Minnesota Department of Transportation

RESEARCH SERVICES

Office of Policy Analysis, Research & Innovation

Statistical Analysis of the Soil Chemical Survey Data

Vladimir Cherkassky, Principal Investigator
Department of Electrical and Computer Engineering
University of Minnesota

June 2010

Research Project
Final Report #2010-22

Your Destination...Our Priority

Technical Report Documentation Page

1. Report No.	2.	3. Recipients Accession No.
MN/RC 2010-22		

4. Title and Subtitle	5. Report Date
Statistical Analysis of the Soil Chemical Survey Data	June 2010
	6.
7. Author(s) Sauptik Dhar, Vladimir Cherkassky	8. Performing Organization Report No.
9. Performing Organization Name and Address University of Minnesota Department of Electrical and Computer Engineering 200 Union St. SE Minneapolis, MN 55455	10. Project/Task/Work Unit No. 11. Contract (C) or Grant (G) No. (c) 89261 (wo) 104
12. Sponsoring Organization Name and Address Minnesota Department of Transportation Research Services Section 395 John Ireland Boulevard, MS330 St. Paul, MN 55155-1899	13. Type of Report and Period Covered Final Report 14. Sponsoring Agency Code

15. Supplementary Notes

http://www.lrrb.org/pdf/201022.pdf

16. Abstract (Limit: 250 words)

This report describes data-analytic modeling of the Minnesota soil chemical data produced by the 2001 metro soil survey and by the 2003 state-wide survey. The chemical composition of the soil is characterized by the concentration of many metal and non-metal constituents, resulting in high-dimensional data. This high dimensionality and possible unknown (nonlinear) correlations in the data make it difficult to analyze and interpret using standard statistical techniques. This project applies a machine learning technique, called Self Organizing Map (SOM), to present the high-dimensional soil data in a 2D format suitable for human understanding and interpretation. This SOM representation enables analysis of the soil chemical concentration trends within the metro area and in the state of Minnesota. These trends are important for various Minnesota regulatory agencies concerned with the concentration of polluting chemical elements due to both (a) human activities, i.e., different industrial land usage, and (b) natural geological factors, such as the geomorphic codes and provenance of glacial sediments.

17. Document Analysis/Descriptors	18. Availability Statement
Self-organizing maps (SOM), Multi-dimensional scaling, Pollution, Soil chemical survey data, Geological surveying, Cluster analysis	No restrictions. Document available from: National Technical Information Services, Springfield, Virginia 22161

19. Security Class (this report)	20. Security Class (this page)	21. No. of Pages	22. Price
Unclassified	Unclassified	202	

Statistical Analysis of the Soil Chemical Survey Data

Final Report

Prepared by:

Sauptik Dhar
Vladimir Cherkassky

Department of Electrical and Computer Engineering
University of Minnesota

June 2010

Published by:

Minnesota Department of Transportation
Research Services Section
395 John Ireland Boulevard, MS 330
St. Paul, MN 55155

ACKNOWLEDGEMENTS

We would like to thank Dr. Robert Edstrom, who served as the Mn/DOT technical lead on this project, for numerous discussions regarding the goals of data modeling and interpretation of results. We also gratefully acknowledge contributions from:

- Mr. James Seaberg who provided the list of rock forming elements and assisted in interpreting the pattern of clustering of these elements;
- Mr. Steven Hennes who provided the list of elements of regulatory interest and helped to interpret the pattern of clustering of these elements;
- Mr. Feng Cai who performed preprocessing of the soil survey dataset, including the analysis of missing data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

EXECUTIVE SUMMARY

Here we provide a brief overview of the report and a summary of main results. The report is organized as follows:

- Chapter 2 describes sampling scheme used for data collection;
- Chapter 3 describes preprocessing steps, this involves handling missing data;
- Chapter 4 describes main modeling results, including description of SOM clustering method, the goals of data modeling and actual SOM modeling results (implementing these goals);
- Conclusions are given in Chapter 5;
- Several appendices describe additional technical results.

The main part of the report (Chapter 4) describes SOM modeling results for preprocessed Metro 2001 Soil Survey Data and MN Statewide 2003 Soil Survey Data. The goal is to perform clustering of 45-dimensional data samples where each sample represents 45 inorganic contaminants collected at different locations (sites). Such clustering shows how different sites are related (similar) to each other based on their soil characteristics, and/or concentration of some selected elements of interest. The resulting SOM model is a set of clusters represented as a 2D grid, which is suitable for visual interpretation. Data modeling is performed separately for three data sets.

Metro 2001 Soil Survey Data

This data represent the data samples collected at different sites along the Mn/DOT roads within the Minneapolis-St. Paul metropolitan area. The purpose of analysis is to understand how the concentration of elements changes with the distance from major highways and their location in different circles representing different areas. A detailed description of the goals of modeling is provided in Section 4.2.

RESULTS and CONCLUSIONS

Based on the results obtained we observe that the soil samples that are collected near to the roads have similar soil characteristics and the soil samples that are collected far away from the roads have similar characteristics. Further we also observe that the background samples seem to have similar soil characteristics as the samples that were collected far away from the roads. We also observe that the soil samples that are collected near to the roads have high concentration values for the elements Pb, W, Zn and Cu, in contrast to the soil samples collected far away from the roads. Detailed description of the results is provided in Section 4.3.1 and the summary of the results is provided in Section 4.3.2.

MN Statewide 2003 Soil Survey Data

This data represent samples collected at different sites throughout Minnesota. The sites are selected so that they represent the statewide geomorphologic units. The purpose of this analysis is to explore the pattern of clustering based on the geomorphologic units and the provenance of the glacial sediments. Detailed description of the goals of modeling is provided in Section 4.2.

RESULTS and CONCLUSIONS

Based on the results obtained we observe a good pattern of clustering based on the provenance. In particular, there is a good pattern of clustering based on the provenance when the soil characteristic is exclusively determined by the rock forming elements Al, Fe, Mg, Ca, Na, K, and Mn. Moreover, we observe that the soil samples collected from northeastern Minnesota have higher concentration values for the elements As, Cr, Cu, Ni, W in contrast to the soil samples collected from other parts of the state. Detailed description of the results is provided in Section 4.3.3 and the summary of the results is provided in Section 4.3.4.

Metro 2001 Soil Survey Data and MN Statewide 2003 Soil Survey Data (joint analysis of both data sets)

For this case, we only use the elements that are common to both the datasets.

RESULTS

In general, we observe that the Statewide 2003 soil samples have similar soil characteristics as the Metro 2001 soil samples that are collected far away from the roads. Further we also observe some interesting patterns of clustering for the elements of regulatory interests' i.e. As, Cr, Cu, Pb, Ni, W, and Zn. A detailed description of the results is provided in Section 4.3.5 and the summary of the results is provided in Section 4.3.6.

1 INTRODUCTION

The soil chemical data produced by the 2001 metro soil survey and by the 2003 state-wide survey have not been analyzed and reported. This data potentially contains useful information about chemical concentration trends as well as the effects of native geology contributions to various chemicals in the soil. However, the data is not straightforward to analyze, because each sample measures concentration of 45 different chemicals. This high dimensionality and possible unknown (nonlinear) correlations in the data make it difficult to analyze via standard statistical techniques. In this report we aim at applying nonlinear machine learning techniques to derive statistically meaningful and interpretable models based on available data. These statistical models would help the Minnesota Department of Transportation (Mn/DOT) to:

1. Present available high-dimensional data in a form suitable for human understanding and interpretation.

2. Answer some questions from regulatory agencies about soil chemical concentration trends within metro area and in the state of Minnesota.

Basically the report is organized as described below:

In Chapter 2 we provide a brief description of the soil sampling scheme used to collect the soil survey data. In chapter 3 we describe about the initial preprocessing of the datasets for the SOM modeling. Chapter 4 contains the main content of the report. In this chapter we describe our goals of modeling and present the results and a summary of the conclusions. Finally we conclude the report in chapter 5.

2 SOIL SAMPLING SCHEME

In this chapter we describe the soil sampling scheme used to collect the data.

2.1 Metro 2001 Soil Survey Data

This dataset represent the data samples collected at different sites along the Mn/DOT roads within the Minneapolis-St. Paul metropolitan area. The soil sampling scheme for this data set is reproduced below from [1]. This scheme uses three concentric circles encompassing the Minneapolis-St. Paul metropolitan area. The center corresponds to Minneapolis Post Office, and each circle generally covers the following areas:

Circle I: Downtown
Circle II: First and second ring suburbs.
Circle III: The edge of the second ring suburbs and the rural lands.
(A schematic representation is shown in Figure 2.1.1)

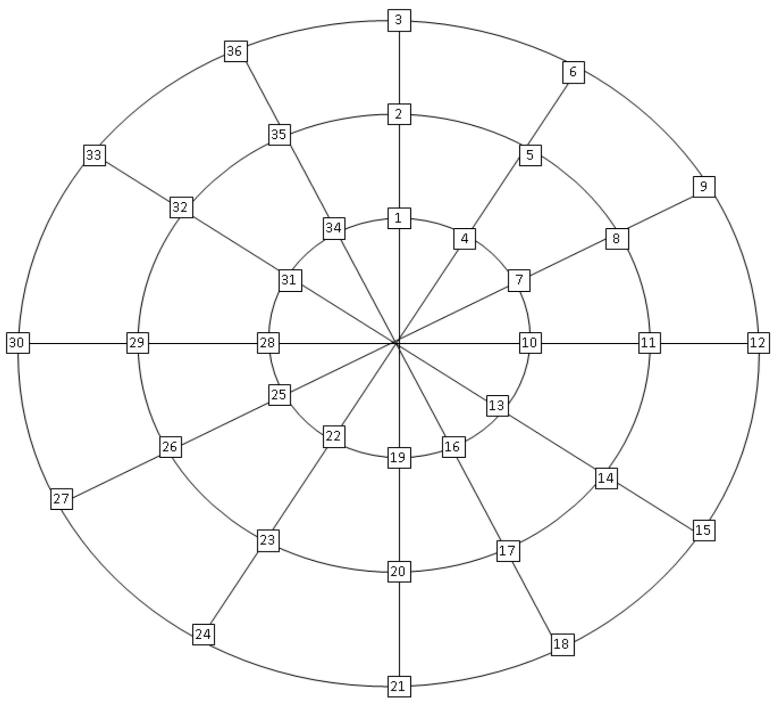


Figure 2.1.1. Map used for the soil sampling scheme for the Metro 2001 Soil Survey Data.

The circles are divided by 12 line segments (each 30^{o} apart). And the points at which the lines intersect the circles are identified as sites (numbered from 1-36 shown inside the squares in Figure 2.1.1). For each site three soil samples are collected at different distances from the road. These samples are named as Type-A, Type-D and Type-E, where
Type-A: Represents the data samples that are nearest to the road.
Type-D: Represents the data samples that are relatively further away from the road. It generally indicates ditch out-slope.

Type-E: Represents the data samples that are farthest from the road. These are basically the samples collected from the end of Right-of-Way.

A schematic representation of the map used to illustrate the different soil sample types is given in Figure 2.1.2. (This figure is reproduced from [2]).

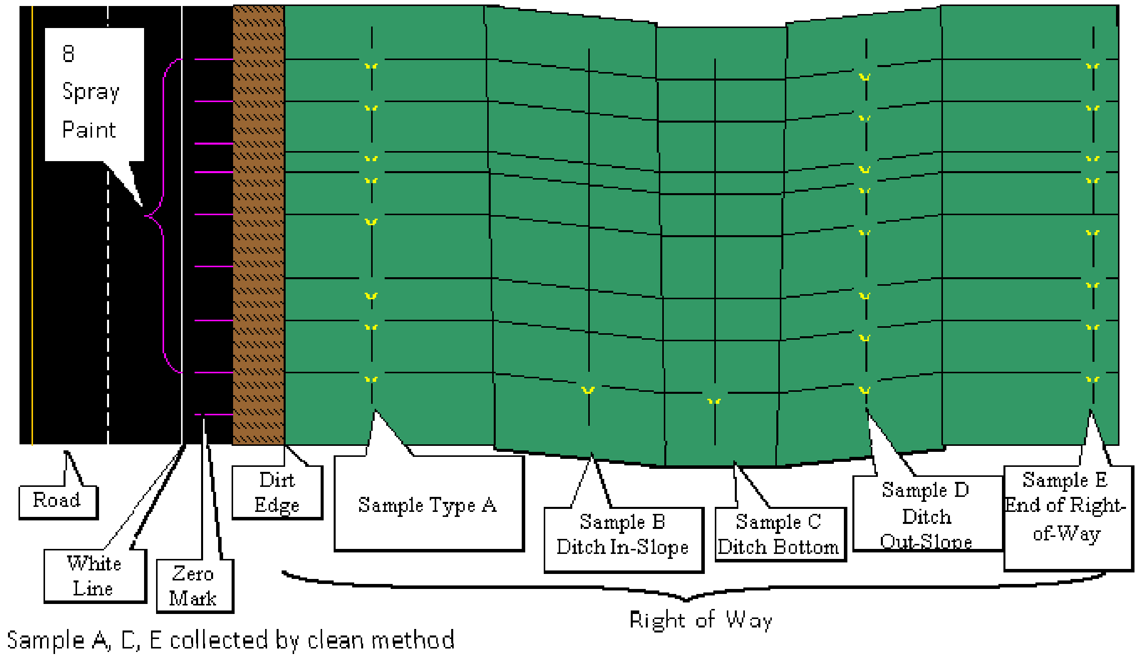


Figure 2.1.2. Map explaining different soil sample types for the Metro 2001 Soil Survey Data.

Note: Apart from this we also have 8 background soils samples for each of the Circle I, Circle-II and Circle III. These soil samples were not along Mn/DOT roads and have been collected near a parking lot or minor city road. A brief description of the background samples is provided in Table 2.1 below.

Table 2.1.1. Background sample sites description

BACKGROUND SAMPLE SITE DESCRIPTIONS		
Ring	Mn/DOT sample ID	Descriptions of sample site
I	37-A	Evergreen Park Roseville (Baseball Field)
I	38-A	Roseville Park (Playground lawn)
II	39-A	Woodland lawn perhaps in a park
II	40-A	Bare spot in wooded area (Perhaps near a school)
III	41-A	Forest Lake truck station near woods.
III	42-A	Carlos Avery Wildlife Mgmnt Area North Metro Office near woods
I	43-A	Minnehaha Falls wooded lawn near nature trail
I	44-A	Metro lake wooded lawn
II	45-A	Wooded lawn with fence
II	46-A	Wooded lawn residential
III	47-A	Unknown state park
III	48-A	Grassy lawn near a farm
I	49-A	W 44th St. and Lake Harriet Parkway grassy roadside
I	50-A	City roadside possibly lake area
II	51-A	Bryant Lake regional park
II	52-A	Bush Lake Park grassy area
III	53-A	Wooded lawn residential
III	54-A	Cleary Lake Regional Park grassy roadside
I	55-A	Bassett Creek Park wooded lawn
I	56-A	North Bass Lake Park grassy lawn near building
II	57-A	Clifton French Regional Park along grassy woodland
II	58-A	Grassy woodland dirt road
III	59-A	Elm Creek Park Reserve Horse Camp grassy field
III	60-A	Brushy grass along dirt road

2.2 MN Statewide 2003 Soil Survey Data

The Mn/DOT soil sampling scheme for the MN Statewide 2003 Soil Survey Data is reproduced below from [3]. In this sampling scheme different sites are selected so that they represent the statewide geomorphologic units. Graphical representation of the geomorphologic codes (overlaid on the map of MN) is shown in Figure 2.2.1.

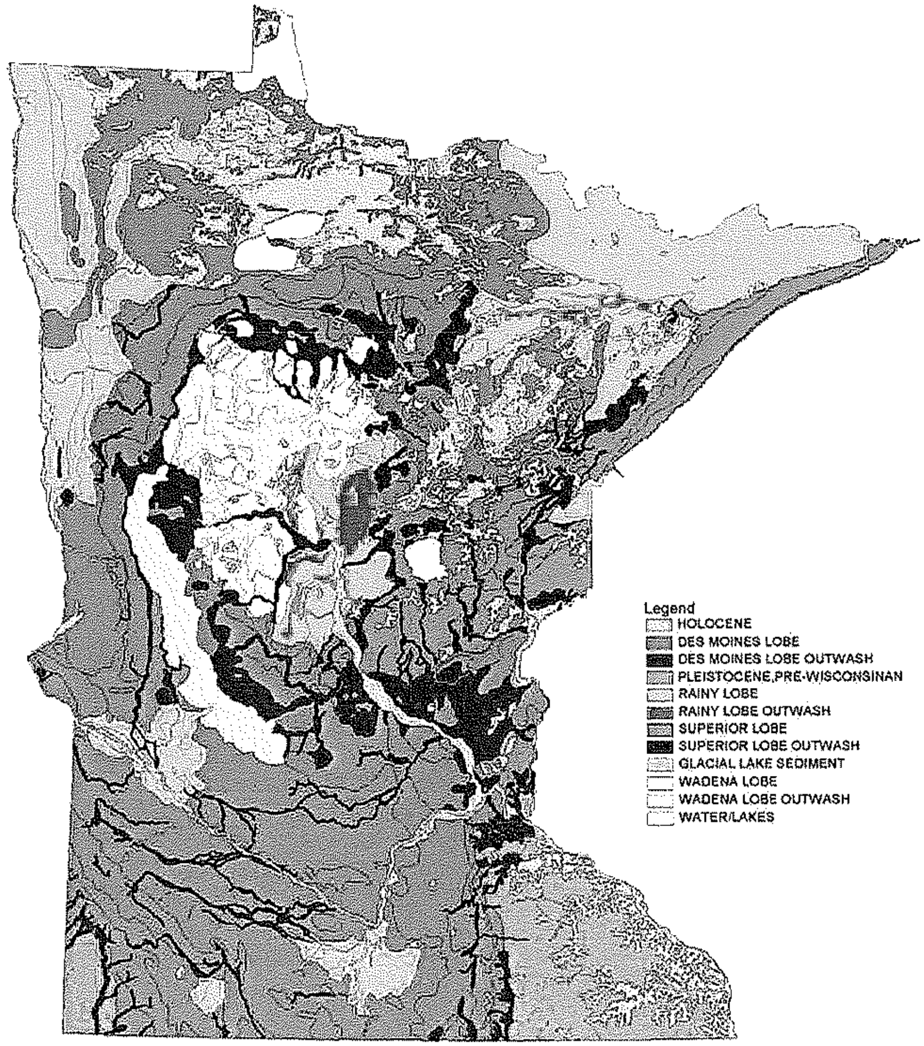


Figure 2.2.1. Statewide geomorphologic units.

Moreover, for each site, three soil transects are collected based on the land use and topography of that site. The transects are mainly used to capture the variance of the soil characteristics within a particular site. So the soil characteristic of a particular site can be captured by averaging transects for that site.

3 PREPROCESSING OF DATASET

In this chapter we provide a description of the preprocessing of the data. In section 3.1 we provide the characterization of the missing data. We provide a brief description of the summary of missing data and then explain how we handle the missing data. In section 3.2 we provide a characterization of the data after scaling.

3.1 Characterization of missing data

3.1.1 Summary of missing data

Metro 2001 Soil Survey Data
A table for the missing data for the Metro 2001 Soil Survey Data is provided in Table 3.1.1. The abbreviation used for the different elements for this data is explained in Appendix E.

Table 3.1.1. Table showing the number of samples for the different elements for Metro 2001 Soil Survey Data

Elements	Total no. of samples	No. of missing samples	% Missing samples
P10	130	0	0.00
Silt	130	1	0.77
Clay	130	1	0.77
SG	130	1	0.77
OM	130	0	0.00
PH	130	0	0.00
EP	130	0	0.00
eK	130	0	0.00
dAl	130	82	63.08
dAs	130	9	6.92
dBa	130	82	63.08
dBr	130	9	6.92
dCa	130	82	63.08
dCe	130	9	6.92
dCo	130	9	6.92
dCr	130	9	6.92
dCs	130	9	6.92
dCu	130	82	63.08
dEu	130	82	63.08
dFe	130	9	6.92
dHf	130	82	63.08
dK	130	82	63.08
dLa	130	9	6.92
dMg	130	82	63.08
dMn	130	9	6.92
dMo	130	9	6.92
dNa	130	82	63.08
dNi	130	9	6.92

dPb	130	9	6.92
dRb	130	9	6.92
dSb	130	9	6.92
dSc	130	9	6.92
dSe	130	27	20.77
dSm	130	9	6.92
dSr	130	9	6.92
dTa	130	82	63.08
dTb	130	82	63.08
dTh	130	9	6.92
dTi	130	82	63.08
dU	130	9	6.92
dV	130	9	6.92
dW	130	9	6.92
dY	130	57	43.85
dZn	130	9	6.92
iCd	130	2	1.54
iCu	130	2	1.54
iPb	130	2	1.54

A brief summary of the missing data can be given as:

i) There are 8 nonmetal parameters. Data for three nonmetal parameters are missing at site 3-E.

ii) There are 36 ‘DUBNA’ metal parameters. Data are missing at 9 sites for all these metal parameters. Measurements for 12 metal parameters are missing at the same 82 sites (There are total of 130 sites, 63% of the data are missing). Metal parameter ‘dSe’ has 27 (20.8%) measurements missing and ‘dY’ has 57 (43.8%) missing.

iii) There are 3 ‘Interpoll’ metal parameters. These parameters give duplicate analysis for quality assurance. Interpoll data has (almost) no missing measurements.

MN Statewide 2003 Soil Survey Data

A table for the missing data for the MN Statewide 2003 Soil Survey Data is provided in Table 3.1.2.The abbreviation used for the different elements for this data is explained in Appendix E.

Table 3.1.2. Table showing the number of samples for the different elements of MN Statewide 2003 Soil Survey Data

Elements	Total no. of samples	No. of missing samples	% Missing samples
P10	189	37	19.58
S&G	189	41	21.69
SILT	189	41	21.69
CLAY	189	41	21.69
OM	189	37	19.58
PH	189	37	19.58
EP	189	37	19.58
EK	189	37	19.58
Ssalt	189	37	19.58
TOT_SOL	189	36	19.05
CU	189	36	19.05
NI_ICP	189	36	19.05
PB	189	36	19.05
TI	189	2	1.06
AL	189	2	1.06
FE	189	2	1.06
MG	189	2	1.06
CA	189	2	1.06
NA	189	2	1.06
K	189	2	1.06
MN	189	2	1.06
BR	189	2	1.06
SC	189	2	1.06
V	189	2	1.06
CR	189	3	1.59
CO	189	2	1.06
NI_NAA	189	23	12.17
ZN	189	2	1.06
AS	189	2	1.06
SB	189	5	2.65
SE	189	114	60.32
RB	189	2	1.06
CS	189	2	1.06
SR	189	10	5.29
BA	189	2	1.06
LA	189	2	1.06
CE	189	2	1.06
ND	189	2	1.06

SM	189	2	1.06
EU	189	2	1.06
TB	189	3	1.59
YB	189	2	1.06
LU	189	2	1.06
ZR	189	2	1.06
HF	189	2	1.06
TA	189	2	1.06
MO	189	3	1.59
W	189	43	22.75
HG	189	189	100.00
TH	189	3	1.59
U	189	2	1.06
AG	189	15	7.94
AU	189	94	49.74
IR	189	169	89.42

A brief summary of the missing data can be given as:

i) There are 189 sampling sites, 31 of which are Interpoll replicates.
ii) There are 10 nonmetal parameters. Measurements at 6 sites are missing for 7 nonmetal parameters and for the rest 3 nonmetal parameters, each has 10 missing measurements.
iii) There are 44 metal parameters. At site 00053-4, measurements for 41 metal parameters are missing. Four metal parameters have more than 50% measurements missing. These are ‘SE’,’HG’, ‘AU’ and ‘IR’. There are 12 metal parameters with fewer than 30% measurements missing.

3.1.2 Possible reasons for missing data

Below detection level
Some missing data may be due to measurements below detection level. In this case, missing measurements can be filled in by some small values. Note that this cause of missing data (below detection level) suggests random pattern of missing values. However, most missing data points have regular pattern.

Recording error or measurement error
Some missing data may be due to measurement error or recording error. In this case, missing entries can be approximated using measurements from neighboring sites.

Budget cuts
Most missing data seems to be due to budget cuts. We found for some parameters such as ‘SE’,’HG’, ‘AU’ and ‘IR’ in MN Statewide 2003 Soil Survey Data (iii), over 30% measurements are missing and those missing measurements are very regularly distributed. So we think the missing measurements are due to budget cuts. If this is the case, a possible way to deal with missing data is to exclude such inputs (with many missing values) or to exclude site(s) with missing measurements, from data analysis.

General guidelines for handling missing data

In general, data-analytic models depend on (a) available data, and (b) a priori knowledge about the true model and about the data. When the data is missing, the corresponding samples need to be either discarded or filled in (using a priori knowledge). So if we know that the reason for missing data is below-detection-level, we fill in reasonable values. For this report we use a value of 0.5 times detection limit. If the reason for missing data is budget cuts, a large portion of measurements, say over 50% is missing. In this case, the missing data cannot be replaced with some reasonable values, and we remove them from our modeling.

3.1.3 Handling of missing data

Metro 2001 Soil Survey Data

For our purpose we perform the following steps:

i. Remove the elements for which the percent of missing elements is greater than 50%.
ii. Remove all the 9 sites (9-A,12-A,9-D,10-D,12-D,13-D,9-E,12-E,25-E) for which data is missing for all the metal parameters.
iii. We replace all the missing data for SE and Y with 0.5 times the detection limit 0.25 parts per million (ppm) and 0.5 ppm respectively.(This has been approved by the Technical Advisory Panel (TAP) committee of Mn/DOT).
iv. Further we remove the 'Interpoll' metal parameter iPb from our analysis. This is because we already have the readings for its Dubna counterpart. Moreover as we have removed the Dubna measurement of Cu (because of more than 50% missing data) we use its Interpoll measurement for our analysis.

MN Statewide 2003 Soil Survey Data

For our purpose we perform the following steps:

i. The 31 'Interpoll' samples are excluded from Self-Organizing Maps (SOM) data analysis.
ii. At site 00053-4, measurements for 41 metal parameters are missing. This site is excluded from our analysis.
iii. Moreover the metals 'SE','HG', 'AU' and 'IR' which has more than 50% missing samples are excluded from our analysis.
iv. Further we replace all the missing data for the 12 metal parameters (with fewer than 30% measurements missing) with their equivalent 0.5 times the detection limit values. (This has been approved by the TAP committee of Mn/DOT).

(Moreover we remove the sample 80 from analysis. We do not have any details about this sample).

Note: The histogram of all the elements for both Metro 2001 Soil Survey Data and MN Statewide 2003 Soil survey Data after handling the missing data is provided in Appendix C.

3.2 Characterization of data after scaling

We then scale both the datasets to a range of [0-1]. It is important to scale the data before the SOM modeling because the ranges of concentration values for different elements are quite different. For instance, in the Metro 2001 Soil Survey Data for Fe(iron) the concentration values ranges from [min=8486 (ppm) to max=36900(ppm)] ;whereas for Cs(cesium) the concentration values ranges from [min=0.2 (ppm) to max=2.6 (ppm)].As such the 2-D representation of the SOM map will be mainly dominated by the elements with the higher range(in this case Fe).So it is necessary to scale the concentration values of all the elements to the range of [0,1] so that the 2-D SOM map can effectively capture the high dimensional data.

Metro 2001 Soil Survey Data

We provide the pair wise Spearman correlation coefficient for all the elements for the Metro 2001 Soil Survey Data after scaling the data to a range of [0-1]. In this case we do not use the background samples. This is provided in Appendix D.

MN Statewide 2003 Soil Survey Data

Unlike the 2001 dataset, the absolute values for the 2003 soil samples at each site are the averaged soil sample concentration of the transects at that site. The primary reason for doing so is that, the transects were mainly collected to capture the information on intra-site variability. For instance, if a site consisted of an upland grassed area, a lowland grassed area and a lowland woodland area; a transect was sampled in each of these areas. As such, it makes sense to consider the average of all the transects of a site to be a representative of the overall soil characteristic for that site. We provide the pair wise Spearman correlation coefficient for all the elements for the MN Statewide 2003 Soil Survey Data after scaling the data to a range of [0-1]. This is provided in Appendix D.

4 SOM MODELING

4.1 Description of SOM modeling

In this section we provide a general description of the Self-Organizing Maps (SOM). SOM model gives visual representation of high-dimensional data as a number of clusters in low-dimensional space (called *Topological Map* or *SOM Map*).Typically SOM map is one-dimensional or two-dimensional grid. Our analysis uses 2D-grid topology shown below:

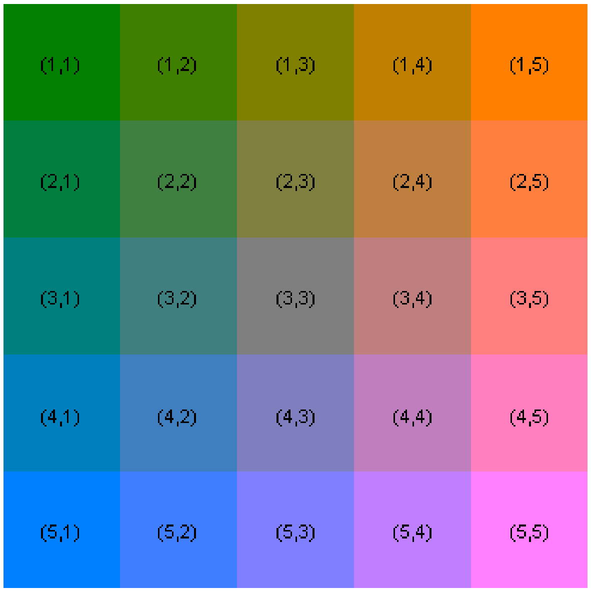


Figure 4.1.1. 2D SOM grid with 25(5x5) units.

In this case each colored square represent an SOM unit. The numbering of each unit is shown in Figure 4.1.1. The units are numbered so that the distance between the nearby units in the topological map is smaller than the distance between the farther units. For example, the distance between the unit (2,5) from the unit (1,5) is {(2-1)+(5-5)}=1 unit. And the distance between the unit (4,2) from the unit (1,5) is {(4-1)+(5-2)}=6 unit.

Example of SOM modeling [4]

Consider modeling the doughnut distribution (shown in Figure 4.1.2), where each point is a two-dimensional data sample with coordinates (x1,x2).This data set can be modeled using 2D map (with 25 units) as given in Figure 4.1.1. Basically the SOM map tries to position its units in the sample space according to non-uniform distribution of the data. And in the process it maintains the topological (neighborhood) relationship between its units. As shown in Figure 4.1.3 the SOM map units are positioned non-uniformly in the input space, as they approximate non-uniform distribution of the data. Yet the topological relationship between the neighboring units (shown in Figure 4.1.1) is maintained for the trained map (shown in Figure 4.1.3). Notice, that in the sample space the units (1, 4) and (2, 5) are still the neighboring units of (1, 5).

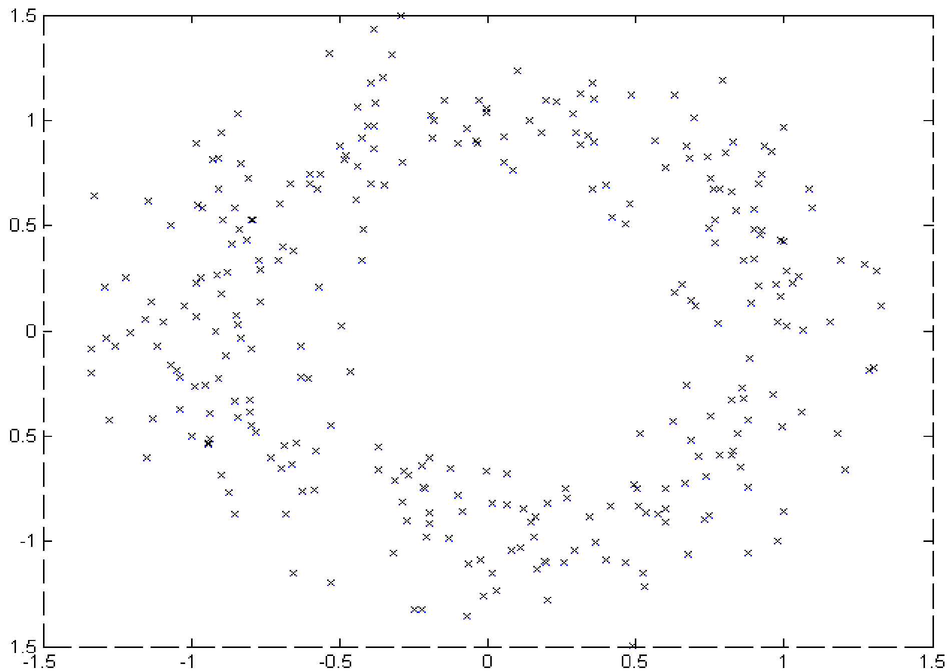


Figure 4.1.2. 300 samples from the Doughnut Distribution corrupted by a Gaussian noise (σ=0.2).

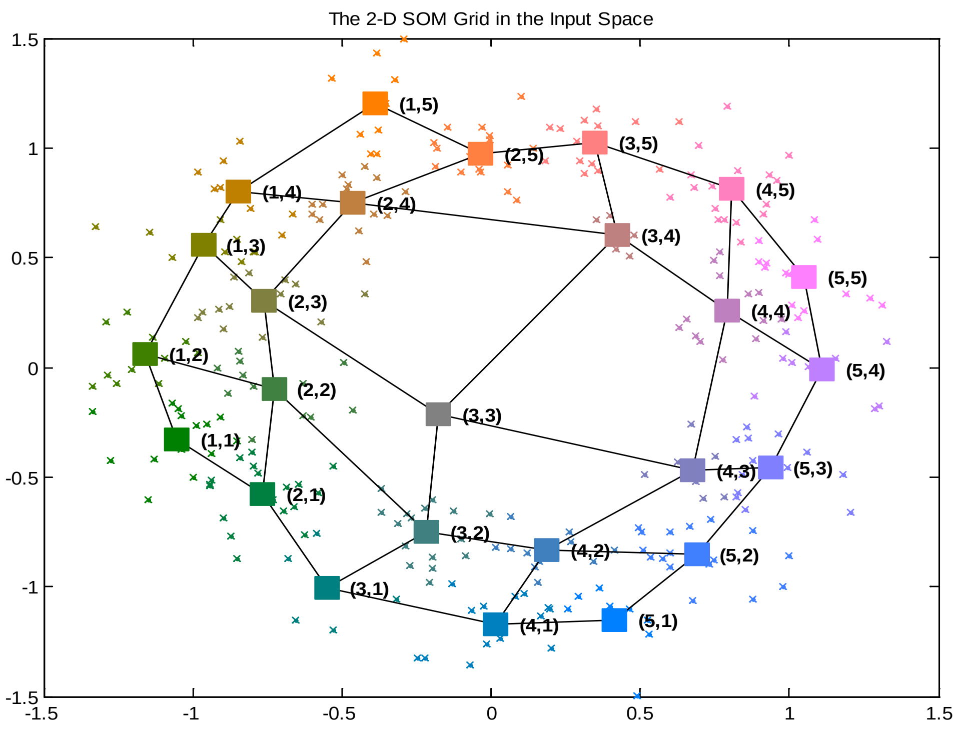


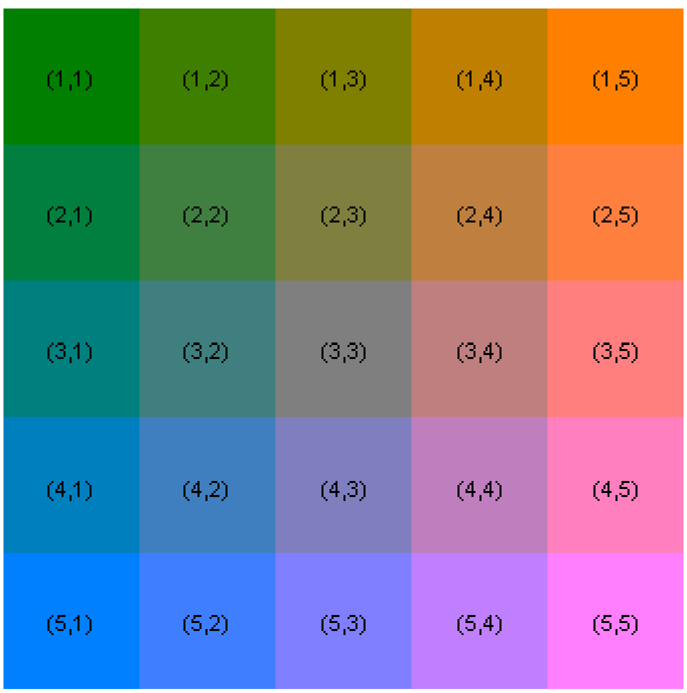
Figure 4.1.3. 5x5 SOM Grid approximating 300 data samples from the Doughnut Distribution shown in Figure 4.1.2.

In the Figure 4.1.3 the data samples are represented by the color of the SOM unit that it falls into. It is evident that the data samples are mapped to the closest SOM unit. Moreover, as seen from the figure, the data samples close to each other in the input or sample space (x1, x2) is mapped onto 'neighboring' units of a 2D SOM grid. For instance, the data samples that are colored in green are mapped into the nearby green SOM units as against the data samples that are colored in pink.
(For additional details on the Self Organizing Maps please refer to [4]).

Note: The example shown in Figures 4.1.2 and 4.1.3 uses 2D data for illustration purpose, but in our application the data is high-dimensional. When modeling high-dimensional data, such as soil sample data, the map displays 2D grid topology, along with soil samples closest to each SOM unit. Visual inspection of data samples mapped onto each SOM units provides useful interpretation of high-dimensional data.

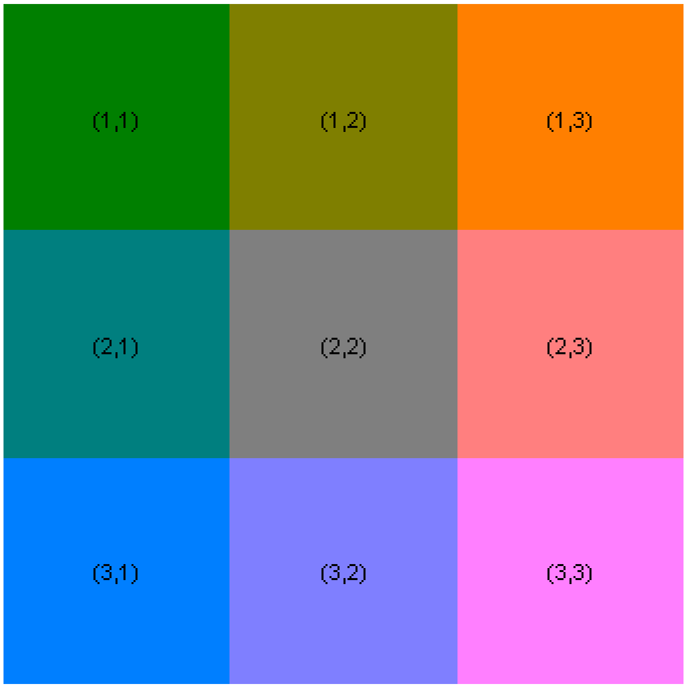
For this report, we have used two different SOM models for the two different datasets.

Metro 2001 Soil Survey Data



2-Dimensional SOM
SOM Algorithm: Batch Version.
Initial radius=1, Final Radius=0.05.
Map Dimension=2.
Number of Units=25 (5x5).
Total Iteration=75.

MN Statewide 2003 Soil Survey data



2-Dimensional SOM
SOM Algorithm: Batch Version.
Initial radius=1, Final Radius=0.05.
Map Dimension=2.
Number of Units=9(3x3).
Total Iteration=75.

4.2 GOALS OF MODELING

In this section we provide a brief description of our goals of modeling. We identify the different objectives of modeling for the Metro 2001 Soil Survey Data and Statewide 2003 Soil Survey Data separately.

4.2.1 Goals of modeling pertaining to the Metro 2001 Soil Survey Data

[**Goal 1**] To estimate how the Metro samples of Circle (I, II, III) and Type (A, D, E) are clustered.
[**MOTIVATION**] The main motivation behind exploring this goal is to understand how the soil samples are related based on their location within the circles (downtown, suburbs or rural lands). Another motivation is to explore if there is any pattern of clustering based on their distance from the Mn/DOT roads. A clustering based on the circles would indicate that the soil samples which are collected at these different circles (downtown, suburbs or rural lands) bear similar characteristics. Moreover, a clustering based on the distance from roads would indicate that the soil samples that are collected at different distances from the roads (Type A, D and E) bear similar characteristics. This pattern of clustering can be attributed to the enrichment of some elements due to proximity to the road. Further it is also important to understand the relation

between the soil samples where the soil characteristics is determined by some specific elements. So we try to explore these patterns of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the selected list of elements: Al aluminum, As arsenic, Ba barium, Ce cerium, Co cobalt, Cr chromium, Cs cesium, Cu copper, Eu europium, Fe iron, Hf hafnium, La lanthanum, Mg magnesium, Mo molybdenum, Ni nickel, Pb lead, Rb rubidium, Sb antimony, Sc scandium, Se selenium, Sm samarium, Sr strontium, Ta tantalum, Tb terbium, Th thorium, Ti titanium, U uranium, V vanadium, W tungsten, Zn zinc. This element list is provided by Dr. Robert Edstrom. We shall call this as *Subset 1*. This subset consists of metals excluding the alkaline earth metals in the full list. The alkaline earth metals (e.g. Na, K, Mn, etc.) are of less importance toxicologically and in regulations.
c) Based on the selected list of elements: As, Cr, Cu, Pb, Ni, W, and Zn. This element list is provided by Mr. Steven Hennes. The underlying intent for using this element list is to explore the pattern of clustering based on the elements that are of regulatory interest. We shall call this as *Subset 2*. For a better understanding of the pattern we further perform SOM modeling taking each of the elements separately.

[**Goal 2**] To estimate how the Metro samples of Circle (I, II, III) and Type (A, D, E) are related to the background samples from Circle (I, II, III).
[**MOTIVATION**] The main motivation for this goal is that we need to understand how the background samples are related to the different soil samples based on the location of the circles or the distance from the roads. Further it is also important to understand the relation between the Metro soil samples with the background soil samples where the soil characteristics is determined by some specific elements. So we try to explore the pattern of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the *Subset 1*.
c) Based on the *Subset 2*. For a better understanding of the pattern we further perform SOM modeling taking each of the elements separately.

4.2.2 Goals of modeling pertaining to the MN Statewide 2003 Soil Survey Data

[**Goal 3**] To estimate how the statewide samples are related based on 5 geomorphologic codes.
[**MOTIVATION**] The main motivation behind this goal is to understand how the samples are clustered based on the geomorphologic codes. Any pattern of clustering would indicate a dependence on the geomorphologic codes which might be useful for soil analysis. We try to explore the pattern of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the *Subset 1*.
c) Based on the *Subset 2*. For a better understanding of the pattern we further perform SOM modeling taking each of the elements separately.
d) Based on the selected list of elements: Al,Fe,Mg,Ca,Na,K,Mn. This list of elements has been provided by Mr. James Seaberg. The underlying intent for using this element list is

to capture the pattern of clustering based on the rock forming elements. We shall call this as *Subset 3*.

e) Based on the selected list of elements: S/G, Clay, Silt, EK, and EP. This element list hasbeen provided by Dr. Robert Edstrom. The underlying intent for using this element list is to capture the pattern of clustering based on the non-metal constituents of the soil. We shall call this as *Subset 4*.

[**Goal 4**] To estimate how the statewide samples are related based on provenance (NW/NE/N).
[**MOTIVATION**] The main motivation behind this goal is to understand how the samples are clustered based on the provenance of the glacial sediments. Any pattern of clustering would indicate a dependence on the provenance which might be useful for soil analysis. We try to explore the pattern of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the *Subset 1*.
c) Based on the *Subset 2*. For a better understanding of the pattern we further perform SOM modeling taking each of the elements separately.
d) Based on the *Subset 3*.
e) Based on the *Subset 4*.

[**Goal 5**] To estimate how the statewide samples are related based on Lacustrine, Till, Outwash (L/T/O).
[**MOTIVATION**] The main motivation behind this goal is to understand how the samples are clustered based on the soil composition Lacustrine, Till and Outwash. Any pattern of clustering would indicate a dependence on this code based on soil composition which might be useful for soil analysis. We try to explore the pattern of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the *Subset 1*.
c) Based on the *Subset 2*. For a better understanding of the pattern we further perform SOM modeling taking each of the elements separately.
d) Based on the *Subset 3*.
e) Based on the *Subset 4*.

4.2.3 Goals of modeling pertaining to both the Metro 2001 Soil Survey Data and MN Statewide 2003 Soil Survey Data

[**Goal 6**] To estimate how the Metro samples are related to the statewide samples and background samples.
[**MOTIVATION**] This goal will further help us to understand the relationship between the Metro 2001 Soil Survey Data samples and the MN Statewide 2003 Soil Survey Data samples. This will further help us to understand how the background samples are related to each of these datasets. Any distinct pattern of clustering between the Metro 2001 samples and the Statewide 2003 samples can be attributed to the enrichment of some elements due to different land usage. We try to explore the pattern of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the *Subset 1*.
c) Based on the *Subset 2*.

[**Goal 7**] To estimate how the Metro 2001 Soil Survey Data samples of Type D& E are related to the MN Statewide 2003 Soil Survey Data samples and background samples.
[**MOTIVATION**] This will help us to understand the relationship between the Metro 2001 Type D & E soil samples and the Statewide 2003 soil samples. We try to explore the pattern of clustering where the soil characteristic is determined:

a) Based on all elements.
b) Based on the *Subset 1*.
c) Based on the *Subset 2*.

4.3 MODELING RESULTS

In this section we describe the results obtained for the different goals. We divide this section into three different subsections where we address the results obtained for the goals pertaining to each dataset.

Note: The dataset used for modeling are the preprocessed Metro 2001 Soil Survey Data and the preprocessed MN Statewide 2003 Soil Survey Data. For the sake of convenience we shall refer to these datasets as the Metro 2001 Soil Survey Data and the MN Statewide 2003 Soil Survey Data in our future discussions.

4.3.1 Results for SOM modeling pertaining to Metro 2001 Soil Survey Data

[**Goal 1a**] To estimate how the Metro samples of Circle (I, II, III) and Types (A, D, E) are clustered based on all elements.

DATASET USED: Metro 2001 Soil Survey Data. All the background samples are removed for this analysis.

NAMING CONVENTION: In order to understand the results better we introduce a naming convention for the different soil samples of the data. This naming convention is described below.

The naming convention used for the soil samples in Metro 2001 Soil Survey Data is specially designed to capture the information implicit to the soil sampling scheme (described in Section 2.1). So, we design a tag to identify the different Circles**,** Site ID and Sample Type. The naming scheme that we used to identify each sample is: **Circle-Site ID-Sample Type**

For Example:
A sample with its tag as **I-4-A** would mean

Circle = Circle I (Downtown)
Site ID (for the sample) = 4
Sample Type= Type A

Note: In this goal we shall only investigate the pattern of clustering based on the Circle and the Sample Type. The Site ID is used just as a tag for the different samples.

SOM MODELING

The SOM modeling results for this goal is shown in Figure 4.3.1.

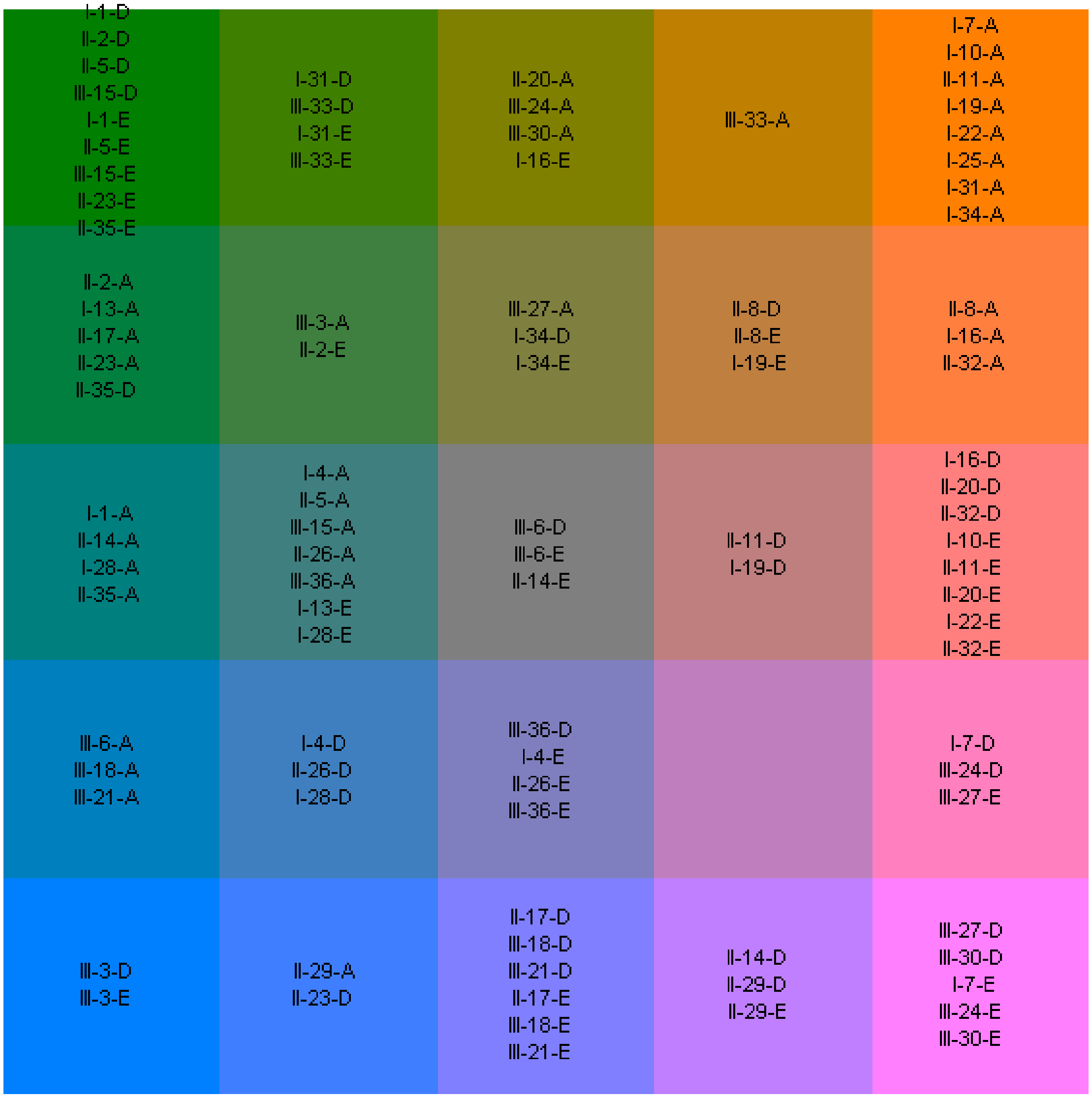


Figure 4.3.1. SOM model for Metro 2001 Soil Survey Data based on all elements.

