Weighted Model-Based Clustering for Remote Sensing Image Analysis

Joseph W. Richards
Department of Statistics
Carnegie Mellon University
Pittsburgh, PA 15213
(jwrichar@stat.cmu.edu)

Johanna Hardin
Department of Mathematics
Pomona College
Claremont, CA 91711
(jo.hardin@pomona.edu)

Eric B. Grosfils
Department of Geology
Pomona College
Claremont, CA 91711
(egrosfils@pomona.edu)

Abstract

We introduce a weighted method of clustering the individual units of a segmented image. Specifically, we analyze geologic maps generated from quantitative analysis of remote sensing images, and provide geologists with a powerful method to numerically test the consistency of a mapping with the entire multidimensional dataset of that region. Our weighted model-based clustering method (WMBC) employs a weighted likelihood and assigns fixed weights to each unit corresponding to the number of pixels located within the unit. WMBC characterizes each unit by the means and standard deviations of the pixels within each unit, and uses the Expectation-Maximization (EM) algorithm with a weighted likelihood function to cluster the units. With both simulated and real data sets, we show that WMBC is more accurate than standard model-based clustering.

KEY WORDS: Weighted likelihood; Mixture model; EM algorithm; Geologic map.

1 INTRODUCTION

As advancements in technology increase our ability to collect massive data sets, statisticians are in constant pursuit of efficient and effective methods to analyze large amounts of information. There is no better example of this than in the study of multi- and hyperspectral images that commonly contain millions of pixels. Powerful clustering methods that automatically

classify pixels into groups are in high-demand in the scientific community. Image analysis via clustering has been used successfully with problems in a variety of fields, including tissue classification in biomedical images, unsupervised texture image segmentation, analysis of images from molecular spectroscopy, and detection of surface defects in manufactured products (see Fraley and Raftery (1998) for more references).

Model-based clustering (Banfield and Raftery 1993; Fraley and Raftery 2002) has demonstrated very good performance in image analysis (Campbell, Fraley, Murtagh, and Raftery 1997; Wehrens, Buydens, Fraley, and Raftery 2004). Model-based clustering uses the Expectation-Maximization (EM) algorithm to fit a mixture of multivariate normal distributions to a data set by maximum likelihood estimation. A combination of initialization via model-based hierarchical clustering and iterative relocation using the EM algorithm has been shown to produce accurate and stable clusters in a variety of disciplines (Banfield and Raftery 1993).

In this paper, we examine the case where manual partitioning of the image has been performed prior to attempts to classify each resulting partition. This situation often arises in the analysis of remote sensing data where geologic maps, divisions of regions of land into units, are created by geologists based on analysis of radar and physical property images (see USGS 2005). In these examples, although the regions are already subdivided into disjoint material units, our goal as statisticians is to allocate

the units into groups defined by the quantitative pixel measurements. Clustering the numeric pixel values permits us to quantitatively evaluate the (usually qualitative) work performed by the geologists, and gives geologists a powerful method to numerically validate their work, compare different geologic maps of the same region, and test the consistency of the defined material units with respect to the entire available multi-dimensional dataset.

A geologic map is meant to convey the mapmaker's interpretation of the region depicted. If multiple geologists map the same area and then compare their results, it is likely that some percentage of their boundaries and unit definitions will be very closely matched, while other areas will bear little resemblance from one map to the next. To improve the mapping process and enhance what can be learned from the maps that are generated, it is necessary to develop new approaches that can be used to evaluate whether material units, defined qualitatively on the basis of geological criteria within a given region, also have robust, self-similar quantitative properties that can be used to characterize the nature of the surface more completely. This is particularly critical for maps generated on the basis of radar data interpretation, as the quantitative properties recorded by the data depend strongly upon the sub-pixel scale physical characteristics of the planet's surface.

The thesis of our paper is that by using the means and standard deviations of the pixel values within each unit of a seg-

mented image, one obtains accurate clustering results from a model-based clustering likelihood that weights each unit by the number of pixels contained within the unit. Using the means and standard deviations of the pixel values simultaneously reduces the size of our data set (from millions of pixels to a few hundreds of groups) and gives information about the central tendencies and variability of the pixels in a unit. Geologically, this combination can yield important quantitative insight into the properties of the surface. For instance, in topography data a smooth, flat plains unit and a highly deformed unit may lie at the same mean elevation, but the high standard deviation for the deformed unit provides a quantitative way to assess the amount and pervasiveness of deformation which has occurred. Similarly, in backscatter data a uniform, flat plains unit formed by regional flooding by lavas may share a mean value with a heavily mottled plains unit formed by overlapping deposits erupted from thousands of small volcanoes but will have distinct variances. In this paper, we show that our weighted clustering method highly outperforms an analogous non-weighted method and generally yields better results than a technique that downweights outliers based on distances (Markatou, Basu, and Lindsay 1998).

In Section 2, we briefly describe model-based clustering and the weighted likelihood function and integrate the two into a weighted model-based clustering method. In Section 3, we design and perform simulations to compare our weighted model-based clustering technique to other model-based clustering tech-

niques in a variety of situations. In Section 4, we apply our technique to a real remote sensing data set. Finally, we conclude with a few comments in Section 5.

2 WEIGHTED MODEL-BASED CLUSTERING (WMBC)

In standard model-based clustering, multivariate observations $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ are assumed to come from a mixture of G multivariate normal distributions with density

$$f(\mathbf{x}) = \sum_{k=1}^{G} \tau_k \ \phi(\mathbf{x}|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_\mathbf{k}), \tag{1}$$

where the τ_k's are the strictly-positive mixing proportions of the model that sum to unity and $\phi(\mathbf{x}|\boldsymbol{\mu}, \boldsymbol{\Sigma})$ denotes the multivariate normal density with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$ evaluated at $\mathbf{x}$.

The general framework for the geometric constraints across clusters was proposed by Banfield and Raftery (1993) through the eigenvalue decomposition of the covariance matrix in the form

$$\boldsymbol{\Sigma}_k = \lambda_k D_k A_k D_k^T, \tag{2}$$

where D_k is an orthogonal matrix of eigenvectors, A_k is a diagonal matrix whose entries are proportional to the eigenvalues, and λ_k is a constant that describes the volume of cluster k. These parameters are treated as independent and can either be constrained to be the same for each cluster or allowed to vary across clusters. For example, the model $\boldsymbol{\Sigma}_k = \lambda_k D_k A D_k^T$ (de-

noted VEV) assumes varying volumes, equal shapes, and varying orientations for each cluster. The completely unconstrained model is denoted VVV. For a thorough discussion of these and other models and the MLE derivation for $\mathbf{\Sigma}$, see Celeux and Govaert (1995).

