September 2, 2004

REMOTE SENSING FOR ENVIRONMENTAL BASELINING and MONITORING

David Reister

1. INTRODUCTION

Hyperspectral remote sensing is an emerging technology with the potential to identify plant species, map vegetation, characterize soil properties, identify contamination, classify ecological units and habitat characteristics, and differentiate causes of vegetation stress. The main benefit of this project, both to the DOE Fossil Energy Program and to the petroleum industry, will be an improved ability to provide detailed environmental data for a region, rapidly, at low cost. Increasingly, environmental data are needed to assess present conditions of lands owned, leased or managed by petroleum companies, and to characterize and quantify changes in the environmental conditions of these lands through time. Present methods of assessing large areas depend extensively on field surveys, which can take weeks or months to complete. Such methods are inconvenient and can be expensive. Further, some areas that are inaccessible or very large are difficult or virtually impossible to monitor accurately using field-survey techniques only. Hyperspectral sensing can be used to: identify plant species, map vegetation, characterize soil properties, identify contamination, classify ecological units and habitat characteristics, and differentiate causes of vegetation stress. The petroleum industry needs reliable analysis tools for rapid quantitative measurements and identification of type of stress causing environmental impact.

The oil industry is responsible for large tracts of land all over the world, used for production, refining, or marketing. If this project is successful at developing appropriate algorithms for the oil industry, remote sensing offers tremendous opportunities for the industry. Although many algorithms have been and are being developed, their focus is primarily on resource management (forest fires) and agriculture (crop nutrient and water needs). Algorithms needed by the oil industry (identification of hydrocarbons and differentiation of different types of stress relevant to our operations) are not currently available. With remote sensing, there is the ability to reduce the time to acquire important environmental information from months to days. Further, the data acquired can have a resolution of meters as opposed to current field monitoring, where sampling sites often are tens of meters apart. Remote data can be collected on a regular basis through time to detect and quantify environmental change. Regular data collection has a huge impact on the ability of the oil industry to baseline and monitor environmental issues. By understanding environmental conditions, capital and remediation projects can be accelerated, saving millions of dollars. By monitoring conditions more closely and in a more timely manner, we can improve management and protection of the environment.

An approximate estimate of the cost of current field monitoring is $20,000/acre. The cost of HyVista hyperspectral imagery is about $60 K for 80,000 acres (324 square km). Using commercial software (ENVI), preliminary analysis of the measurements to produce geobotanical maps costs $10 K. Thus, after R&D is completed, the total cost of producing geobotanical maps may be about $1/acre. Advanced methods using satellite data may be able to reduce the cost of hyperspectral imagery component of the total cost.

Produced water is the largest volume waste product in upstream activities. The oil industry needs low

cost accurate methods to quickly detect environmental discharges of produced water in oil fields that they operate or plan to purchase. DOE is supporting this project because it is developing an important environmental monitoring technique that does not currently exist.

2. SITE SELECTION

The US Geological Survey (USGS), the USDA Agricultural Research Service (ARS), and ORNL are collaborating to develop remote (airplane or satellite) hyperspectral sensor techniques to identify areas impacted by oil production. This is the final year of a multiyear project. Initially the project was focused on an oil spill site at the Jornada Experimental Range near Las Cruces, NM. New methods were developed to analyze hyperspectral field data measurements that were collected at Jornada. This year our focus shifted to Osage county, OK, which is the Osage Indian Reservation and has been a major oil producing area (38,500 oil wells) since 1896. The county is large (2,260 square miles) and 1,480 square miles are within a quarter mile of an oil well. The Osage Nation owns the mineral rights and there are 500 independent operators in the county. Many areas in the county have brine scars or weathered oil pits (The USGS has sample photographs: http://ok.water.usgs.gov/skiatook/Skiatook.Photo.html). The objective of the project was to collect hyperspectral remote data in selected regions of Osage county and analyze the data to detect brine scars, oil pits, and plant stress associated with brine and oil.

In November 2002, the USGS and ORNL visited Osage County to select potential regions to image. The USGS has been conducting research at two sites (Site A and Site B) that are adjacent to Skiatook Lake and owned by the Army Corps of Engineers. We selected the 2 USGS sites, 6 sites on the John Zink Ranch, 4 sites on the Tallgrass Prairie Preserve (owned by the Nature Conservancy), and 5 sites on the Bluestem Ranch.

ORNL selected the HyVista Corporation (http://www.hyvista.com/) to measure the hyperspectral data. In May 2003, ORNL contacted HyVista and requested that they collect images of seven regions that total 39 square kilometers. One of the seven regions is shown in Figure 1. The red markers on the figure are Global Positioning System (GPS) waypoints that were measured during the field visit in November 2000. The region in Figure 1 includes USGS Site B (114), and three Zink Ranch sites: 115, 124-125, and 126-127. On October 12, HyVista made the images. On October 27 and 28, a team from the USGS, ARS, and ORNL measured hyperspectral field data at 13 sites in the seven areas where the remote data was measured.

On November 10, ORNL received the data from HyVista. HyVista created a flight path that imaged three regions with overlapping strips. Each strip is 512 three-meter-pixels wide, for a total of 1536 meters. The North region has three strips with a total length of 12 km and a total area of 18 square km. The South region has 6 strips with a length of 69 km and an area of 106 square km. The West region has 5 strips with a total length of 40 km and an area of 61 square km. The total area is 185 square km. The data was received on five DVD and totals 14.4 gigabytes. The latitude and longitude of each pixel was provided. ORNL verified that the data covered our 7 study regions.

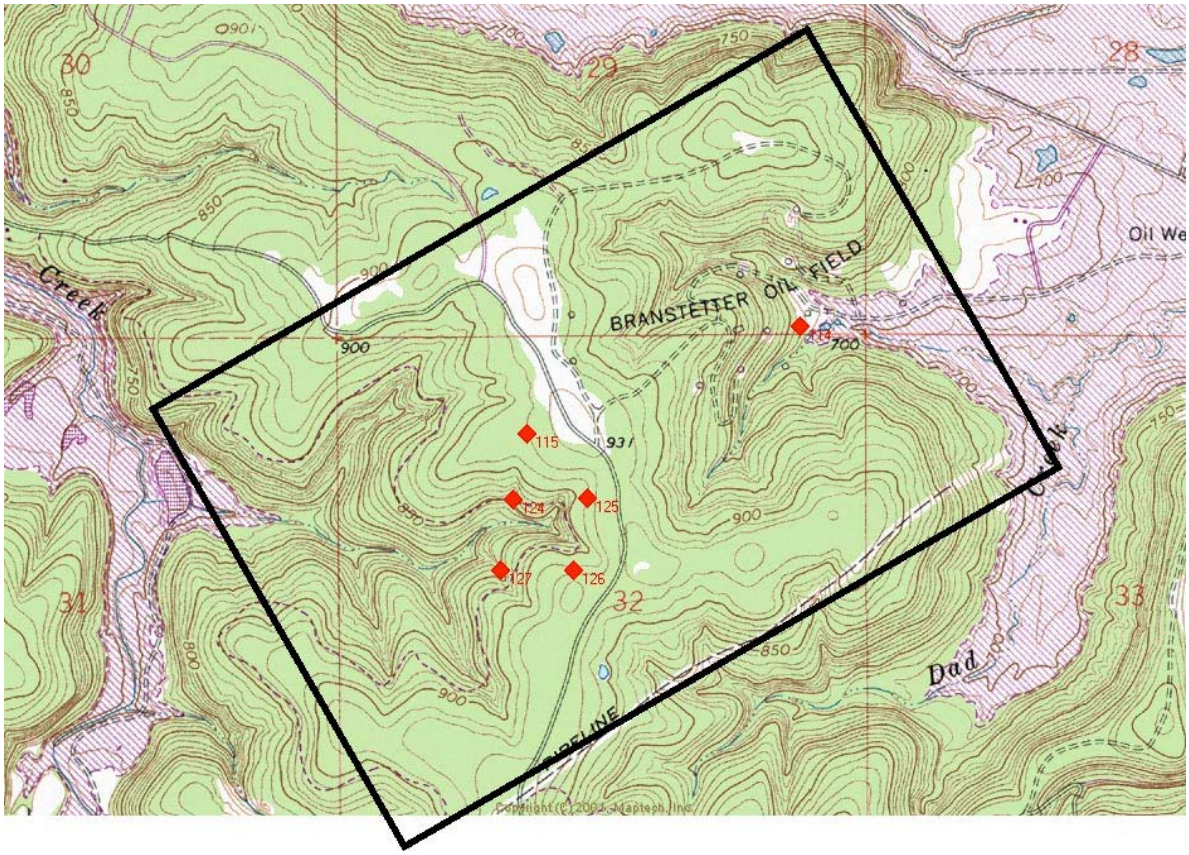


Figure 1. Selected region for Site B and three of the Zink Ranch sites.

The HyVista data has 126 spectral bands. Using three of the bands we can create the color images in Figures 2 and 3 (which cover the same region as Figure 1).

Figure 2. Three band HyVista image of the three Zink Ranch sites on the left side of Figure 1.

Figure 3. Three band HyVista image of Site B on the right side of Figure 1.

3. ANALYSIS

Our objective is to group hyperspectral measurements into as many distinct classes as possible. Richards and Jia [1] draw the distinction between information classes and spectral classes, where information classes are human classifications and spectral classes are groups or clusters of measurements made by a computer. A botanist uses many features to classify a plant in a particular information class: leaf type (flat, needle, deciduous, succulent), leaf geometry, attributes of branches, flowers and seeds, and smell. A hyperspectral reflectance measurement consists of hundreds of numbers that can range between zero and one. Spectral classes are groups of measurements with similar values for their reflectance spectra.

Hyperspectral sensors measure radiance and the data can be transformed to reflectance. For each of N spatial measurements, we transform the data by subtracting the mean of the M spectral values and normalizing to unit length. After the transformation, the data [a_{ij}, where $i = 1, M$ and $j = 1, N$] have the following properties: the spectral sum of data values is 0 and the spectral sum of the squares of the data values equals 1.

$$\sum_i a_{ij} = 0 \tag{1}$$

$$\sum_i a_{ij}^2 = 1 \tag{2}$$

The components of the correlation matrix (c_{jk}) are the correlation coefficient between the spectrum at pixel j and the spectrum at pixel k:

$$c_{jk} = \sum_i a_{ij} a_{ik} \ . \tag{3}$$

For hyperspectral data, the correlation matrix can be huge (If the data has millions of pixels, the full correlation matrix will have trillions of elements). Ultimately, we need to calculate the correlation matrix between a few measurements that will be the root vectors for the clusters and all of the measurements (a matrix with 100 rows and 100,000 columns). Iterative methods can be used to screen the measured vectors and find candidates for the root vectors.

The Euclidean distance (d_{jk}) between two measurements is defined by:

$$\left(d_{jk}\right)^2 = \sum_i \left(a_{ij} - a_{ik}\right)^2 . \tag{4}$$

Using (2) and (3), the Euclidean distance is directly related to the components of the correlation matrix

$$\left(d_{jk}\right)^2 = 2\left(1 - c_{jk}\right) . \tag{5}$$

We will define the correlation cutoff parameter (ε) by:

$$\varepsilon = \min\left[c_{jk}\right] . \tag{6}$$

Our clustering algorithm depends on two parameters, the correlation cutoff (ε) and the minimum number of members in a cluster (n_c). For each value of the correlation cutoff, we can define the set of spectral neighbors (ψ_k) for each spatial measurement point:

$$\psi_k = \left\{ j \mid j \in [1, N], c_{jk} > \varepsilon \right\} \tag{7}$$

Since the limits on the correlation coefficient are ± 1, the correlation cutoff will be in the interval [-1,1]. For a sufficiently large correlation cutoff, all of the neighborhoods ($\boldsymbol{\psi}_{\mathbf{k}}$) will have one member (the point k). For a sufficiently small correlation cutoff, all of the neighborhoods will have all N points. When the correlation cutoff is given, no iterations are required to calculate the members of each neighborhood.

We start with an arbitrary root vector and calculate the correlation coefficients with all of the other vectors. We can count the members of the neighborhood. If there are more than n_c members, we have a cluster. To choose the next root vector, we look for a vector that is not close to any of the current root vectors. We need a criterion for "not close". Each cluster is a ball in an M dimensional space with a radius of *r*. A necessary condition for two spheres to be non-overlapping is that the distance between their root vectors must be at least 2 *r*. A necessary condition for many spheres to be non-overlapping is that the minimum distance between all possible root vectors must be at least 2 *r*. Using Eq. (5), when the radius is doubled, the right side of the equation will increase by a factor of four. Thus, criterion for "not close" is that the next root vector will have a correlation with all of the current root vectors that is less than ε_a, where ε_a is defined by: $\varepsilon_a = 1 - 4\left(1 - \varepsilon\right)$. The results in this report were calculated using an erroneous version of Eq. (5), in which the right side of the equation was not squared. Consequently, we concluded that the factor of four was a factor of two. Empirically, we found that the factor of two was not sufficient and used a value of 2.8. Since we had a larger value of "not close", we are reporting results with more clusters than we will find when we repeat the calculation using the factor of four.

We are attracted to the terminology for human population patterns on the surface of the Earth. A cluster is like a neighborhood in a large city or a subdivision in a rural county. The next larger unit is the group,

which is analogous to a city or county. The largest unit is the domain, which is analogous to a state, country, or the world. We define two additional correlation cutoffs: group (ε_g) and domain (ε_d). If the correlation between a measurement and a root vector is greater than ε_g, the measurement is in the group of the root vector. If the correlation is greater than ε_d, the measurement is in the domain of the root vector. We assume that: $\varepsilon > \varepsilon_a > \varepsilon_g > \varepsilon_d$. In this report, we have used the domain correlation cutoff to exclude measurements that corresponded to reflection from water.

Each cluster is a ball in an M dimensional space. The clusters do not fill the space and the fraction of the space filled by the clusters becomes less as M increases. We compare the volume of a hypersphere to the volume of its bounding box. When M = 1, the hypersphere and the box are identical and the ratio is 100%. When M = 2, the ratio is $\pi/4$ = 79%. When M = 3, the ratio is $\pi/6$ = 52%. When M = 6, the ratio is 3%. We will fill the space by associating each measurement with the nearest cluster.

If we cluster a set of measurements, we will find a set of root measurements and many of the measurements will be members of one of the clusters, but many of the measurements will not be a member of any of the clusters. We will say that a measurement is associated with the root vector that has the highest correlation with the measurement if the correlation is greater than ε_g. If we assume that the minimum number of members in a cluster is one (and $\varepsilon_a > \varepsilon_g$), all of the measurements will be members or associates of one of the clusters. If the minimum number of members in a cluster is more than one, some of the measurements may be unclassified.

As we calculate each row of the correlation matrix, we update the closest root vector. Let, g_j be defined by: $g_j = \max_k c_{jk}$. As each correlation coefficient is calculated, we can update g_j. After we calculate each row of the correlation matrix, we look for a vector that is "not close" to any of the current root vectors and add it to the list of root vectors. Thus, the next root vector satisfies: $g_j < \varepsilon_a$ and $c_{1k} > \varepsilon_d$. When we get to the end of the list of vectors, we stop clustering.

We will store the root vectors in an array: $\varphi(m)$. Before we begin clustering, we choose the first root vector [$\varphi(1)$] and set the cluster counter (mc) to zero. During the cluster algorithm, we will choose the next root vector.

Cluster Algorithm

```
1.  Loop over mg.                    {Continue until all root vectors chosen.}
        mc = mc + 1.
        ma = φ(mc).
        mem = 0.                     {mem = members of the cluster.}
2.  Loop over mb.                    {All vectors.}
        Calculate C(mb,mc).          {C(mb,mc) is the mc row of the correlation matrix.}
        If C(mb,mc) > ε, mem = mem + 1.
End of mb loop.
3.  If mem > 1, loop over mb         {Have a cluster.}

        If C(mb,mc) > ε              {mb in cluster.}
            Check for cluster overlap.   {Is mb in another cluster?}
            Λ(mb) = mc.                  {Set the cluster map.  Used to check for overlap.}
        End if.
        If C(mb,mc) > g(mb)
            g(mb) = C(mb,mc).        {Highest correlation.}
            θ(mb) = mc.              {Closest root vector}
        End if.
        Set the color map for the figures.
End if.
4.  If mem < n_c, mc = mc - 1.       {Do not have a cluster.}
5.  Loop over mb.  Start at ma.      {Choose the next root vector.}
        If C(mb,1) < ε_b, next mb.
        If g(mb) < ε_a,
            φ(mc+1) = mb.                {Next root vector.}
            Go to 1.                     {Continue mg loop.}
End mb loop.
6.  Exit from the mg loop.               {Clustering is completed.}
7.  End of mg loop.
8.  Loop over mb                         {Set the final cluster map.}
        if g(mb) > ε_g, Λ(mb) = θ(mb).
End mb loop.
```

4. RESULTS

We have analyzed the hyperspectral images for the nine sites listed in Table 1. To compare Site B with its neighborhood, we expanded the area from 12,740 pixels to 90,720 pixels. The number of pixels ranges from 23,100 for Site A to 104,160 for Bluestem 134. The two rows for each of the sites present the results for two different cluster radii: 0.990 and 0.996. For the smaller cutoff, the number of clusters ranges from 35 for Zink Ranch 127 to 133 for Bluestem 134. For the larger cutoff, the number of clusters ranges from 100 for Site A to 418 for Bluestem 134.

Name	Clusters	Members	Associated	Unclustered	Pixels
Blue 134	133	76120	28027	13	104160
	418	55271	48785	104	104160
Blue 135	81	53687	21366	19	75072
	238	31379	43625	68	75072
Blue 136	61	49244	10755	1	60000
	185	27522	32454	24	60000
Site A	46	13702	3337	6061	23100
	100	7573	8856	6671	23100
Site B	82	68290	11202	11288	90720
	219	44506	34709	11505	90720
Old Site B	38	8787	2427	1526	12740
TGP 118	49	20324	8053	9	28416
	125	11737	16652	27	28416
TGP 122	74	25059	23951	10	49020
	257	25420	23566	34	49020
Zink 127	35	25923	11030	7	36960
	115	25821	11132	7	36960
Zink 131	50	68171	19724	9	87904
	174	41595	46298	11	87904

Table 1. Summary of the clustering results.

At the end of the clustering algorithm, we eliminate overlap by assigning each pixel to the closest root vector. We associate pixels that are not members of a cluster with the nearest root vector, if they are within the group cutoff (r_g) of a root vector. In Table 1, the Members column is the sum of all of the cluster members, the Associated column is the sum of all of the associated pixels, and the Unclustered column is the sum of all pixels that are neither members nor associates.

The Unclustered values are high for Sites A and B because we used r_b to exclude the water pixels.

For each site, we have more than 10, 000 pixels and each pixel has 126 bands of spectral data. If the hyperspectral measurements have sufficient variance, we would be able to find a basis for the measurements with 126 members. Typically, the variance is not sufficient and the number of basis vectors (the numerical rank) is 20 or less. Singular value decomposition (SVD) can be used to create both an uncorrelated orthonormal basis for a collection of spectral measurements and the value of the

numerical rank (K). The SVD basis vectors are the principal axis of the hyperellipsoid that bounds the measurements. The first basis vector is the mean value for the measurements; the second and higher components are the directions with the most variance in the data. We will define the numerical rank by determining the number of basis vectors required to explain a given level (for example 99%) of the variance in the data.

We have applied SVD to the root vectors for each of the nine sites and the results are summarized in Table 2. The table shows the numerical rank required to achieve three levels of variance: 99%, 99.5%, and 99.9%. At all three levels, the Blue 134 site has the highest numerical rank. This high level of variance is consistent with the Blue 134 site having the highest number of clusters. Site B also has high values for numerical rank. Site A, TGP 118, and Zink 127 have low values for numerical rank and small numbers of clusters. In all cases, the numerical rank is lower or the same for the cases with the smaller correlation cutoff and higher number of clusters. My intuition suggests that the rank would be higher.

Name	Clusters	99%	99.5%	99.9%
Blue 134	133	8	10	17
	418	6	8	13
Blue 135	81	7	9	15
	238	5	6	12
Blue 136	61	6	7	13
	185	5	7	12
Site A	46	6	8	14
	100	4	6	11
Site B	82	7	9	17
	219	5	7	13
Old Site B	38	6	7	12
TGP 118	49	5	7	12
	125	5	6	11
TGP 122	74	6	8	14
	257	5	7	12
Zink 127	35	5	7	11
	115	5	6	11
Zink 131	50	6	8	13
	174	5	7	12

Table 2. Numerical rank of root vectors using the SVD basis.

This section will present the results for the nine sites.

4.1. USGS Site A

The USGS Site A is in the upper right corner of Figure 4. The boundaries of analysis region are displayed in Figure 5. Figure 5 is based on the same three band HyVista reflectance data as in Figure 4. However, the magnitude of the colors for each pixel has been increased until one of the colors is at the maximum value. We call it a "Bright Pixel" image. The unvegetated areas disturbed by oil production and changes in the lake level are orange on the image.

Figure 4. USGS Site A Three Band HyVista image.

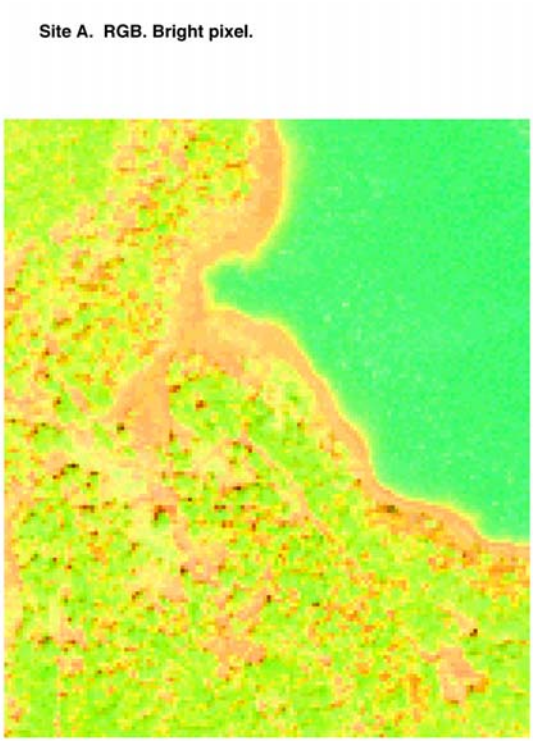


Figure 5. Bright Pixel Image of the Site A Analysis Region.

As displayed in Table 1, we have created clusters. The next step is to produce geobotanical maps in which each cluster has a different color and the cluster position is plotted on a map that covers the same region as Figure 5. In Table 1, the number of clusters ranges from 35 to 418. We have created a 27 value color table by allowing each of the three primary colors (red, green, blue) have three values (zero, middle, high). The color table is displayed in Figure 6. Since it is difficult to distinguish all of the 27 colors on a computer screen and even harder on printed output, we cannot create a unique color for each of the clusters.

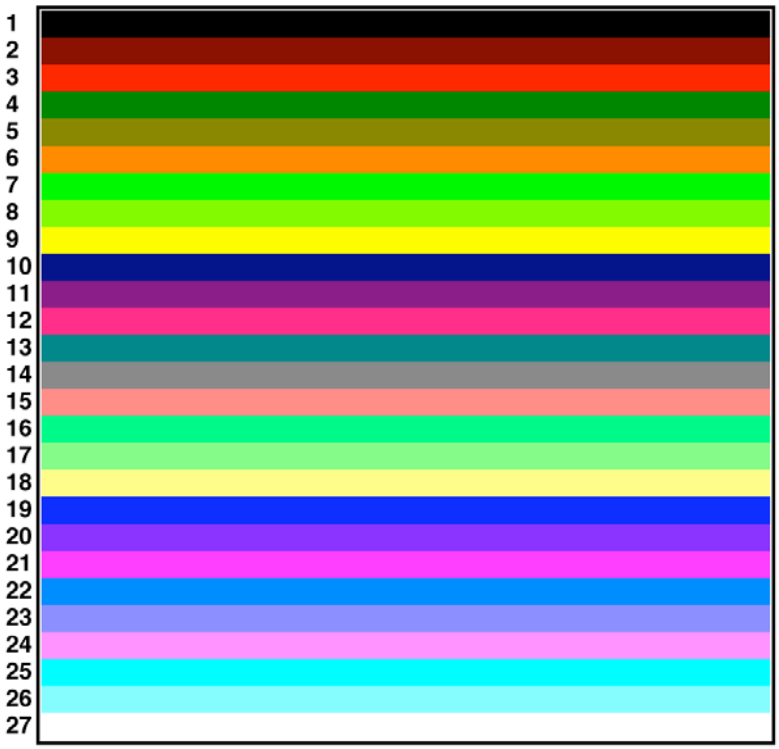


Figure 6. Color Table.

We can display all of the big clusters with a unique color if we set the definition of big high enough. The big clusters for Site A are displayed in Figure 7. For Site A, big means members greater than 3. For Bluestem 134 with 133 clusters, big means members greater than 80. We are primarily interested in brine scars. We can create an image where most vegetation is green, and we can use the colors for all of the brown (not green) clusters. For a large number of brown clusters the color scale repeats after all of the colors are used. Figure 8 displays all of the brown clusters for Site A.

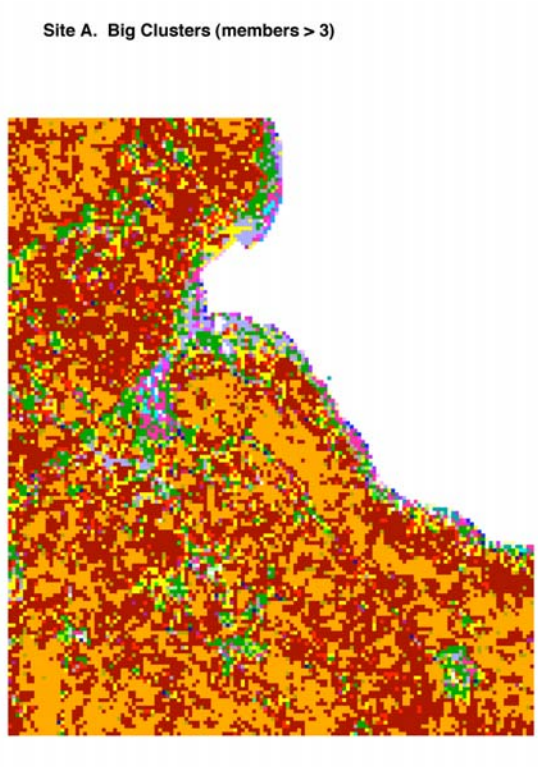


Figure 7. Site A Big Clusters.

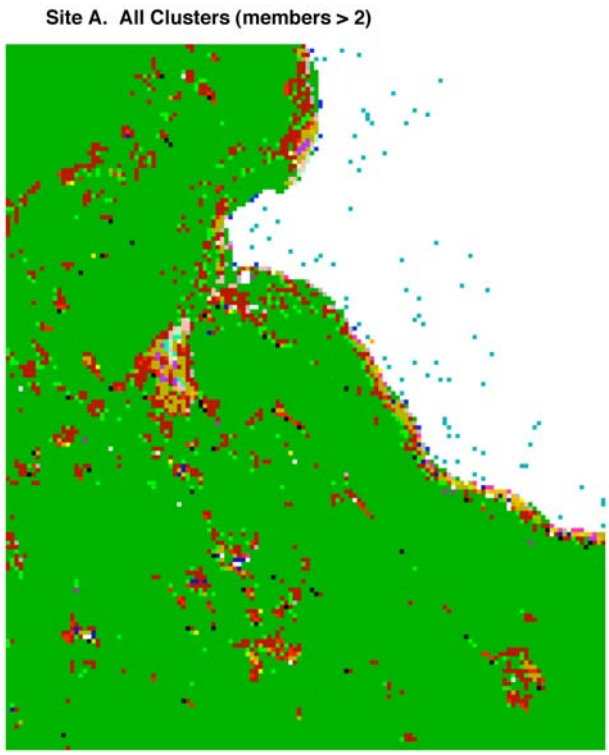
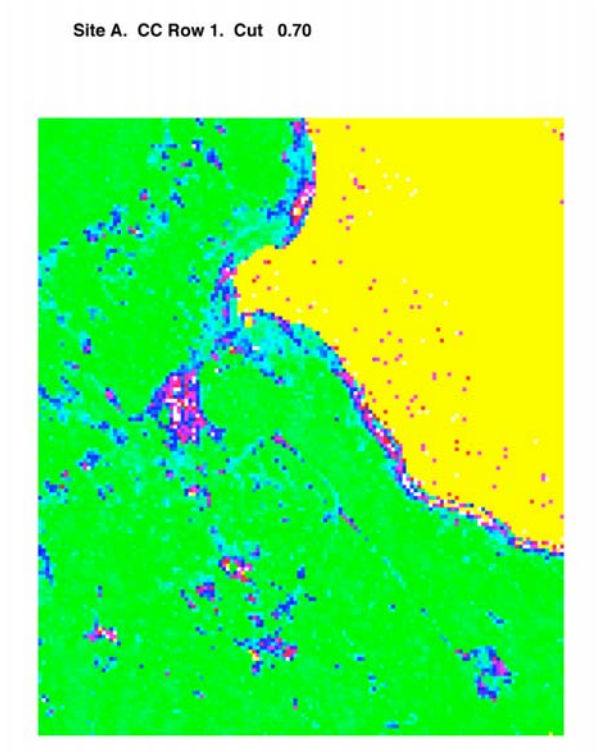


Figure 8. All Brown Clusters for Site A.

We have explored several methods for visualizing the data. We plot the correlation coefficient with respect to one of the root vectors. We will use either green vectors or brown vectors. We have created a continuous color scale from green through blue to red (see Figure 10). We use a cutoff and all measurements below the cutoff are yellow. When we correlate with a green vector, high correlation is green and low is red (Figure 9). When we correlate with a brown vector, high correlation is red and low is green (Figure 11). While both Figures tell similar stories, the red area is much larger in Figure 11.



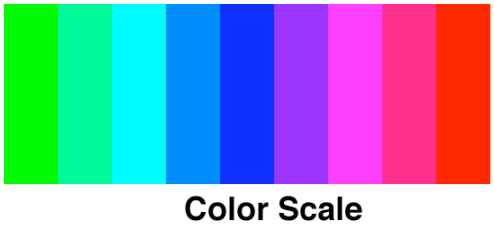


Figure 10. Color Scale

Figure 9. Correlation with a Green vector for Site A

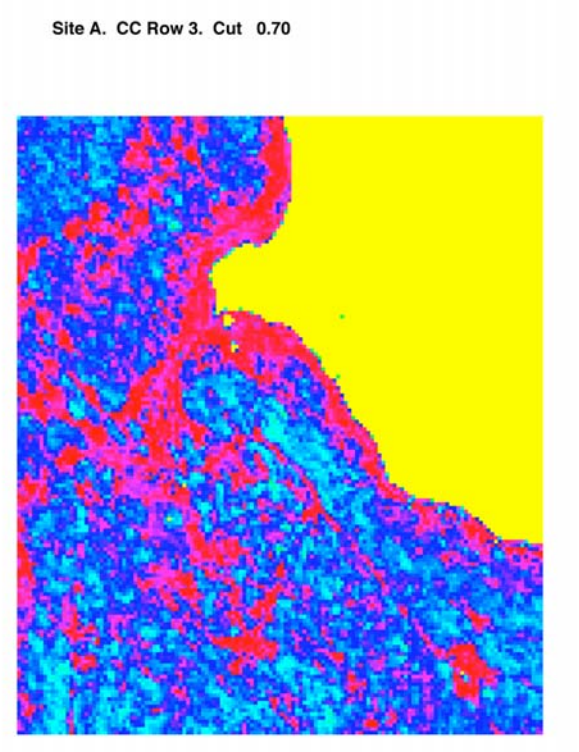


Figure 11. Correlation with a Brown vector for Site A

4.2. USGS Site B

The USGS Site B is in the upper middle of Figure 3. The bright pixel image of the initial Site B analysis area is displayed in Figure 12. The unvegetated regions are orange on the image. This region has 38 clusters. The big clusters with more than 4 members are shown in Figure 13. The all brown cluster image is displayed in Figure 14. One of our challenges is to differentiate between different sources of environmental damage. For this site, we want to distinguish roads from areas damaged by produced water spills. An enlargement of the road in the upper left area of the study area is displayed in Figure 15 and an enlargement of the area down stream from the produced water holding pit is shown in Figure 16. The holding pit is the white region in Figure 16 (and all of the other figures). From the pit, the site has a gentle slope to the lake shore on the right. Comparing the two images, we find two clusters that are only in the road image, nine clusters that are only in the pit image, and 7 clusters that are in both images. Thus, we can distinguish road damage from produced water damage.

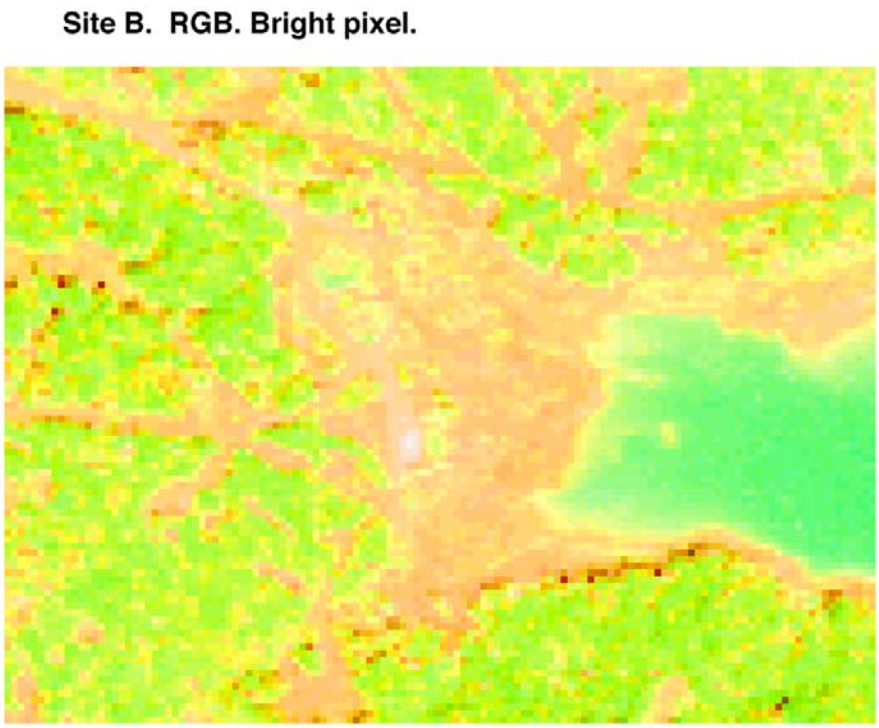


Figure 12. Bright Pixel Image of the Initial Site B Analysis Region.