ANALYSIS

From Figure 4.3.1 we do not see any clustering of soil samples based on their location in Circle I, Circle II or Circle III. This indicates lack of any similarity between the sites based on their location in a Circle. However we can clearly see that the samples of Type A are all clustered with each other and the samples of Type D & E are clustered differently. These results suggest that the *samples collected close to the road (Type A) have similar properties and the samples far away from the road (Type D&E) also have similar characteristics.*

[**Goal 1b**] To estimate how the Metro samples of Circle (I, II, III) and Types (A, D, E) are clustered based on the *Subset 1*.

DATASET USED: Same as Goal 1a.In this case we use only the measurements for the selected *Subset 1* of elements. However, the elements Al, Ba, Eu, Hf, Mg, Ta, Tb, and Ti are not present in the list. The elements Al, Ba, Eu, Hf, Ta, Tb, and Ti were removed during the missing data analysis. There was no measurement for the element Mg in the original data.

NAMING CONVENTION: Same as Goal 1a.

SOM MODELING
The SOM modeling results for this goal is shown in Figure 4.3.2.

Figure 4.3.2. SOM model for Metro 2001 Soil Survey Data based on *Subset 1*.

ANALYSIS

From Figure 4.3.2 we do not see any clustering of soil samples based on their location in Circle I, Circle II or Circle III. This indicates lack of any similarity between the sites based on their location in a Circle. However we can see that the samples of Type A are somewhat clustered with each other and the samples of Type D & E are clustered differently. These results suggest that the *samples collected close to the road (Type A) have somewhat similar properties and the samples far away from the road (Type D&E) also have similar characteristics.* Of course this pattern is not as distinct as seen in the previous case.

[**Goal 1c**] To estimate how the Metro samples of Circle (I, II, III) and Types (A, D, E) are clustered based on the *Subset 2*.

DATASET USED: Same as Goal 1a.In this case we use only the measurements for the selected *Subset 2* of elements.

NAMING CONVENTION: Same as in Goal 1a.

SOM MODELING

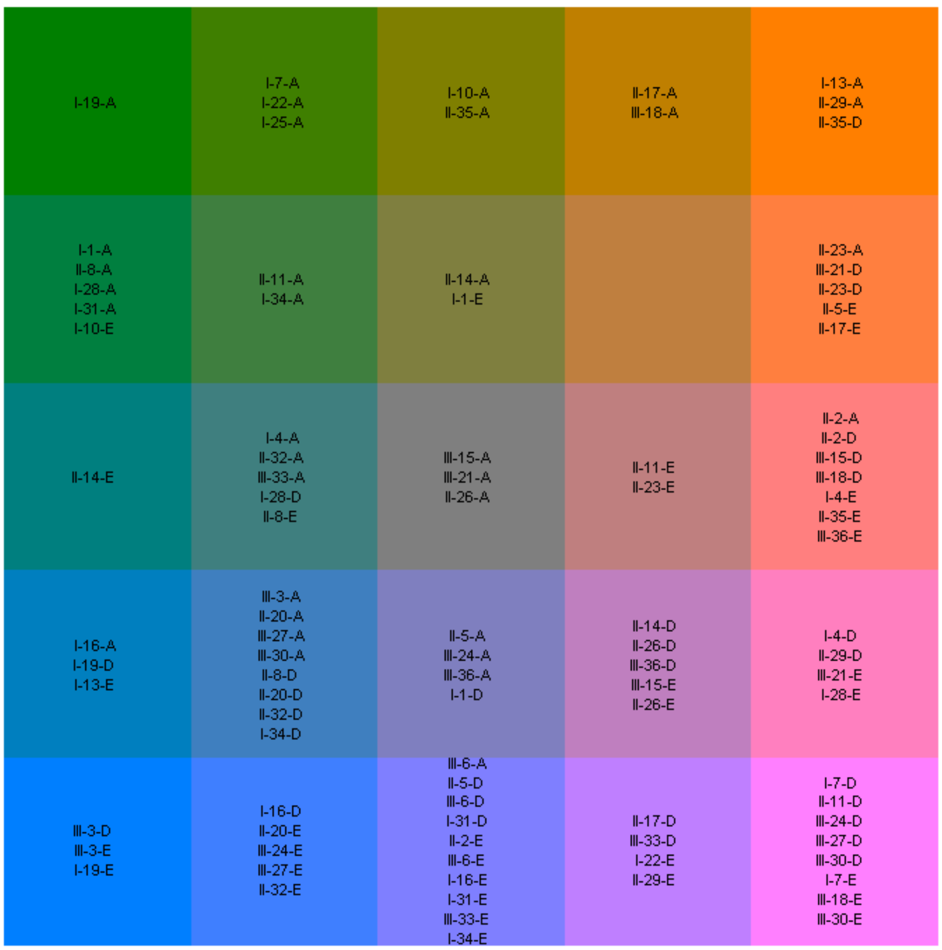


Figure 4.3.3. SOM model for Metro 2001 Soil Survey Data based on *Subset 2.*

ANALYSIS

From Figure 4.3.3 we do not see any clustering of soil samples based on their location in Circle I, Circle II or Circle III. This indicates lack of any similarity between the sites based on their location in a Circle. Moreover although the clustering based on the sample type is not very apparent; we do observe a pattern in clustering. We observe that the samples of Type A are mostly present in the Top half of the Map and the samples of Type D& E are clustered in the bottom half. This indicates that based on the selected subset the *samples collected close to the road (Type A) have similar properties and the samples far away from the road (Type D&E) also have similar characteristics.*

Note that the elements in the *Subset 2* viz., As, Cr, Cu, Pb, Ni, W, and Zn are the metals of regulatory interest. So, in order to have a better understanding of how the samples are clustered based on each one of these elements we analyze the SOM model for each of the elements taken separately.

ELEMENT=Arsenic (As)

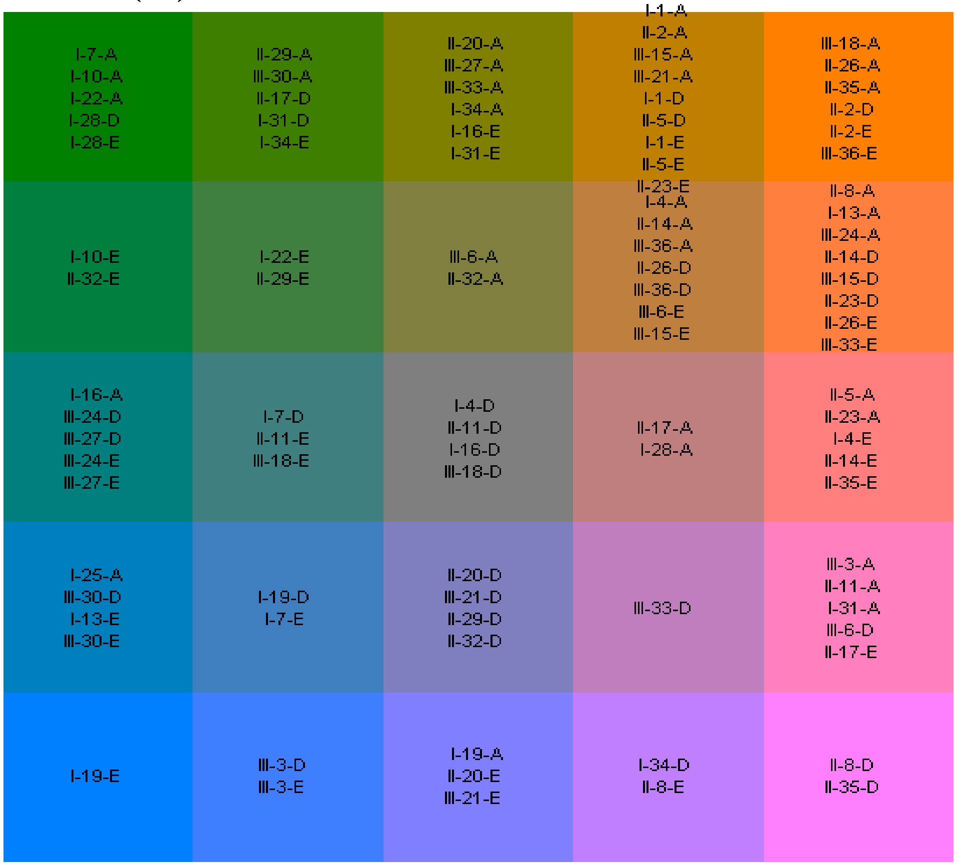


Figure 4.3.4. SOM model for Metro 2001 Soil Survey Data based on Arsenic (As).

ANALYSIS

From Figure 4.3.4 we do not see a very good clustering based on their location in circle. Moreover we do not see any pattern of clustering of the samples of Type A and the samples of Type D & E.

ELEMENT=Chromium (Cr)

Figure 4.3.5. SOM model for Metro 2001 Soil Survey Data based on Chromium (Cr).

ANALYSIS

From Figure 4.3.5 we do not see a very good clustering based on their location in circle or sample type.

ELEMENT= Lead (Pb)

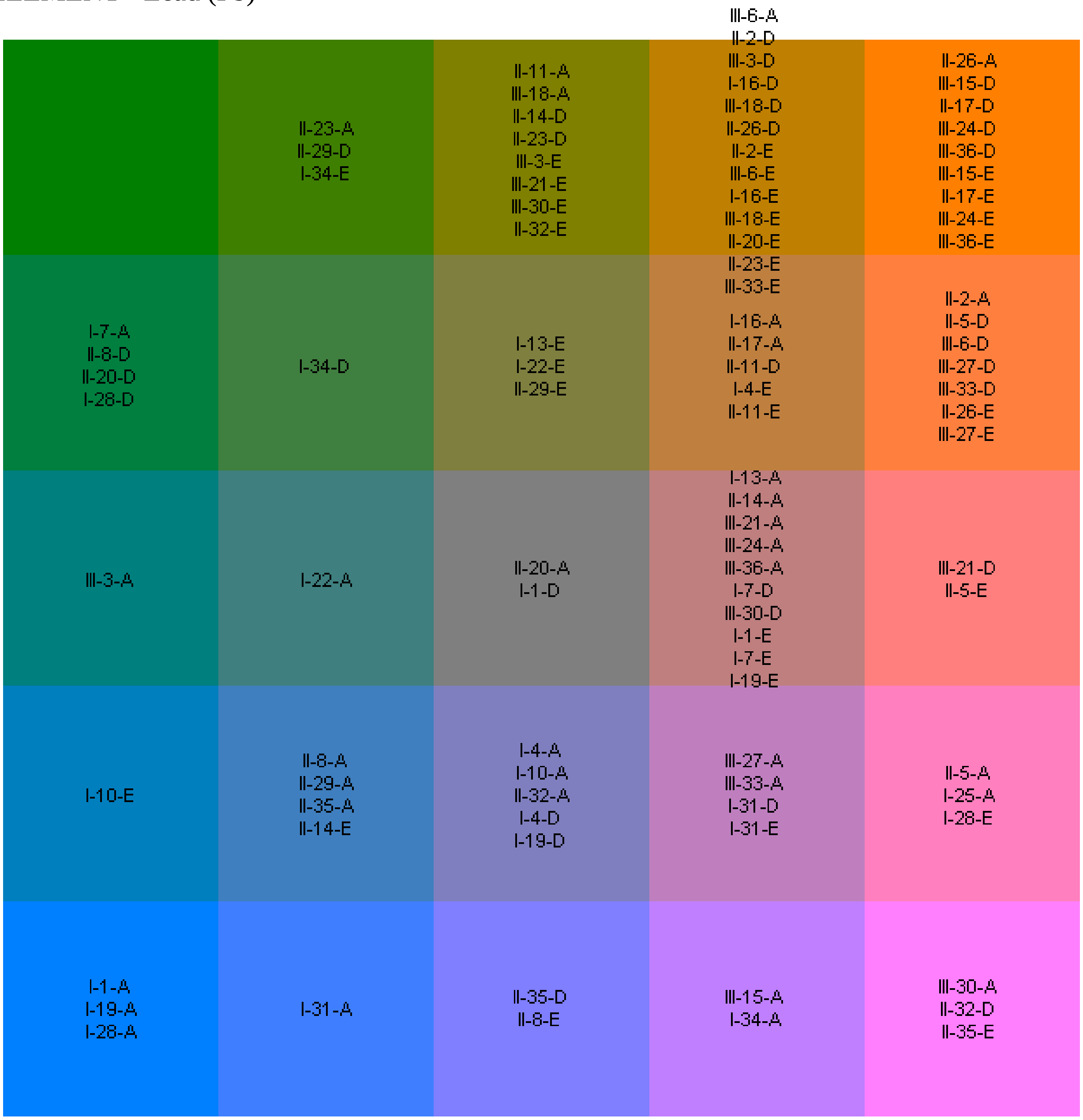


Figure 4.3.6. SOM model for Metro 2001 Soil Survey Data based on Lead (Pb).

ANALYSIS

From Figure 4.3.6 we do not see a very good clustering based on their location in circle. However, we do observe a pattern where we see the samples of Type D&E cluster in the top right of the map. This indicates that based on Pb the soil samples collected far away from the road (Type D&E) could have similar characteristics. Moreover, we found that the samples of Type A have relatively higher Pb concentration than the samples of Type D&E.

ELEMENT=Nickel (Ni)

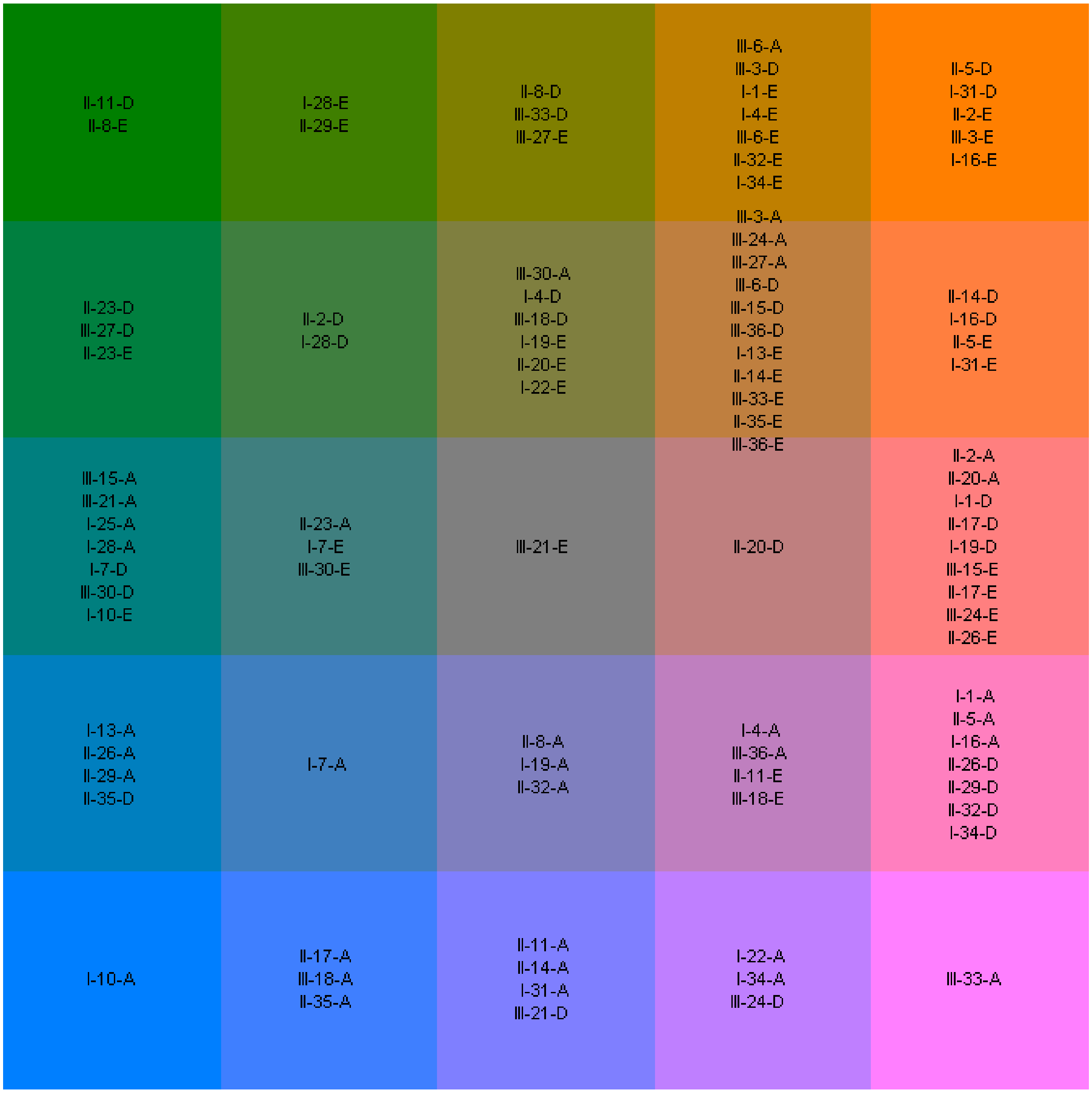


Figure 4.3.7. SOM model for Metro 2001 Soil Survey Data based on Nickel (Ni).

ANALYSIS

From Figure 4.3.7 we do not see a very good clustering based on their location in circle. Although, we do observe some clustering of the samples of Type A (bottom left) against Type D&E (top right of the map) however this clustering is not very strong.

ELEMENT=Tungsten (W) (For the sake of better view ability the SOM Map has been rotated)

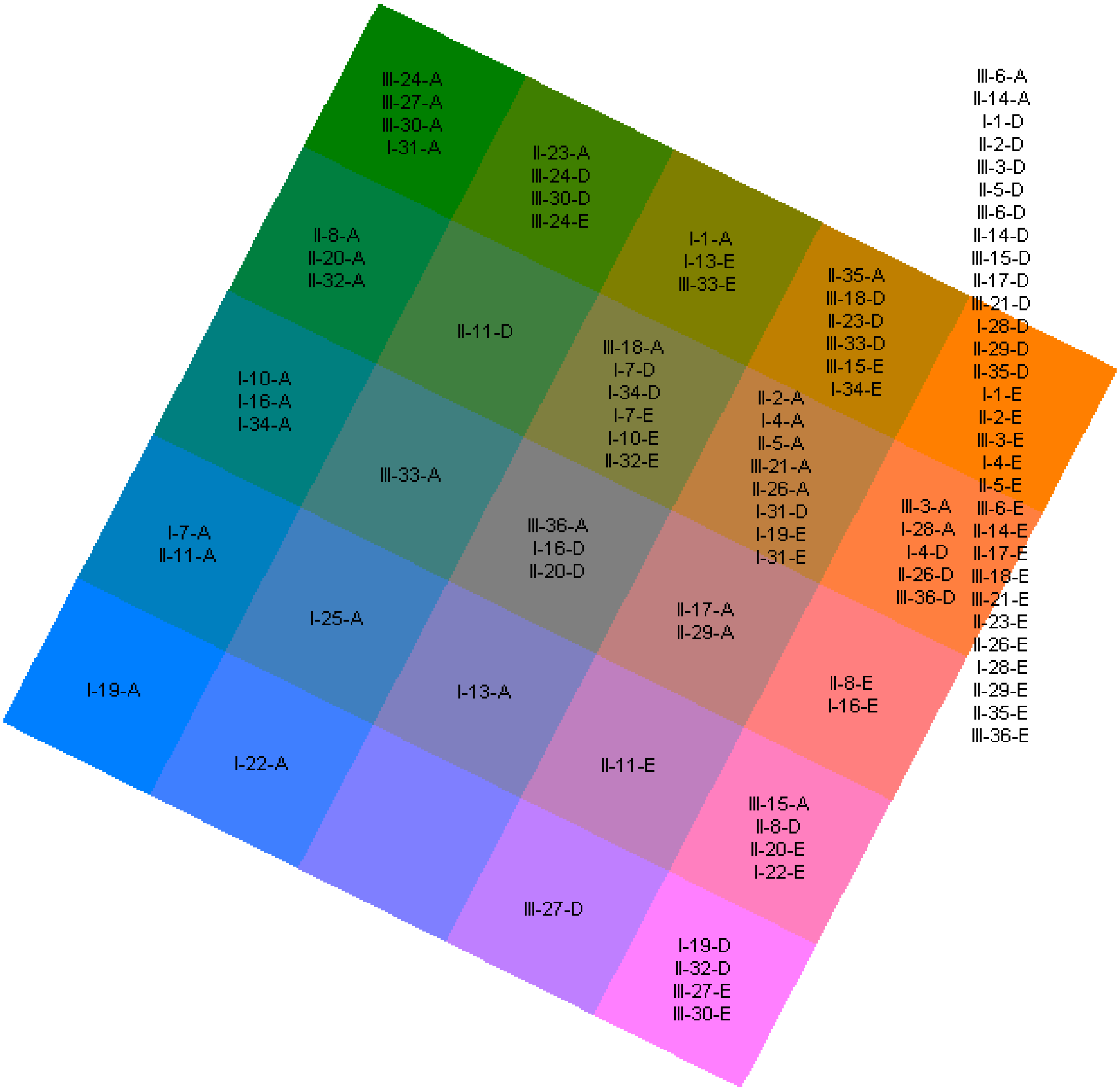


Figure 4.3.8. SOM model for Metro 2001 Soil Survey Data based on Tungsten (W).

ANALYSIS

From Figure 4.3.8 we observe a good clustering of the samples of Type A (left) against Type D&E (top right of the map). This pattern suggest that based on W the soil samples collected near to the road(Type A) have similar W concentration and those far away from the road (Type D&E) have similar W concentration. Further, we have checked that the samples of Type A have relatively higher W concentration values.

ELEMENT=Zinc (Zn)

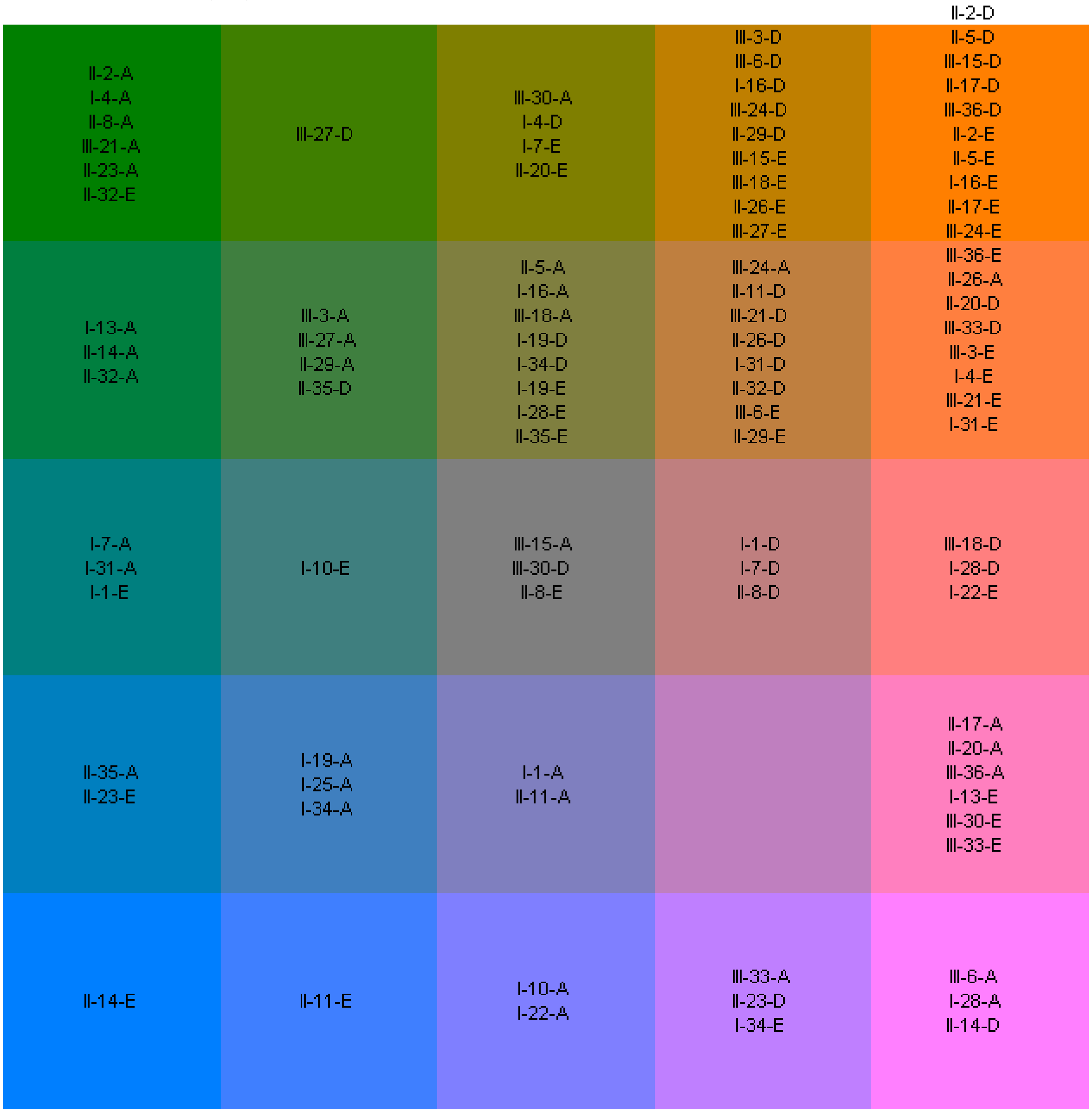


Figure 4.3.9. SOM model for Metro 2001 Soil Survey Data based on Zinc (Zn).

ANALYSIS

From Figure 4.3.9 we do not see a very good clustering based on their location in circle. However, we do observe a cluster of the samples of Type A against Type D&E (top right of the map). This indicates that based on Zn most of the soil samples collected near to the road (Type A) and those far away from the road (Type D&E) have similar characteristics. Further, we have checked that except for some samples like II-14-E, II-11-E most of the Type D&E samples have lower Zn concentration values.

ELEMENT=Copper (Cu)

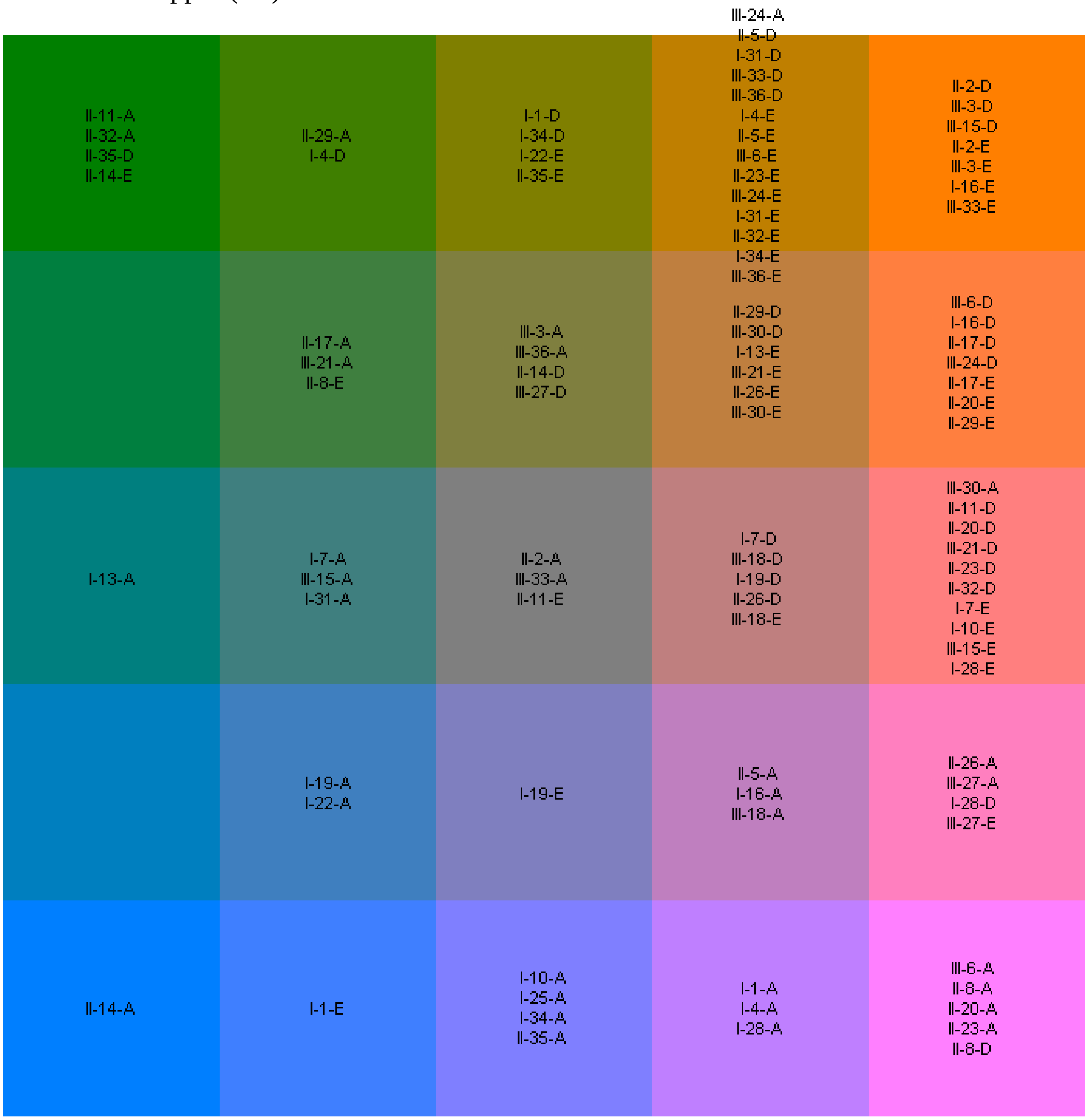


Figure 4.3.10. SOM model for Metro 2001 Soil Survey Data based on Copper (Cu).

ANALYSIS

From Figure 4.3.10 we do not see a very good clustering based on their location in Circle. However we do observe a good cluster of the samples of Type A against Type D&E. This indicates that based on Cu, samples collected close to the road (Type A) have similar properties and the samples far away from the road (Type D&E) also have similar characteristics. Further it seems like the samples of Type A have higher Cu concentration values.

[**Goal 2a**] To estimate how the Metro samples of Circle (I, II, III) and Types (A, D, E) are related to the background samples from Circle (I, II, III) based on all the elements.

DATASET USED: Metro 2001 Soil Survey Data.

NAMING CONVENTION: In order to understand the results better we introduce a naming convention for the different soil samples of the data. This naming convention is described below.

The naming convention used for the soil samples in Metro 2001 Soil Survey Data is specially designed to capture the information implicit to the soil sampling scheme (described in Section 2.1). So, we design a tag to identify the different Circles**,** Sites and Sample Type. The naming scheme that we used to identify each sample is: **Circle-Site ID-Type**

For Example:
A sample with its tag as **I-4-A** would mean

Circle = Circle I (Downtown)
Site ID (for the sample) = 4
Sample-Type= Type A.

Moreover, to identify the background samples we append the tag (Back) to each of the background samples.

For Example:
A sample with its tag as **I-4(Back)** would mean

Circle = Circle I (Downtown).
Site ID (for the sample) = 4.
Sample-Type= background.

SOM MODELING

The SOM modeling results for this goal is shown in Figure 4.3.11.

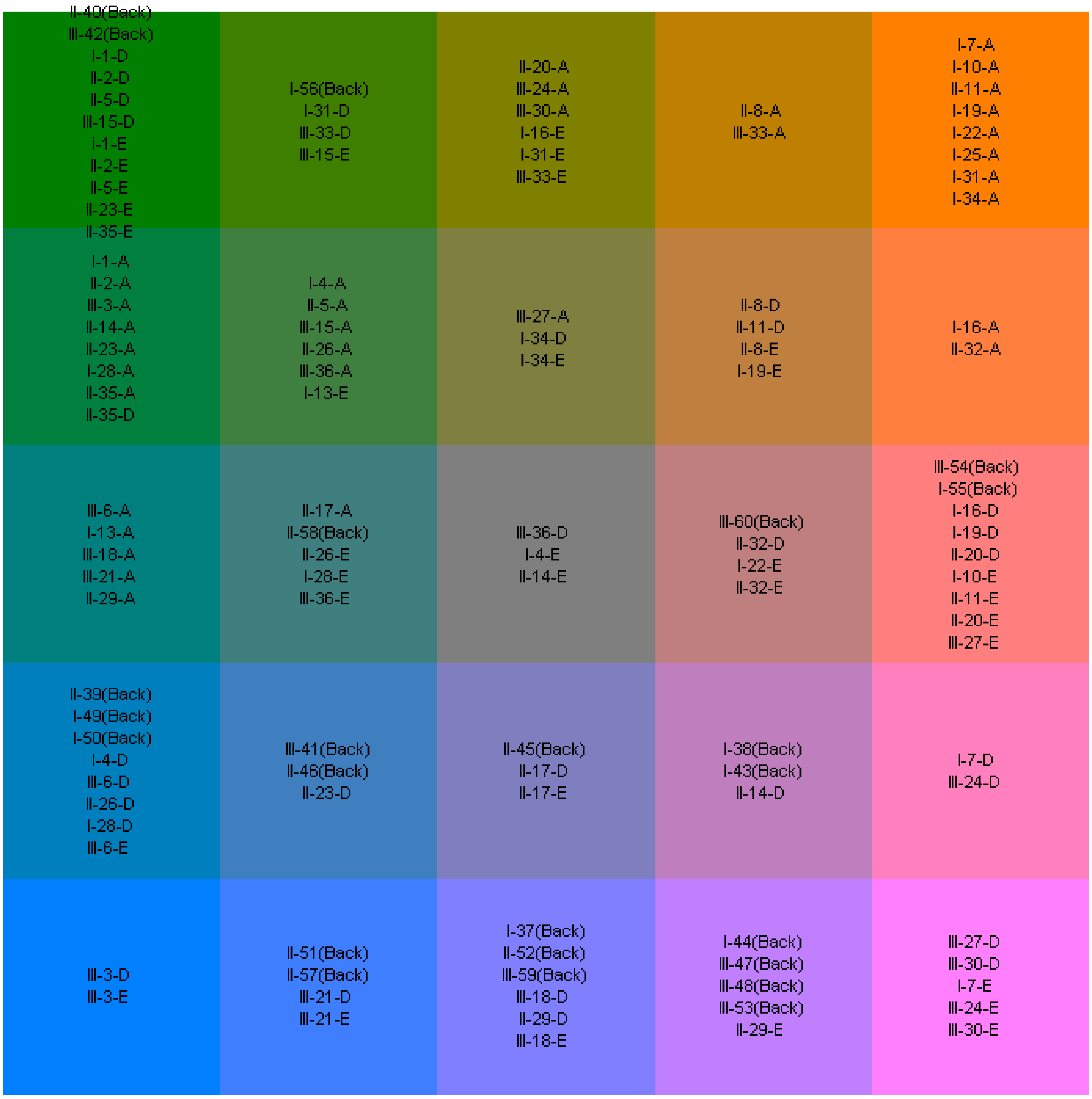


Figure 4.3.11. SOM model for Metro 2001 Soil Survey Data based on all elements (with background samples).

ANALYSIS

From Figure 4.3.11 we do not see any clustering of soil samples based on their location in Circle I, Circle II or Circle III. This indicates lack of any similarity between the sites based on their location in a circle. However we can clearly see (as before) that the samples of Type A are mostly clustered with each other and the samples of Type D & E are clustered differently. Moreover, the background samples seem to be clustered with the samples of Type D&E rather than samples of Type A. Thus, the additional insight that we gain from this model is that based on all the metals the background samples seem to be more related to the samples of Type D& E than the samples of Type A.

[**Goal 2b**] To estimate how the Metro samples of Circle (I, II, III) and Types (A, D, E) are related to the background samples for Circle (I, II, III) based on the selected *Subset 1*.

DATASET USED: Metro 2001 Soil Survey Data. However in this case we use the selected *Subset 1* for modeling.

NAMING CONVENTION: Same as in [**Goal 2a**].

SOM MODELING
The SOM modeling results for this goal is shown in Figure 4.3.12.

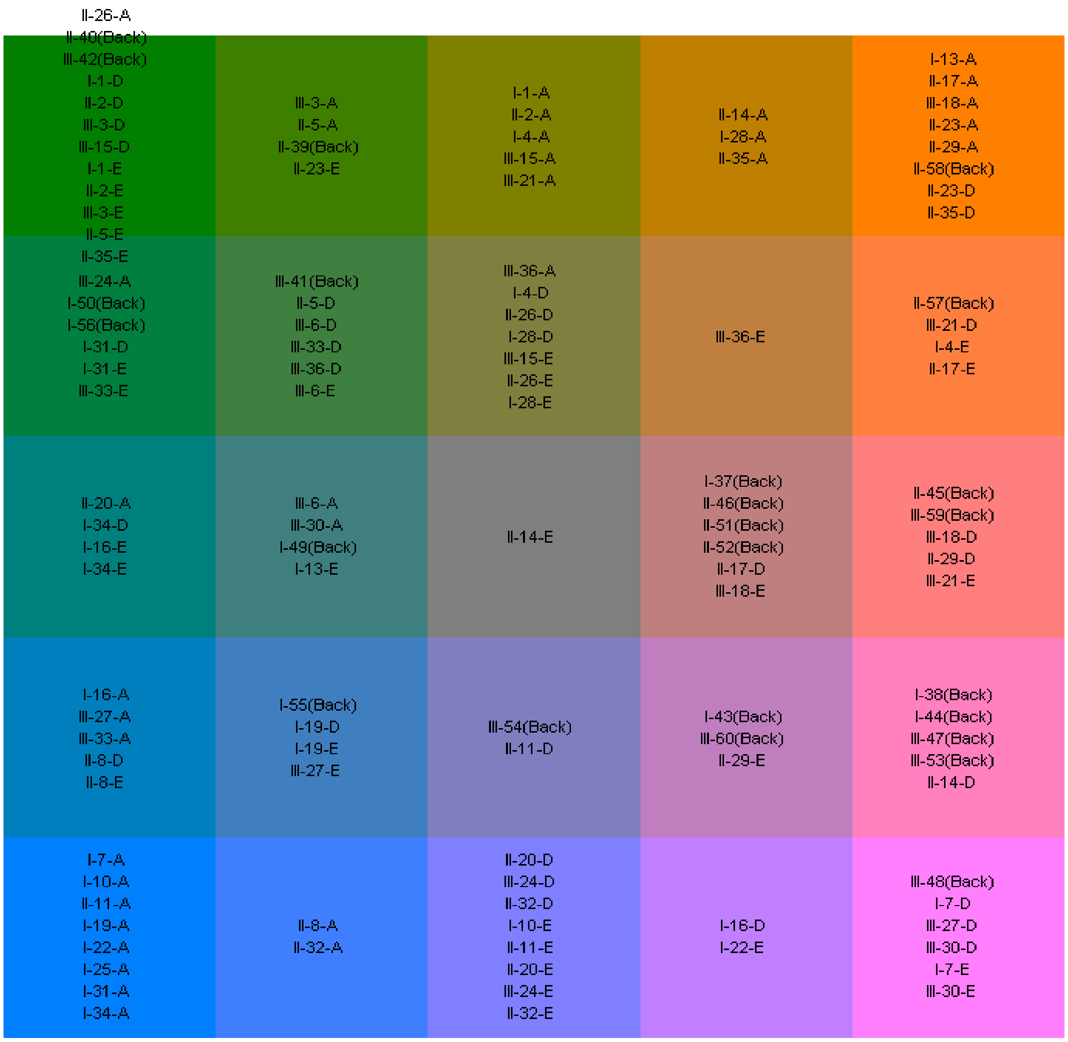


Figure 4.3.12. SOM model for Metro 2001 Soil Survey Data based on *Subset 1* (with background samples).

ANALYSIS
From Figure 4.3.12 we do not see any clustering of soil samples based on their location in Circle I, Circle II or Circle III. This indicates lack of any similarity between the sites based on their location in a Circle. Moreover, although as in Goal 2a the clustering based on the sample type is

not very predominant; we do observe some clustering based on the soil samples of Type A vs. the soil samples of Type D&E. We can also see that the background samples are more clustered with the sample Type D&E rather than the samples of Type A. This is in concordance with our results for Goal 2a. Thus, we can say that based on the selected subset of elements the background samples seem to be more related to the samples of Type D& E than the samples of Type A.

[**Goal 2c**] To estimate how the Metro samples of Circle (I, II, III) and Types (A, D, E) are related to the background samples for Circle (I, II, III) based on the selected *Subset 2*.

DATASET USED: Metro *2001 Soil Survey Data.* However in this case we use the selected *Subset 2* for modeling.

NAMING CONVENTION: Same as in [**Goal 2a**].

SOM MODELING
The SOM modeling results for this goal is shown in Figure 4.3.13.

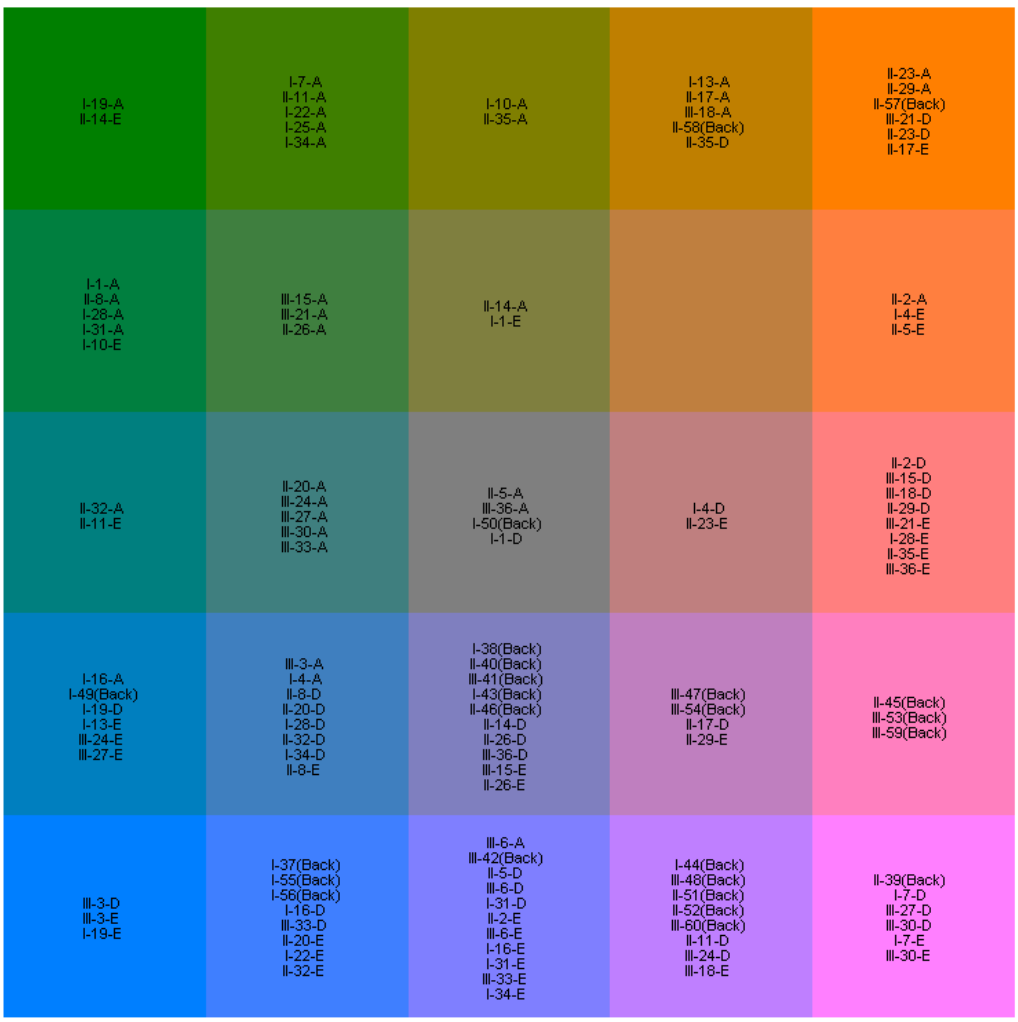


Figure 4.3.13. SOM model for Metro 2001 Soil Survey Data based on *Subset 2* (with background samples).

ANALYSIS

From Figure 4.3.13 we do not see any clustering of soil samples based on their location in Circle I, Circle II or Circle III. This indicates lack of any similarity between the sites based on their location in a circle. Moreover, although as in Goal 2a the clustering based on the sample type is not vey predominant; we do observe some pattern based on the soil samples of Type A(top side of Map) vs. the soil samples of Type D&E(bottom side of the map). We can also see that the background samples are more clustered with the sample Type D&E rather than the samples of Type A. Thus, we can say that based on the selected subset of elements the background samples seem to be more related to the samples of Type D& E than the samples of Type A.

Similarly as in [**Goal 1c**] we analyze the SOM model built for each of the elements taken separately.

ELEMENT=Arsenic (As)

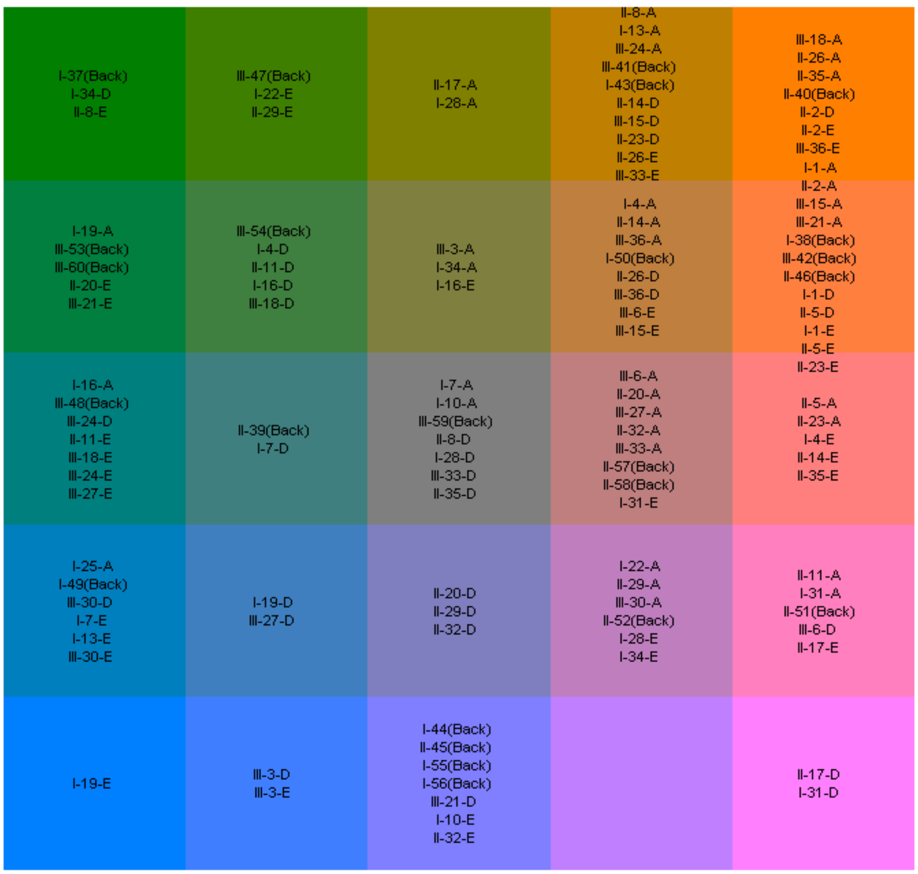


Figure 4.3.14. SOM model for Metro 2001 Soil Survey Data based on Arsenic (As).

ANALYSIS

From Figure 4.3.14 we do not see a good pattern of clustering for the background samples. Thus, we infer that there is no relationship between the background samples and the Metro 2001 samples.

ELEMENT=Chromium(Cr)

Figure 4.3.15. SOM model for Metro 2001 Soil Survey Data based on Chromium (Cr).

ANALYSIS

From Figure 4.3.15 we do not see any pattern of clustering for the background based on Cr.