Starting with some initial partition of the n units into G groups, we use the Expectation-Maximization (EM) algorithm (Dempster, Laird, and Rubin 1977; McLachlan and Krishnan 1997) to update our partition such that the parameter estimates of the clusters maximize the mixture likelihood. Hierarchical agglomeration has been used successfully to obtain an initial partition (Banfield and Raftery 1993) . The EM algorithm iterates between an M-step and an E-step. The M-step calculates the cluster parameters $\boldsymbol{\mu}$, $\mathbf{\Sigma}$ and $\boldsymbol{\tau}$ using the maximum likelihood estimates (MLEs) of the complete-data loglikelihood,

$$l(\boldsymbol{\mu}, \mathbf{\Sigma}, \boldsymbol{\tau}|\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n}\sum_{k=1}^{G} z_{ik}[\log(\tau_k \ \phi(\mathbf{x}_i|\boldsymbol{\mu}_k, \mathbf{\Sigma}_k))] \tag{3}$$

based on the current allocation of the units into groups, $\mathbf{z}$. These MLEs are

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^{n} \hat{z}_{ik}\mathbf{x}_i}{\sum_{i=1}^{n} \hat{z}_{ik}}, \tag{4}$$

$$\hat{\tau}_k = \frac{\sum_{i=1}^{n} \hat{z}_{ik}}{n}, \tag{5}$$

and a model-dependent estimate of $\hat{\mathbf{\Sigma}}_k$ (Celeux and Govaert 1995). The E-step calculates the conditional probability that a unit $\mathbf{x}_i$ comes from the k^{th} group using the equation

$$\hat{z}_{ik} = \frac{\hat{\tau}_k \ \phi(\mathbf{x}_i|\hat{\boldsymbol{\mu}}_k, \hat{\mathbf{\Sigma}}_k)}{\sum_{j=1}^{G} \hat{\tau}_j \ \phi(\mathbf{x}_i|\hat{\boldsymbol{\mu}}_j, \hat{\mathbf{\Sigma}}_j)}, \tag{6}$$

based on the current cluster parameters. The M-E iteration continues until the value of the loglikelihood function converges.

In standard model-based clustering (SMBC), each data point is given equal importance in the model. However, there are situations in which some data points are more accurately measured than others, and therefore deserve higher weight in the model. For example, in segmented pixelated data, those units with more pixels will have means and standard deviations that better approximate the true parameters of the underlying distribution. In SMBC, the ability of data point $\mathbf{x}_i$ to determine the parameters of cluster k only depends on z_{ik}, the posterior probability that the unit belongs to that group. To give units unequal weights, we introduce the weighted likelihood (WL) (Newton and Raftery 1994; Markatou et al. 1998; Agostinelli and Markatou 2001), where each data point receives a fixed weight, $w_i \in (0, 1]$ based on the number of pixels located inside the unit, where higher weights give more influence in estimating the parameters. In general, the WL function for n independent data points is

$$\tilde{L}(\theta) = \prod_{i=1}^{n} f_i(x_i|\theta)^{w_i}, \tag{7}$$

where f_i is the density function for point x_i and θ is a set of parameters. The weighted maximum likelihood estimator (WLE) has been shown to be consistent and asymptotically normal under fixed weights (Wang, van Eeden, and Zidek 2004).

The weighted mixture model loglikelihood equation (Marka-

tou 2000) is

$$\tilde{l}(\boldsymbol{\mu}, \boldsymbol{\Sigma}, \boldsymbol{\tau}|\mathbf{x}, \mathbf{z}) = \sum_{i=1}^{n} \sum_{k=1}^{G} w_i z_{ik} [\log(\tau_k \ \phi(\mathbf{x}_i|\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_\mathbf{k}))], \quad (8)$$

whose only difference from (3) is the additional weights, w_i. As in SMBC, weighted model-based clustering (WMBC) begins with some partition of the data points and proceeds to the M-step, where the WLEs are computed. For each $k = 1, \ldots, G$, the WLE for $\boldsymbol{\mu}_k$ is

$$\hat{\boldsymbol{\mu}}_k = \frac{\sum_{i=1}^{n} w_i \hat{z}_{ik} \mathbf{x}_i}{\sum_{i=1}^{n} w_i \hat{z}_{ik}}, \quad (9)$$

compared to the MLE for $\boldsymbol{\mu}_k$, (4). Similarly, the WLE for the mixing proportion τ_k is

$$\hat{\tau}_k = \frac{\sum_{i=1}^{n} w_i \hat{z}_{ik}}{\sum_{i=1}^{n} w_i}, \quad (10)$$

compared to the MLE for $\tau_\mathbf{k}$, (5), while the WLE of the covariance matrix depends on the model selected. The E-step uses these estimates exactly as in the standard E-step (6), and the algorithm continues until the weighted loglikelihood (8) converges.

3 SIMULATED DATA

3.1 Simulation Design

Before using our WMBC technique to cluster real data sets, we first use simulated data to compare the accuracy of WMBC clusters to those of other model-based clustering techniques in a variety of situations. In each simulation, we generate several

units, where each unit consists of a random number of pixels generated from a uniform [500,50000] distribution and each pixel is assigned a value from a predefined bivariate normal distribution.

We are justified in simulating the pixel values with a normal distribution (when in actuality pixel values need not be distributed normally) because the data summaries we use in the mixture likelihood are the means and standard deviations of these pixels. Regardless of the distribution of the pixel values, their mean is asymptotically normally distributed by the Central Limit Theorem, and by a combination of Slutsky's Theorem, the Central Limit Theorem, and the Delta Method, their standard deviation is also asymptotically normally distributed. Therefore, no matter the distribution of the pixel values, a multivariate normal mixture model is appropriate for modeling the summary statistics used in clustering the units.

We simulate units from different bivariate normal distributions corresponding to different groups. Since we are simulating the data, we know from which distribution (population) each data point is generated. Therefore we can compare different clustering techniques by comparing the number of points that are correctly-classified in each. Throughout this section we assume that the number of groups is known, and we initialize the clusters with unsupervised model-based hierarchical classification. We use the covariance model VEV described in Section 2.