Site B. Big Clusters (members > 4)

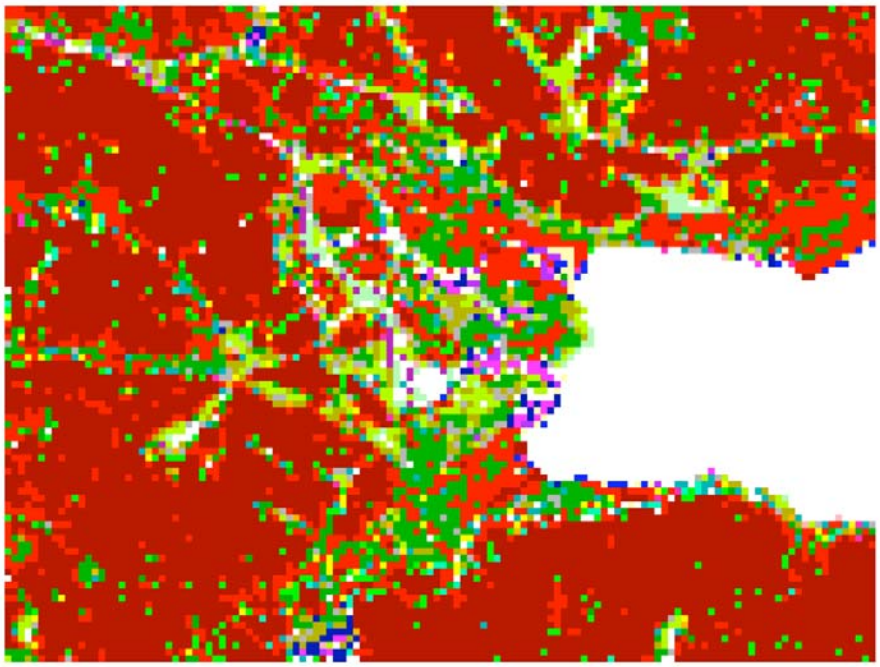


Figure 13. Site B Big Clusters.

Site B. All Clusters (members > 2)

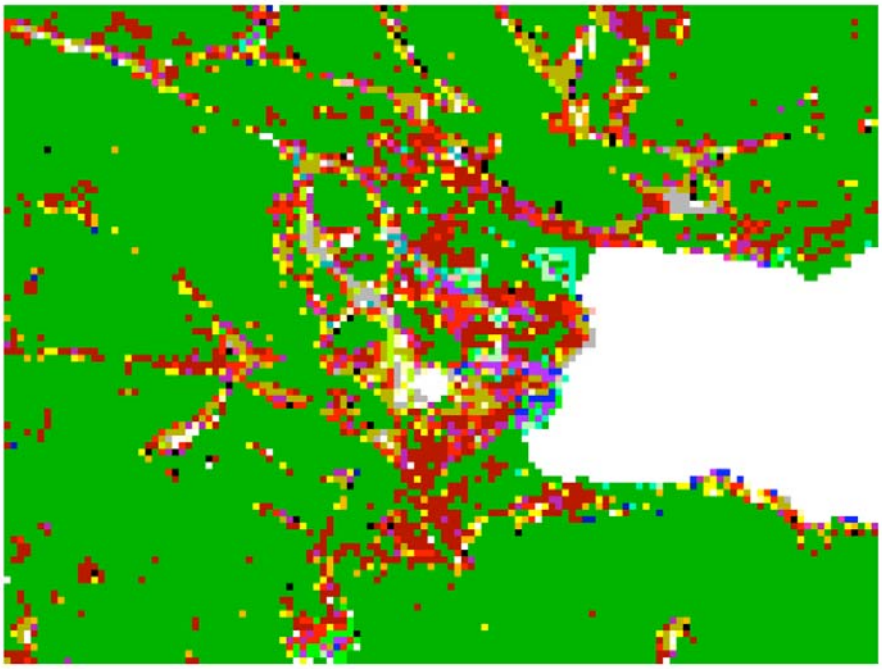


Figure 14. All Brown Clusters for Site B.

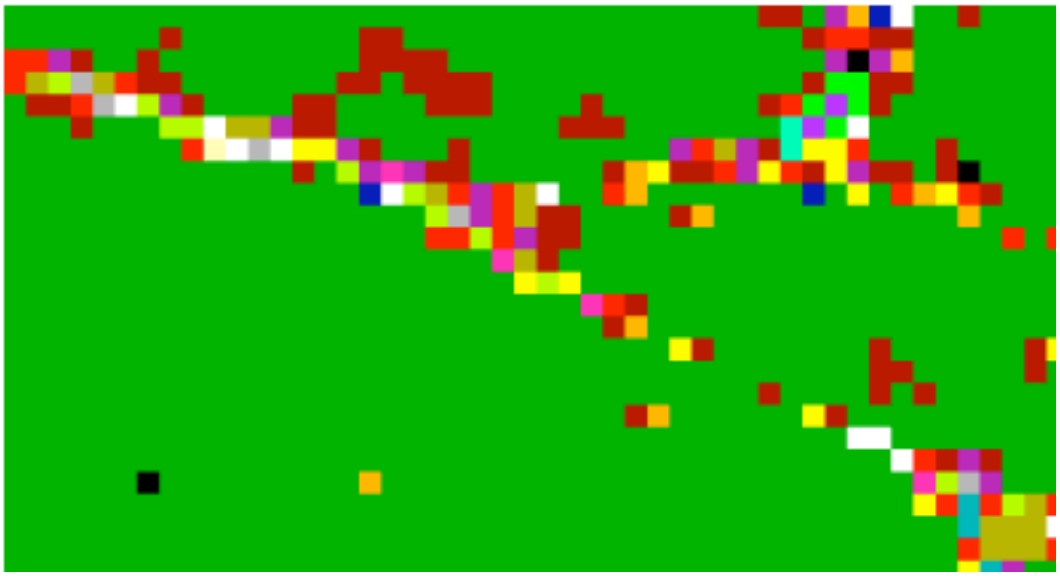


Figure 15. Enlargement of Road in upper left of Figure 14.

Figure 16. Enlargement for Site B.

4.3. Bluestem Site 134

The color image for the Bluestem Site 134 is displayed in Figure 17. The bright pixel image of the analysis area is shown in Figure 18. The big (more than 80 members) cluster map is displayed in Figure 19. The brown cluster map is shown in Figure 20. Figures 21 to 24 present enlargements of various regions of Figure 20. The all cluster map is shown in Figure 25 for the 131 cluster case and in Figure 26 for the 418 cluster case. We use single cluster maps to display the spatial location of each cluster. Most of the brown and peach pixels in Figure 26 are from clusters 1 and 14, which are shown in Figures 27 and 28.

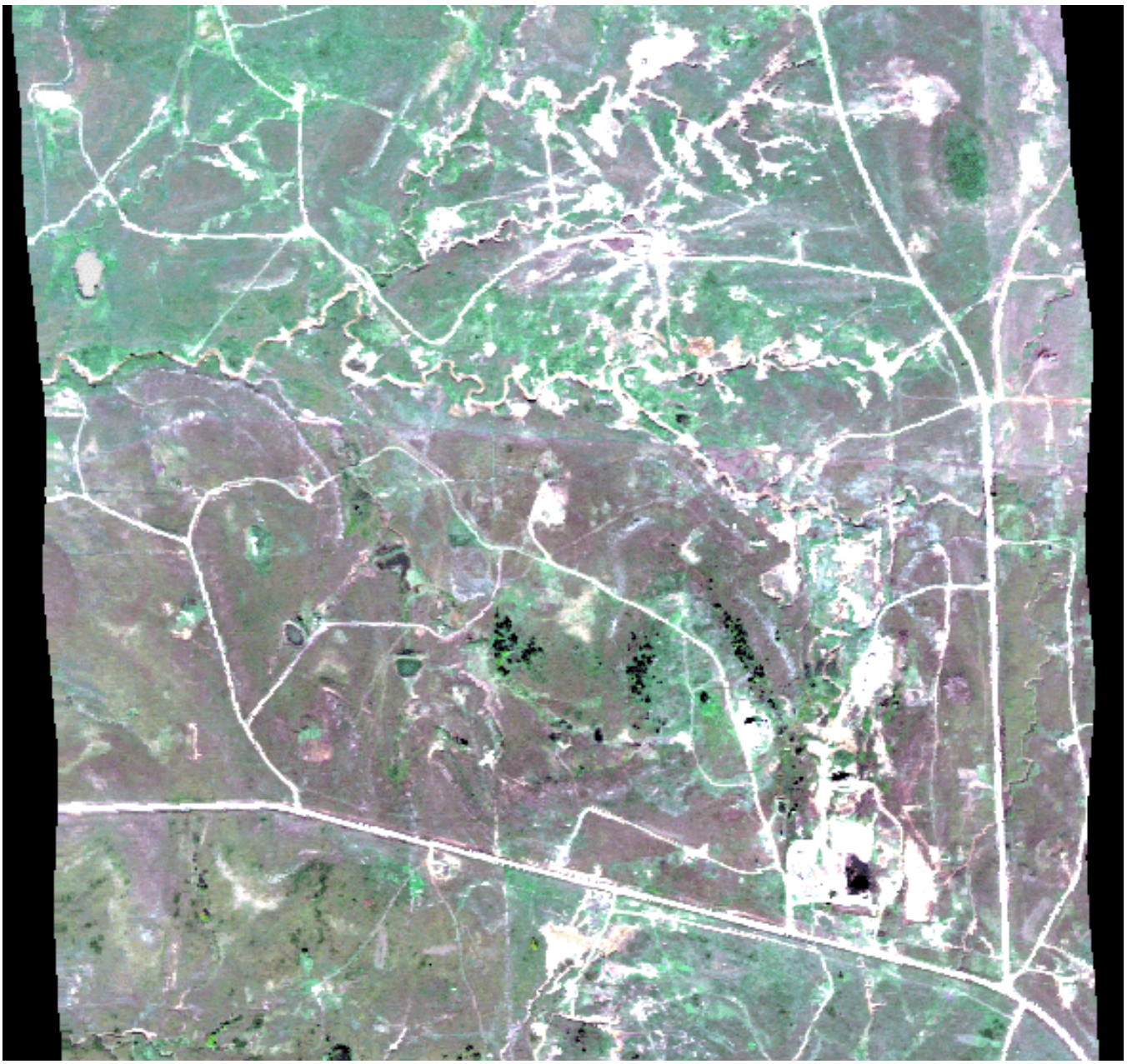


Figure 17. The color image for the Bluestem Site 134.

Bluestem 134. RGB. Bright pixel.

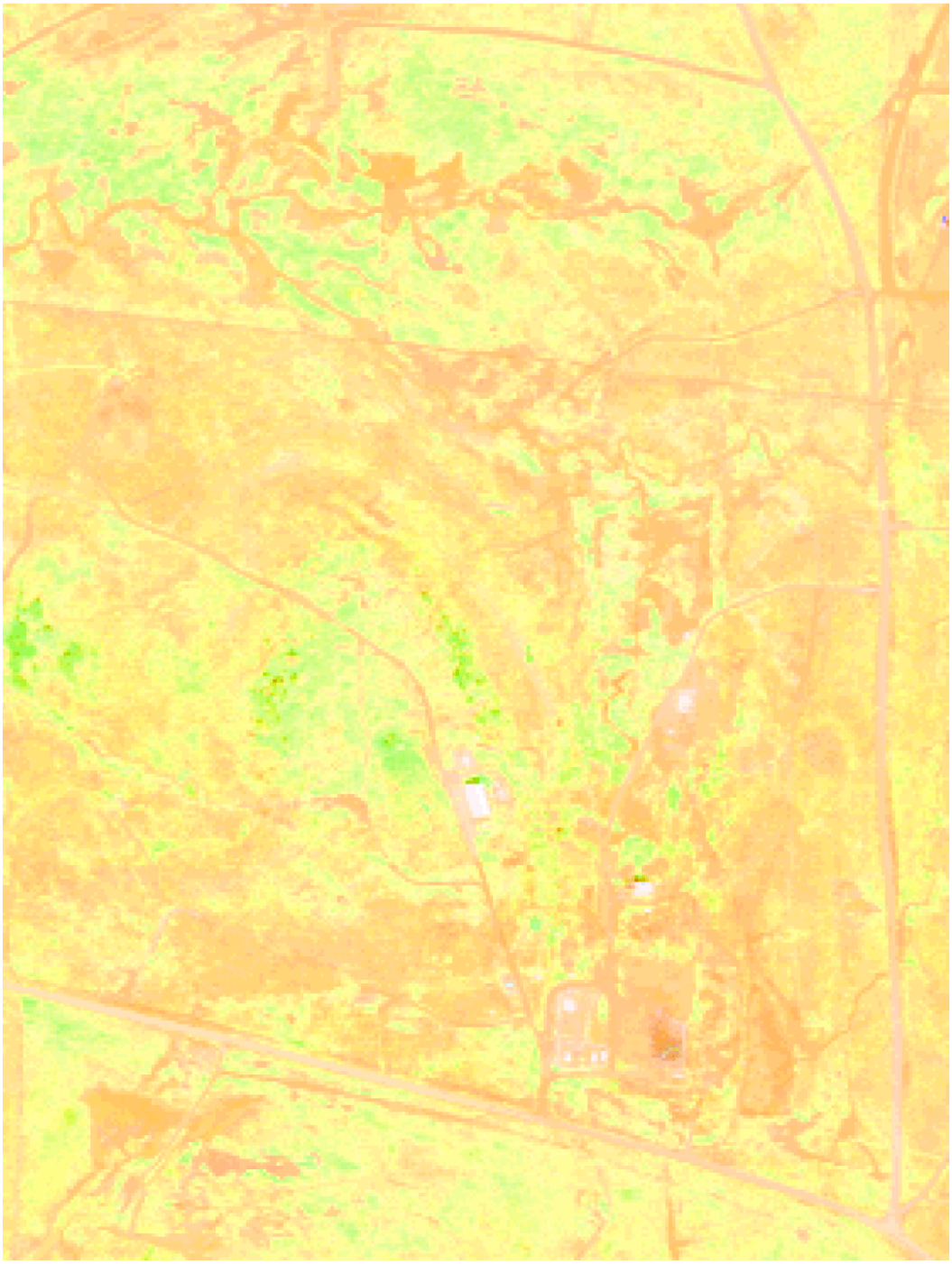


Figure 18. The Bright Pixel Image of the Bluestem Site 134.

Bluestem 134. Big Clusters (members > 80)

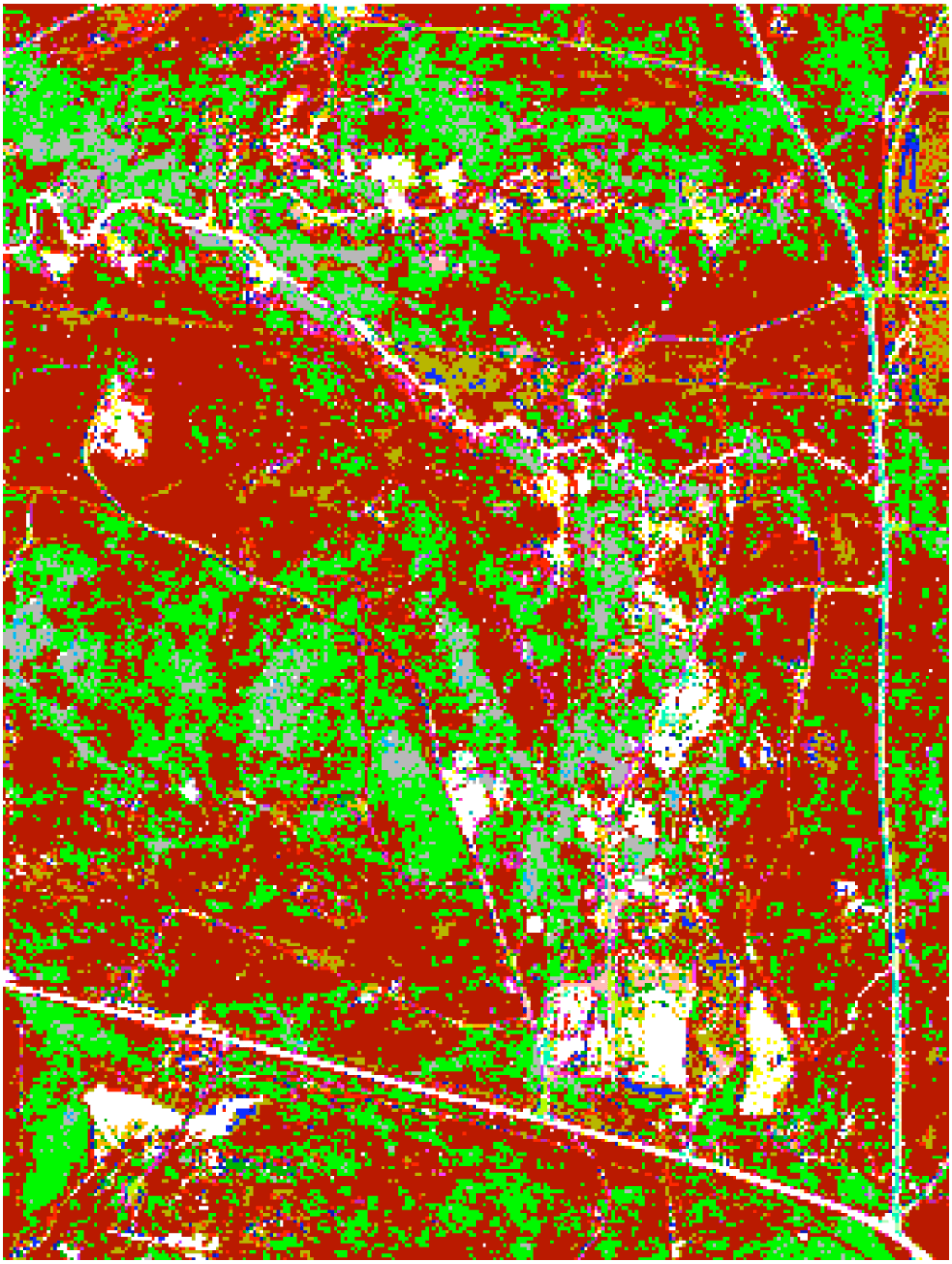


Figure 19. The Big Cluster map for the Bluestem Site 134.

Bluestem 134. All Brown Clusters (members > 2)

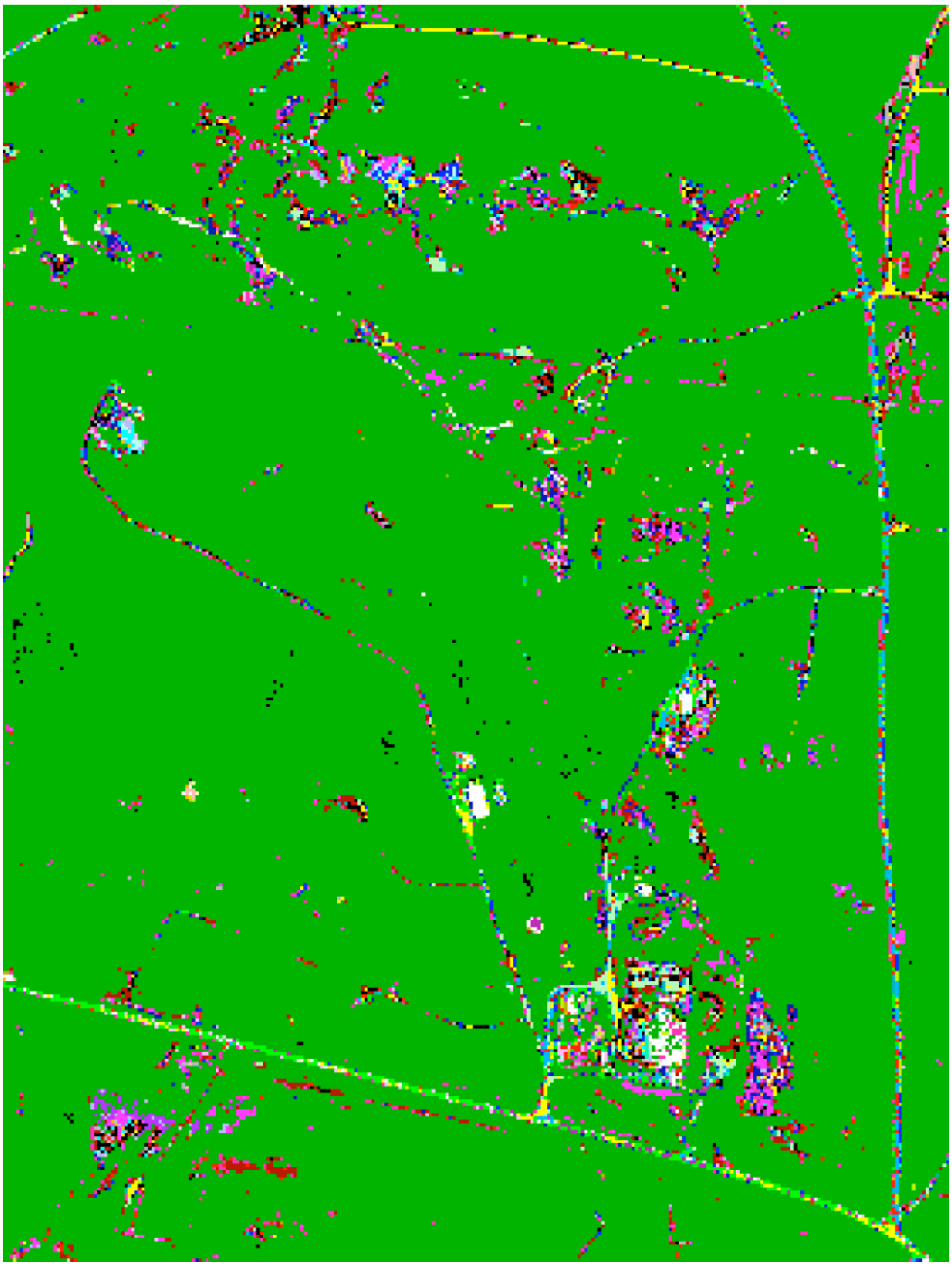


Figure 20. The Brown Cluster Map for the Bluestem Site 134.

Figure 21. Enlargement of the Bluestem Site 133 in the lower left of Figure 20.

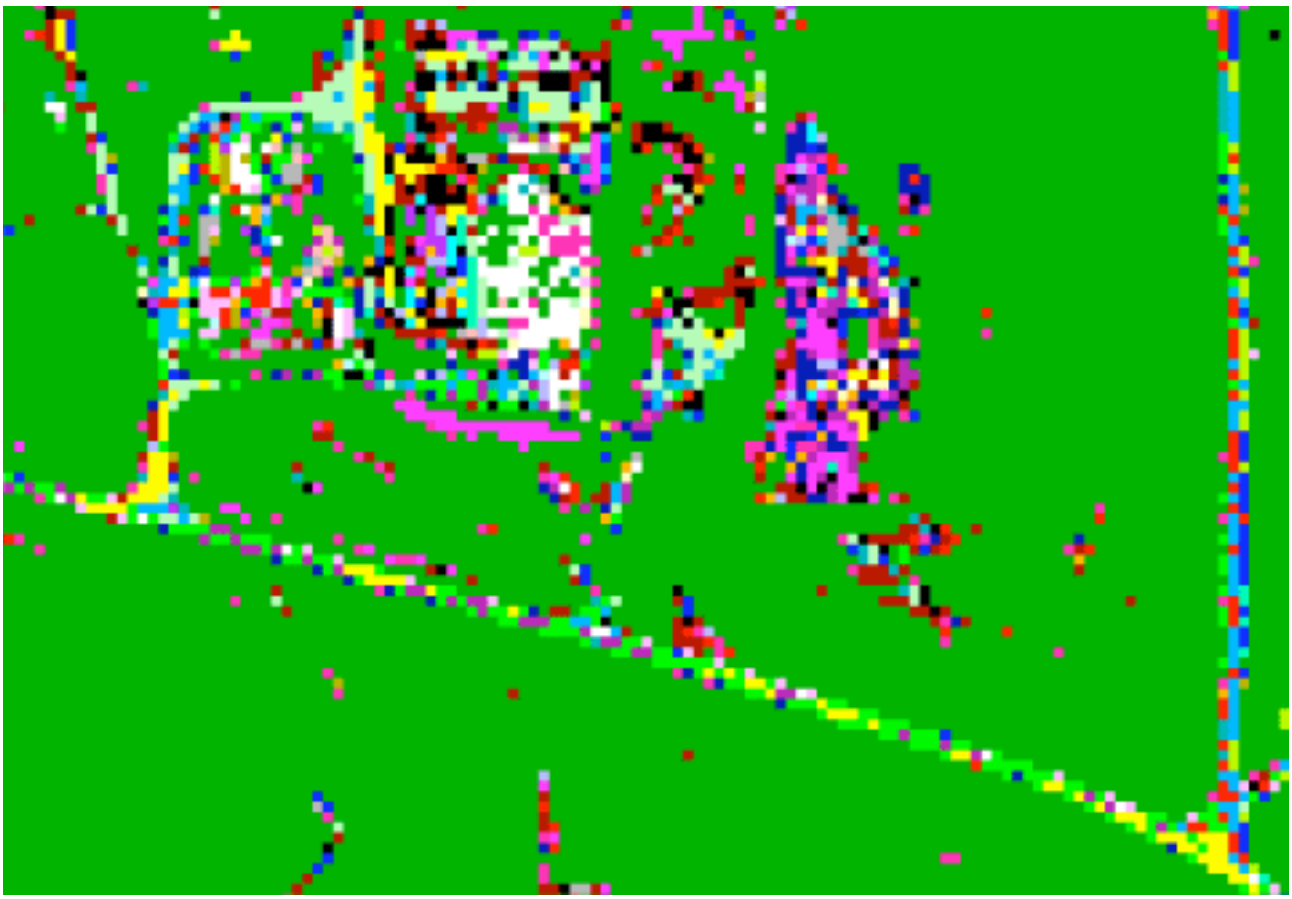


Figure 22. Enlargement of the Pit Site in the lower right of Figure 20.

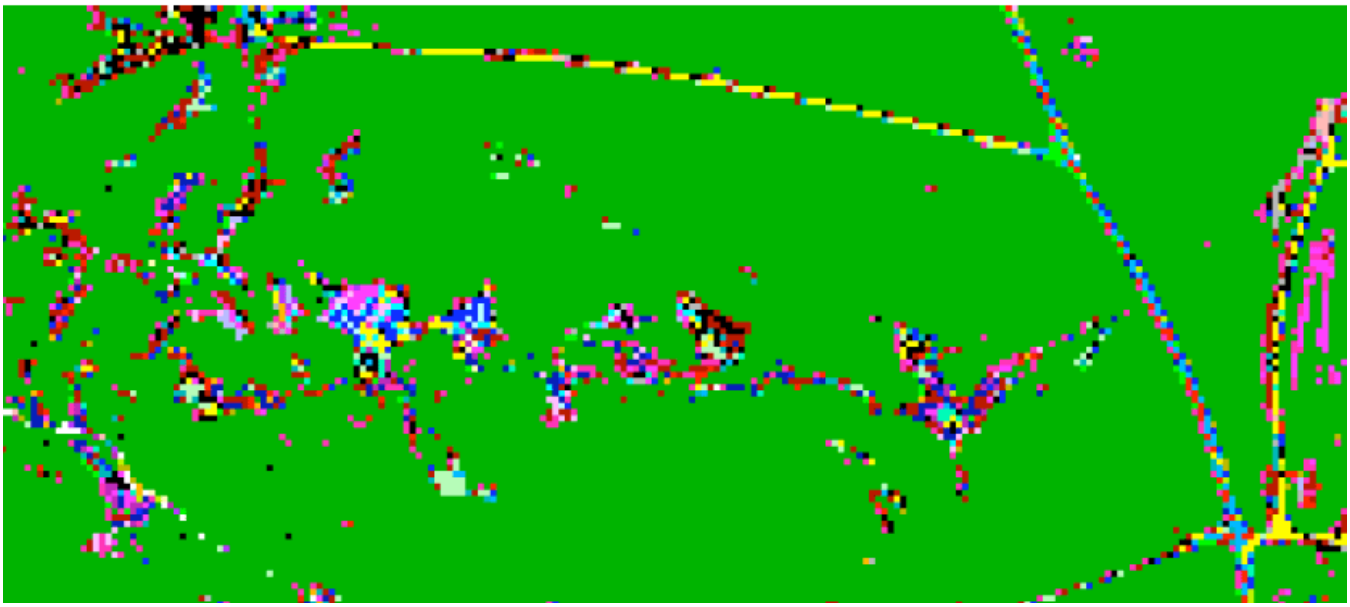


Figure 23. Enlargement of the road in the upper part of Figure 20.

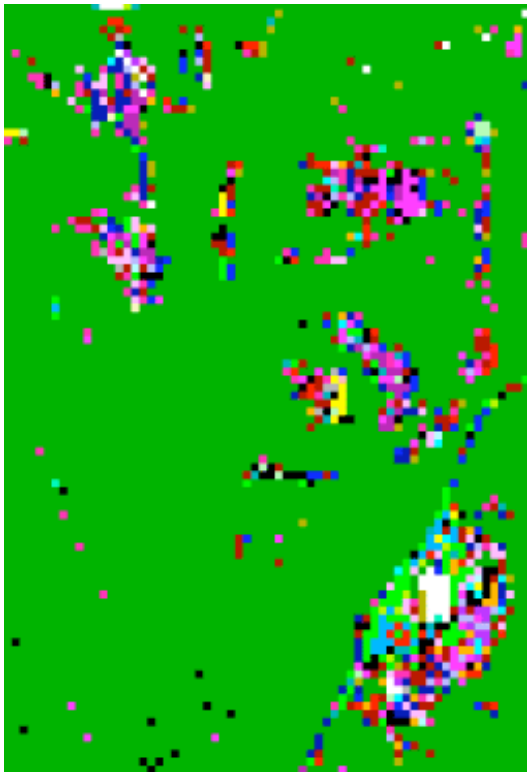


Figure 24. Enlargement of the center of Figure 20.

Blue_134. All Clusters

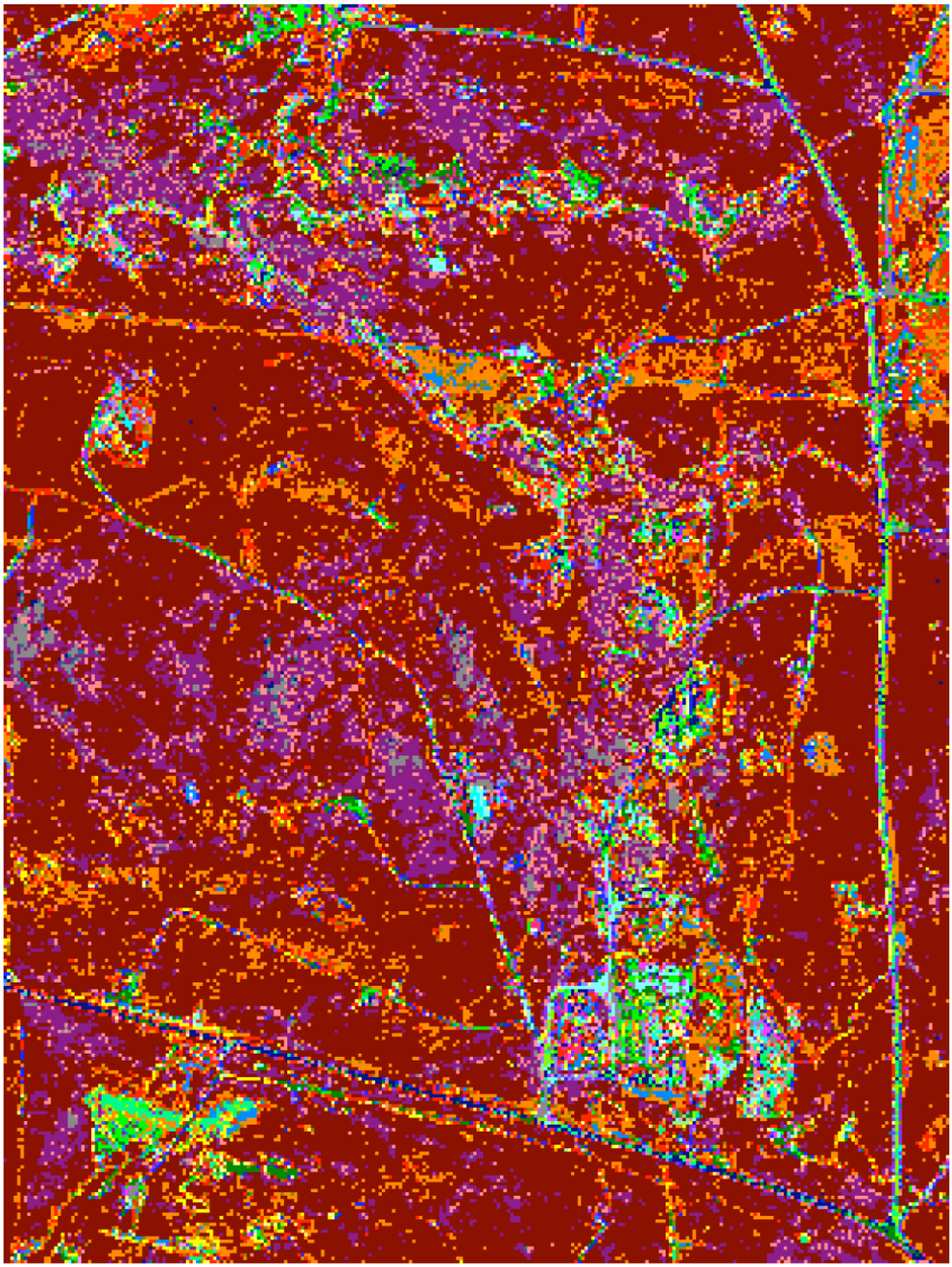


Figure 25. The 133 Cluster Map for the Bluestem Site 134.