ELEMENT=Lead (Pb) (Rotated the SOM map for better view ability)

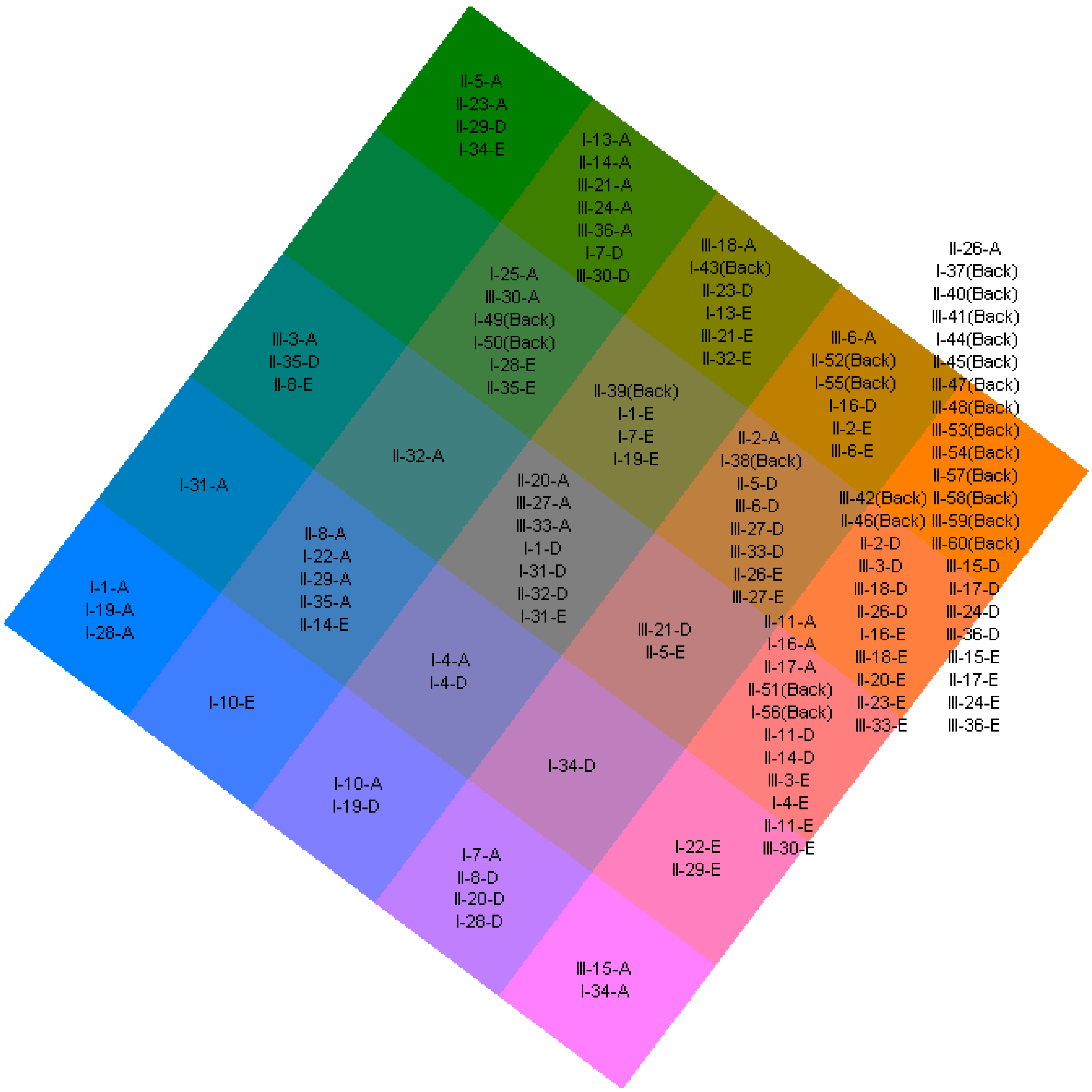


Figure 4.3.16. SOM model for Metro 2001 Soil Survey Data based on Lead (Pb).

ANALYSIS

From Figure 4.3.16 we can clearly see that the background samples are more related to the samples of Type D&E based on Lead (Pb). Further, we have checked that the samples of Type A have higher Pb concentration values than the samples of Type D&E and background samples.

ELEMENT=Nickel (Ni)

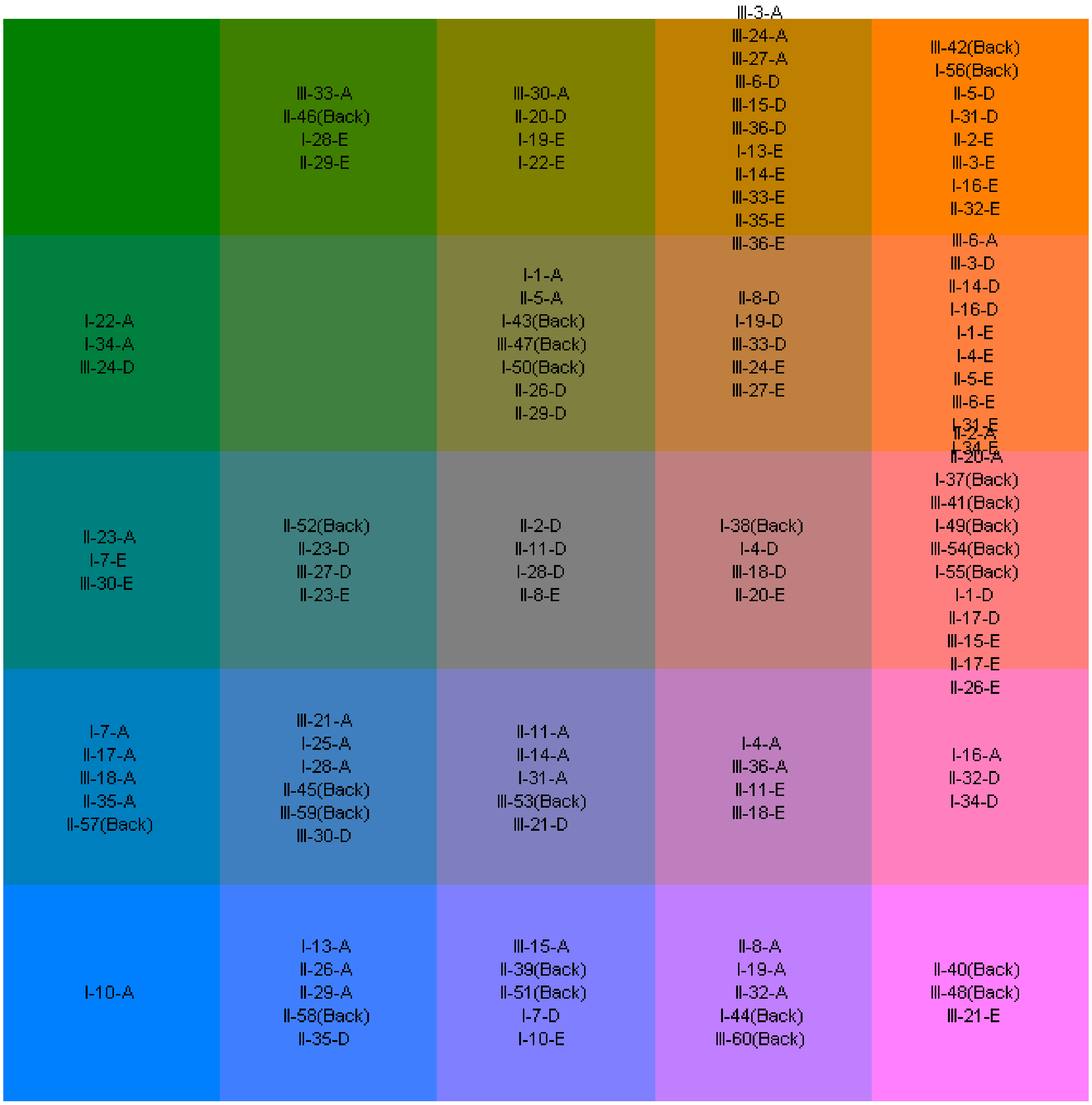


Figure 4.3.17. SOM model for Metro 2001 Soil Survey Data based on Nickel (Ni).

ANALYSIS

From Figure 4.3.17 we cannot see a good pattern of clustering for the background samples based on Nickel (Ni).

ELEMENT= Tungsten (W) (Rotated the SOM map for better view ability)

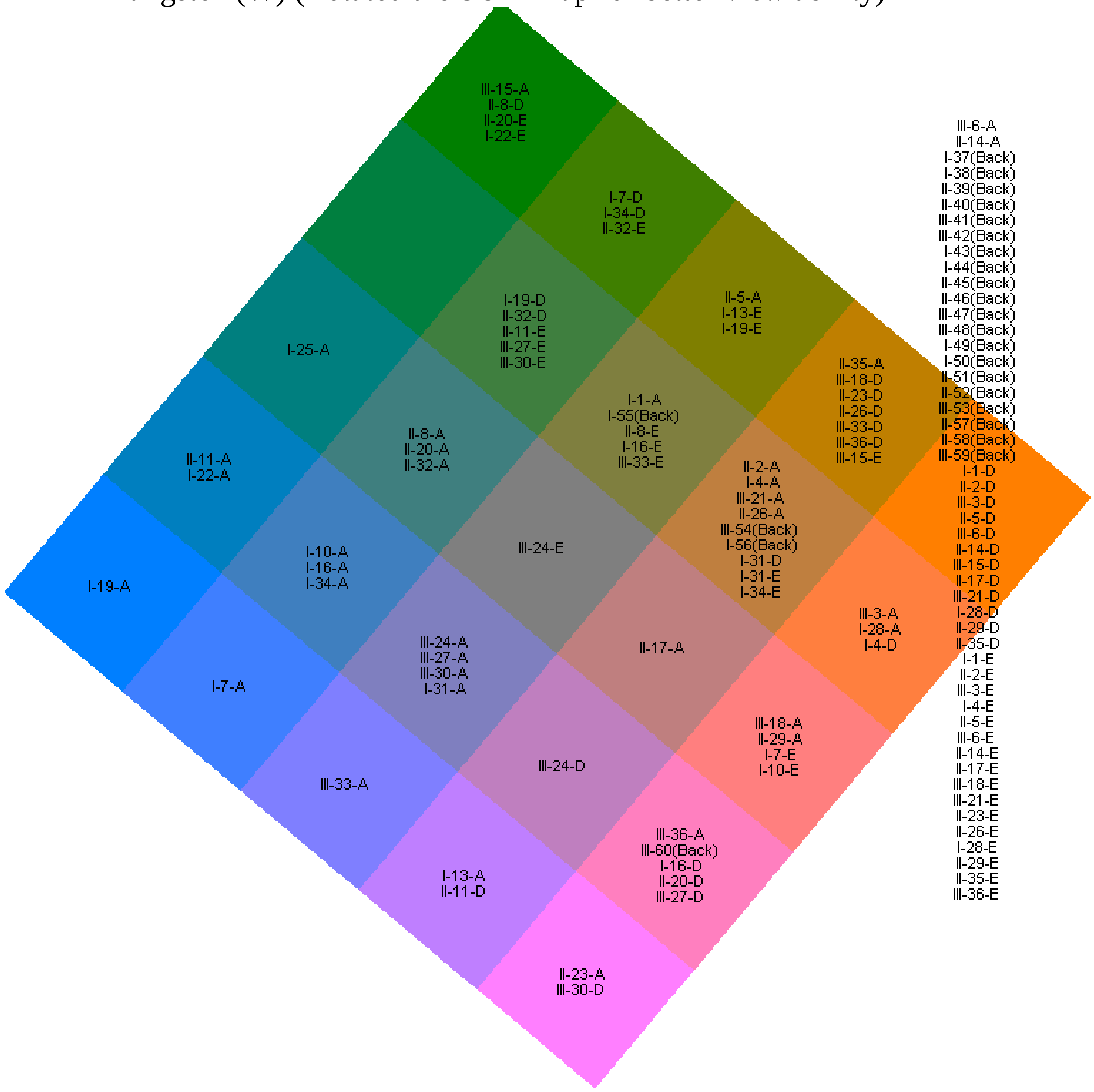


Figure 4.3.18. SOM model for Metro 2001 Soil Survey Data based on Tungsten (W).

ANALYSIS

From Figure 4.3.18 we can clearly see that the background samples are strongly related to the samples of Type D&E based on Tungsten (W). Further, we have checked that the samples of Type A have higher W concentration values than the samples of Type D&E and the background samples.

ELEMENT=Zinc (Zn)

Figure 4.3.19. SOM model for Metro 2001 Soil Survey Data based on Zinc (Zn).

ANALYSIS

From Figure 4.3.19 we can clearly see that the background samples are strongly related to the samples of Type D&E based on Zinc (Zn). We have checked except for the samples II-14-E, II-11-E which are clustered at the bottom left the samples of Type A have relatively higher Zn concentrations than the Type D&E samples and the background samples.

ELEMENT=Copper (Cu)

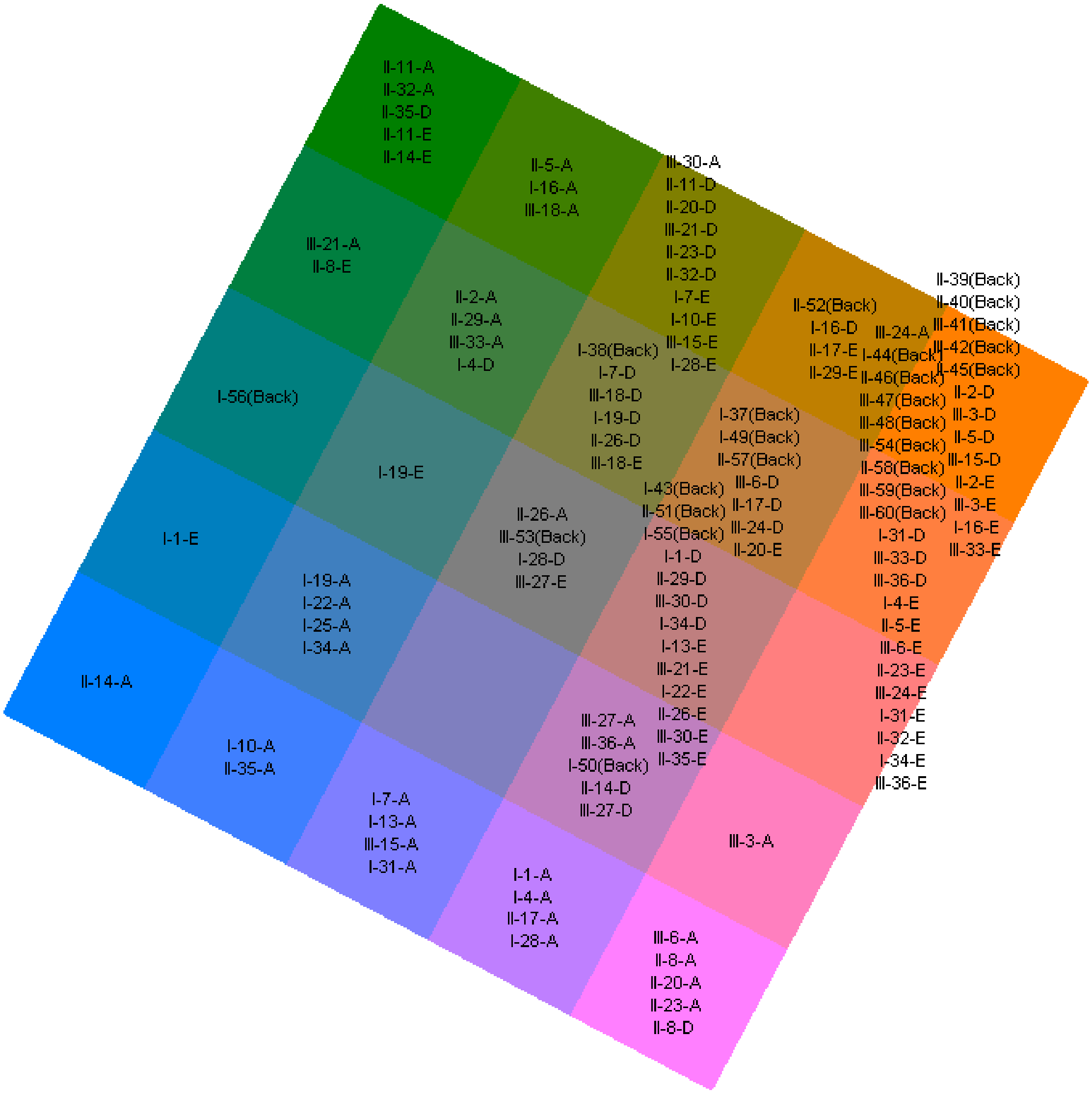


Figure 4.3.20. SOM model for Metro 2001 Soil Survey Data based on Copper (Cu).

ANALYSIS

From Figure 4.3.20 we can clearly see that the background samples are strongly related to the samples of Type D&E based on Copper (Cu). Further, we have checked that the samples of Type-A have higher Cu concentration values than the samples of Type D&E and the background samples.

4.3.2 Summary of the results obtained for Metro 2001 Soil Survey Data

1. Based on all elements, *Subset 1* and *Subset 2* the soil samples collected close to the road (Type A) have similar properties and the samples far away from the road (Type D&E) also have similar characteristics.

2. Based on all elements, *Subset 1* and *Subset 2* the background samples have similar characteristics to the samples far away from the road (Type D&E) than the samples collected near to the road (Type A).

3. Moreover we can also highlight the following pattern of clustering based on the concentration values of some of the regulatory metals of interest:

Arsenic:
- No clustering based on the location in the circle.
- No clustering based on the distance from the roads.
- No pattern of clustering for the background samples.

Chromium:
- No clustering based on the location in the circle.
- No clustering based on the distance from the roads.
- No pattern of clustering for the background samples.

Lead:
- No clustering based on the location in the circle.
- There is clustering based on the distance from the roads. (This is even more evident from Figure 4.3.16). Samples close to the road have higher Pb concentration.
- Background samples are clustered to the samples of type D & E which are collected far away from the road.

Nickel:
- No clustering based on the location in the circle.
- There is no clustering based on the distance from the roads.
- No pattern of clustering for the background samples.

Tungsten:
- No clustering based on the location in the circle.
- There is clustering based on the distance from the roads i.e. the soil samples collected near to the road have similar W concentration and the soil samples collected far away from road have similar W concentration. Samples close to the road have higher W concentration.
- Background samples are clustered to the samples which are far away from the road.

Zinc:
- No clustering based on the location in the circle.
- There is clustering based on the distance from the roads; i.e. the soil samples collected near to the road have similar Zn concentration and the soil samples collected far away from road have similar Zn concentration. Samples close to the road have higher Zn concentration.
- Background samples are clustered to the samples which are far away from the road.

Copper:

- No clustering based on the location in the circle.
- There is clustering based on the distance from the roads .Samples close to the roads have higher Cu concentration values.
- Background samples are clustered to the samples which are far away from the road.

4.3.3 Results for SOM modeling pertaining to MN Statewide 2003 Soil Survey Data.

[**Goal 3a**] To estimate how the statewide samples are related based on 5 geomorphologic codes based on all elements.

DATASET USED: MN Statewide 2003 Soil Survey Data.

NAMING CONVENTION: In order to understand the results better we introduce a naming convention for the different soil samples of the data. This naming convention is described below.

Basically in order to understand the pattern of clustering based on the geomorphologic codes we introduce a map of reduced geomorphologic codes as shown in Figure 4.3.21. (This map has been provided by Dr. Robert Edstrom).Basically now we divide the map into 5 geomorphologic categories (as shown in Table 4.3.1).

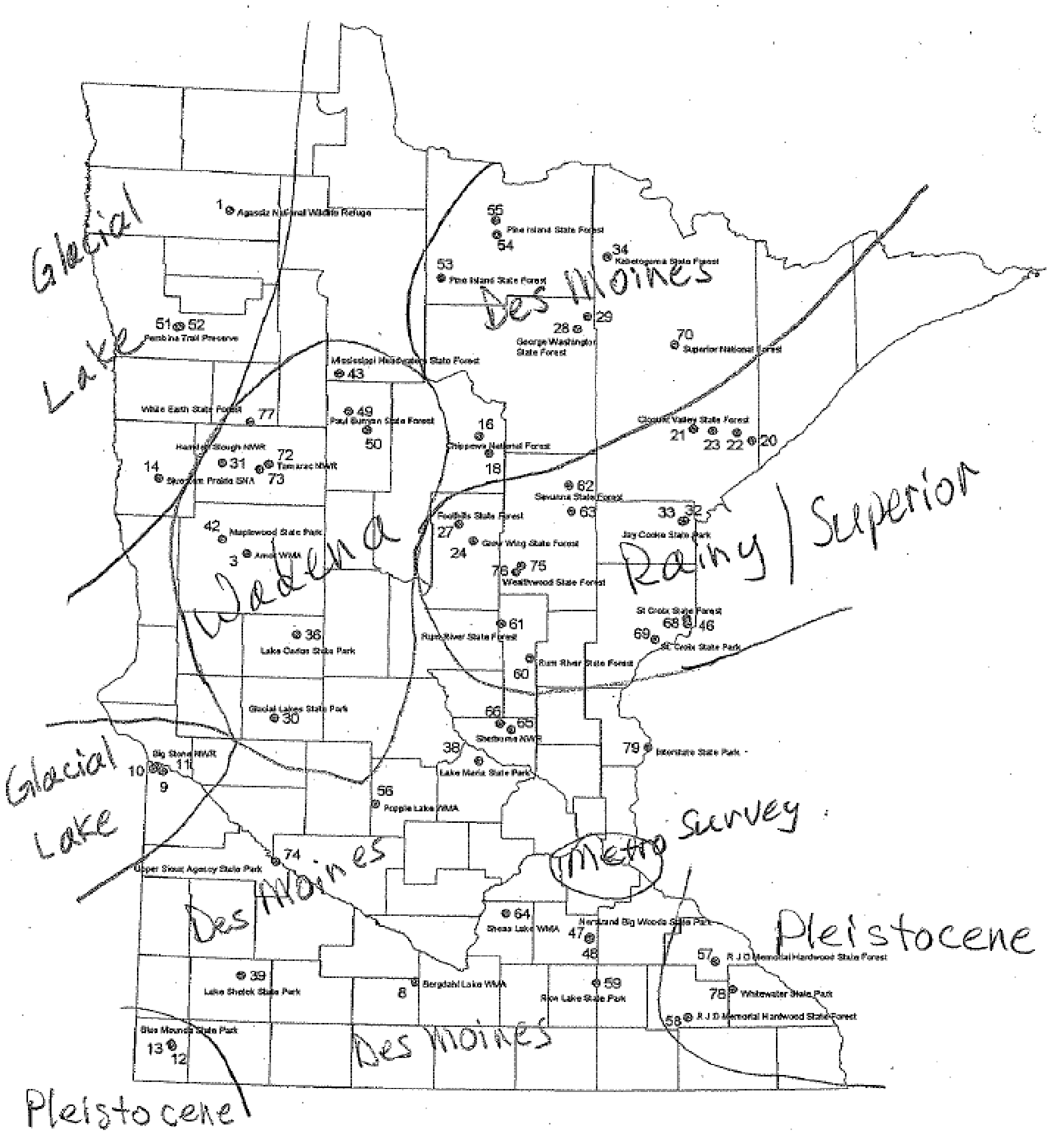


Figure 4.3.21. Map with reduced geomorphologic units (provided by Dr. Robert Edstrom).

Table 4.3.1 Geomorphologic codes for the reduced geological units shown in Figure 4.3.21

GEOMORPHOLOGIC CODE	GEOMORPHOLOGIC CATEGORY
1	Des Moines
2	Glacial Lakes
3	Rainy Superior
4	Wadena
5	Pleistocene

The naming scheme that we used to identify each sample is:
Site ID (GEOMORPHOLOGIC CODE)

For example:
A sample with its tag as **63(3)** encodes

Site ID (for the sample) = 63.
Geomorphologic category= Rainy Superior (3).

Moreover when we consider the transects of each sample we edit the tag to indicate the transect number as **Site ID-TRANSECT NUMBER (GEOMORPHOLOGIC CODE)**

For example:
A sample with its tag as **63-1(3)** encodes

Site ID (for the sample) = 63.
Transect number=1.
Geomorphologic category= Rainy Superior (3).

SOM MODELING
The SOM modeling results for this goal is shown below.

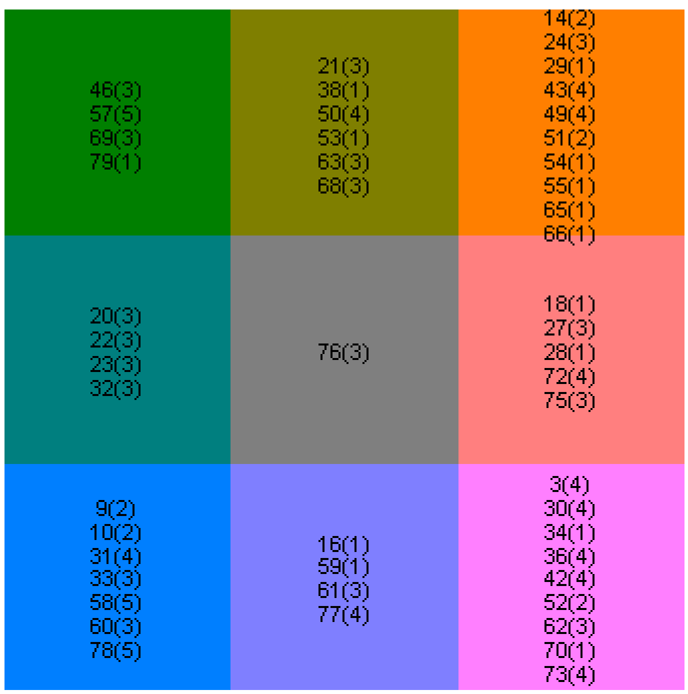


Figure 4.3.22. SOM model for MN Statewide 2003 Data based on all elements.

Note: For a better understanding we also provide the output of Multi-Dimensional Scaling (MDS) for this data. The goal of MDS is to produce a low-dimensional (typically 2D) representation of the inter-point distance information between multivariate samples (in our case, 45-dimensional samples). This method preserves the pair wise distance relationship between 45-

dimensional samples (sites), and represents them in a 2-Dimensional space. So the samples that are far/close from/to each other in the MDS representation are likely to be far/close apart in the original high-dimensional space. A brief description of the MDS method is provided in Appendix F.

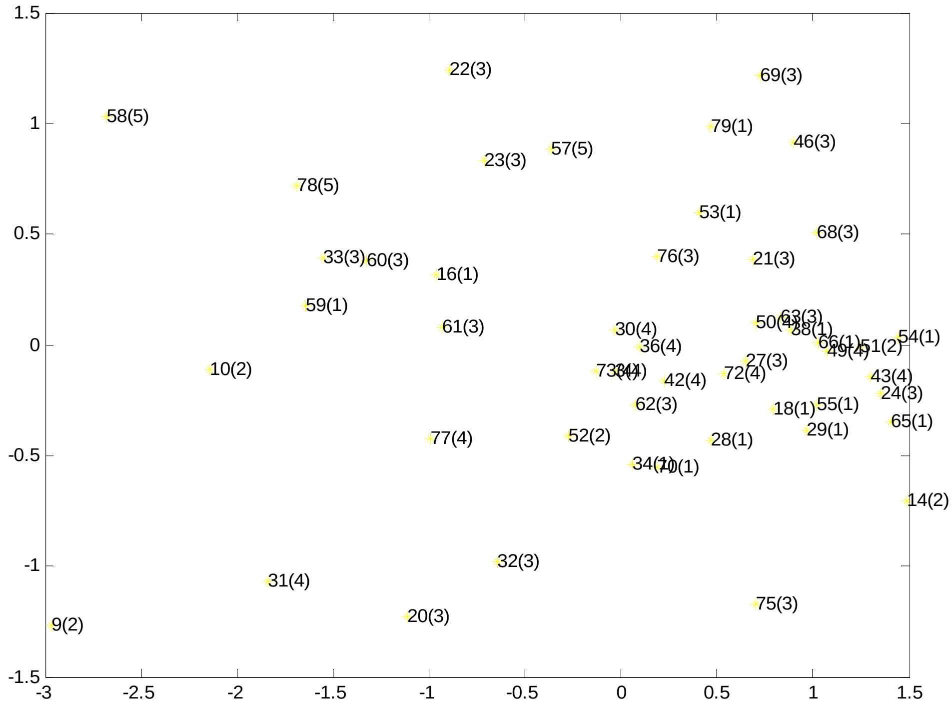


Figure 4.3.23. MDS output for MN Statewide 2003 Data based on all elements.

The Figure 4.3.22 and Figure 4.3.23 provide the SOM and MDS results for the MN Statewide 2003 Soil Survey Data based on the average of the 3-transects for each sites. We further provide the SOM model and the MDS output for this data without averaging the transects. This is useful for understanding the pattern of clustering because the number of samples now will be much more and any distinct clustering can be clearly captured. Note that the caption of the Figures would indicate if the results are obtained with/without averaging the transects.

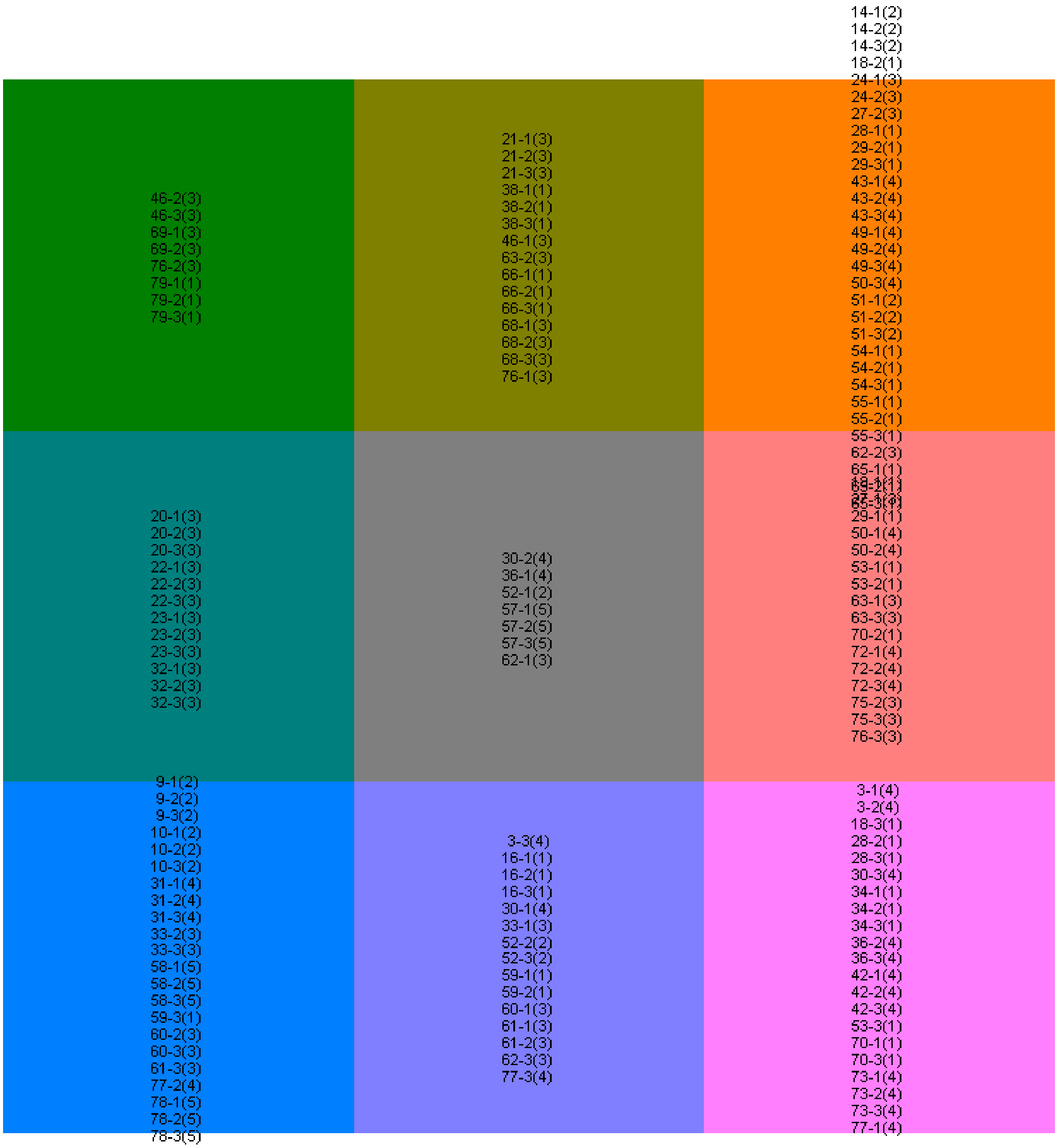


Figure 4.3.24. SOM model for MN Statewide 2003 Data based on all elements (without averaging the transects).

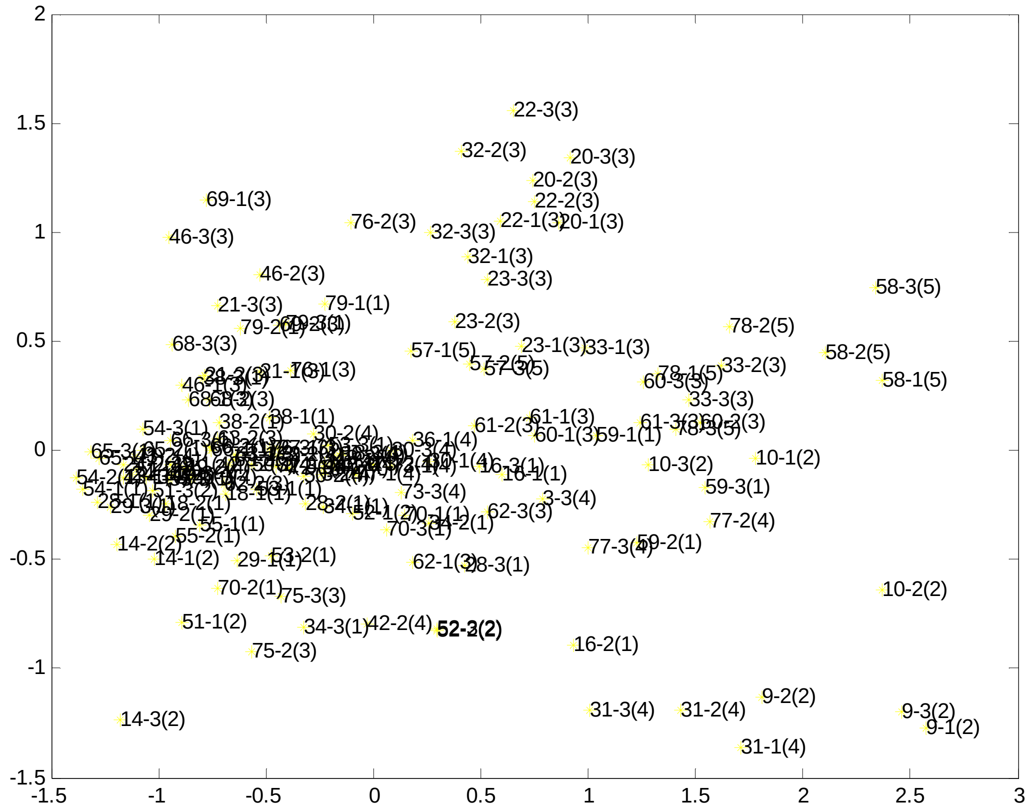


Figure 4.3.25. MDS output for MN Statewide 2003 Data based on all elements (without averaging the transects).

ANALYSIS

Although it is quite hard to perceive the pattern but based on the Figure 4.3.22, Figure 4.3.23, Figure 4.3.24 and Figure 4.3.25 we can observe some form of clustering of the soil samples based on the geomorphologic code 3 (Rainy Superior) as against that of 1(Des Moines) & 4(Wadena). We do not see any distinct pattern of clustering for the soil samples belonging to the geomorphologic codes 2(Glacial Lakes) and 4(Pleistocene).

[**Goal 3b**] To estimate how the statewide samples are related based on 5 geomorphologic codes based on the *Subset 1*.

DATASET USED: MN Statewide 2003 Soil Survey Data. However, in this case we use only the elements present in *Subset 1*.

NAMING CONVENTION: Same as in Goal 3a.

SOM MODELING
The SOM modeling results for this goal is shown below.

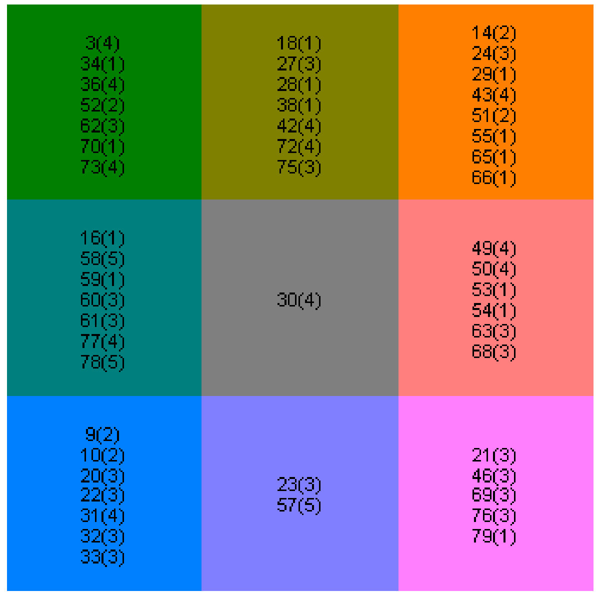


Figure 4.3.26. SOM model for MN Statewide 2003 Data based on *Subset 1*.

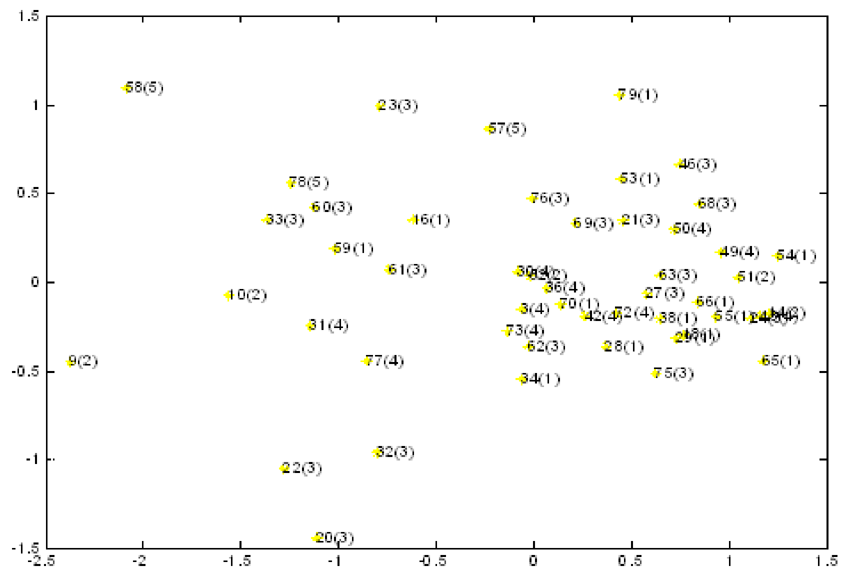


Figure 4.3.27. MDS output for MN Statewide 2003 Data based on *Subset 1*.

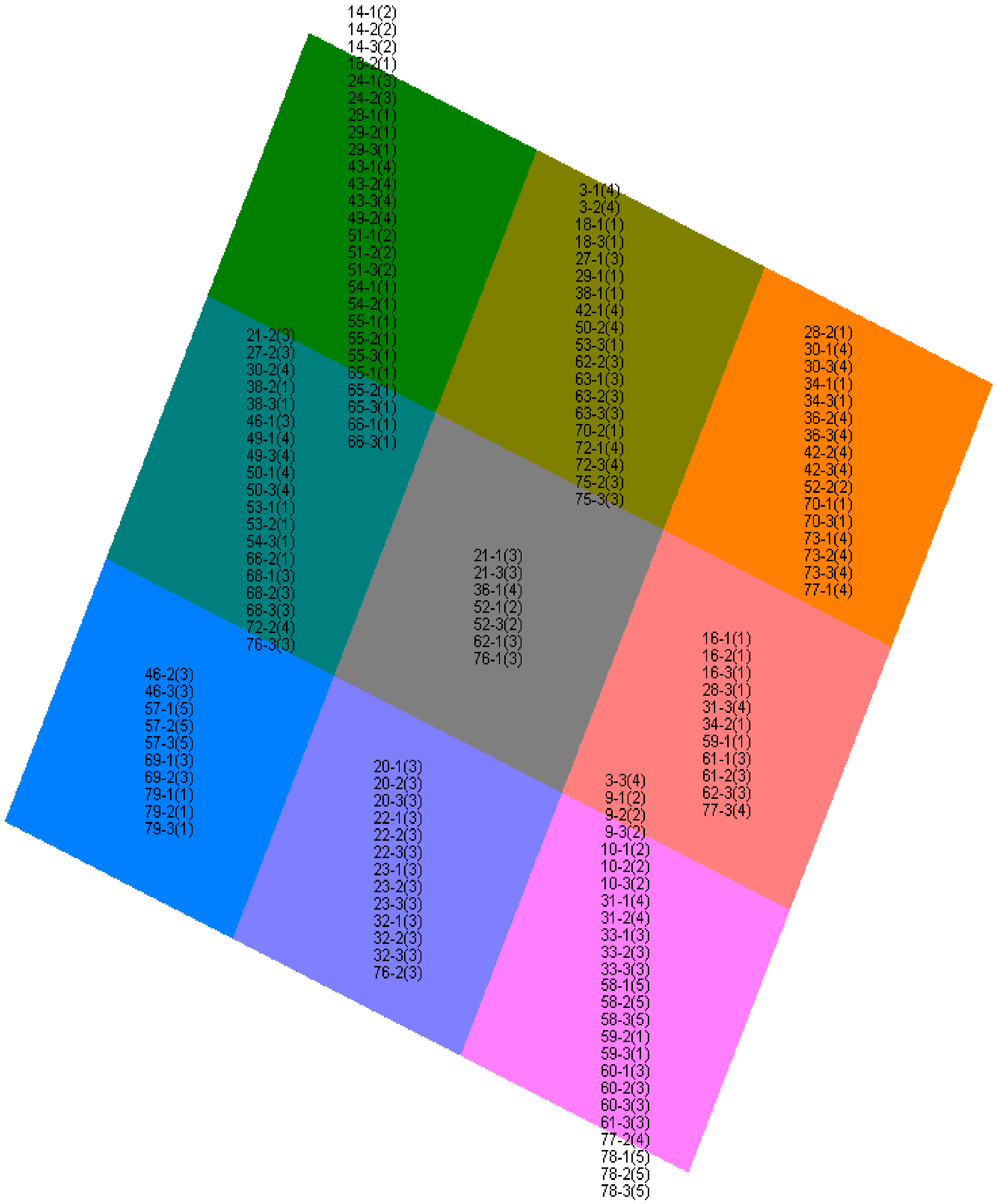


Figure 4.3.28. SOM model for MN Statewide 2003 Data based on *Subset 1* (without averaging the transects).

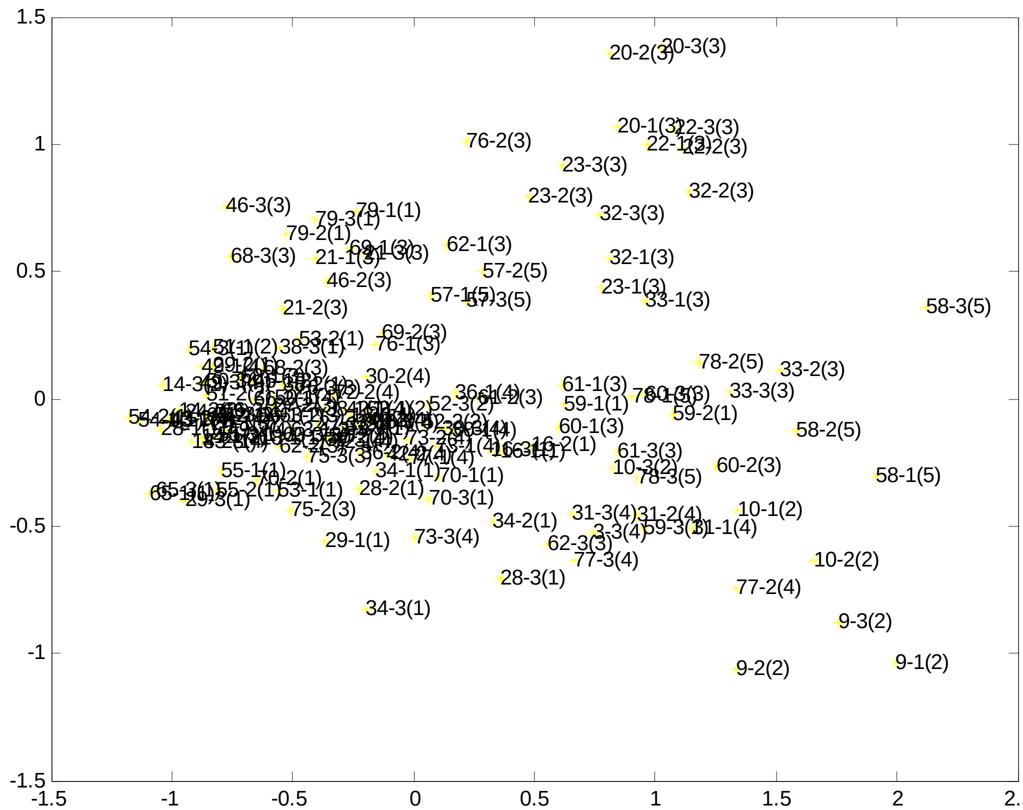


Figure 4.3.29. MDS output for MN Statewide 2003 Data based on *Subset 1* (without averaging the transects).

ANALYSIS

Results in this section are almost similar to that in Goal 3a. From Figure 4.3.26, Figure 4.3.27, Figure 4.3.28 and Figure 4.3.29 we can see some form of clustering of the soil samples based on the geomorphologic code 3 (Rainy Superior) as against that of 1(Des Moines) & 4(Wadena). We do not see any distinct pattern of clustering for the soil samples belonging to the geomorphologic codes 2(Glacial Lakes) and 4(Pleistocene). However for this case the pattern is not as prevalent as was seen for Goal 3a.

[**Goal 3c**] To estimate how the statewide samples are related based on 5 geomorphologic codes based on the *Subset 2*.

DATASET USED: MN Statewide 2003 Soil Survey Data. However in this case we use only the elements present in *Subset 2* for the SOM modeling.

NAMING CONVENTION: Same as in Goal 3a.

SOM MODELING

The SOM modeling results for this goal is shown below.

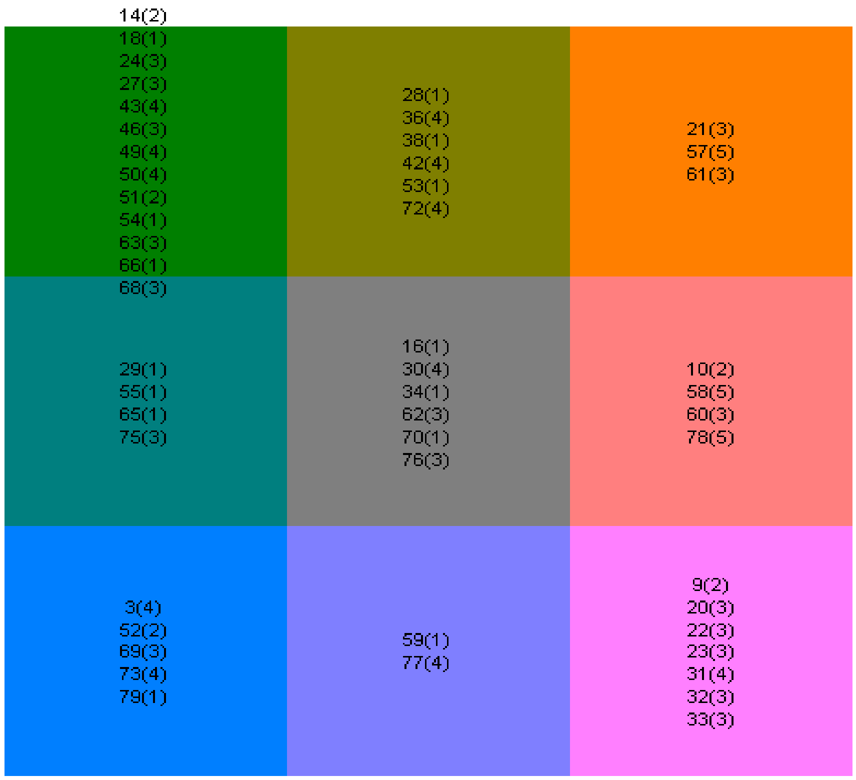


Figure 4.3.30. SOM model for MN Statewide 2003 Data based on *Subset 2*.

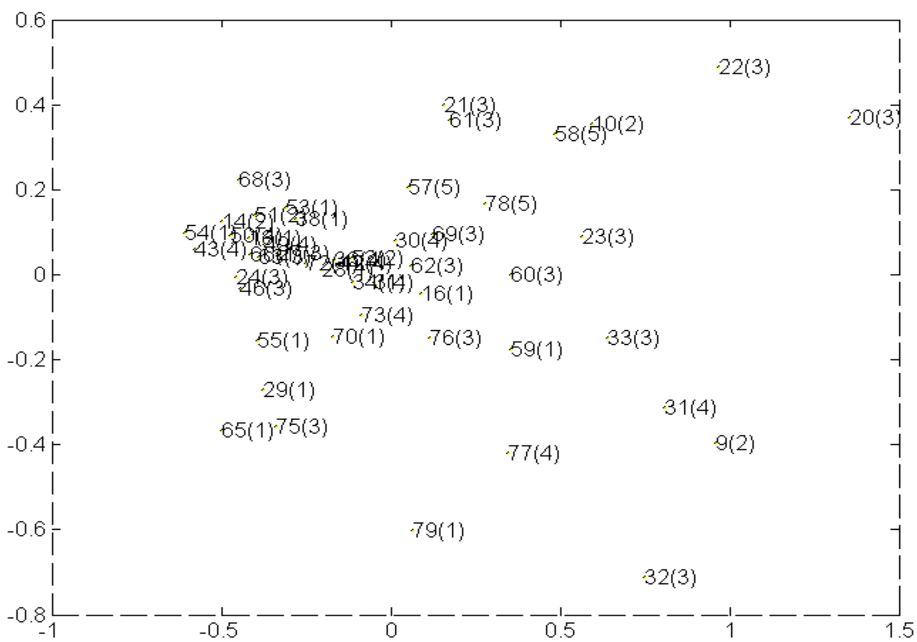


Figure 4.3.31. MDS output for MN Statewide 2003 Data based on *Subset 2*.

