number of data points is large. Also, the selection of the clustering criterion is somewhat arbitrary, and different results are possible for different criteria [3].

In this project, we will cluster with mixture models. Mixture model clustering is a type of relocation clustering technique where it is assumed that each cluster has a unique probability density function. The optimal clustering is obtained by maximizing a data likelihood function based on these probability distributions[4]. Mixture models allow us to

talk about the classification of data points probabilistically. Also, there are algorithms that can perform this maximization very quickly. In this paper we will explore these methods and apply them to the Nemesis Tessera problem. In Chapter 2 we provide a detailed description and justification of the model and methods used. In Chapter 3 we use these methods to analyze our geologic units. Chapter 4 presents a model that takes into account the area of each unit and a model that includes weights on the importance of each variable. We conclude with Chapter 5, where we present our final results to the geologists.

Chapter 2

Methods

2.1 Mixture Models

The idea behind finite mixture models is that our data x_1, x_2, \dots, x_n are generated from multiple underlying probability distributions, with each distribution corresponding to a different unit type. Because each unit type has a unique set of characteristics that defines it, we expect that each distribution will be different and hence have a distinct set of parameters. With mixture models, each data point, x_i , is generated from a population that is a mixture of G subpopulations in proportions τ_k ($k = 1, \dots, G$). The probability density function (p.d.f.) of an observation x_i is

$$f(x_i|\tau, \theta) = \sum_{k=1}^G \tau_k f_k(x_i|\theta_k) \quad (2.1)$$

which is simply the weighted sum of the probability density functions, $f_k(x_i|\theta_k)$, of each of the G groups, where θ_k represents the vector of parameters that defines the k^{th} density function [8].

We assume that our data are independently and identically distributed (i.i.d.) with common distribution f . Independence is an appropriate assumption in our case because each geologic unit is a separate entity whose physical characteristics are for the most part not affected by other units. By the independence of our data, we can devise a likelihood function for the set of parameters based on the data by simply multiplying the densities for each data point. This likelihood is

$$L_{MIX}(\theta_1, \dots, \theta_G, \tau_1, \dots, \tau_G|\mathbf{x}) = \prod_{i=1}^n f(x_i|\tau, \theta) = \prod_{i=1}^n \sum_{k=1}^G \tau_k f_k(x_i|\theta_k) \quad (2.2)$$

which we maximize in order to determine the most likely set of parameters based on the data [8]. These parameters will define an optimal set of density functions based on the data.

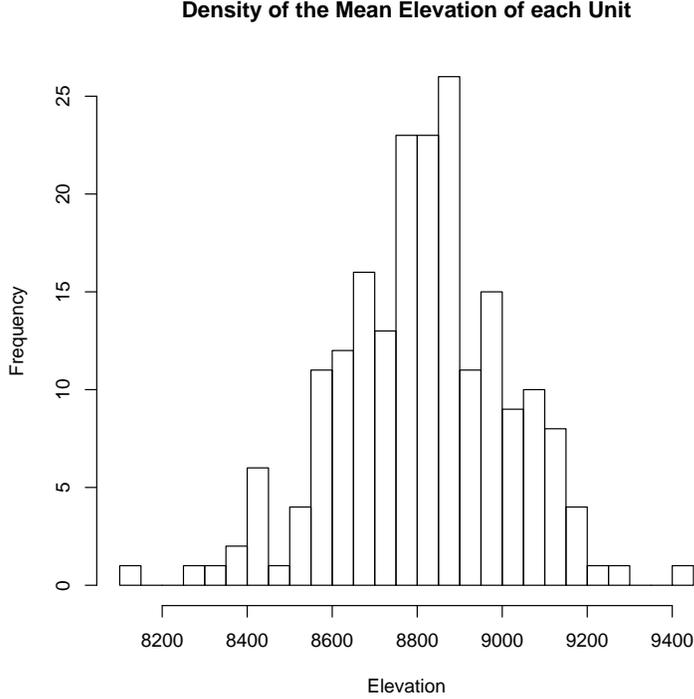


Figure 2.1: Distribution of the mean elevations for each unit. The data appear to follow a normal distribution. Because the mean elevation of each unit follows a normal distribution, it is reasonable the elevations for each group will be distributed normally.

2.1.1 Multivariate Normal Distribution

In our model, we assume that the underlying distribution for each cluster is the multivariate normal distribution. In other words, we assume that the set of data points belonging to each cluster will be distributed normally with respect to each variable. This is a valid assumption because this model has been used successfully in a great range of scientific fields [4] and because our data on the whole appears to be distributed normally with respect to each variable (e.g., see Figure 2.1).

We can now define our density functions for each of the G groups. For group k , the density of the i^{th} data point is the density of the multivariate normal distribution,

$$f_k(x_i|\mu_k, \Sigma_k) = \frac{\exp\left(-\frac{1}{2}(x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k)\right)}{\left(\frac{2}{\pi}\right)^{\frac{v}{2}} |\Sigma_k^{\frac{1}{2}}|} \quad (2.3)$$

where μ_k , the mean vector, and Σ_k , the covariance matrix, completely define the density function for group k . The dimension of each μ_k is $v \times 1$, where v is the number of variables.

The dimension of each Σ_k is $v \times v$.

Each of the G probability density functions defines a cluster. The multivariate normal distribution creates v -dimensional ellipsoidal clusters. For group k , μ_k determines the center of the cluster and Σ_k determines other geometric features of the cluster, such as shape, volume, and orientation.

The eigenvalue decomposition of the k^{th} covariance matrix can be written

$$\Sigma_k = \lambda_k D_k A_k D_k^T \quad (2.4)$$

where λ_k is a scalar, D_k is the orthogonal matrix of the eigenvectors of Σ_k , and A_k is a diagonal matrix whose entries are proportional to the eigenvalues of Σ_k . This decomposition is important because each element of the decomposition determines a property of the cluster. For cluster k , the value of λ_k determines the volume of the cluster, A_k specifies the cluster's shape, and D_k determines the cluster's orientation [4]. It is possible to let the values of some or all of these parameters either vary between groups or be fixed across groups. For example, if we forced the value of A to be the same for each group and let the other parameters vary between groups, our clusters would all have the same shape but would vary in volume and orientation.

2.1.2 Complete Data Likelihood

When clustering with mixture models, we wish to put the n data points into G groups by maximizing a data likelihood function. Thus far we do not have a variable that specifies to which group each data point belongs. Therefore, we have incomplete data. The complete data are (x_i, \mathbf{z}_i) where $\mathbf{z}_i = (z_{i1}, \dots, z_{iG})$ such that

$$z_{ik} = \begin{cases} 1 & \text{if } x_i \text{ belongs to group } k \\ 0 & \text{else} \end{cases} \quad (2.5)$$

Recall that for each k , τ_k is the mixture proportion of the probability density function for group k . A group will contribute to the overall density function proportionally to the number of data points in the group. Therefore, τ_k is the proportion of data points belonging to group k . Alternatively, τ_k can be thought of as the probability that a randomly selected data point belongs to group k . Hence, for each data point, \mathbf{z}_i is distributed multinomially i.i.d. with one draw on G categories with probabilities τ_1, \dots, τ_G . [4] The probability density function of \mathbf{z}_i given τ is

$$f(\mathbf{z}_i | \tau) = \tau_1^{z_{i1}} \tau_2^{z_{i2}} \dots \tau_G^{z_{iG}} \quad (2.6)$$

which is the p.d.f. of a multinomial($1, \tau$) distribution.

Now we can write the probability density function of each data point, x_i , in terms of \mathbf{z}_i . Recall that in equation 2.1 we wrote the p.d.f. of x_i as a mixture of G distributions in proportions τ_k as if every data point had the same probability of belonging to group k for $k = 1, \dots, G$. However, with the variable \mathbf{z}_i , we now know to which group each data point

belongs. The p.d.f. of data point x_i is simply the p.d.f. of that point in the group to which it belongs. We can write this as

$$f(x_i|\mathbf{z}_i, \mu, \Sigma) = \prod_{k=1}^G f_k(x_i|\mu_k, \Sigma_k)^{z_{ik}} \quad (2.7)$$

where $f_k(x_i|\mu_k, \Sigma_k)^{z_{ik}}$ will be 1 if data point x_i is not in group k . Thus, if x_i is in group p , then $f(x_i|\mathbf{z}_i, \mu, \Sigma) = f_p(x_i|\mu_p, \Sigma_p)$, which is what we want.

We are now in a position to construct the complete data likelihood equation. Recall that we want to find the parameters (μ, Σ, τ) , of the clusters that best fit the data. Therefore we want to maximize the likelihood of our parameters based on the data and \mathbf{z}_i , our knowledge of the allocation of data points into groups. For data point x_i , the likelihood is $L(\mu, \Sigma, \tau|x_i, \mathbf{z}_i)$. By the definition of likelihood [10], we can express this as

$$L(\mu, \Sigma, \tau|x_i, \mathbf{z}_i) = f(x_i, \mathbf{z}_i|\mu, \Sigma, \tau) = f(x_i|\mu, \Sigma, \tau, \mathbf{z}_i)f(\mathbf{z}_i|\mu, \Sigma, \tau), \quad (2.8)$$

where the last equality comes from the definition of conditional probability. We also know that \mathbf{z}_i only depends on τ and is independent of μ and Σ . Therefore

$$f(\mathbf{z}_i|\mu, \Sigma, \tau) = f(\mathbf{z}_i|\tau). \quad (2.9)$$

Consequently, the likelihood for our i^{th} observation is

$$L(\mu, \Sigma, \tau|x_i, \mathbf{z}_i) = f(x_i|\mu, \Sigma, \tau, \mathbf{z}_i)f(\mathbf{z}_i|\tau). \quad (2.10)$$

Substituting equations 2.6 and 2.7 into this expression yields the likelihood equation for data point x_i ,

$$L(\mu, \Sigma, \tau|x_i, \mathbf{z}_i) = \prod_{k=1}^G f_k(x_i|\mu_k, \Sigma_k)^{z_{ik}} \tau_k^{z_{ik}}. \quad (2.11)$$

By the independence of our data, the complete data likelihood is

$$L(\mu, \Sigma, \tau|\mathbf{x}, \mathbf{z}) = \prod_{i=1}^n \prod_{k=1}^G f_k(x_i|\mu_k, \Sigma_k)^{z_{ik}} \tau_k^{z_{ik}} \quad (2.12)$$

and the complete data loglikelihood is

$$l(\mu, \Sigma, \tau|\mathbf{x}, \mathbf{z}) = \sum_{i=1}^n \sum_{k=1}^G z_{ik} [\log(\tau_k f_k(x_i|\mu_k, \Sigma_k))], \quad (2.13)$$

which is easier to work with.

2.2 The EM Algorithm

Our ultimate goal is to find the best allocation of data points into clusters based on the data. Maximizing our complete data loglikelihood (2.13) gives us the optimal set of clusters given \mathbf{z} . However, \mathbf{z} is unknown! In fact, it is the missing data that we are looking for in the first place. Therefore, we need a means of maximizing equation 2.13 in the face of missing data while at the same time solving for that missing data. The general approach to this kind of problem is the expectation-maximization (EM) algorithm [4]. The EM algorithm iterates between an M-step, which maximizes equation 2.13 with respect to the other parameters while holding \mathbf{z} constant, and an E-step, which calculates the expected value of \mathbf{z} given the parameters from the M-step.

The idea of the EM algorithm is that we have missing information and a likelihood function that we want to maximize with respect to a set of unknown parameters. Because of the missing information, we cannot do the maximization directly using maximum likelihood estimators. Therefore, we instead alternate between finding the expected value of our missing information based on the parameters and maximizing the expectation of the likelihood function with respect to the data and current values of the parameters.

2.2.1 EM Initialization

The EM algorithm must either begin with some initial value of \mathbf{z} and then proceed to the M-step, or begin with some initial values of the other parameters, μ , Σ , and τ and proceed to the E-step. It is much easier to initialize with \mathbf{z} : we can obtain a starting matrix \mathbf{z} by constructing some initial clustering of the data points into groups. The most common initialization is typically given by non-model based methods such as hierarchical clustering (see Section 1.1.2) [4]. In our problem, we have an existing allocation of units into groups that was created by geologists. We have confidence in this clustering because it was performed over a span of three years. Additionally, we do not want to lose the large amount of information that was used to construct the existing map. Therefore, we will initialize the EM algorithm with $\hat{\mathbf{z}}$ where

$$\hat{z}_{ik} = \begin{cases} 1 & \text{if the geologists say that unit } x_i \text{ belongs to group } k \\ 0 & \text{else} \end{cases} \quad (2.14)$$

so as to use the geologists' work as a starting point. Any differences that arise in our analysis, then, can be attributed to properties of the data.

2.2.2 M-step (Maximization)

After initialization, the EM algorithm proceeds to the M-step. Here, $\hat{\mathbf{z}}$ is held constant and equation 2.13 is maximized with respect to μ , Σ , and τ . The optimization is performed by means of maximum likelihood estimation. In this section I will derive the maximum likelihood estimator for τ and μ and discuss the model-based estimator for Σ .