Blue_134. All Clusters

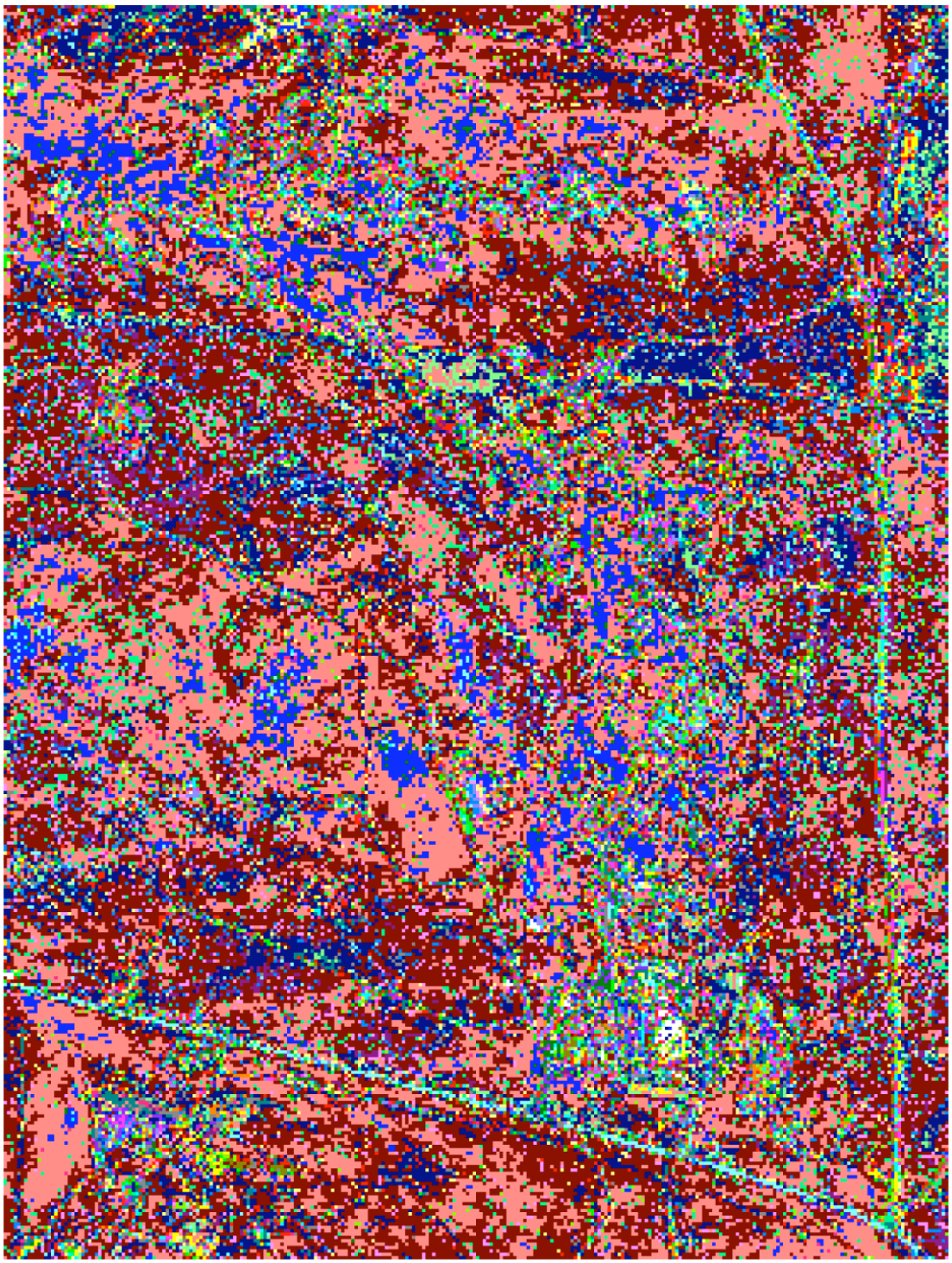


Figure 26. The 418 Cluster Map for the Bluestem Site 134.

Blue_134. Cluster 1.

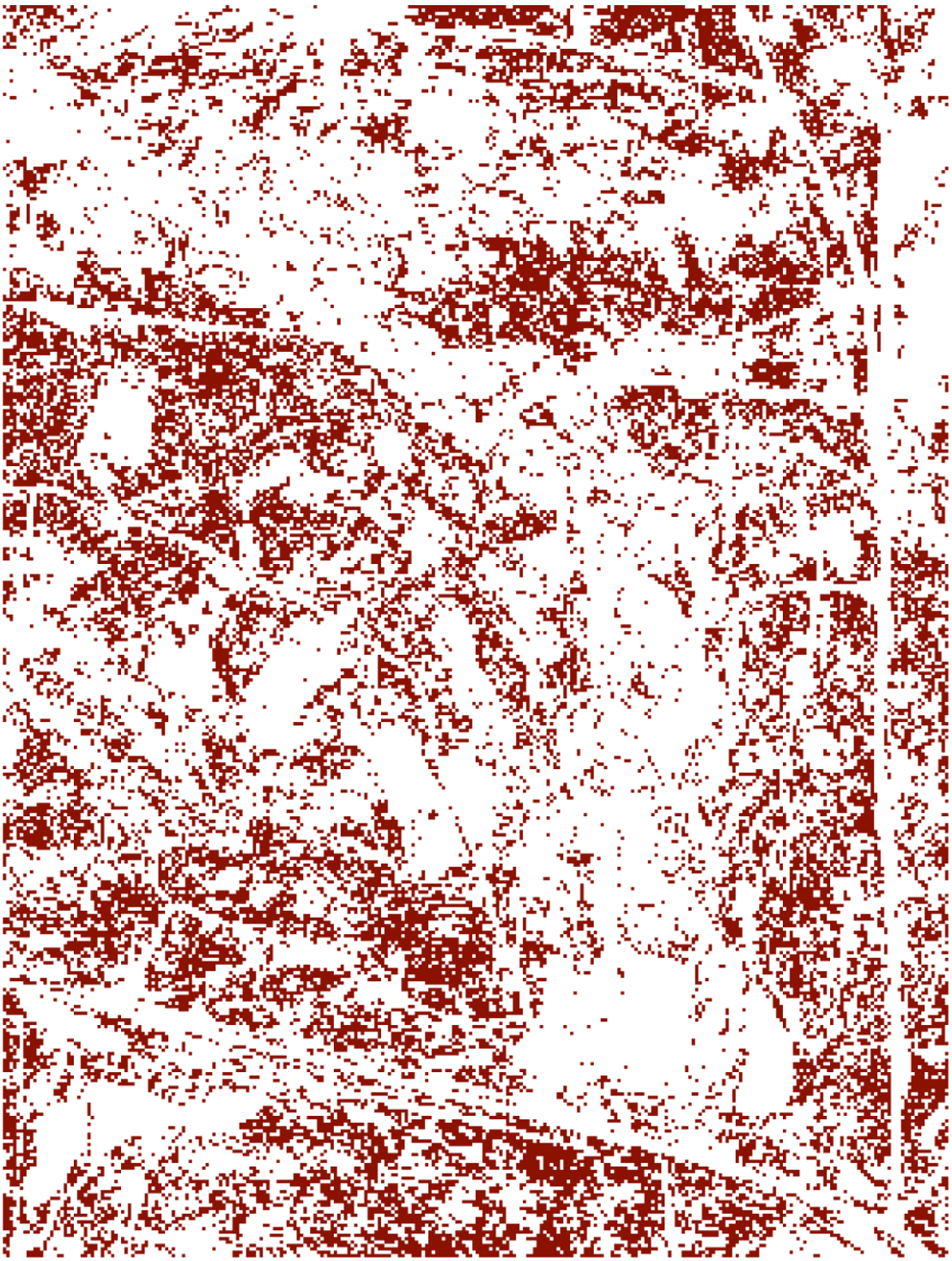


Figure 27. The Cluster 1 (of 418) Map for the Bluestem Site 134.

Blue_134. Cluster 14.

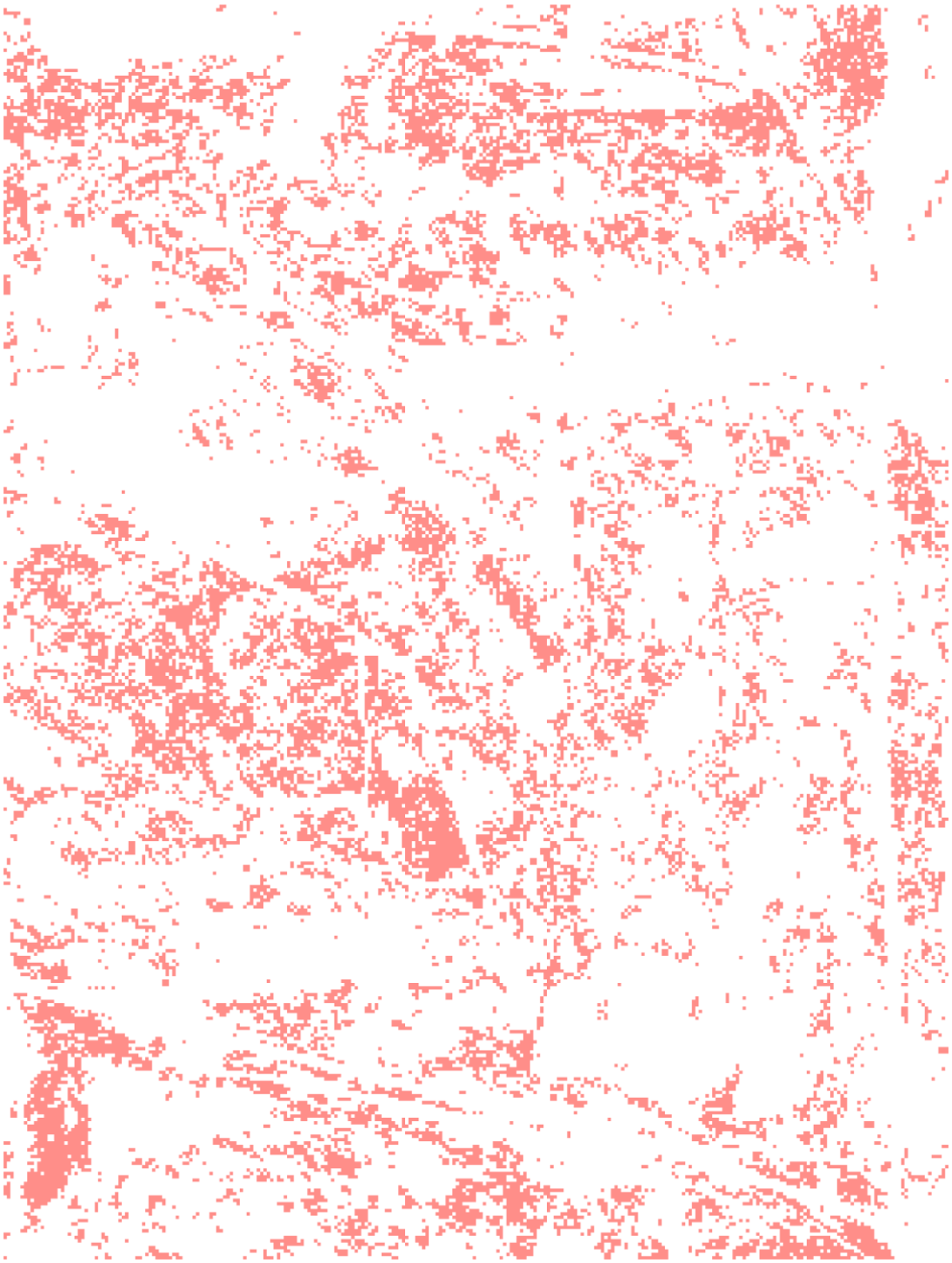


Figure 28. The Cluster 14 (of 418) Map for the Bluestem Site 134.

4.4. Bluestem Site 135

The color image for the Bluestem Site 135 is displayed in Figure 29. The bright pixel image of the analysis area is shown in Figure 30. The big (more than 24 members) cluster map is displayed in Figure 31. The brown cluster map is shown in Figure 32. The all cluster map is shown in Figure 33 for the 81 cluster case and in Figure 34 for the 238 cluster case.

Figure 29. The color map for the Bluestem Site 135.

Bluestem 135. RGB. Bright pixel.

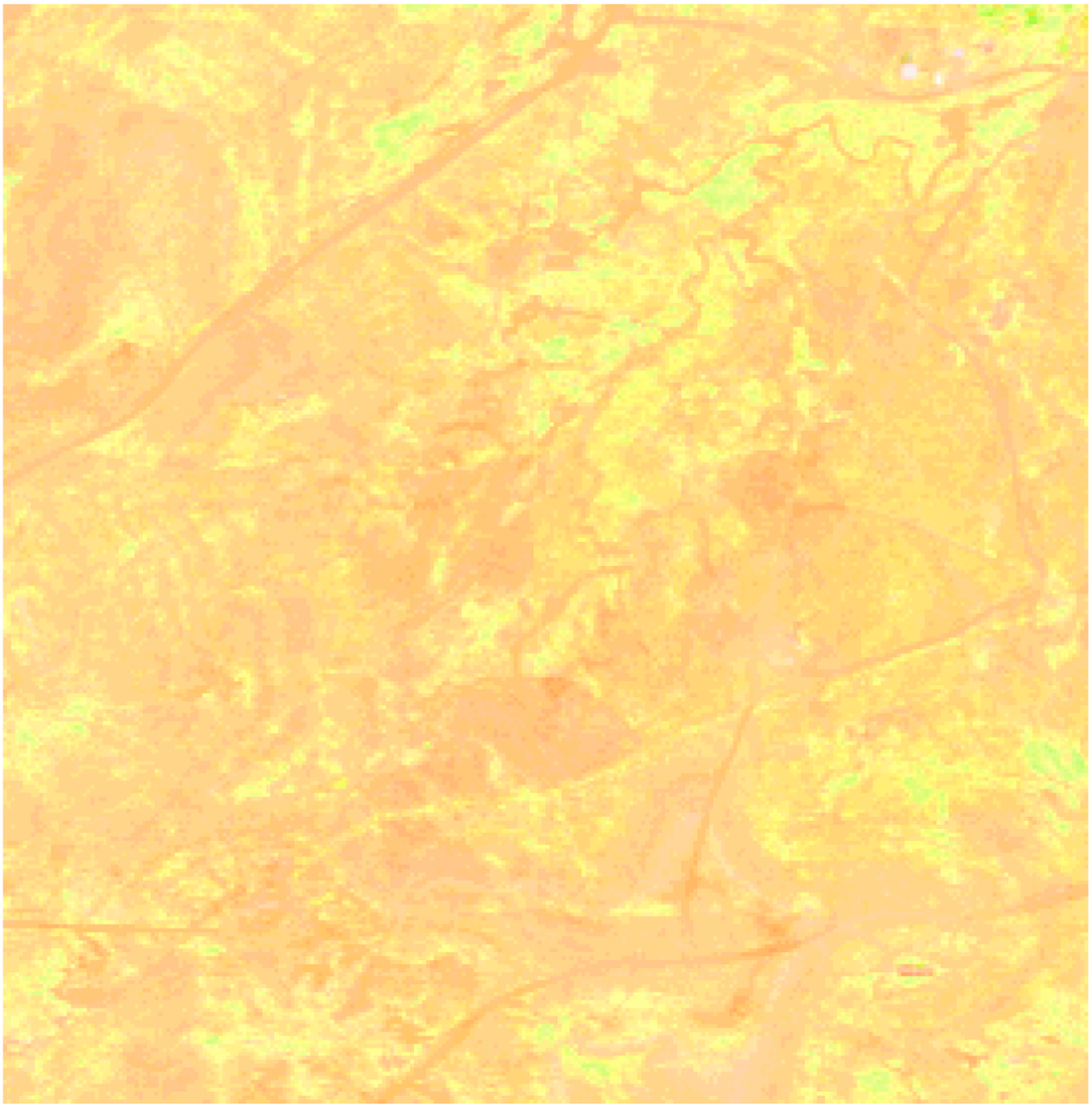


Figure 30. The Bright Pixel Image of the Bluestem Site 135.

Bluestem 135. Big Clusters (members > 24)

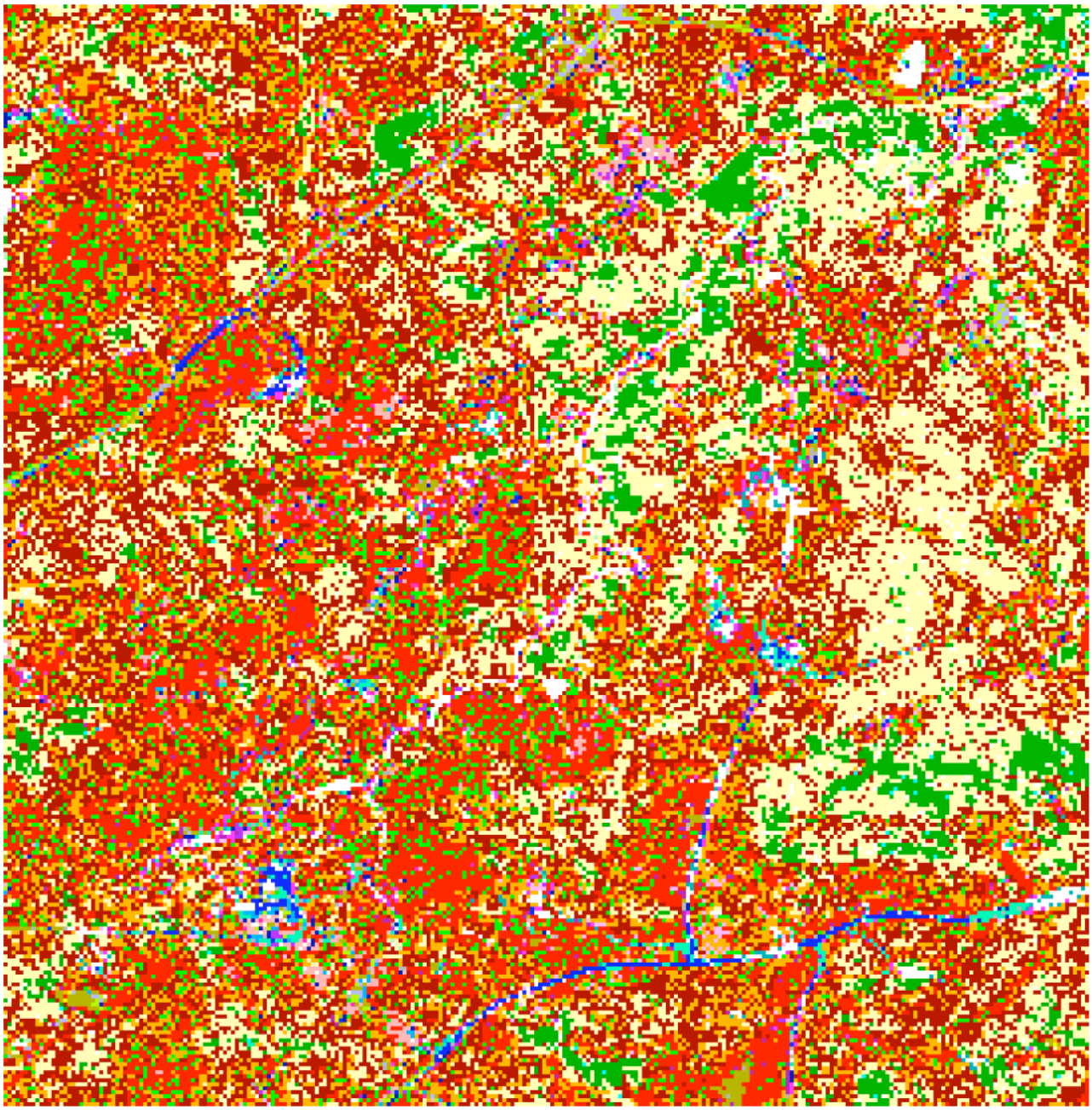


Figure 31. The Big Cluster map for the Bluestem Site 135.

Bluestem 135. All Brown Clusters (members > 2)

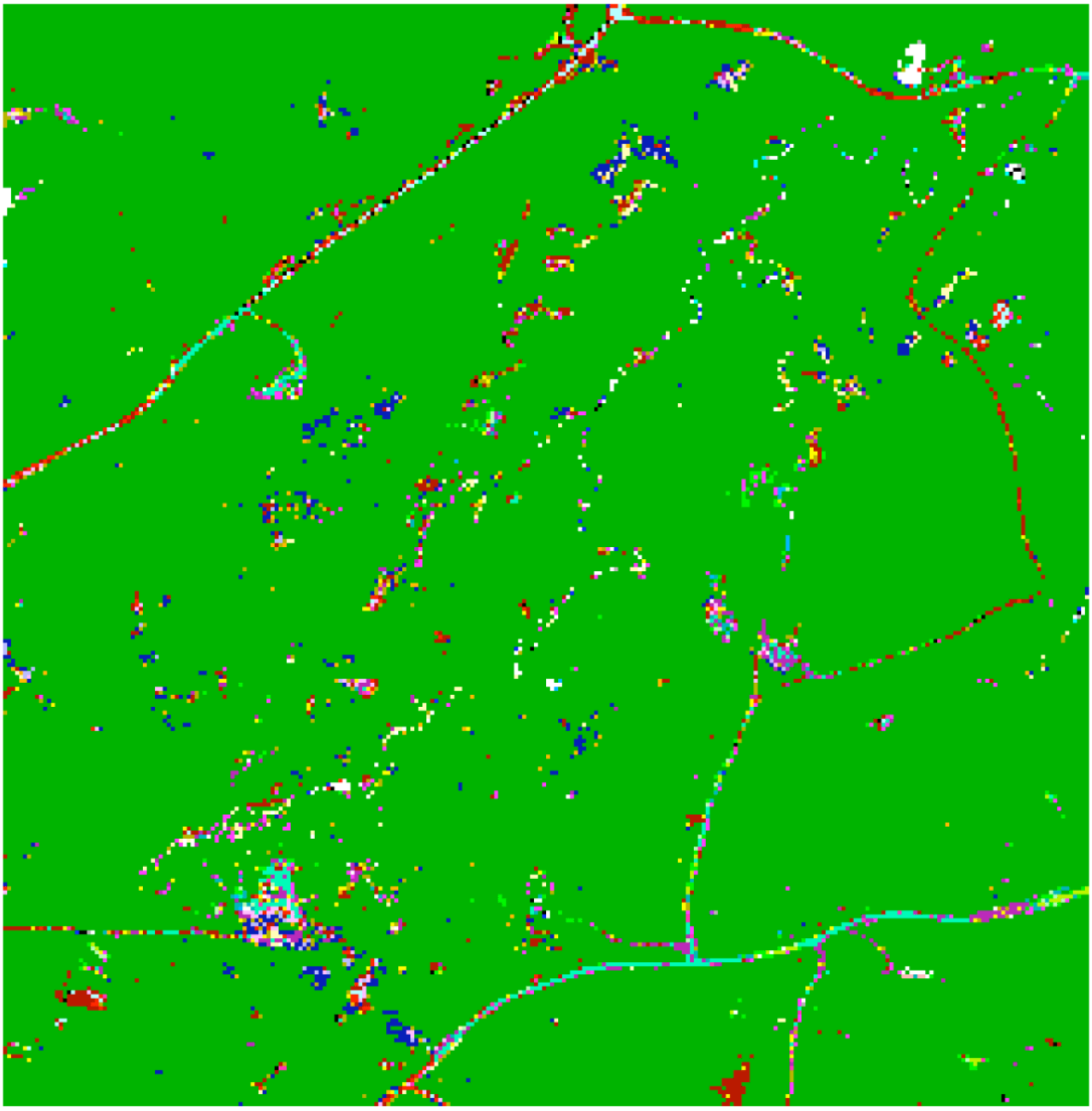


Figure 32. The Brown Cluster Map for the Bluestem Site 135.

Blue_135. All Clusters

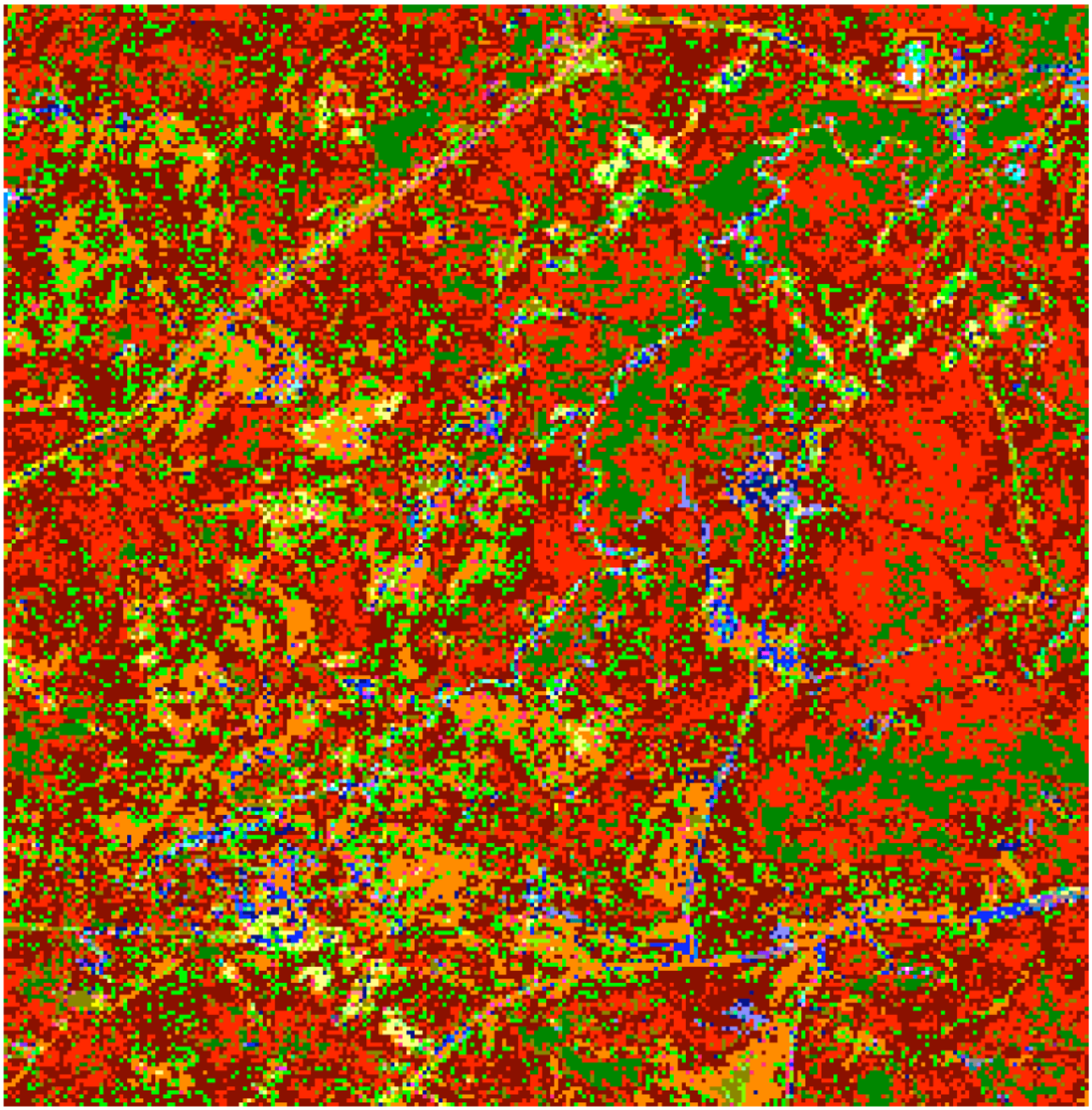


Figure 33. The 81 Cluster Map for the Bluestem Site 135.

Blue_135. All Clusters

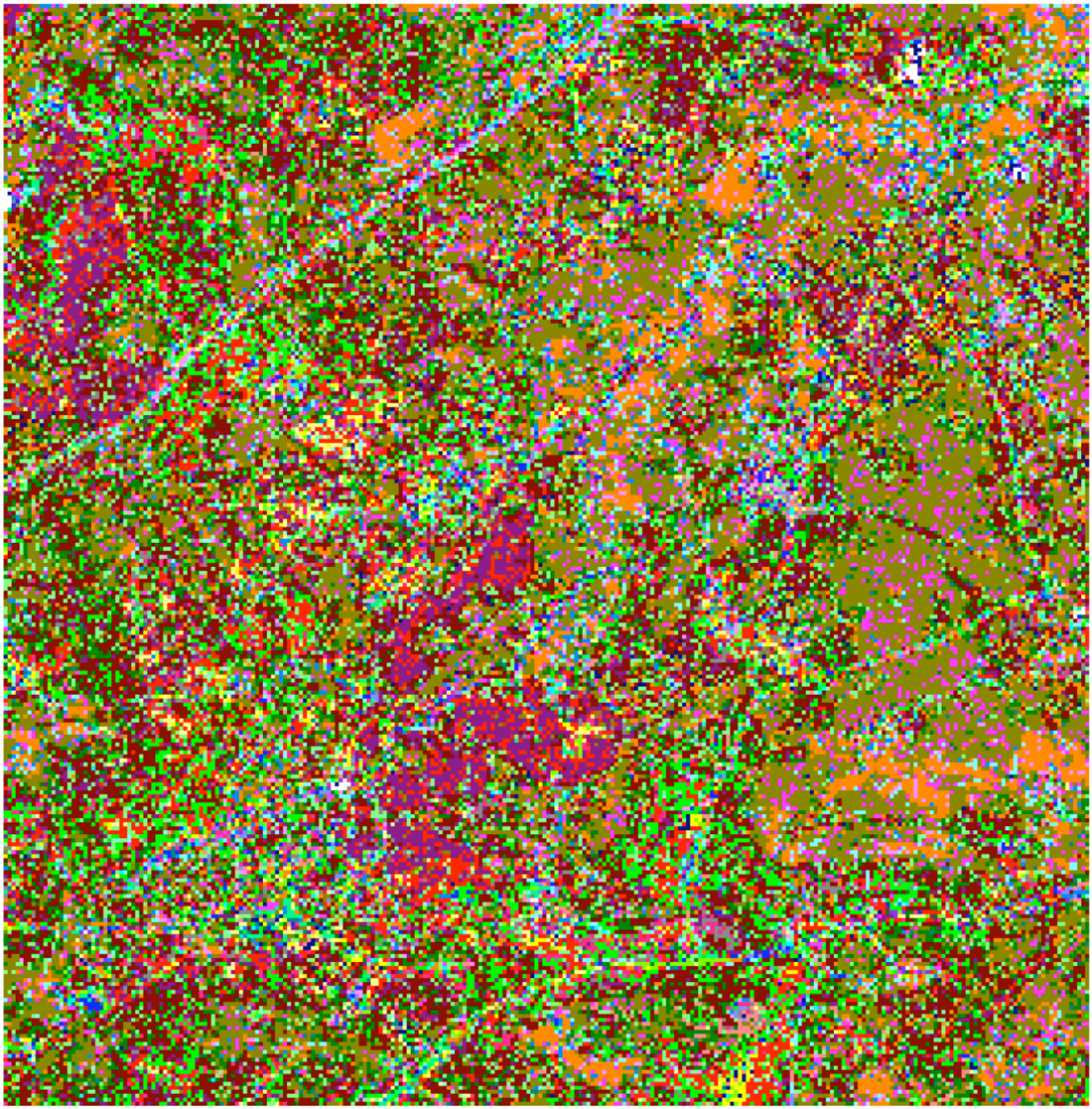


Figure 34. The 238 Cluster Map for the Bluestem Site 135.

4.5. Bluestem Site 136

The color image for the Bluestem Site 136 is displayed in Figure 35. The bright pixel image of the analysis area is shown in Figure 36. The big (more than 14 members) cluster map is displayed in Figure 37. The brown cluster map is shown in Figure 38. The all cluster map is shown in Figure 39 for the 61 cluster case and in Figure 40 for the 185 cluster case.

Figure 35. The color map for the Bluestem Site 136.

Bluestem 136. RGB. Bright pixel.