3.2 Two Cluster Simulations

In this section, we compare WMBC to SMBC for situations where there are two groups (i.e. unit types). In each trial we simulate 100 units from each of two bivariate normal distributions. These distributions have parameters

$$\boldsymbol{\mu}_1 = \begin{bmatrix} x \\ 5 \end{bmatrix}, \boldsymbol{\Sigma}_1 = \begin{bmatrix} 180 & r_1\sqrt{180 * 170} \\ r_1\sqrt{180 * 170} & 170 \end{bmatrix}$$

$$\boldsymbol{\mu}_2 = \begin{bmatrix} 4 \\ 5 \end{bmatrix}, \boldsymbol{\Sigma}_2 = \begin{bmatrix} 170 & r_2\sqrt{170 * 160} \\ r_2\sqrt{170 * 160} & 160 \end{bmatrix}$$

where r_1 and r_2 are independent, random (uniform on -1 to 1) correlations, and x takes on each of 21 values ranging from 2 to 4, in steps of 0.1. For each of these 21 spacings of the means of the two groups, we generate 1000 data sets and cluster each one using both the weighted and standard model. Because we cluster each data set with both WMBC and SMBC, we can directly compare the two techniques for a variety of situations (ranging from widely spaced to heavily overlapping clusters).

Results show that WMBC is more accurate for each separation of the means of the two groups, and is far superior than SMBC when the groups are closer together. Table 1 reveals that for each separation in the two groups, the average number of correct classifications for WMBC is greater than the average number of correct classifications for SMBC, and each difference is significant at the 0.0001 level using both a paired t-test and a non-parametric paired Wilcoxon test. Figure 1 shows that

for each of the 21 separations of the group means, WMBC produces a more accurate clustering than SMBC in a higher proportion of data sets than vice versa. When cluster means are close together, WMBC is highly superior, averaging more than 4.5 more correctly-classified units per data set and better clusterings in over 75% of simulations. When clusters are widely-spaced, WMBC is also significantly better but loses much of its superiority because the majority of simulations result in ties between WMBC and SMBC.

WMBC performs better than SMBC because it is not easily distracted by outlying data points. Outliers generally come from data generated from a small number of pixels, and thus are downweighted by WMBC, and largely ignored by the clusters. In SMBC, however, clusters react more strongly to outliers, growing in volume and subsequently claiming points that belong to other groups. When clusters are close or overlapping, outliers can cause a cluster to grow to encompass a large part of another cluster, producing a highly erroneous classification. In WMBC this is avoided because points with large weights are generated from many pixels, and thus are extremely likely to be near the true cluster center. When clusters are widely spaced, the advantage enjoyed by WMBC is somewhat lost, as clusters are less likely to grow so much as to claim data points belonging to another cluster.

Next, using the same simulation model described above, we simulate clusters of several different sizes to show that WMBC

is superior to the SMBC under varied conditions. To simplify our results, instead of considering all 21 spacings of the clusters as we did above, we will only look at three: widely spaced (separation of means of 1.5), intermediately spaced (separation of 0.7), and overlapping (separation of 0.1).

When there are an equal number of units in each group, WMBC produces more accurate classifications than SMBC for each of several group sizes (Table 2). For each separation in the centers of the groups, a much higher percentage of the simulations result in more accurate clusters by the WMBC method. The average number of correct classifications is higher for the weighted method in each simulation and for all but the smallest group size (10) is significant at the 0.0001 level using a paired Wilcoxon test. Again, WMBC performs best when the cluster centers are very close together.

When the groups have an unequal number of units, we again observe that WMBC outperforms SMBC(Table 3). In each simulation, we randomly assigned which group had more data points. The mean number of correct classifications was greater for the weighted method in every situation, with larger discrepancies when the clusters overlapped, and each was significant at the 0.0001 level.

3.3 Distance Weights

A weighted-likelihood model that downweights observations inconsistent with the model (outliers) was introduced by Marka-

tou et al. (1998). They introduce weights based on the Pearson residual, δ, where the weights are defined as

$$w(\delta) = 1 - \frac{\delta^2}{(\delta + 2)^2}. \tag{11}$$

The weights take on values on the interval [0,1], with smaller weights corresponding to data points with high Pearson residuals. For a thorough discussion of the construction of the weight equation, see Markatou et al. (1998).

We compare a clustering method that weights based on Mahalanobis distance (DW) using (11) to our pixel-weighting technique (PW). Like the DW technique, PW downweights outliers, since any point that is an outlier is likely to come from a unit with a small number of pixels. Hence, we postulate that these two methods will produce similar results.

Results in Table 4 show that relative performances of the two methods are dependent on the amount of separation in the clusters. When the clusters are widely spaced, DW tends to do better: in 5 of the 6 simulations DW had a higher average number of correct classifications than PW. However, only one of these simulations yielded a significant result at the 0.1 level (simulation with 2 groups of 20 units each). Additionally, over 96% of the simulations resulted in ties in each widely-spaced comparison. When the clusters are intermediately-spaced, PW outperformed DW in 5 of the 6 simulations, and produced significant differences at the 0.05 level in each of these five. When the clusters were closely spaced, PW outperformed DW in all six simulations, with significant differences in 5 of the 6 at the

0.0001 level.

Overall, PW outperformed DW: in 10 of our simulations PW yielded significantly better results (at the 0.05 level) as compared to only 2 simulations where DW significantly outperformed PW. Relative advantage in PW depends largely on the spacing in the clusters. Highly-spaced clusters produce insignificant advantages for DW, while closer clusters give significant and highly-significant advantages to PW. There was one anomalous situation, where the two group sizes were 20 and 20, in which DW consistently performed consistently better than PW.

A critical drawback to DW is that it requires many more iterations to converge. In 100 simulations, it took PW an average of 7.49 iterations to converge and DW an average of 18.68 iterations. Also, because the weights in DW are based on the Mahalanobis distance from each data point to the center of its cluster, these values continually change as points are reallocated and covariance matrices change and thus have to be recalculated, causing each iteration to take longer. The changing weights also account for the difficulty of the algorithm to converge. For example, if a point is reallocated, it will cause its new cluster to stretch somewhat in its direction, subsequently causing the point's Mahalanobis distance to decrease and its weight to rise. On the next iteration, the point's higher weight will cause the cluster to stretch even more and the pattern to continue, resulting in clusters that are more unstable and less accurate than those produced by the fixed-weight, PW method.