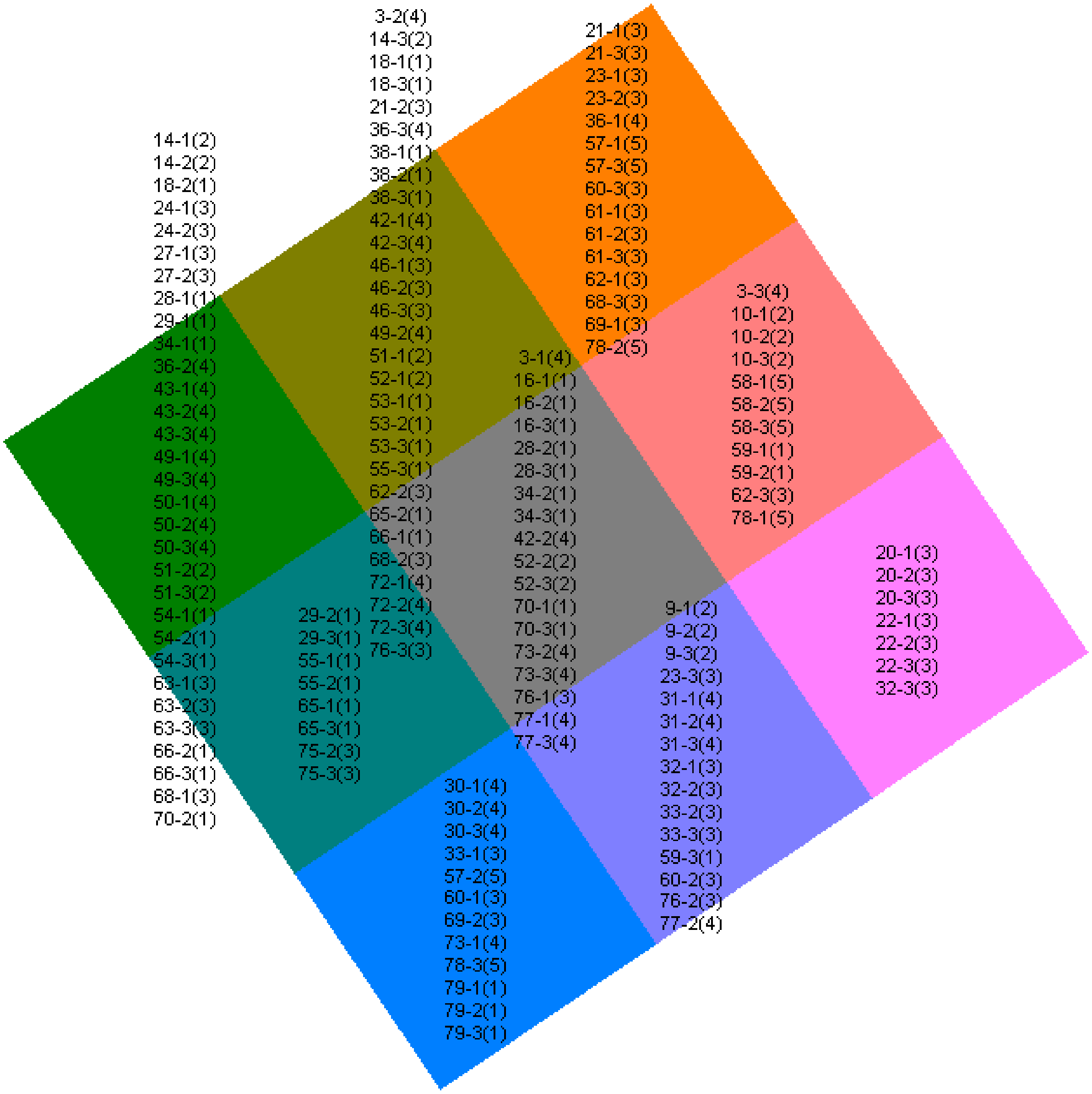


Figure 4.3.32. SOM model for MN Statewide 2003 Data based on *Subset 2* (without averaging the transects).

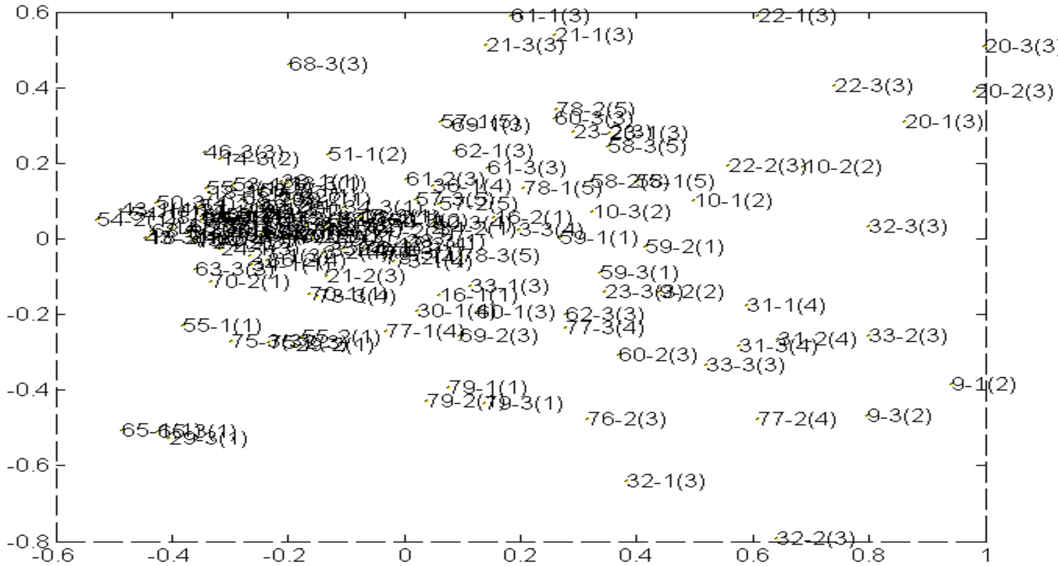


Figure 4.3.33. MDS output for MN Statewide 2003 Data based on *Subset 2* (without averaging the transects).

ANALYSIS

Unlike the results for Goal 3a and Goal 3b based on the Figure 4.3.26, Figure 4.3.27, Figure 4.3.28 and Figure 4.3.29 we cannot see a very good pattern of clustering of the soil samples based on the geomorphologic codes.

Similarly as before we analyze the SOM model built for each of the elements taken separately.

ELEMENT= Arsenic (As)

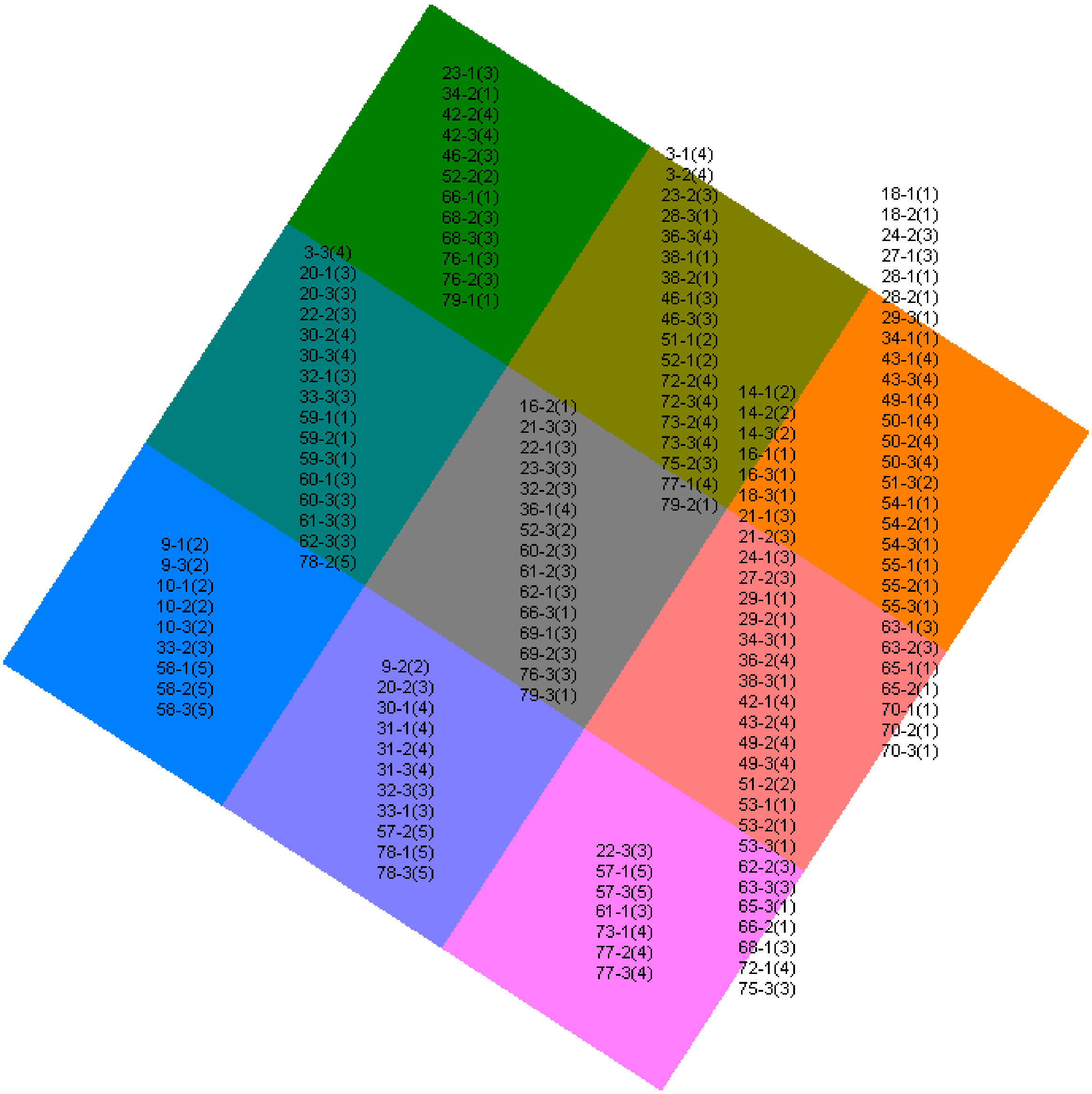


Figure 4.3.34. SOM model for MN Statewide 2003 Data based on Arsenic (As) (without averaging the transects).

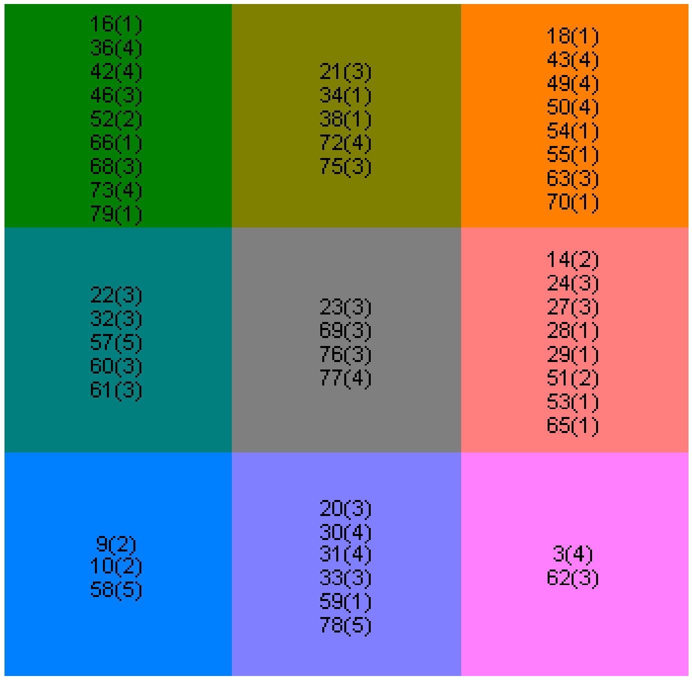


Figure 4.3.35. SOM model for MN Statewide 2003 Data based on Arsenic (As).

ANALYSIS

Based on Figure 4.3.34 and Figure 4.3.35 we do not observe any pattern of clustering of the samples based on the concentration values of Arsenic (As).

ELEMENT=Chromium (Cr)

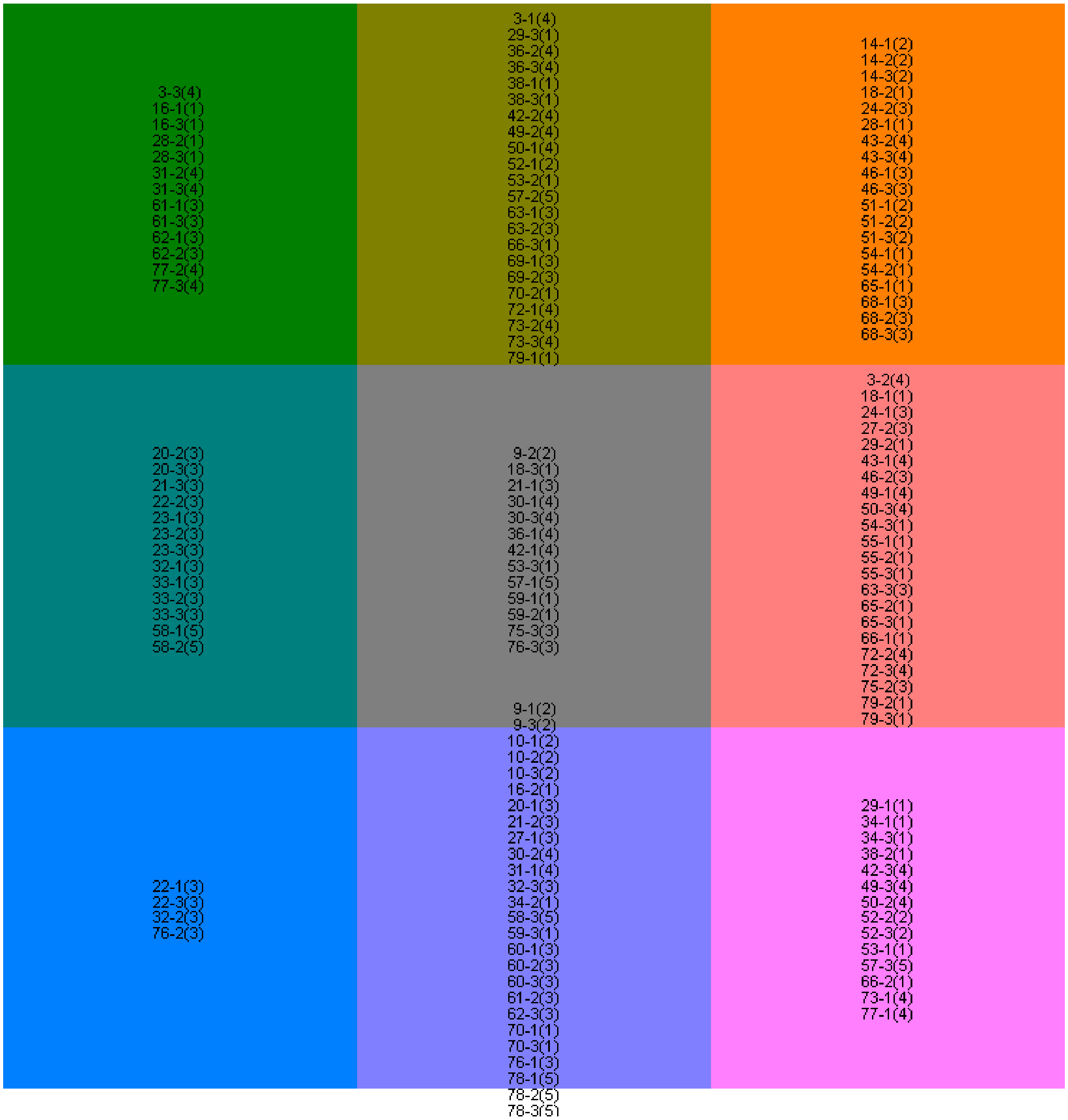


Figure 4.3.36. SOM model for MN Statewide 2003 Data based on Chromium (Cr) (without averaging the transects).

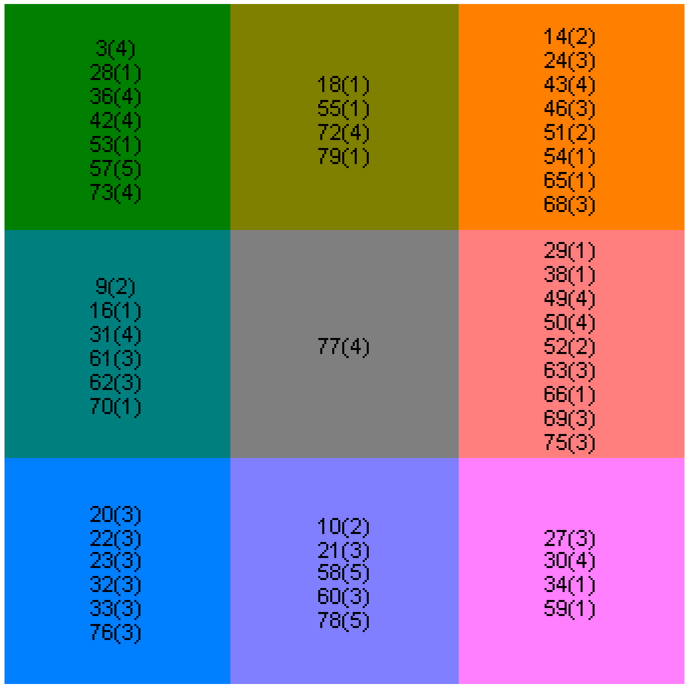


Figure 4.3.37. SOM model for MN Statewide 2003 Data based on Chromium (Cr).

ANALYSIS

Based on Figure 4.3.36 and Figure 4.3.37 we observe that the samples 20, 22,23,32,33,76,61,62 are all clustered together. All these sites have relatively higher Cr concentration values. On careful observation from Figure 4.3.21 we see that all these samples lie on the North-East of the state of Minnesota. This is somewhat in concordance with the bedrock geological map of MN (provided in Appendix B Figure B.1). Based on the map we observe that all these sites fall into the *Duluth Complex and associated rocks*. The possible reason for this is that the drilling in this terrain has indicated the presence of the mineral Chromium (Cr). This could be a possible reason why we observe such a pattern of clustering based on the element Chromium (Cr).

ELEMENT= Copper (Cu)

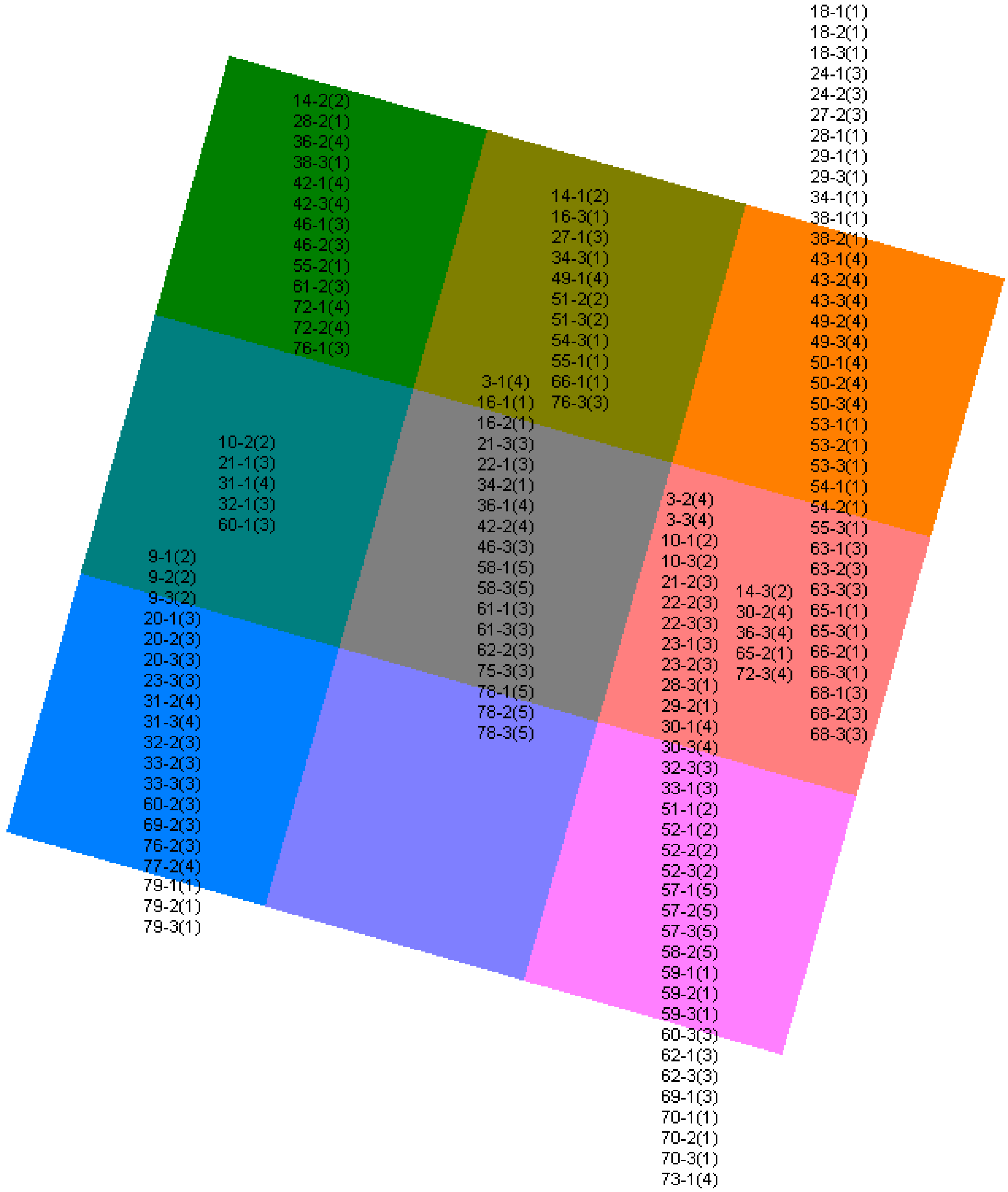


Figure 4.3.38. SOM model for MN Statewide 2003 Data based on Copper (Cu) (without averaging the transects).

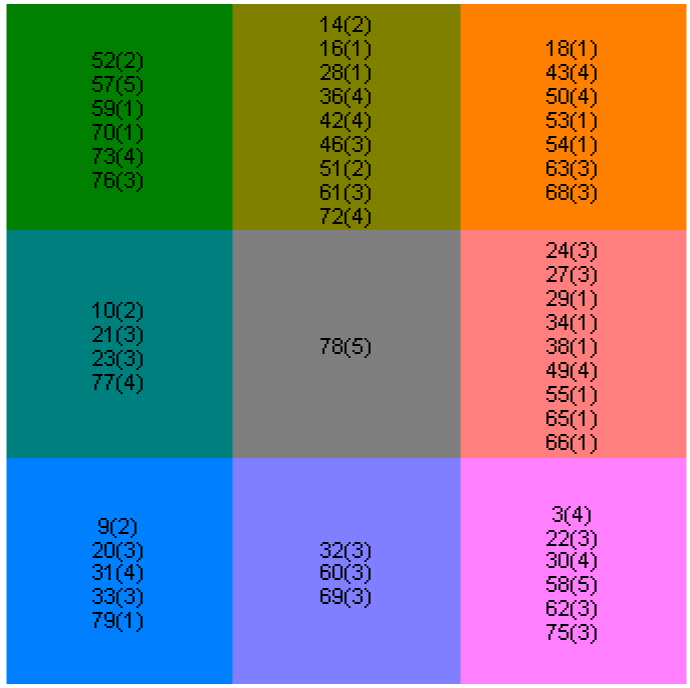


Figure 4.3.39. SOM model for MN Statewide 2003 Data based on Copper (Cu).

ANALYSIS

Based on Figure 4.3.38 and Figure 4.3.39 we observe a weak clustering between the samples that are on the north east of the state of MN. These are the samples that belong to the *Duluth Complex and associated rocks*. Thus it seems like that the samples belonging to this terrain might have similar values of Copper (Cu) concentration. However this is not very clear from this map. So we shall not try to make any conclusion from this.

ELEMENT= Lead (Pb)

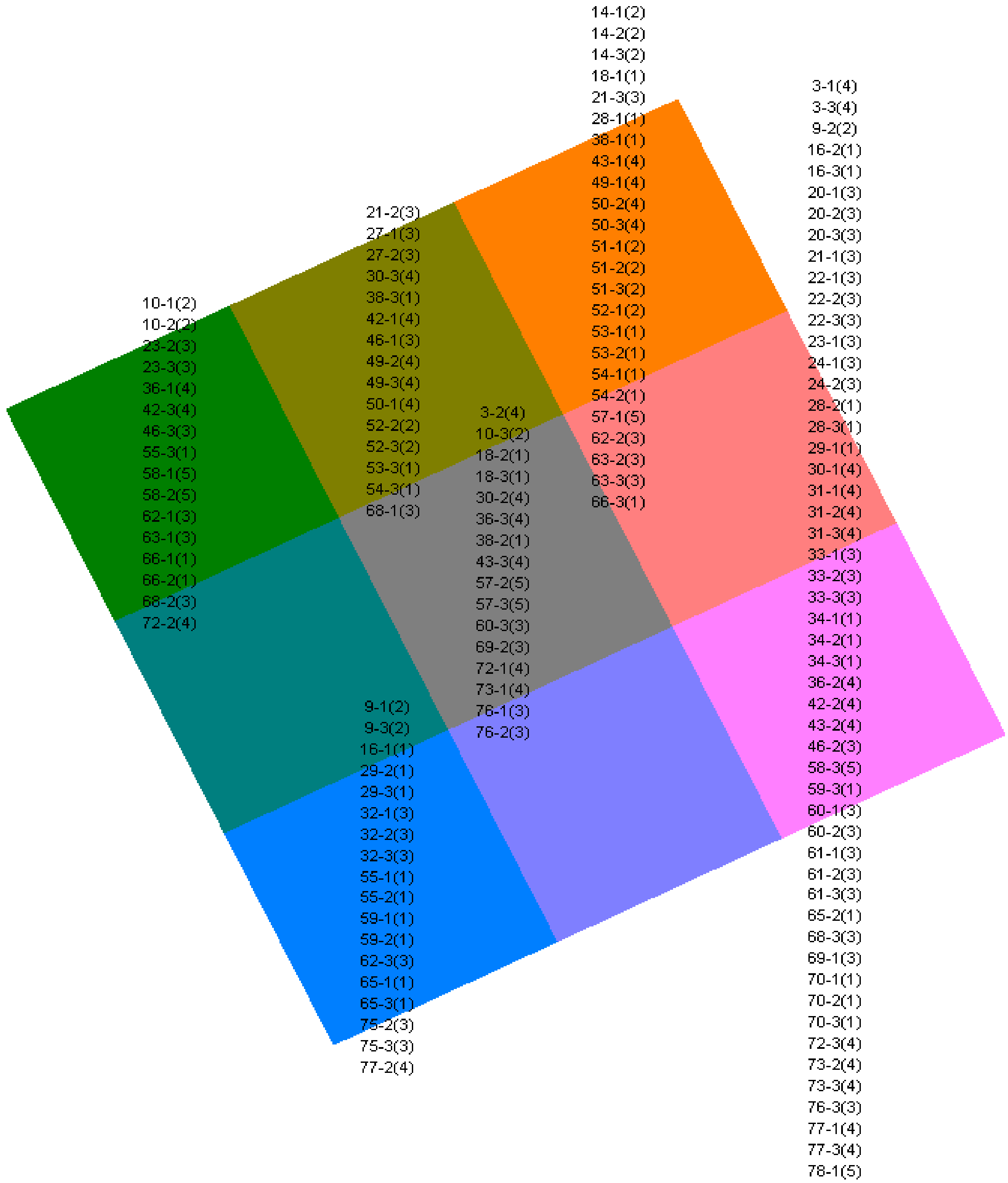


Figure 4.3.40. SOM model for MN Statewide 2003 Data based on Lead (Pb) (without averaging the transects).

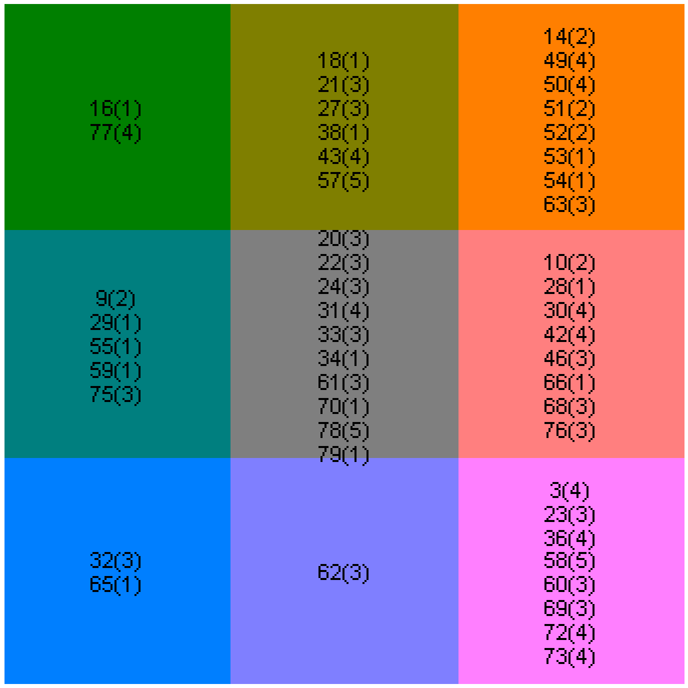


Figure 4.3.41. SOM model for MN Statewide 2003 Data based on Lead (Pb).

ANALYSIS

In this case we do not observe any pattern of clustering

Note: Based on the map B.1 we could expect some form of clustering based on the samples belonging to the Southern Minnesota's Archean migmatitic gneisses, younger sedimentary rocks, and batholithic granitic rocks and deformed volcanic, which are rich in lead and zinc. However, in our data set the number of samples from the south east of MN is quite less. So we cannot make any conclusion based on the results obtained.

ELEMENT= Nickel (Ni)

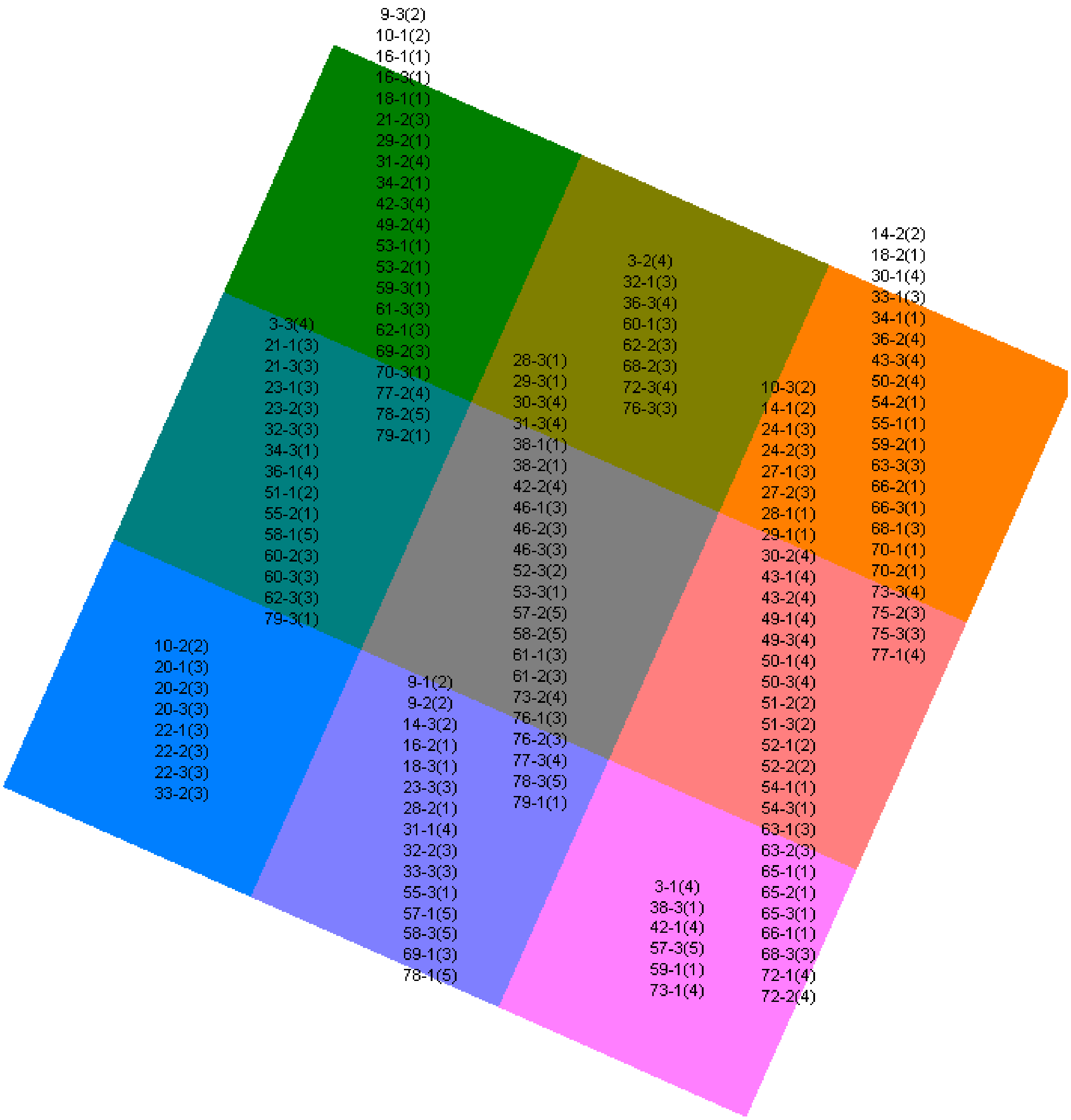


Figure 4.3.42. SOM model for MN Statewide 2003 Data based on Nickel (Ni) (without averaging the transects).

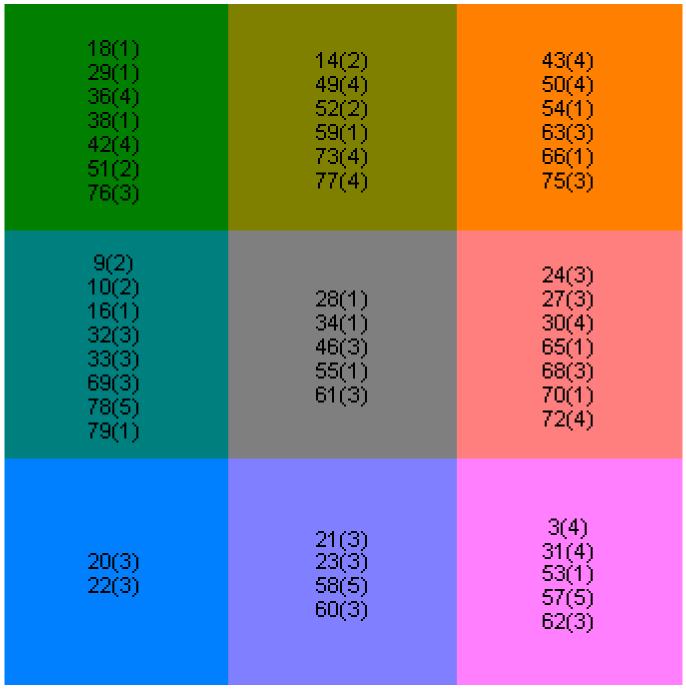


Figure 4.3.43. SOM model for MN Statewide 2003 Data based on Nickel (Ni).

ANALYSIS

Based on Figure 4.3.42 and Figure 4.3.43 we observe a clustering between the samples 20,21,22,23,32,33,79 that are on the north east of the state of MN. These samples have relatively higher Nickel (Ni) concentration. These are the samples that belong to the *Duluth Complex and associated rocks.* Thus it seems like that the samples belonging to this terrain might have similar values of Ni concentration.

ELEMENT=Zinc (Zn)

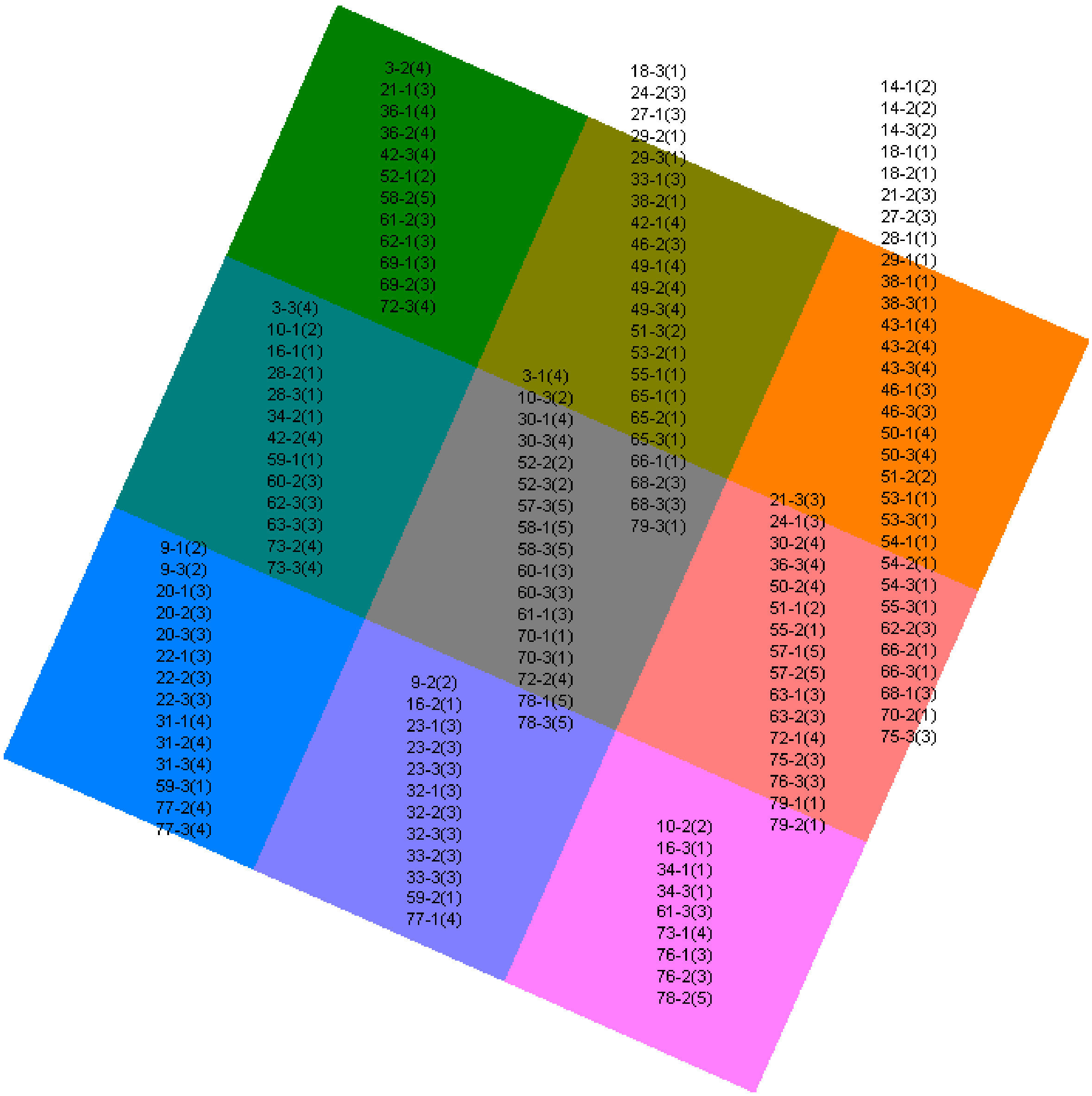


Figure 4.3.44. SOM model for MN Statewide 2003 Data based on Zinc (Zn) (without averaging the transects).

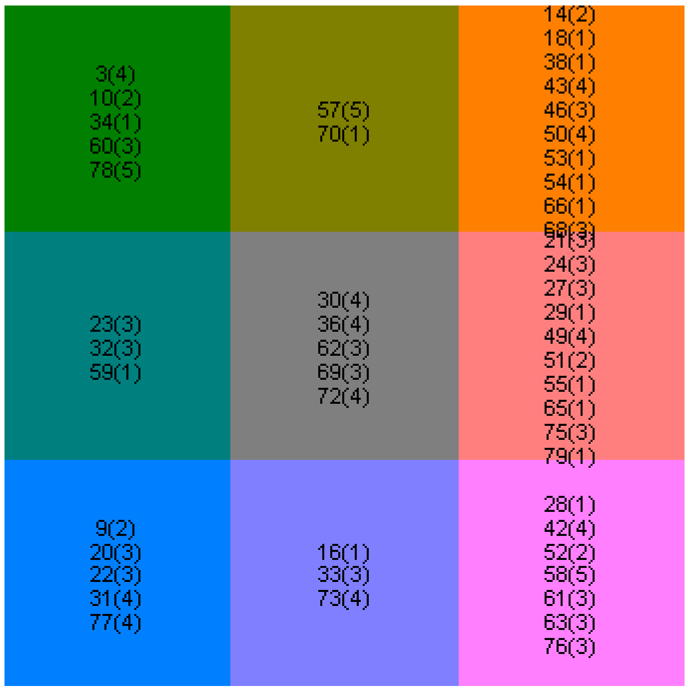


Figure 4.3.45. SOM model for MN Statewide 2003 Data based on Zinc (Zn).

ANALYSIS

In this case we do not observe any pattern of clustering.

Note: Based on the map B.1 we could expect some form of clustering based on the samples belonging to the Southern Minnesota's Archean migmatitic gneisses, younger sedimentary rocks, and batholithic granitic rocks and deformed volcanic, which are rich in lead and zinc. However, in our data set the number of samples from the south east of MN is quite less. So we cannot make any conclusion based on the results obtained.

ELEMENT= Tungsten (W)

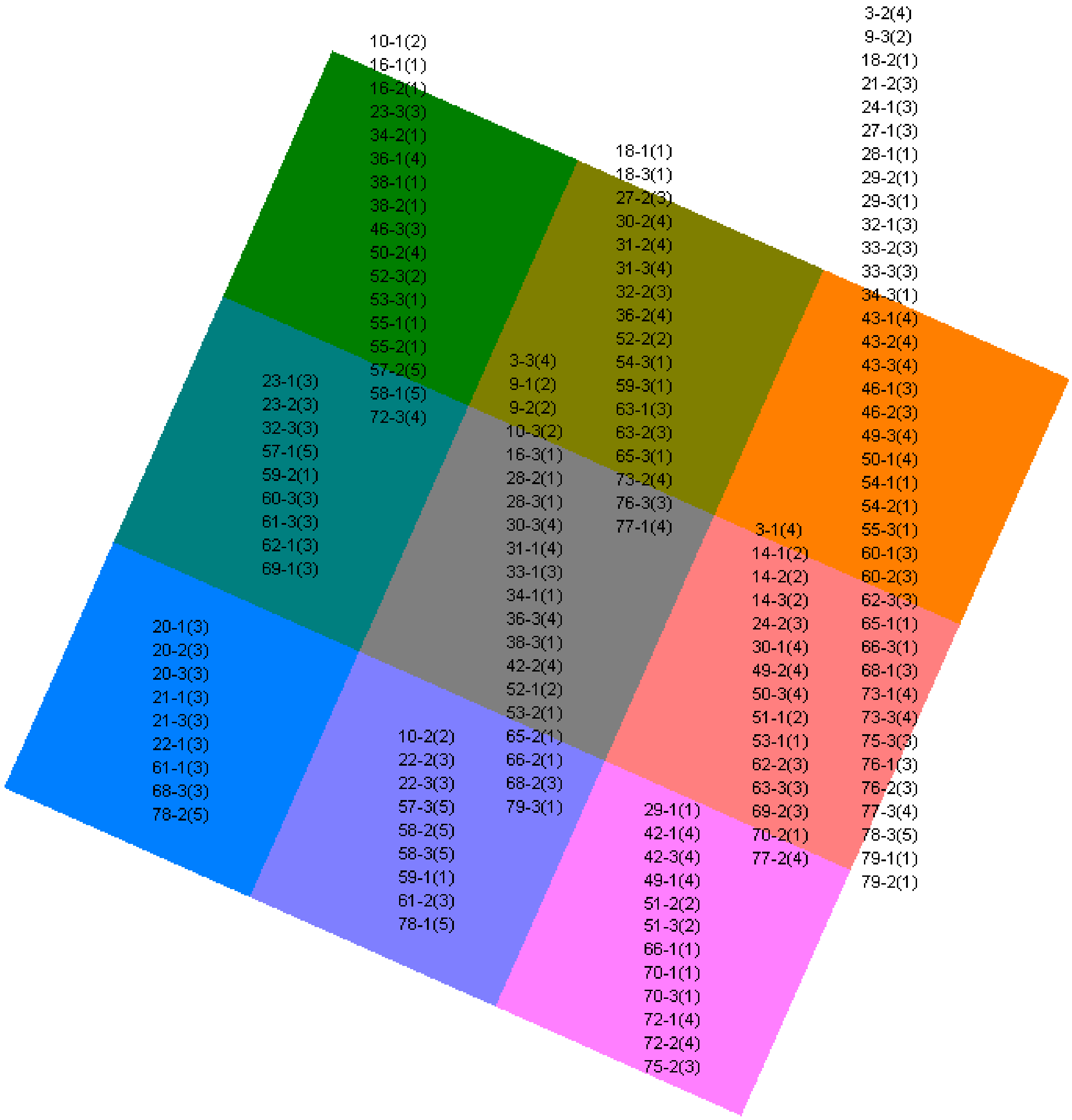


Figure 4.3.46. SOM model for MN Statewide 2003 Data based on Tungsten (W) (without averaging the transects).

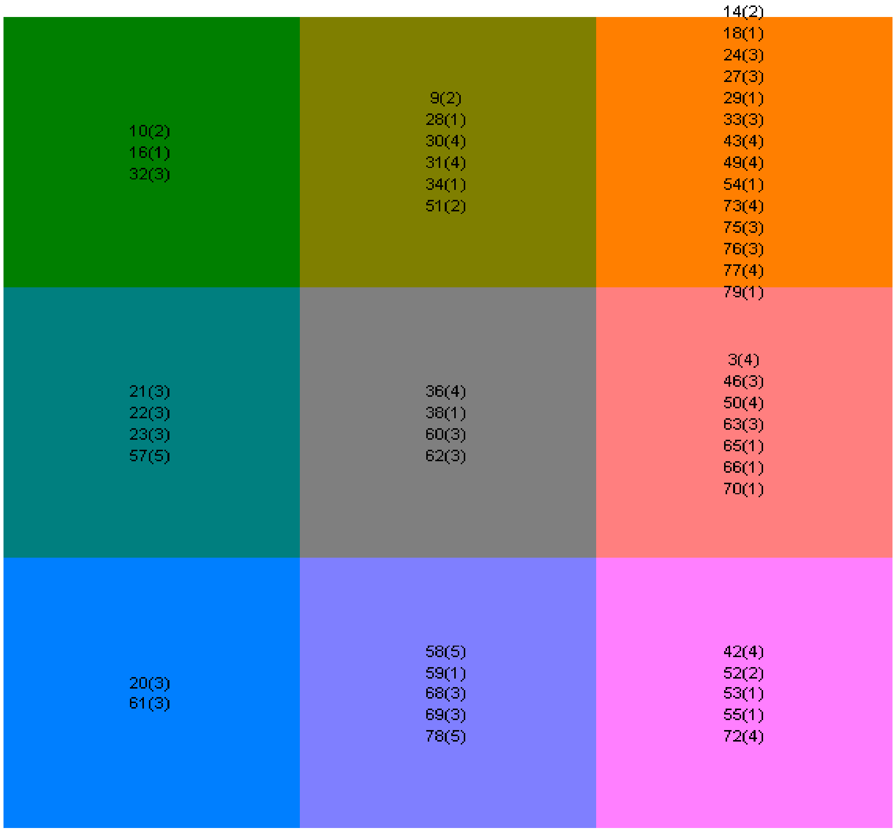


Figure 4.3.47. SOM model for MN Statewide 2003 Data based on Tungsten (W).

ANALYSIS
Based on Figure 4.3.46 and Figure 4.3.47 we observe a clustering between the samples 20,21,22,23,68,69,78 that are on the north east of the state of MN. These samples have higher Tungsten(W) concentration values. These are the samples that belong to the *Duluth Complex and associated rocks*. Thus it seems like that the samples belonging to this terrain might have higher values of W concentration.

[**Goal 3d**] To estimate how the statewide samples are related based on 5 geomorphologic codes based on the *Subset 3*.

DATASET USED: MN Statewide 2003 Soil Survey Data. In this case we use only the elements from the *Subset 3* for SOM modeling.

NAMING CONVENTION: Same as in Goal 3a.

SOM MODELING

The SOM modeling results for this goal is shown below.

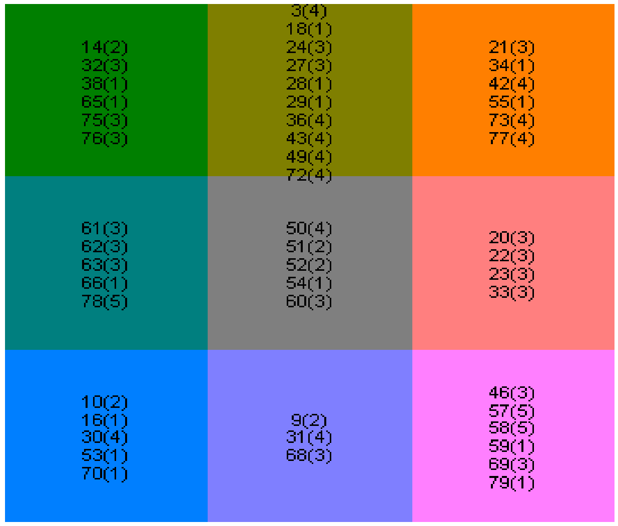


Figure 4.3.48. SOM model for MN Statewide 2003 Data based on *Subset 3*.

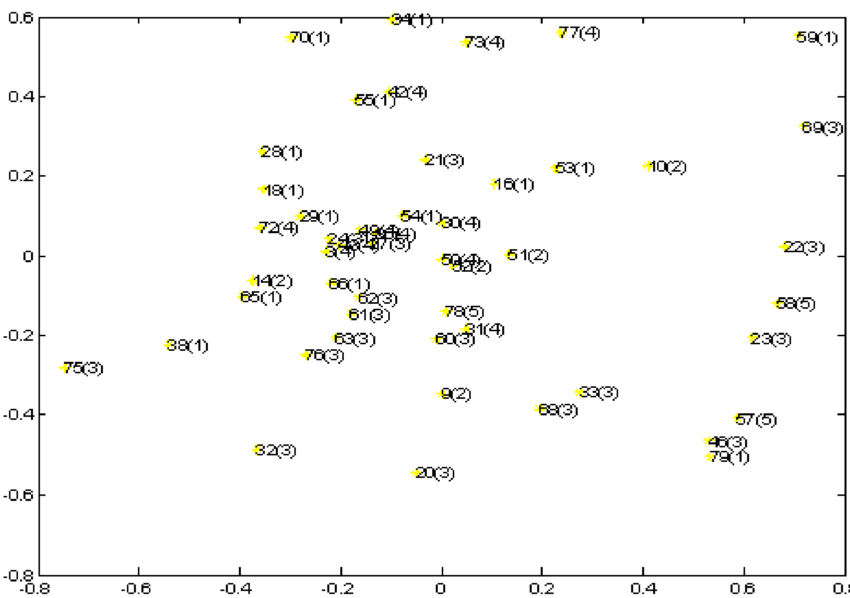


Figure 4.3.49. MDS output for MN Statewide 2003 Data based on *Subset 3*.