Claim: $\hat{\tau}_k = \frac{\sum_{i=1}^n \hat{z}_{ik}}{n}$, for $k = 1, \dots, G$ is the MLE for τ_k for our likelihood equation

Proof:

let $p \in (1, \dots, G-1)$

$$\begin{aligned}
\frac{\partial l}{\partial \tau_p} &= \frac{\partial}{\partial \tau_p} \sum_{i=1}^n \sum_{k=1}^G \hat{z}_{ik} [\ln(\tau_k) + \ln(f_k(x_i | \mu_k, \Sigma_k))] \\
&= \sum_{i=1}^n \left[\frac{\partial}{\partial \tau_p} \sum_{k=1}^{G-1} \hat{z}_{ik} [\ln(\tau_k) + \ln(f_k(x_i | \mu_k, \Sigma_k))] \right. \\
&\quad \left. + \frac{\partial}{\partial \tau_p} \hat{z}_{iG} [\ln(1 - \sum_{k=1}^{G-1} \tau_k) + \ln(f_G(x_i | \mu_G, \Sigma_G))] \right] \\
&= \sum_{i=1}^n \left(\frac{\hat{z}_{ip}}{\tau_p} + (-1) \frac{\hat{z}_{iG}}{(1 - \sum_{k=1}^{G-1} \tau_k)} \right) = 0 \\
\Rightarrow \frac{\sum_{i=1}^n \hat{z}_{ip}}{\tau_p} &= \frac{\sum_{i=1}^n \hat{z}_{iG}}{1 - \sum_{k=1}^{G-1} \tau_k} \\
\tau_p &= \sum_{i=1}^n \hat{z}_{ip} \left[\frac{1 - \sum_{k=1}^{G-1} \tau_k}{\sum_{i=1}^n \hat{z}_{iG}} \right]
\end{aligned}$$

$1 - \sum_{k=1}^{G-1} \tau_k = \tau_G = \Pr(\text{a randomly chosen unit is in group } G)$
 $\sum_{i=1}^n \hat{z}_{iG} = \text{the number of units in group } G$

$$\text{Therefore, } \frac{1 - \sum_{k=1}^{G-1} \tau_k}{\sum_{i=1}^n \hat{z}_{iG}} = \frac{1}{n}$$

$$\text{And by substitution } \hat{\tau}_p = \frac{\sum_{i=1}^n \hat{z}_{ip}}{n}$$

We show that $\hat{\tau}_p$ is an MLE by taking second partial with respect to τ_p :

$$\begin{aligned}
\frac{\partial^2 l}{\partial \tau_p^2} &= \sum_{i=1}^n \left(-\frac{\hat{z}_{ip}}{\tau_p^2} - \frac{\hat{z}_{iG}}{(1 - \sum_{k=1}^{G-1} \tau_k)^2} \right) \\
&< 0
\end{aligned}$$

For the case when $p = G$, the proof is analogous. ■

Our maximum likelihood estimator $\hat{\tau}_k$, the probability that a random unit is in group k , is intuitive because it simply divides the total number of units in group k by the total number of units overall.

Claim: $\hat{\mu}_k = \frac{\sum_{i=1}^n \hat{z}_{ik} x_i}{\sum_{i=1}^n \hat{z}_{ik}}$, for $k = 1, \dots, G$ is the MLE for μ_k for our likelihood equation

Proof:

let $p \in (1, \dots, G)$

$$\begin{aligned} \frac{\partial l}{\partial \mu_p} &= \frac{\partial}{\partial \mu_p} \sum_{i=1}^n \sum_{k=1}^G \hat{z}_{ik} [\ln(\tau_k) + \ln(f_k(x_i | \mu_k, \Sigma_k))] = 0 \\ \frac{\partial}{\partial \mu_p} \sum_{i=1}^n \sum_{k=1}^G \hat{z}_{ik} \left[\ln(\tau_k) + \left(-\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) + \ln(2\pi^{m/2} |\Sigma|^{1/2}) \right] &= 0 \\ \sum_{i=1}^n -\frac{1}{2} \hat{z}_{ip} \frac{\partial}{\partial \mu_p} \left((x_i - \mu_p)^T \Sigma_p^{-1} (x_i - \mu_p) \right) &= 0 \\ \sum_{i=1}^n -\frac{1}{2} \hat{z}_{ip} (2\Sigma_p^{-1} (-1) (x_i - \mu_p)) &= 0 \\ \sum_{i=1}^n \hat{z}_{ip} \Sigma_p^{-1} x_i - \sum_{i=1}^n \hat{z}_{ip} \Sigma_p^{-1} \mu_p &= 0 \\ \Sigma_p^{-1} \mu_p \sum_{i=1}^n \hat{z}_{ip} &= \Sigma_p^{-1} \sum_{i=1}^n \hat{z}_{ip} x_i \\ \hat{\mu}_p &= \frac{\sum_{i=1}^n \hat{z}_{ip} x_i}{\sum_{i=1}^n \hat{z}_{ip}} \end{aligned}$$

We show that $\hat{\tau}_p$ is an MLE by taking second partial with respect to τ_p :

$$\begin{aligned} \frac{\partial^2 l}{\partial \mu_p^2} &= \frac{\partial}{\partial \mu_p} \left(\sum_{i=1}^n \hat{z}_{ip} \Sigma_p^{-1} x_i - \sum_{i=1}^n \hat{z}_{ip} \Sigma_p^{-1} \mu_p \right) \\ &= -\sum_{i=1}^n \hat{z}_{ip} \Sigma_p^{-1} \\ &< 0 \end{aligned}$$

So it is proved. ■

Again, this is an intuitive formula for the mean because we are simply taking the weighted sum of the values of each variable for those units in group p and then dividing by the total number of units in group p .

The maximum likelihood estimation calculations for Σ_k are not as straightforward. The method of finding the M.L.E. depends on the model used. We will primarily use the model with clusters of variable volume, equal shape, and variable orientation (VEV). We use VEV

because it is the most unconstrained model that we can use. Less constrained models are unavailable to us because they require more data to estimate additional parameters, and in our situation the initial clustering of units contains some groups that have fewer units than the number of parameters needed. For example, if we were to attempt to use a model where the covariance was completely unconstrained (VVV), then our covariance matrix would become singular (determinant equal to 0) on the first iteration because the number of parameters would exceed the number of units in the smallest group. In general, it is better to use unconstrained models because this allows clusters more flexibility in adjusting to the values of the data points (see Section 3.3.2). With our data, better results came from more unconstrained models.

In the VEV model, we begin maximum likelihood estimation with the eigenvalue decomposition of Σ_k (equation 2.4). However, because we are assuming that our clusters have equal shape, we use A , a diagonal matrix of normalized eigenvalues that does not vary across groups. With the VEV model, maximizing equation 2.13 with respect to Σ_k is equivalent to minimizing [1]

$$F(\lambda, \mathbf{D}, A) = \sum_{k=1}^G \frac{1}{\lambda_k} \text{tr}(W_k D_k A^{-1} D_k^T) + d \sum_{k=1}^G \sum_{i=1}^n \hat{z}_{ik} \ln(\lambda_k) \quad (2.15)$$

where

$$W_k = \sum_{i=1}^n \hat{z}_{ik} (x_i - \mu_k)(x_i - \mu_k)^T. \quad (2.16)$$

If we let the eigenvalue composition of W_k be $W_k = L_k \Omega_k L_k^T$, then it can be shown that the parameters λ_k , D_k , and A that minimize equation 2.15 are found by iteratively solving [1]

$$\hat{\lambda}_k = \frac{\text{tr}(W_k D_k A^{-1} D_k^T)}{d \sum_{i=1}^n \hat{z}_{ik}} \quad (2.17)$$

$$\hat{D}_k = L_k \quad (2.18)$$

$$\hat{A} = \frac{\sum_{k=1}^G \frac{1}{\lambda_k} \Omega_k}{|\sum_{k=1}^G \frac{1}{\lambda_k} \Omega_k|^{1/d}} \quad (2.19)$$

And finally, $\hat{\Sigma}_k = \hat{\lambda}_k \hat{D}_k \hat{A} \hat{D}_k^T$ is the maximum likelihood estimate for the k^{th} covariance matrix. Hence we have found the MLEs for all of our parameters of interest, and now the EM algorithm proceeds to the expectation step (E-step).

2.2.3 E-step (Expectation)

The M-step allowed us to find the values of μ , Σ , and τ that maximized our likelihood equation while keeping $\hat{\mathbf{z}}$ constant. This determined the most likely set of clusters based on our previous classification. Now, we wish to update $\hat{\mathbf{z}}$ to fit these clusters given the mixture

likelihood parametrized by the MLEs. That is, we want to find the expected classification based on the clusters defined by the M-step. Therefore, we will keep $\hat{\mu}$, $\hat{\Sigma}$, and $\hat{\tau}$ constant and find the expected value of z_{ik} for all $i = 1, \dots, n$ and $k = 1, \dots, G$.

Recall that z_{ik} is 1 if unit i is in group k and 0 if unit i is not in group k . Therefore, the expected value of z_{ik} given the fixed values of $\hat{\mu}$, $\hat{\Sigma}$, and $\hat{\tau}$ and the data \mathbf{x} is

$$\begin{aligned}\hat{z}_{ik} = E[z_{ik}] &= \Pr(x_i \in \text{unit } k | \hat{\mu}, \hat{\Sigma}, \hat{\tau}, x_i) \cdot 1 + \Pr(x_i \notin \text{unit } k | \hat{\mu}, \hat{\Sigma}, \hat{\tau}, x_i) \cdot 0 \\ &= \Pr(x_i \in \text{unit } k | \hat{\mu}, \hat{\Sigma}, \hat{\tau}, x_i).\end{aligned}$$

Observe that z_{ik} , which began as an indicator variable, is estimated using a probability: it is the probability that unit i is in group k given the parameters $\hat{\mu}$, $\hat{\Sigma}$, $\hat{\tau}$, and x_i . Now, we can apply Bayes' rule to rewrite this equation as

$$\hat{z}_{ik} = \Pr(x_i \in \text{unit } k | \hat{\mu}, \hat{\Sigma}, \hat{\tau}, x_i) = \frac{\Pr(x_i | x_i \in \text{unit } k, \hat{\mu}, \hat{\Sigma}, \hat{\tau}) \Pr(x_i \in \text{unit } k | \hat{\mu}, \hat{\Sigma}, \hat{\tau})}{\Pr(x_i | \hat{\mu}, \hat{\Sigma}, \hat{\tau})}. \quad (2.20)$$

Notice that $\Pr(x_i | x_i \in \text{unit } k, \hat{\mu}, \hat{\Sigma}, \hat{\tau}) = f_k(x_i | \hat{\mu}_k, \hat{\Sigma}_k)$ because given that a unit is in group k , we know it is described by the density function f_k . Also, $\Pr(x_i \in \text{unit } k | \hat{\mu}, \hat{\Sigma}, \hat{\tau}) = \hat{\tau}_k$, the probability that a randomly-drawn unit is in group k . Therefore we can rewrite equation 2.20 as

$$\hat{z}_{ik} = \frac{\hat{\tau}_k f_k(x_i | \hat{\mu}_k, \hat{\Sigma}_k)}{f(x_i | \hat{\mu}, \hat{\Sigma}, \hat{\tau})}, \quad (2.21)$$

and substituting equation 2.1 into the denominator yields the standard form of the equation for \hat{z}_{ik} [4],

$$\hat{z}_{ik} = \frac{\hat{\tau}_k f_k(x_i | \hat{\mu}_k, \hat{\Sigma}_k)}{\sum_{j=1}^G \hat{\tau}_j f_j(x_i | \hat{\mu}_j, \hat{\Sigma}_j)}. \quad (2.22)$$

Equation 2.22 says the probability that a unit i is in a particular group k given the parameters that define the clusters is the density of that point in cluster k multiplied by the probability that a randomly drawn unit is in group k divided by the sum of these products across all units. The original probability density function for \mathbf{x} was a weighted sum of the density functions of all the groups (with weights τ), so the probability that a particular unit is in group k is simply the proportion of the overall density of x_i comprised by group k .

Our new estimate, $\hat{\mathbf{z}}$, replaces equation 2.14 and we start again with the M-step. The EM algorithm iterates between the M-step and E-step until some tolerance level in the log likelihood function is reached. The algorithm finally stops when the value of the change in the log likelihood function in an entire iteration divided by the absolute value of the likelihood function is less than some predetermined value (we use a relative tolerance of 0.00001).

2.3 Properties of the EM Algorithm

In this section, we aim to show that by using the EM algorithm, the likelihood function, L , converges to some value L^* under suitable regularity conditions. We will give sufficient conditions for L^* to be a local maximum of L and sufficient conditions such that the set the parameters $\Phi = \{\mu, \Sigma, \tau\}$ converge to a stationary point.

2.3.1 General form of EM algorithm

Keeping with our notation from section 2.2, we will let \mathbf{Y} be the complete data and \mathbf{X} be the observed data, where $\mathbf{Y}=(\mathbf{X},\mathbf{Z})$. Then, there is a many-to-one mapping $\mathbf{X} = f(\mathbf{Y})$, because each observation X has many possible values of Z , and thus many corresponding values of Y . Then, we define

$$Y(\mathbf{X}) = \{\mathbf{Y} : f(\mathbf{Y}) = \mathbf{X}\} \quad (2.23)$$

as the set of all possible Y s corresponding to a particular X . We can write the probability of our data \mathbf{X} , given a set of parameters Φ as

$$g(\mathbf{X}|\Phi) = \int_{Y(\mathbf{X})} g_c(\mathbf{Y}|\Phi)d\mathbf{Y} \quad (2.24)$$

where we wish to maximize $L(\Phi) = g(\mathbf{X}|\Phi)$ [9] (Note: $L(\Phi)$ is given in equation 2.2 for the general finite mixture model setting). In equation 2.24, $g(\mathbf{X}|\Phi)$ is the probability density function of the observed data \mathbf{X} given Φ , while $g_c(\mathbf{Y}|\Phi)$ is the p.d.f. of the complete data \mathbf{Y} given Φ . In general, maximization of $g_c(\mathbf{Y}|\Phi)$ has an analytical solution, but maximization of $L(\Phi) = g(\mathbf{X}|\Phi)$ does not [9]. We will assume that the maximization of $g_c(\mathbf{Y}|\Phi)$ has an analytical solution. Therefore we use the EM algorithm, which iteratively attempts to maximize L .