3.4 Three Cluster Simulations

We also applied our method to the situation with three clusters. As before, we considered three possibilities: highly spaced clusters, intermediately spaced clusters, and overlapping clusters. We compared our method to the standard, unweighted model-based clustering method for a variety of different sample sizes.

Again, WMBC is superior to SMBC (Table 5). For each situation, WMBC outperforms SMBC at a highly significant level. Also, WMBC is particularly good when groups are large and/or overlapping. These results are important because in most circumstances, including the remote sensing example in Section 4, there will be more than two groups present.

4 EXAMPLE: MAGELLAN VENUS DATA

4.1 Data Background

On May 4, 1989 the National Aeronautics and Space Administration (NASA) launched the Magellan Spacecraft to study the surface of Venus. From September 15, 1990 until September 14, 1992, Magellan radar-mapped 97% of the planet's surface at resolutions that were ten times better than any previous mapping of the planet, transmitting back to Earth more data than that from all past planetary missions combined (Saunders et al.

1992). A set of about 30,000, 1024 x 1024 pixel, synthetic aperture radar (SAR), 75m/pixel resolution images were transmitted by Magellan.

The Ganiki Planitia V14 quadrangle (180°-210° E, 25°-50° N) is a section of Venus that has been studied by geologists (Grosfils et al. 2005) as part of a global mapping effort (see USGS 2003). Situated between regions where extensive tectonic and volcanic activity has occurred in the past, Ganiki Planitia consists of what are interpreted as volcanically-formed plains which embay older units and are themselves modified by tectonic, impact and volcanic processes. Before studying complex geological issues such as whether there have been systematic changes in the volcanic and tectonic activity in the V14 quadrangle over time, a working geologic map of the region was created on the basis of standard geological criteria, dividing the continent-sized area into 200 material units (Figure 3).

To create the geologic map (e.g., Grosfils et al. (2005)), standard planetary mapping techniques (use of crosscutting and superposition relationships, unit geomorphology, etc.) were used to analyze the full resolution SAR map (called the FMAP) of V14 as well as four physical property data images; however, the numerical information encoded in the data was not used quantitatively when defining the material units. The FMAP for V14 is a mosaicked SAR data set consisting of 131,316,652 pixels. The physical property data sets are: surface reflectivity (gredr), emissivity (gedr), elevation (gtdr), and RMS slope (gsdr), and

each contain between 380,585 and 382,324 pixels. See Figure 2 for the pixelated FMAP and three physical property data sets. We will only consider three of the physical property datasets: gedr, gtdr, and gsdr, because gredr and gedr are close to inversely proportional.

Throughout this section we will take the geologists' classification (Figure 3) to be correct. Then, we can compare the accuracy of WMBC and SMBC by observing how close the clusters are to the geologists' classification. Plots of the raw data show that clusters overlap heavily, and are essentially indiscernible to the eye (Figure 4). Hence, we expect that WMBC will outperform SMBC, as it did in simulations where clusters were extremely close together.

4.2 Clustering Entire Data Set

Starting from the geologists' classification, we cluster the 200 units and observe the error rate for different methods. The material units on V14 vary widely in size, as the largest unit has 22,000 times the number of FMAP pixels than the smallest. Moreover, the areas of the units are very highly skewed: there are a handful of units that are extremely large compared to the mean size (Figure 5 (a)). If we assign weights directly proportional to unit area, the very large units are given weights that completely dominate over the vast majority of material units, rendering extremely insignificant the ability of small and even medium-sized units to affect group parameters. To allevi-

ate this, we take the log of the pixel weights before clustering, which results in a symmetric distribution of weights (Figure 5 (b)) and preserves the order of the unit areas. Clustering under this weighting system results in WMBC clusters that have a lower error percentage than SMBC clusters (Table 6).

We also attempt to cluster the geologic material units starting with a hierarchical classification. However, because the clusters are so close together, hierarchical initialization tends to place most units into one group. Consequently, the final clusters are not very accurate when compared to the geologists' classification. However, WMBC slightly outperforms SMBC (Table 7). To compare the hierarchical-initialized clusterings to the geologists' classification, we use the adjusted Rand statistic (Hubert and Arabie 1985). The adjusted Rand statistic compares any two classifications of the same data set, with higher values signifying closer concordance.

4.3 Clustering Background Plains

One important problem on the V14 quadrangle is classifying its 54 background plains units. Background plains, inferred to be of volcanic origin, dominate V14, containing 62.3% of the pixels of the FMAP. They are divided into three types: a, b, and c, corresponding to three general states of appearance (caused by surface morphology, modification, etc.) in the radar backscatter images. Determining which units belong to each type is important to constrain the characteristics and possibly the evolution

of each unit. However, it is also a difficult problem because it is primarily based on a geologist's interpretation of the brightness of the FMAP image.

We clustered the background plains units with WMBC and SMBC. Again, because of the presence of a very large unit, we used the log of the pixel weights in WMBC. Results show extremely close concordance of clustering and geologist classifications for both techniques (Table 6), with no advantage for either.

5 CONCLUSIONS

In this paper, we have introduced a weighted model-based clustering method that can be used to classify groups of pixels in previously-segmented images by employing the means and standard deviations of the pixel values within each segment. We have shown, with both simulated and real data sets, that one obtains more accurate clustering results using our WMBC method than with SMBC. WMBC is superior to SMBC in the segmented-image context because it both ignores outliers and strongly-defines cluster centers. It performs comparatively best when cluster centers are close because whereas SMBC clusters tend to merge into one another, WMBC clusters have a stronger ability to stay separated since they pay stronger attention to those points situated near the true group center.

Weighted mixture models that downweight outliers based on distance had previously been introduced (Markatou et al.

1998). However, our method is preferable because it produces more accurate results for close and overlapping clusters, and because it uses fixed weights, creates more stable results and converges in fewer iterations.

Our method is a powerful tool for planetary mappers who wish to numerically validate their qualitative analyses. The results from the application of WMBC to the V14 quadrangle demonstrate that most units remain classified the same way as specified by the original geologic map, meaning, for example, that all areas mapped as background plains b units (prb) quantitatively resemble one another more than any of the other unit types mapped. Under WMBC, 41 units (20.5% of the total) were assigned to different groups, and for each case the geologists then examined the unit to determine if it had been mapped incorrectly. In all but one instance, misclassification resulted when a geologically important piece of information integrated into definition of the unit during the mapping process was not quantitatively distinctive enough to be perceived by the statistical algorithm. For instance, five units created by extensive flow of lavas from a large but very flat central edifice were reclassified as regional plains units because in each instance the topography was gentle enough that the presence of the edifice was not detected quantitatively. Similarly, plains characterized by overlapping systems of eruptions from small (1-10 km diameter) shield volcanoes were in some instances reclassified because the subtle morphology of the small shield volcanoes yields no

quantitatively robust signature with which the classification algorithm can work.