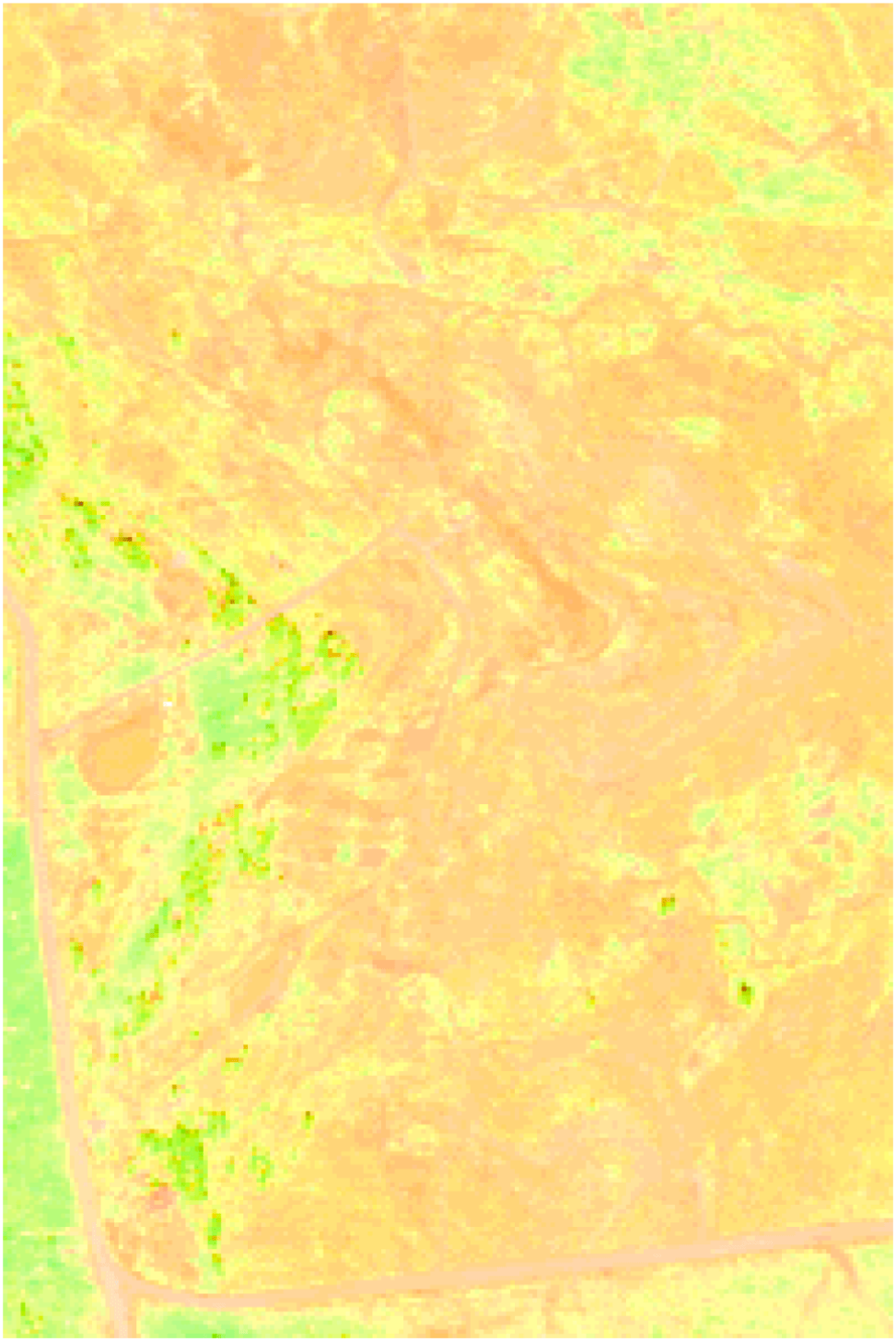


Figure 36. The Bright Pixel Image of the Bluestem Site 136.

Bluestem 136. Big Clusters (members > 14)

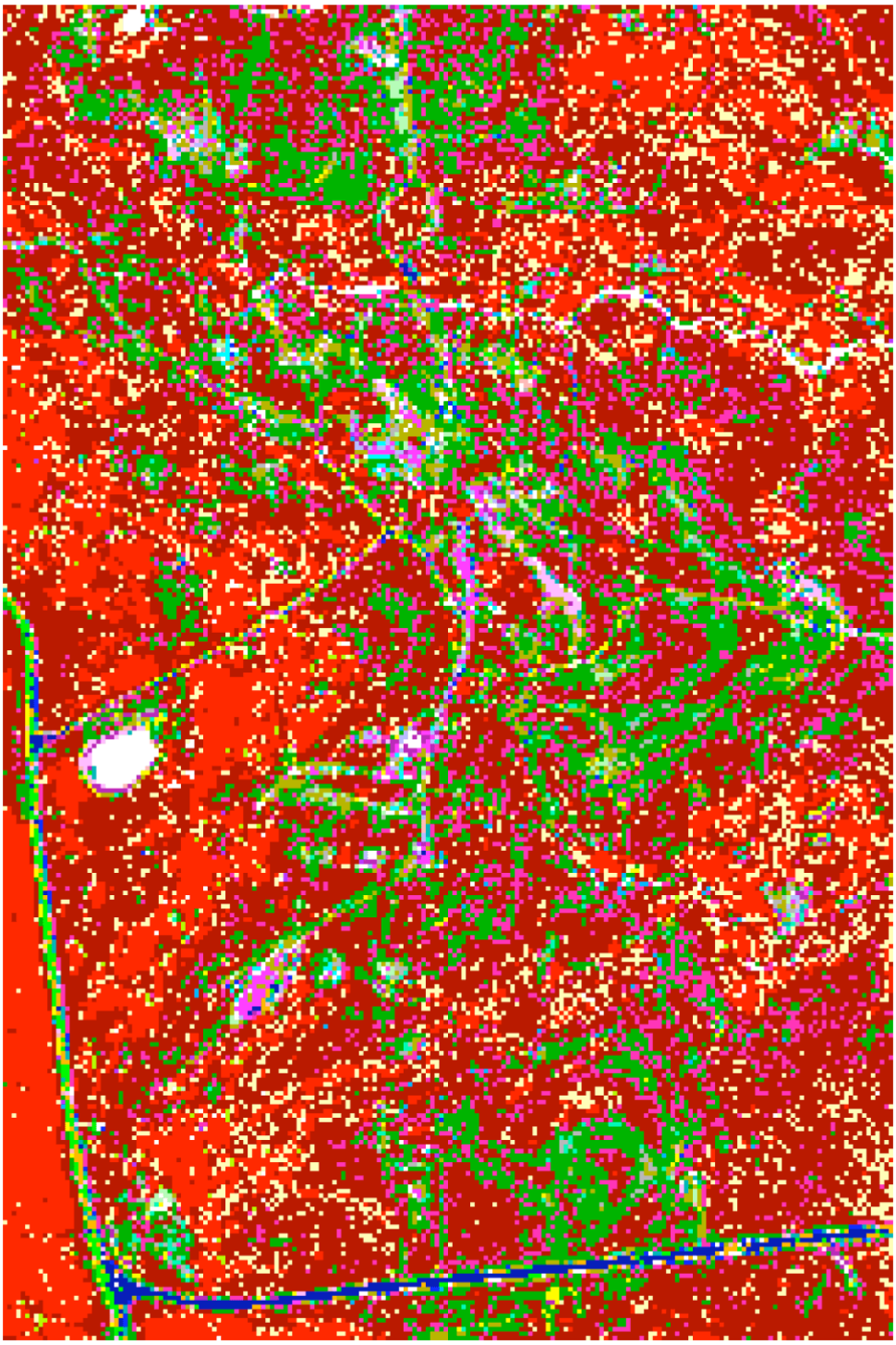


Figure 37. The Big Cluster map for the Bluestem Site 136.

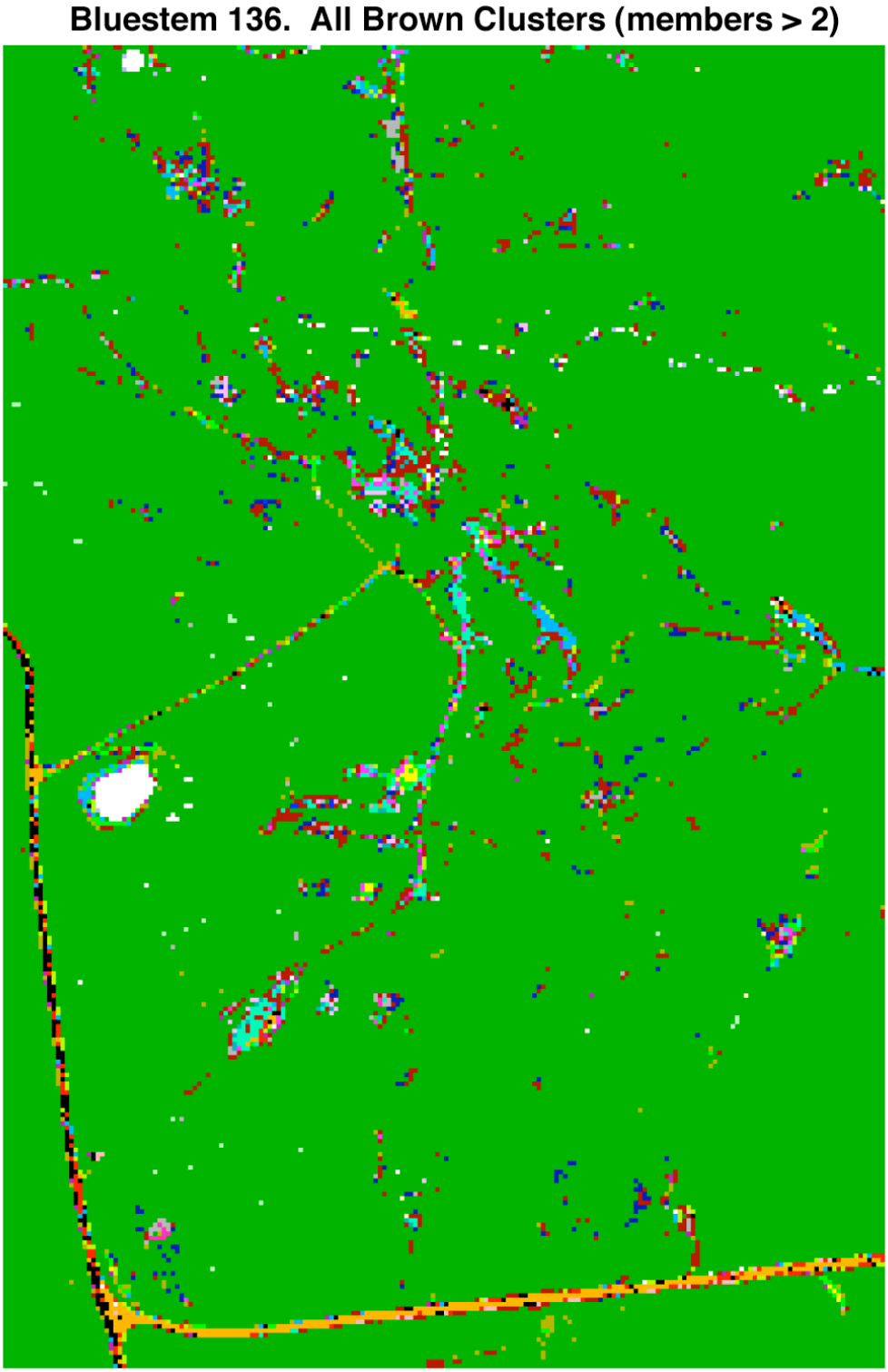


Figure 38. The Brown Cluster Map for the Bluestem Site 136.

Blue_136. All Clusters

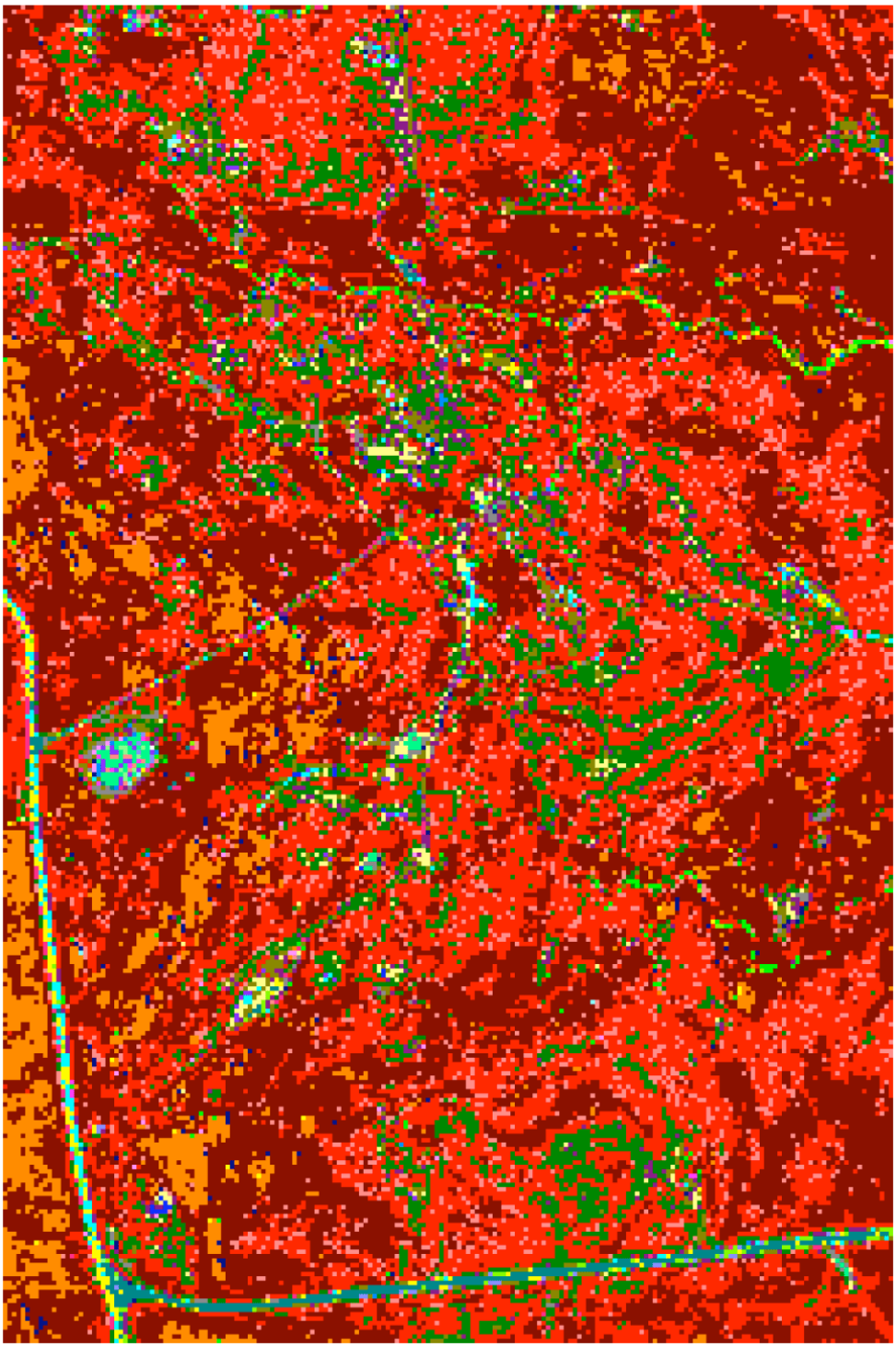


Figure 39. The 61 Cluster Map for the Bluestem Site 136.

Blue_136. All Clusters

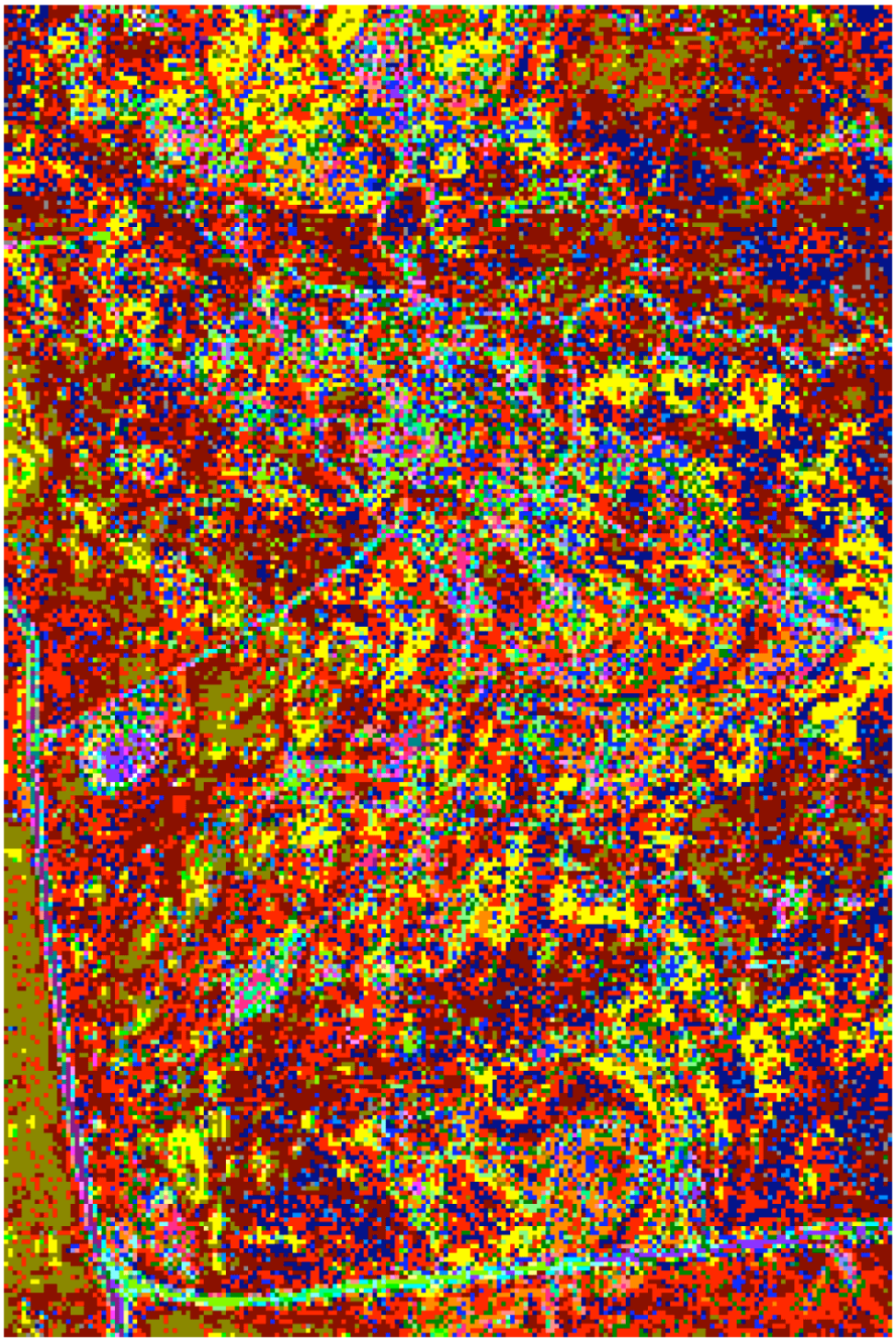


Figure 40. The 185 Cluster Map for the Bluestem Site 136.

4.6. Tallgrass Prairie Site 118

The color image for the Tallgrass Prairie Site 118 is displayed in Figure 41. The bright pixel image of the analysis area is shown in Figure 42. The big (more than 13 members) cluster map is displayed in Figure 43. The brown cluster map is shown in Figure 44. The all cluster map is shown in Figure 45 for the 49 cluster case and in Figure 46 for the 125 cluster case.

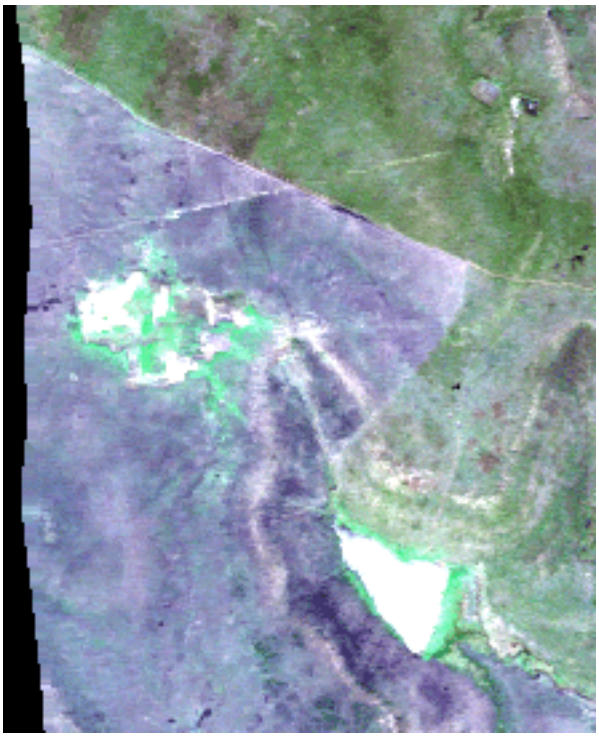


Figure 41. The color map for the Tallgrass Prairie Site 118.

TGP 118. RGB. Bright pixel.

Figure 42. The Bright Pixel Image of the Tallgrass Prairie Site 118.

TGP 118. Big Clusters (members > 13)

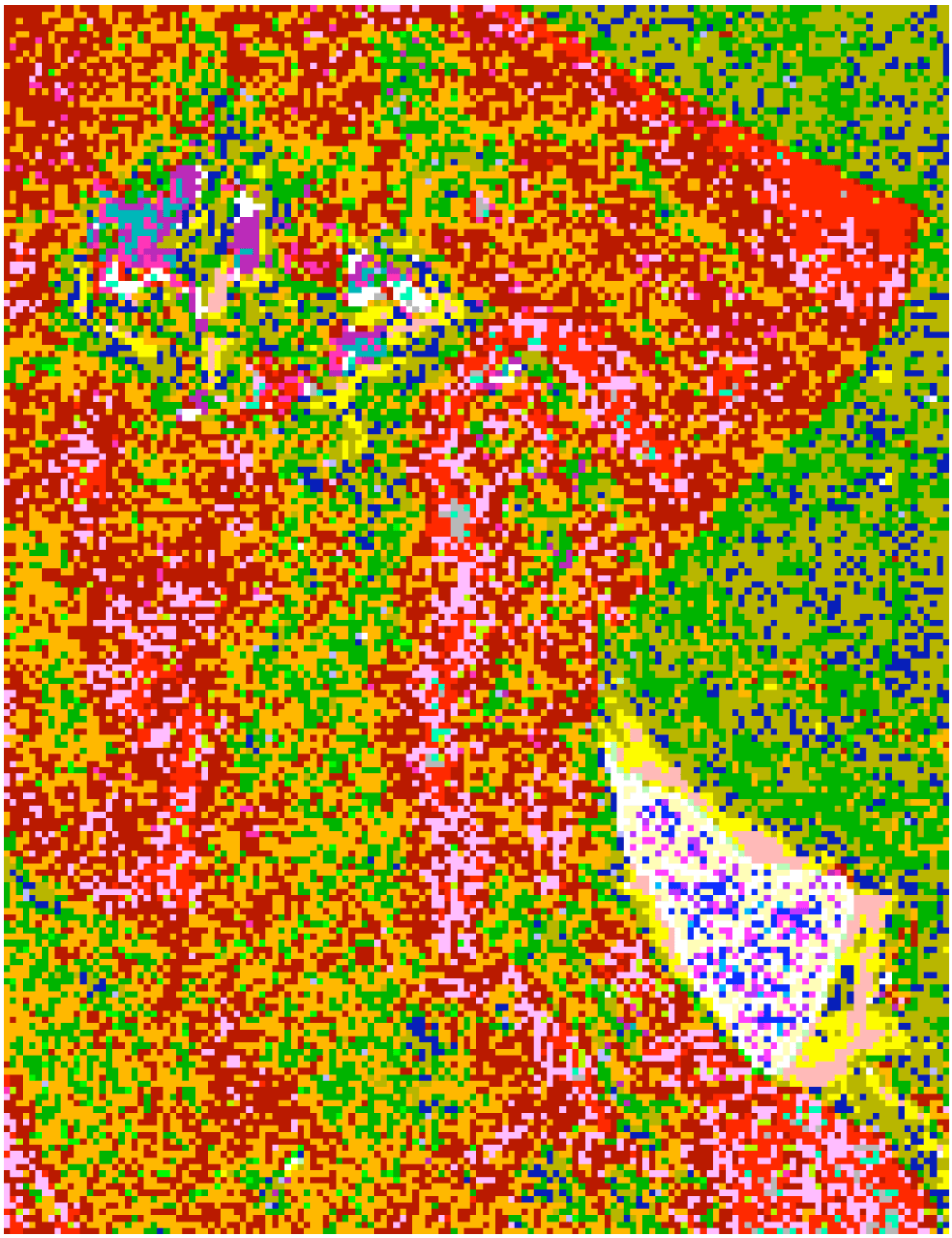


Figure 43. The Big Cluster map for the Tallgrass Prairie Site 118.

TGP 118. All Clusters (members > 2)

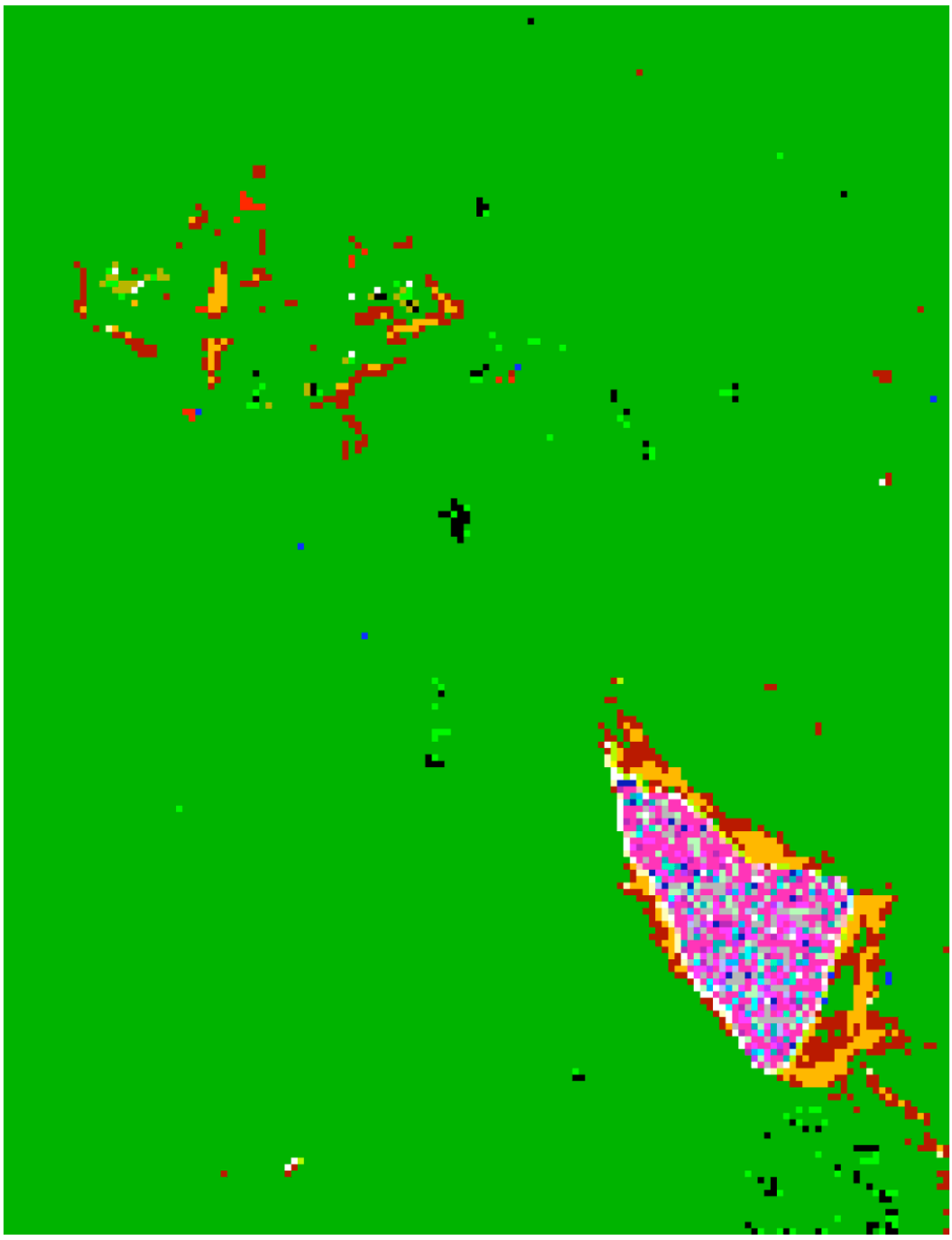


Figure 44. The Brown Cluster Map for the Tallgrass Prairie Site 118.

TGP_118. All Clusters

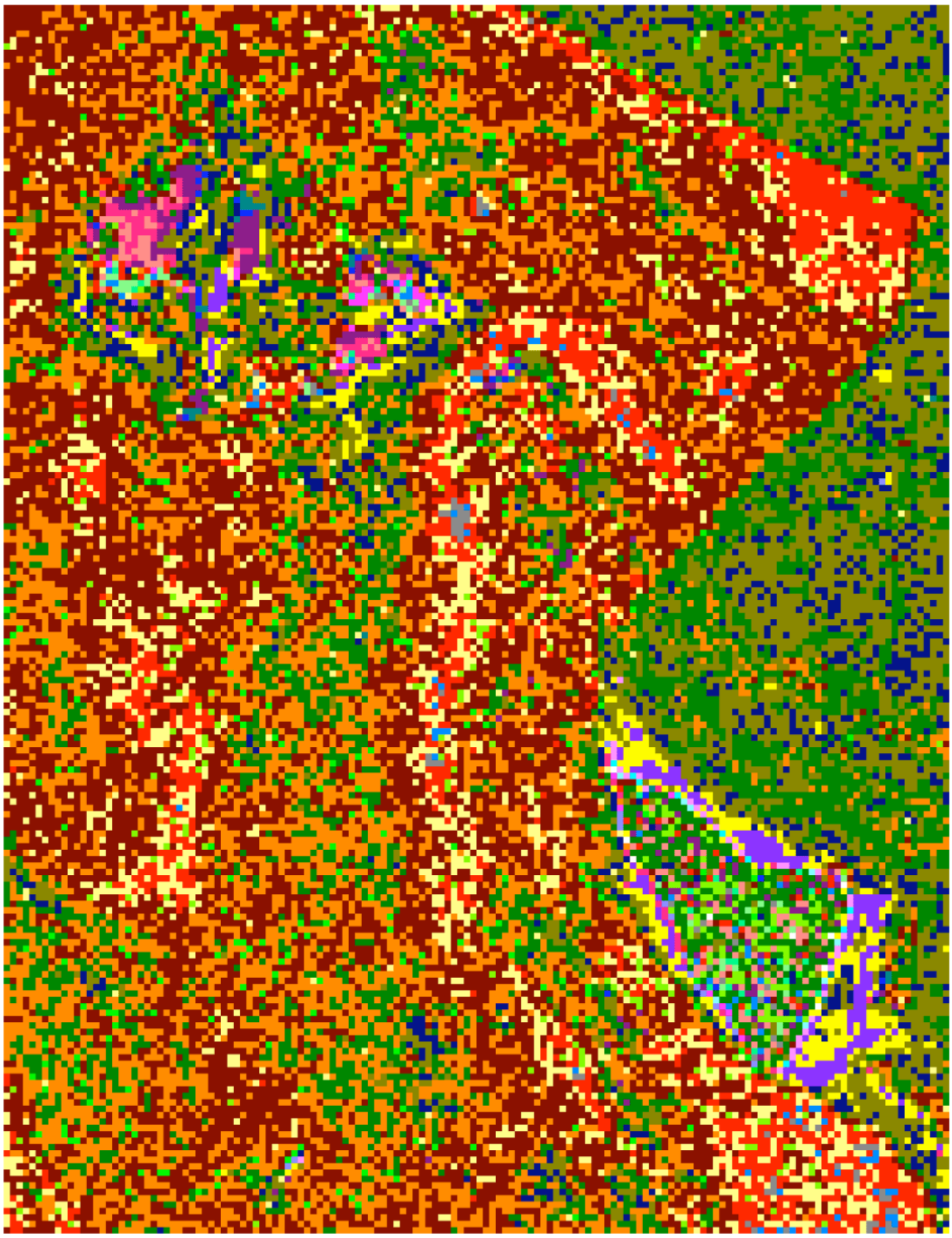


Figure 45. The 49 Cluster Map for the Tallgrass Prairie Site 118.

TGP_118. All Clusters

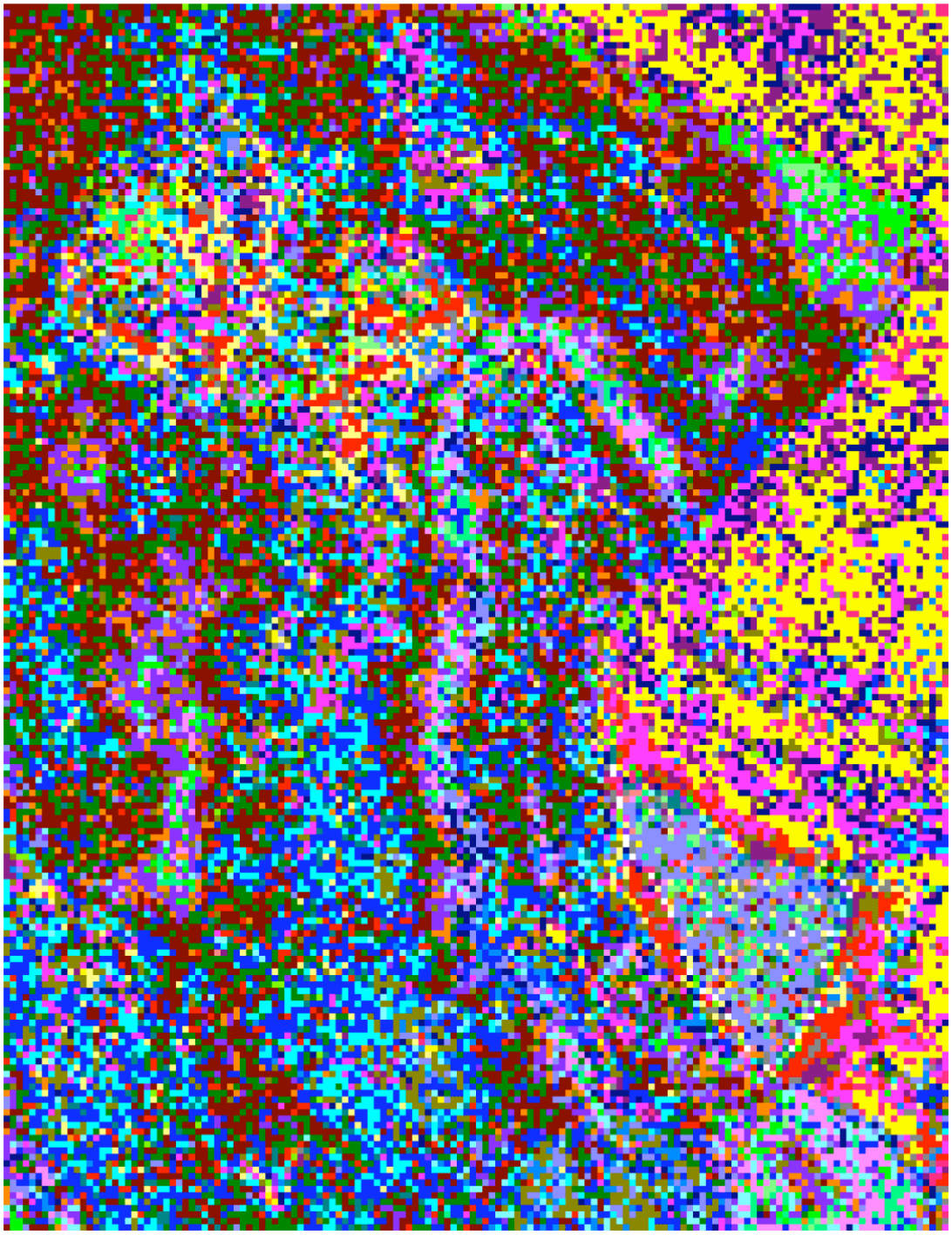


Figure 46. The 125 Cluster Map for the Tallgrass Prairie Site 118.