Let $\Phi^{(p)}$ be the value of Φ after p iterations. Then, the next iteration will be:

E-step: Calculate $Q(\Phi|\Phi^{(p)})$, where

$$Q(\Phi|\Phi^{(p)}) = E_{\Phi^{(p)}}[\ln g_c(\mathbf{Y}|\Phi)|\mathbf{X}] \quad (2.25)$$

M-step: We now maximize the value of this expectation with respect to Φ . Therefore we choose $\Phi^{(p+1)}$ to be any value of Φ that maximizes $Q(\Phi|\Phi^{(p)})$. Thus

$$Q(\Phi^{(p+1)}|\Phi^{(p)}) \geq Q(\Phi|\Phi^{(p)}) \quad (2.26)$$

for all $\Phi \in \Omega$. This can be done because as stated above by assumption, the maximization of $g_c(\mathbf{Y}|\Phi)$ has an analytical solution.

2.3.2 $L(\Phi)$ is Non-decreasing for each Iteration

Claim:

$$L(\Phi^{(p+1)}) \geq L(\Phi^{(p)}) \text{ for all } p \in (0, 1, \dots) \quad (2.27)$$

Proof: We let p be a fixed natural number.

We define k to be the conditional density of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$:

$$k(\mathbf{y}|\mathbf{x}, \Phi) = \frac{g_c(\mathbf{y}|\Phi)}{g(\mathbf{x}|\Phi)}. \quad (2.28)$$

Then,

$$\begin{aligned} L(\Phi) &= g(\mathbf{x}|\Phi) \\ &= \frac{g_c(\mathbf{y}|\Phi)}{k(\mathbf{y}|\mathbf{x}, \Phi)} \end{aligned}$$

So,

$$\ln L(\Phi) = \ln g_c(\mathbf{y}|\Phi) - \ln k(\mathbf{y}|\mathbf{x}, \Phi) \quad (2.29)$$

Now, we take the expectation of each side of equation 2.29 with respect to the conditional distribution of \mathbf{Y} given $\mathbf{X} = \mathbf{x}$, using our current parameters, $\Phi^{(p)}$, which yields

$$\begin{aligned} E_{\Phi^{(p)}}[\ln L(\Phi)|\mathbf{x}] &= E_{\Phi^{(p)}}[\ln g_c(\mathbf{y}|\Phi)|\mathbf{x}] - E_{\Phi^{(p)}}[\ln k(\mathbf{y}|\mathbf{x}, \Phi)|\mathbf{x}] \\ \ln L(\Phi) &= Q(\Phi|\Phi^{(p)}) - H(\Phi|\Phi^{(p)}) \\ \text{where } H(\Phi|\Phi^{(p)}) &= E_{\Phi^{(p)}}[\ln k(\mathbf{Y}|\mathbf{x}, \Phi)|\mathbf{x}] \quad [9] \end{aligned}$$

Then,

$$\begin{aligned} \ln L(\Phi^{(p+1)}) - \ln L(\Phi^{(p)}) &= [Q(\Phi^{(p+1)}|\Phi^{(p)}) - Q(\Phi^{(p)}|\Phi^{(p)})] \\ &\quad - [H(\Phi^{(p+1)}|\Phi^{(p)}) - H(\Phi^{(p)}|\Phi^{(p)})] \end{aligned} \quad (2.30)$$

We now aim to show that $\ln L(\Phi^{(p+1)}) - \ln L(\Phi^{(p)}) \geq 0$.

By definition, the M-step maximizes Q over all values of Φ , so $\Phi^{(p+1)}$ is chosen such that

$$Q(\Phi^{(p+1)}|\Phi^{(p)}) \geq Q(\Phi^{(p)}|\Phi^{(p)}) \quad (2.31)$$

Therefore to prove the result, we must show that necessarily

$$[H(\Phi^{(p+1)}|\Phi^{(p)}) - H(\Phi^{(p)}|\Phi^{(p)})] \leq 0 \quad (2.32)$$

For any Φ ,

$$\begin{aligned} [H(\Phi|\Phi^{(p)}) - H(\Phi^{(p)}|\Phi^{(p)})] &= E_{\Phi^{(p)}}[\ln k(\mathbf{Y}|\mathbf{x}, \Phi)|\mathbf{x}] - E_{\Phi^{(p)}}[\ln k(\mathbf{Y}|\mathbf{x}, \Phi^{(p)})|\mathbf{x}] \\ &= E_{\Phi^{(p)}}\left[\ln \frac{k(\mathbf{Y}|\mathbf{x}, \Phi)}{k(\mathbf{Y}|\mathbf{x}, \Phi^{(p)})}|\mathbf{x}\right] \end{aligned}$$

$$\leq \ln E_{\Phi^{(p)}} \left[\frac{k(\mathbf{Y}|\mathbf{x}, \Phi)}{k(\mathbf{Y}|\mathbf{x}, \Phi^{(p)})} \middle| \mathbf{x} \right]$$

by Jensen's inequality and the concavity of \ln

$$= \ln \int_{Y(\mathbf{x})} \frac{k(\mathbf{Y}|\mathbf{x}, \Phi)}{k(\mathbf{Y}|\mathbf{x}, \Phi^{(p)})} k(\mathbf{Y}|\mathbf{x}, \Phi^{(p)}) d\mathbf{y}$$

by the definition of expected value and using the fact that $k(\mathbf{Y}|\mathbf{x}, \Phi^{(p)})$ is the conditional density of \mathbf{Y} given \mathbf{x} with respect to $\Phi^{(p)}$

Finishing the proof,

$$\begin{aligned} [H(\Phi|\Phi^{(p)}) - H(\Phi^{(p)}|\Phi^{(p)})] &\leq \ln \left(\int_{Y(\mathbf{x})} k(\mathbf{Y}|\mathbf{x}, \Phi) d\mathbf{y} \right) \\ &= \ln \left(\int_{Y(\mathbf{x})} \frac{g_c(\mathbf{y}|\Phi)}{g(\mathbf{x}|\Phi)} d\mathbf{y} \right) \\ &= \ln \left(\frac{1}{g(\mathbf{x}|\Phi)} \int_{Y(\mathbf{x})} g_c(\mathbf{y}|\Phi) d\mathbf{y} \right) \\ &= \ln \left(\frac{g(\mathbf{x}|\Phi)}{g(\mathbf{x}|\Phi)} \right) \\ &= \ln(1) \\ &= 0. \end{aligned}$$

Using this result and the result in equation 2.31, we see that substitution into equation 2.30 yields

$$\begin{aligned} \ln L(\Phi^{(p+1)}) - \ln L(\Phi^{(p)}) &\geq 0 \\ \Rightarrow L(\Phi^{(p+1)}) &\geq L(\Phi^{(p)}). \end{aligned}$$

2.3.3 Convergence of the EM Algorithm

We have proven that in each iteration of the EM algorithm, the likelihood equation necessarily increases in value or stays the same. In this section, we will discuss the conditions under which the likelihood converges to a value, L^* , and under which conditions L^* is a local maximum. Moreover, we will discuss the conditions under which the parameters converge to some value Φ^* as the likelihood equation converges.

If the sequence of loglikelihoods, $\{L^{(p)}\}$ is bounded above, then the loglikelihood will increase monotonically to some L^* . It can be shown that if (1) Ω is a subset of R^r , (2) the set of parameters Φ such that $L(\Phi) \geq L(\Phi_0)$ for any $L(\Phi_0) > -\infty$ is compact, and (3) L is continuous in Ω and differentiable in the interior of Ω , then $\{L^{(p)}\}$ is bounded above [11].

The compactness assumption above can be restrictive when no realistic compactification of the original parameter space is available, but this usually occurs only in variance components models or factor analysis [11], so will not be a concern with our model.

Therefore, in the EM process, the loglikelihood function converges to some value L^* . However, we would like to know if this convergence is to a global maximum, local maximum, or other stationary value such as a saddle point. Although global maximization of Q is guaranteed in the M-step, there is no guarantee of similar optimization of H . In general, the convergence point of the EM algorithm depends on the initialization of the algorithm when several local maxima or stationary points exist [11].

For L to converge to a local maximum, the following two conditions must be met: [11]

$$\begin{aligned}
 & Q(\Psi|\Phi) \text{ is continuous in both } \Psi \text{ and } \Phi \\
 & \sup_{\Phi'} Q(\Phi'|\Phi) > Q(\Phi|\Phi) \text{ for any } \Phi \in \{\text{stationary points that are not local maxima}\}
 \end{aligned}$$

The first condition is weak and satisfied in most situations, but the second is usually hard to verify. Therefore it is suggested that different initial values of the parameters be tried to determine maxima [11]. If the algorithm is stuck on some stationary point such as a saddle point, a small random perturbation of Φ away from the saddle point will usually cause the algorithm to diverge from the saddle point [9].

Convergence of L to a local or global maximum does not automatically imply that the sequence of parameters, $\{\Phi\}$ will converge to some point Φ^* . For the algorithm to give meaningful results, the parameters should approach some point as the likelihood does because ultimately we do not care about the value of the likelihood, only the values of the parameters.

It has been shown [11] that $\{\Phi\}$ converges to a local maximum if there are a discrete number of stationary points with a given L and $\|\Phi_{p+1} - \Phi_p\| \rightarrow 0$ as $p \rightarrow \infty$. These conditions are quite restrictive, and the requirements for them to hold are many and outside the scope of this discussion.

Chapter 3

Results

Our overall goal in clustering the units into unit type is to test if the geologic map of V14 is consistent with the numerical data. We aim to identify units on V14 whose classification does not fit the data, and therefore might be incorrectly classified. We will not be the final authority on decisions of correct classification, we only wish to present the geologists with a set of units that should be further investigated.

3.1 Data Collection

We begin by assuming that the units are well-drawn and that the geologists' classifications are correct. These are reasonable assumptions because the geologists have spent three years creating and revising the geologic map using standard planetary mapping techniques. The assumption that the units are well-drawn is very important because it allows us to treat each unit as a well-defined entity. Therefore, once we have extracted our data from each unit, we do not have to concern ourselves with the possibility that the unit is misdrawn. If we allow for the units to change, our model becomes much too complicated, as there are uncountably many ways to draw the units.

The assumption that the existing geologic map is correct is essential because if we find a handful of units whose classification does not fit the data, then these differences can be attributed to characteristics of the data, and not some other initial classification. The assumption is reasonable because we are not attempting to propose an entire overhaul of the map, as this would be foolish given the amount of time and effort that went into creating the initial map.

We extracted the data from the units using Arcview GIS software. With the "Zonal Statistics" tool, we obtained the mean, standard deviation, and median of pixel values within the boundary of each unit for each data set. The five data sets are surface emissivity, surface reflectivity, topography (elevation), root mean squared (RMS) slope, and backscatter (FMAP radar image). The backscatter has the highest resolution (225 m/pixel) and represents an overhead picture of V14. The backscatter raster was latitude-corrected to eliminate the dependency of pixel values on latitude [7].

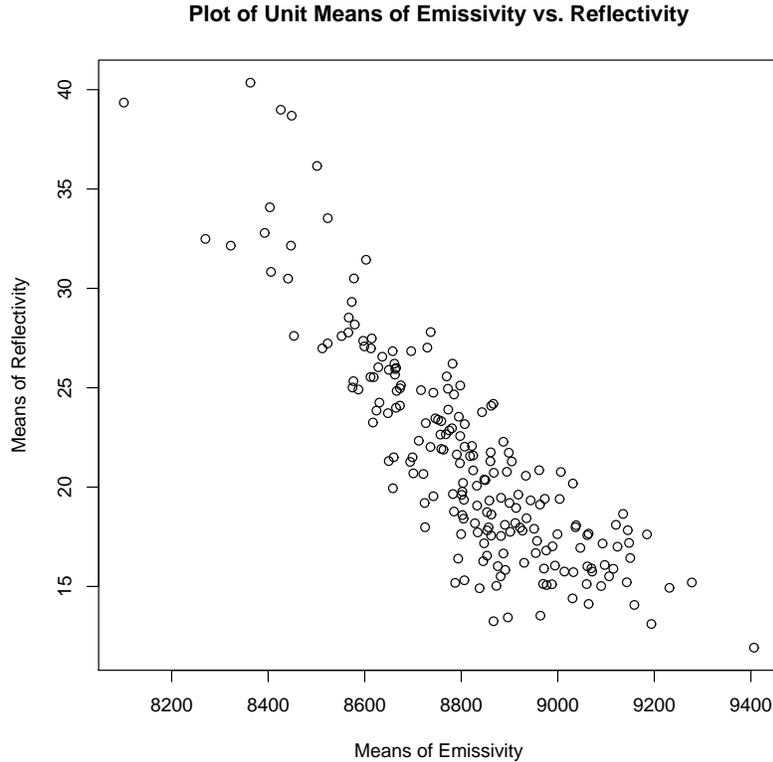


Figure 3.1: Plot of the mean of emissivity for each unit versus that unit’s mean of reflectivity. The variables should be exactly inversely related. However, we observe that although there is a significant inverse trend, there is much variation.

Although we have five data sets, two of them are measurements of the same property. Emissivity and reflectivity both measure the percent of light that is absorbed (or reflected) by the surface. Therefore, their values should correspond exactly. However, somewhere between Venus and Pomona College, their values have been incrementally altered (e.g. through imperfect extraction of the data) such that they are not perfect inverses (see Figure 3.1). Therefore, we expect to obtain slightly different results clustering with emissivity versus clustering with reflectivity. However, we cannot cluster with both variables because in that case we would double-weight the property, which would be reflected in the reassignment of a large number of units with slightly anomalous values of emissivity/reflectivity.