Ultimately, while user insight is still required to examine any possible misclassifications that get called out, the strength of the statistical technique we have developed is that it quantitatively uses all available raster data to test the internal self-consistency of the map units defined within the quadrangle. This is of great value to the mappers, demonstrating for the first time that each type of unit is statistically distinctive from all the others when the full suite of quantitative data at our disposal is employed, and thus validating independently the robustness of the material units defined qualitatively using standard geological mapping techniques.

Our method can only be used with previously-segmented images, such as geologic maps, and therefore relies heavily on the initial partitioning of an image. It is primarily used to assess and analyze work that has already been manually performed instead of as a tool to automatically classify pixels. However, it can be a powerful tool for planetary geologists that desire to numerically analyze the classification of geologic units by standard, non-quantitative analysis and determine if the material units, as defined, are consistent with the total available set of numeric data.

References

[1] Agostinelli, C., and Markatou, M., (2001), "Test of Hypotheses Based on the Weighted Likelihood Methodology," *Statistica Sinica*, 11, 499-514.

[2] Banfield, J. D., and Raftery, A. E., (1993), "Model-Based Gaussian and Non-Gaussian Clustering," *Biometrics*, 49, 803-821.

[3] Campbell, J. G., Fraley, C., Murtagh, F., and Raftery, A. E., (1997), "Linear Flaw Detection in Woven Textiles Using Model-Based Clustering," *Pattern Recognition Letters*, 18, 1539-1548.

[4] Celeux, G., and Govaert, G., (1995), "Gaussian Parsimonious Clustering Models," *Pattern Recognition*, 28, 781-793.

[5] Dempster, A. P., Laird, N. M., and Rubin, D. B., (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society. Series B (Methodological)*, 39, 1-38.

[6] Dupuis, D. J., and Morgenthaler, S., (2002), "Robust weighted likelihood estimators with an application to bivariate extreme value problems," *The Canadian Journal of Statistics*, 30, 17-36.

[7] Fraley, C., and Raftery, A. E., (1998), "How Many Clusters? Which Clustering Method? Answers Via Model-Based Cluster Analysis," *The Computer Journal*, 41, 378-388.

[8] ——— (2002), "Model-Based Clustering, Discriminant Analysis, and Density Estimation," *Journal of the American Statistical Association*, 97, 611-631.

[9] Green, P. J., (1984), "Iteratively Reweighted Least Squares for Maximum Likelihood Estimation, and some Robust and Resistant Alternatives," *Journal of the Royal Statistical Society. Series B (Methodological)*, 46, 149-192.

[10] Grosfils, E. B., Drury, D. E., Hurwitz, D. M., Kastl, B., Long, S. M., Richards, J. W., and Venechuk, E. M., (2005),

"Geological Evolution of the Ganiki Planitia Quadrangle (V14) on Venus, Abstract No. 1030," *LPSC, XXXVI.*

[11] Hu, F., and Zidek, J. V., (2002), "The Weighted Likelihood," *The Canadian Journal of Statistics*, 30, 347-371.

[12] Hubert, L., and Arabie, P. (1985), "Comparing Partitions," *Journal of Classification*, 193-218.

[13] Markatou, M., (2000), "Mixture Models, Robustness, and the Weighted Likelihood Methodology," *Biometrics*, 56, 483-486.

[14] Markatou, M., Basu, A., and Lindsay, B. G., (1998), "Weighted Likelihood Equations With Bootstrap Root Search," *Journal of the American Statistical Association*, 93, 740-750.

[15] McLachlan, G. J., and Krishnan, T., (1997), *The EM Algorithm and Extensions*, New York, NY: John Wiley & Sons, Inc.

[16] Newton, M. A., and Raftery, A. E., (1994), "Approximate Bayesian Inference with the Weighted Likelihood Bootstrap," *Journal of the Royal Statistical Society. Series B (Methodological)*, 56, 3-48.

[17] Rukhin, A. L., and Vangel, M. G., (1998), "Estimation of a Common Mean and Weighted Means Statistics," *Journal of the American Statistical Association*, 93, 303-308.

[18] Saunders, R. S., Spear, A. J., Allin, P. C., Austin, R. S., Berman, A. L., Chandlee, R. C., Clark, J. deCharon, A. V., De Jong, E. M., Griffith, D. G., Gunn, J. M., Hensley, S., Johnson, W. T. K., Kirby, C. E., Leung, K. S., Lyons, D. T., Michaels, G. A., Miller, J., Morris, R. B., Morrison, A. D., Piereson, R. G., Scott, J. F., Shaffer, S. J., Slonski, J. P., Stofan, E. R., Thompson, T. W., and Wall, S. D., (1992), "Magellan mission summary," *Journal of Geophysical Research*, 97, 13067-13090.

[19] U. S. Geological Survey, (2003), "USGS Astrogeology: Planetary Geologic Mapping Home Page," http://astrogeology.usgs.gov/Projects/PlanetaryMapping/.

[20] ———, (2005), "USGS National Geologic Map Database," ngmdb.usgs.gov/.

[21] Wang, X., van Eeden, C., and Zidek, J. V., (2004), "Asymptotic properties of maximum weighted likelihood estimators," *Journal of Statistical Planning and Inference*, 119, 37-54.

[22] Wang, X., and Zidek, J. V., (2005), "Selecting Likelihood Weights by Cross-Validation," *The Annals of Statistics*, 33, 463-500.

[23] Wehrens, R., Buydens, L. M. C., Fraley, C., and Raftery, A. E., (2004), *Journal of Classification*, 21, 231-253.

Table 1: Comparison of the accuracy of WMBC versus SMBC for 21 different separations of the means of the two groups. There are 200 total units in each simulation. Averages are from 1000 simulated data sets. Standard deviations are in parentheses.