4.7. Tallgrass Prairie Site 122

The color image for the Tallgrass Prairie Site 122 is displayed in Figure 47. The bright pixel image of the analysis area is shown in Figure 48. The big (more than 24 members) cluster map is displayed in Figure 49. The brown cluster map is shown in Figure 50. Figures 51 to 54 are enlargements of four regions of Figure 50. The all cluster map is shown in Figure 55 for the 74 cluster case and in Figure 56 for the 257 cluster case.

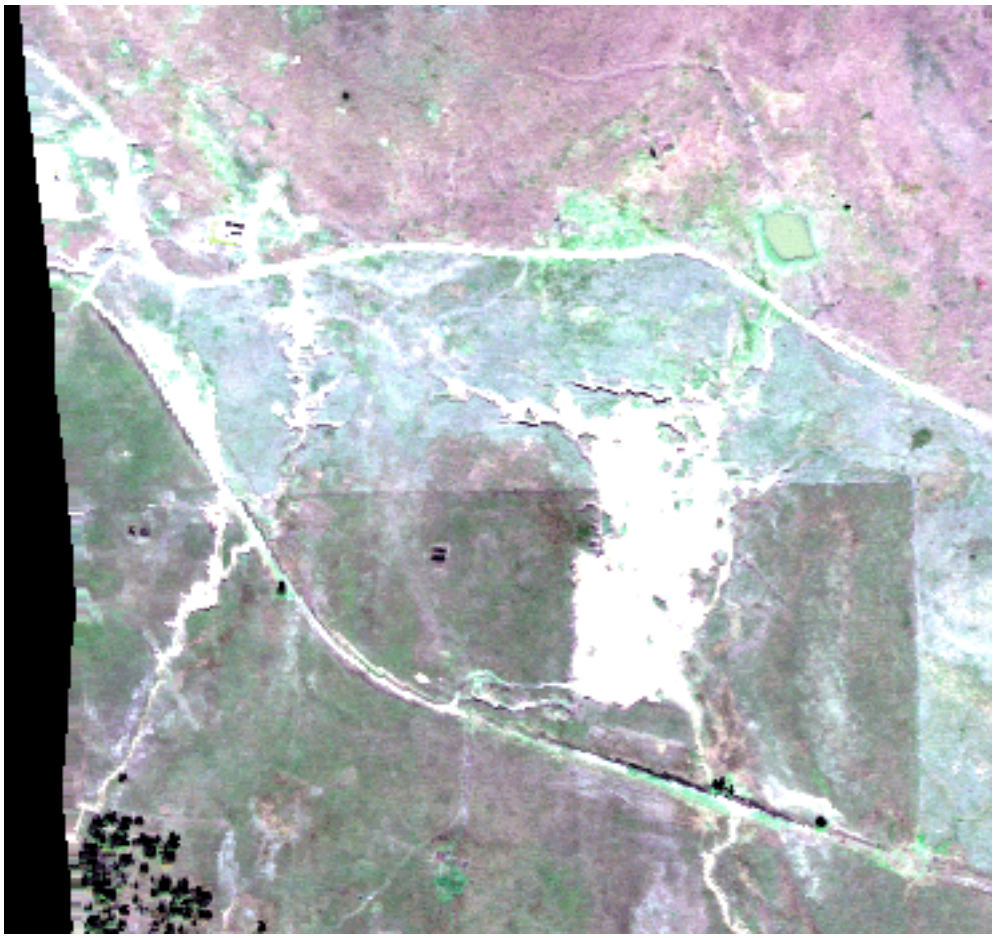


Figure 47. The color map for the Tallgrass Prairie Site 122.

TGP 122. RGB. Bright pixel.

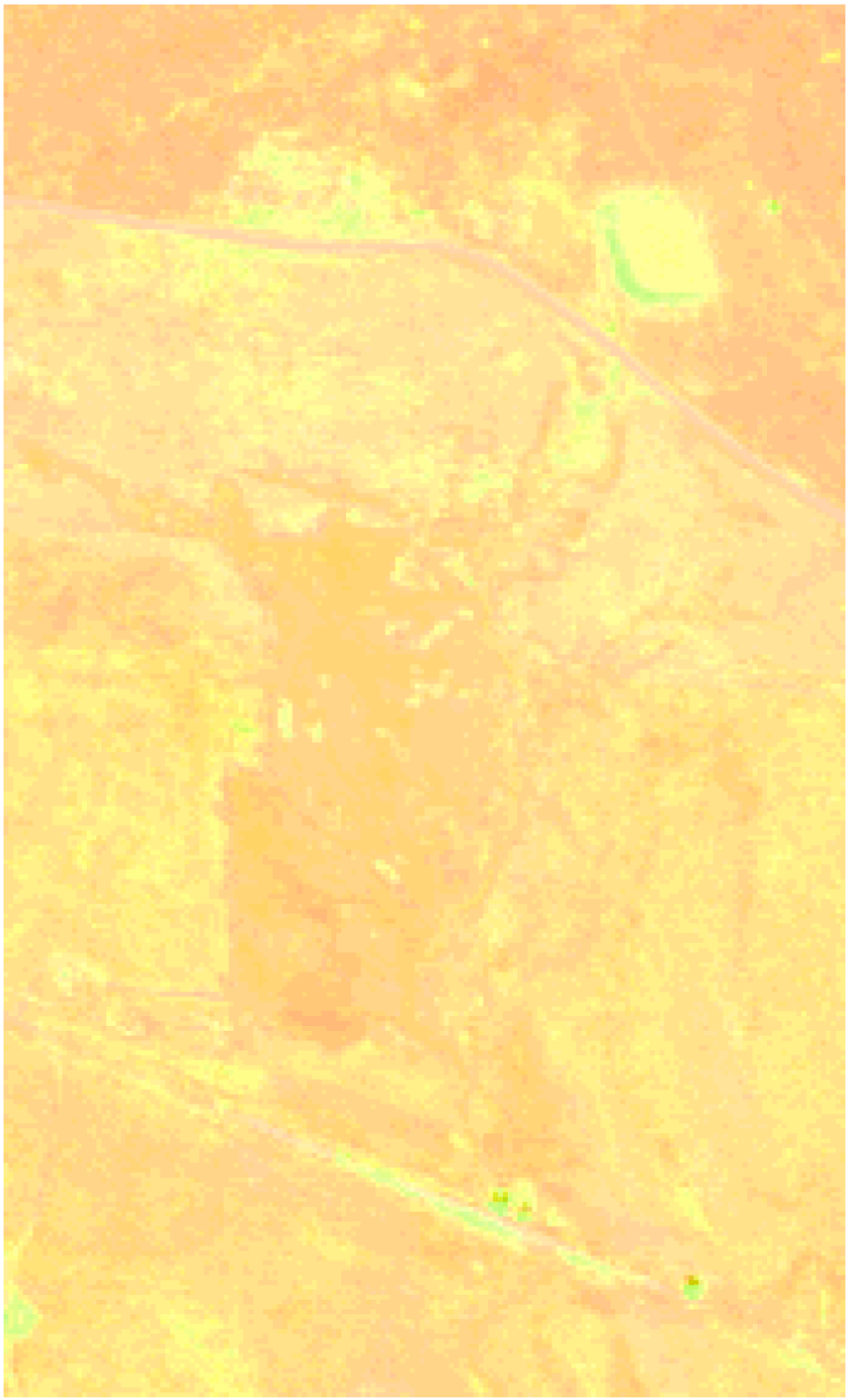


Figure 48. The Bright Pixel Image of the Tallgrass Prairie Site 122.

TGP 122. Big Clusters (members > 24)

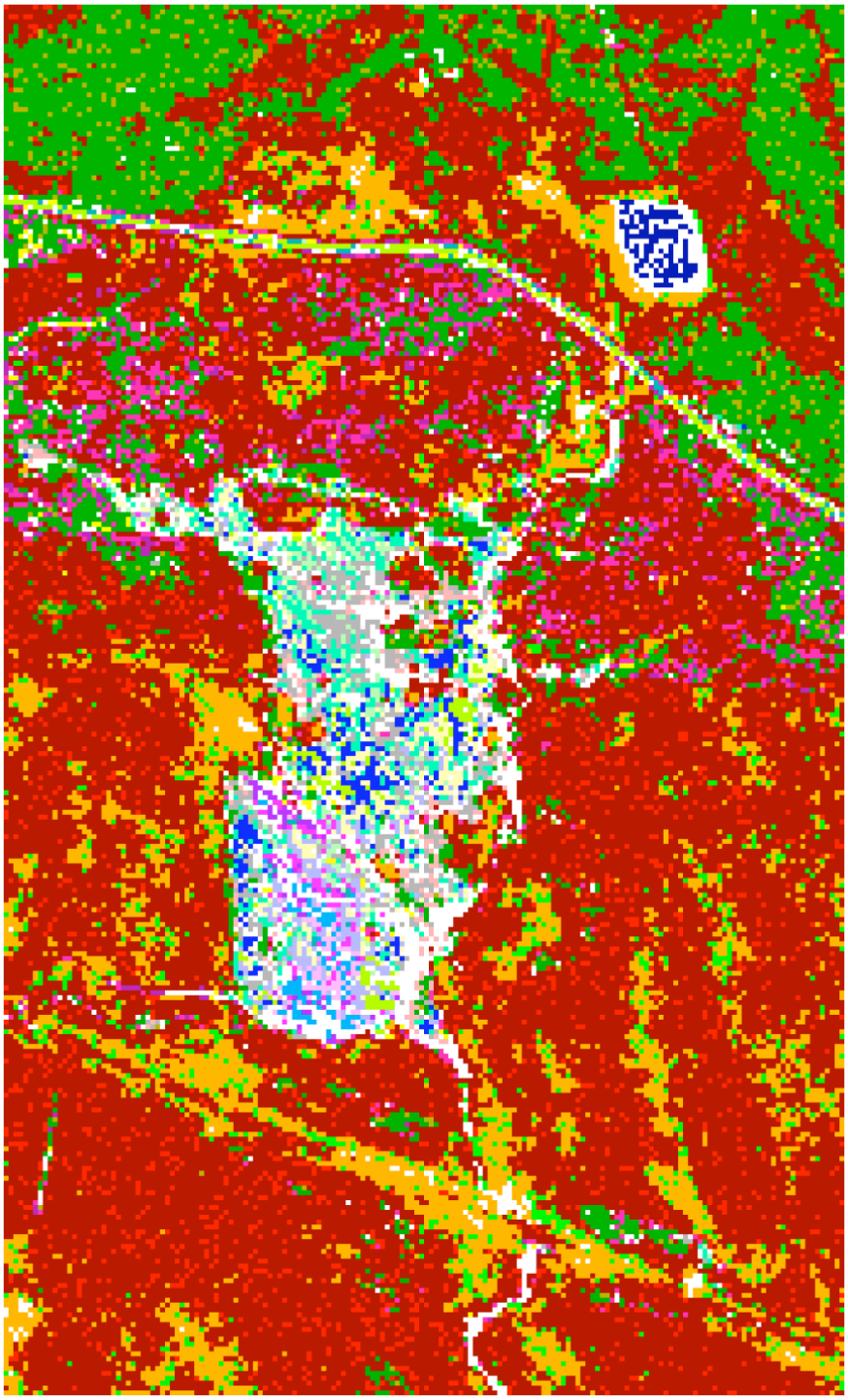


Figure 49. The Big Cluster map for the Tallgrass Prairie Site 122.

TGP 122. All Clusters (members > 2)

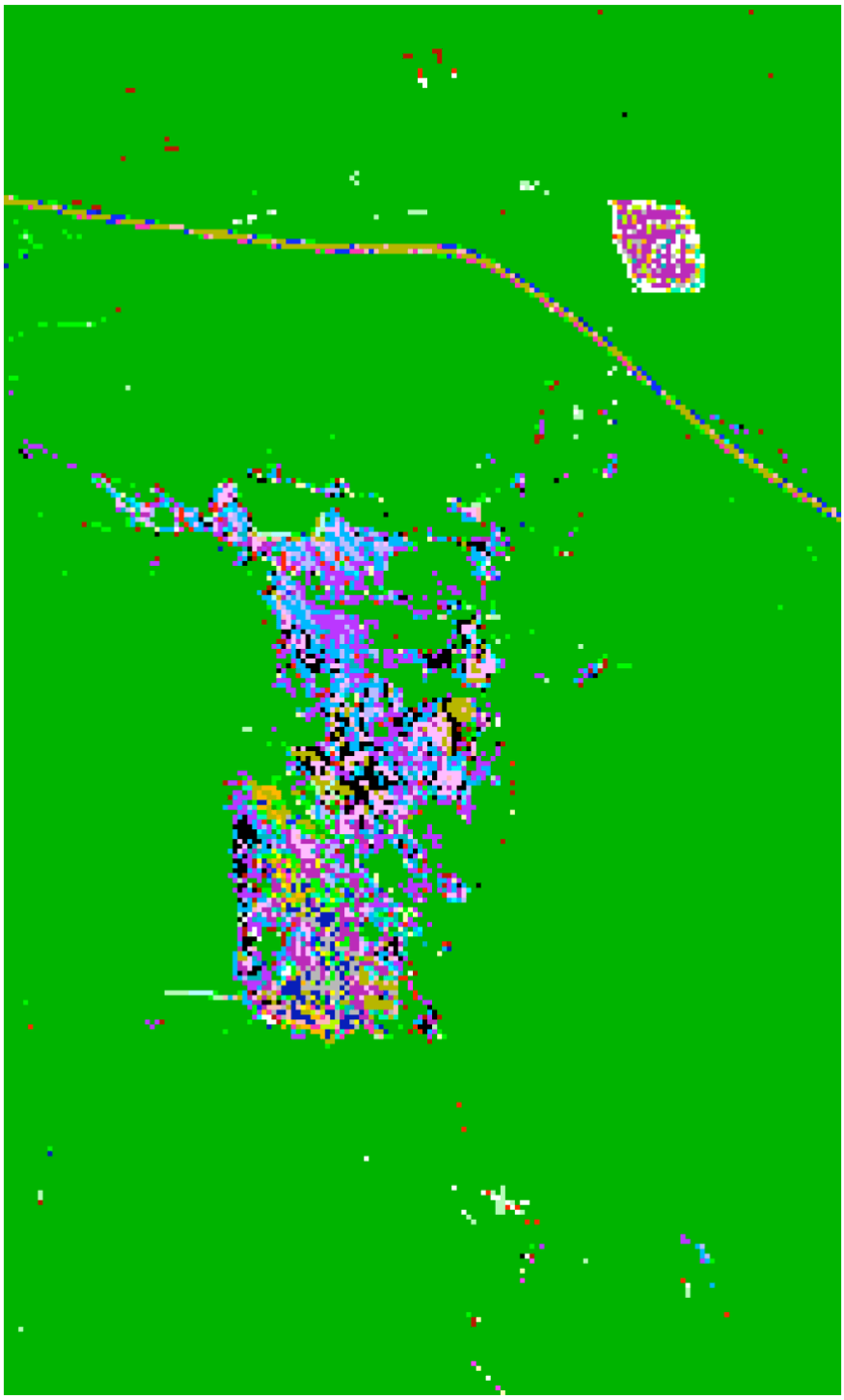


Figure 50. The Brown Cluster Map for the Tallgrass Prairie Site 122.

Figure 51. Enlargement of the pond in the upper right of Figure 50.

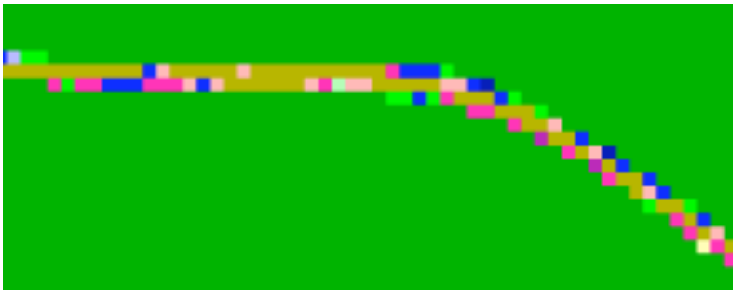


Figure 52. Enlargement of the road in the upper center of Figure 50.

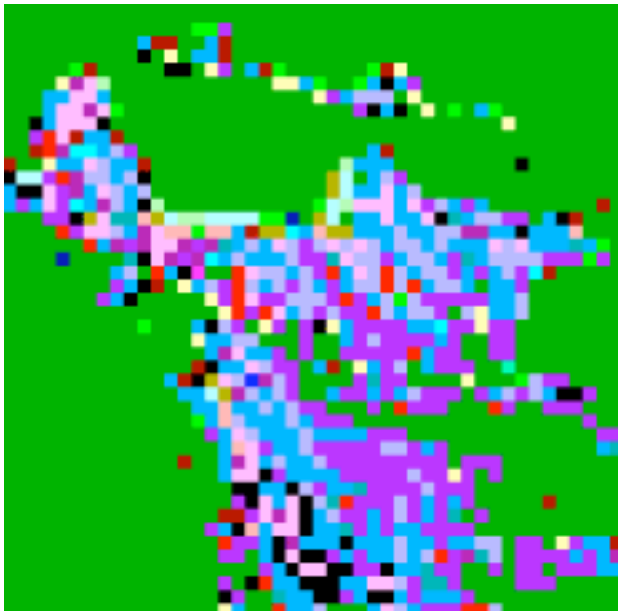


Figure 53. Enlargement of the Northwest region of the scar in Figure 50.

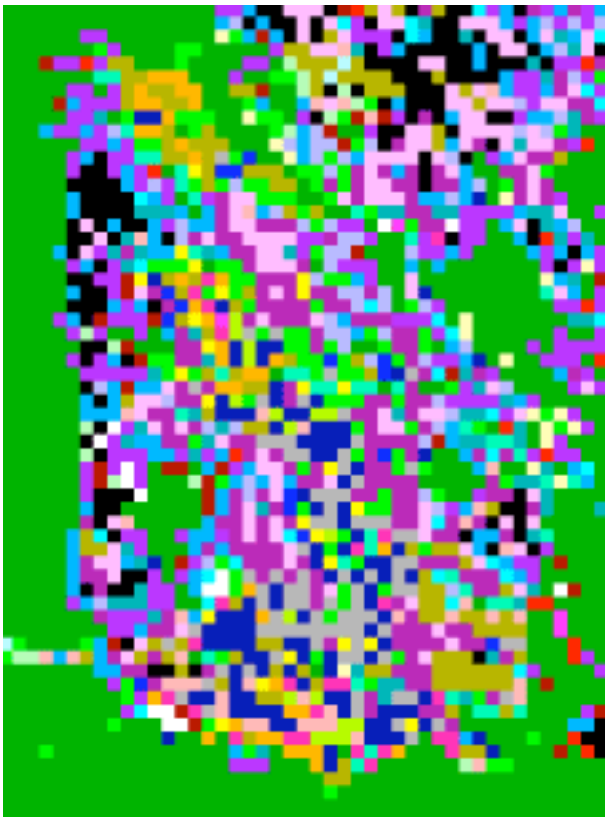


Figure 54. Enlargement of the Southwest region of the scar in Figure 50.

TGP_122. All Clusters

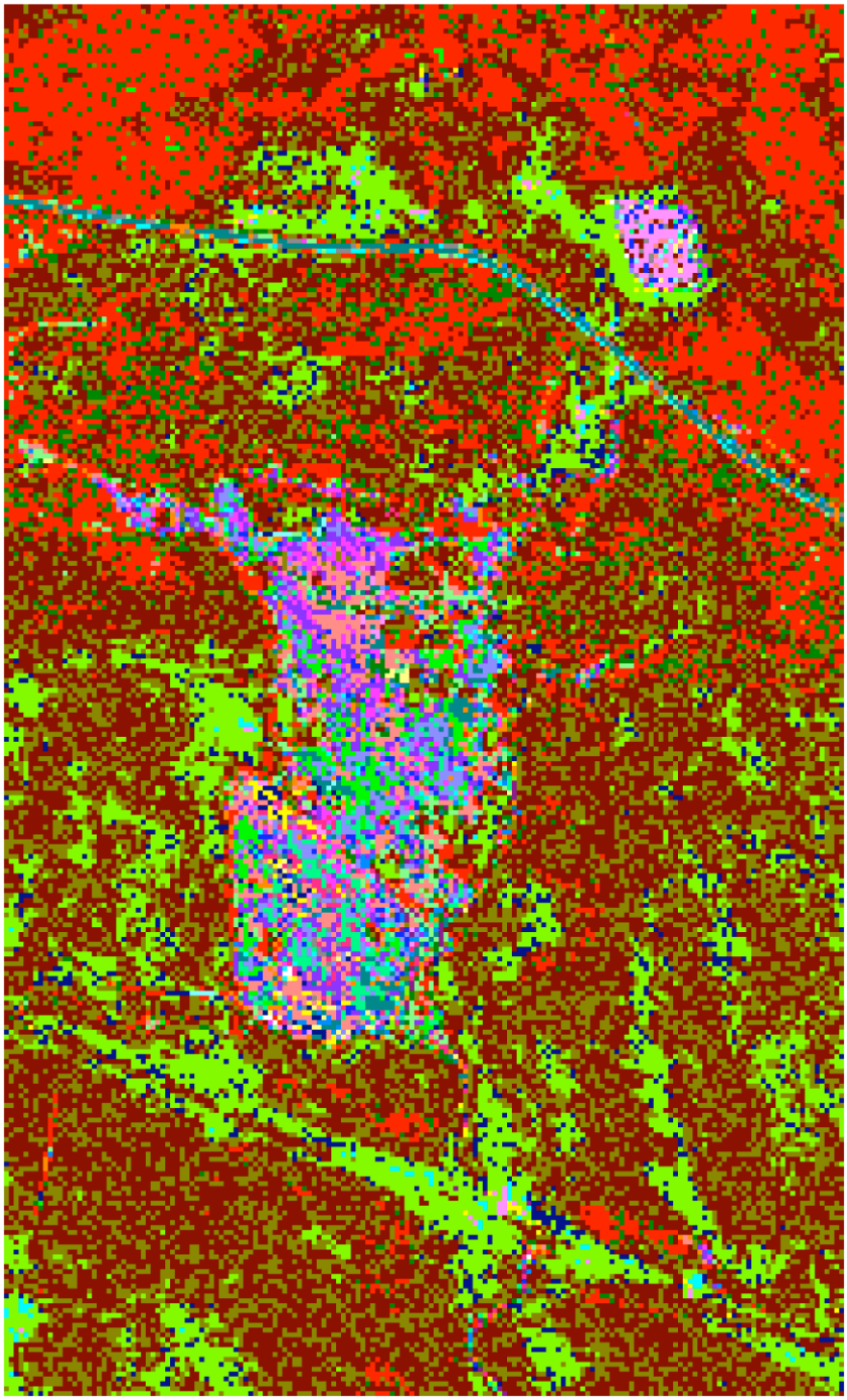


Figure 55. The 74 Cluster Map for the Tallgrass Prairie Site 122.

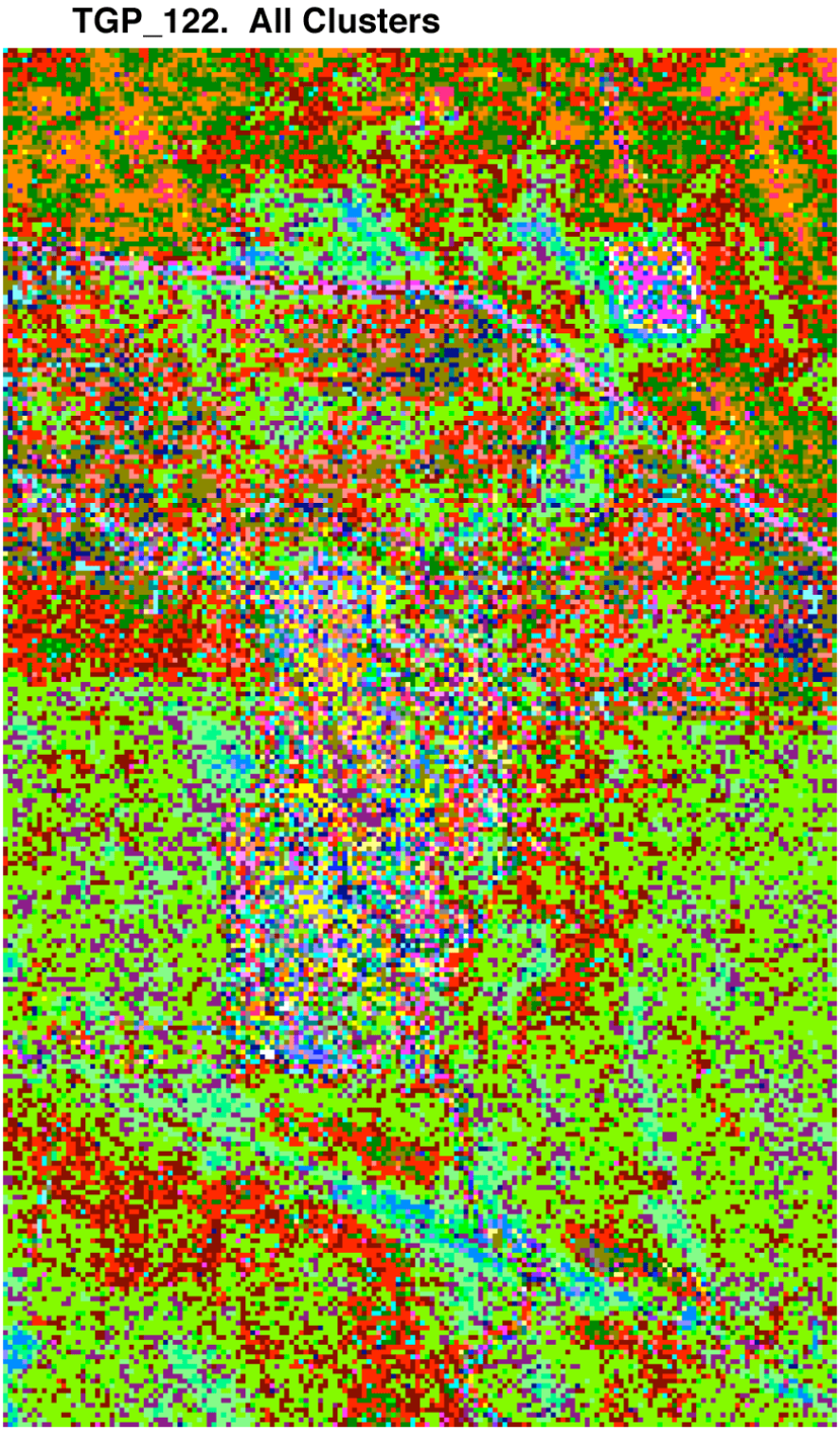


Figure 56. The 257 Cluster Map for the Tallgrass Prairie Site 122.

4.8. Zink Ranch Site 127

The color image for the Zink Ranch Site 127 is displayed on the left side of Figure 57 (which is identical to Figure 3). The bright pixel image of the analysis area is shown in Figure 58. The big (more than 14 members) cluster map is displayed in Figure 59. The brown cluster map is shown in Figure 60. The all cluster map is shown in Figure 61 for the 35 cluster case and in Figure 62 for the 115 cluster case.

Figure 57. The color map for the Zink Ranch Site 127.

Site 127. RGB. Bright pixel.

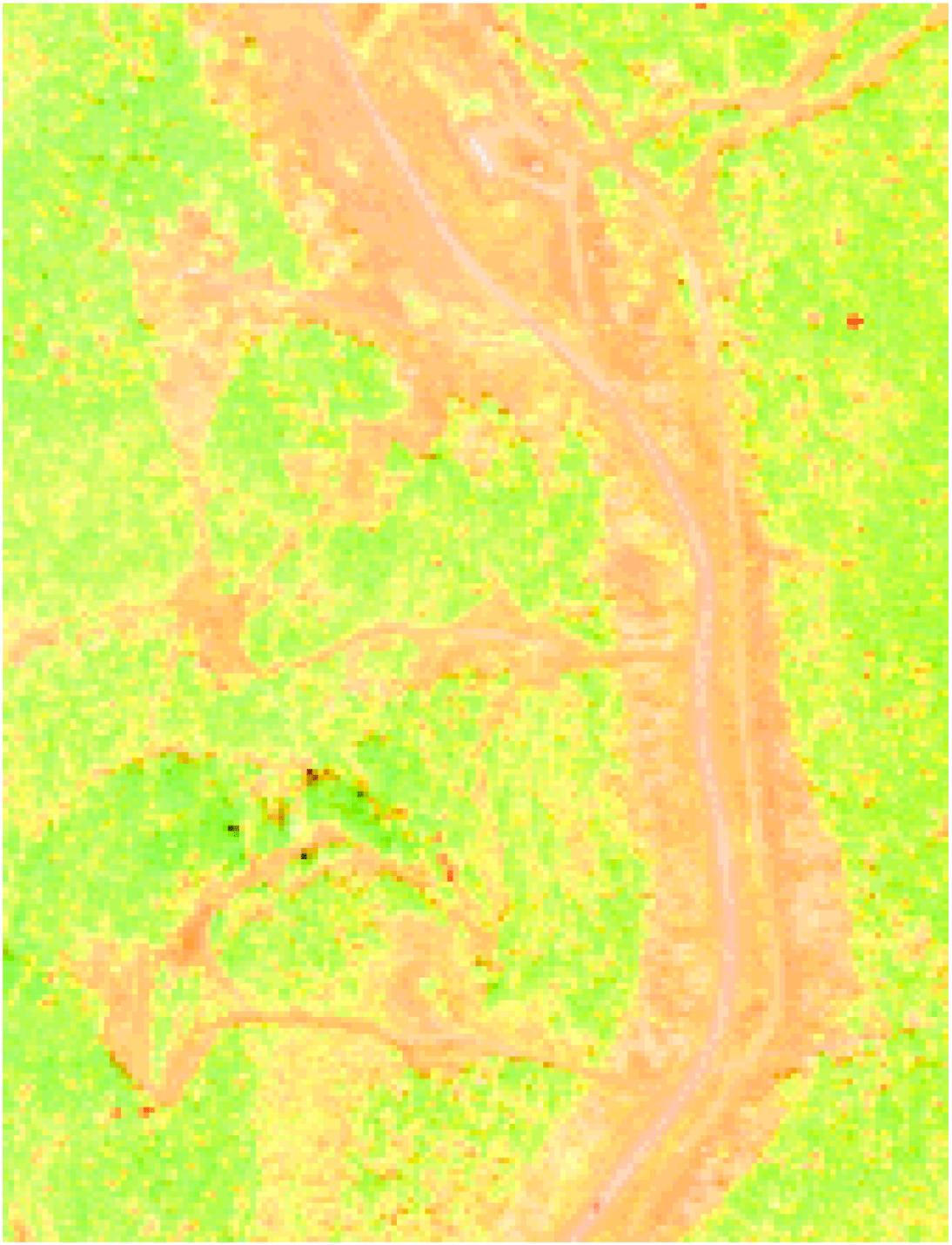


Figure 58. The Bright Pixel Image of the Zink Ranch Site 127.