The breakdown for the number of units in each group from the initial classification is in Table 3.1. For the remainder of this paper, we will use the notation introduced by Grosfils [5]. We notice that there are several groups with more than 15 units (*t* (56 units), *prc* (25), *fe* (20), *pe* (18), and *prb* (16)). However, there are also five groups with only one unit

Name of Group	Symbol	Number of Initial Members
Linear belt	<i>bl</i>	5
Crater	<i>c</i>	11
Crater deposit	<i>cf</i>	1
Corona rim	<i>co</i>	1
Corona flow	<i>cof</i>	1
Airburst	<i>cs</i>	4
Volcanic flow	<i>fe</i>	20
Edifice plains	<i>pe</i>	18
Homogeneous plains	<i>ph</i>	1
Lineated plains	<i>pl</i>	3
Lineated plains 1	<i>pla</i>	7
Lineated plains 2	<i>plb</i>	10
Lahevhev plains	<i>pLla</i>	1
Lahevhev plains 1	<i>pLlb</i>	7
Light background plains	<i>pra</i>	13
Intermediate background plains	<i>prb</i>	16
Dark background plains	<i>prc</i>	25
Tessera	<i>t</i>	56

Table 3.1: Initial allocation of the 200 units by the geologists. Five units only have one member, which is a problem for model-based clustering.

member (*cf*, *co*, *cof*, *ph*, and *pLla*). Single unit clusters are a problem in our model, because the covariance matrices that we find for each of these five groups in the M-step of the EM algorithm will be singular, and the algorithm will be forced to stop before completion. We have two options: (1) we can attempt to reallocate these units into other groups based on geologic knowledge, or (2) we can remove them altogether and continue with the remaining units.

Relocation of the single units can be done by combining groups based on geologic knowledge. The first single group we see is *cf*, or crater deposit. This is formed from the ejecta of an impactor event (like a meteor hitting the surface). We will put this group into the airburst group (*cs*). Airbursts on Venus are formed when a large impactor does not reach the surface because it is broken up by the planet’s extremely thick clouds, yet still transfers energy to the surface of the planet via a powerful shock wave, thus changing the geology of the surface [5]. Although these two geologic unit types are not exactly the same, they are formed in similar manners, and are suitable to be grouped together. Next we group the *co* unit into the group *bl*. Corona, prevalent features on the surface of Venus, are circular formations created by rising hot magma that does not reach the surface [5]. The structure of the corona is a central region surrounded by concentric rings. The rings, which are a series of tectonic ridges, are similar in nature to linear belts, which are linear regions with much tectonic activity. We place the *cof* and *ph* units into the *fe* group. This is reasonable

because the *fe* group contains units of relatively young volcanic flows. Both the *cof* and *ph* units are, though slightly different geomorphologically than the *fe* group, well within the comfort range of being relatively young flows units. Finally, we regroup *pLla* into *pLlb*. Both of these unit types are a particular type of plains unit, and although slightly different in nature, they fit better together than with any other group.

3.2 Clustering Results

3.2.1 Relocation of Single Units

We begin clustering by considering the adjusted groups (those where the single-unit groups are relocated based on geologic knowledge). We will cluster with four different combinations of variables: (1) means and standard deviations of every property except emissivity, (2) means and standard deviations of every property except reflectivity, (3) means, standard deviations, and medians of every property except emissivity, and (4) means, standard deviations, and medians of every property except reflectivity. These four are logical combinations of variables because we cannot include both emissivity and reflectivity in one clustering attempt or we risk double-weighting the same property. Also, it is advantageous to include medians in two clusterings because the medians may carry additional information not present in the means data, such as skewness of the distribution of pixel values. It is also important to note that there is no median information for the backscatter data set because, unlike the other properties, each backscatter pixel has a non-integer value (due to latitude-correction), and ArcGis does not compute medians for non-integer data.

Results for the four clustering attempts are in Table 3.2. When we use more variables in our model, fewer units are moved to different groups and the algorithm converges more quickly. This occurs because with more variables, the units are more rigidly bound to their original groups since to be reallocated they now need to strongly differ from their original group in more variables. This also explains why the algorithm converges more

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated	# of Iterations to Convergence
means & standard devs., no emissivity	61	30.5%	23
means & standard devs., no reflectivity	67	33.5%	22
means, st. devs., and medians, no emissivity	37	18.5%	12
means, st. devs., and medians, no reflectivity	29	14.5%	8

Table 3.2: Basic results for the four clustering attempts. When more variables are included, the algorithm stops after fewer iterations and results are closer to the geologists’.

quickly: if fewer units are changing groups, they will settle into a stable arrangement quickly. The fact that models with more variables “force” fewer units to be reallocated is not problematic. With each clustering result, we obtain different information about the units that are reallocated. For instance, if we find that a particular unit is reallocated under one set of variables and not another, we can identify the variables in which the unit is anomalous with respect to its original classification.

The final allocation after the convergence of the EM algorithm is determined by mapping each unit to the group for which its value of z is highest. Therefore, theoretically we can be uncertain about the mapping of some units (for instance if the largest z_{ik} is less than 0.5). However, in our case this is not a problem. For the two clustering attempts that use medians, the smallest maximum values of z_{ik} are .899 and .893 respectively. For the two clustering attempts that do not include the medians, the maximum values drop to .551 and .542, respectively. However, this is still not a great worry because, for example, 195 of the 200 units in the no medians nor emissivity clustering result have maximum z_{ik} s greater than 0.9, and 188 have maximum z_{ik} s greater than 0.95. Therefore there are only a few units whose final classification is somewhat uncertain, and these can be easily identified and studied.

Each of the four clusterings gives us different information. We want to combine this information to get an overall idea of what the data sets tell us about the units. Also, keeping with our original assumption that the geologic map is accurate, we want to present a small subset of units for the geologists to reexamine. Therefore we will look at trends across the four clustering results. For each unit, we see how many times that unit was reallocated in the four clusterings (Table 3.3).

These results show confidence in the original geologic map, as over half of the units (103) did not change at all in any clustering. Also, we can pinpoint 14 units (7.0% of the units) that were reallocated every time and an additional 8 that were reallocated in three of the four clustering attempts. For each of these units, we know the group in which it started, where it is situated on the map, and to which group(s) it is allocated by clustering. This is a wealth of information about a small number of units that can be easily transferred to the geologists. For the units that were reallocated every time in the series of four clusterings we performed, the area of the unit, its original classification and its classification(s) by model based clustering are listed in Table 3.4. For the clustering classifications, the number in parentheses denotes the number of times (out of four) that the unit was classified to that particular group.

The two most commonly reallocated units are *fe* (6) and *pe* (4). Geologically, it is sensible that these units may be distributed differently than the geologist-created map

Number of Times Reallocated Across the Four Different Variable Combinations	0	1	2	3	4
Number of Units	103	36	39	8	14

Table 3.3: Results across the four combinations of variables for the individual units.

Unit Number	Unit Area (meters ²)	Original Classification	Clustering Classifications
50	7.15×10^9	<i>t</i>	<i>pLlb</i> (4)
81	3.09×10^{10}	<i>pe</i>	<i>prb</i> (2), <i>fe</i> (1), <i>prc</i> (1)
98	2.24×10^9	<i>prb</i>	<i>fe</i> (1), <i>pe</i> (1), <i>pra</i> (1), <i>prc</i> (1)
100	1.63×10^{10}	<i>fe</i>	<i>plb</i> (2), <i>prb</i> (1), <i>prc</i> (1)
102	4.18×10^9	<i>pe</i>	<i>prb</i> (4)
105	2.37×10^{11}	<i>fe</i>	<i>pe</i> (2), <i>prc</i> (2)
111	8.81×10^{10}	<i>pe</i>	<i>prb</i> (3), <i>fe</i> (1)
133	1.50×10^{10}	<i>pe</i>	<i>prb</i> (3), <i>prc</i> (1)
136	3.55×10^{10}	<i>fe</i>	<i>prc</i> (3), <i>pra</i> (1)
139	1.92×10^9	<i>prc</i>	<i>pe</i> (3), <i>prb</i> (1)
142	6.06×10^{10}	<i>fe</i>	<i>prc</i> (2), <i>pra</i> (1), <i>prb</i> (1)
143	5.57×10^{10}	<i>fe</i>	<i>prb</i> (3), <i>prc</i> (1)
167	2.07×10^9	<i>fe</i>	<i>prb</i> (2), <i>prc</i> (1), <i>c</i> (1)
193	3.66×10^8	<i>prc</i>	<i>t</i> (4)

Table 3.4: The 14 units that were reallocated in each of the four clusterings.

when the classification is performed with only the numerical data. The *fe* unit is a volcanic flows unit, meaning that a relatively recent volcanic event created it. However, without analyzing the geomorphology of these units, it is virtually impossible to tell volcanic flows apart from other plains units, which were formed by older volcanic events. The *pe* units are edifice plains, meaning plains filled with shield volcanoes. Again, without recognizing the presence of these volcanoes, it would be hard to differentiate them from other plains units. At the same time, only a handful of each of these units are reallocated every time. This means that something starkly different is occurring on these particular *fe* and *pe* units than on others, and therefore the processes that formed them may be different.

Additionally, two relatively small *prc* units, a small *prb* unit, and a slightly bigger *t* unit are all reallocated every time. These units clearly have properties that are out of the ordinary for their unit type. However, on very small units like unit number 193, the size of a physical property pixel is very large compared to the size of the unit. There may be strange things happening to the data from smaller units like these because a small amount of pixels means a higher variability in the mean, median, and standard deviation of these pixel values. Therefore, there may be characteristics in the data that are not actually present geologically.

3.2.2 Removal of Single Units

Next, we cluster without the units that belong to single-member groups. This allows us to cluster without first depending on a reallocation of some units to groups that are, although similar in nature to the groups specified by the geologists, not actually the groups that the geologists assigned these units to. By removing these five units, their data will no longer

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated	# of Iterations to Convergence
means & standard devs., no emissivity	53	27.2%	14
means & standard devs., no reflectivity	68	34.9%	18
means, st. devs., and medians, no emissivity	40	20.5%	13
means, st. devs., and medians, no reflectivity	30	15.4%	13

Table 3.5: Statistics for the four clustering attempts with single units removed.

Number of Times Reallocated Across the Four Different Variable Combinations	0	1	2	3	4
Number of Units	96	43	29	18	9

Table 3.6: Results across the four combinations of variables.

affect the parameters of the groups into which they were relocated. In turn, clustering will be initiated with a set of groups that is truer to the geologists’ original interpretation. Hence we tie ourselves closer to the original null assumption that the geologists have correctly partitioned the units into groups.

Using the same combinations of variables as the previous section, we obtain the results in Table 3.5, which are similar to the results in Table 3.2. Looking at trends across groups yields Table 3.6.

First notice that the number of units that are never reallocated has dropped to 96 (49.2% of the units—down from 51.5% in the relocation clusterings). At the same time, only 9 units (4.6%) are reallocated from the geologists’ classification every time. The proportion of units that are reallocated in over half of the clustering attempts is slightly higher with single units removed (13.8%) than with single units relocated (11.0%). Overall, the results under removal of single units are very similar to the results under relocation. However, we expect that a different set of units will be reallocated every time under single-unit removal because the removed units no longer will affect the parameters of any cluster. This set of units is in Table 3.7.

In reality, we see that of the nine units that are reallocated every time by our clusterings (with single units removed), seven of these were also reallocated every time by the first set of clusterings (with single units relocated). The two newly-reallocated units are a large *t* unit and a very small *prc* unit. The reallocation of the large *t* (tessera) unit is unexpected, especially since it is reallocated into the *fe* (volcanic flow) group, a unit type that does not at all resemble extremely deformed tessera. The reclassification of the *prc* unit is not troubling because the unit is very small.

An interesting result from this set of clusterings is that only one *fe* unit is reallocated

Unit Number	Unit Area (meters ²)	Original Classification	Clustering Classifications
39	1.29×10^{10}	<i>t</i>	<i>fe</i> (4)
81	3.09×10^{10}	<i>pe</i>	<i>fe</i> (3), <i>prb</i> (1)
95	6.68×10^8	<i>prc</i>	<i>fe</i> (3), <i>t</i> (1)
102	4.18×10^9	<i>pe</i>	<i>prb</i> (4)
111	8.81×10^{10}	<i>pe</i>	<i>prb</i> (3), <i>fe</i> (1)
133	1.50×10^{10}	<i>pe</i>	<i>prb</i> (2), <i>prc</i> (1), <i>fe</i> (1)
139	1.92×10^9	<i>prc</i>	<i>pe</i> (4)
167	2.07×10^9	<i>fe</i>	<i>prb</i> (3), <i>prc</i> (1)
193	3.66×10^8	<i>prc</i>	<i>t</i> (4)

Table 3.7: The nine units that were reallocated every time by the single unit removal clusterings.

every time compared to six that were reallocated every time by the clustering results where single units were relocated. In the relocation of single groups, two units (one *cof* and one *ph*) were placed into the *fe* group. This may have caused an artificial distortion in the parameters of the *fe* group, thus causing a large amount of *fe* units to be incorrectly reallocated.

3.3 More Clustering Results

3.3.1 Background Plains

We will briefly turn our attention to the background plains (*pra*, *prb*, and *prc*). Background plains make up 62.3% of the area of the map and were created by region-wide volcanic events. There are three categories of background plains corresponding to the brightness of the radar backscatter. The *pra* units have bright, the *prb* units have intermediate, and the *prc* units have dark radar backscatters. Essentially, the only way that geologists differentiated between these three groups was by the color of the radar backscatter. However, they performed this analysis before the backscatter data was latitude-corrected. Therefore, while they have confidence in the comparison of brightness between two neighboring units, they do not hold this same confidence for units that lie in significantly different latitudes.