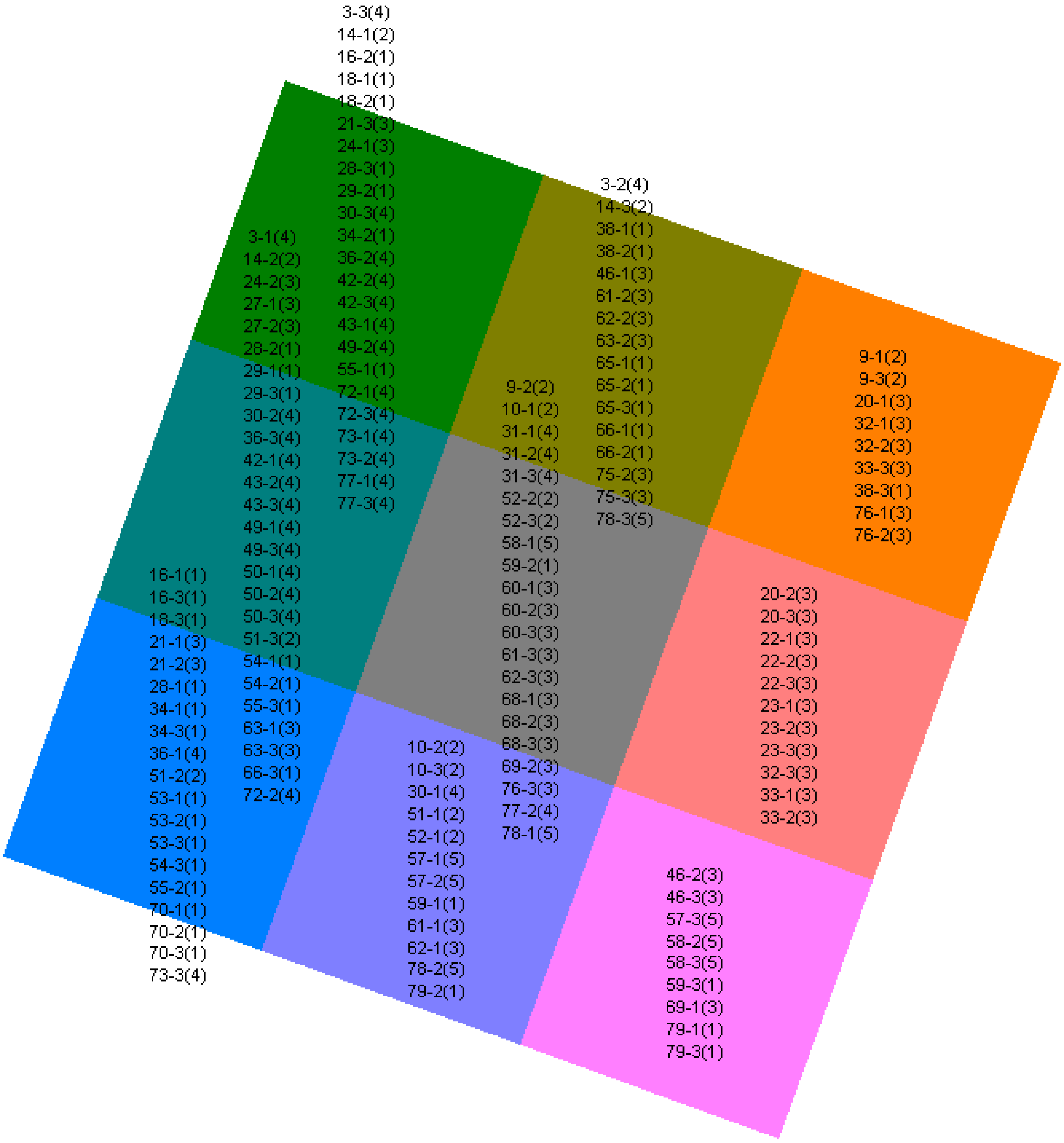


Figure 4.3.50. SOM model for MN Statewide 2003 Data based on *Subset 3* (without averaging the transects).

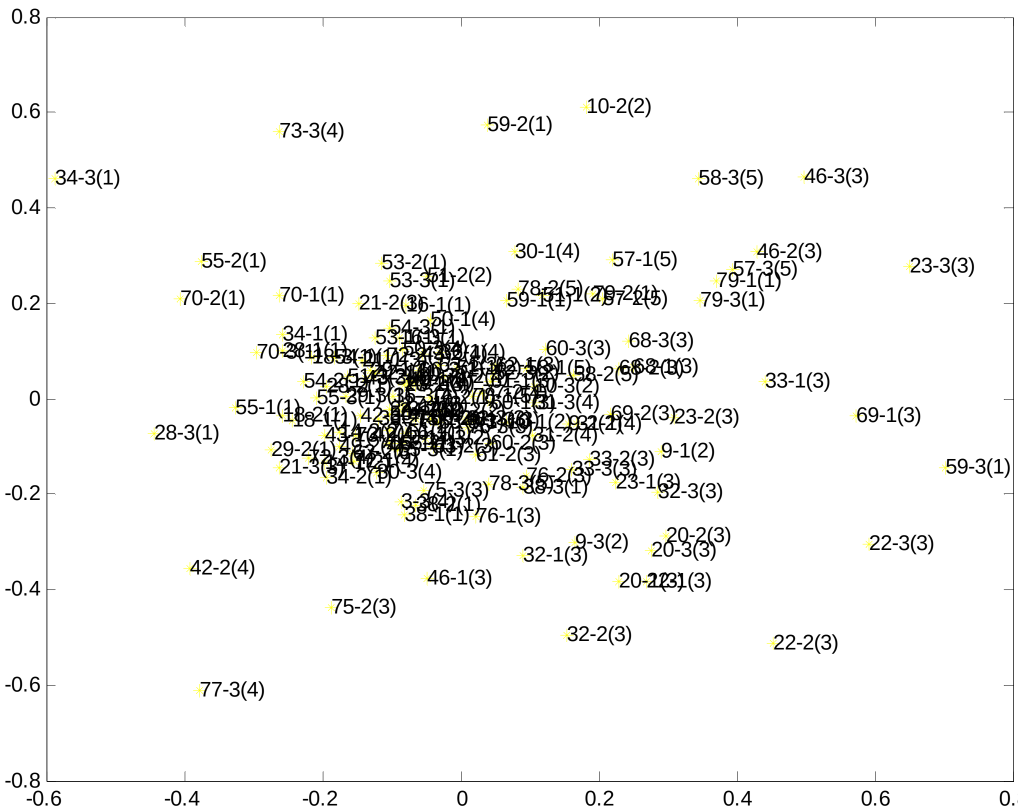


Figure 4.3.51. S output for MN Statewide 2003 Data based on *Subset 3* (without averaging the transects).

ANALYSIS
Figure 4.3.48, Figure 4.3.49, Figure 4.3.50, and Figure 4.3.51 suggest some pattern of clustering based on the geomorphologic codes. Specifically, from Figure 4.3.50 and Figure 4.3.51 it seems like samples belonging to geomorphologic codes 3(Rainy Superior) seem to cluster together and the samples belonging to 1(Des Moines) & 4(Wadena) seem to cluster together.

[**Goal 3e**] To estimate how the statewide samples are related based on 5 geomorphologic codes based on the *Subset 4*.

DATASET USED: *MN Statewide 2003 Soil Survey data*. In this case we use only the elements from the Subset 4 for SOM Modeling.

NAMING CONVENTION: Same as in Goal 3a

SOM MODELING

The SOM modeling results for this goal is shown below.

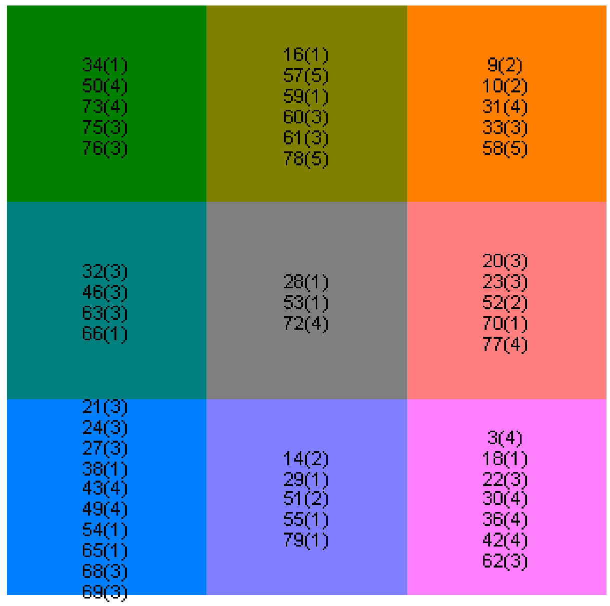


Figure 4.3.52. SOM Model for MN Statewide 2003 Data based on *Subset 4*.

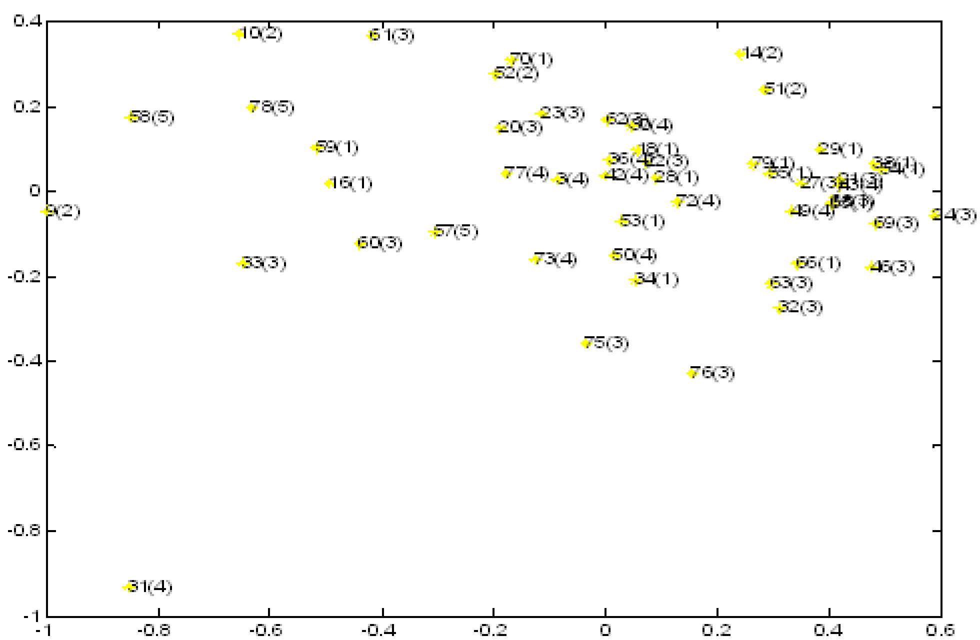


Figure 4.3.53. MDS output for MN Statewide 2003 Data based on *Subset 4*.

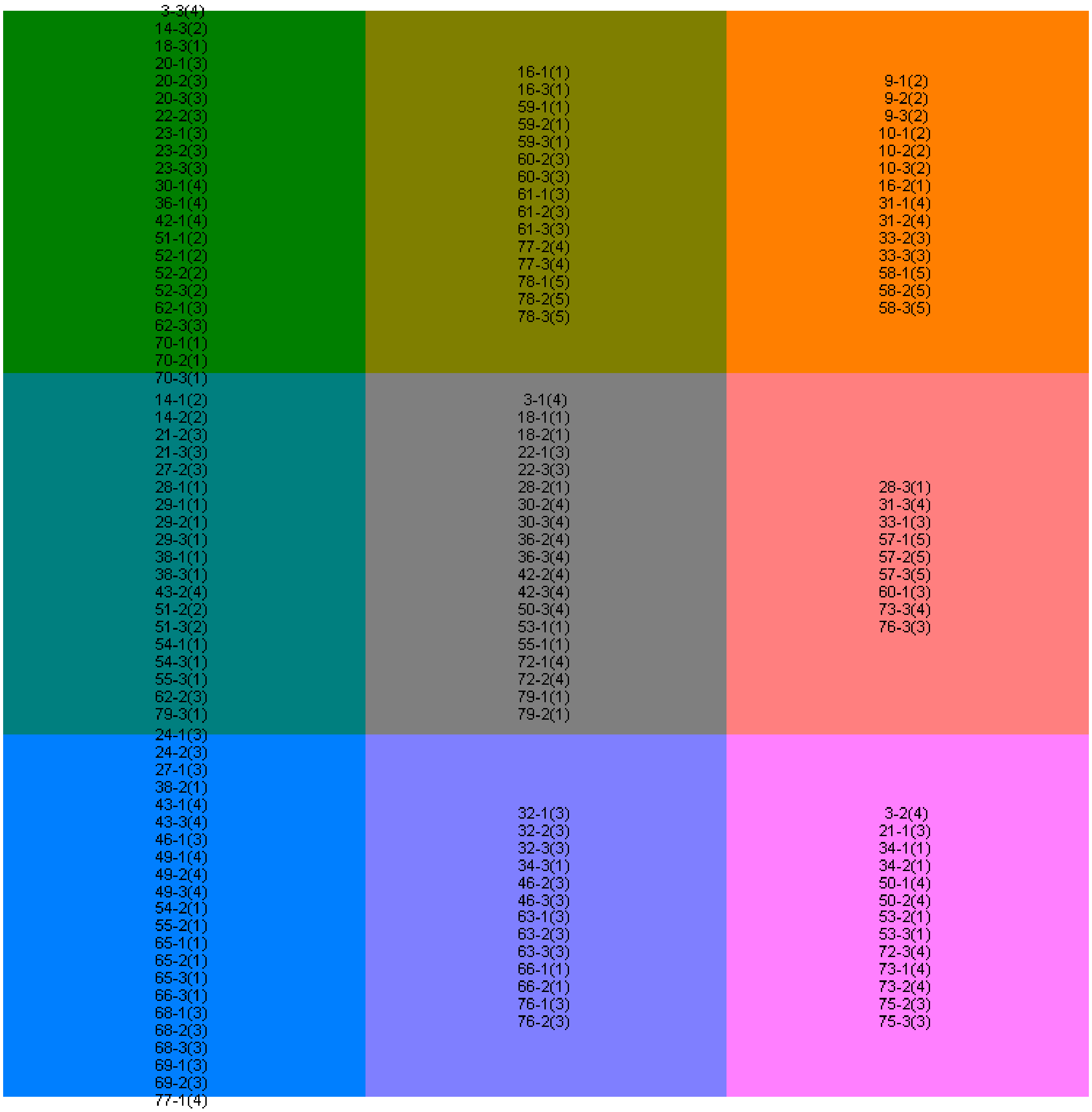


Figure 4.3.54. SOM map for MN Statewide 2003 Data based on *Subset 4* (without averaging the transects).

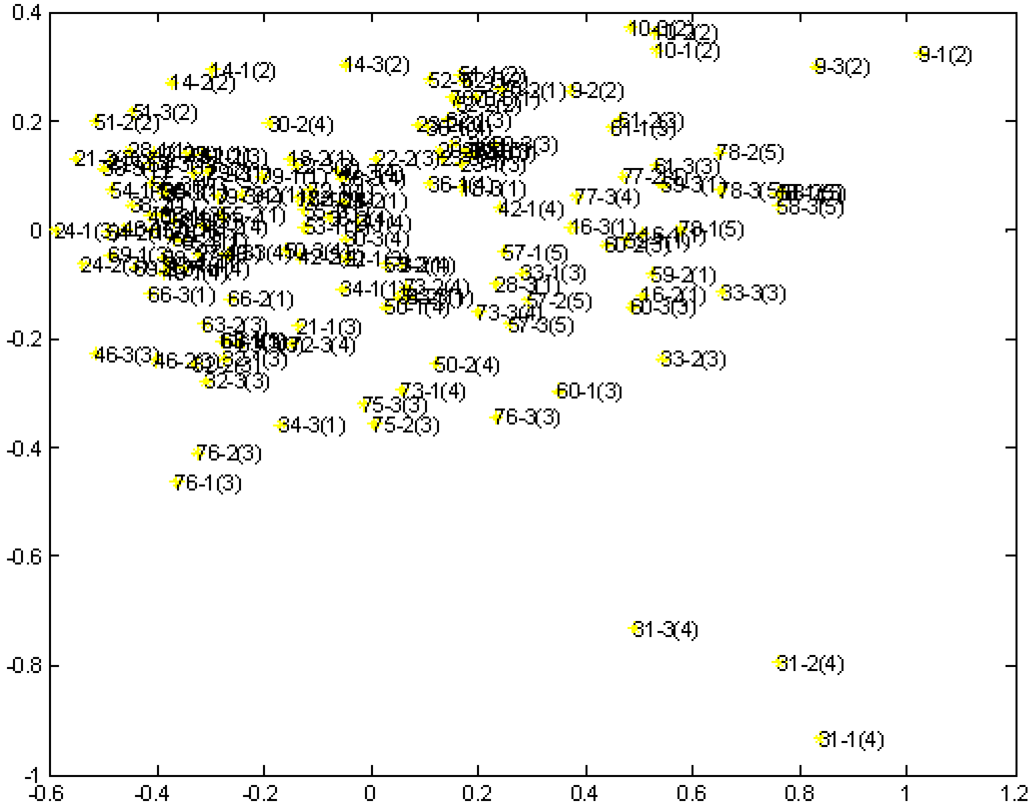


Figure 4.3.55. MDS output for MN Statewide 2003 Data based on *Subset 4* (without averaging the transects).

ANALYSIS

Figure 4.3.52, Figure 4.3.53, Figure 4.3.54 and Figure 4.3.55 suggest that there is no pattern in clustering based on the geomorphologic codes.

[**Goal 4a**] To estimate how the statewide samples are related based on provenance (NW/NE/N) based on all elements.

DATASET USED: MN Statewide 2003 Soil Survey Data. However in this case we use only the samples which belong to NW, N and NE provenance.

NAMING CONVENTION: In order to understand the results better we introduce a naming convention for the different soil samples of the data. This naming convention is described below: Basically in order to understand the pattern of clustering based on the provenance we introduce a map as shown in Figure 4.3.56. (This map has been provided by Mr. James Seaberg).Basically now we divide the map based on their provenance of the glacial sediments.

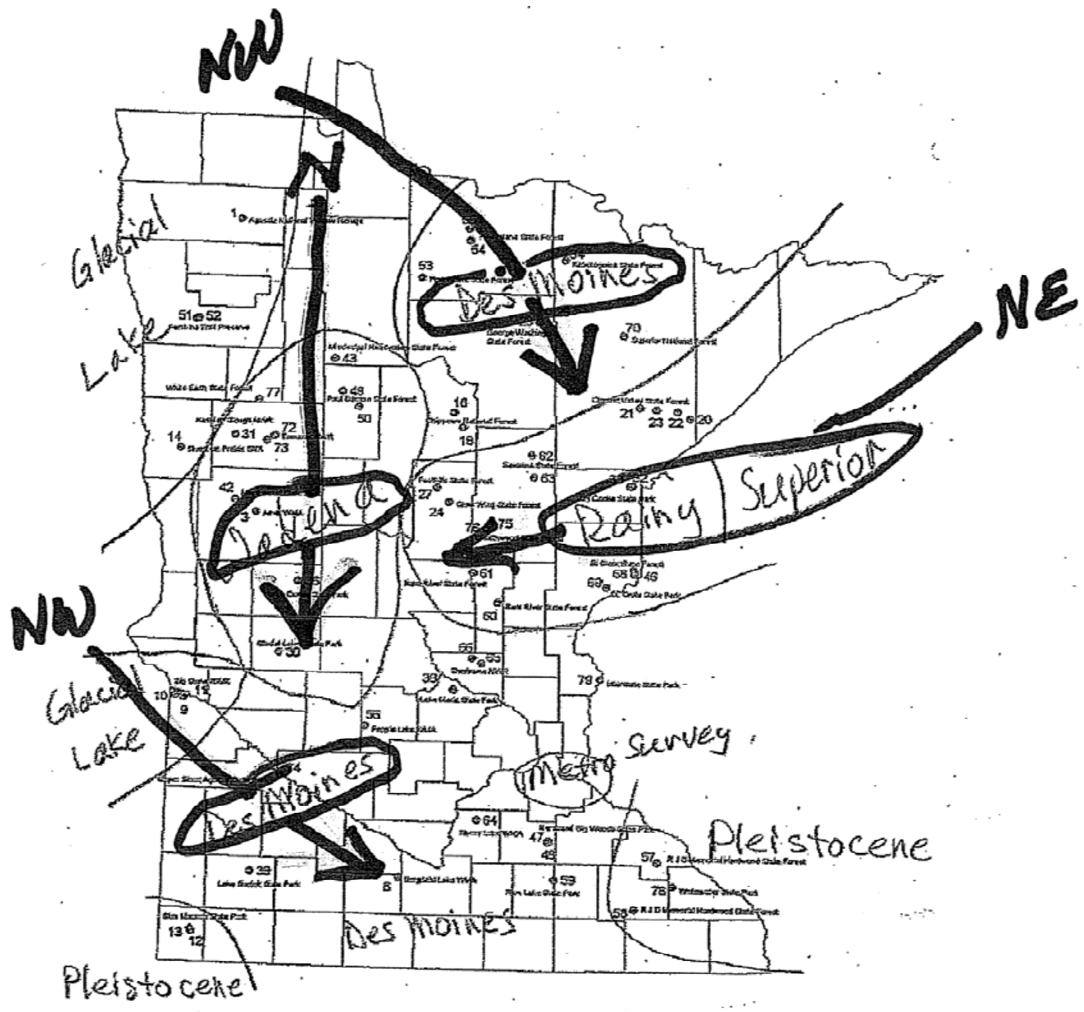


Figure 4.3.56. Map showing the different categories based on the provenance (provided by Mr. James Seaberg).

The naming scheme that we used to identify each sample is: **Site ID (PROVENANCE)**

For example:
A sample with its tag as **63(NE)** encodes

Site ID (for the sample) = 63.
Provenance= North East.

Moreover when we consider the transect of each sample we edit the tag to indicate the transect number as **Site ID-TRANSECT NUMBER (PROVENANCE)**

For example:
A sample with its tag as **63-1(NE)** encodes

Site ID (for the sample) = 63.
Transect number=1.
Provenance= North East.

SOM MODELING
The SOM modeling results for this goal is shown below.

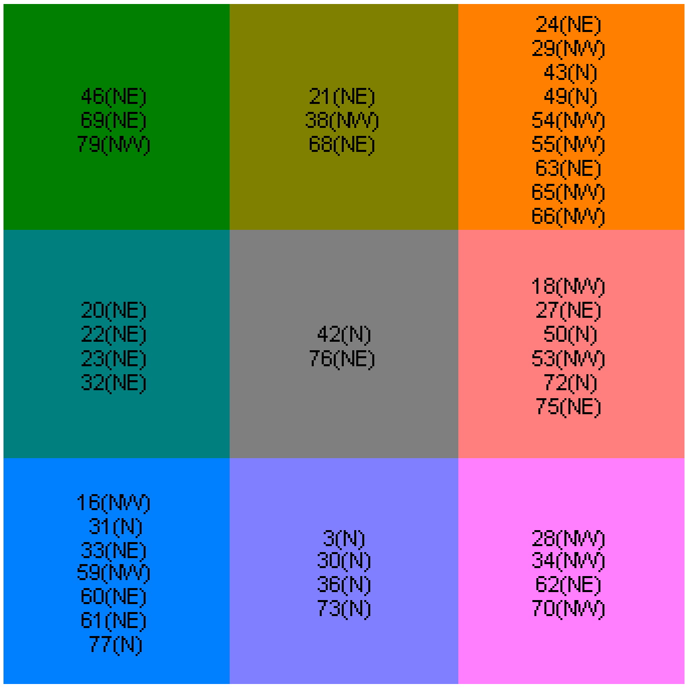


Figure 4.3.57. SOM model for MN Statewide 2003 Data based on all elements.

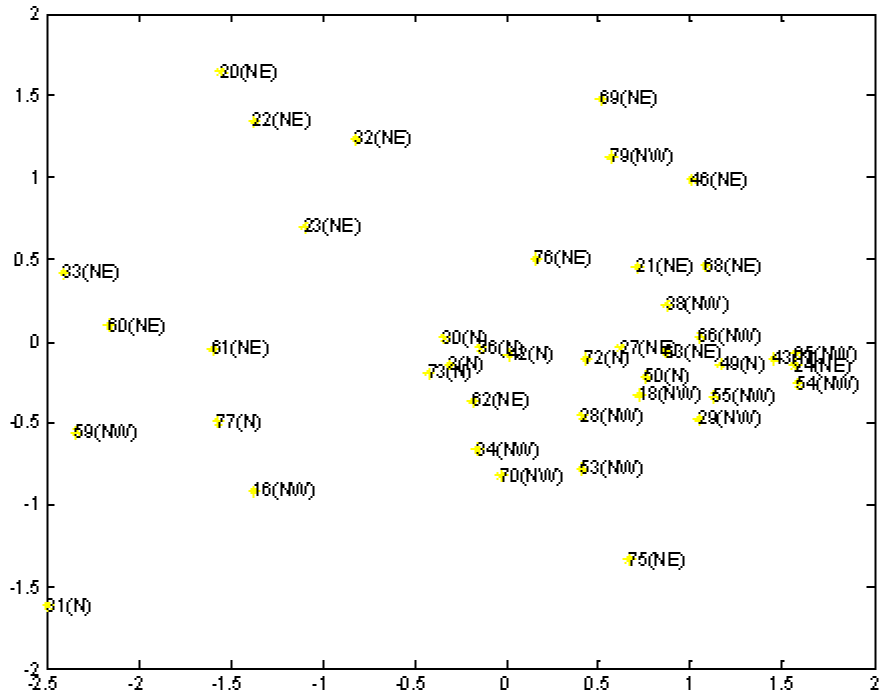


Figure 4.3.58. MDS output for MN Statewide 2003 Data based on all elements.

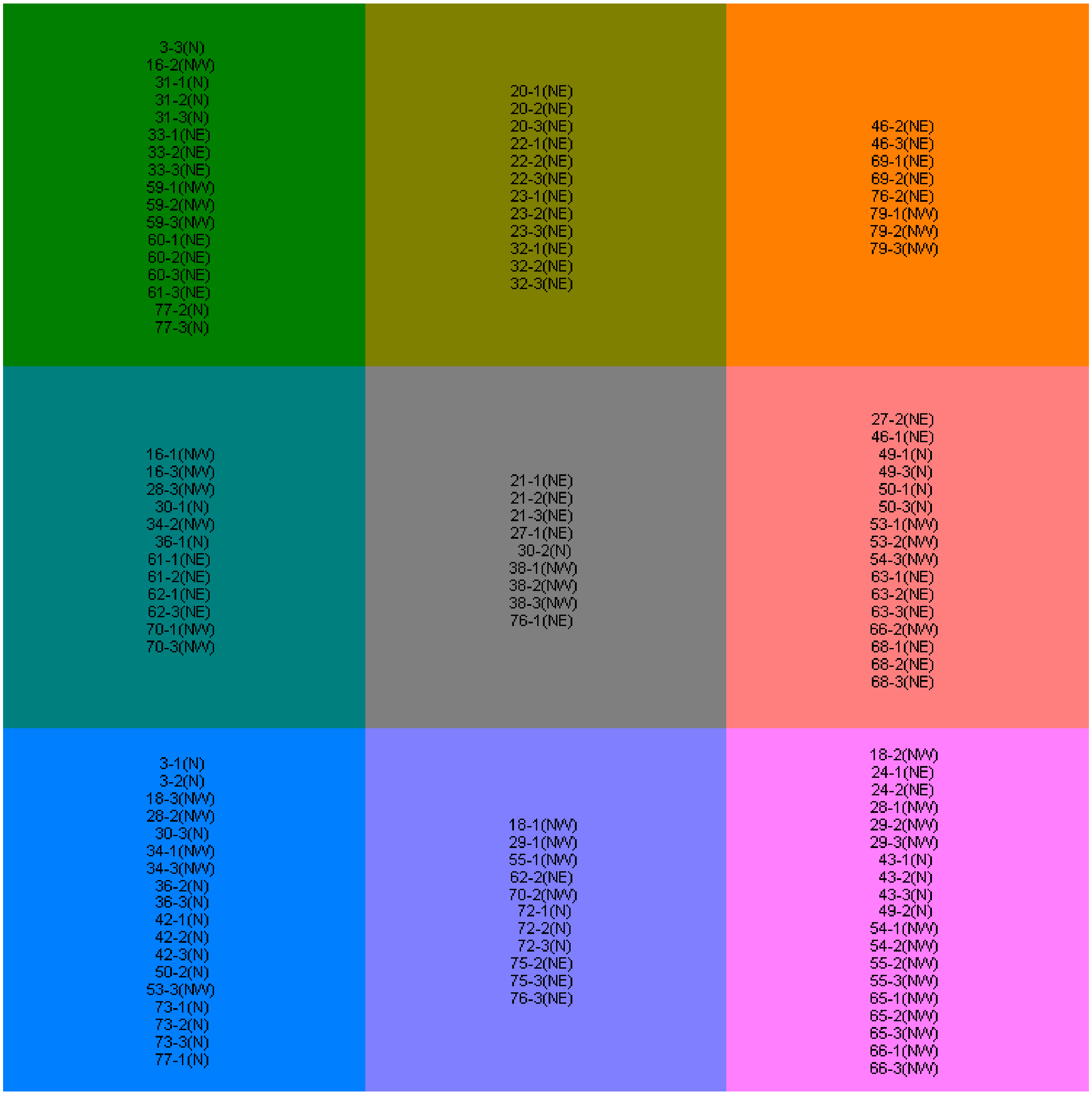


Figure 4.3.59. SOM model for MN Statewide 2003 Data based on all elements (without averaging the transects).

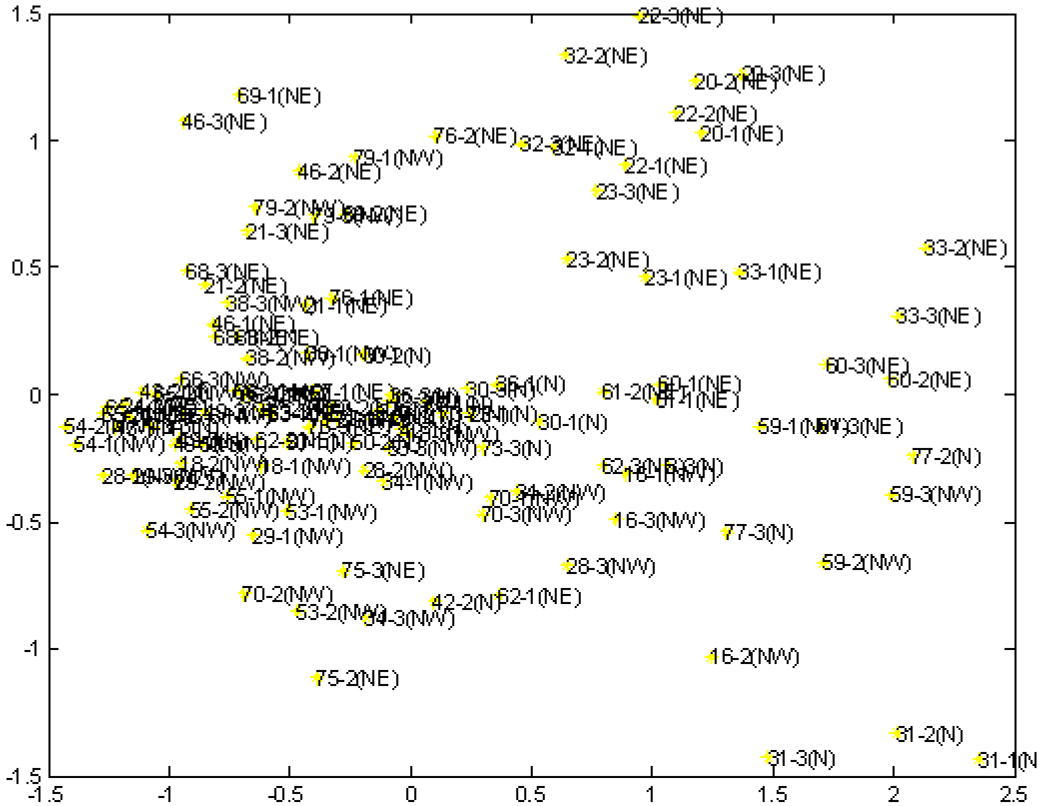


Figure 4.3.60. SOM model for MN Statewide 2003 Data based on all elements (without averaging the transects).

ANALYSIS

From Figure 4.3.57, Figure 4.3.58, Figure 4.3.59 and Figure 4.3.60 we can see that the soil samples belonging to the provenance (NE) seem to cluster together and the soil samples belonging to provenance (N/NW) form a different cluster. Moreover, it may be worth noticing from Figure 4.3.57 and Figure 4.3.59 that the samples of provenance NE are spread outside whereas the samples of type (NW/N) are strongly clustered. This is a possible indication that the samples of type (NW/N) are strongly related and have very similar characteristics among each other as compared to the samples of type (NE).

[**Goal 4b**] To estimate how the statewide samples are related based on provenance (NW/NE/N) based on *Subset 1*.

DATASET USED: Same as the data used for Goal 4a except that in this case we use only the elements as mentioned in *Subset 1*.

NAMING CONVENTION: Same as in Goal 4a.

SOM MODELING
The SOM modeling results for this goal is shown below.

Figure 4.3.61. SOM model for MN Statewide 2003 Data based on *Subset 1*.

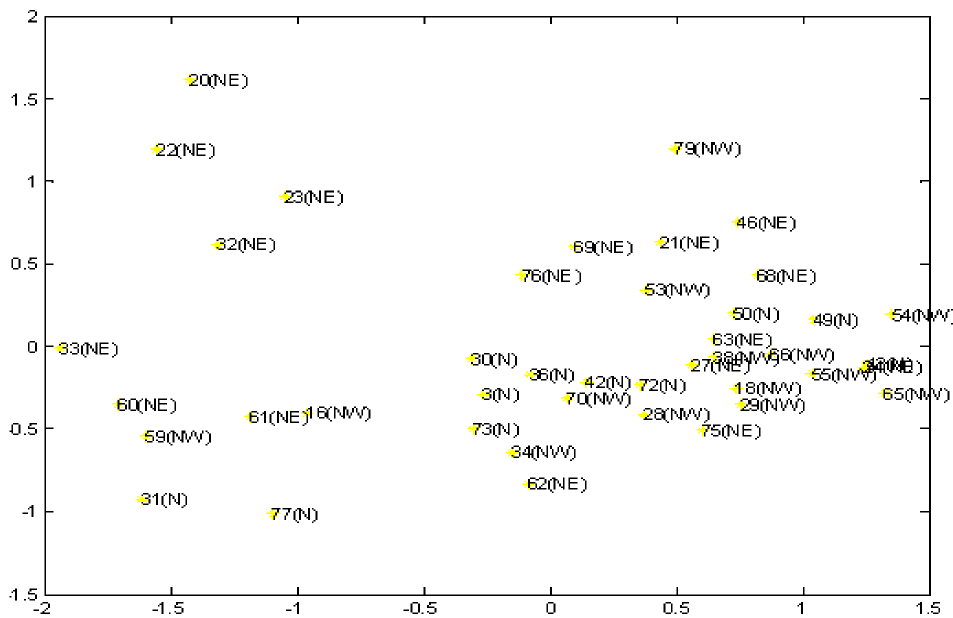


Figure 4.3.62. MDS output for MN Statewide 2003 Data based on *Subset 1*.

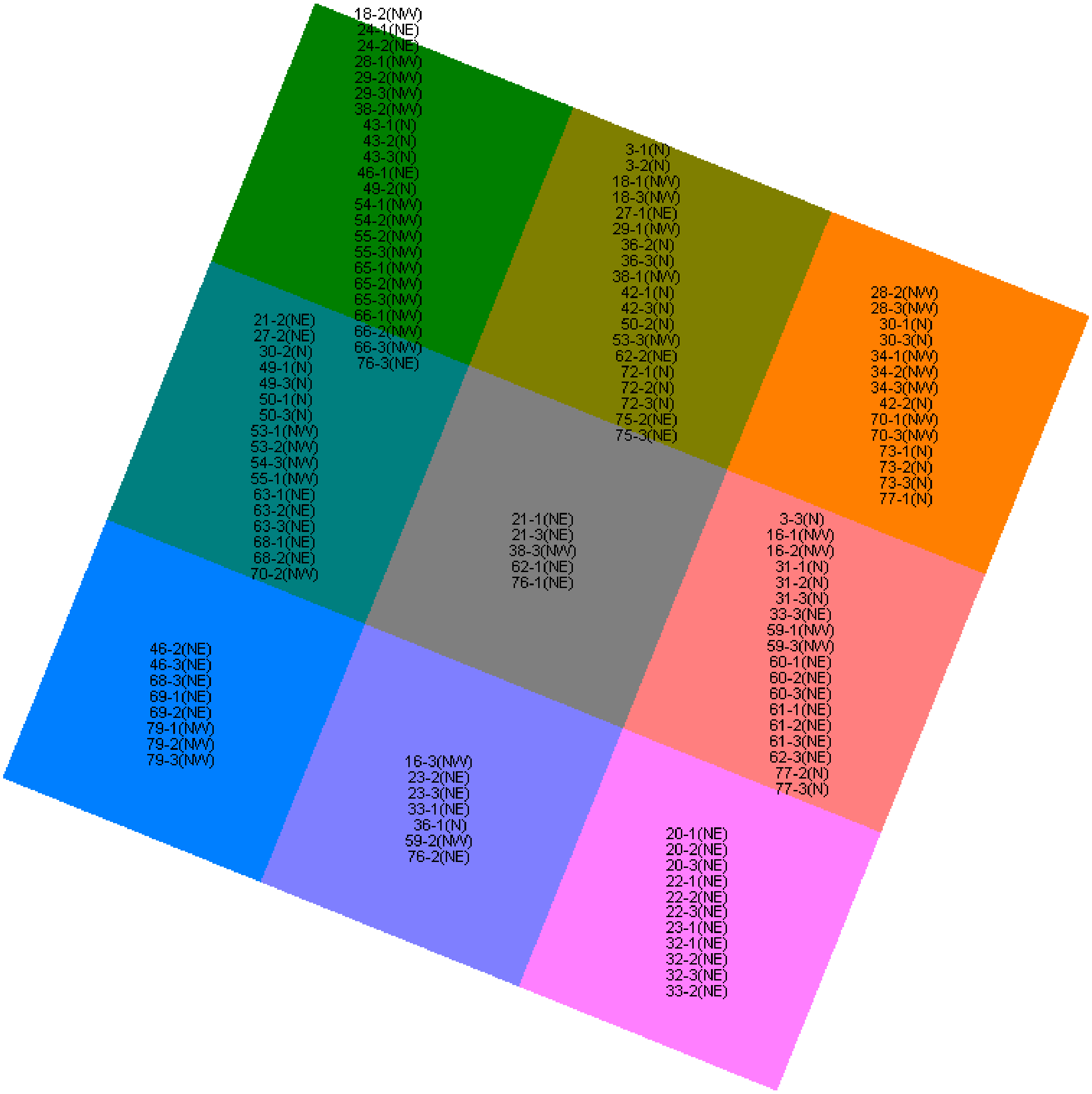


Figure 4.3.63. SOM Model for MN Statewide 2003 Data based on *Subset 1* (without averaging the transects).

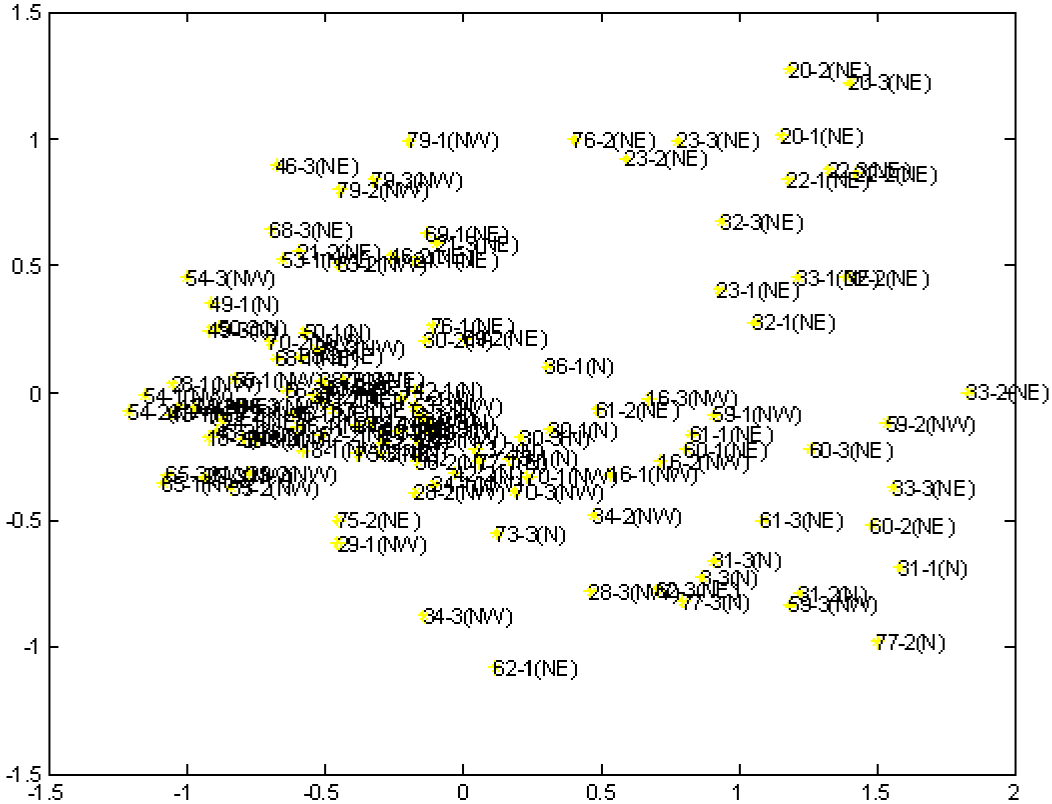


Figure 4.3.64. MDS output for MN Statewide 2003 Data based on *Subset 1* (without averaging the transects).

ANALYSIS
Based on the results obtained from Figure 4.3.61, Figure 4.3.62, Figure 4.3.63, and Figure 4.3.64 we can observe the similar results as seen for Goal 4a. Thus based on *Subset 1* there is some clustering between the soil samples of type NW/N and the soil samples of type NE are differently clustered.

[**Goal 4c**] To estimate how the statewide samples are related based on Provenance (NW/NE/N) based on *Subset 2*.

DATASET USED: Same as the data used for Goal 4a except that in this case we use only the elements as mentioned in *Subset 2*.

NAMING CONVENTION: Same as in Goal 4a.

SOM MODELING

The SOM modeling results for this goal is shown below.

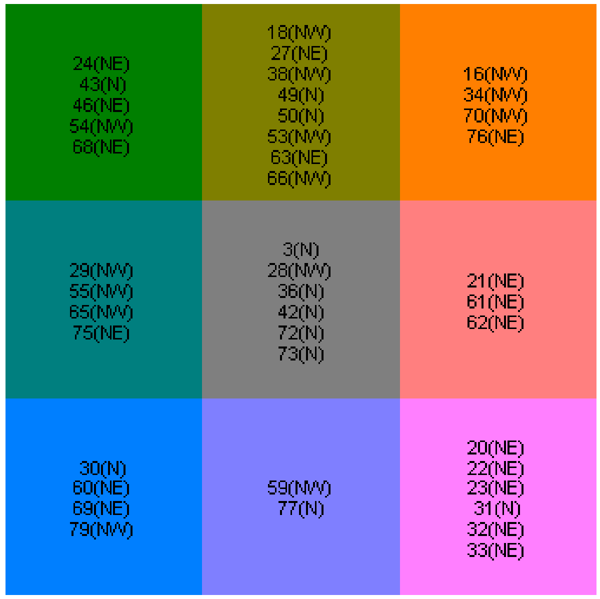


Figure 4.3.65. SOM model for MN Statewide 2003 Data based on *Subset 2*.

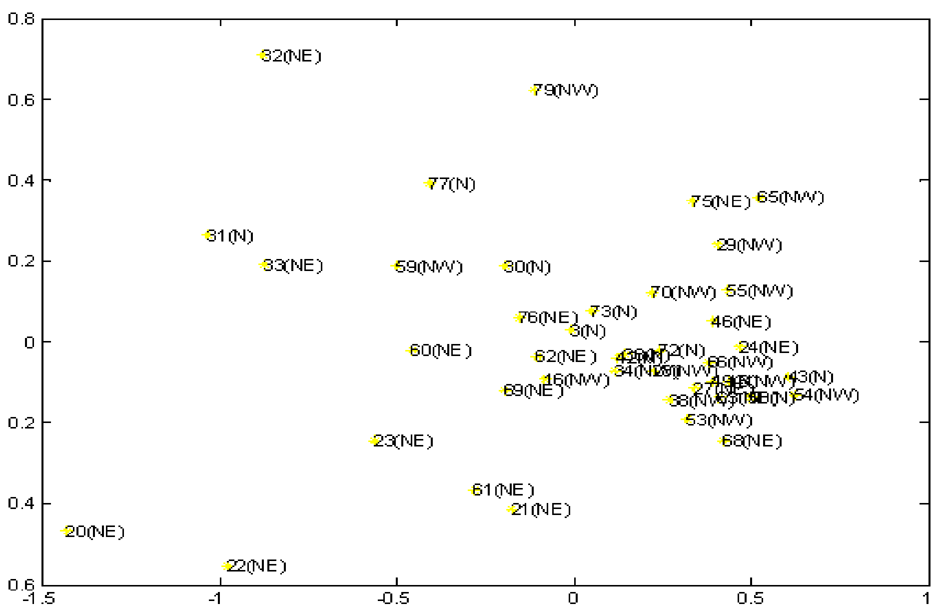


Figure 4.3.66. MDS output for MN Statewide 2003 Data based on *Subset 2*.

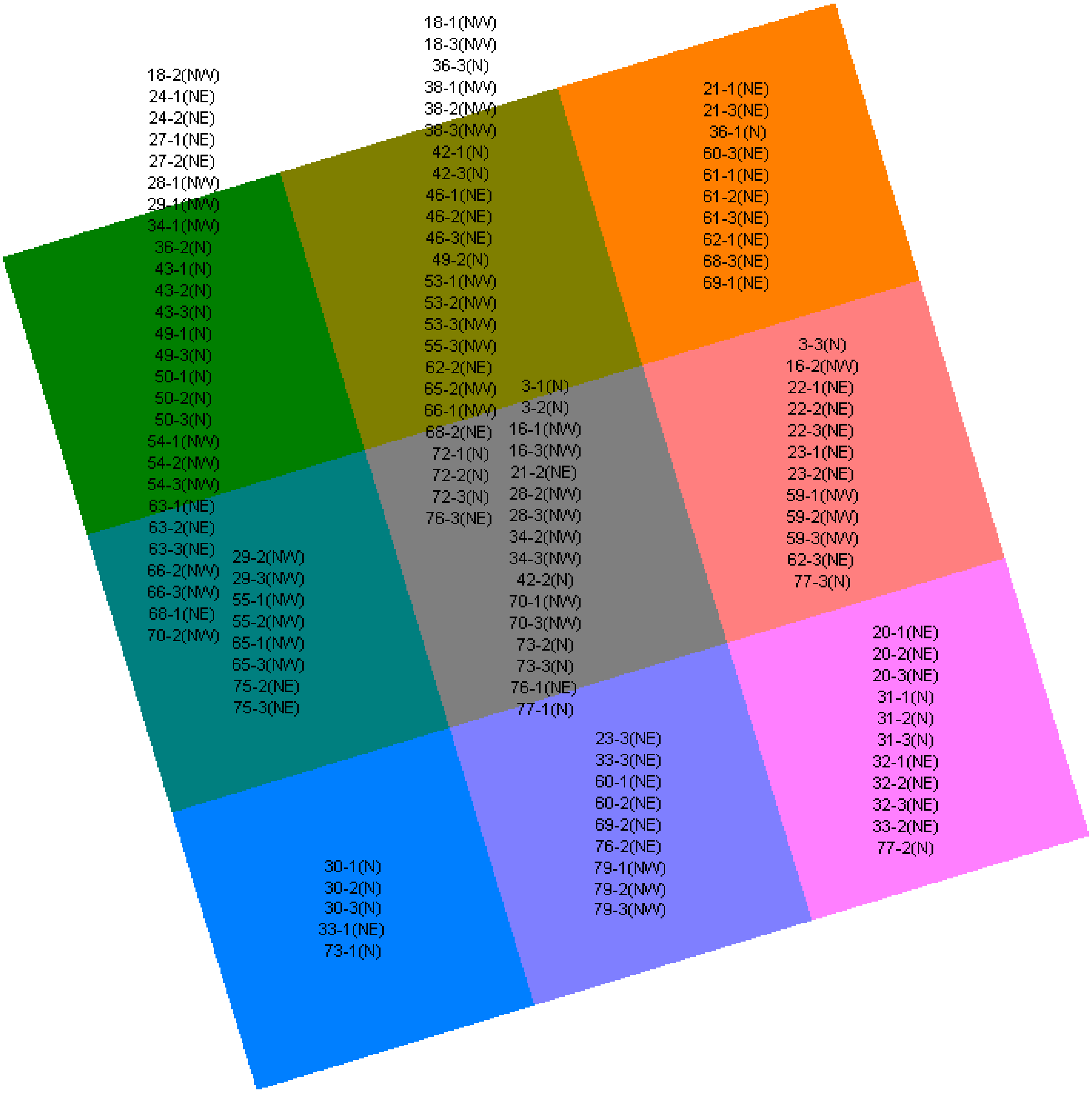


Figure 4.3.67. SOM Model for MN Statewide 2003 Data based on *Subset 2* (without averaging the transects).