Site 127. Big Clusters (members > 8)

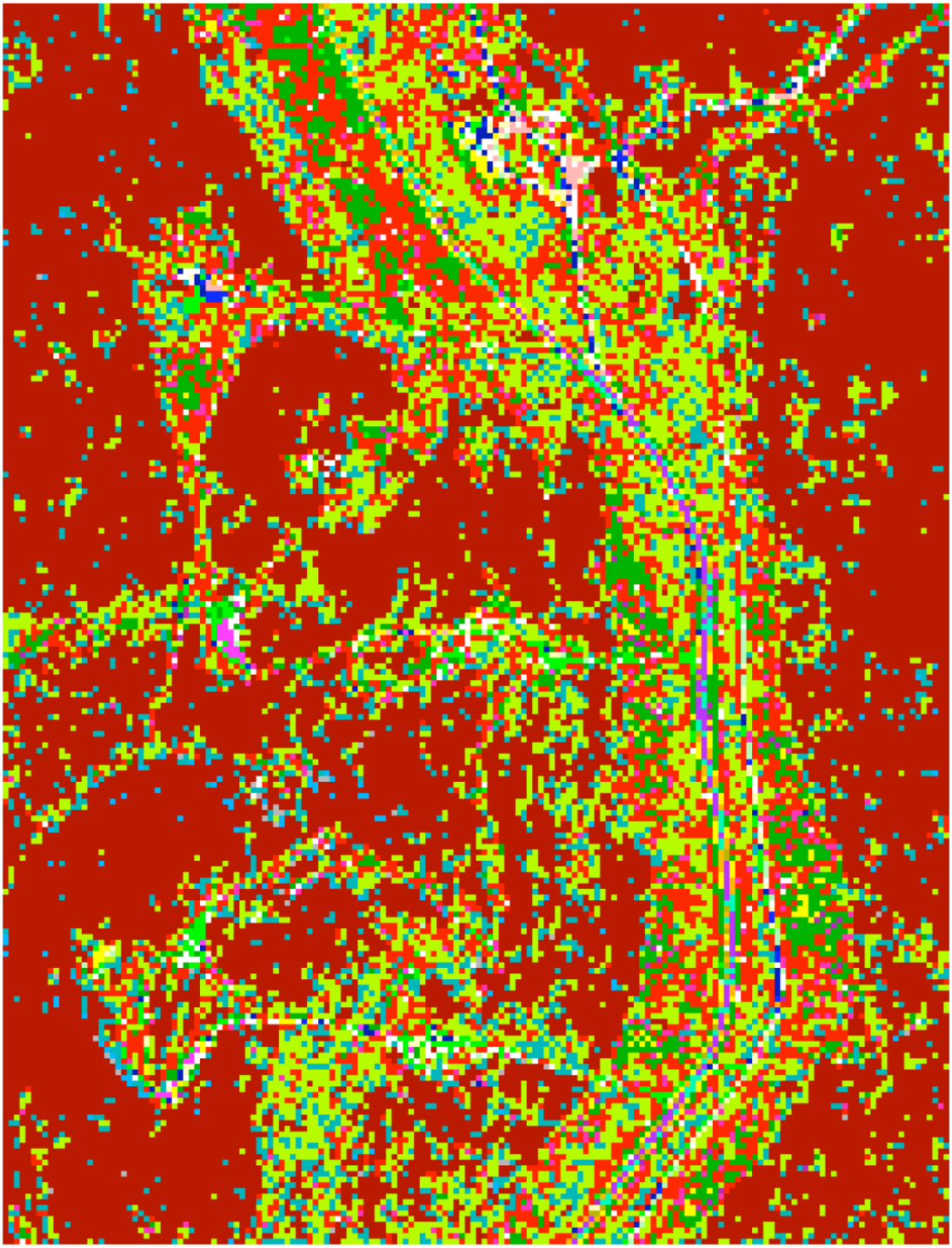


Figure 59. The Big Cluster map for the Zink Ranch Site 127.

Site 127. All Clusters (members > 2)

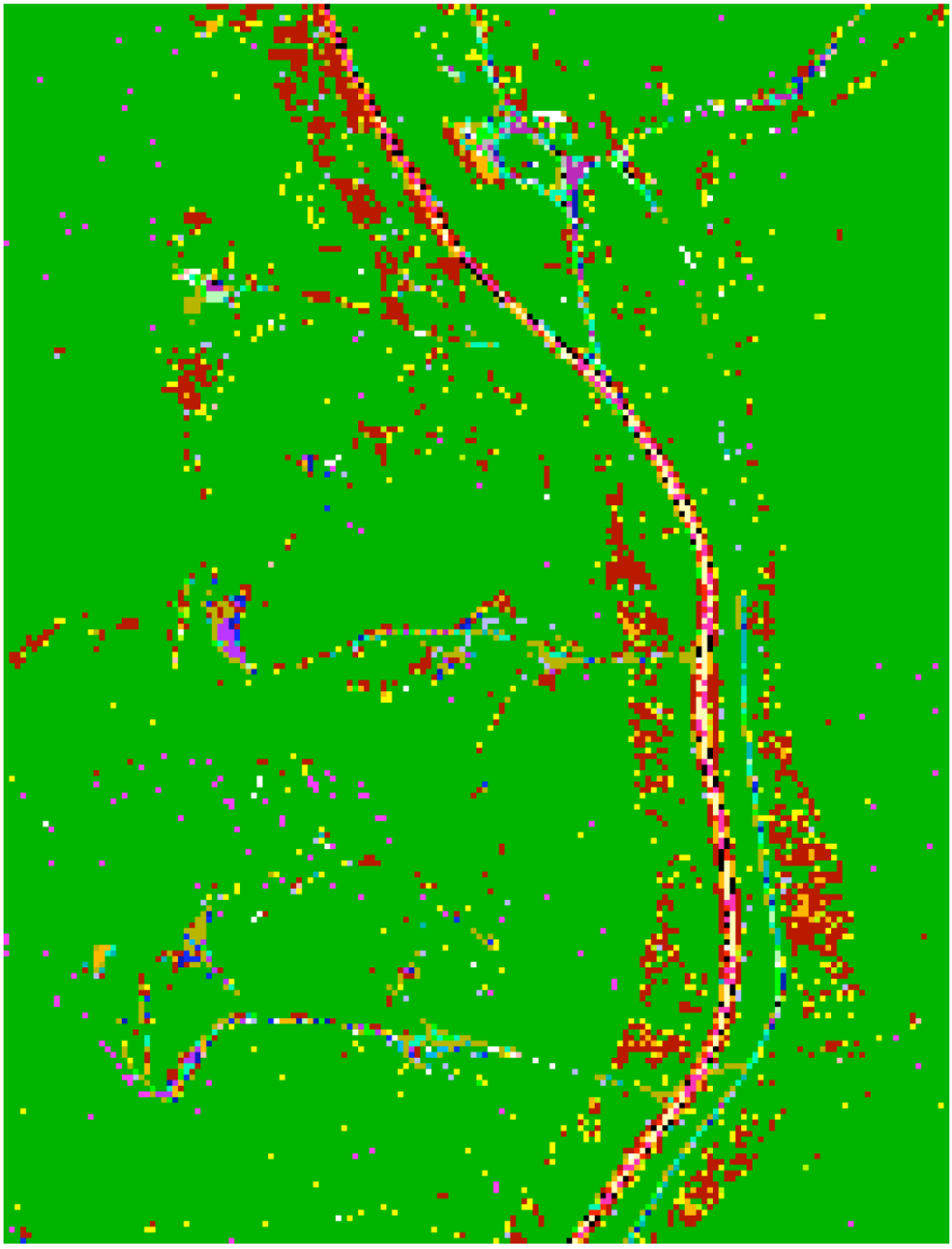


Figure 60. The Brown Cluster Map for the Zink Ranch Site 127.

Zink_127. All Clusters

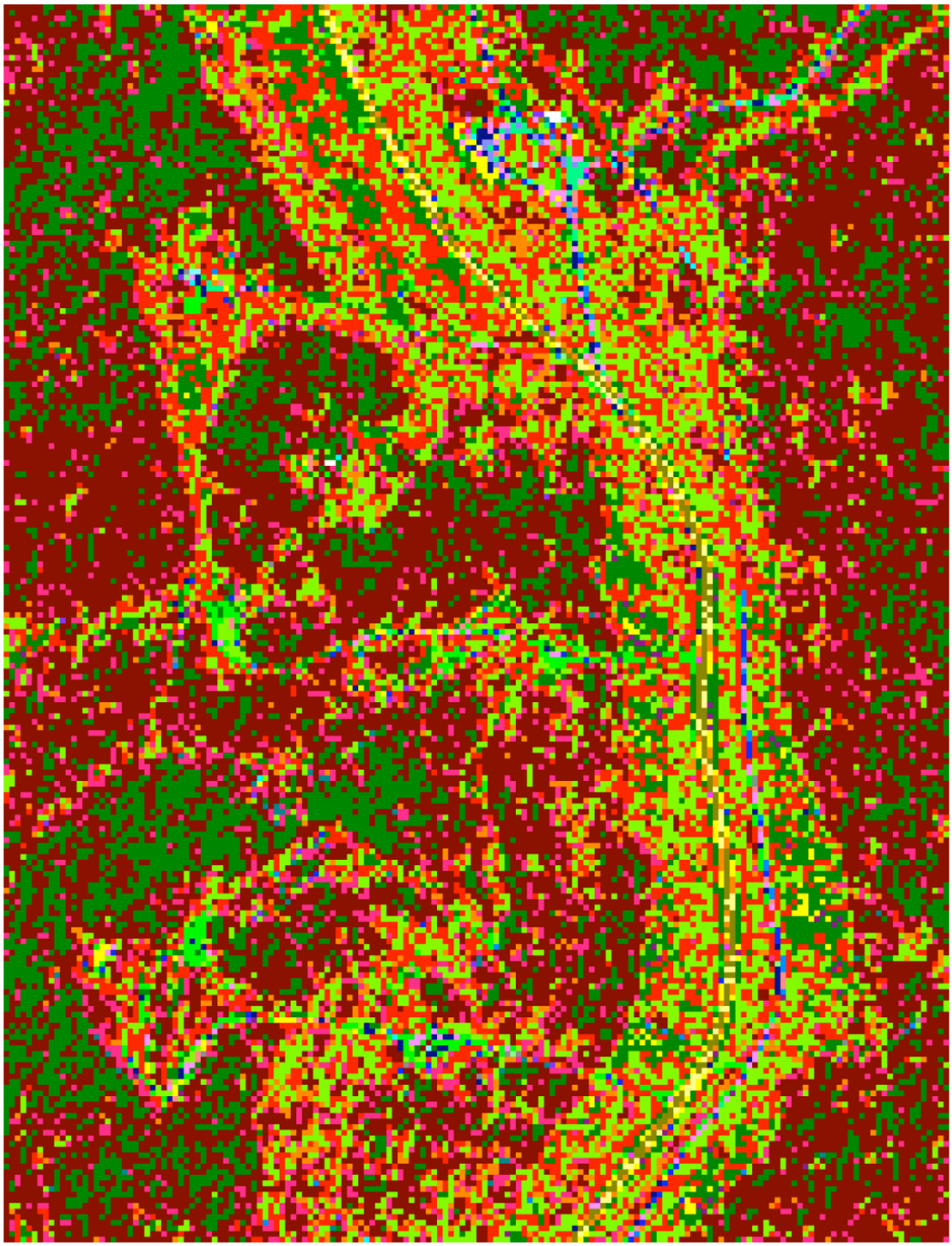


Figure 61. The 35 Cluster Map for the Zink Ranch Site 127.

Zink_127. All Clusters

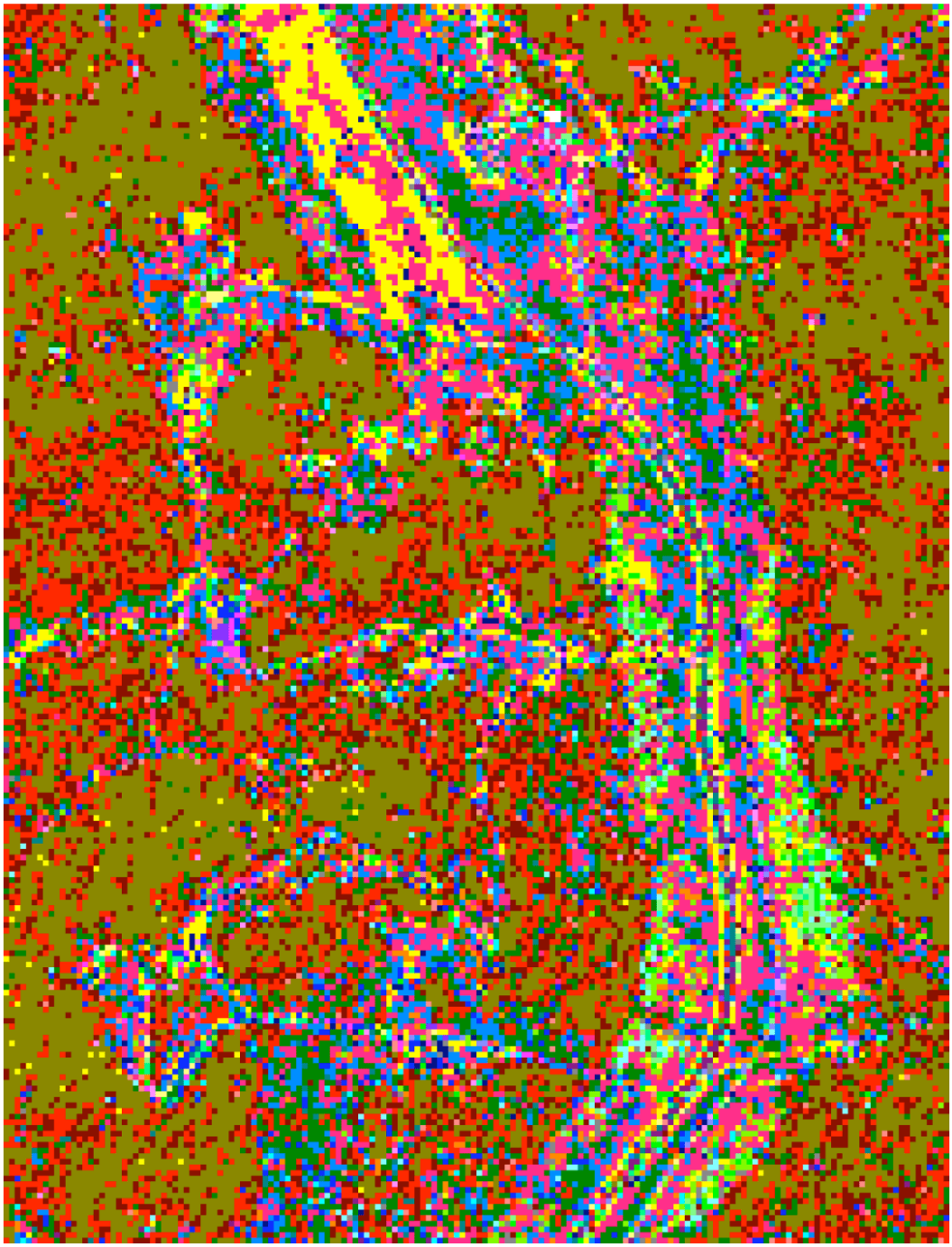


Figure 62. The 115 Cluster Map for the Zink Ranch Site 127.

4.9. Zink Ranch Site 131

The color image for the Zink Ranch Site 131 is displayed in Figure 63. The bright pixel image of the analysis area is shown in Figure 64. The big (more than 14 members) cluster map is displayed in Figure 65. The brown cluster map is shown in Figure 66. Enlargements of five regions of Figure 66 are displayed in Figures 67 to 71. The all cluster map is shown in Figure 72 for the 50 cluster case and in Figure 73 for the 174 cluster case.

Figure 63. The color map for the Zink Ranch Site 131.

Zink 131. RGB. Bright pixel.

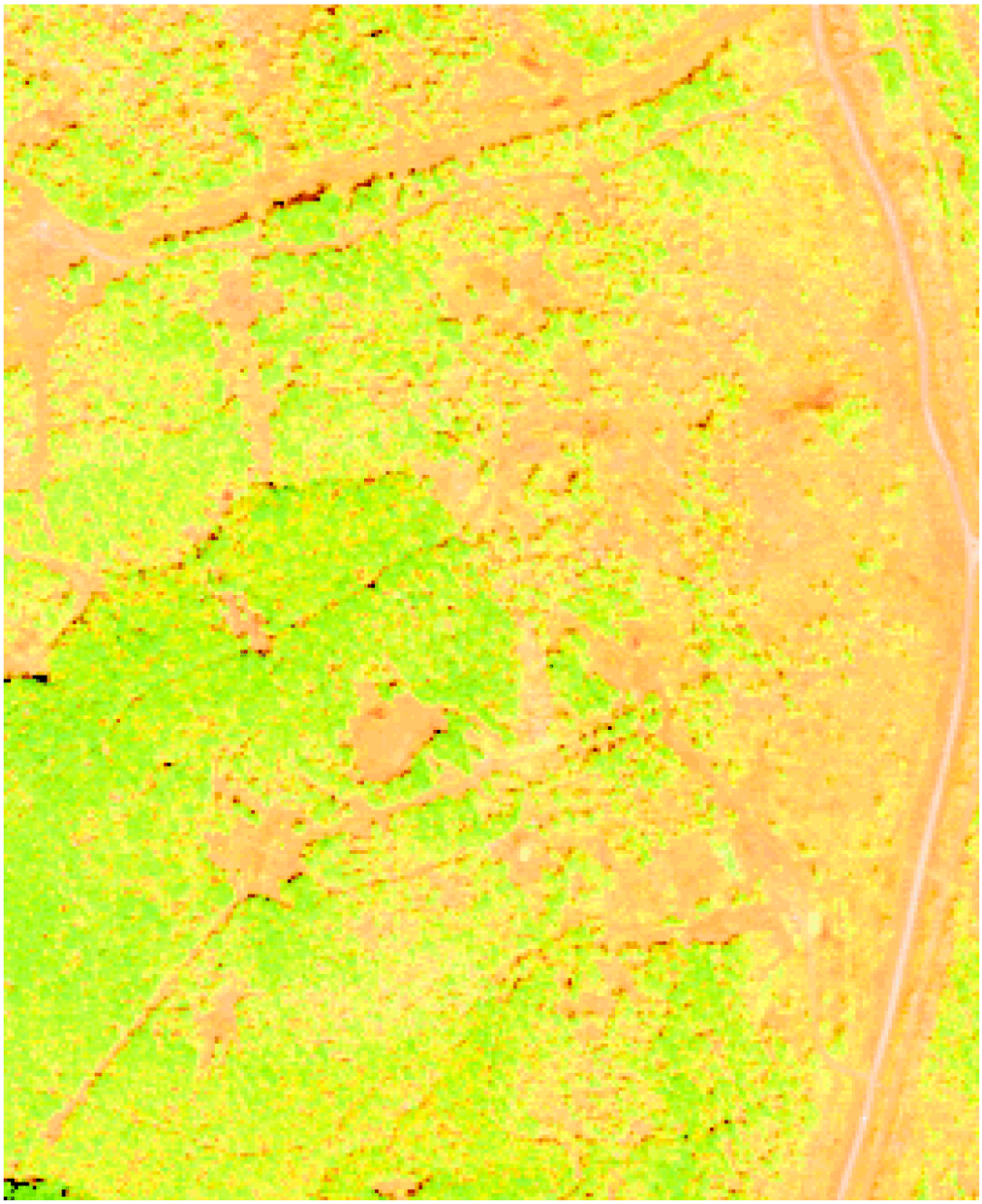


Figure 64. The Bright Pixel Image of the Zink Ranch Site 131.

Zink 131. Big Clusters (members > 18)

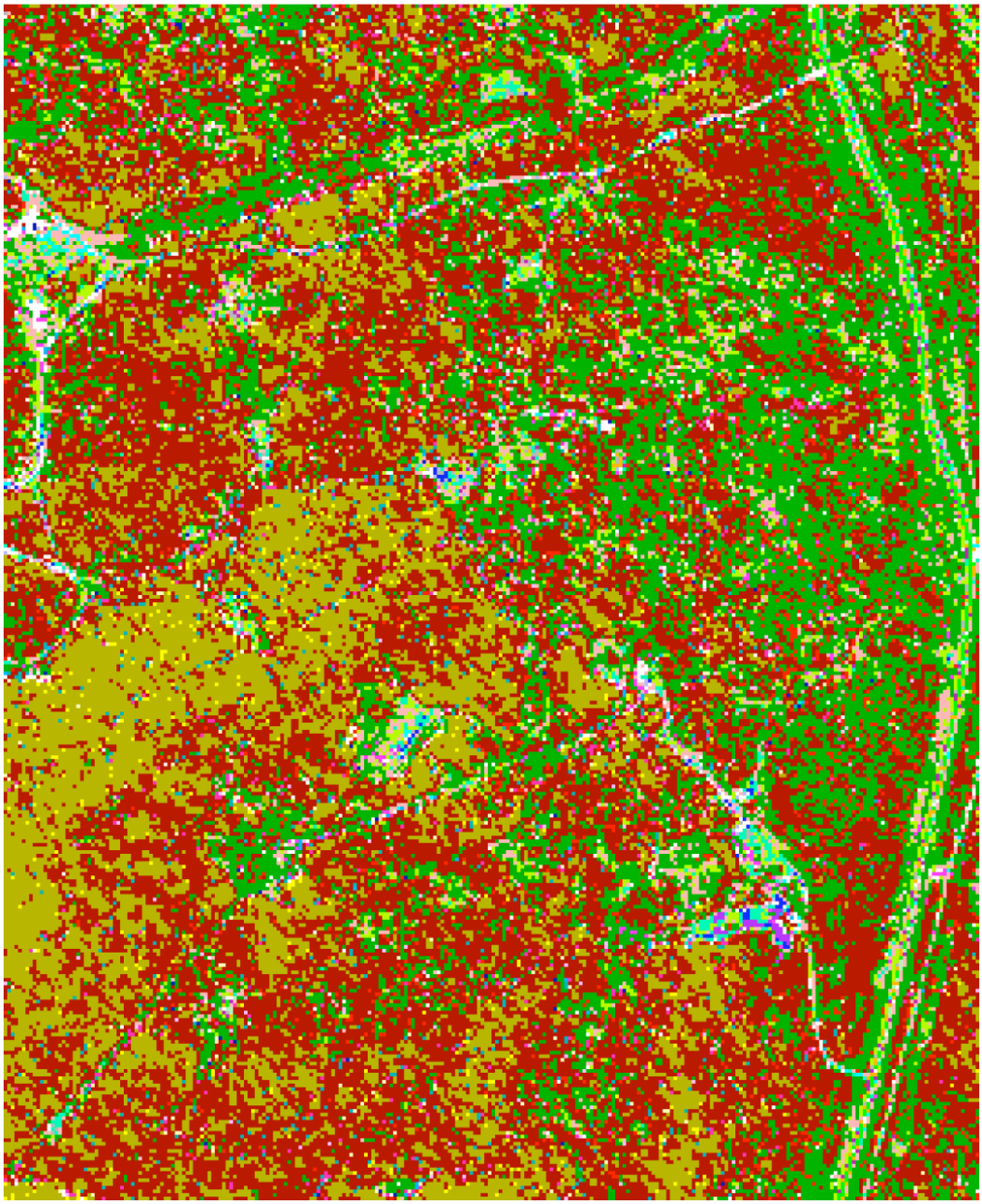


Figure 65. The Big Cluster map for the Zink Ranch Site 131.

Zink 131. All Brown Clusters (members > 2)

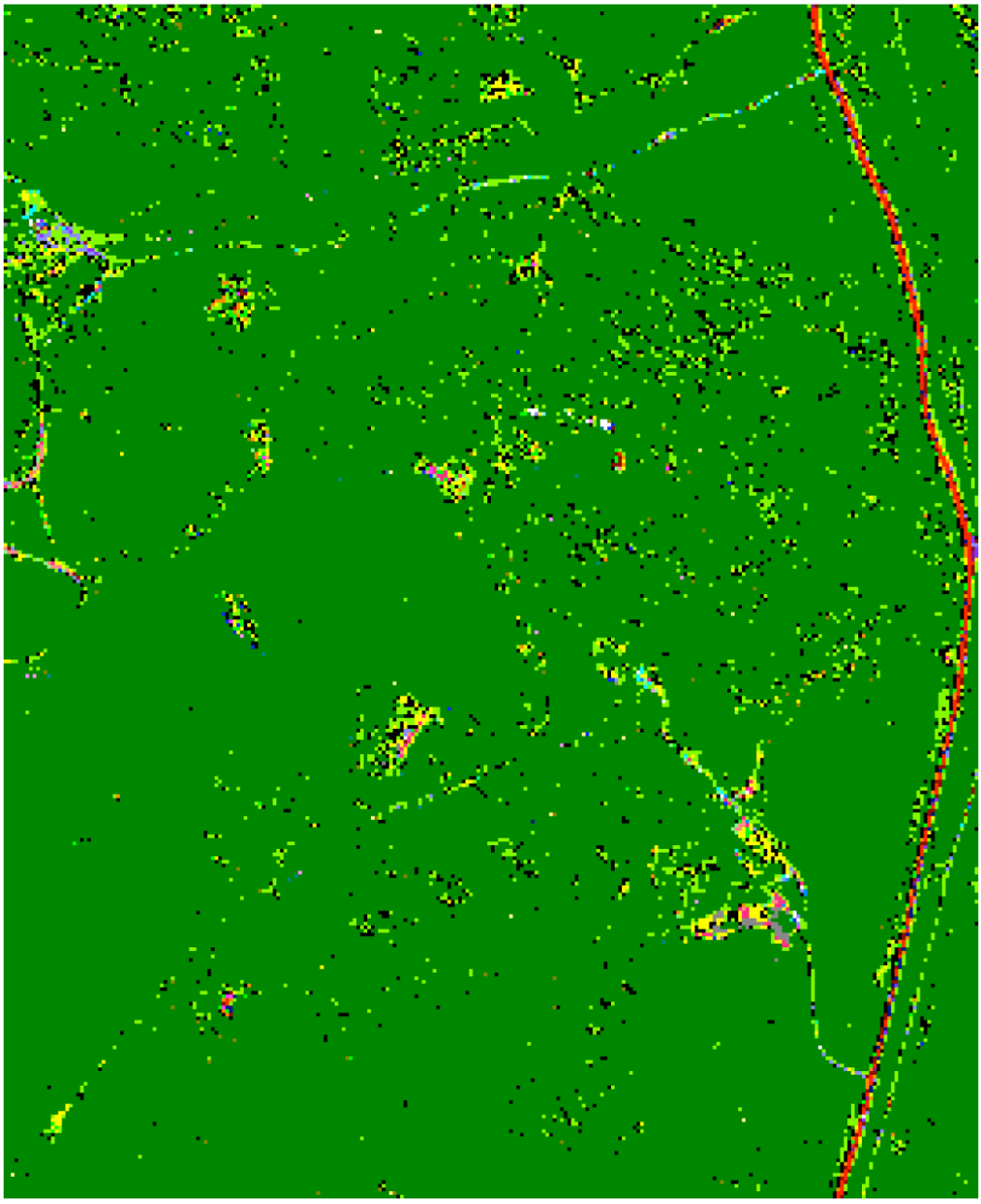


Figure 66. The Brown Cluster Map for the Zink Ranch Site 131.

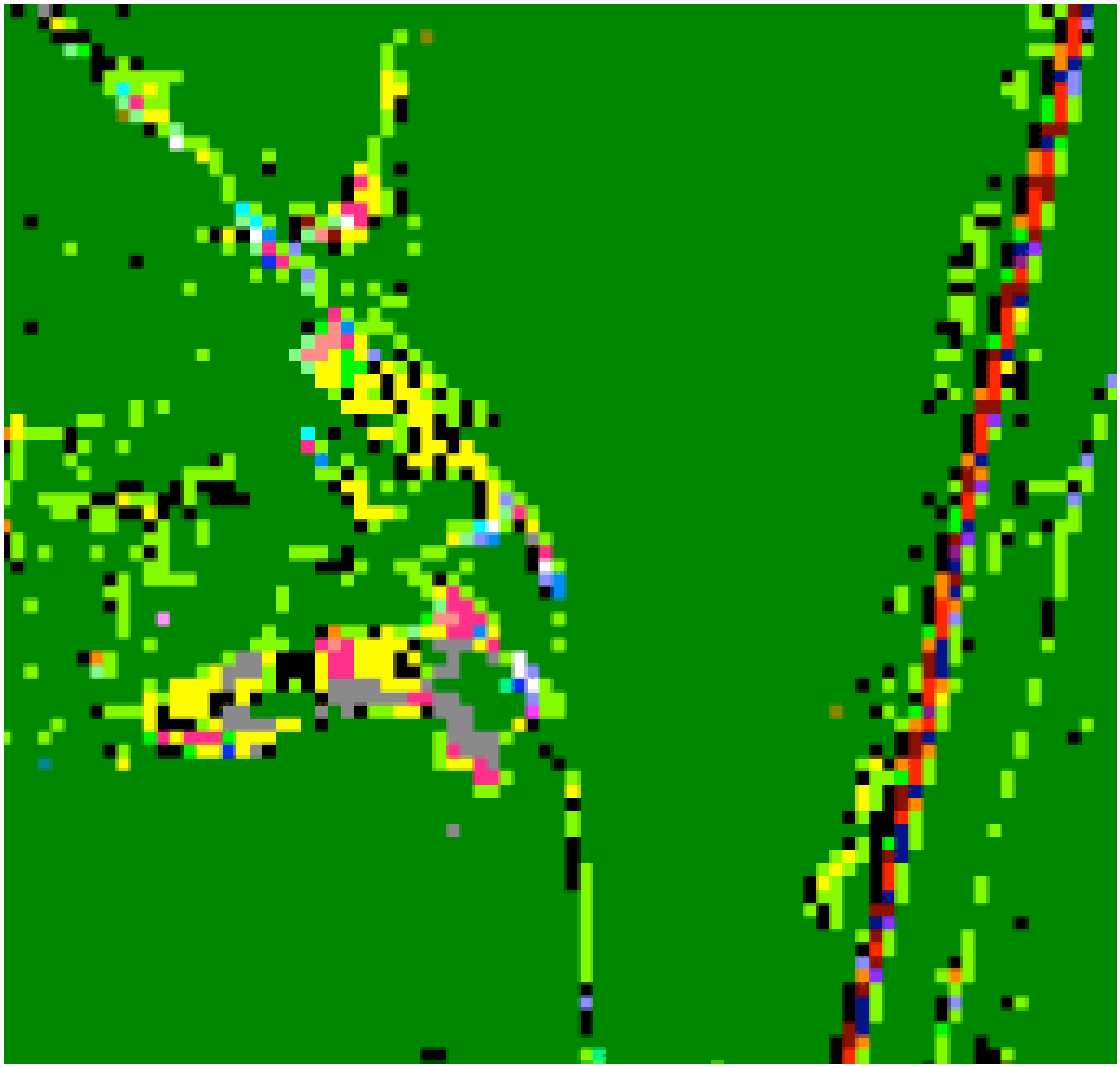


Figure 67. Enlargement of the lower right of Figure 66.

Figure 67 is in the lower right corner of Figure 66 and the left side is the Zink 131 oil production site and the right side is the road. Some of the pixels (green, black, and yellow) are associated with both oil and road. The brown, red, orange, and blue pixels are only associated with the road. The oil linked pixels are: magenta (C21), peach (C29), and grey (C23).

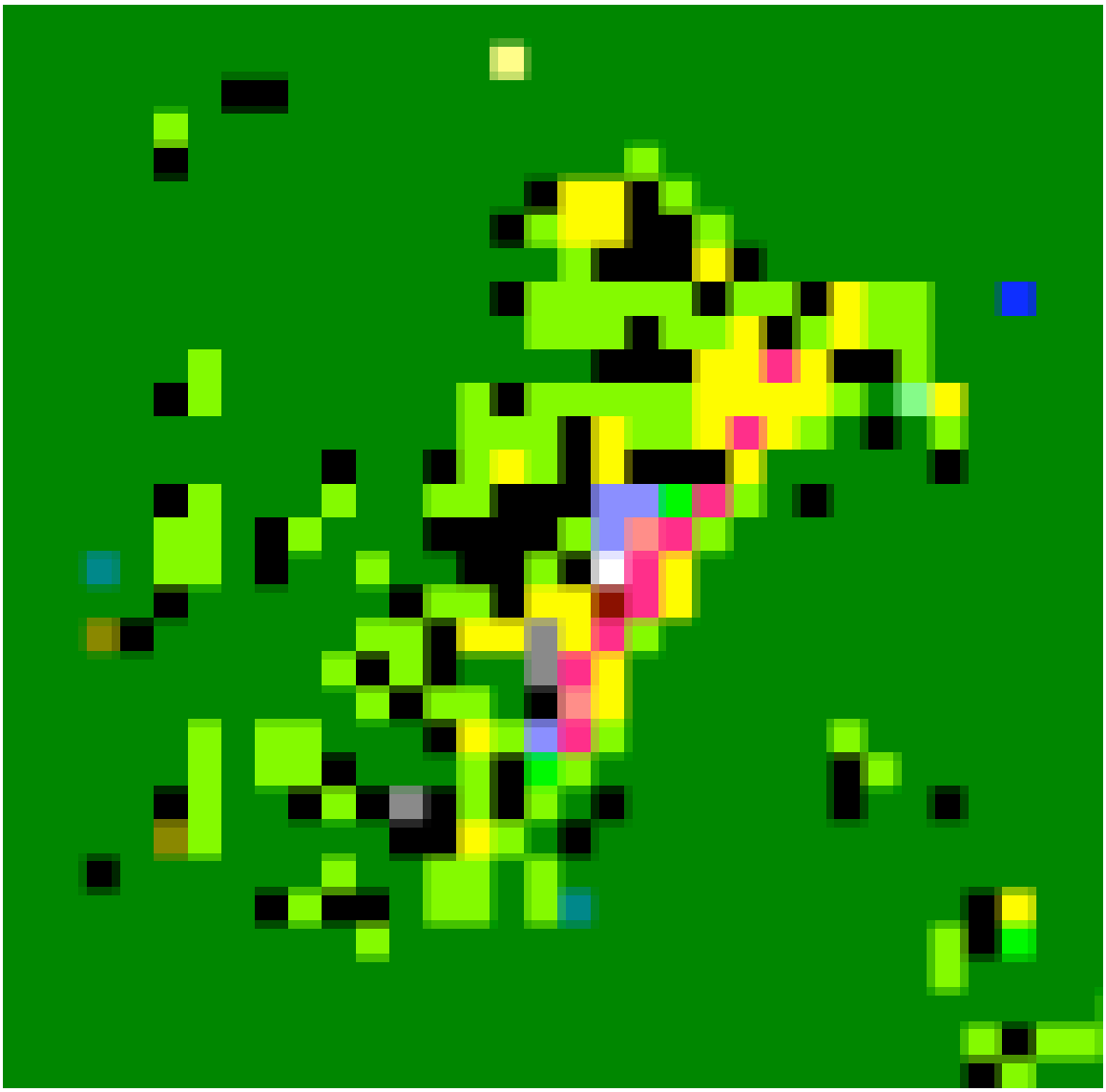


Figure 68. Enlargement of the lower middle of Figure 66.

Figure 68 is in the lower middle of Figure 66. The magenta, peach, and grey pixels are present.