To analyze the geologists' final classification of background plains units, we will cluster with only *pra*, *prb*, and *prc* units. Again, we will begin with the null assumption that the geologists are correct so that any reclassifications can be attributed to properties of the data. We will also use the same combinations of variables that we used in section 3.2 and compare results between clusterings.

When we perform this clustering, we get results that suggest that the geologists have classified these units well. For any one clustering, the maximum number of units that are reclassified is nine (16.7% of background plains units). The best individual clustering

Number of Times Reallocated Across the Four Different Variable Combinations	0	1	2	3	4
Number of Units	38	10	3	2	1

Table 3.8: Results across clusterings for the 54 background plains units. Only one unit was reclassified every time.

Unit Number	Unit Area (meters ²)	Number of Times Reclassified	Original Classification	Clustering Classifications
138	1.43×10^{11}	4	<i>prc</i>	<i>prb</i> (4)
101	1.21×10^9	3	<i>pra</i>	<i>prc</i> (2), <i>prb</i> (1)
150	4.31×10^8	3	<i>prb</i>	<i>prc</i> (3)
98	2.24×10^9	2	<i>prb</i>	<i>pra</i> (1), <i>prc</i> (1)
130	1.40×10^9	2	<i>prc</i>	<i>prb</i> (2)
162	4.01×10^{10}	2	<i>prc</i>	<i>prb</i> (2)

Table 3.9: The six units that were reallocated in at least two of the four clustering attempts. Most of these units have very small areas compared to the average background plains unit.

reclassifies five units (9.3%). Results across clusterings are in Table 3.8.

Across clusterings, we see that well more than half of all background plains units (70.4%) are never reclassified by model-based clustering. Also, only one unit is reclassified every time, and only five others are reclassified in at least two of the four clustering attempts.

Unit number 138, a large unit in the northwest corner of V14, is the only background plains unit that was reclassified every time. This unit, which is surrounded on three sides by a *pra* unit and does not touch any other background plains unit, was originally classified as a *prc* unit by the geologists. Our clusterings unanimously conclude that this unit should be classified as a *prb* unit. Because the unit does not touch either of these unit types and since the geologists performed the clustering with non-corrected backscatter data, this unit definitely needs geologic reexamination with the latitude-corrected data.

Four of the other five units in the table are very small (the entire map has area 6.55×10^{12}), and none is reclassified every time. Therefore they do not warrant any geologic reexamination with adjusted backscatter. Unit number 162 is somewhat sizable and was classified as *prb* each time emissivity was used and as *prc* each time reflectivity was used. Therefore, there is something anomalous in the unit’s data. The unit shares a long boundary with a large *prb* unit, and is obviously darker. Thus, if it is indeed a background plains unit, it should be classified as *prc*. An explanation for the reclassification by the clusterings is a very large concentration of shield volcanoes on the unit.

3.3.2 A Less Constrained Model

Thus far we have been using the model where our clusters have variable volume, equal shape, and variable orientation (VEV). This is the most unconstrained model available to

us with the initial classification of geologic units (see section 2.2.2) because some groups do not initially have enough units. If we want to use the most unconstrained model: variable volume, shape, and orientation of clusters (VVV), we need to either combine or eliminate groups such that each group initially has enough units.

The more unconstrained model we use, the more flexibility the individual clusters have to fit the structure of the data. We want to use the most unconstrained model available because we have no preconceived notion that all clusters should have the same volume, shape, or orientation. We want the clusters to be able to fit the data without these restrictions.

We begin by relocating the single units as in section 3.2.1. Then, because craters and airbursts are features that are essentially unmistakable under geologic scrutiny, we will remove groups *c* and *cs* from our model. We will also remove the six linear belt units (*bl*) because there are very few of them and because they are easily recognized with geologic analysis. Next, we will combine all of the deformed plains units (*pl*, *pla*, *plb*) because they are all similar geologically. Finally, we will remove the brushed plains (*pLlb*) because there are very few of these units and they only occur in one small, isolated area of the map.

This leaves us with 170 units and seven groups, where initially the smallest group (*pra*) has 13 unit members. When we run VVV clustering with the same combination of variables as before, we run into problems with singularity of covariance matrices in the clustering attempt that uses means, standard deviations, and medians with no emissivity (we do not have this problem in any other combination of variables). The problem occurs because in VVV clustering we need more units in each group than number of variables, and in this clustering one of the groups (*fe*) drops below 11 members. We can resolve this problem by using fewer variables. Therefore, in this combination of variables we will remove the slope variable (arguably it is the least important in deciding group membership). We now cluster using the VEV and then the VVV model.

VEV Results

Tables 3.10 and 3.11 present the results for each of the four clustering attempts using the VEV model and our condensed groups.

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated	# of Iterations to Convergence
means & standard devs., no emissivity	74	43.5%	32
means & standard devs., no reflectivity	58	34.1%	22
means, st. devs., and medians, no emissivity or slope	65	38.2%	19
means, st. devs., and medians, no reflectivity	46	27.1%	22

Table 3.10: Results for each clustering attempt using the combined groups and VEV model.

Number of Times Reallocated Across the Four Different Variable Combinations	0	1	2	3	4
Number of Units	65	29	28	34	14

Table 3.11: Results for individual units for the VEV clustering.

VVV Results

The corresponding results for our VVV clusters are in the tables 3.12 and 3.13.

The increased flexibility of the clusters in the VVV model yields results that are slightly closer to the classification by the geologists. We see that 80 units (47.1% of the units) stay classified the same as the original geologic map with the VVV model compared to only 65 (38.2%) with the VEV model. We are not surprised that VVV yields results closer to the geologists' because we initialize the EM algorithm by assuming the geologists are correct. Thus in a less constrained model the clusters will have more flexibility to conform to the data and the initial classification.

Five of the six units that were reclassified every time both the VEV and VVV models

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated	# of Iterations to Convergence
means & standard devs., no emissivity	54	31.8%	18
means & standard devs., no reflectivity	61	35.9%	40
means, st. devs., and medians, no emissivity or slope	52	30.6%	19
means, st. devs., and medians, no reflectivity	40	23.5%	13

Table 3.12: Cluster results for the VVV model.

Number of Times Reallocated Across the Four Different Variable Combinations	0	1	2	3	4
Number of Units	80	31	17	26	16

Table 3.13: Results across clusterings for the VVV model.

Unit Number	Unit Area (meters ²)	Original Classification	Clustering Classifications
95	6.68×10^8	<i>prc</i>	<i>t</i> (3), <i>fe</i> (2), <i>pra</i> (2), <i>pe</i> (1)
105	2.37×10^{11}	<i>fe</i>	<i>prc</i> (3), <i>pe</i> (3), <i>prb</i> (2)
111	8.81×10^{10}	<i>pe</i>	<i>prb</i> (7), <i>pl</i> (1)
133	1.50×10^{10}	<i>pe</i>	<i>prc</i> (5), <i>prb</i> (3)
167	2.07×10^9	<i>fe</i>	<i>prb</i> (6), <i>prc</i> (1), <i>pe</i> (1)
196	2.58×10^{11}	<i>prb</i>	<i>fe</i> (7), <i>t</i> (1)

Table 3.14: Six units were reallocated every time by both the VEV and VVV models. These units are listed in the table above.

with combined groups were also reallocated by one of the original clusterings (see section 3.2). The sixth, unit number 196 is a large intermediate background plains unit. It is continually being classified as a volcanic flows unit by the clusterings. As discussed before, our clustering method’s failure to observe the geomorphological geometry of the unit may be causing this reallocation.

There are some problems with the techniques we used to run the VVV model. We removed 30 units, and although we only eliminated units whose original classification was confident, we still lose much of the structure of the entire data set by eliminating them. Also, we combined all of the deformed plains *pl* groups into one supergroup. While the deformed plains are all similar, there are differences between the three subgroups. Hence a *pla* unit may be very close to the *pla* cluster’s center but far from the center of the combined *pl* cluster, causing it to be reallocated when in reality it should not have been. These distortions cause changes throughout all groups because once a unit is reallocated it subsequently changes the parameters of its new group, which may in turn cause more reallocations.

Therefore, while the condensing of units into seven groups was an easy way to be able to use the VVV model, major alterations to the original classification should not be performed. We did show that more flexible models yield clusterings that are closer to the original geologic map. In fact, as we use stricter and stricter models, our results stray further from the geologists’ work. Therefore, if we had more geologic units at our disposal it would be advantageous to use the VVV model, but with our data we must stay with the VEV model.

Chapter 4

Weighting

4.1 Weighting by Unit Area

4.1.1 Motivation

In our original model, the ability of unit i to determine the parameters of cluster k only depended on z_{ik} , the probability that unit i belongs to group k . For example, our equation for the maximum likelihood estimate of μ_k ,

$$\hat{\mu}_k = \frac{\sum_{i=1}^n \hat{z}_{ik} x_i}{\sum_{i=1}^n \hat{z}_{ik}}$$

was only a function of the data, \mathbf{x} , and \hat{z}_{ik} . In fact, $\hat{\mu}_k$ was the weighted average of the data, with the weights being the \hat{z}_{ik} s.

Under this model, if two units belong to the same group k with probability 1, they will both have the same power in determining $\hat{\mu}_k$, $\hat{\Sigma}_k$, and $\hat{\tau}_k$, even if one unit is 100 times the size of the other unit. However, we want the larger unit to have more influence in determining the center, size, and shape of the cluster because it is a more consequential unit for the region as a whole, is easier for the geologists to study and categorize, and produces more accurate data for the mean, standard deviation, and median of pixel values for each variable because it contains more pixels.

On the V14 quadrangle, unit size is a major issue. The quadrangle is dominated by a few extremely large units with many smaller units scattered throughout. The largest unit on V14 makes up 28.6% of the area of the map while the smallest makes up only 0.0013% of the map, a difference of 22,000 times the size! Figure 4.1 shows the large range of unit areas on the quadrangle. In our original model-based clustering we heavily biased toward the smaller units by not taking into account the large differences in unit areas. Our goal in this section, then, is to devise a way to take into account these tremendous differences in unit area in our model.

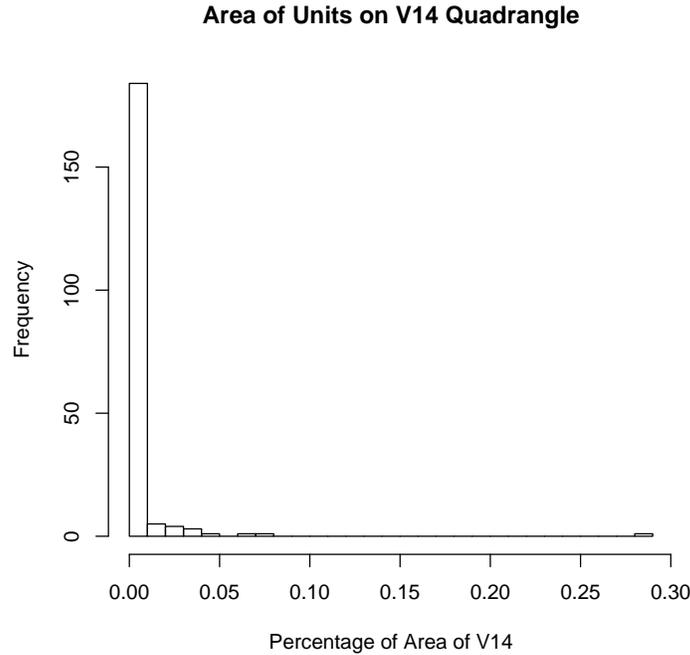


Figure 4.1: Distribution of the areas of the 200 units on the Ganiki Planitia (V14) Quadrangle. The vast majority of units are very small. There are a few very large units, including one (*pre*) that takes up 28.6% of the map.

4.1.2 Methods

A sensible way to approach the unit area problem is to simply cut the map into more units. Doing this would not only solve the area problem but also create more units and consequently allow us to use more flexible models like VVV. There are several ways to cut the map into smaller units. First, we can make a series of latitudinal and longitudinal cuts across the existing geologic map, thus cutting all units that the lines intersect. This method gives us some control over the areas of the new units, but we will also cut some smaller units into even smaller parts. A second option is to set a maximum value of the area of a unit and subsequently cut all units bigger than this threshold into the correct amount of pieces so that each piece is smaller than the threshold. Here, although we reduce the range of the areas of the units, there will still be a range of areas from the smallest unit to the threshold value. A third option is to cut the map into equal-area squares and then assign each square to the unit type which most of its area belongs to. This technique eliminates the range of areas, but now we are ignoring the individual unit boundaries that the geologists created, which is foolish because our original hypothesis is that the original geologic map is correct.

Any method we choose to cut the map into smaller pieces introduces critical problems. First, unless we ignore the geologists' original boundaries altogether, we can never actually completely equalize the areas. Also, there are an infinite number of ways we can cut the map, and each will create different data that will probably produce different clustering results. There is no correct way to cut the map, so unless we try every plausible combination of cuts (which is impossible) we will not obtain robust results. Lastly, and probably the most important reason we should not try to cut the map into smaller pieces is that the geologists have drawn the units on the map in such a way that each unit is a single entity. Although there may be variation across the unit, there are similarities across the entire unit that the geologists have recognized. For example, when we attempted the second cutting technique above, there was one *pe* unit that, although never reallocated once in nonweighted clustering, had each of its six parts reallocated every time. The reason for this is that *pe* units are plains units with large concentrations of small shield volcanoes. And, although the characteristics of these volcanoes may vary across the unit, the fact that they are all located on this single plains region is reason enough to call the region a single unit. Therefore, we will not alleviate the unit area problem by cutting the map into smaller pieces. We must stay with our original assumption that the geologists have correctly drawn the units.