Separation of group means	Average number of correct classifications		
	WMBC	SMBC	Difference *
2.0	199.957 (0.208)	199.854 (0.524)	0.103
1.9	199.924 (0.273)	199.800 (0.655)	0.124
1.8	199.940 (0.280)	199.764 (0.733)	0.176
1.7	199.923 (0.278)	199.721 (0.823)	0.202
1.6	199.888 (0.346)	199.728 (0.723)	0.16
1.5	199.857 (0.398)	199.627 (0.888)	0.23
1.4	199.829 (0.427)	199.507 (1.050)	0.322
1.3	199.778 (0.507)	199.443 (1.123)	0.335
1.2	199.735 (0.541)	199.336 (1.208)	0.399
1.1	199.686 (0.571)	199.094 (1.570)	0.592
1.0	199.602 (0.650)	198.895 (1.717)	0.707
0.9	199.501 (0.771)	198.634 (1.852)	0.867
0.8	199.377 (0.852)	198.291 (2.281)	1.086
0.7	199.232 (0.888)	197.738 (2.957)	1.494
0.6	198.899 (1.244)	196.904 (3.526)	1.995
0.5	198.689 (1.394)	196.239 (4.028)	2.45
0.4	198.451 (1.632)	195.458 (4.610)	2.993
0.3	198.281 (1.584)	194.690 (5.101)	3.591
0.2	197.807 (2.105)	193.596 (5.645)	4.211
0.1	197.577 (2.214)	193.062 (6.207)	4.515
0.0	197.490 (2.537)	192.873 (6.584)	4.617

*Each difference significant at 0.0001 for two-sided paired t-test and paired Wilcoxon test

Table 2: Percentage of simulations (out of 1000) each clustering method outperformed the other for various equal-sized groups. Groups are widely-spaced (a), intermediately spaced (b), and overlapping (c).

(a)

	% of times better		average diff. in # of correct	two-sided p-value
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC)	(Paired Wilcoxon)
90	18.3	2.7	0.247	< 0.0001
80	15.1	2.7	0.203	< 0.0001
70	14.2	2.8	0.178	< 0.0001
60	13.2	1.4	0.232	< 0.0001
50	13.7	1.7	0.224	< 0.0001
40	13.0	1.4	0.196	< 0.0001
30	13.7	1.0	0.194	< 0.0001
20	8.2	0.9	0.094	< 0.0001
10	1.0	0.4	0.006	0.117

(b)

	% of times better		average diff. in # of correct	two-sided p-value
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC)	(Paired Wilcoxon)
90	47.2	7.9	1.318	< 0.0001
80	47.2	4.5	1.304	< 0.0001
70	40.5	6.3	0.972	< 0.0001
60	39.5	5.8	0.898	< 0.0001
50	38.7	5.6	0.817	< 0.0001
40	31.4	4.8	0.588	< 0.0001
30	27.2	4.6	0.412	< 0.0001
20	17.6	3.7	0.205	< 0.0001
10	3.5	2.1	0.022	0.051

(c)

	% of times better		average diff. in # of correct	two-sided p-value
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC)	(Paired Wilcoxon)
90	70.9	6.0	3.948	< 0.0001
80	73.0	6.5	3.825	< 0.0001
70	66.7	6.3	3.050	< 0.0001
60	62.6	7.5	2.488	< 0.0001
50	58.2	7.6	1.916	< 0.0001
40	54.3	7.0	1.500	< 0.0001
30	41.1	7.6	0.852	< 0.0001
20	28.0	7.5	0.335	< 0.0001
10	5.2	4.6	0.331	0.736

Table 3: Percentage of simulations (out of 1000) each clustering method outperformed the other for six uneven groups. Groups are widely-spaced (a), intermediately spaced (b), and overlapping (c).

(a)

	% of times better		average diff. in # of correct
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC) *
75 / 25	15.8	1.7	0.451
90 / 10	27.4	0.3	1.577
50 / 25	12.5	1.4	0.202
40 / 10	9.6	0.5	0.219
25 / 10	5.4	0.1	0.083
25 / 5	6.9	0.7	0.087

(b)

	% of times better		average diff. in # of correct
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC) *
75 / 25	43.7	5.5	2.152
90 / 10	60.1	6.0	3.658
50 / 25	33.9	5.3	0.814
40 / 10	26.3	3.8	0.576
25 / 10	15.3	3.1	0.173
25 / 5	15.6	4.2	0.206

(c)

	% of times better		average diff. in # of correct
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC) *
75 / 25	63.1	8.2	4.096
90 / 10	56.3	24.3	2.167
50 / 25	53.0	8.1	1.802
40 / 10	37.7	13.6	0.801
25 / 10	24.4	9.3	0.277
25 / 5	20.2	12.4	0.137

*Each difference significant at 0.0001 for two-sided paired t-test and paired Wilcoxon test

Table 4: Percentage of simulations (out of 1000) our pixel weighting method (PW) outperformed distance weighting based on the Pearson residual (DW) and vice versa. Groups are widely-spaced (a), intermediately spaced (b), and overlapping (c).

(a)

	% of times better		average diff. in # of correct	two-sided p-value
Group sizes	PW	DW	classifications (PW - DW)	(Paired Wilcoxon)
100 / 100	1.1	2.0	-0.009	0.138
50 / 50	0.9	1.2	-0.002	0.721
20 / 20	1.2	2.6	-0.025	0.005
75 / 25	1.6	2.2	-0.002	0.841
50 / 25	1.3	1.5	-0.003	0.617
25 / 10	1.0	1.2	0.029	0.931

(b)

	% of times better		average diff. in # of correct	two-sided p-value
Group sizes	PW	DW	classifications (PW - DW)	(Paired Wilcoxon)
100 / 100	7.2	4.6	0.031	0.021
50 / 50	7.7	5.3	0.031	0.024
20 / 20	4.0	6.3	-0.029	0.019
75 / 25	9.5	5.8	0.578	< 0.0001
50 / 25	8.5	4.5	0.152	0.0005
25 / 10	7.1	5.2	0.152	0.005

(c)

	% of times better		average diff. in # of correct	two-sided p-value
Group sizes	PW	DW	classifications (PW - DW)	(Paired Wilcoxon)
100 / 100	18.2	10.5	0.314	< 0.0001
50 / 50	15.4	9.6	0.227	< 0.0001
20 / 20	11.5	9.5	0.034	0.350
75 / 25	36.4	6.6	4.015	< 0.0001
50 / 25	19.6	10.9	1.042	< 0.0001
25 / 10	20.3	9.1	0.531	< 0.0001

Table 5: Results of simulations (1000 trials each) comparing performance of WMBC and SMBC for three groups. Groups are widely-spaced (a), intermediately spaced (b), and overlapping (c).