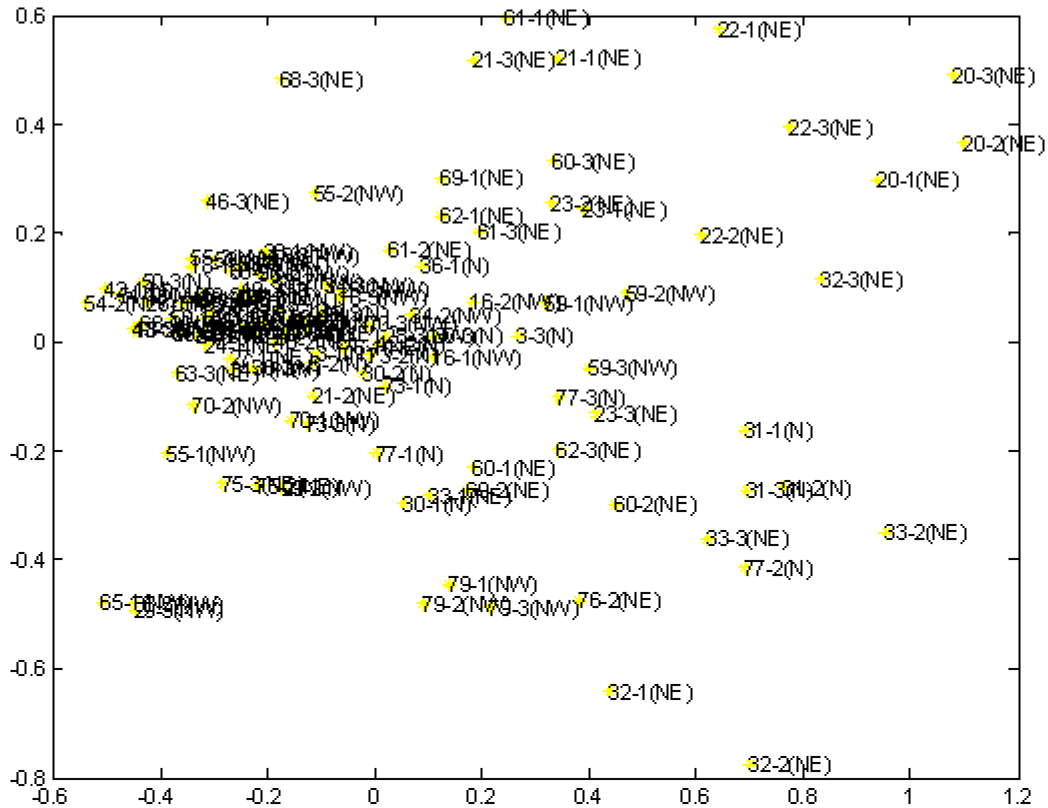


Figure 4.3.68. MDS output for MN Statewide 2003 Data based on *Subset 2* (without averaging the transects).

ANALYSIS

Figure 4.3.65 and Figure 4.3.66 suggest a pattern that the NE samples are clustered together as against the N/NW samples. Further from Figure 4.3.67 we can observe that the NE samples tend to cluster more towards the right bottom of the SOM map and the N/NW samples are more towards the top left. In fact this behavior is also evident from the Figure 4.3.68 where we observe that the N/NW samples form a tight cluster in the middle, whereas the NE samples are mostly spread outside. Thus we can see some form of clustering based on the provenance.

We further build the SOM model for each of the elements taken separately.

ELEMENT=Arsenic (As)

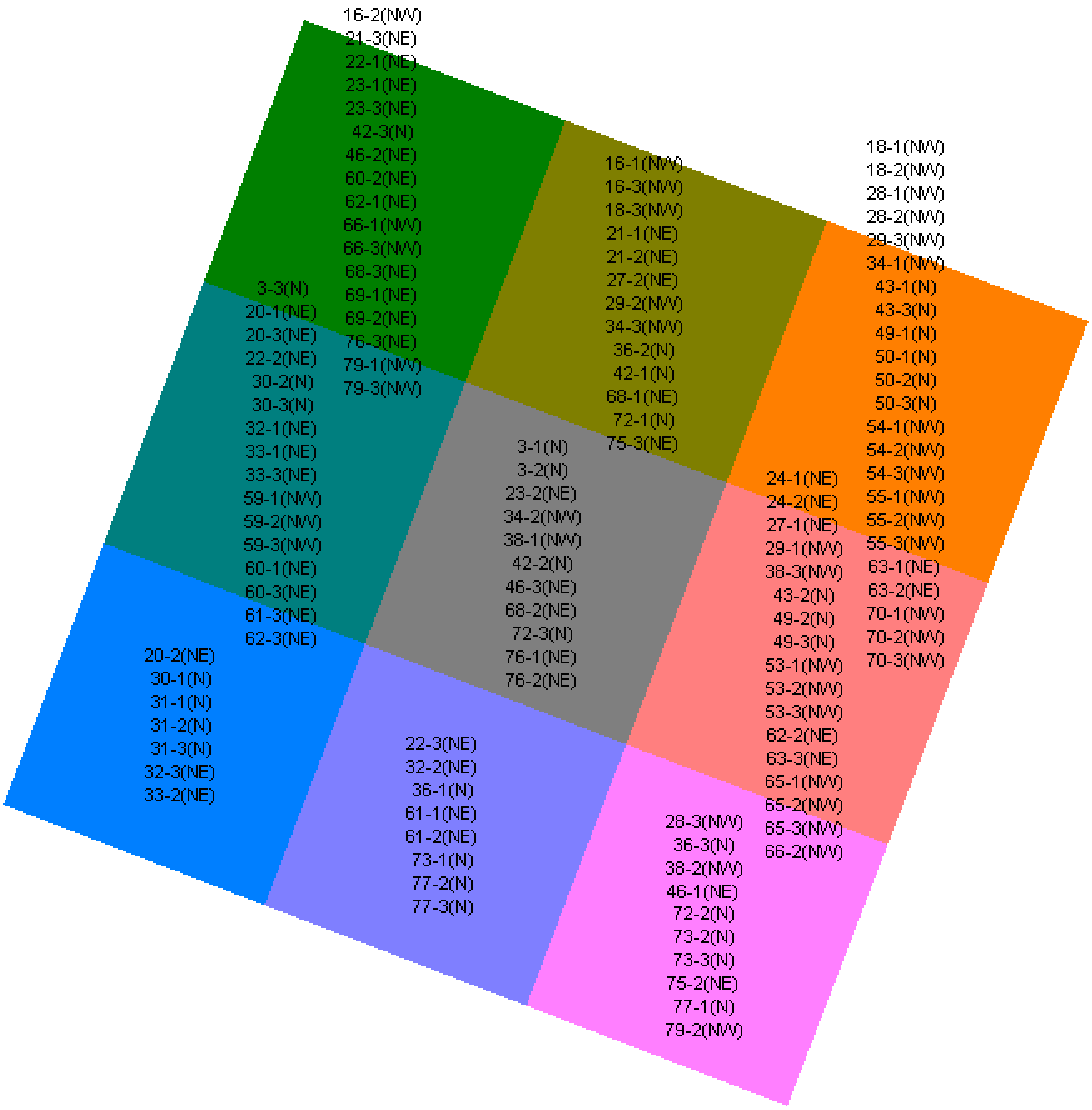


Figure 4.3.69. SOM model for MN Statewide 2003 Data based on As (without averaging the transects).

Figure 4.3.70. SOM model for MN Statewide 2003 Data based on As.

ANALYSIS

As we can observe from Figure 4.3.69 and Figure 4.3.70 we can see that there is a weak pattern of clustering based on provenance. However, if we observe from Figure 4.3.70 the samples 22, 32, 60, 61, 62, 69, 20 seem to cluster together. These samples have higher As concentration values. We observe from Figure 4.3.56 that these samples are geographically on the North-East of MN. Thus there seem to be some form of clustering based on the geographical location of the soil samples. Further based on the bedrock geological Map of MN (provided in Appendix B) we see that the samples 22, 32, 60, 61, 62, 69, 20 belong to *Duluth Complex and associated rocks*. Thus this clustering could in fact indicate some dependence on the mineral deposit of MN.

ELEMENT=Chromium (Cr)

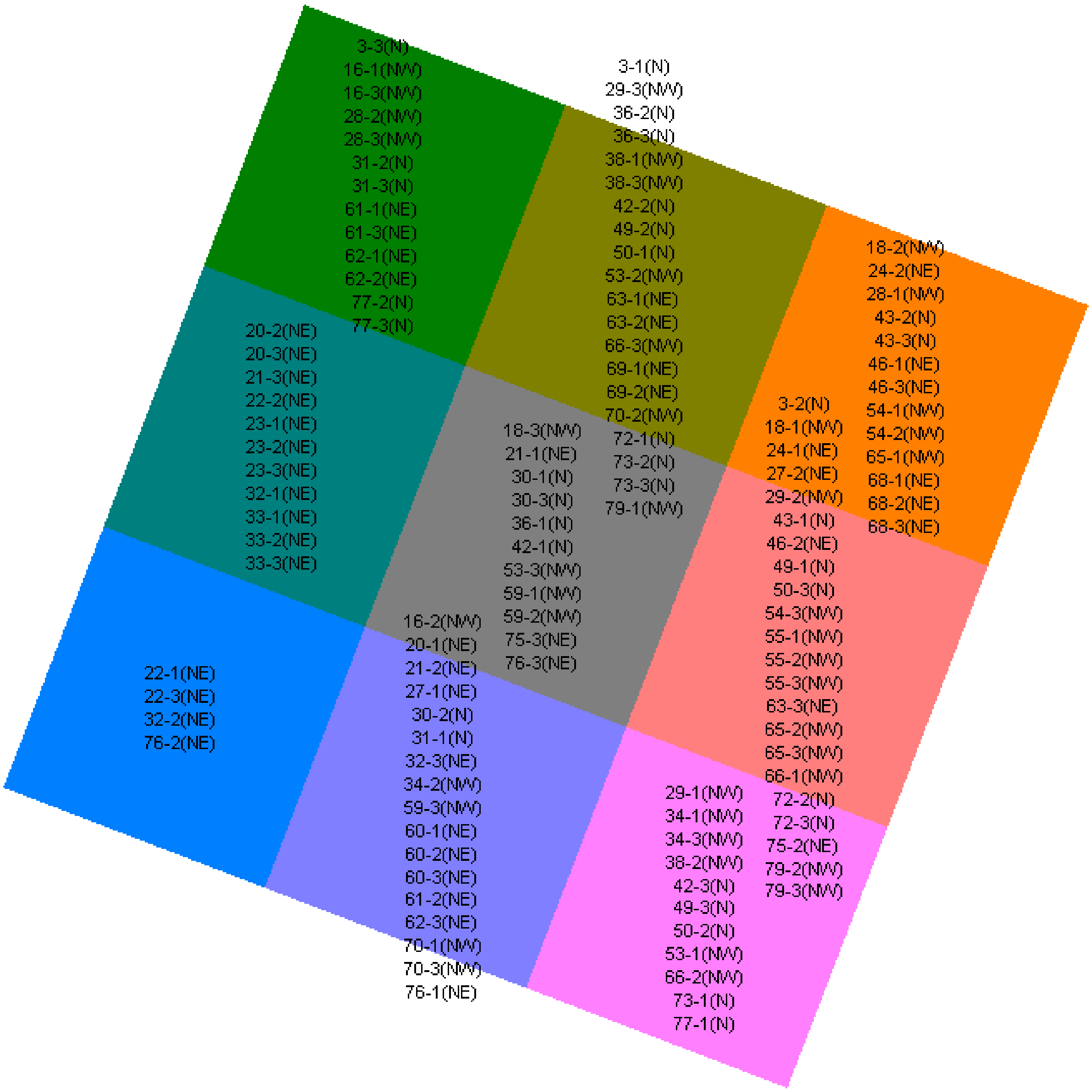


Figure 4.3.71. SOM model for MN Statewide 2003 Data based on Cr (without averaging the transects).

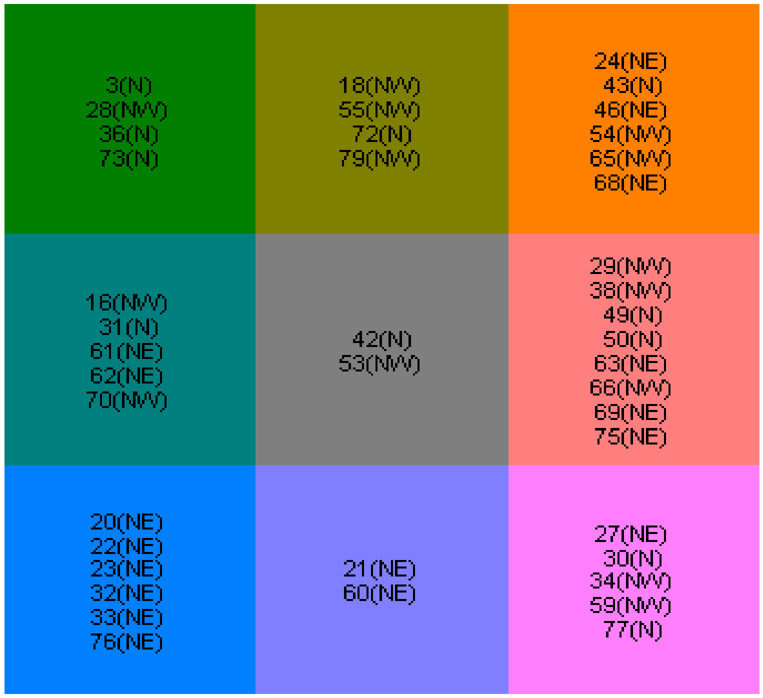


Figure 4.3.72. SOM model for MN Statewide 2003 Data based on Cr.

ANALYSIS

In this case there seems to be weak clustering based on the provenance. However, on careful inspection it seems that the clustering is pretty much geographical. Similar to As in this case too it seems like the samples that are collected on the North-East of the state of Minnesota have the similar concentration values of Chromium (Cr). These samples have higher concentration values for Cr. Further if we check most of the samples that are clustered at the bottom left unit of the SOM belong to the *Duluth Complex and associated rocks* (as explained in Appendix B). Drilling in this terrain has indicated presence of the mineral Cr [5]. Thus it makes sense to observe a clustering of these samples based on the Cr concentration values.

ELEMENT= Copper (Cu)

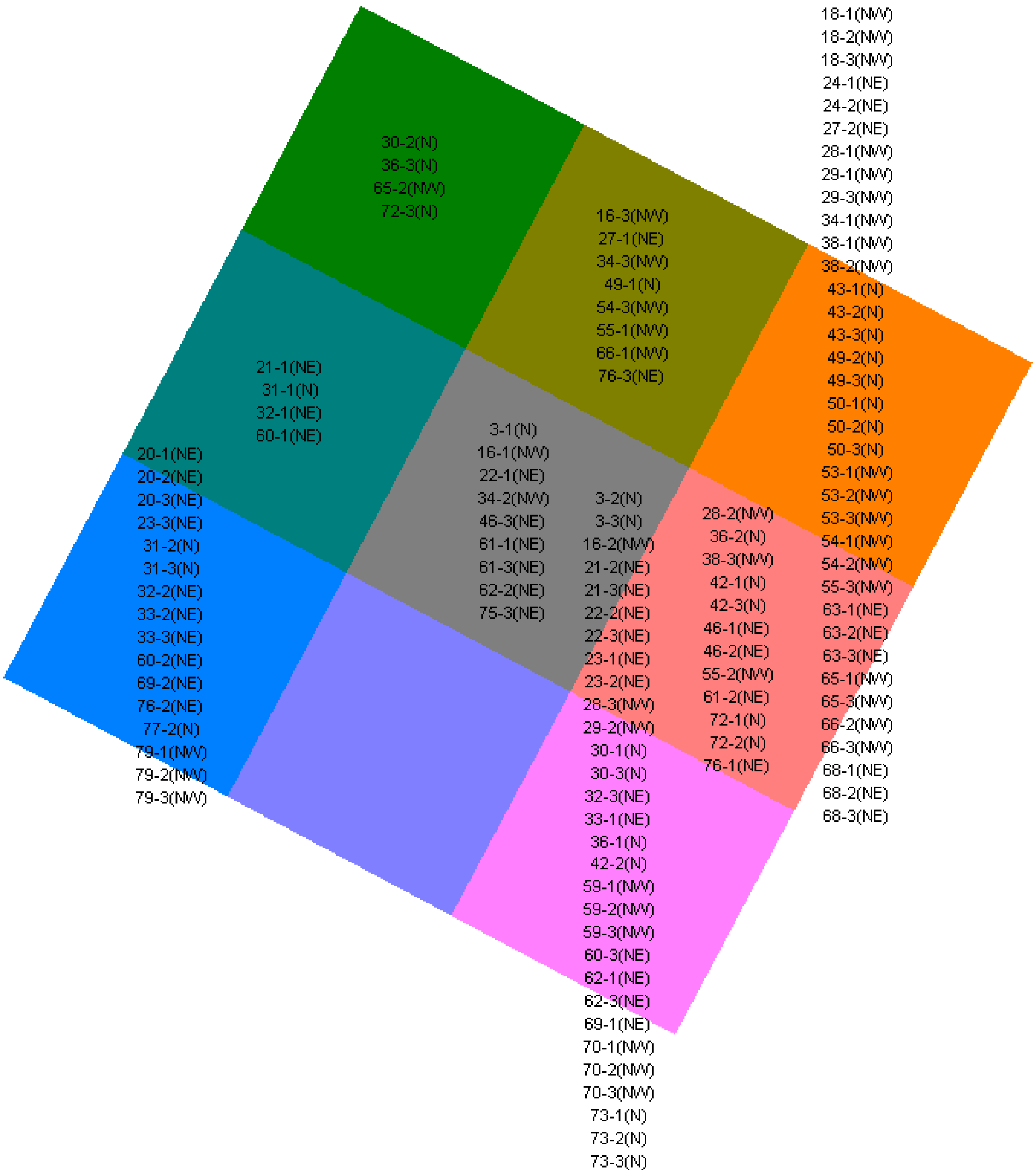


Figure 4.3.73. SOM model for MN Statewide 2003 Data based on Cu (without averaging the transects).

Figure 4.3.74. SOM model for MN Statewide 2003 Data based on Cu.

ANALYSIS

In this case there seems to be weak clustering based on the provenance. However, on careful inspection it seems that the clustering is pretty much geographical. i.e Similar to Cr in this case too it seems like the samples that are collected on the North-East of the state of MN have the similar concentration values of Copper(Cu). These samples have higher Cu values. Further if we check most of the NE samples that are clustered at the left units of the SOM belong to the *Duluth Complex and associated rocks* (as explained in Appendix B). Minnesota Department of Natural Resources estimated the identified Copper-Nickel resource of this area at about 4.4 billion Tons averaging 0.66% Cu and 0.2% Ni [5]. Thus it makes sense to observe a clustering of these samples based on the Cu concentration values.

ELEMENT= Lead (Pb)

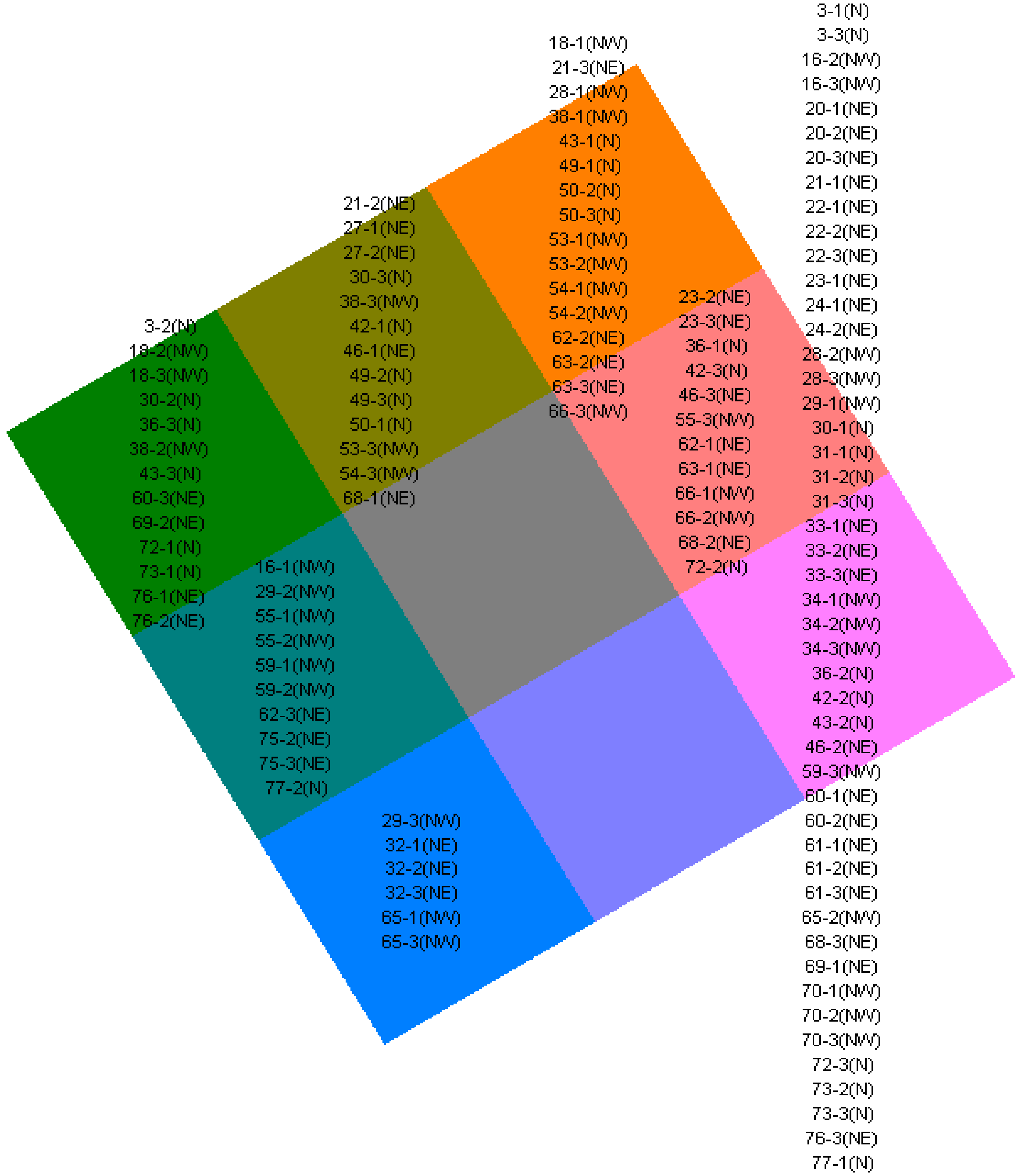


Figure 4.3.75. SOM model for MN Statewide 2003 Data based on Pb (without averaging the transects).

Figure 4.3.76. SOM model for MN Statewide 2003 Data based on Pb.

ANALYSIS

We do not observe any form of clustering in this case.

ELEMENT =Nickel (Ni)

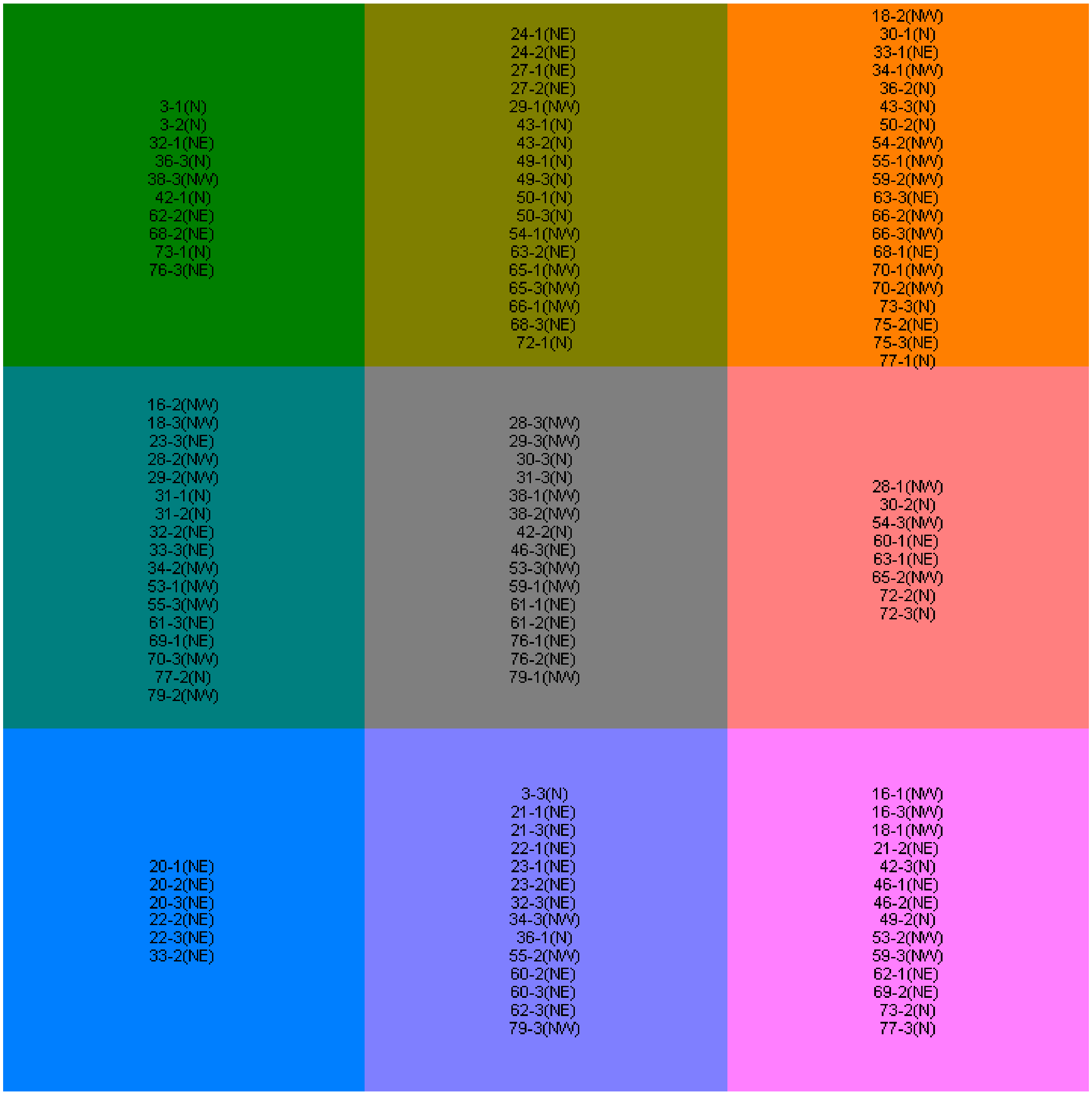


Figure 4.3.77. SOM model for MN Statewide 2003 Data based on Ni (without averaging the transects).

Figure 4.3.78. SOM model for MN Statewide 2003 Data based on Ni.

ANALYSIS

In this case there seems to be weak clustering based on the provenance. However, on careful inspection it seems that the clustering is pretty much geographical. i.e Similar to Copper(Cu) in this case too it seems like the samples that are collected on the North-East of the state of MN have a higher concentration value of Nickel. Further if we check most of the NE samples that are clustered at the left units of the SOM belong to the *Duluth Complex and associated rocks* (as explained in Appendix B). Minnesota Department of Natural Resources estimated the identified Copper-Nickel resource of this area at about 4.4 billion Tons averaging 0.66% Cu and 0.2% Ni. Thus it makes sense to observe a clustering of these samples based on the nickel concentration values.

ELEMENT= Zinc (ZN)

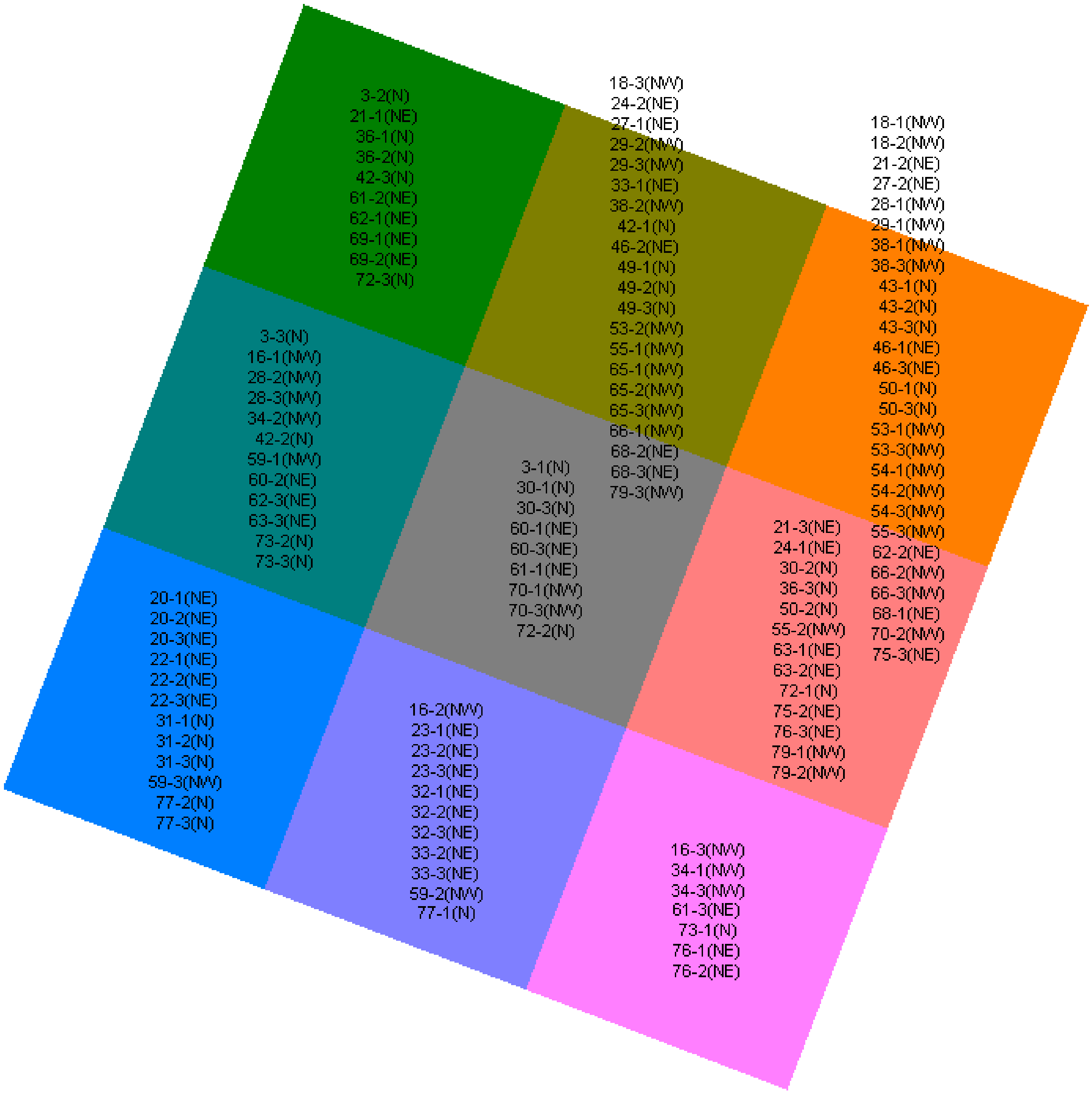


Figure 4.3.79. SOM model for MN Statewide 2003 Data based on Zn (without averaging the transects).

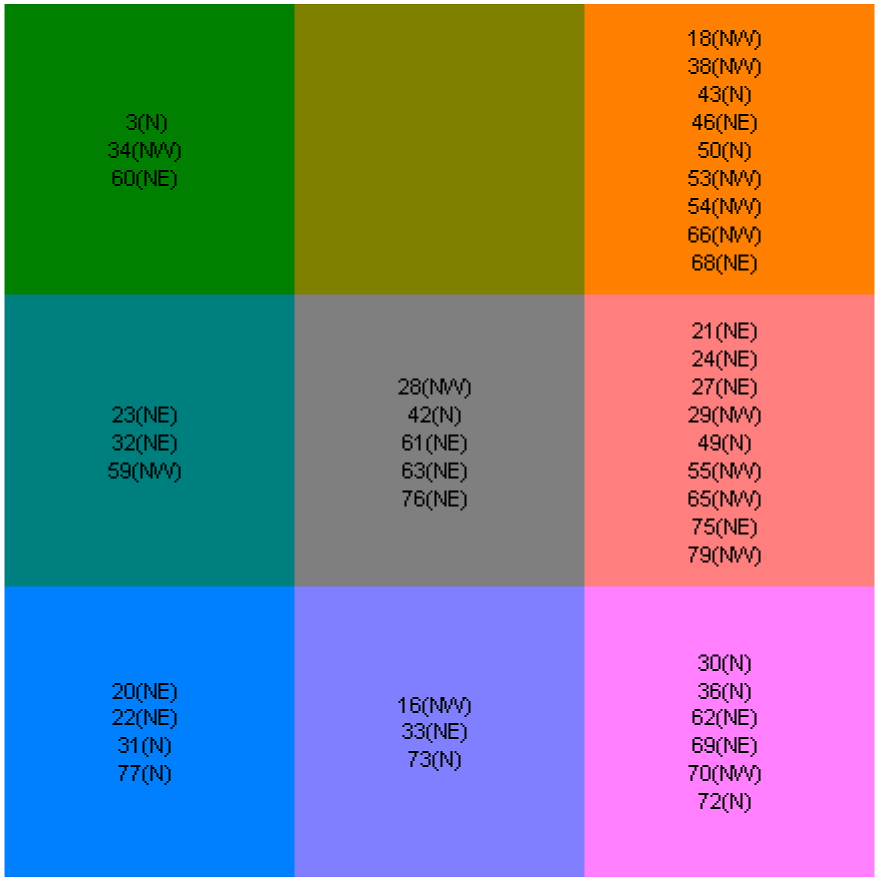


Figure 4.3.80. SOM model for MN Statewide 2003 Data based on Zn.

ANALYSIS

We do not observe any good pattern of clustering for this case.

ELEMENT= Tungsten (W)

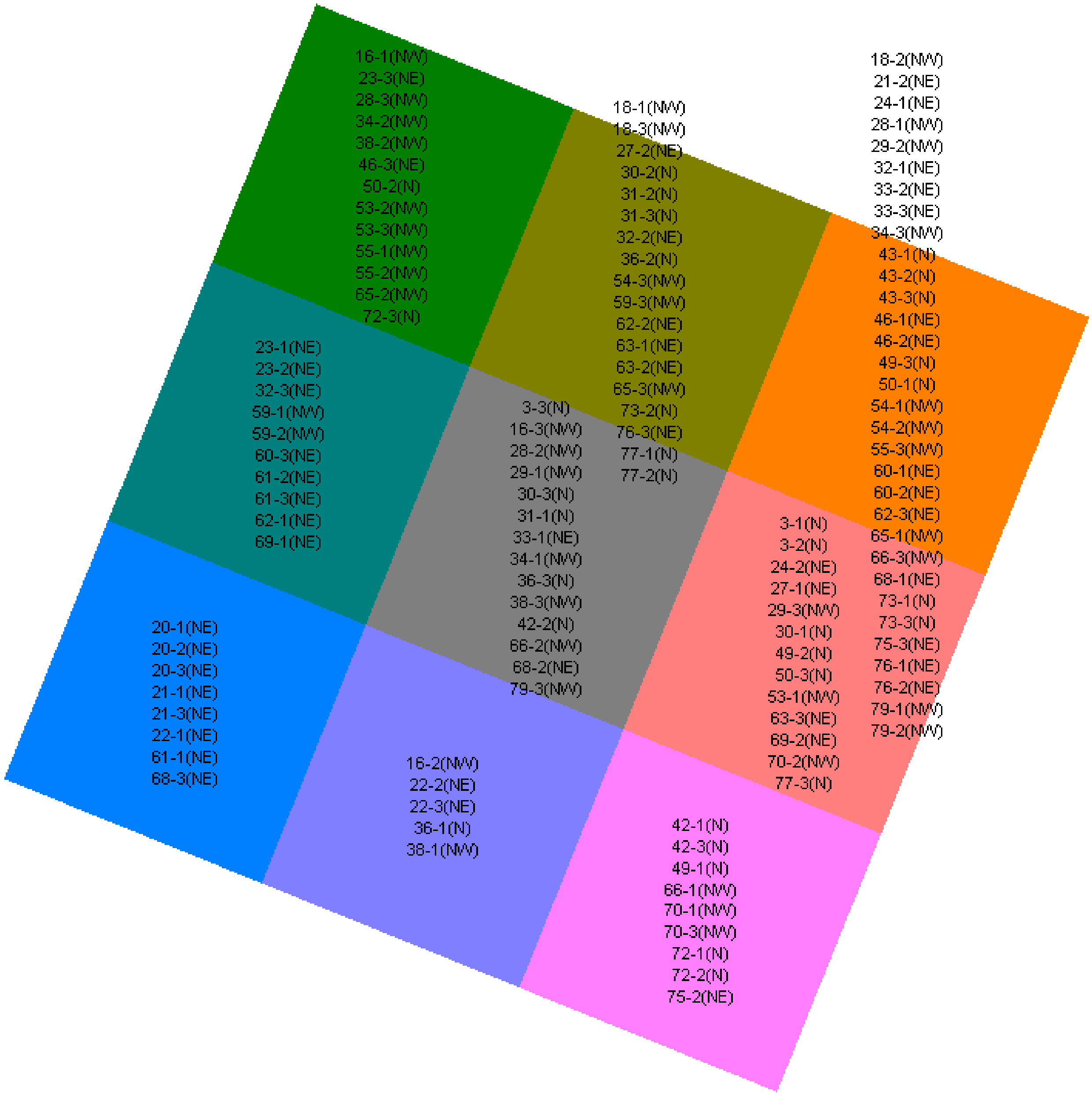


Figure 4.3.81. SOM model for MN Statewide 2003 Data based on W (without averaging he transects).

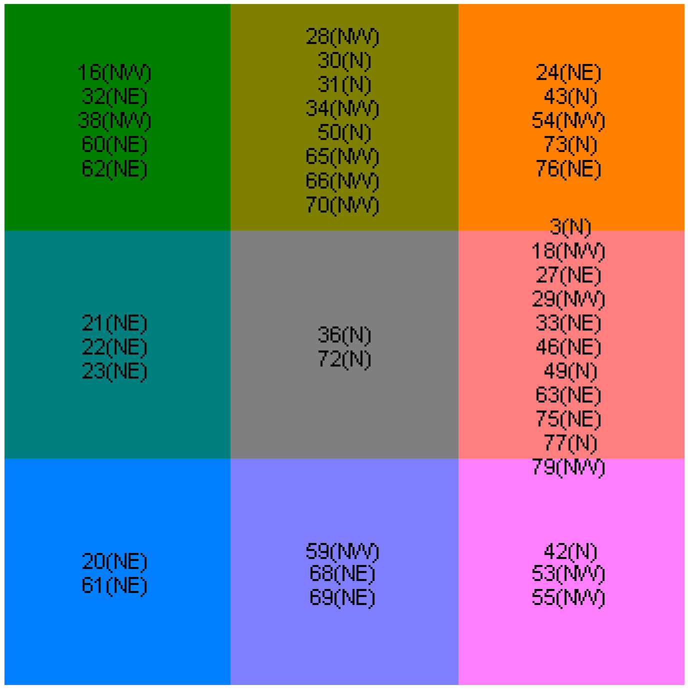


Figure 4.3.82. SOM model for MN Statewide 2003 Data based on W.

ANALYSIS

In this case there seems to be weak clustering based on the provenance. However, on careful inspection it seems that the clustering is pretty much geographical. i.e Similar to As in this case too it seems like the samples that are collected on the North-East of the state of MN have a higher concentration value for Tungsten(W). Further if we check most of the samples that are clustered at the bottom left unit of the SOM belong to the *Duluth Complex and associated rocks* (as explained in Appendix B). Thus this clustering could in fact be as a result of drilling in the mining regions.

[**GOAL 4d**] To estimate how the statewide samples are related based on provenance (NW/NE/N) based on the *Subset 3*.

DATASET USED: Same as the data used for Goal 4a except that in this case we use only the elements as mentioned in Subset3. (Rock forming elements)

NAMING CONVENTION: Same as in Goal 4a.

SOM MODELING

The SOM modeling results for this goal is shown below.

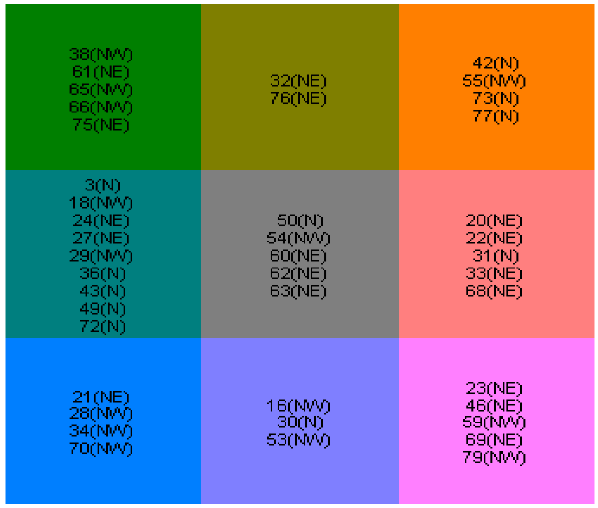


Figure 4.3.83. SOM model for MN Statewide 2003 Data based on *Subset 3*.

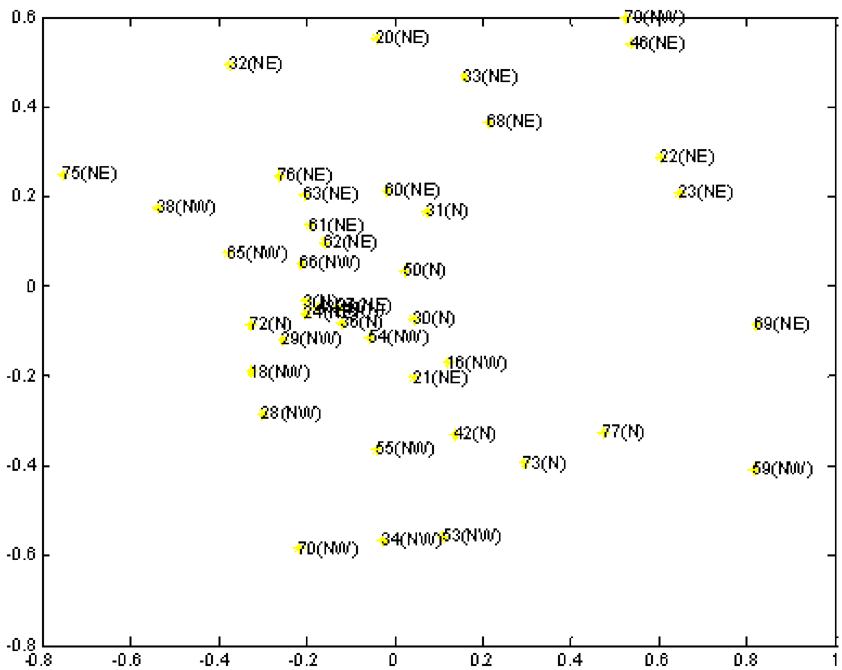


Figure 4.3.84. MDS output for MN Statewide 2003 Data based on *Subset 3*.

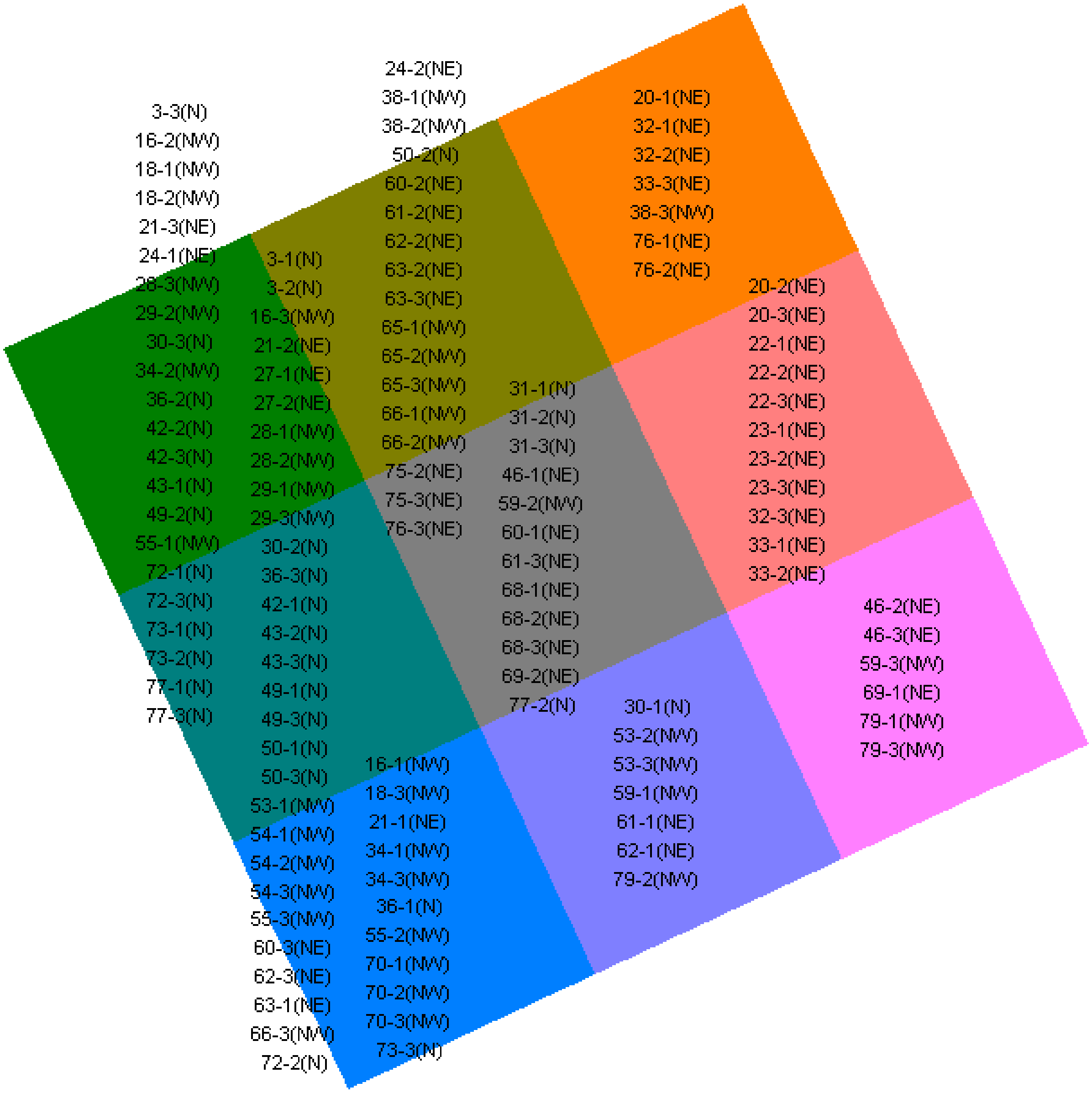


Figure 4.3.85. SOM Model for MN Statewide 2003 Data based on *Subset 3* (without averaging the transects).

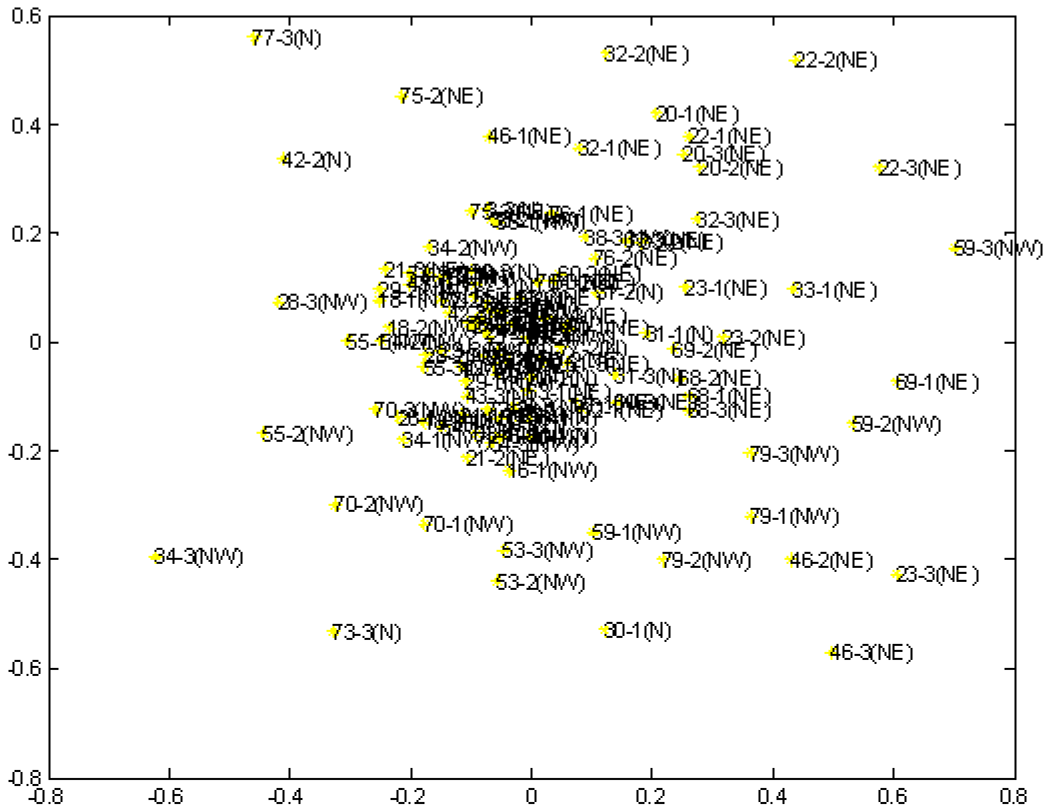


Figure 4.3.86. MDS output for MN Statewide 2003 Data based on *Subset 3* (without averaging the transects).

ANALYSIS

Based on Figures 4.83-86 it seems like there is a good pattern of clustering based on the provenance when we select the rock forming elements. This result indicates that based on the rock forming elements the soil samples have a strong relationship with their provenance of glacial sediments.

[**GOAL 4e**] To estimate how the statewide samples are related based on provenance (NW/NE/N) based on the *Subset 4*.

DATASET USED: Same as the data used for Goal 4a except that in this case we use only the elements as mentioned in Subset 4.

NAMING CONVENTION: Same as in Goal 4a.

SOM Modeling
The SOM modeling results for this goal is shown below.