Creating a New Likelihood

A better way to handle the unit area problem is to introduce a vector of weights, \mathbf{w} , that will indicate the size of the unit for each unit i . Similar to how our probability matrix \mathbf{z} determined how much each unit affected the parameters of each group, our vector \mathbf{w} will control how much influence each unit has over the parameters of the clusters. For instance, we want the maximum likelihood estimate for our mean vector for unit k , $\hat{\mu}_k$ to be a weighted average of the data that belong to the group and the corresponding area weights of those units. Therefore we want

$$\hat{\mu}_k = \frac{\sum_{i=1}^n w_i \hat{z}_{ik} x_i}{\sum_{i=1}^n w_i \hat{z}_{ik}} \quad (4.1)$$

for all $k = 1, \dots, G$. If two units belong to group k with probability 1, but one is 100 times the size of the other, we can scale our weights w_i accordingly so the larger unit has much more power to determine the value of $\hat{\mu}_k$ than the smaller unit.

Now, we need to determine how to integrate this MLE for μ_k (equation 4.1) back into our original model. Recall from chapter 2 that we devised a likelihood equation (equation 2.12) using mixture models. This likelihood equation defined the likelihood of our clusters based on the data and our current allocation. We now want to construct a similar likelihood equation that produces an MLE for $\hat{\mu}_k$ that is equation 4.1.

Claim: $\hat{\mu}_k = \frac{\sum_{i=1}^n w_i \hat{z}_{ik} x_i}{\sum_{i=1}^n w_i \hat{z}_{ik}}$ is the maximum likelihood estimator for $\hat{\mu}_k$ under the likelihood equation

$$L(\mu, \Sigma, \tau | \mathbf{x}, \mathbf{z}) = \prod_{i=1}^n \prod_{k=1}^G f_k(x_i | \mu_k, \Sigma_k)^{w_i z_{ik}} \tau_k^{w_i z_{ik}}. \quad (4.2)$$

Proof:

$$\begin{aligned}
\frac{\partial l}{\partial \mu_p} &= \frac{\partial}{\partial \mu_p} \sum_{i=1}^n \sum_{k=1}^G w_i \hat{z}_{ik} [\ln(\tau_k) + \ln(f_k(x_i | \mu_k, \Sigma_k))] = 0 \\
\frac{\partial}{\partial \mu_p} \sum_{i=1}^n \sum_{k=1}^G w_i \hat{z}_{ik} &\left[\ln(\tau_k) + \left(-\frac{1}{2} (x_i - \mu_k)^T \Sigma_k^{-1} (x_i - \mu_k) \right) + \ln(2\pi^{m/2} |\Sigma|^{1/2}) \right] = 0 \\
\sum_{i=1}^n -\frac{1}{2} w_i \hat{z}_{ip} \frac{\partial}{\partial \mu_p} &\left((x_i - \mu_p)^T \Sigma_p^{-1} (x_i - \mu_p) \right) = 0 \\
\sum_{i=1}^n -\frac{1}{2} w_i \hat{z}_{ip} (2\Sigma_p^{-1} (-1) &(x_i - \mu_p)) = 0 \\
\sum_{i=1}^n w_i \hat{z}_{ip} \Sigma_p^{-1} x_i - \sum_{i=1}^n &w_i \hat{z}_{ip} \Sigma_p^{-1} \mu_p = 0 \\
\Sigma_p^{-1} \mu_p \sum_{i=1}^n w_i \hat{z}_{ip} &= \Sigma_p^{-1} \sum_{i=1}^n w_i \hat{z}_{ip} x_i \\
\hat{\mu}_p &= \frac{\sum_{i=1}^n w_i \hat{z}_{ip} x_i}{\sum_{i=1}^n w_i \hat{z}_{ip}}
\end{aligned}$$

■

Therefore the likelihood equation we want to use for unit area weighting is equation 4.2. Our loglikelihood equation is

$$\sum_{i=1}^n \sum_{k=1}^G w_i \hat{z}_{ik} [\ln(\tau_k) + \ln(f_k(x_i | \mu_k, \Sigma_k))]. \quad (4.3)$$

The MLE for τ_k for our new likelihood equation is

$$\tau_k = \frac{\sum_{i=1}^n w_i \hat{z}_{ik}}{n}. \quad (4.4)$$

However, there is a potential problem that the sum of the τ_k s over all k will not equal 1. Remember from section 2.1 that τ_k represents the probability that a randomly drawn unit is in group k . Thus,

$$\begin{aligned}
\sum_{k=1}^G \hat{\tau}_k &= 1 \\
\frac{\sum_{k=1}^G \sum_{i=1}^n w_i \hat{z}_{ik}}{n} &= 1 \\
\sum_{k=1}^G \sum_{i=1}^n w_i \hat{z}_{ik} &= n
\end{aligned}$$

$$\sum_{i=1}^n w_i \sum_{k=1}^G \hat{z}_{ik} = n$$

$$\sum_{i=1}^n w_i = n$$

Therefore, we must be sure that the sum of our weights, w_i , is equal to the number of units n . To do this we scale our original weights (that may or may not sum to 1) by the factor $\frac{n}{\sum_{i=1}^n w_i}$.

4.1.3 Choosing the Weighting Vector

The most logical choice of weights is to set w_i to the proportion of the total area of the map that unit i contains. Then, the weights truly take into account the differences in area, as each unit will have influence over the cluster parameters exactly proportional to its area. Under these weights, τ_k is the probability that a randomly chosen point on the map is in group k .

However logical this weighting system is, it does not work. The main problem is that there is one huge unit on the map that, when given a weight proportional to its area, has so much control over the parameters of its group that it essentially makes the group its own, stripping off all the units that were initially in that group. This leaves the group with only one member, and its covariance matrix goes to singularity, forcing the algorithm to stop. The units that were initially in that group are allocated to groups in which they probably do not belong, wrongfully changing the parameters of those groups and causing many more unwarranted classifications.

To alleviate this problem we can do one of several things. We can construct our initial vector of weights and then put a restriction on how big any weight can be. This will force the very large units to have more reasonable weights and cause the algorithm to run to completion. However, the value of this bound is highly arbitrary. Also, we lose information about the relationship of the areas of the big groups to one another and information of the actual differences in area between the large groups and smaller groups. Alternatively, we can put a restriction on how small any weight can be. This essentially says that any one unit, no matter how small, has importance of at least this lower bound. Again, this will cause the algorithm to run to completion while sacrificing information about the relationship of smaller units to one another.

Another possibility is creating a number of bins to allocate the units into based on their area. For instance, if we made four bins we could give them weights $1/4$, $1/2$, $3/4$, and 1 and then assign each unit into a bin. This, again, eliminates the problem of extreme weights. However, it is very difficult to decide which units belong to which bin, how many bins there should be, and what weights the bins should receive. Also, by grouping the units we lose information about the ordering of units based on area for those units that are assigned to the same bin. We also lose all information about the actual sizes of the units.

Lastly, we can choose to scale down the weights. For example, we can take the square root of the unit areas and then assign weights as the proportion of these square root areas. This method is good because it maintains the original ordering of the units. The method scales down the actual differences between units, allowing the algorithm to run to completion. A drawback is that the choice of the scaling function (e.g. logs, exponents, etc.) is arbitrary and greatly affects results.

4.1.4 Results

For each clustering result below we used the method of relocating single units discussed in chapter 3.

Bound on Maximum Weight

First, we clustered using an upper bound on the value of the weights. To choose the upper bound, we picked a high value that still made the algorithm continue to convergence. For our data we chose .0075 (.01 caused singular covariance), meaning that the 21 units that had area of more than .75% of the map were given the weight .0075 while the other 179 units kept their original weights. Then, we clustered using the four combinations of variables that we have used throughout this project.

Results for the individual clustering attempts (Table 4.1) show that more units are being reallocated in each clustering attempt than in our nonweighted clusters. Results across clusters (Table 4.2) show that more units (24) are being reallocated every time than in the nonweighted clusters and fewer units (90) are never reallocated. Most of the 24 units that are reallocated every time are very small units. Of these 24, 18 have areas that are less than the median area of a unit (.076% of the map). Most of the reassigned units are *prb* (8). This was a unit type that had very few reassignments in the nonweighted case. Also a few *prc* (4), *pe* (4), and *plb* (3) units were reallocated. Interestingly, only 2 *fe* units were reallocated, much fewer than in the nonweighted clusterings.

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated
means & standard devs., no emissivity	85	42.5%
means & standard devs., no reflectivity	76	38.0%
means, st. devs., and medians, no emissivity	55	27.5%
means, st. devs., and medians, no reflectivity	46	23.0%

Table 4.1: Cluster results for weighted clusters with a bound on the maximum weight.

Number of Times Reallocated	0	1	2	3	4
Number of Units	90	30	32	24	24

Table 4.2: Results for individual units for the weighted clusterings with an upper bound on weight.

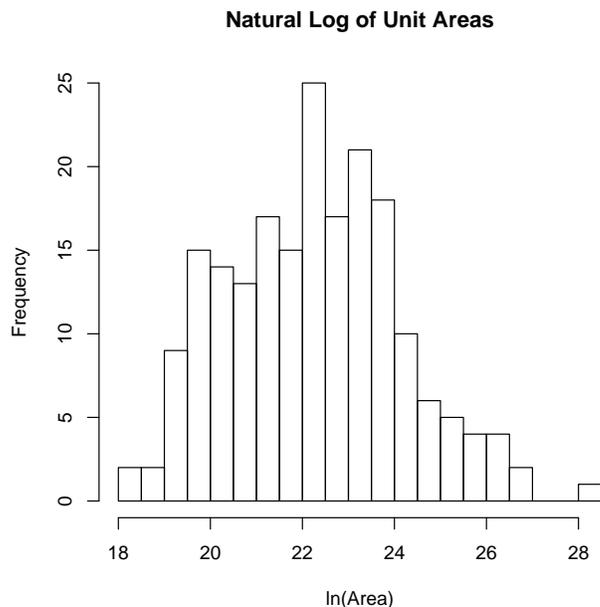


Figure 4.2: Histogram of the natural log of the area for each of the 200 units. Adjusting by natural logs is a conservative way to assign weights, as the difference between the largest and smallest unit is not very big.

Scaled Down Weights: Natural Logs

Next we scaled down the weights using natural logs. We chose natural logs because it is an easy function to implement and it greatly decreases the range of the weights (see figure 4.2). Taking the natural log of the weights may be an overly-drastic adjustment, as it decreases the maximum weight from 28.6% to 0.75%. However, it still is useful because it retains the ordering of the areas and gives bigger units slightly more power to determine the clusters.

In Table 4.4 are the results for our four clustering attempts. Compared to the non-weighted clusterings, more units were kept in the groups as specified by the geologists (106). Also, only 12 units were reallocated every time. Because the weights were not very extreme, these results are fairly similar to the nonweighted results. However, the fact that they are slightly closer to the geologists' classification is reason to believe that heavier weights may result in classifications even closer to the geologists'.

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated
means & standard devs., no emissivity	63	31.5%
means & standard devs., no reflectivity	60	30.0%
means, st. devs., and medians, no emissivity	37	18.5%
means, st. devs., and medians, no reflectivity	30	15.0%

Table 4.3: Clustering results using the natural log scaled weights.

Number of Times Reallocated	0	1	2	3	4
Number of Units	106	36	32	14	12

Table 4.4: Results across clusterings for the natural log scales weights.

Unit Number	Unit Area (meters ²)	Original Classification	Clustering Classifications
50	7.15×10^9	<i>t</i>	<i>pLlb</i> (4)
81	3.09×10^{10}	<i>pe</i>	<i>fe</i> (2), <i>prb</i> (1), <i>plb</i> (1)
102	4.18×10^9	<i>pe</i>	<i>prb</i> (4)
105	2.37×10^{11}	<i>fe</i>	<i>pe</i> (2), <i>pra</i> (1) <i>prc</i> (1)
111	8.81×10^{10}	<i>pe</i>	<i>prb</i> (3), <i>fe</i> (1)
133	1.50×10^{10}	<i>pe</i>	<i>prb</i> (3), <i>fe</i> (1)
136	3.55×10^{10}	<i>fe</i>	<i>prc</i> (2), <i>pra</i> (1), <i>pla</i> (1)
138	1.48×10^{11}	<i>prc</i>	<i>prb</i> (4)
139	1.92×10^9	<i>prc</i>	<i>pe</i> (3), <i>prb</i> (1)
142	6.06×10^{10}	<i>fe</i>	<i>prc</i> (2), <i>pra</i> (1), <i>prb</i> (1)
167	2.07×10^9	<i>fe</i>	<i>prb</i> (2), <i>prc</i> (1), <i>pl</i> (1)
193	3.66×10^8	<i>prc</i>	<i>t</i> (4)

Table 4.5: The 12 units that were reallocated every time are in the table. Only one of these units (unit number 138) was not reallocated by either of the nonweighted clusterings.