(a)

	% of times better		average diff. in # of correct
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC) *
50 / 50 / 50	33.5	1.3	0.746
25 / 25 / 25	28.8	1.0	0.489
10 / 10 / 10	3.0	0.5	0.027
50 / 25 / 25	32.9	1.1	0.793
50 / 25 / 10	24.6	1.7	0.700
50 / 10 / 10	17.5	1.5	0.429
25 / 25 / 10	21.6	1.0	0.462
25 / 10 / 10	9.5	1.1	0.134

(b)

	% of times better		average diff. in # of correct
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC) *
50 / 50 / 50	48.5	5.9	1.288
25 / 25 / 25	34.8	3.3	0.615
10 / 10 / 10	5.6	1.5	0.047
50 / 25 / 25	41.7	4.4	1.136
50 / 25 / 10	37.8	4.8	1.165
50 / 10 / 10	26.5	5.7	0.619
25 / 25 / 10	25.0	5.7	0.427
25 / 10 / 10	18.9	3.4	0.26

(c)

	% of times better		average diff. in # of correct
Group sizes	WMBC	SMBC	classifications (WMBC - SMBC) *
50 / 50 / 50	63.5	7.0	2.278
25 / 25 / 25	44.5	8.8	0.854
10 / 10 / 10	8.6	5.9	0.039 **
50 / 25 / 25	50.8	9.9	1.549
50 / 25 / 10	44.9	13.6	1.087
50 / 10 / 10	41.7	14.5	0.707
25 / 25 / 10	33.7	9.4	0.592
25 / 10 / 10	23.7	6.7	0.304

*Each difference significant at 0.0001 for two-sided paired t-test and paired Wilcoxon test

** Result significant at 0.01

Table 6: Error rate for clustering the Venus V14 Quadrangle geologic units with WMBC and SMBC. Truth is taken to be the geologists' classification.

Situation	Error rate %	
	WMBC	SMBC
All 200 units	20.5	27.5
All 54 background units	9.3	9.3

Table 7: Adjusted Rand of WMBC and SMBC for V14 Venus data when initialization is model-based hierarchical clustering instead of geologists' classification.

Situation	Adjusted Rand	
	WMBC	SMBC
All 200 units, hierarchical initialization	0.0352	0.0310

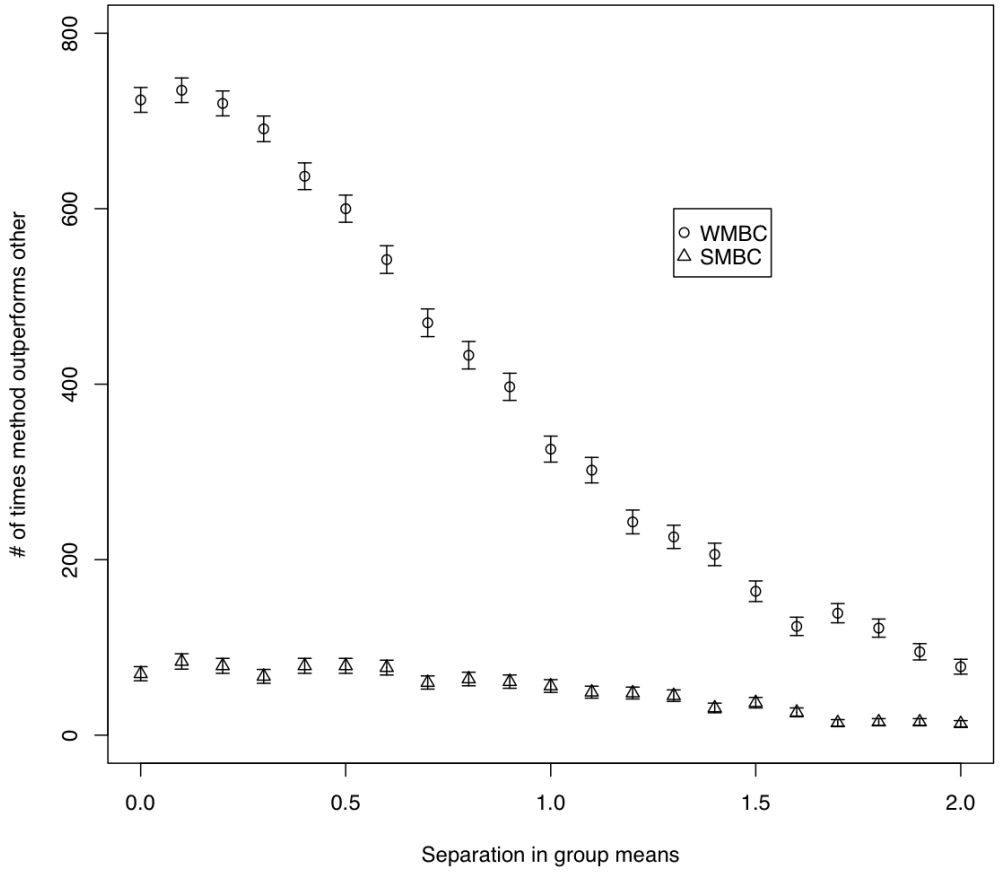

Figure 1: The number of times WMBC (∘) and SMBC (△) produced more accurate results in each of 1000 simulated data sets at 21 different separations of the means of each group. One-sigma error bars have been plotted.