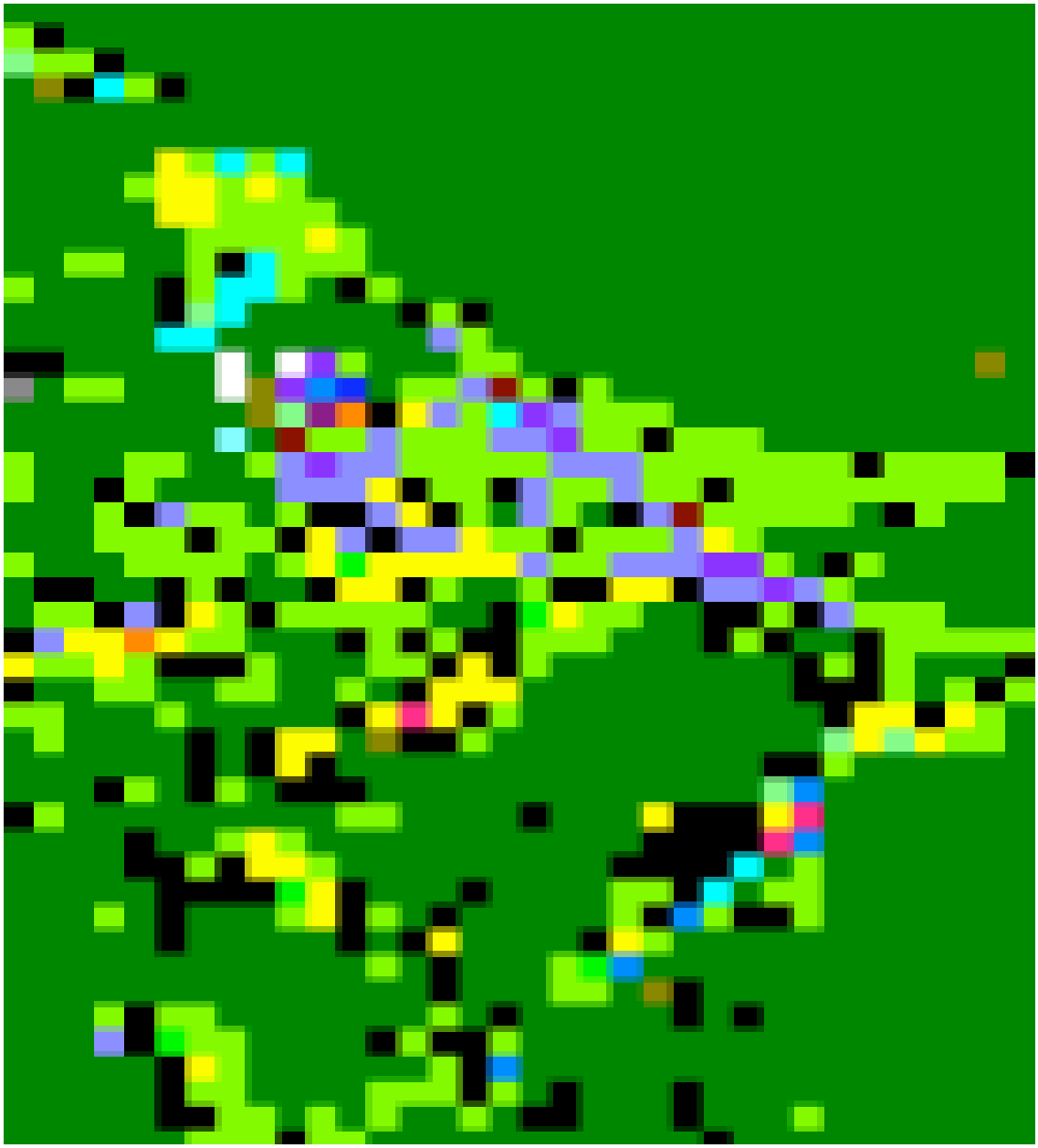


Figure 69. Enlargement of the upper left of Figure 66.

Figure 69 is in upper left of Figure 66 and is Zink site 130. There are three magenta and one grey pixels. There are many violet pixels in the figure, but we shall find that they are associated with roads.

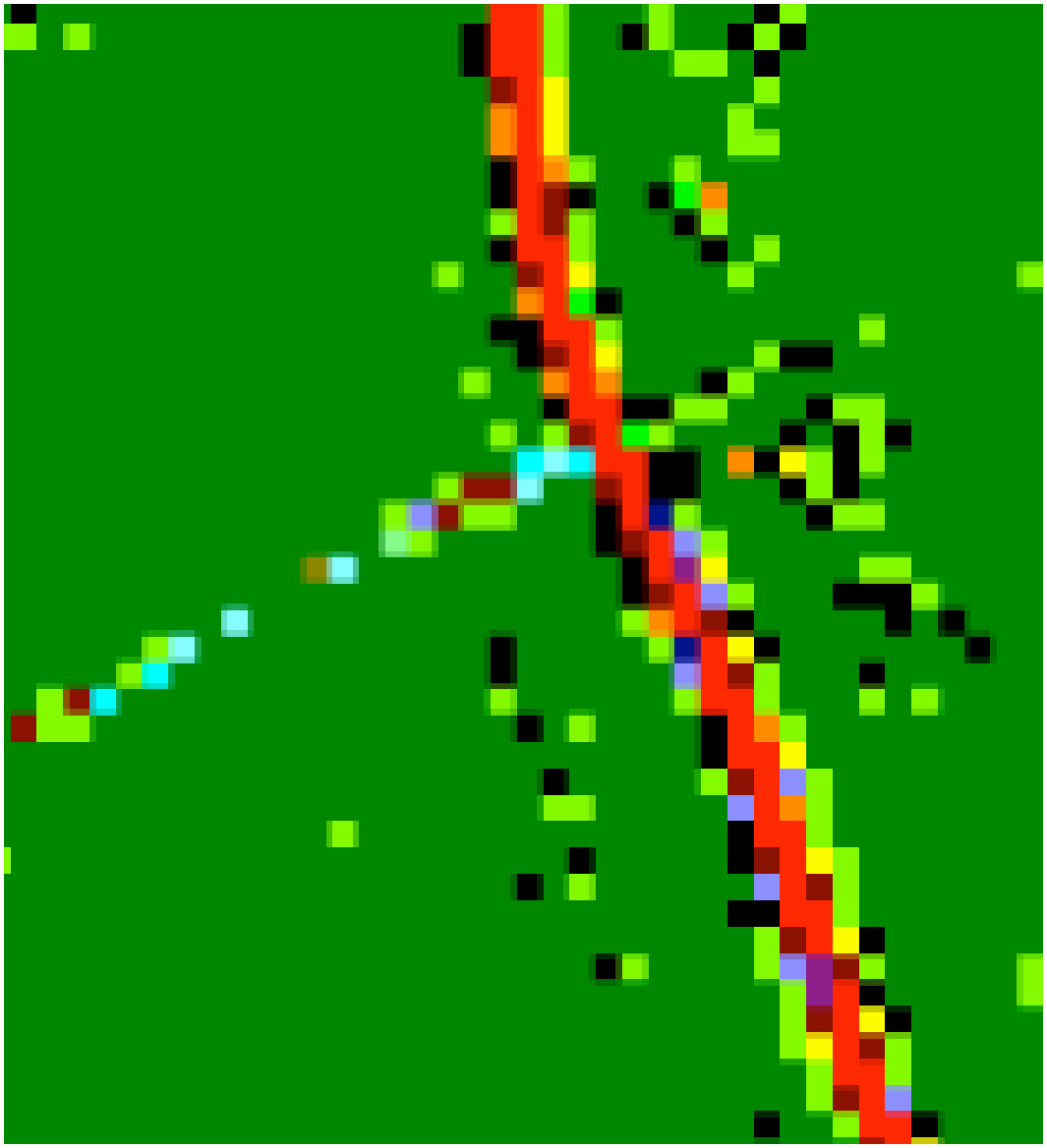


Figure 70. Enlargement of the upper right of Figure 66.

Figure 70 displays the intersection of a small road with the paved road in the upper right corner of Figure 66. The violet pixels are adjacent to the red pixels in the center of the road.

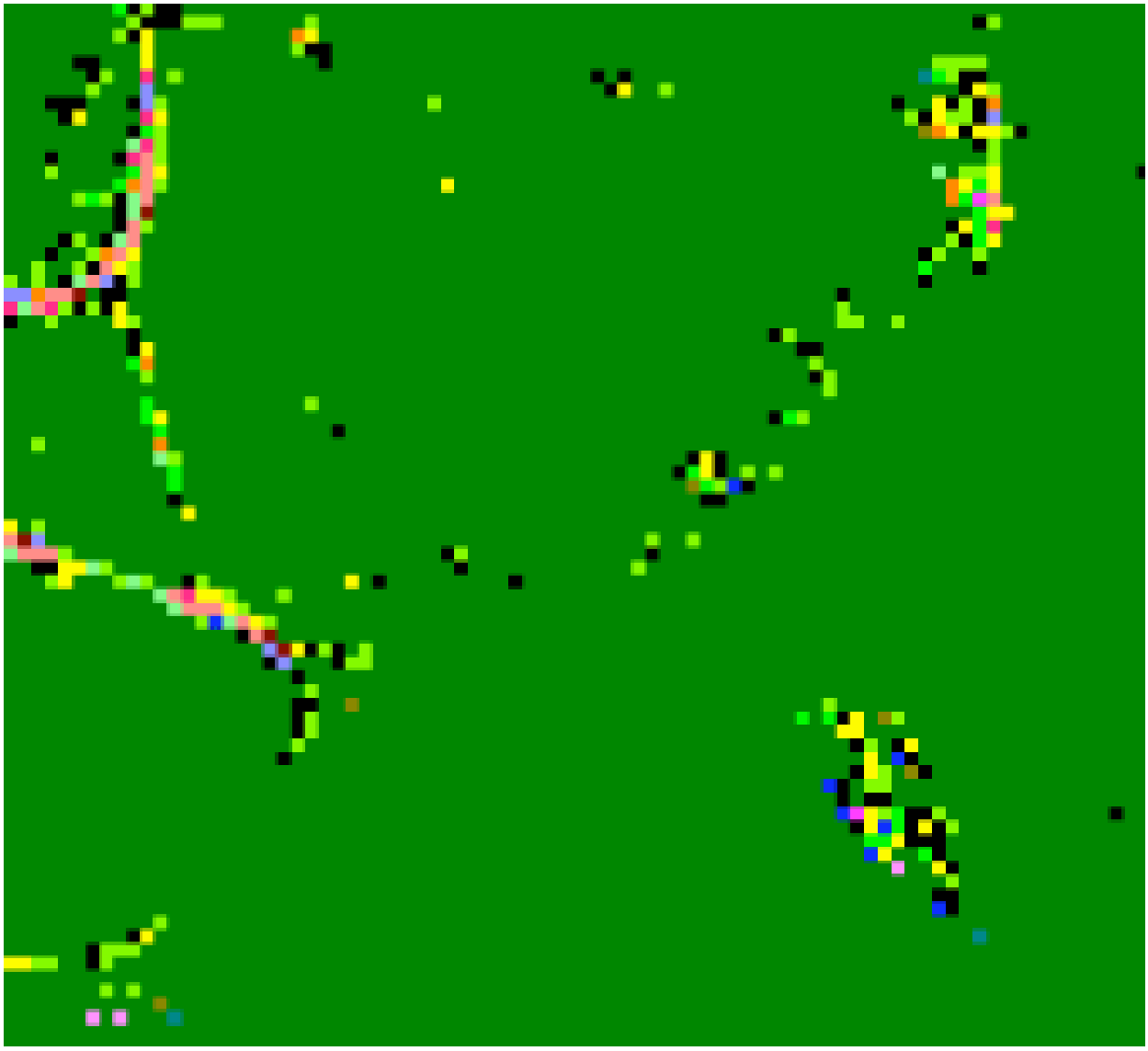


Figure 71. Enlargement of the center of the left edge of Figure 66.

Figure 71 is in the center of the left edge of Figure 66. The figure has 7 magenta pixels on the left and one on the upper right and two pink pixels (Cluster 40) on both the upper and lower left.

Zink 131. All Clusters

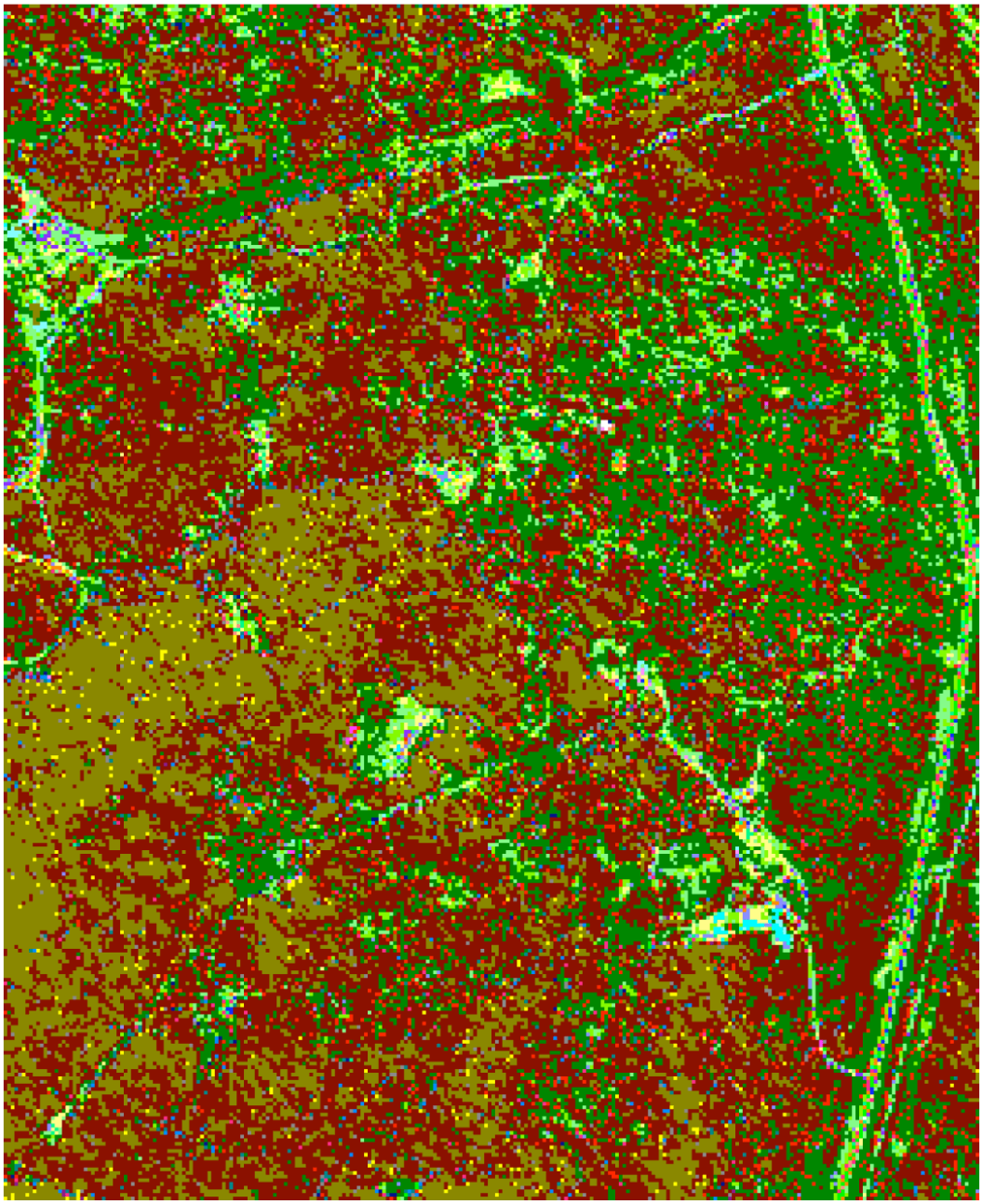


Figure 72. The 50 Cluster Map for the Zink Ranch Site 131.

Zink 131. All Clusters

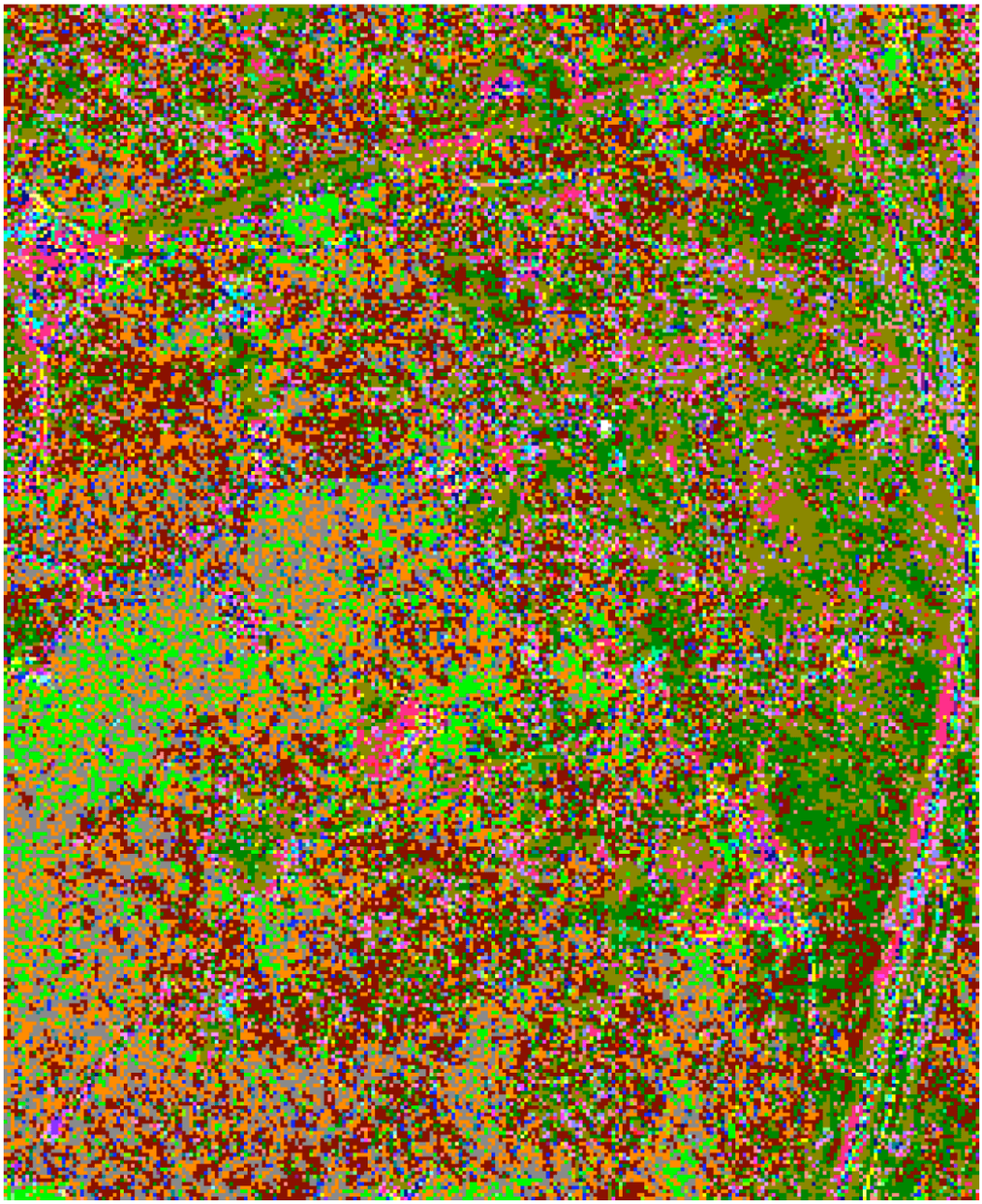


Figure 73. The 174 Cluster Map for the Zink Ranch Site 131.

4.10. Cluster Coordinates.

As mentioned at the start of this section, we have used SVD to create an uncorrelated orthonormal basis for the root vectors for each of the nine sites. Table 4 shows that between 4 and 8 basis vectors are required to capture 99% of the variance in the root vectors. In this subsection, we will use the first 4 basis vectors to plot the locations of the root vectors for four cases: Site A with 46 clusters, Site A with 100 clusters, Site B with 82 clusters, and Site B with 219 clusters. For each site will plot three slices through the 4D space: axis 1 vs axis2, axis 2 vs axis 3, and axis 2 vs axis 4. The results are displayed in Figures 74 to 85.

Since axis 1 is the mean value of the measurements and axis 2 is the direction of greatest variance, the plot of axis 1 vs axis 2 looks like a parabola. The root vectors with the highest value of a1 have a small value for a2, while the root vectors with either the highest or lowest values of a2 have the smallest values for a1. Since we used a r_b to eliminate the water pixels for both Site A and Site B, we are not surprised that there are no negative values for a1 for any of the four cases. For both sites, the range of values of a1 is higher in the case with a small number of clusters than in the case with a large number of clusters. A large number of clusters occurs when the cluster radius is smaller. The total number of cluster members is smaller with the larger number of clusters. Even though the members of each cluster will have less variance, we would expect a larger number of distinct root vectors to have more variance than a smaller number. This result is consistent with the results in Table 2, where the numerical rank is higher for the cases with the smaller number of root vectors.

In general, the plots of axis 2 vs axis 3, and axis 2 vs axis 4 have a significant number of points in all four quadrants. Thus, the root vectors are significantly different from each other are diverse. As we complete our analysis of Sites A and B, we may be able to place a label on each of the SVD basis vectors. For example, axis 1 could distinguish between green plants and disturbed soil.

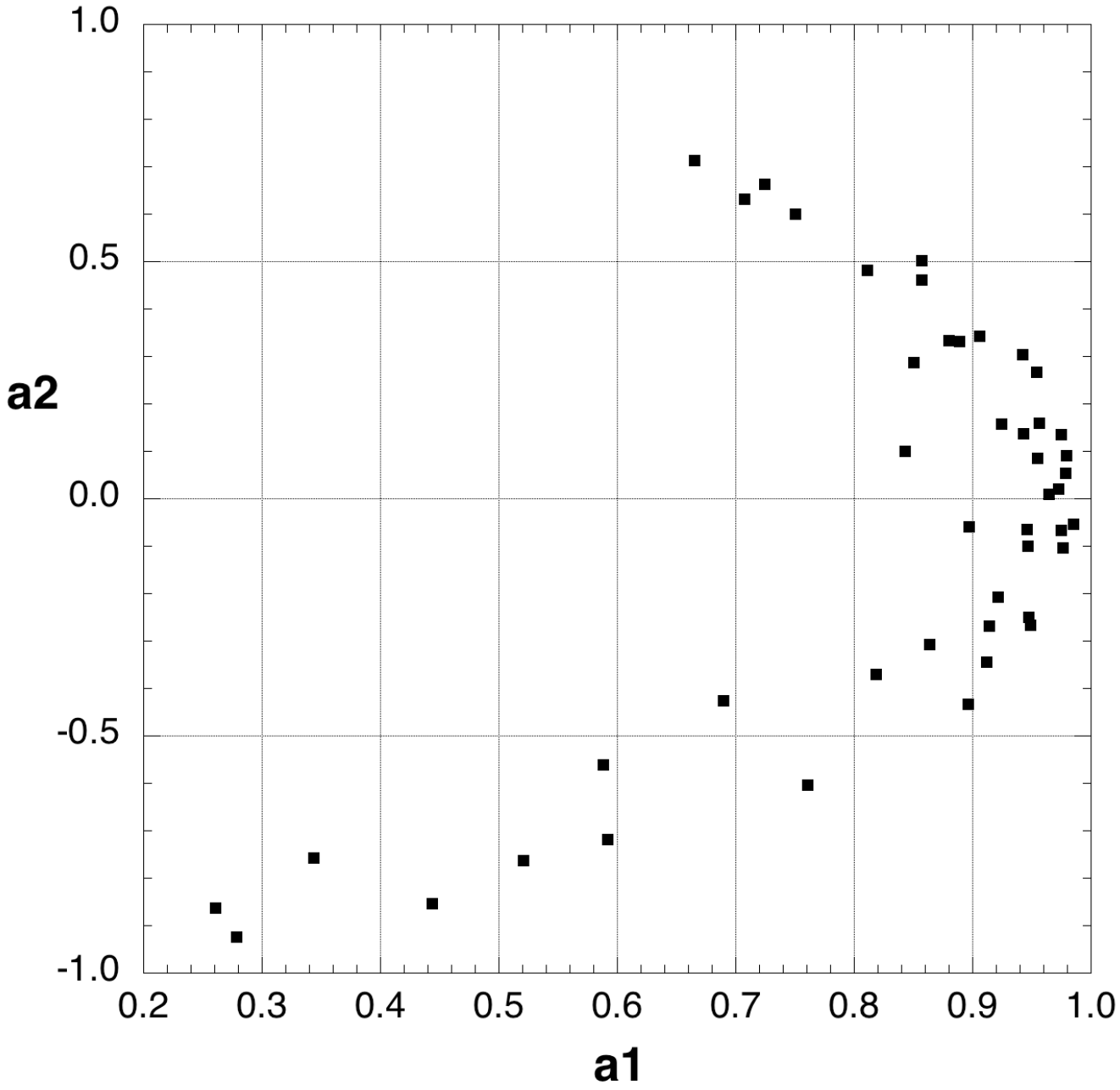


Figure 74. The location of the 46 root vectors for Site A on axis 1 vs axis 2.

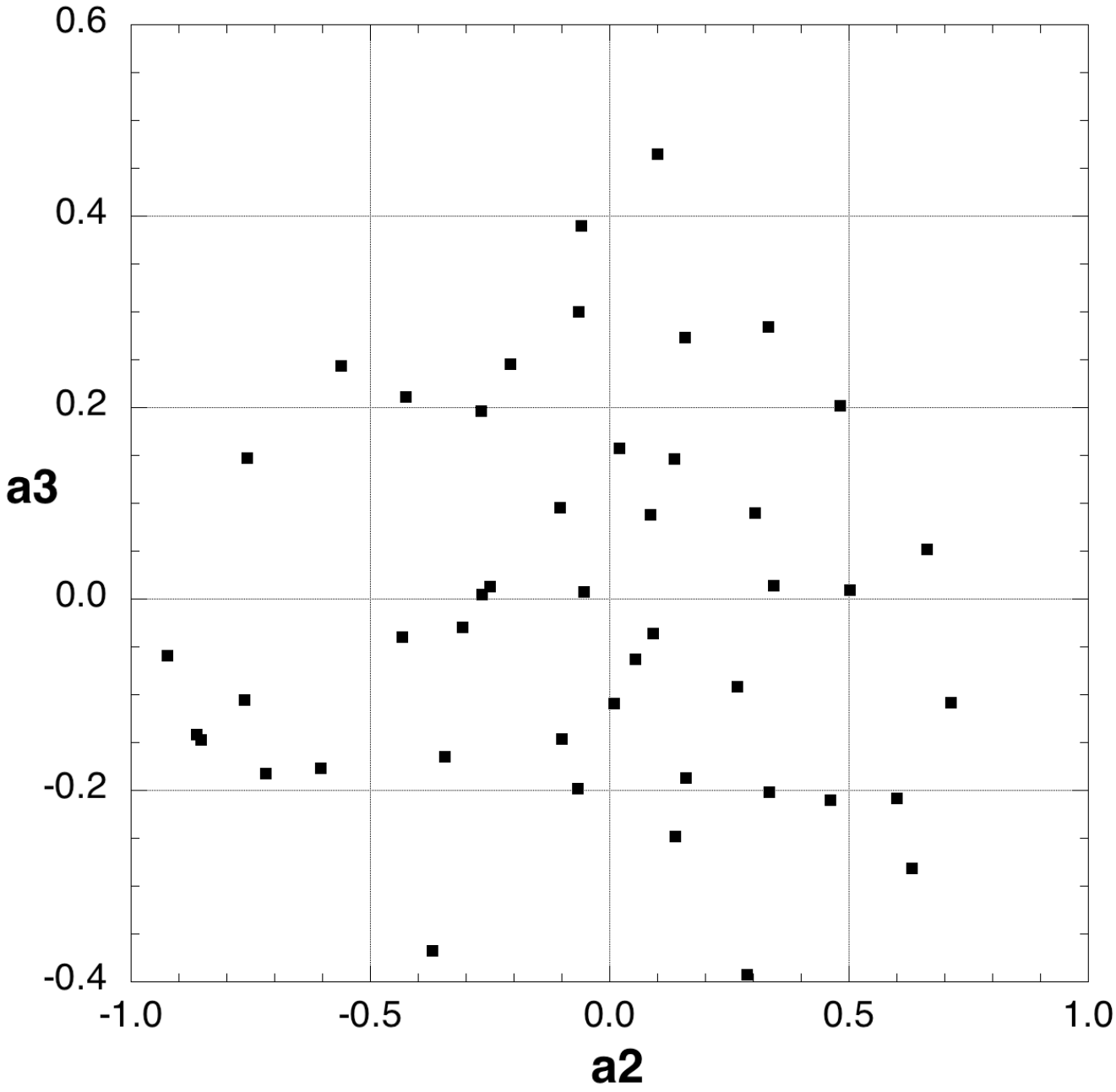


Figure 75. The location of the 46 root vectors for Site A on axis 2 vs axis 3.

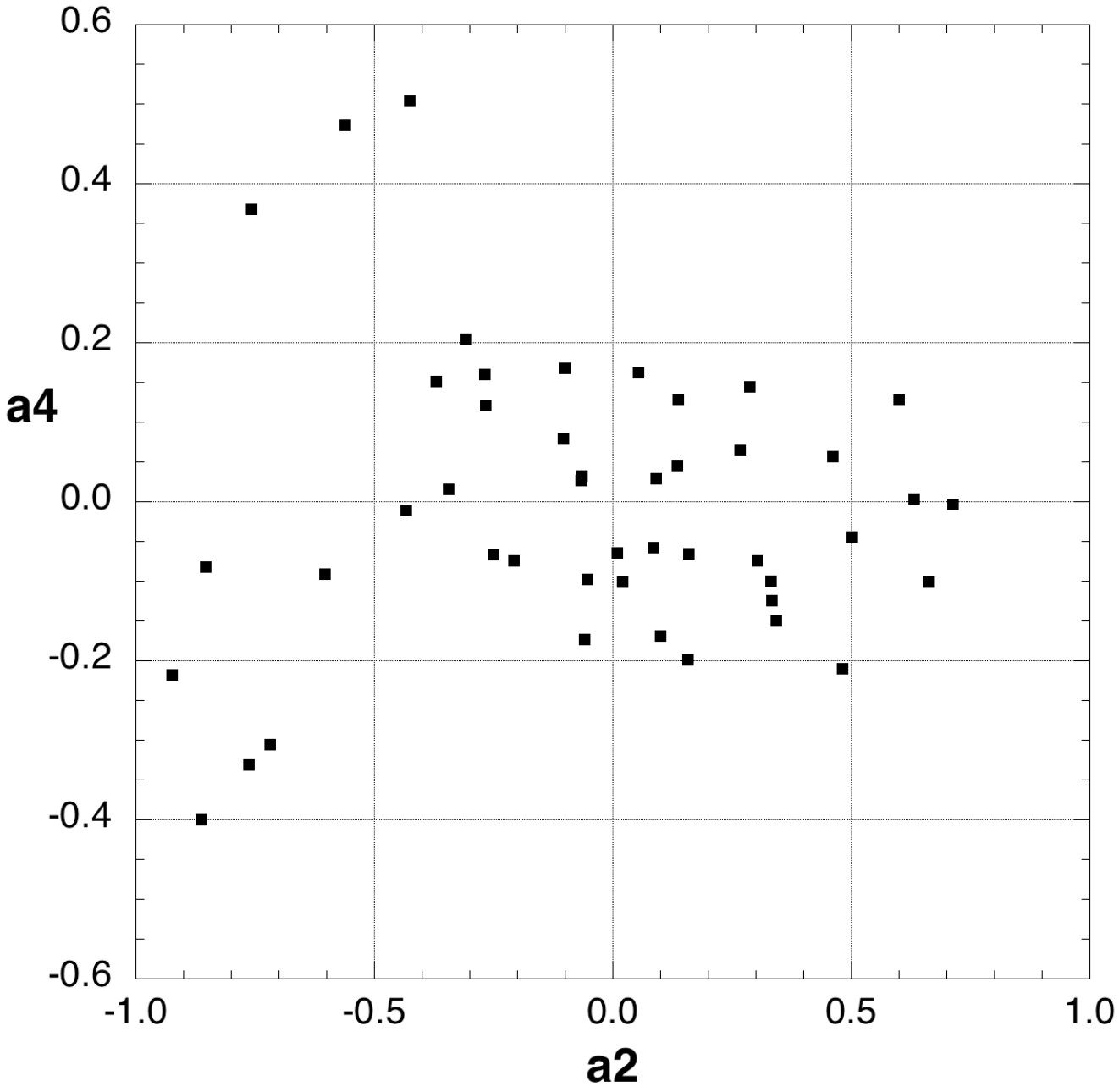


Figure 76. The location of the 46 root vectors for Site A on axis 2 vs axis 4.