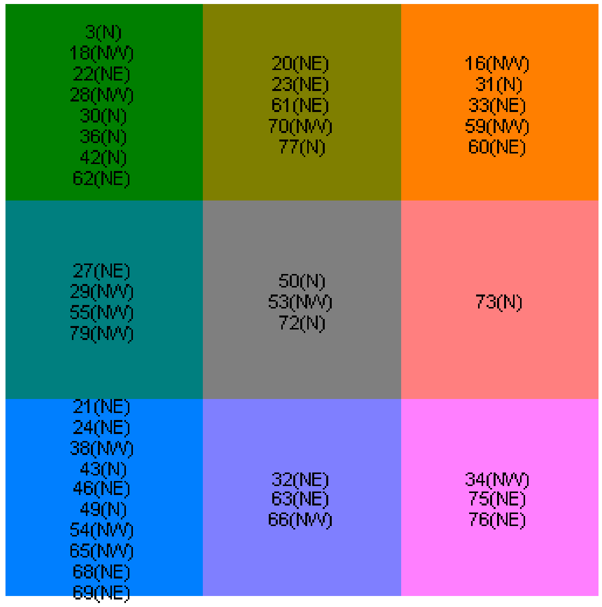


Figure 4.3.87. SOM model for MN Statewide 2003 Data based on *Subset 4*.

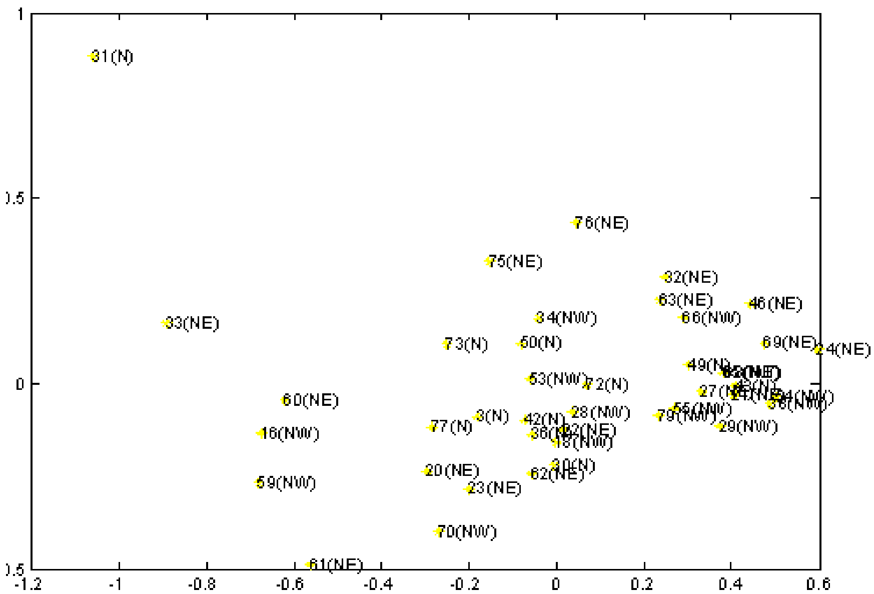


Figure 4.3.88. MDS output for MN Statewide 2003 Data based on *Subset 4*.

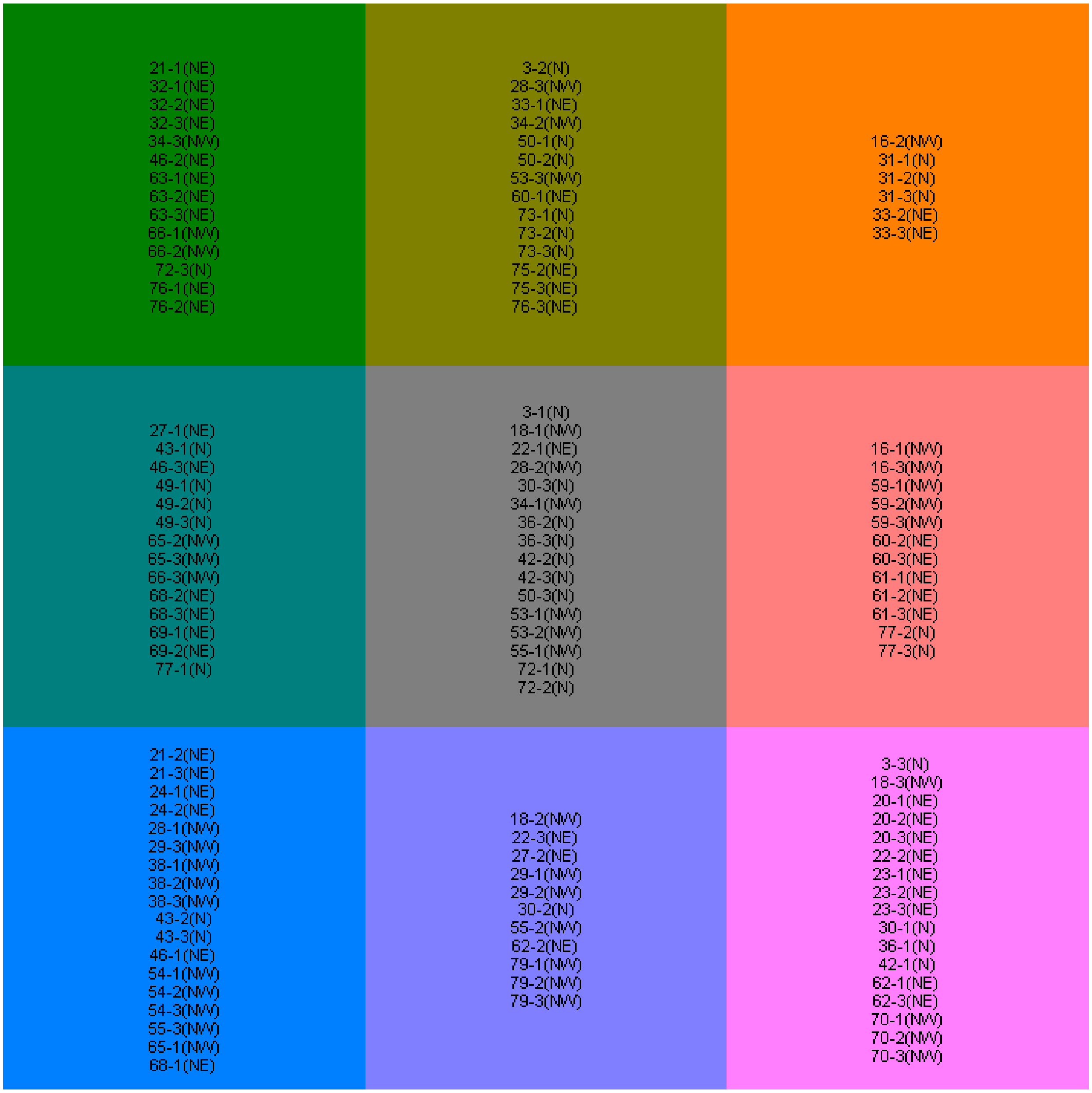


Figure 4.3.89. SOM Model for MN Statewide 2003 Data based on *Subset 4* (without averaging the transects).

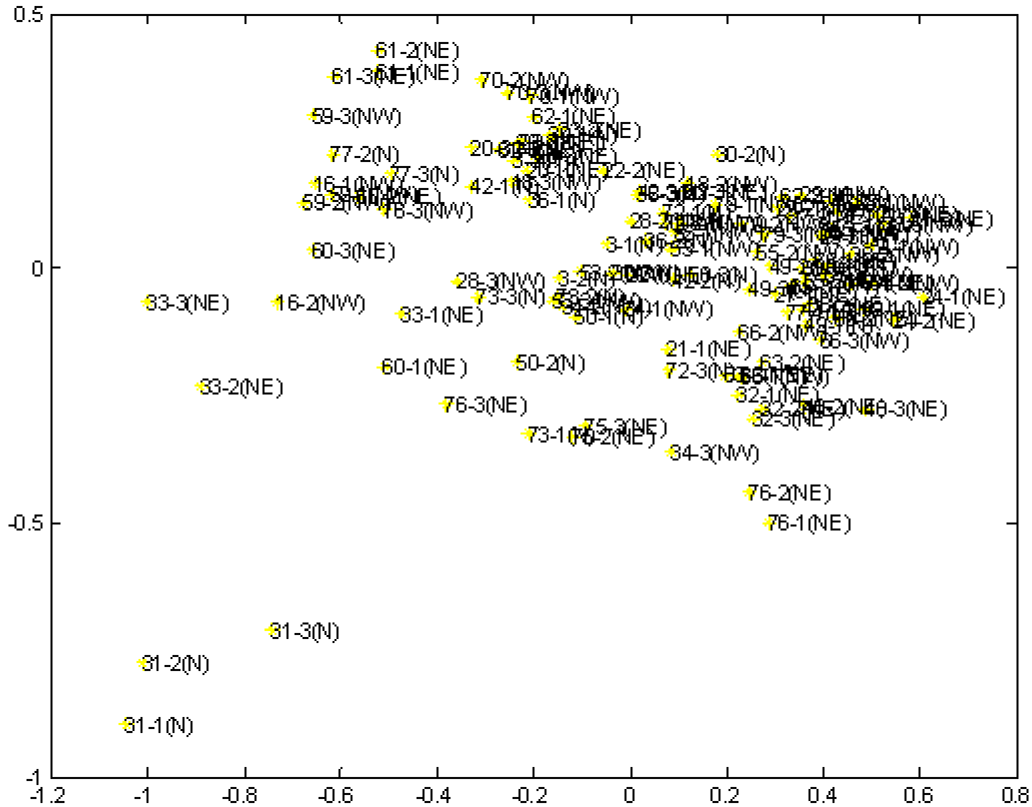


Figure 4.3.90. MDS output for MN Statewide 2003 Data based on *Subset 4* (without averaging the transects)

ANALYSIS
We do not observe any form of clustering for this case.

[**Goal 5a**] To estimate how the statewide samples are related based on Lacustrine, Till, Outwash (L/T/O) based on all elements.

DATASET USED: MN Statewide 2003 Soil Survey data.

NAMING CONVENTION: In order to understand the results better we introduce a naming convention for the different soil samples of the data. This naming convention is described below. Basically in order to understand the pattern of clustering based on the L/T/O we introduce a map as shown in Figure 4.3.85. (This map has been used from [6] and the markings are provided by Mr. James Seaberg).Basically now we divide the map of MN based on Lacustrine/Till/Outwash.

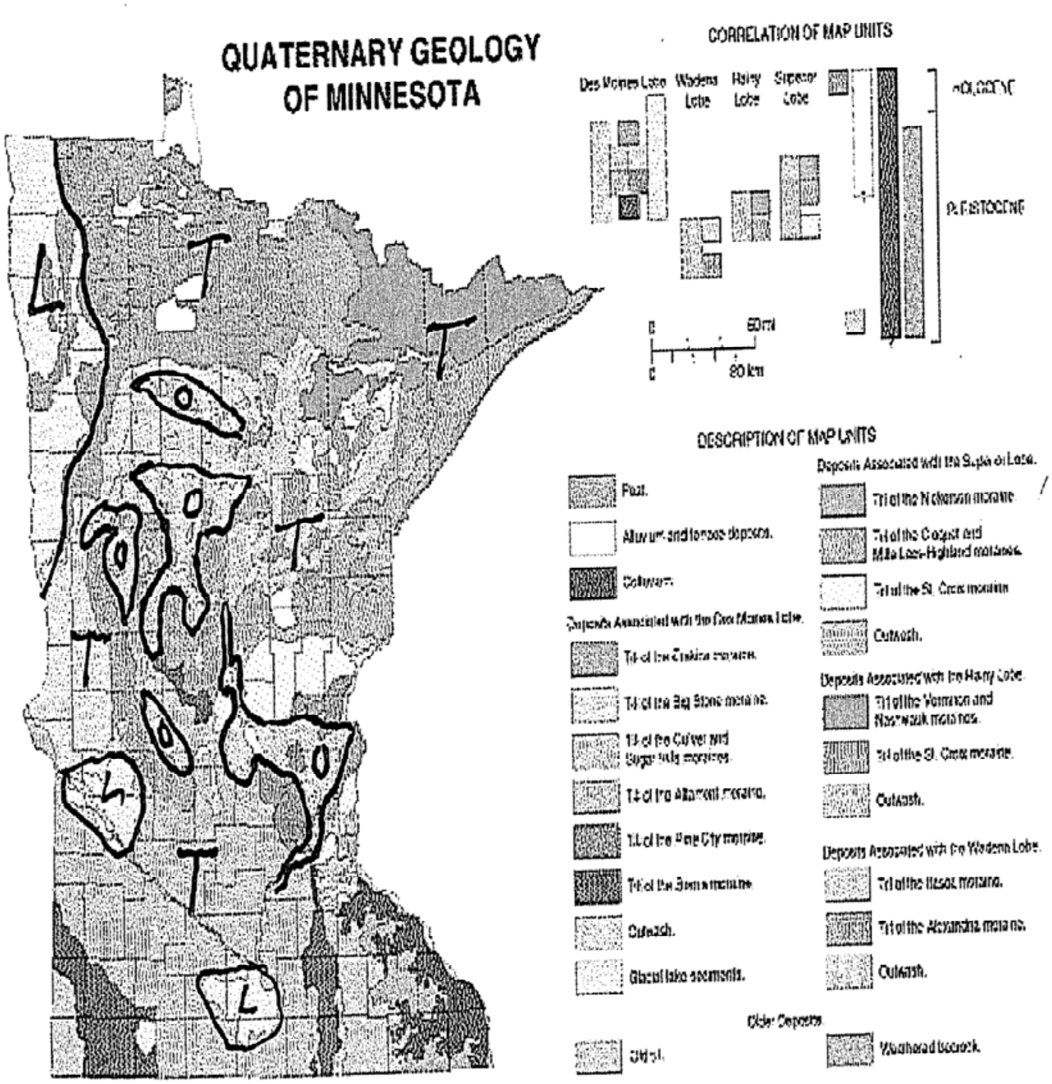


Figure 4.3.91. Map showing the different categories based on the L/T/O. (provided by Mr. James Seaberg)

The naming scheme that we used to identify each sample is: **Site ID (L/T/O)**

For example:
A sample with its tag as **63(T)** encodes.
Site ID (for the sample) = 63.
Soil Type=Till.

SOM MODELING
The SOM modeling results for this goal is shown below. For this case about 93% of the samples belong to the samples of type till. As the number of samples for the other soil type is significantly low it is not possible to derive conclusions about any pattern of clustering. However

we show the SOM output after averaging the transects just to illustrate that with more samples it may be possible to derive good results.

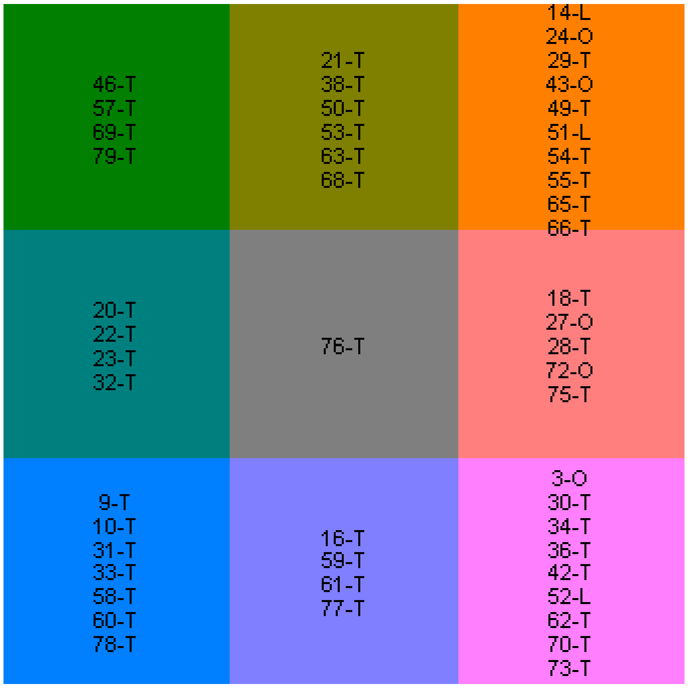


Figure 4.3.92. SOM model for MN Statewide 2003 Data based on all elements.

ANALYSIS

Although the number of samples is pretty low however Figure 4.3.86 gives us an intuition that there may be a pattern of clustering, suggesting that the samples of type Outwash (O) seem to cluster together.

[**Goal 5b**] To estimate how the statewide samples are related based on Lacustrine, Till, Outwash (L/T/O) based on *Subset 1*.

DATASET USED: MN Statewide 2003 Soil Survey Data. Only the elements of the *Subset 1* are used in this case.

NAMING CONVENTION: Same as Goal 5a.

SOM MODELING

The SOM model for this goal is given below.

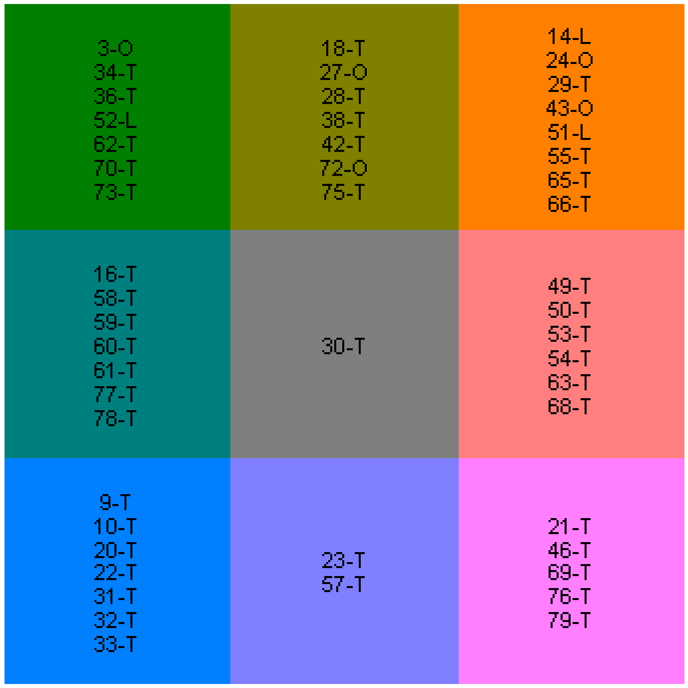


Figure 4.3.93. SOM model for MN Statewide 2003 Data based *Subset 1*.

ANALYSIS

Although the number of samples is pretty low however Figure 4.3.87 gives us an intuition that there may be a pattern of clustering suggesting that the samples of type Outwash (O) seem to cluster together. However we cannot make conclusions from it.

[**Goal 5c**] To estimate how the statewide samples are related based on Lacustrine, Till, Outwash (L/T/O) based on *Subset 2.*

DATASET USED: MN Statewide 2003 Soil Survey Data. Only the elements of the *Subset 2* are used in this case.

NAMING CONVENTION: Same as Goal 5a.

SOM MODELING

The SOM model for this goal is given below.

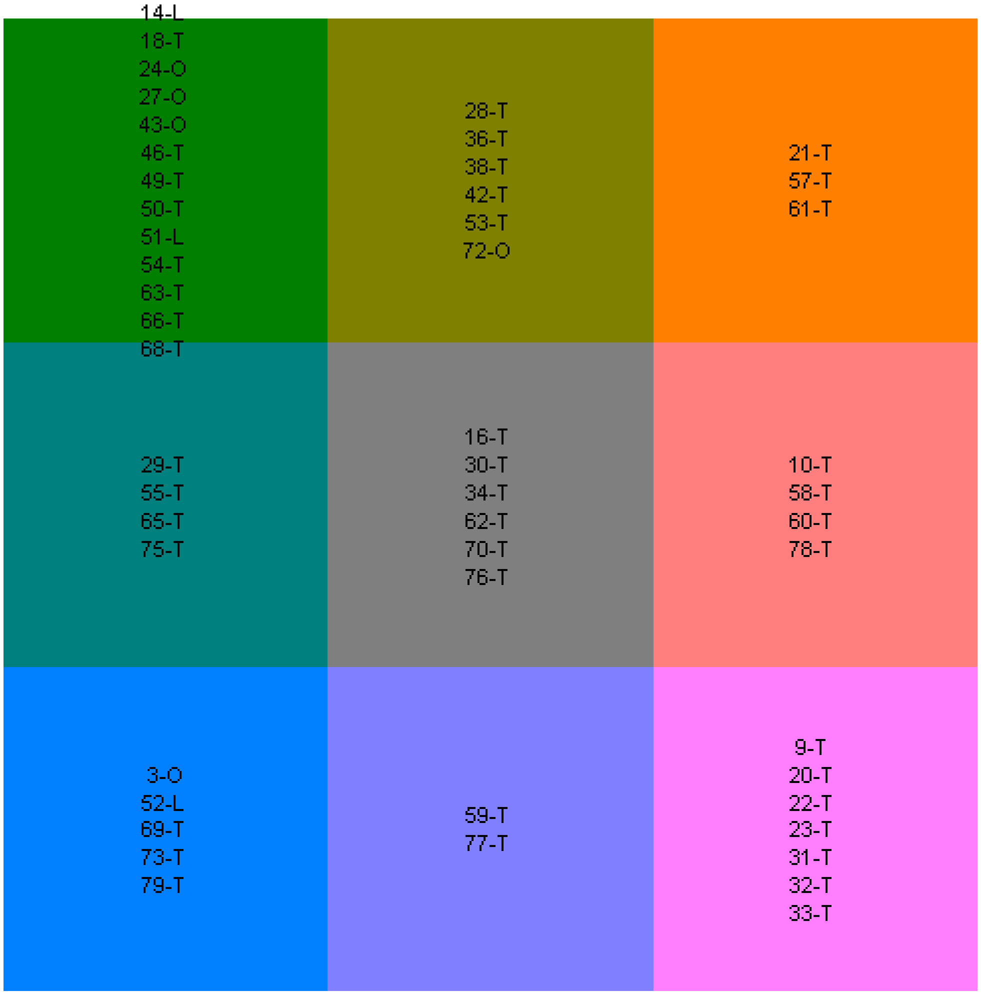


Figure 4.3.94. SOM model for MN Statewide 2003 Data based *Subset 2*.

ANALYSIS

Although the number of samples is pretty low however Figure 4.3.88 gives us an intuition that there may be a pattern of clustering suggesting that the samples of type Outwash (O) seem to cluster together. However we cannot make conclusions from it.

[**Goal 5d**] To estimate how the statewide samples are related based on Lacustrine, Till, Outwash (L/T/O) based on *Subset 3.*

DATASET USED: MN Statewide 2003 Soil Survey Data. Only the elements of the *Subset 3* are used in this case.

NAMING CONVENTION: Same as Goal 5a.

SOM MODELING
The SOM model for this goal is given below.

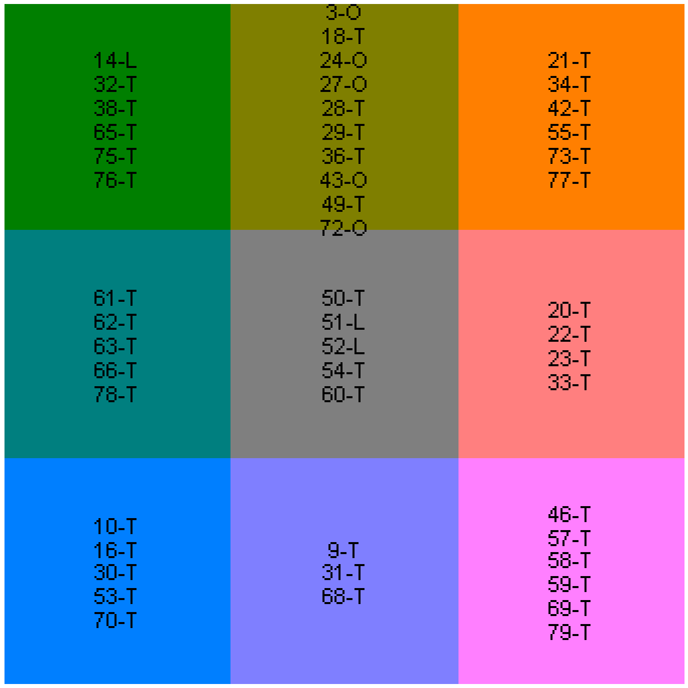


Figure 4.3.95. SOM model for MN Statewide 2003 Data based *Subset 4*.

ANALYSIS
For this case it is worth noticing that the samples of type O are all clustered into one single unit. This is rightly so, because the L/T/O encodes the soil composition, as such it may be expected that based on the soil characteristic being determined by the rock forming elements there may be some good pattern in the clustering. However, this could be confirmed with more soil samples.

[**Goal 5e**] To estimate how the statewide samples are related based on Lacustrine, Till, Outwash (L/T/O) based on *Subset 4.*

DATASET USED: MN Statewide 2003 Soil Survey Data. Only the elements of the *Subset 4* are used in this case.

NAMING CONVENTION: Same as Goal 5a.

SOM MODELING
The SOM model for this goal is given below.

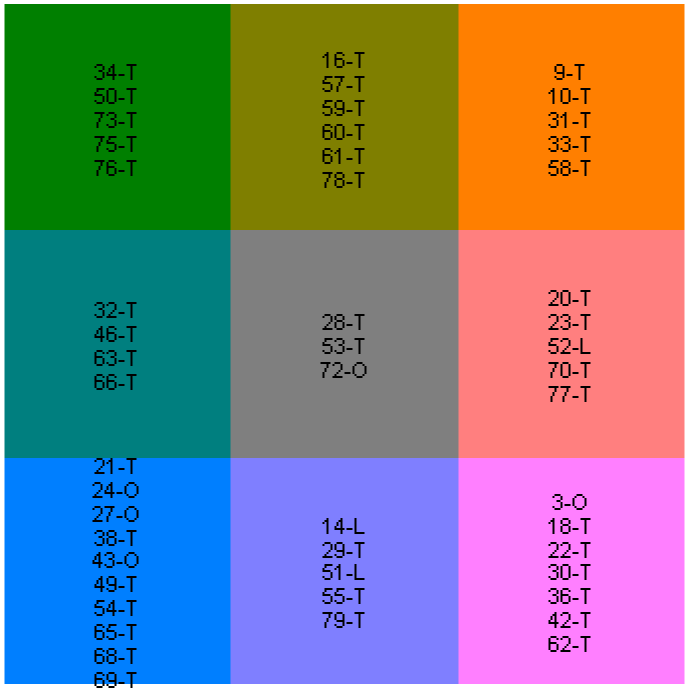


Figure 4.3.96. SOM model for MN Statewide 2003 Data based *Subset 5*.

ANALYSIS
It seems some of the samples of type Outwash cluster together. However we cannot make conclusions from this map.

<u>4.3.4 Summary of results for the MN Statewide 2003 Soil Survey Data</u>

Based on all the elements

- We observe that the samples belonging to the Rainy Superior bear same characteristic and the samples belonging to Des Moines and Wadena together bear same characteristic.
- The samples having a provenance of NE are clustered together and the samples having a provenance N& NW form a different cluster.
- Owing to the less number of samples of Type Lacustrine and Outwash it may not be possible to make any conclusions.

Based on *Subset 1* (element list provided by Dr. Robert Edstrom)

- We observe that the samples belonging to the Rainy Superior bear same characteristic and the samples belonging to Des Moines and Wadena together bear same characteristic.
- The samples having a provenance of NE are clustered together and the samples having a provenance N& NW form a different cluster.
- Owing to the less number of samples of type Lacustrine and Outwash it may not be possible to make any conclusions.

Based on *Subset 2* (elements of regulatory interest)

- We do not observe any pattern of clustering based on the geomorphologic codes.

- The samples having a provenance of NE are clustered together and the samples having a provenance N& NW form a different cluster.
- Owing to the less number of samples of Type Lacustrine and Outwash it may not be possible to make any conclusions.

Based on *Subset 3*(Rock Forming Elements)

- We observe that the samples belonging to the Rainy Superior bear same characteristic and the samples belonging to Des Moines and Wadena together bear same characteristic.
- The samples having a provenance of NE are clustered together and the samples having a provenance N& NW form a different cluster.
- Although it is very interesting to see that all the samples of type Outwash are clustered onto a single unit. However owing to the less number of samples of Type Lacustrine and Outwash it may not be possible to make any conclusions.

Based on *Subset 4* (list of Non Metals)

- There is no pattern of clustering based on the Geomorphologic codes.
- There is no pattern of clustering based on the Provenance.
- Owing to the less number of samples of Type Lacustrine and Outwash it may not be possible to make any conclusions.

Results pertaining to the metals of regulatory interests

ELEMENT=Arsenic (As)

- We do not observe a good pattern of clustering based on the geomorphologic codes.
- We do not observe a good pattern of clustering based on the provenance of the samples.
- We observe some pattern of clustering that is mostly geographic. For instance, the samples that are collected at the North-East of the State of MN seem to have higher concentration values for Arsenic and cluster together.

ELEMENT= Chromium (Cr)

- We do not observe a good pattern of clustering based on the geomorphologic codes.
- We do not observe a good pattern of clustering based on the provenance of the samples.
- We observe some pattern of clustering that is mostly geographic. For instance, the samples that are collected at the North-East of the State of MN seem to have higher concentration values for Chromium and cluster together.

ELEMENT=Copper (Cu)

- We do not observe a good pattern of clustering based on the geomorphologic codes.
- We do not observe a good pattern of clustering based on the provenance of the samples.
- We observe some pattern of clustering that is mostly geographic. For instance, the samples that are collected at the North-East of the State of MN seem to have higher concentration values for Copper and cluster together.

ELEMENT= Nickel (Ni)

- We do not observe a good pattern of clustering based on the geomorphologic codes.
- We do not observe a good pattern of clustering based on the provenance of the samples.
- We observe some pattern of clustering that is mostly geographic. For instance, the samples that are collected at the North-East of the State of MN seem to have higher concentration values for Nickel and cluster together.

ELEMENT= Lead (Pb)

- We do not observe a good pattern of clustering based on the geomorphologic codes.

- We do not observe a good pattern of clustering based on the provenance of the samples.
- We do not observe any pattern of clustering based on the geographical location of the soil samples.

ELEMENT= Tungsten (W)

- We do not observe a good pattern of clustering based on the geomorphologic codes.
- We do not observe a good pattern of clustering based on the provenance of the samples.
- We observe some pattern of clustering that is mostly geographic. For instance, the samples that are collected at the North-East of the State of MN seem to have higher concentration values for Tungsten and cluster together.

ELEMENT=Zinc (Zn)

- We do not observe a good pattern of clustering based on the geomorphologic codes.
- We do not observe a good pattern of clustering based on the Provenance of the samples.
- We do not observe any pattern of clustering based on the geographical location of the soil samples.

4.3.5 Results of SOM modeling for Metro 2001 Soil Survey Data and MN Statewide 2003 Soil Survey Data.

[**Goal 6a**] To estimate how the Metro samples are related to the statewide samples and background samples based on all elements.

DATASET USED: Metro 2001 Soil Survey Data and MN Statewide 2003 Soil Survey Data. In this case both the datasets are merged. Only the elements that are common to both the datasets are retained. These elements are: P10, Silt, Clay, S&G, OM, PH, EP, EK, As, Br, Ce, Co, Cr, Cs, Fe, La, Mn, Mo, Ni, Pb, Rb, Sb, Sc, Se, Sm, Sr, Th, U, V, W, Zn, Cu. Further for the elements Ti, Al, Fe, Mg, Ca, Na, K, and Mn for the 2003 data the concentration values are reported in percentage. So, we convert the values using the following conversion formula: 1% = 10,000 ppm.

NAMING CONVENTION: For this case we use similar naming convention as in Goal 2a (for Metro 2001 Soil Survey Data) and Goal 3a (for MN Statewide 2003 Soil Survey Data). The only change is that we tag an identifier '-2001' for the 2001 dataset and '-2003' for the 2003 dataset.

For example:

A sample with its tag as **I-4-A-2001** would mean

Circle = Circle I (Downtown).
Site ID (for the sample) = 4.
Sample-Type= Type A.
Soil dataset= Metro 2001 Soil Survey Data.

A sample with its tag as **63(3) -2003** encodes

Site ID (for the sample) = 63.
Geomorphologic category= Rainy Superior (3).

Soil dataset= MN Statewide 2003 Soil Survey Data.

SOM MODELING

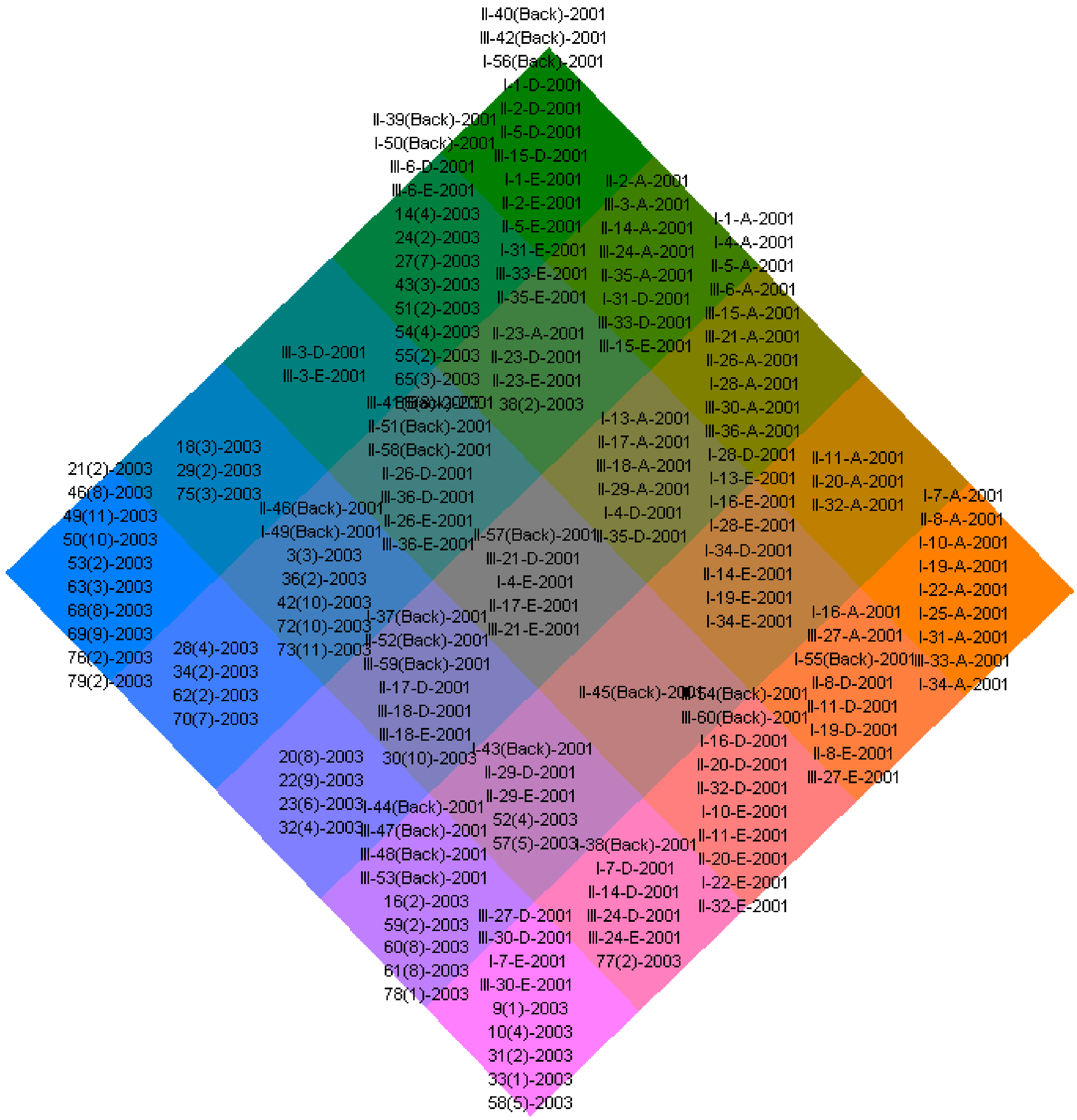


Figure 4.3.97. SOM model for both 2001 and 2003 Data using all the elements.

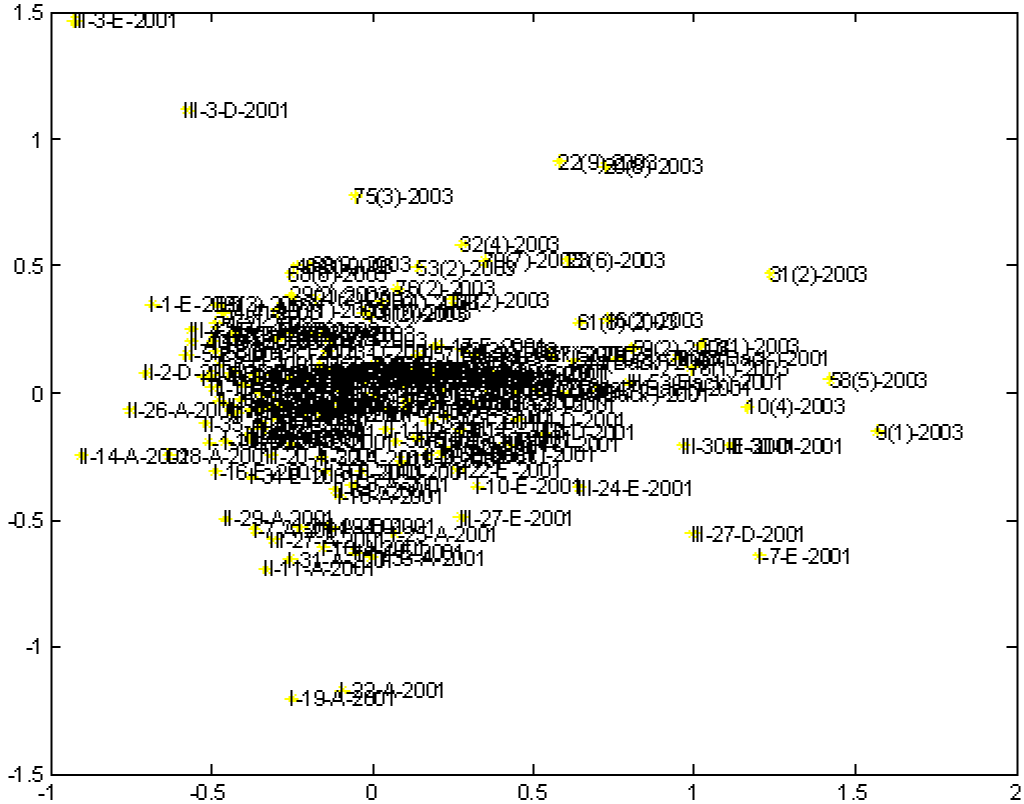


Figure 4.3.98. MDS output for both 2001 and 2003 Data using all the elements.

ANALYSIS

From the Figure 4.3.97 we can see three distinct patterns of clustering. It seems like for the Metro 2001 Soil Survey Data the soil samples of Type A are clustered together and the soil samples of Type D& E form a different cluster. Further it seems like the MN Statewide 2003 Soil Survey Data samples form a different cluster. In fact, when we see the Figure 4.3.98 we can observe that the samples of Type D and E are in between the two different clusters formed by samples of Type A and the statewide samples. Moreover the background samples seem to be clustered along with the samples of Type D and E and Statewide 2003 samples. This suggests that the samples of Type A have similar characteristics and are quite different than the Statewide 2003 samples. Further, it seems like the samples of Type D & E have characteristics that acts somewhat as transition between the samples of Type A and the Statewide 2003 samples. Moreover the background samples seem to be related more to the samples of Type D&E and Statewide 2003 samples rather than samples of Type A.

[**Goal 6b**] To estimate how the Metro samples are related to the statewide samples and background samples based on *Subset 1*.

DATASET USED: Same as Goal 6a. However in this case we use only the elements from the *Subset 1*.

NAMING CONVENTION: Same as in Goal 6a.

SOM MODELING

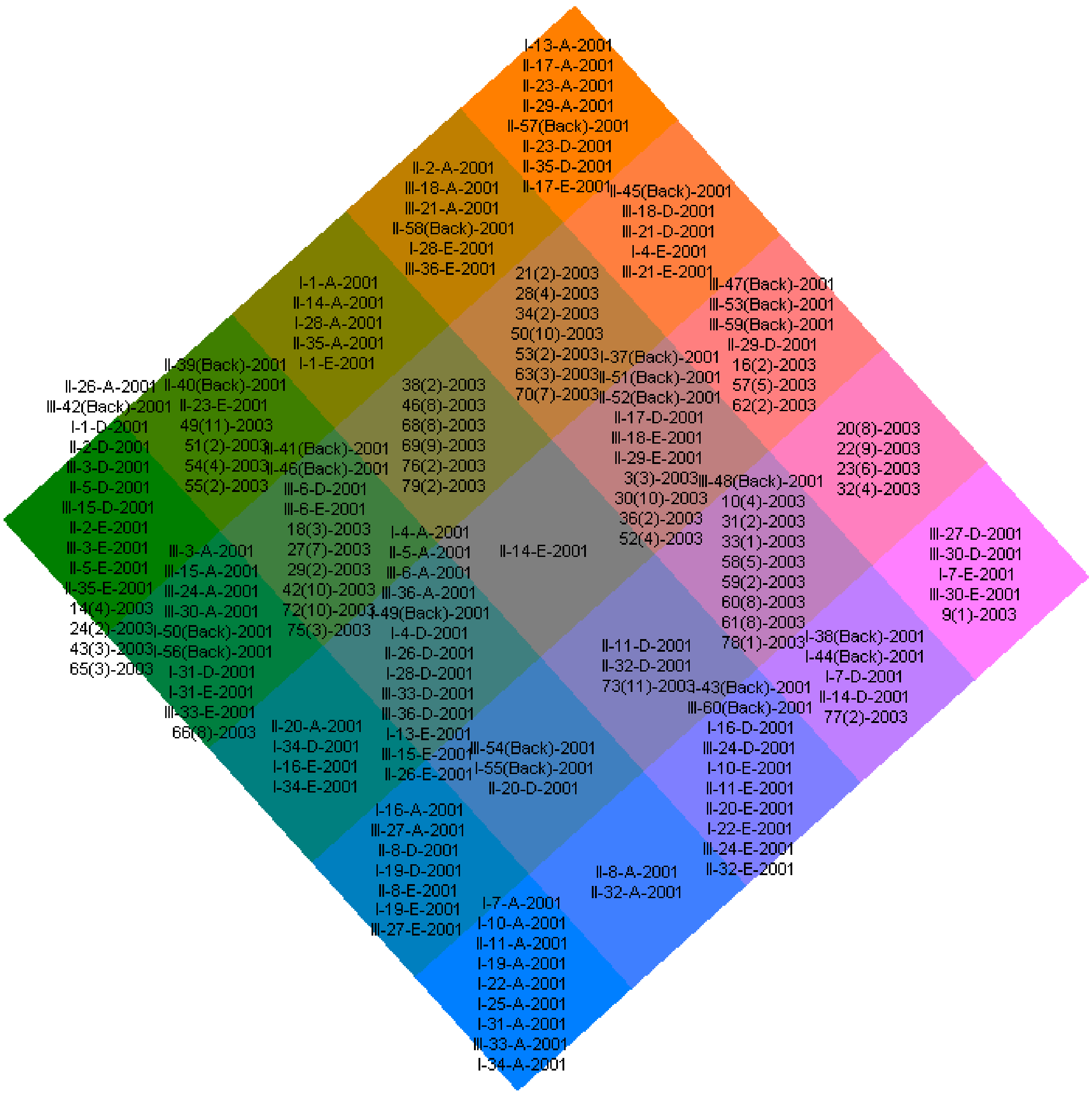


Figure 4.3.99. SOM model for both 2001 and 2003 Data using *Subset 1*.

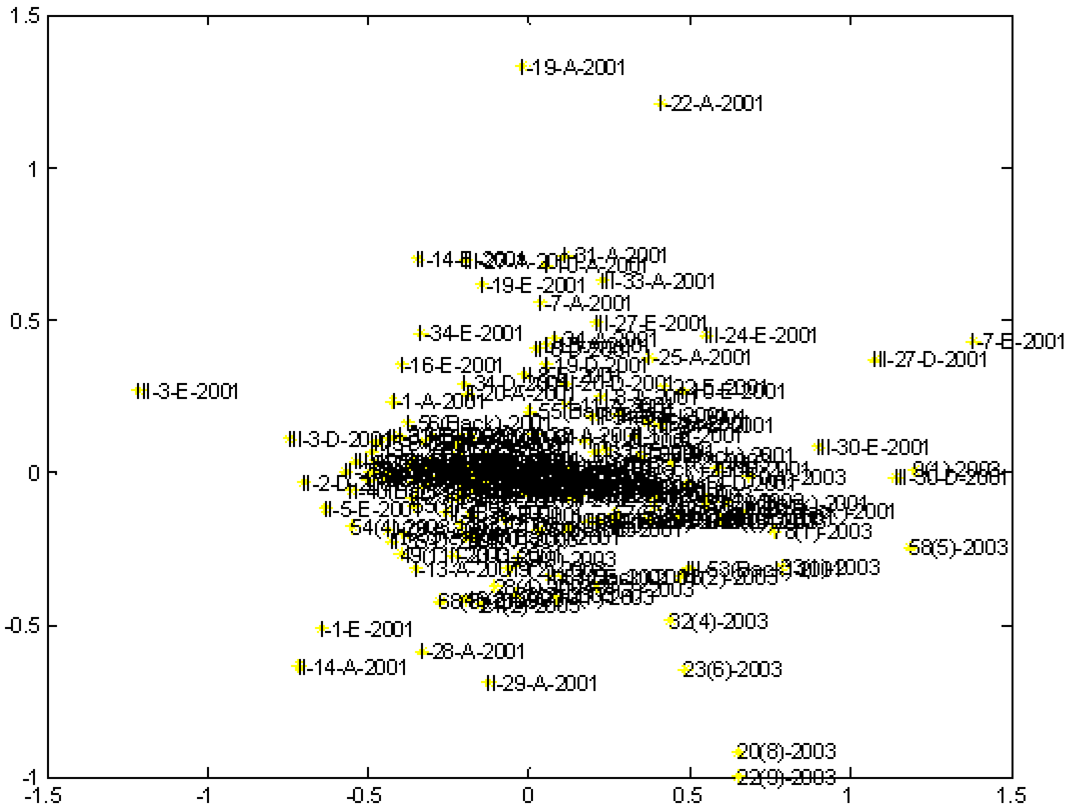


Figure 4.3.100. MDS output for both 2001 and 2003 Data using *Subset 1*.

ANALYSIS

In this case based on Figure 4.3.99 we do not observe a very good pattern of clustering. However based on Figure 4.3.100 we observe a pattern of clustering where we observe that the Metro 2001 samples seem to cluster together and some of the Statewide 2003 samples form a different cluster. Thus, there seems to be a possible clustering of the Metro 2001 samples against the Statewide 2003 samples.

[**Goal 6c**] To estimate how the Metro samples are related to the statewide samples and background samples based on *Subset 2*.

DATASET USED: Same as Goal 6a. However in this case we use only the elements from the *Subset 2*.

NAMING CONVENTION: Same as in Goal 6a.

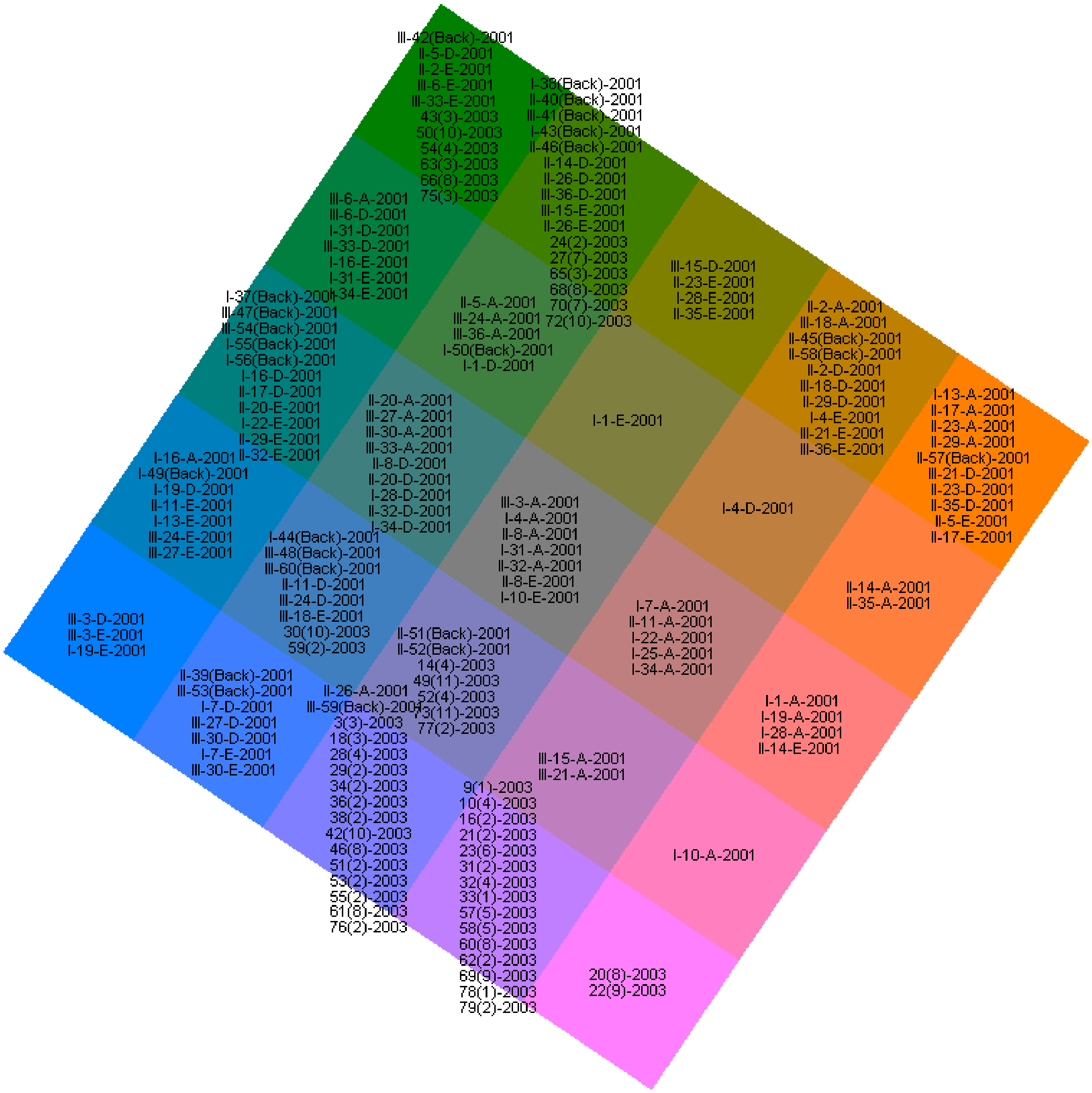


Figure 4.3.101. SOM model for both 2001 and 2003 Data using *Subset 2*.