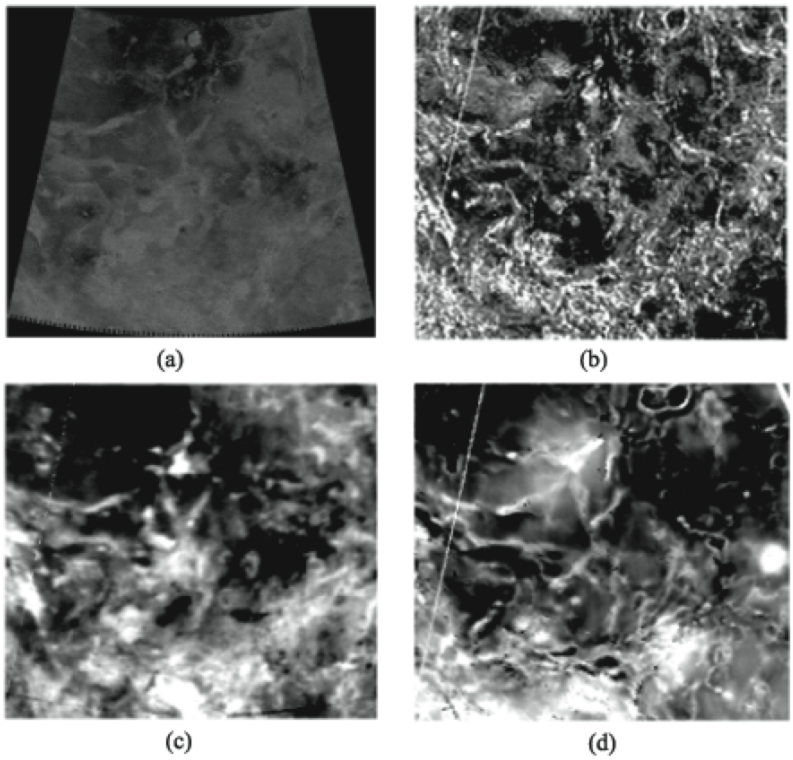

Figure 2: Four data sets that we use: (a) FMAP, (b) RMS slope, (c) emissivity, and (d) elevation. The FMAP image is over 300 times the resolution of the other data sets.

Figure 3: The original geologic map of V14 created by geologists. The region is divided into 200 units, which are distributed into 18 different groups. Each color in the image represents a different group.

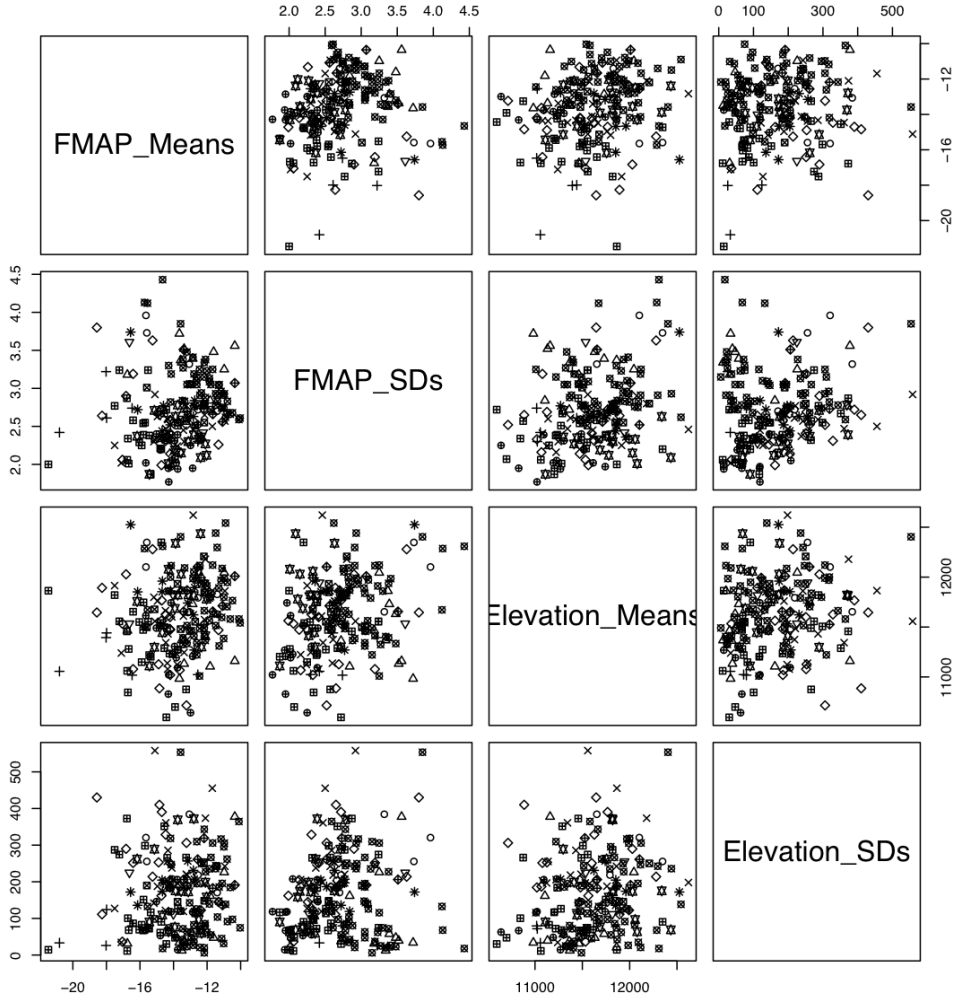

Figure 4: Plots of the means and standard deviations of FMAP and elevation pixels within each unit. The geologists' allocation of each unit is denoted by symbols.

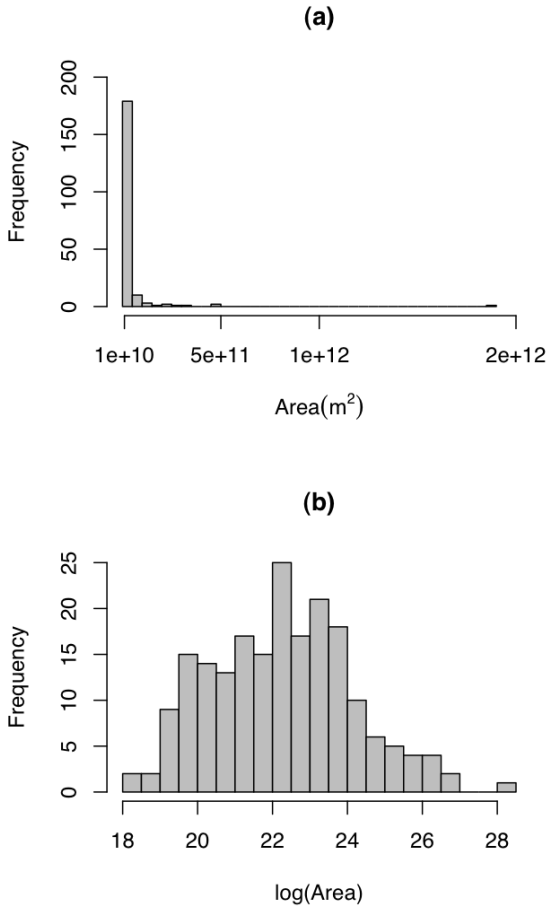

Figure 5: In the histogram of the areas of units on V14 (a), it is apparent that very few units dominate the total area of the quadrangle. Taking the log of these weights (b) preserves their order, but produces a much more symmetric distribution of weights that prohibits any single unit from adversely controlling cluster parameters in WMBC.