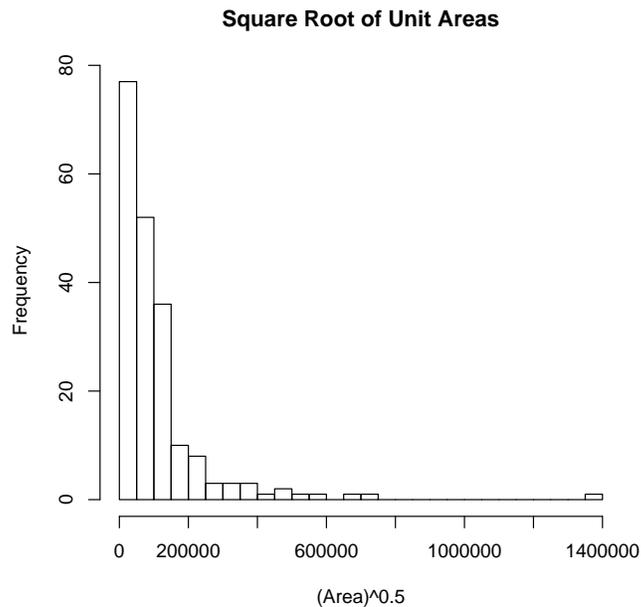


Figure 4.3: Histogram of the square root of the area for each of the 200 units. These weights are less extreme than the original weights (Figure 4.1), but much more extreme than the natural log adjusted weights (Figure 4.2).

Scaled Down Weights: Square Root

Now we scale down the weights with the square root function. The weights produced by this scaling (Figure 4.3) are not as extreme as the unscaled weights but are much more extreme than the natural log scaling. Therefore we expect the weights to have a larger influence on the clustering than in the natural log case and to produce results that are more representative of exactly-weighted areas.

Results for the four clustering attempts are in Tables 4.6 and 4.7. These results are slightly further from the geologists' allocation than the unweighted clusterings, but still very close to the original geologic map. Still, over half of the units (102) are allocated as specified by the geologists every time and only a handful (16) are allocated differently than this specification each time.

The 16 units that were reallocated every time are in Table 4.8. A large number of these units (6) are *prc*. It makes sense that many *prc* units would be reallocated because the large *prc* unit still has a very large weight and continues to dominate the group. Seven of the 16 units were not reallocated by the nonweighted clusterings. This includes one fairly large *plb* unit and a sizable *prc* unit. It is interesting that these units were reallocated under an area-weighted clustering but not a non-weighted clustering because having large size, they

Variables Included in Model	Number of Units Reallocated	% of Units Reallocated
means & standard devs., no emissivity	69	34.5%
means & standard devs., no reflectivity	61	30.5%
means, st. devs., and medians, no emissivity	46	23.0%
means, st. devs., and medians, no reflectivity	33	16.5%

Table 4.6: Clustering results for the square root scaled weights.

Number of Times Reallocated	0	1	2	3	4
Number of Units	102	35	31	16	16

Table 4.7: Results across clusterings for the square root scaled weights.

Unit Number	Unit Area (meters ²)	Original Classification	Clustering Classifications
4	3.02×10^8	<i>c</i>	<i>t</i> (4)
50	7.15×10^9	<i>t</i>	<i>pLlb</i> (4)
95	6.68×10^8	<i>prc</i>	<i>fe</i> (2), <i>t</i> (2)
101	1.21×10^9	<i>pra</i>	<i>prc</i> (2), <i>fe</i> (1), <i>pe</i> (1)
105	2.37×10^{11}	<i>fe</i>	<i>pe</i> (2), <i>prc</i> (2)
111	8.81×10^{10}	<i>pe</i>	<i>prb</i> (3), <i>fe</i> (1)
118	9.91×10^8	<i>prb</i>	<i>pe</i> (3), <i>prc</i>
133	1.50×10^{10}	<i>pe</i>	<i>prb</i> (2), <i>fe</i> (2)
136	3.55×10^{10}	<i>fe</i>	<i>prc</i> (3), <i>pra</i> (1)
139	1.92×10^9	<i>prc</i>	<i>pe</i> (3), <i>fe</i> (1)
160	3.40×10^9	<i>prc</i>	<i>pe</i> (2), <i>fe</i> (1), <i>prb</i> (1)
162	4.01×10^{10}	<i>prc</i>	<i>fe</i> (2), <i>pl</i> (2)
163	1.24×10^{10}	<i>plb</i>	<i>prb</i> (3), <i>fe</i> (1)
167	2.07×10^9	<i>fe</i>	<i>prb</i> (2), <i>pe</i> (1), <i>pl</i> (1)
192	2.15×10^8	<i>prc</i>	<i>t</i> (3), <i>c</i> (1)
193	3.66×10^8	<i>prc</i>	<i>t</i> (4)

Table 4.8: The 16 units that were reallocated every time by the square root scaled clusterings.

have more influence over their clusters in the weighted clustering. Thus, the fact that they are reallocated must mean that their data is significantly different than the other big units in their groups. The other five units that were reallocated by the square-root weighted clusterings but not the non-weighted clusters are all small. Notably, one crater unit (c) is reallocated as tessera. This is the first time we have seen the reallocation of a crater unit. Therefore, the data on this small crater must be significantly different than the data on the larger craters in V14.

4.2 Weighting by Variable

4.2.1 Motivation

In our original model, each of our variables (mean, standard deviation, and median of backscatter and four physical properties) is given the same amount of influence in determining the final clustering. However, each of these variables does not actually have the same importance. For instance, the resolution of the backscatter data is more than 400 times the resolution of the physical property data sets. Also, we expect that the mean of pixel values is probably a better measure of the data than standard deviation of pixel values.

4.2.2 Methods

Saying that we want some variables to have more importance than others is equivalent to saying that we want to change the shapes of our clusters so that they are wider in the directions of less important variables and narrower in the directions of more important variables. To see why this is the case, consider two variables v_1 and v_2 , where v_1 is more important than v_2 . Then, for two data points we do not care about differences in v_2 as much as differences in v_1 . Subsequently, we want to stretch out our clusters in the v_2 direction and constrict them in the v_1 direction so that subtle differences in our more important variable, v_1 , are more important than subtle differences in v_2 .

Recall from chapter 2 that A_k , the diagonal matrix of eigenvalues for Σ_k where $\Sigma_k = \lambda_k D_k A_k D_k^T$ is what controls the shape of the clusters. To introduce variable weights into our model we want to change A_k appropriately. To do this, we simply create a diagonal matrix W , and multiply Σ_k on each side by W . That is, we create a new matrix Σ'_k such that

$$\Sigma'_k = W \Sigma_k W. \tag{4.5}$$

We need to decide the correct W to use to weight the variables. First let us define

$$W = \begin{pmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & w_v \end{pmatrix},$$

where w_h is a scalar for $h = 1, \dots, v$ and v is the number of variables. We want w_h to be large if variable h is not important and w_h to be small if variable h is important. This will stretch out our clusters in the directions of the least important variables and constrict the clusters in the directions of more important variables.

We also have to be careful not to change the areas of the clusters. To do this, we standardize our weights so their product is equal to 1. Then, the determinant of W will be 1, and multiplying Σ_k on either side by W will not change the determinant of Σ_k , which controls the area of cluster k . To perform this standardization, we simply divide each weight by the k^{th} root of the product of the weights.

For example, if we have two variables v_1 and v_2 and v_1 is twice important as v_2 , then we will initially set $w_1 = \sqrt{1/2}$ and $w_2 = 1$. For each weight we take the square root of the ratio because we will multiply Σ_k on both sides by W . Next, we standardize each w , yielding $w_1 = 0.841$ and $w_2 = 1.189$.

4.2.3 Results

When we try to implement the variable weights we are unsuccessful. For all but the most modest of weights (all w_i between 0.9 and 1.1), the algorithm stops after a few iterations due to singularity in the covariance matrices. This happens because we are manually forcing the clusters to take on a certain shape while the data is causing the clusters to have a different shape. Therefore, as the clusters conform to the structure of the data, we continually force them away. Consequently, the clusters do not fit the data well and some groups go to zero unit members quickly.

Weighting variables is not a successful addition to our model. However, it is also not a necessary addition to the model. Left alone, the clusters will conform to the data. Any differences in the data in any variables will be reflected in the final classification. The important differences in the data points will come out in the clusters, so we do not need to explicitly tell the clusters which differences are more important than others.

Chapter 5

Conclusions

In this project our goal was to use the numeric data retrieved by Magellan on the Ganiki Planitia (V14) Quadrangle to analyze an existing geologic map. By doing this, we could cross-validate the geologists' results against the unused physical property data sets of emissivity, reflectivity, and RMS slope. Also, we could cross-validate their results against the numerical data encoded in the images that the geologists used to create the map.

Our ultimate goal was to present to the geologists (1) a set of units whose data consistently does not fit their original classification, and (2) an overall sense of how well the geologic map as a whole fits the data.

In our model, we began with the hypothesis that the existing geologic map was correct. Our clusters, then, will fall at a local maximum of the likelihood equation that is near the original classification. Determining whether or not the geologic map fits the data is a matter of seeing how far away this local maximum is to the starting classification. We will do this by analyzing how many units are allocated differently by our clusters than the geologic map.

The results we will analyze are two nonweighted clusterings (relocation of single units and removal of single units), and two unit-area-weighted clusterings (log scaled weights and square root scaled weights). For any one clustering, between 14.5% and 34.9% of the units were reallocated, showing that a majority of the units are not changing from their original classification in any single clustering. Trends across the clusterings with four different combinations of variables also supported the original geologic map. On average, more than half of all units were not reallocated in any of the four clusterings, giving support of the original classifications of these units. Also, around 70% of the units were reallocated in fewer than half of the four clusterings and about 85% of the units were reallocated either zero times, once, or twice out of the four clusterings. Hence, our results support the original geologic map. We can in turn tell the geologists that as a whole, their results are consistent with the numerical data.

However, there are several units whose original classification does not fit the data. We will consider those units that are reallocated every time by at least one set of clusterings.

Of the five units that are reallocated every time by every set of clusterings (Figure 5.1),

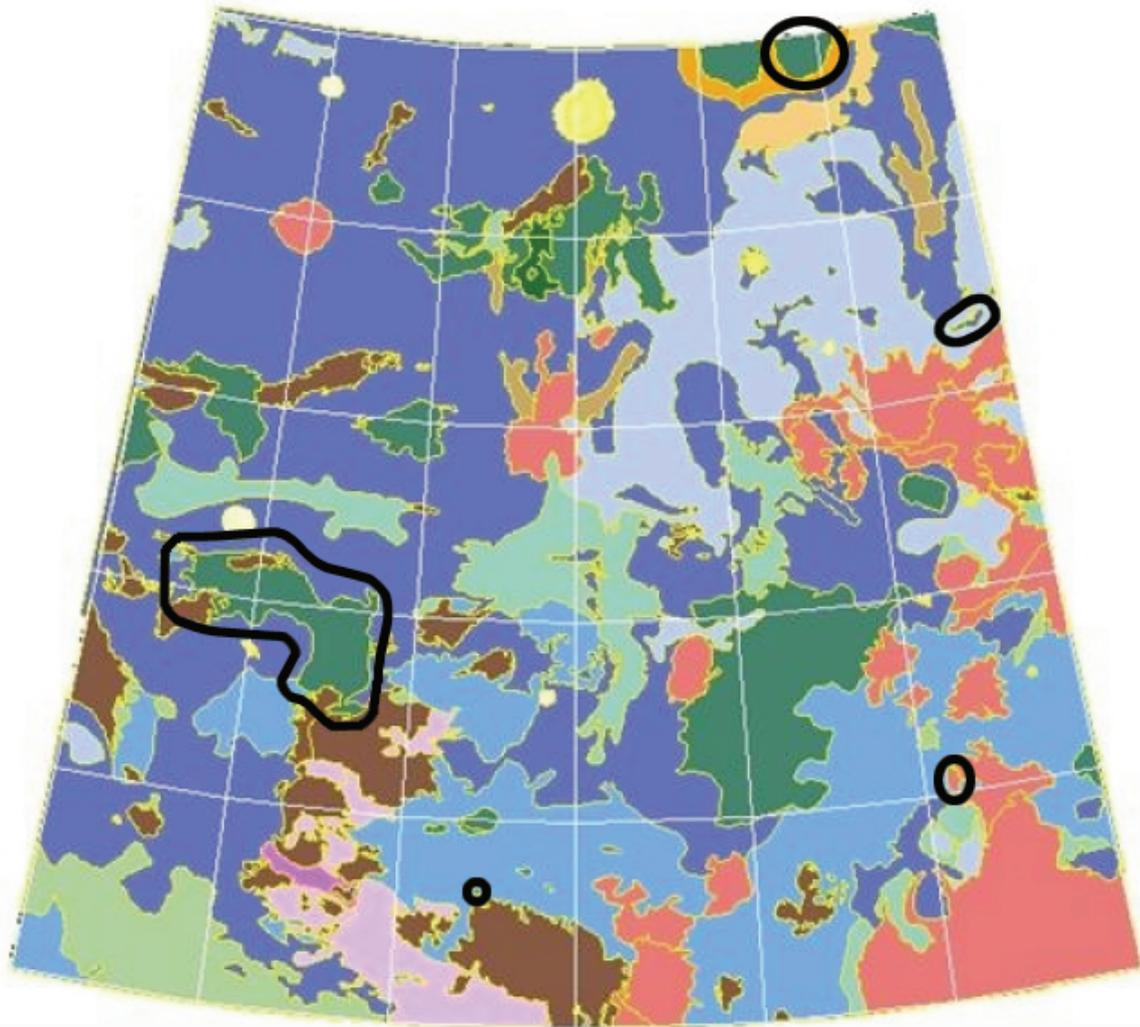


Figure 5.1: The five units that were reallocated every time by all four clusterings (2 without weights, 2 with unit-area weights) are indicated with circles.