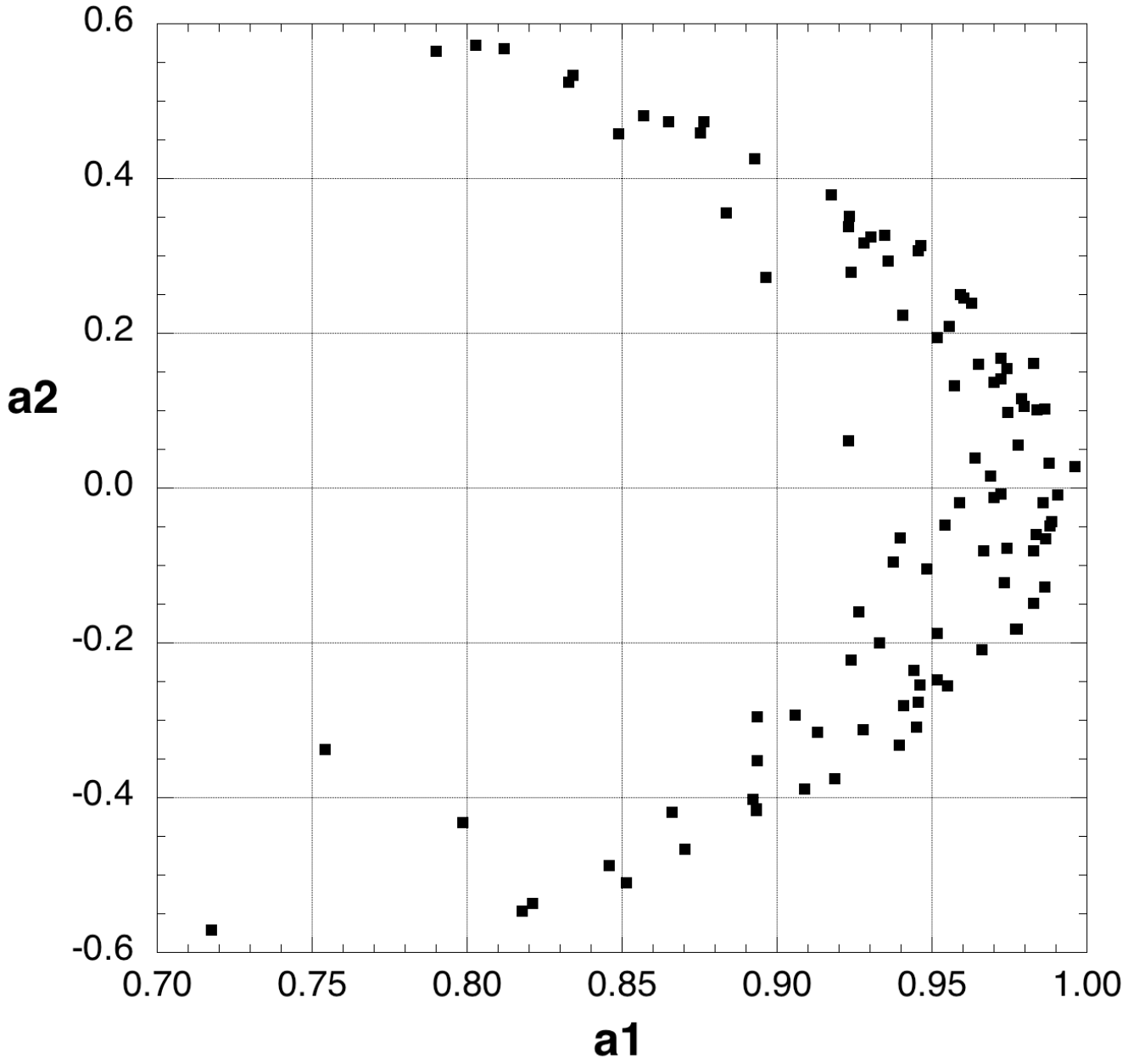


Figure 77. The location of the 100 root vectors for Site A on axis 1 vs axis 2.

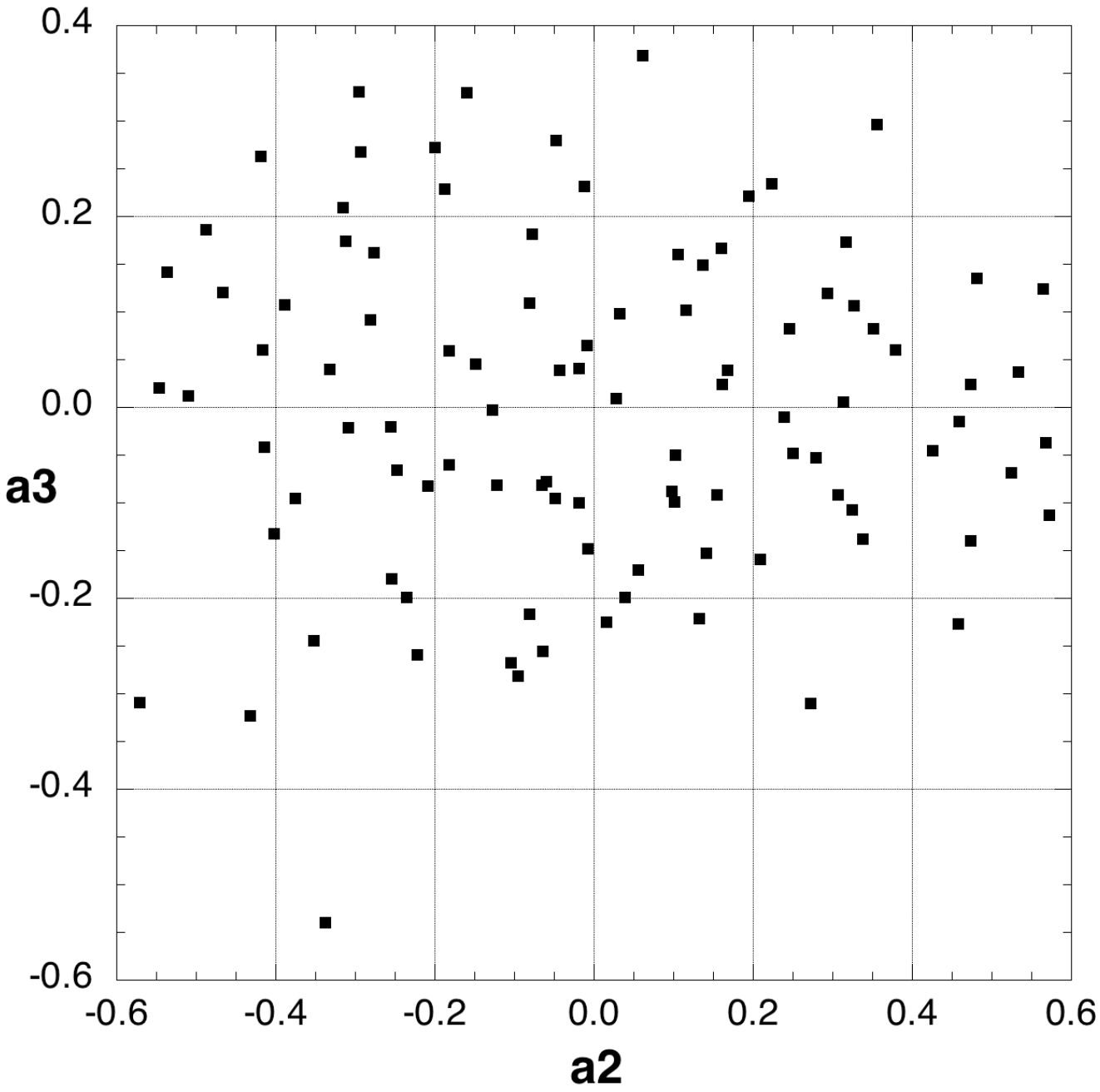


Figure 78. The location of the 100 root vectors for Site A on axis 2 vs axis 3.

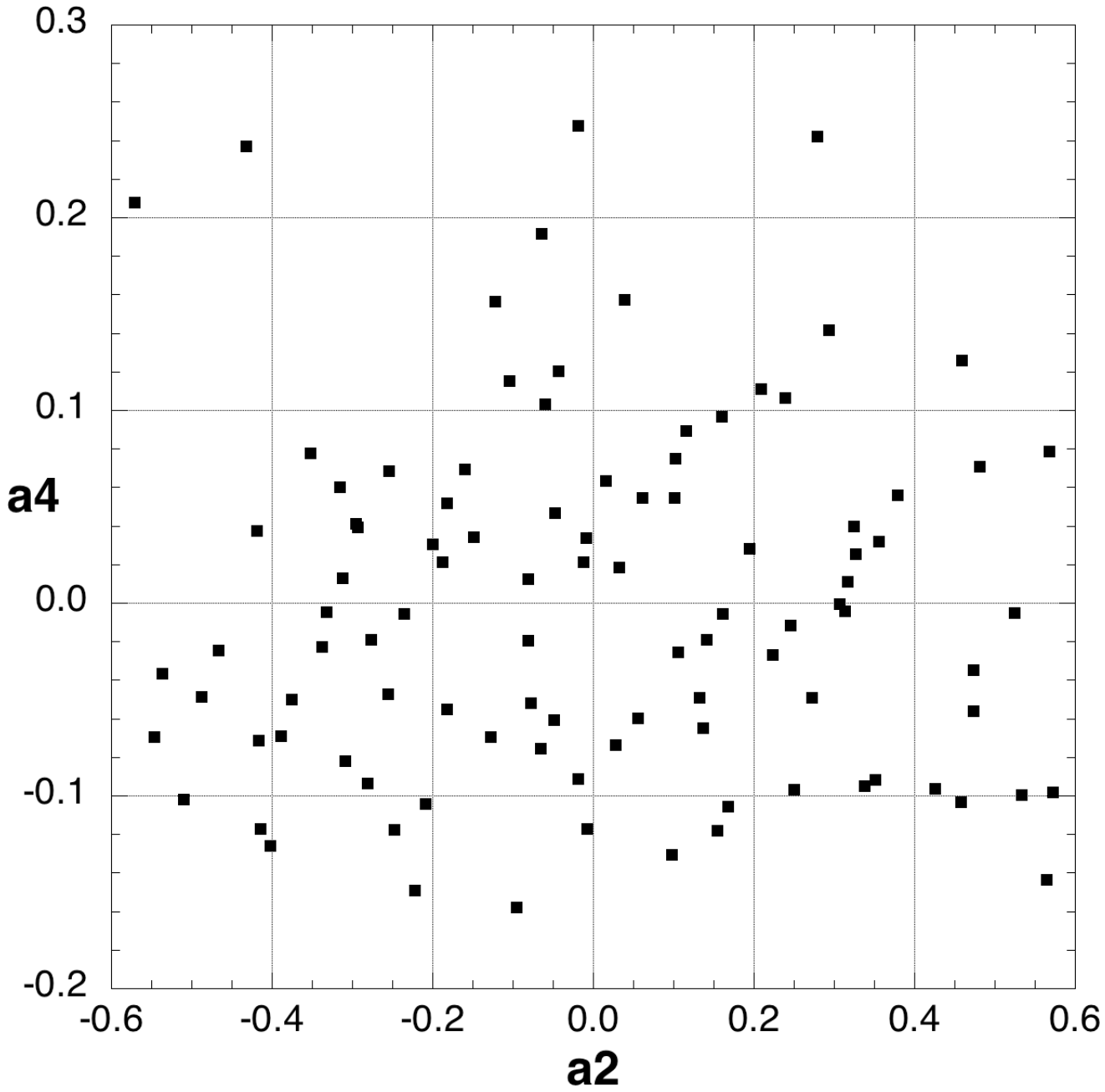


Figure 79. The location of the 100 root vectors for Site A on axis 2 vs axis 4.

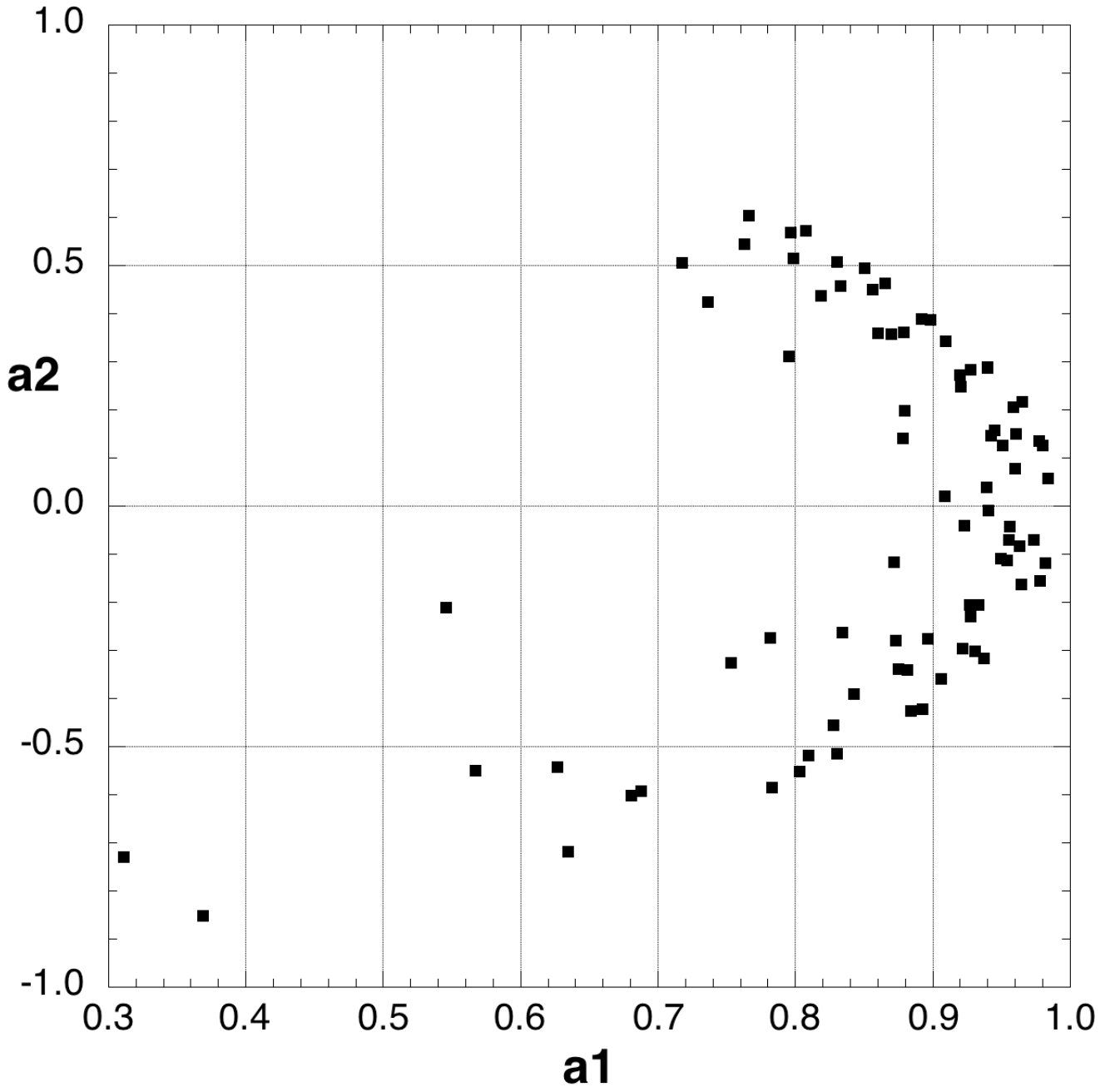


Figure 80. The location of the 82 root vectors for Site B on axis 1 vs axis 2.

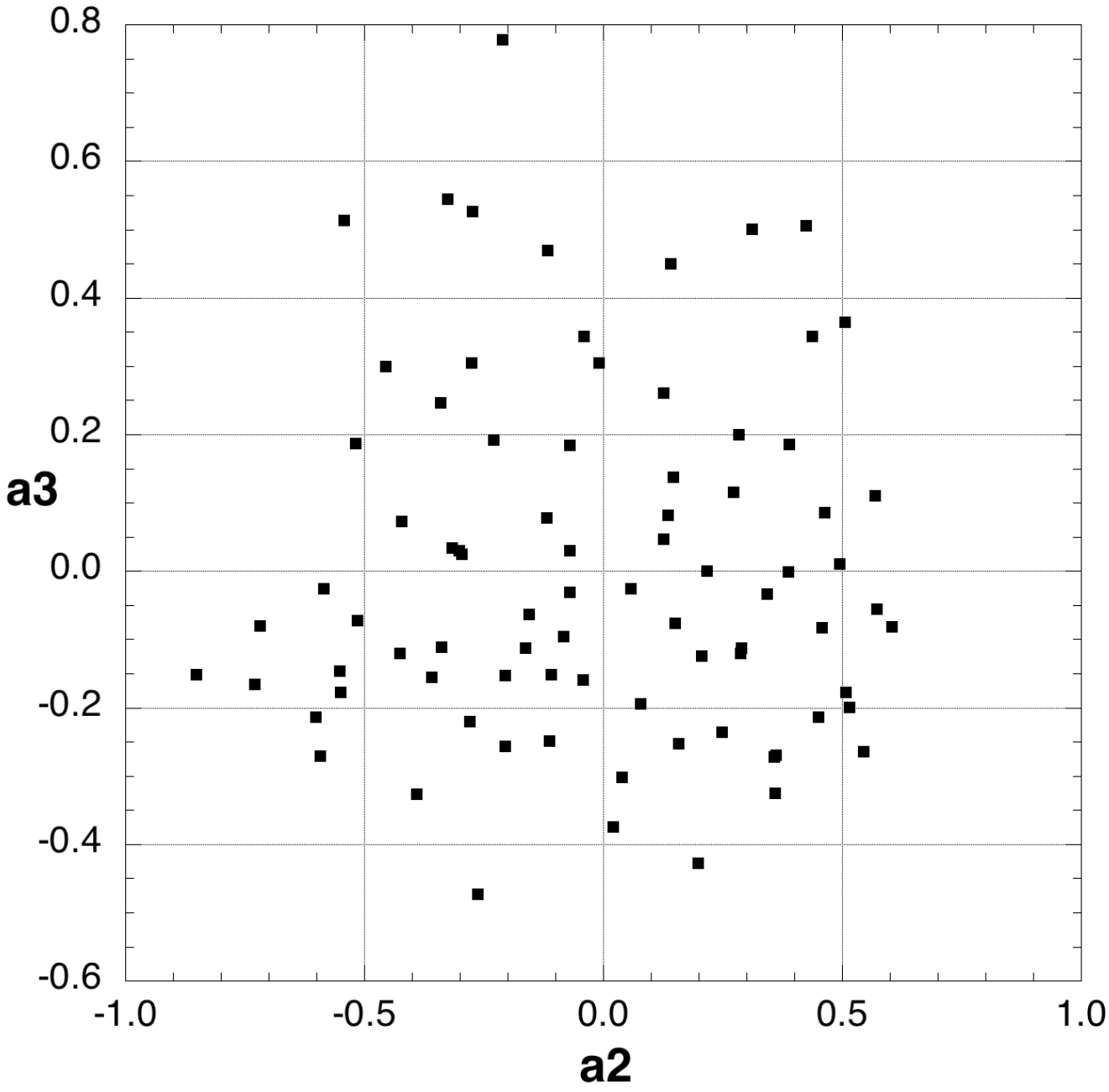


Figure 81. The location of the 82 root vectors for Site B on axis 2 vs axis 3.

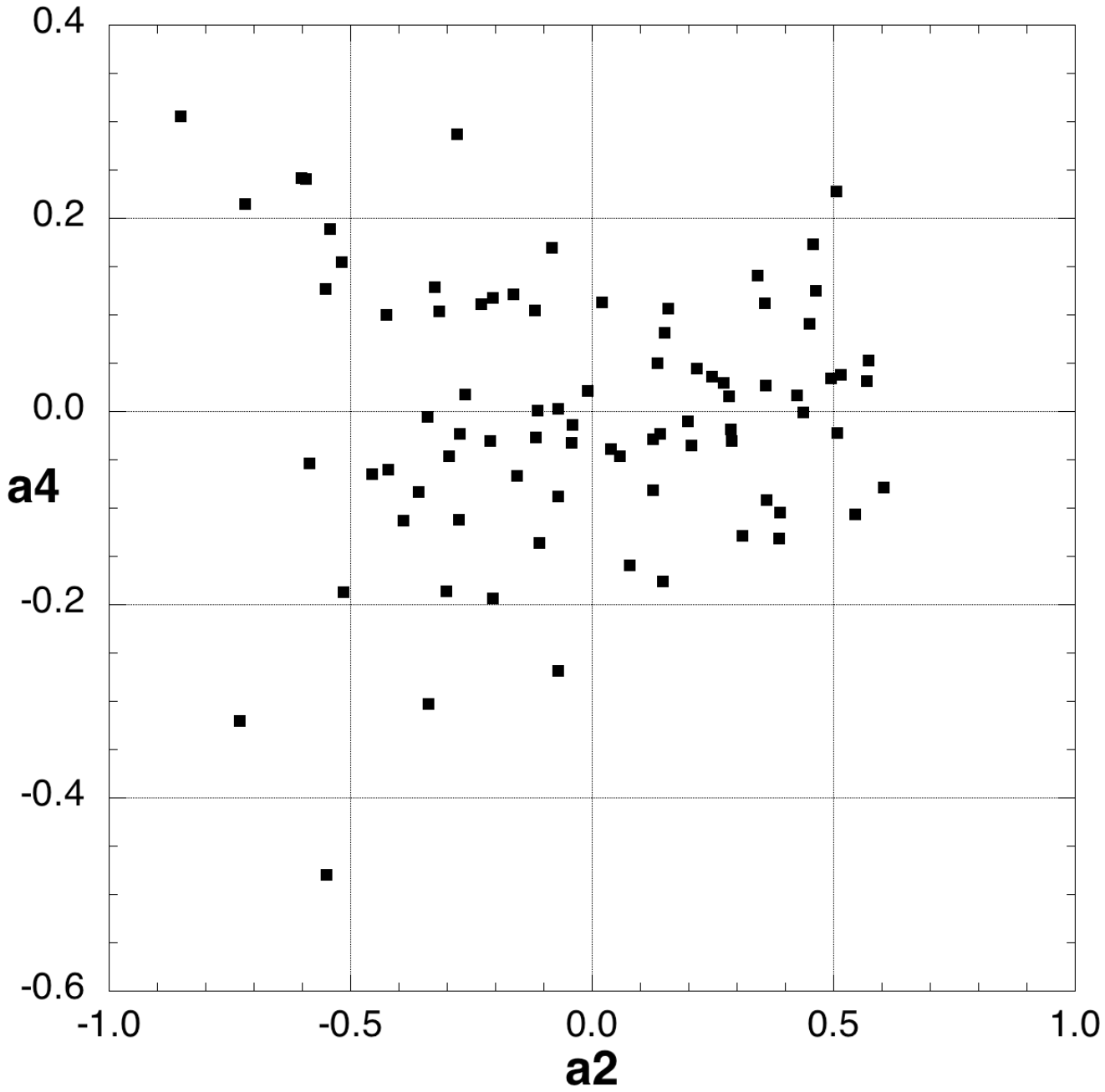


Figure 82. The location of the 82 root vectors for Site B on axis 2 vs axis 4.

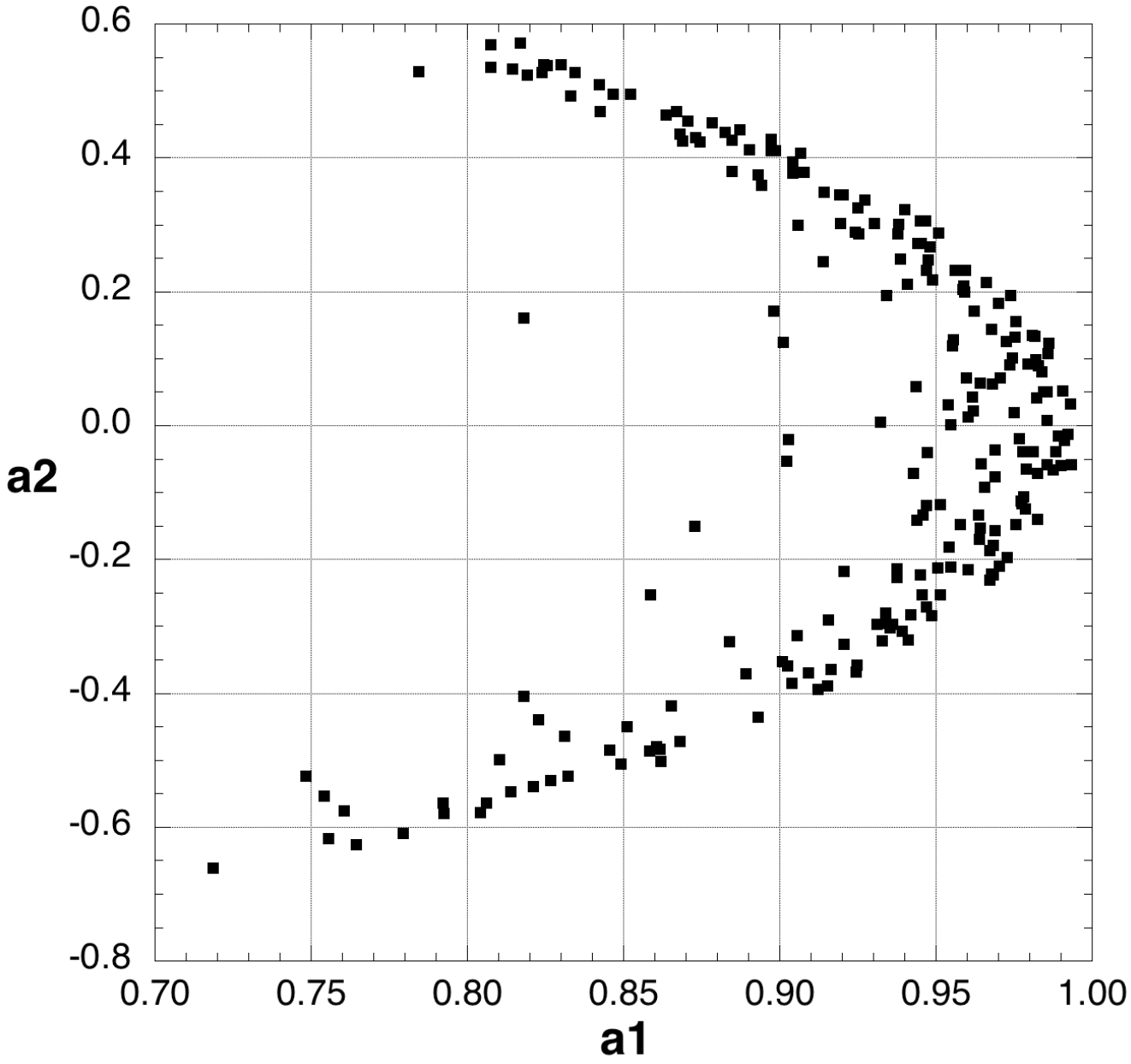


Figure 83. The location of the 219 root vectors for Site B on axis 1 vs axis 2.

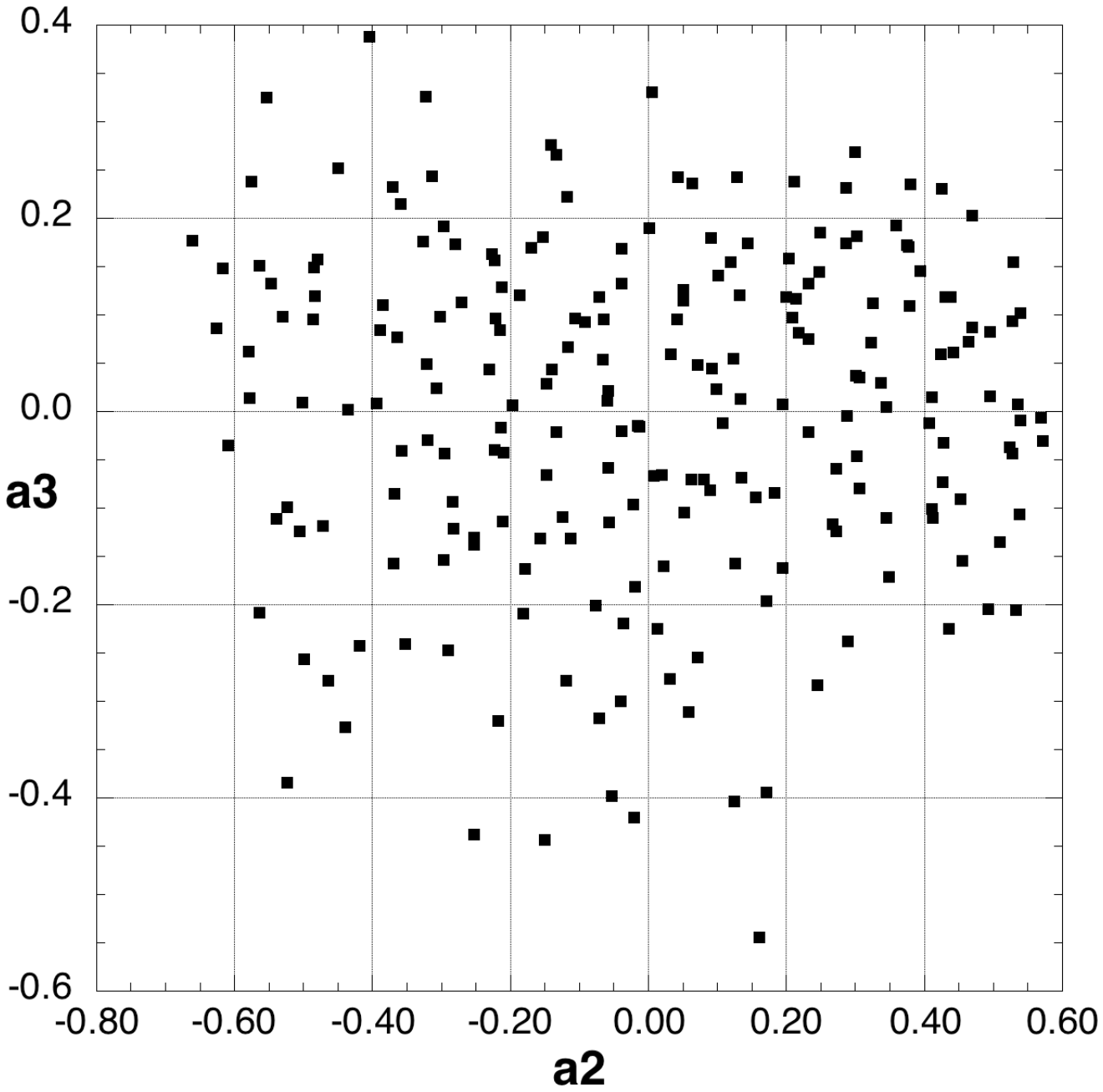


Figure 84. The location of the 219 root vectors for Site B on axis 2 vs axis 3.

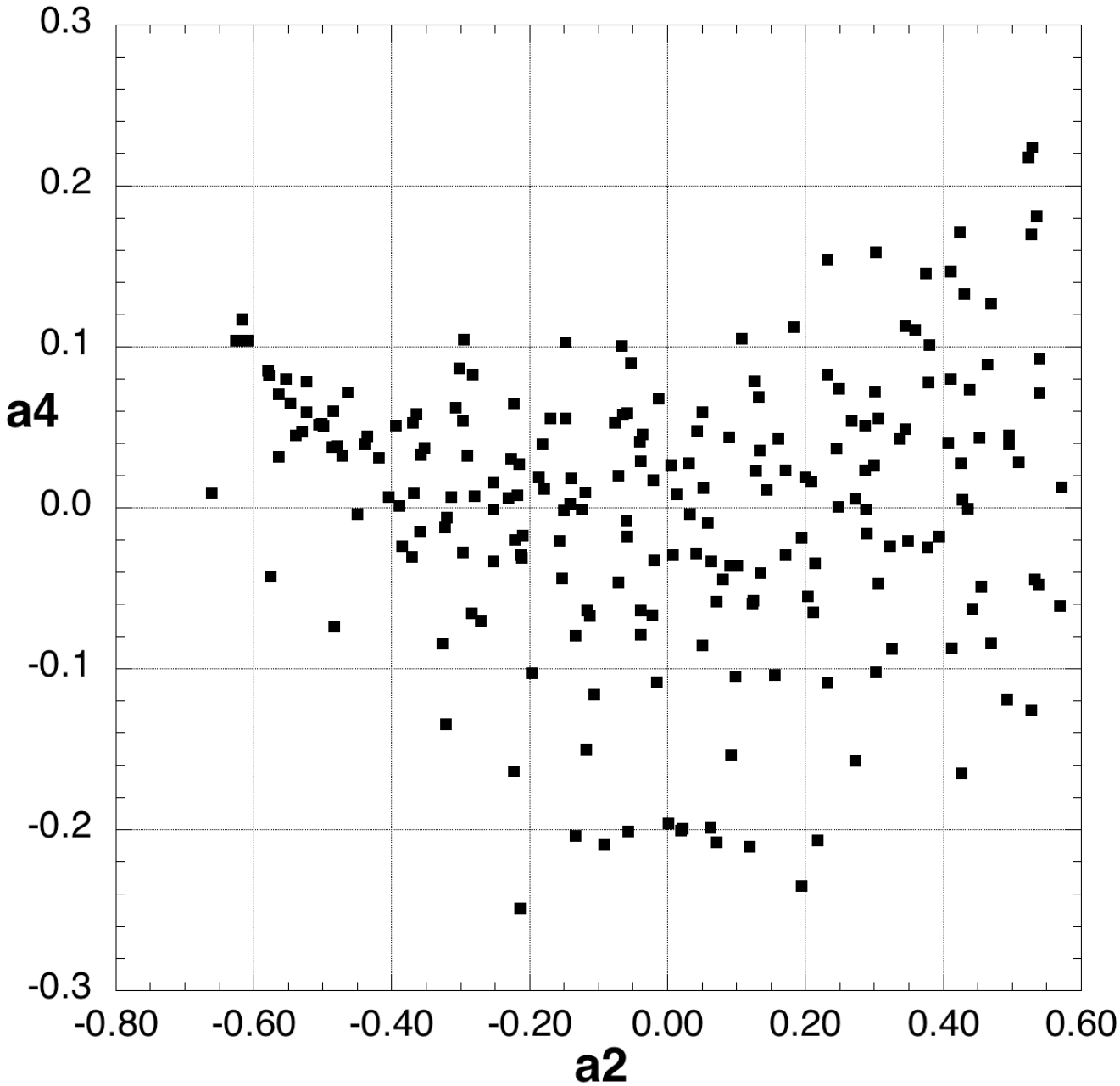


Figure 85. The location of the 219 root vectors for Site B on axis 2 vs axis 4.

5. ACKNOWLEDGMENT

This research was supported by the U. S. Department of Energy (DOE) Office of Fossil Energy (Office of Natural Gas and Petroleum Technology) through the National Energy Technology Laboratory and the DOE Office of Science (Office of Basic Energy Sciences). ORNL is managed by UT-Battelle, LLC for the U. S. Department of Energy under Contract No. DE-AC05-00OR22725.