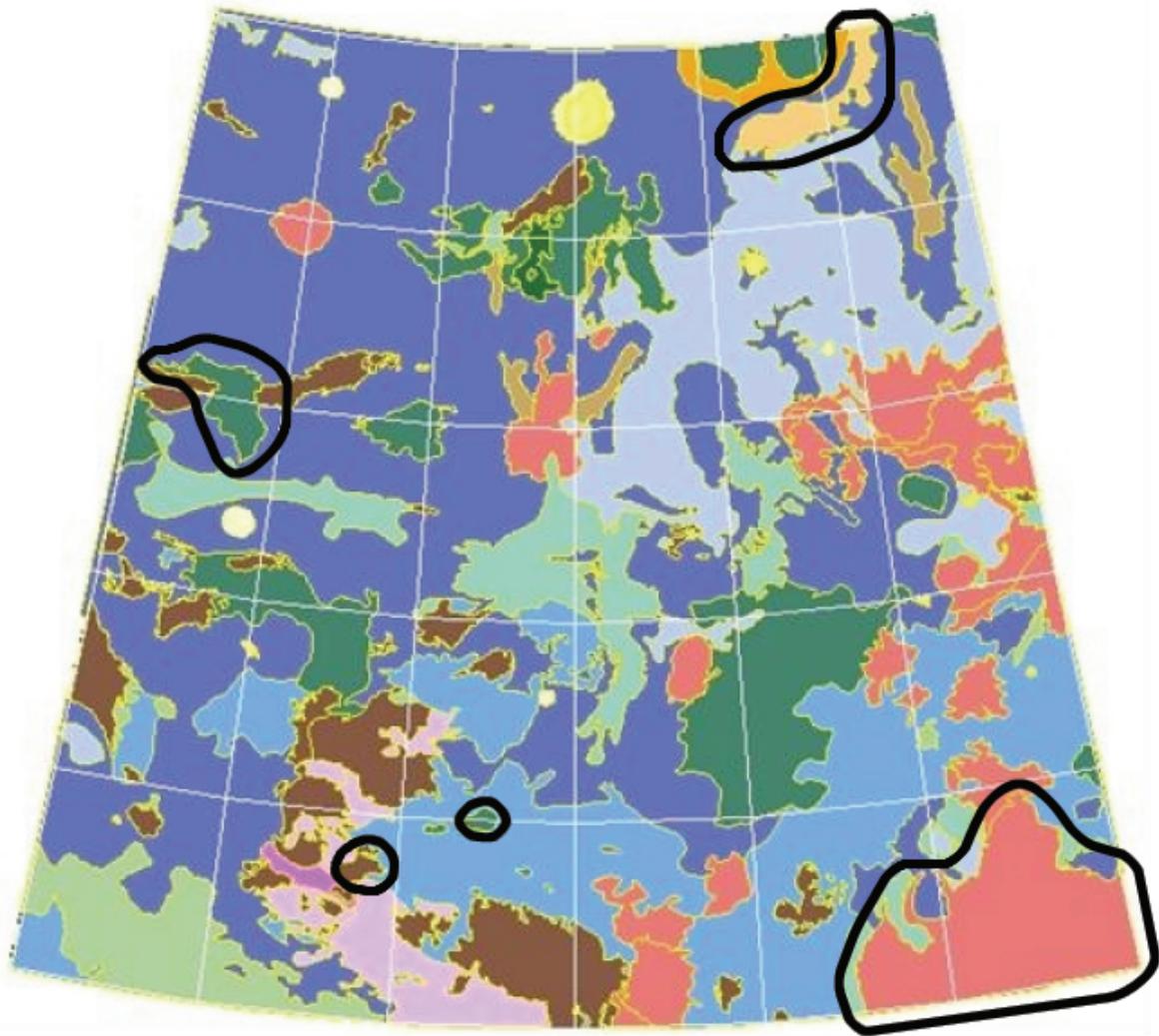


Figure 5.2: The five units that were reallocated by three of the four clusterings are indicated with circles.



Figure 5.3: The two units that were reallocated by two of the four clusterings are indicated with circles.

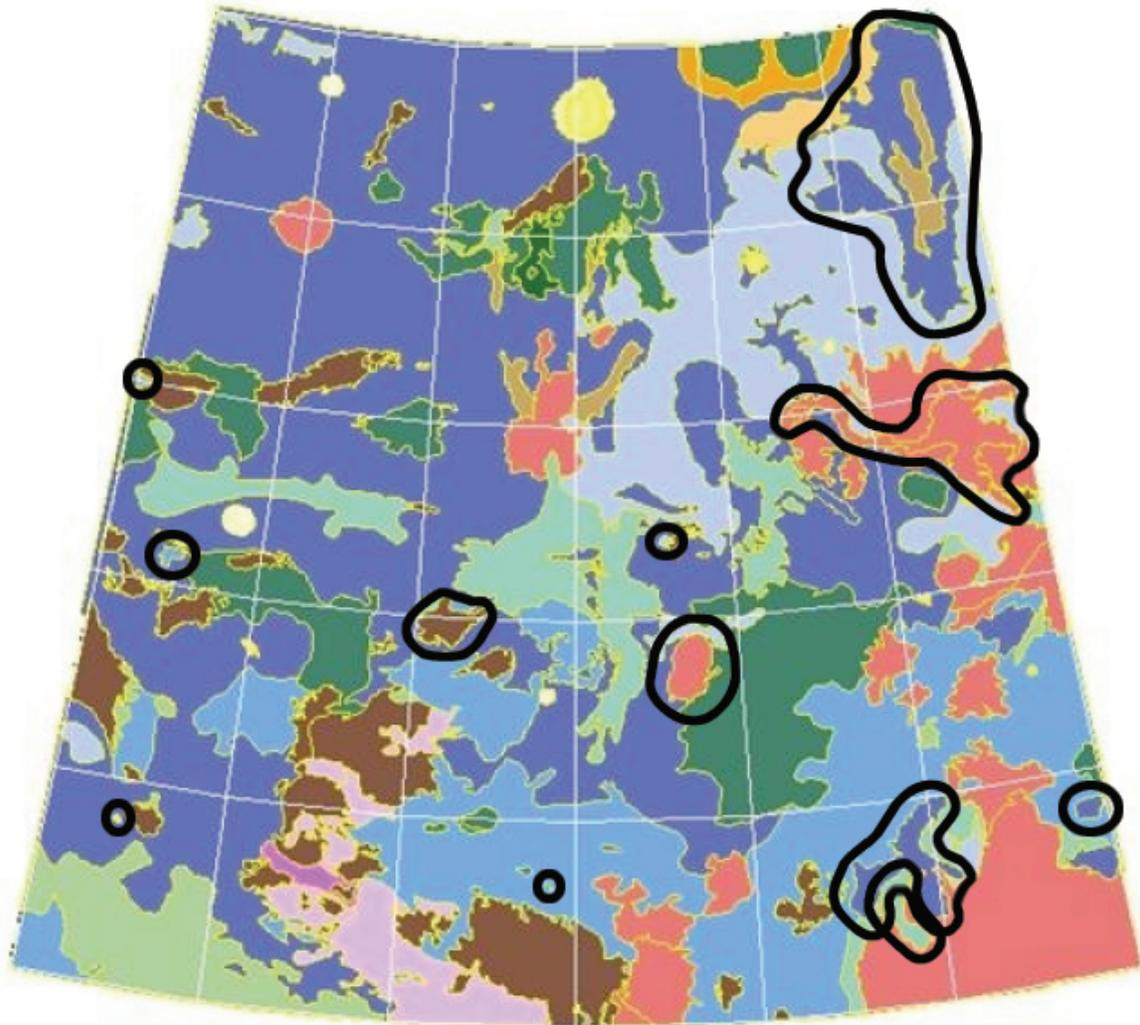


Figure 5.4: The 12 units that were reallocated by one of the four clusterings are indicated with circles.

two are large *pe* (green) units. Due to the significant size of these units, we will advise the geology department that they need to study them more thoroughly. One of the units, a tiny *prc* unit at the bottom of the map, is miniscule in size, so we will ignore it. The other two units are small, but not inconsequential, so we will tell the geologists that they could be reexamined, but it is not imperative that they are.

Five units are reallocated in three of the four clusterings (Figure 5.2). One of these units is the solitary corona flows unit (*cof*) that was relocated to the *fe* group in three of the clusterings and removed from the model in one clustering. A corona flow and a volcanic flow are geologically related, but not exactly the same unit type. This is reflected by the constant reallocation of the *cof* unit from the *fe* group. We cannot conclude, then, that there is anything wrong with the geologists' classification of the *cof* unit, as our results actually support their decision to put this unit in its own group. The other four units that are reallocated three times are a very large *fe* unit, a moderately-large *pe* unit, a small *pe* unit, and a small *t* unit. All of these warrant reexamination by the geologists, as they are all reallocated by at least one nonweighted and at least one area-weighted clustering. The two larger units need to be reexamined because they are so large. The two smaller units could receive further treatment, but it is not necessary that they do.

Two units are reallocated in two of the four clusterings: one large *fe* unit and one small *prc* unit (Figure 5.3). Each unit was reallocated once by a nonweighted clustering and once by a unit-area weighted clustering, showing that the reallocation is robust across the two different models. The *fe* unit is very large and the *prc* unit is small, so we will recommend that the *fe* unit definitely be reanalyzed and that the *prc* could be reanalyzed.

Twelve units are reallocated in one of the four clusterings (Figure 5.4). However, five of these have very small areas and will be ignored. The large *prc* unit in the northeast corner of the map was reallocated in the log-weighted clustering. It was also the only unit that was consistently reallocated when we just considered background plains units. Therefore, we will tell geologists to further study the unit. The large *fe* unit on the eastern part of the map was only reallocated by one of the nonweighted clusterings. Our model has a tendency to reallocate *fe* units because it does not explicitly take into account the geomorphology. Therefore, the geologists' allocation is probably correct, and we will tell them that it is not essential to restudy this unit. The three units in the southeast part of V14 were only reallocated by the square-root weighted map, and we will advise geologists that they could be further studied. The *fe* unit and the *t* units in the middle of the map were only reallocated by one nonweighted clustering, which is also no cause for alarm.

	Unit Number	Unit Area (meters²)	Original Classification	Clustering Classification
Must Reexamine	81	3.09×10^{10}	<i>pe</i>	<i>fe</i> or <i>prb</i>
	105	2.37×10^{11}	<i>fe</i>	<i>pe</i> or <i>prc</i>
	111	8.81×10^{10}	<i>pe</i>	<i>prb</i> or <i>fe</i>
	133	1.50×10^{10}	<i>pe</i>	<i>prb</i>
	138	1.48×10^{11}	<i>prc</i>	<i>prb</i>
	142	6.06×10^{10}	<i>fe</i>	<i>prc</i> , <i>pra</i> , or <i>prb</i>
Can Reexamine	39	1.29×10^{10}	<i>t</i>	<i>fe</i>
	50	7.15×10^9	<i>t</i>	<i>pLlb</i>
	95	6.68×10^8	<i>prc</i>	<i>fe</i> or <i>t</i>
	100	1.63×10^{10}	<i>fe</i>	<i>plb</i> , <i>prb</i> , or <i>prc</i>
	102	4.18×10^9	<i>pe</i>	<i>prb</i>
	139	1.92×10^9	<i>prc</i>	<i>pe</i>
	143	5.57×10^{10}	<i>fe</i>	<i>prb</i>
	160	3.40×10^9	<i>prc</i>	<i>pe</i> , <i>fe</i> , or <i>prb</i>
	162	4.01×10^{10}	<i>prc</i>	<i>fe</i> or <i>pl</i>
	163	1.24×10^{10}	<i>plb</i>	<i>prb</i>
	167	2.07×10^9	<i>fe</i>	<i>prb</i>

Table 5.1: To conclude, we produce a final table of units to be reanalyzed by the geologists. The table is separated into units that the geologists must reexamine and those that they should reexamine. We present six units that the geologists must reexamine and 11 that they should reexamine.

Bibliography

- [1] Celeux, G. and Govaert, G. (1995) Gaussian Parsimonious Clustering Models. *Pattern Recognition*, pp. 781-793.
- [2] Dempster, A.P., Laird, N.M. and Rubin, D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J.R. Statistical Soc. B*, pp. 1-38.
- [3] Everitt, B., et al. (2001) Cluster Analysis, Hodder Headline Group, London, pp. 90-117.
- [4] Fraley, C. and Raftery, A. (2002) Model-based Clustering, Discriminant Analysis, and Density Estimation. *Journal of the American Statistical Association*, June 2002, Vol. 97, pp. 611-631.
- [5] Grosfils, E. B., et.al. (2005) Geologic Evolution of the Ganiki Planitia Quadrangle (V14) on Venus. Lunar and Planetary Science Conference XXXVI.
- [6] Jet Propulsion Laboratory, California Institute of Technology. *Magellan Mission Summary*. www.jpl.nasa.gov/magellan/
- [7] Long, S. M. and Grosfils, E. B. (2005) Quantitative Analysis of Venus Radar Backscatter Data in ArcGIS. Lunar and Planetary Science Conference XXXVI.
- [8] McLachlan, G. and Basford, K. (1988) Mixture Models: Inference and Applications to Clustering. Marcel Dekker, Inc., New York, pp. 1-36.
- [9] McLachlan, G. and Krishnan, T. (1997) The EM Algorithm and Extensions. John Wiley & Sons, Inc., New York, pp. 82-109.
- [10] Tanner, M. (1996) Tools for Statistical Inference. Springer, New York, pp. 64-89.
- [11] Wu, C.F. Jeff (1983) On the Convergence Properties of the EM Algorithm. *The Annals of Statistics*, Mar. 1983, Vol. 11, No. 1, pp. 91-103