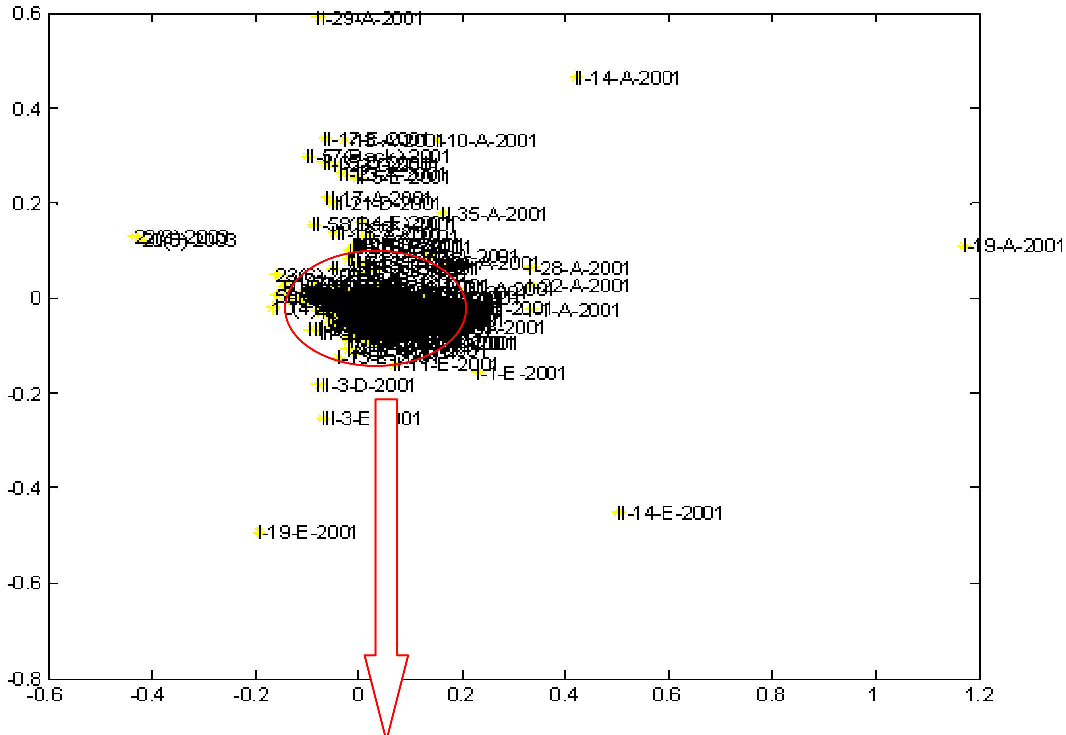


Figure 4.3.102 (a). MDS output for both 2001 and 2003 Data using *Subset 2*.

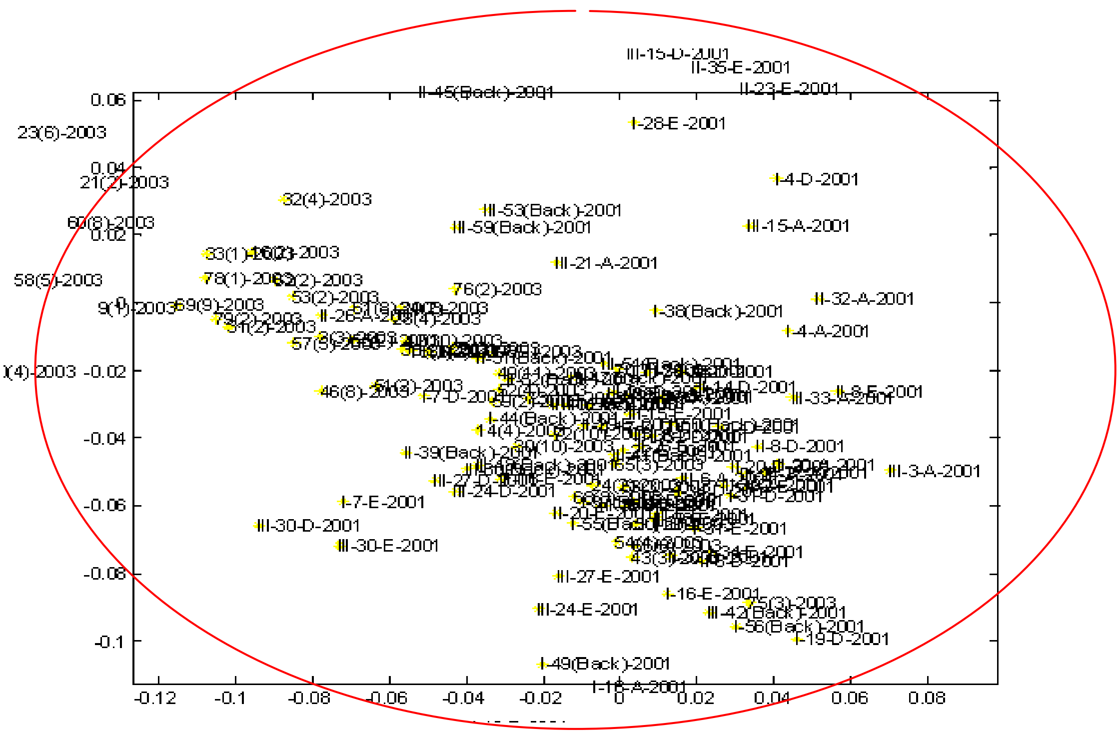


Figure 4.3.102 (b). MDS output zoomed in for the purpose of explanation.

ANALYSIS

From the Figure 4.3.101 we observe that there is some clustering of the Metro 2001 samples vs. the Statewide 2003 samples. In this case the pattern of clustering in not conspicuous from Figure 4.3.102a. However if we zoom in (as shown in Figure 4.3.102b) we do observe that the Statewide 2003 samples are clustered together in the left and the Metro 2001 samples form a different cluster onto the right. This suggests that the Statewide 2003 samples are more related to each other than the Metro 2001 soil samples.

Further, in order to understand the relation between the soil samples in more details we further do the SOM modeling by taking each one of the element separately.

ELEMENT=Arsenic (AS)

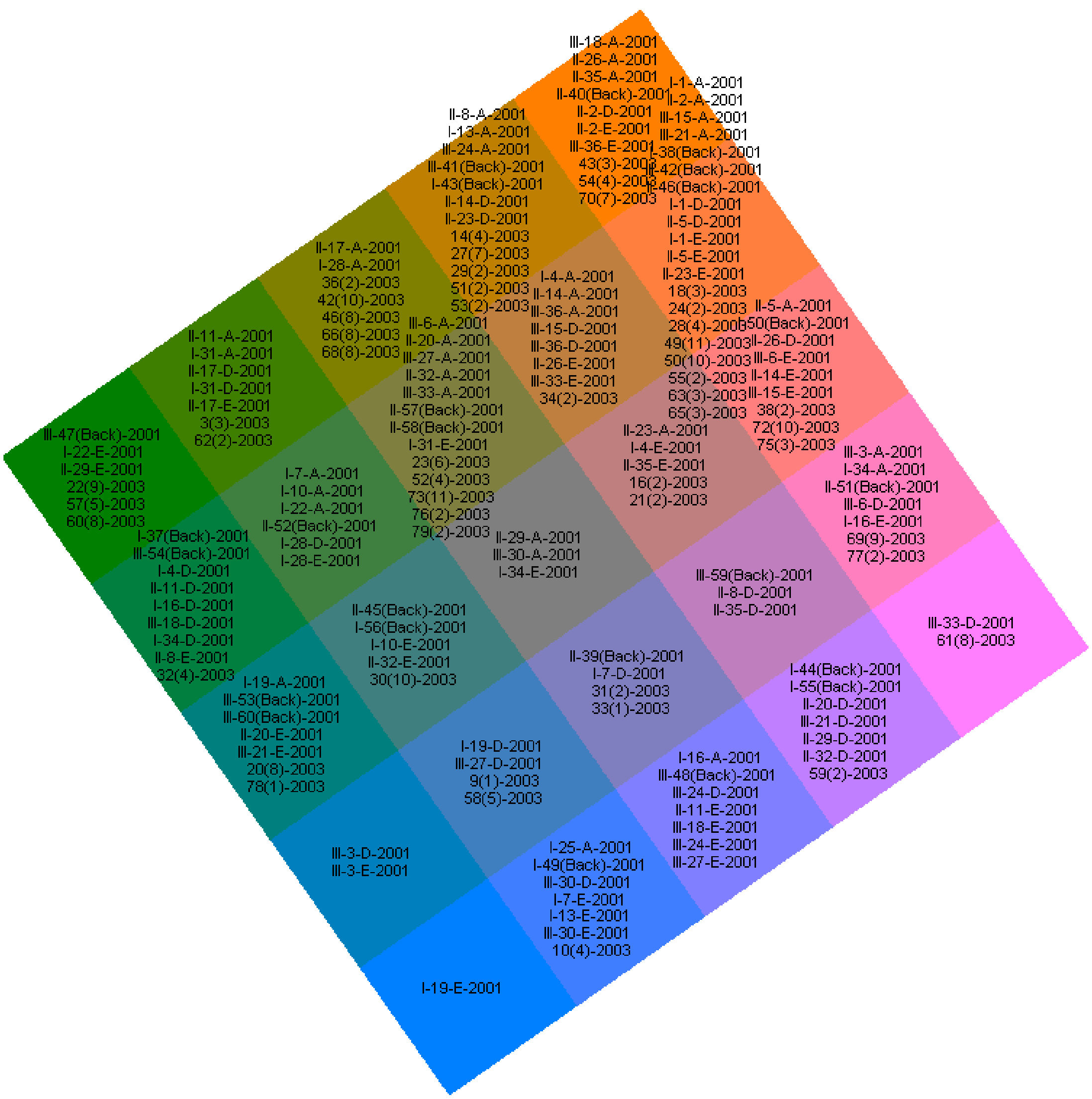


Figure 4.3.103. SOM model for both 2001 and 2003 Data using Arsenic (As).

ANALYSIS
From Figure 4.3.103 we do not observe any good pattern of clustering.

ELEMENT=Chromium (Cr)

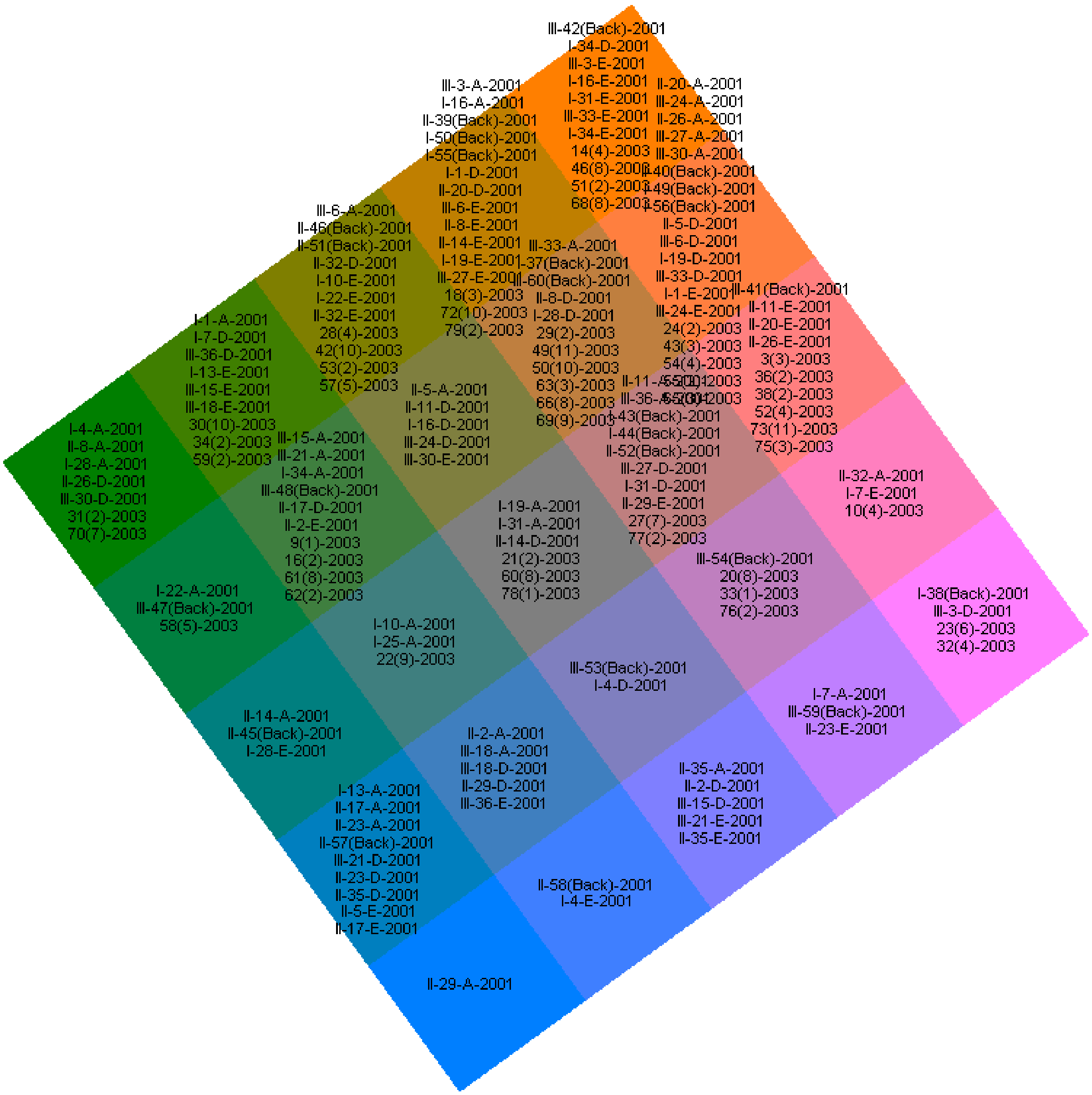


Figure 4.3.104. SOM model for both 2001 and 2003 Data using Chromium (Cr).

ANALYSIS
From Figure 4.3.104 we do not observe any good pattern of clustering.

ELEMENT=Copper (Cu)

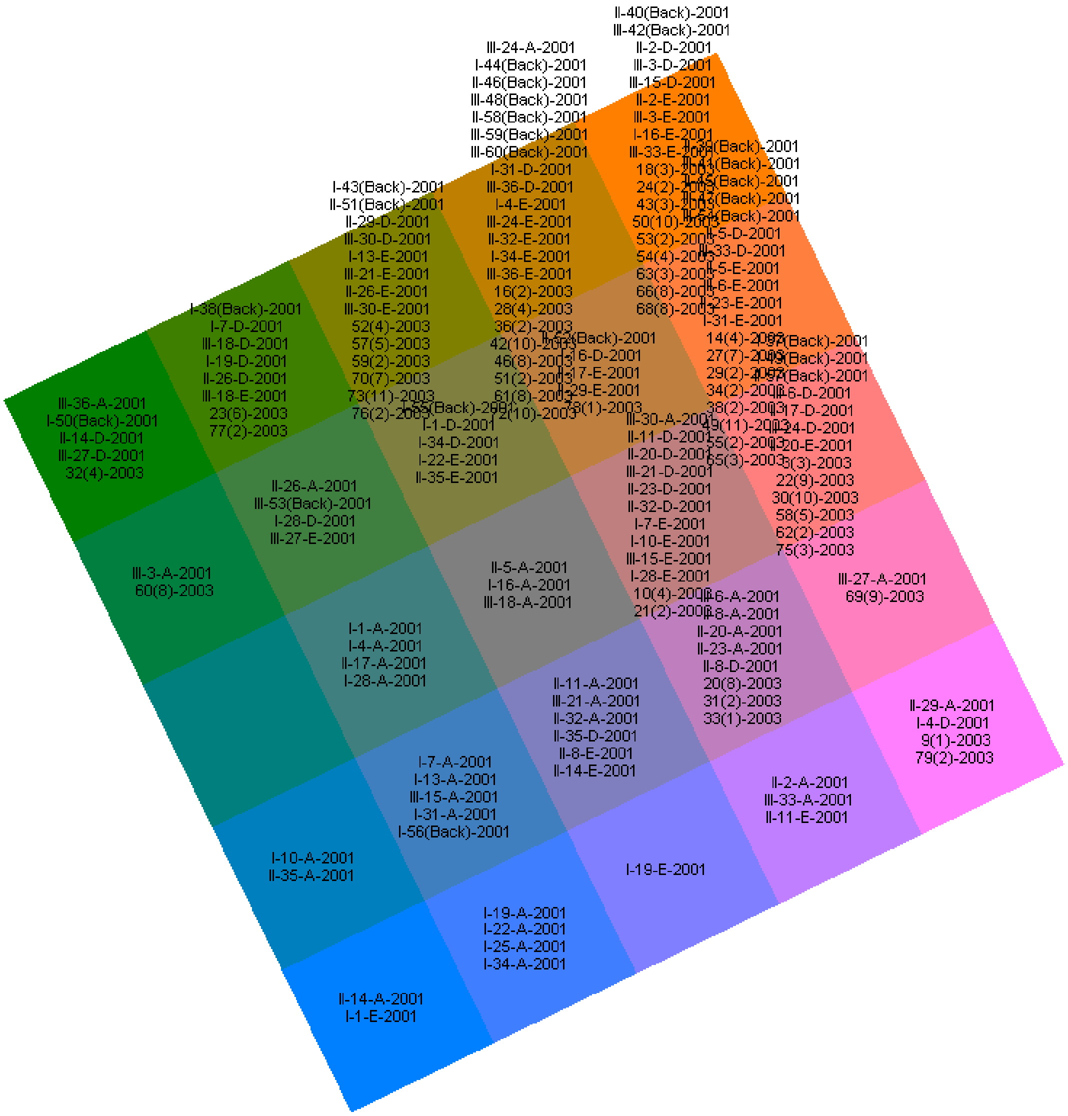


Figure 4.3.105. SOM model for both 2001 and 2003 Data using Copper (Cu).

ANALYSIS

From Figure 4.3.105 we observe that the samples of Type A are all mostly clustered in the bottom of the map. Moreover it seems like the background samples are clustered with the samples of Type D&E or the Statewide 2003 samples. Thus we can say that the background samples are mostly related to the samples of Type D&E or the statewide samples than the Metro samples that are collected near the Mn/DOT roads i.e. Type A samples . In fact the Metro 2001 samples of Type A seem to form a different cluster. We have checked that these Metro 2001 samples of Type A have relatively high Cu concentration values. Thus we can conclude that the samples of Type A form a different cluster than the other samples and are quite different from the Metro 2001 Type D&E samples and the Statewide 2003 samples.

ELEMENT= Lead (Pb)

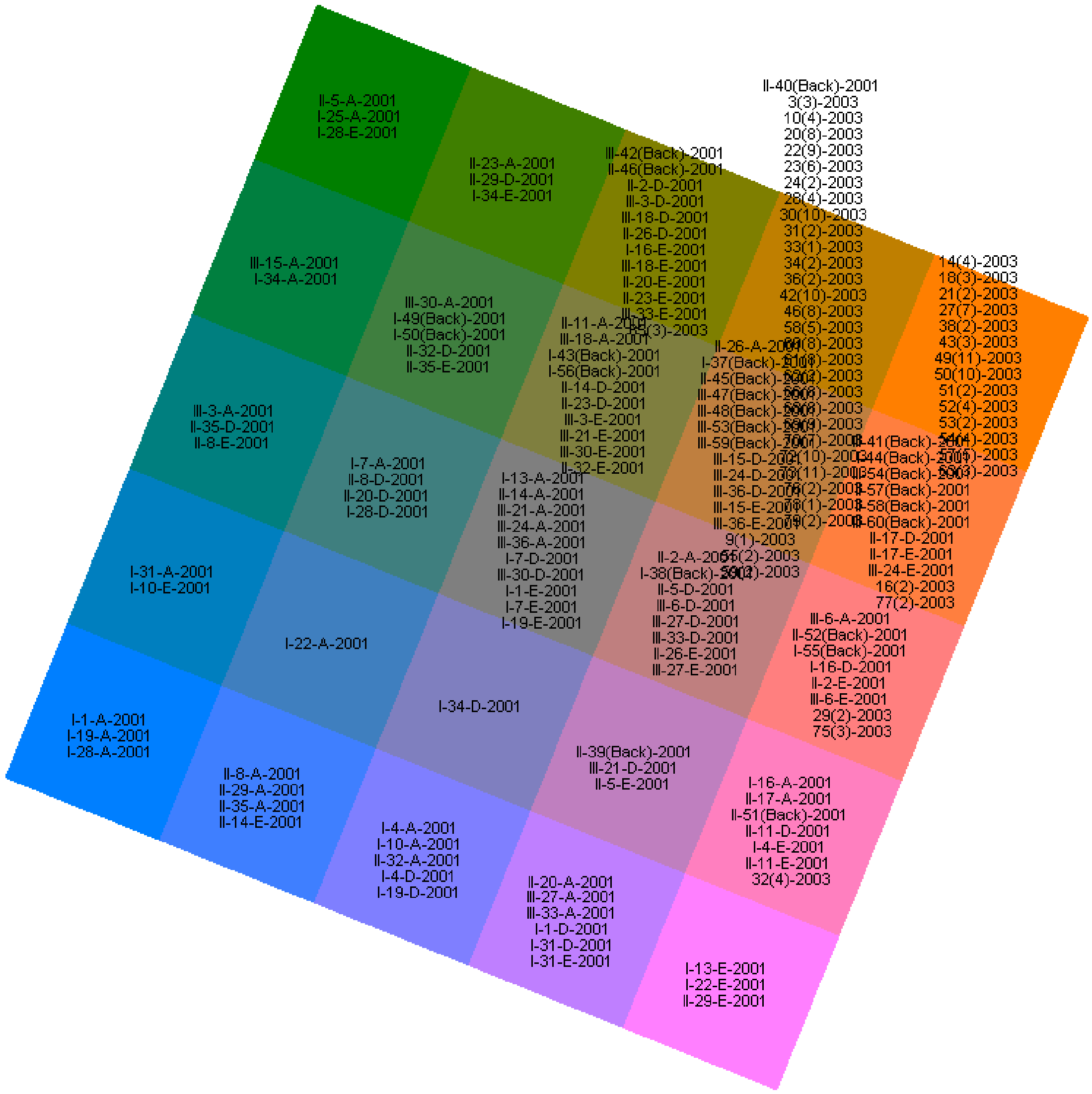


Figure 4.3.106. SOM model for both 2001 and 2003 Data using Lead (Pb).

ANALYSIS

From Figure 4.3.106 we observe that the samples of Type A are all mostly clustered together. Moreover it seems like the background samples are clustered with the samples of Type D&E and the Statewide 2003 soil samples form a different cluster. Thus it seems like the soil samples of Type A have similar Lead concentration values. The Metro 2001 Type D& E soil samples and the background samples have similar Pb concentration values. And the Statewide 2003 soil samples have similar Lead concentration values. We have also checked that the concentration value of Lead (Pb) for Type A samples is high.

ELEMENT=Nickel (Ni)

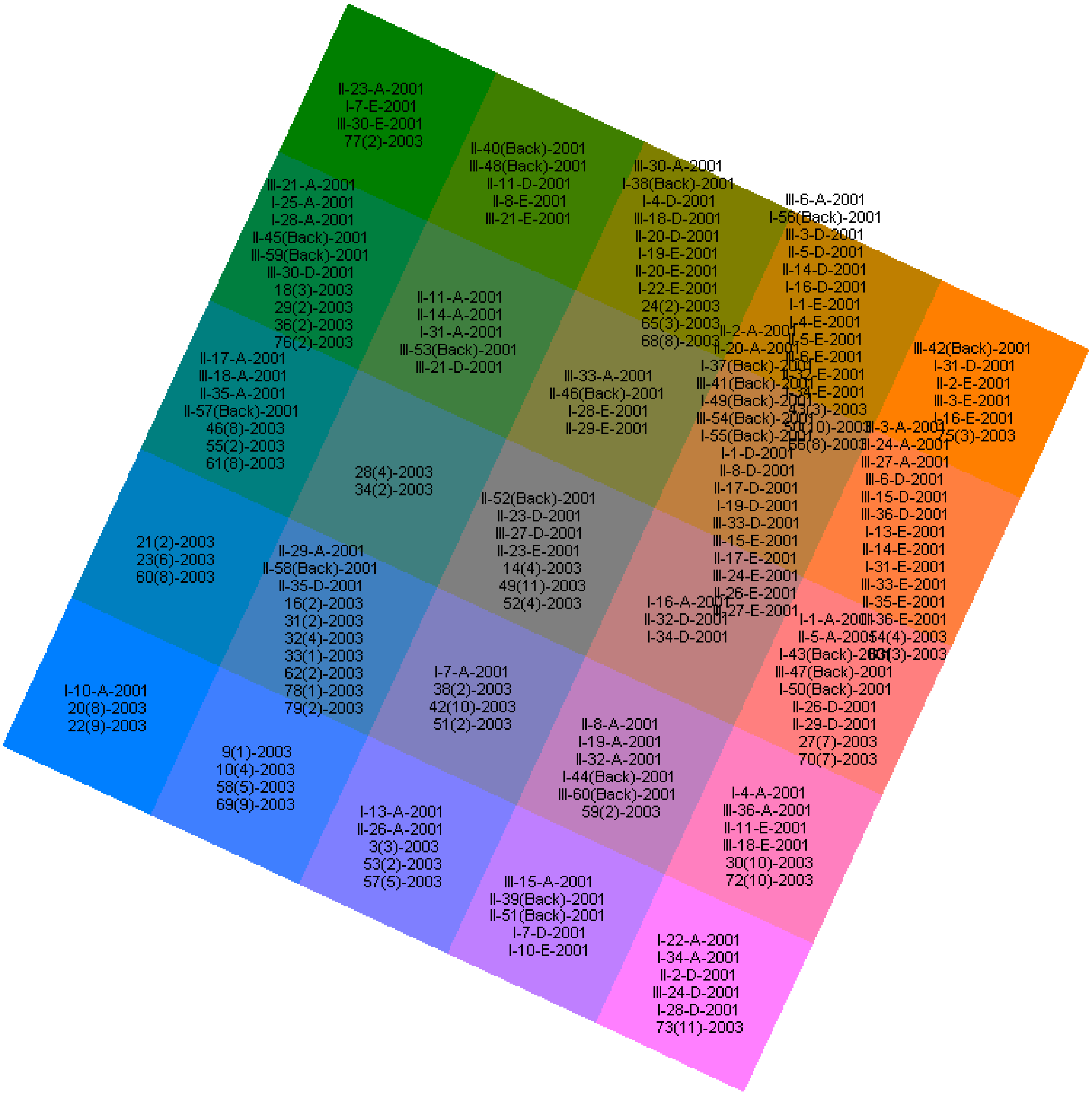


Figure 4.3.107. SOM model for both 2001 and 2003 Data using Nickel (Ni).

ANALYSIS

From Figure 4.3.107 we observe that the Statewide 2003 soil samples are clustered together and the Metro 2001 samples are clustered differently. Moreover we observe that the background samples are clustered together with the Metro 2001 samples. We have also checked that the concentration value of Nickel (Ni) for Statewide 2003 samples is high.

ELEMENT=Tungsten (W)

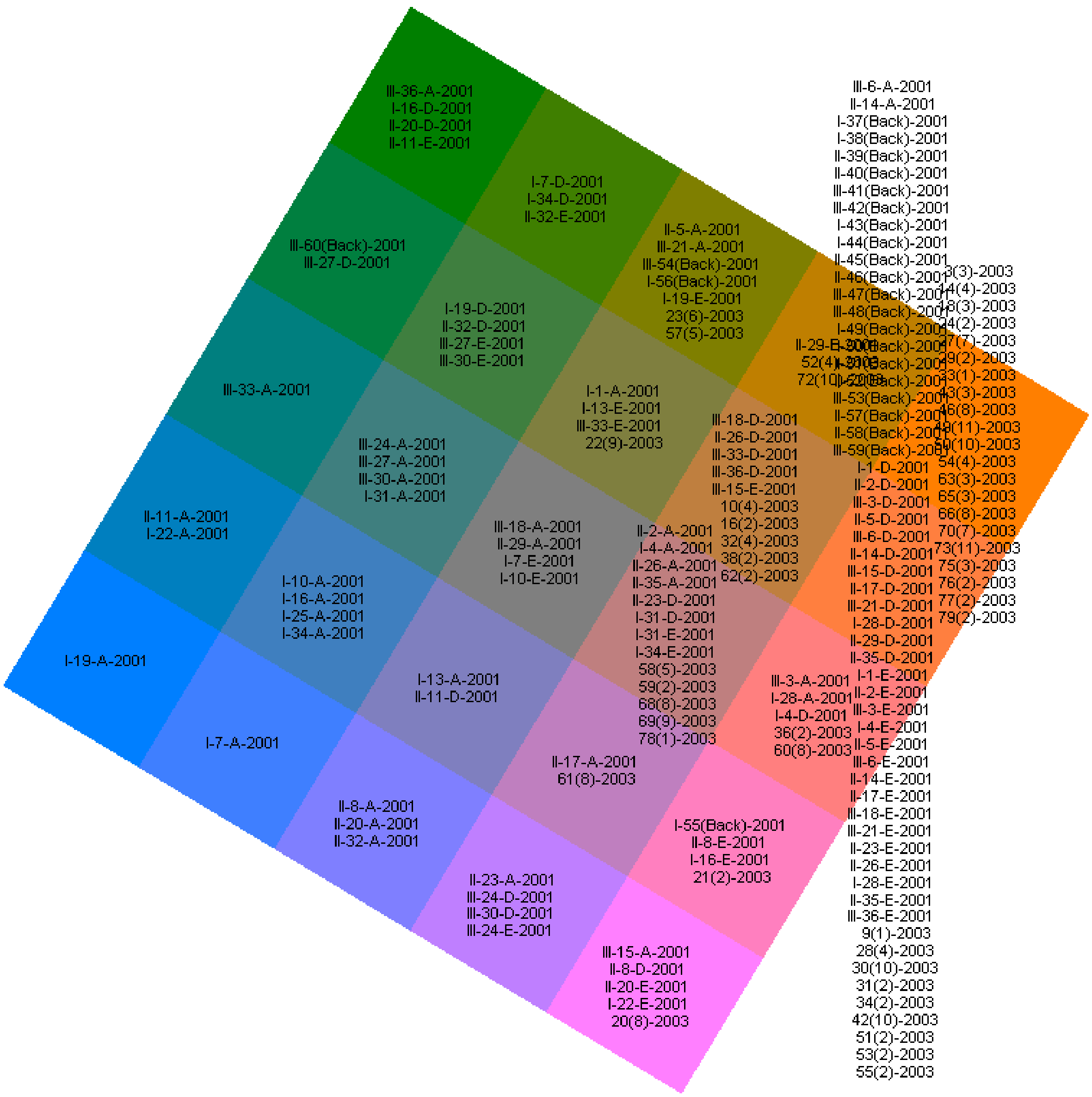


Figure 4.3.108. SOM model for both 2001 and 2003 Data using Tungsten(W).

ANALYSIS

From Figure 4.3.108 we observe that the Metro 2001 samples of Type A are clustered together. Further it seems like the background samples, Metro 2001 (Type D&E) and Statewide 2003 samples form a different cluster. We have also checked that the concentration value of Tungsten for Metro 2001 samples of Type A samples is high.

ELEMENT=Zinc (Zn)

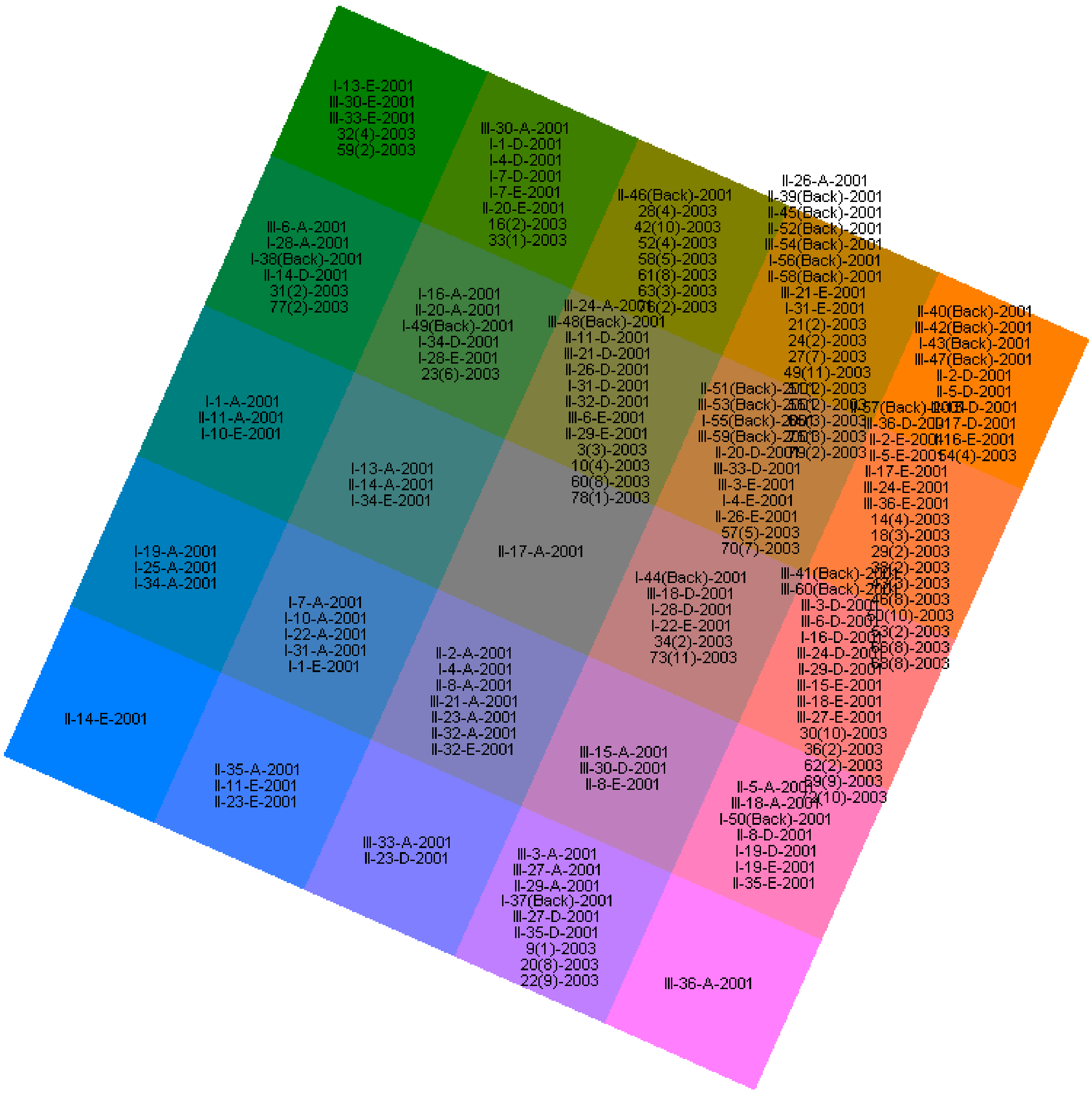


Figure 4.3.109. SOM model for both 2001 and 2003 Data using Zinc (Zn).

ANALYSIS

From Figure 4.3.109 we observe that the Metro 2001 samples of Type A are clustered on the left of the SOM map. Further it seems like the background samples, Metro 2001 (Type D&E) and Statewide 2003 samples form a different cluster.

[**Goal 7a**] To estimate how the Metro samples of Type D& E are related to the statewide samples and background samples based on all elements.

DATASET USED: Same as Goal 6a except that in this case we remove all the samples of Type A.

NAMING CONVENTION: Same as Goal 6a.

For example:
A sample with its tag as **I-19-D-2001** would mean
Circle = Circle I (Downtown).
Site ID (for the sample) = 19.
Sample-Type= Type D.
Soil dataset= Metro 2001 Soil Survey Data.

A sample with its tag as **63(3) -2003** encodes
Site ID (for the sample) = 63.
Geomorphologic category= Rainy Superior (3).
Soil dataset= MN Statewide 2003 Soil Survey Data.

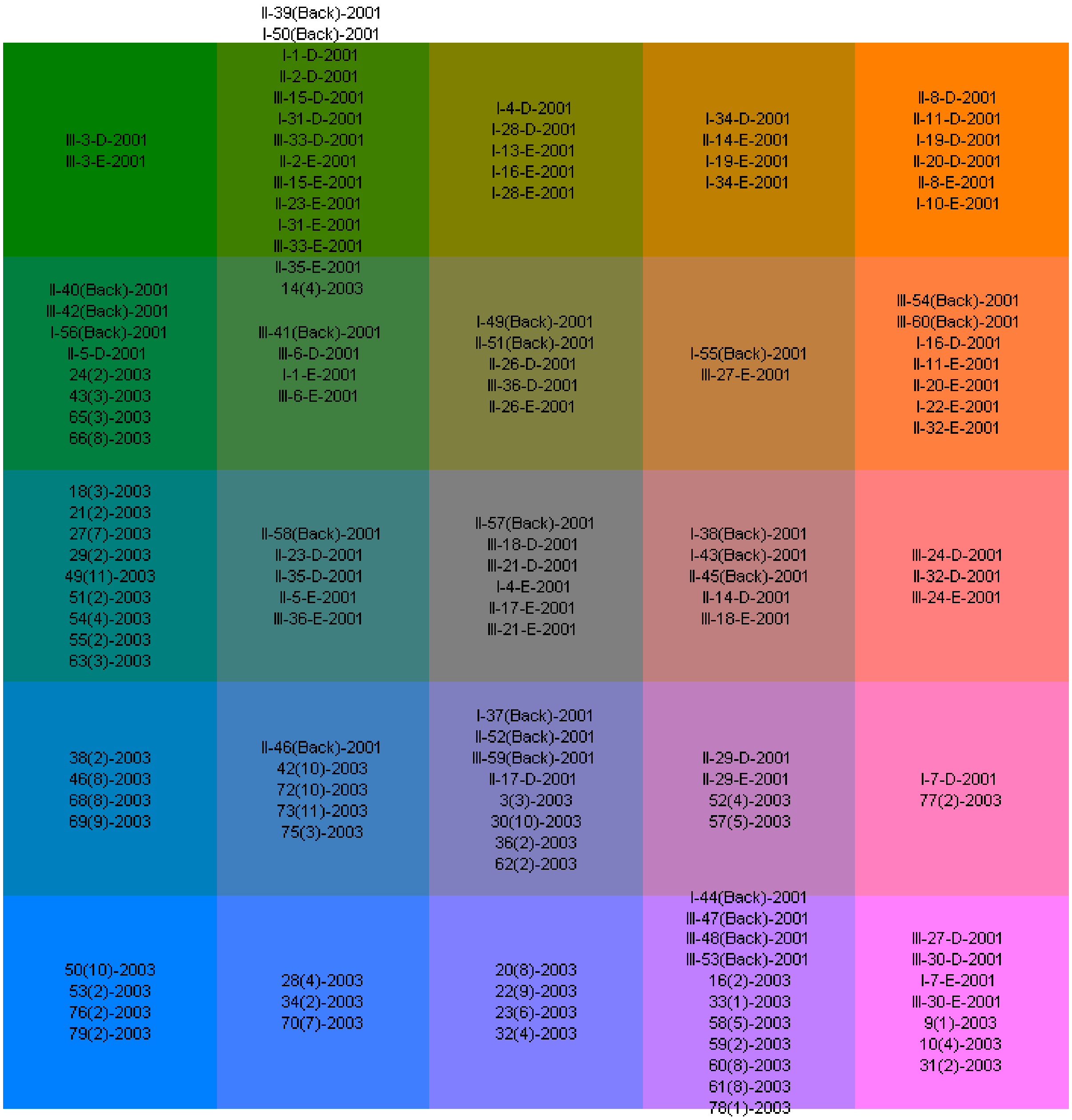


Figure 4.3.110. SOM model for both 2001 and 2003 Data using all the elements (without Type A samples).

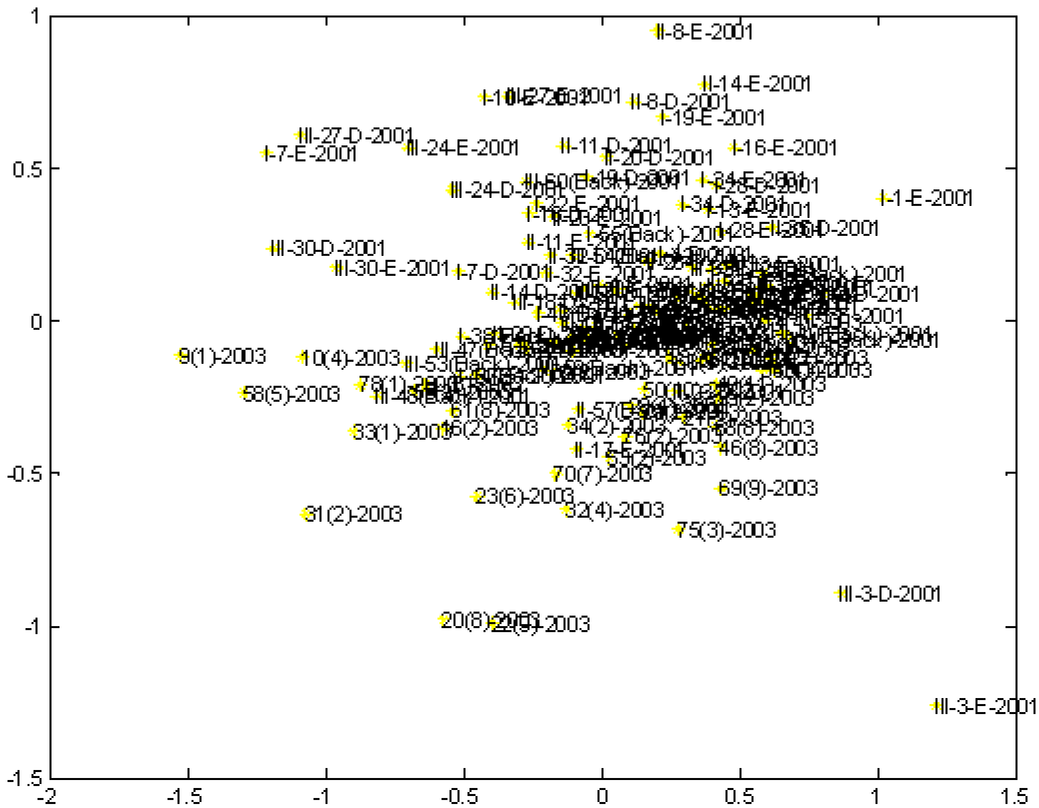


Figure 4.3.111. MDS output for both 2001 and 2003 Data using all the elements (without Type A samples).

ANALYSIS

From Figure 4.3.110 and Figure 4.3.111 we can see that the samples of Type D& E are clustered together and the Statewide 2003 samples form a different cluster. There does not seem to be any distinct pattern of clustering for the background samples. This indicates that the Type D & E samples are closely related to each other and the statewide samples are closely related to each other. The background samples seem to have the soil characteristics which is somewhat related to samples of Type D and E in some places and to the statewide samples at some other places.

[**Goal 7b**] To estimate how the Metro samples of Type D& E are related to the statewide samples and background samples based on the *Subset 1*.

DATASET USED: Same as Goal 6a except that in this case we remove all the samples of Type A. Moreover we use only the elements of *Subset 1*.

NAMING CONVENTION: Same as Goal 6a.

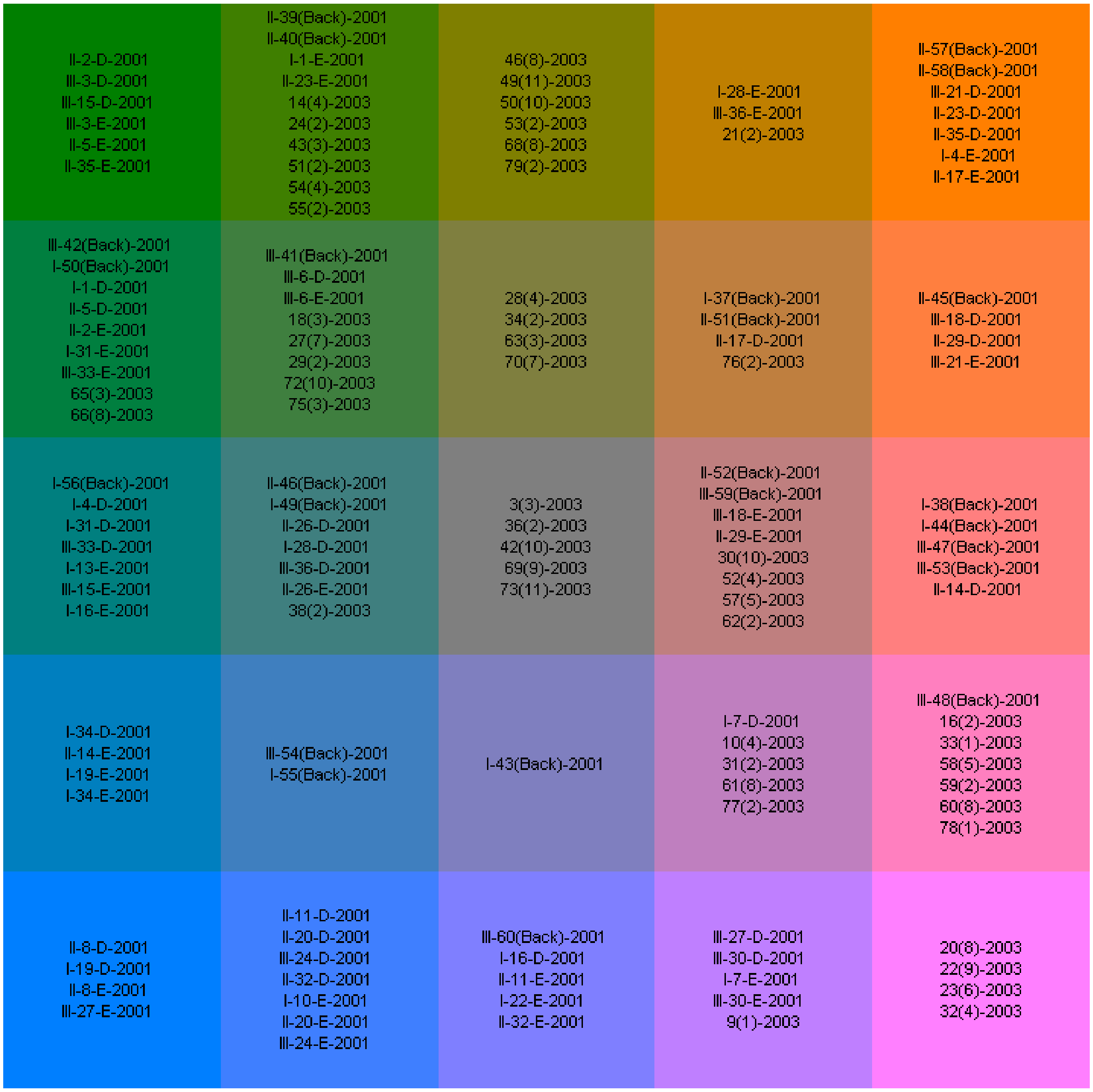


Figure 4.3.112. SOM model for both 2001 and 2003 Data using the *Subset 1* (without Type A samples).

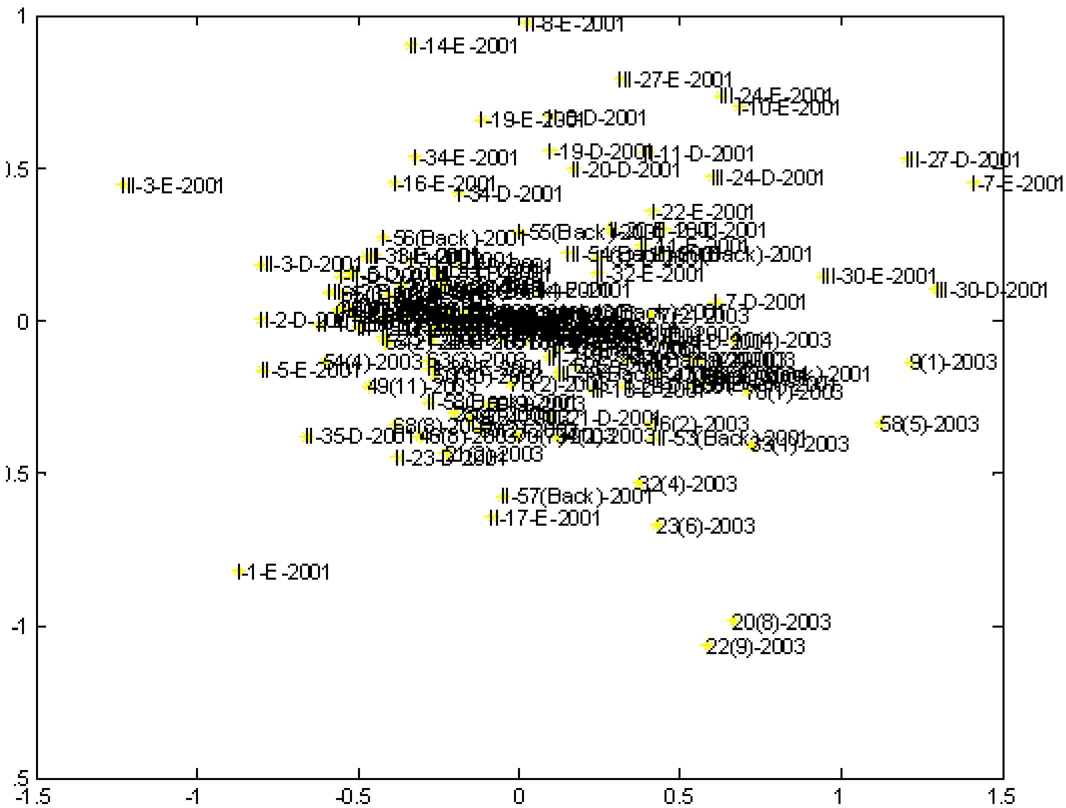


Figure 4.3.113. MDS output for both 2001 and 2003 Data using the *Subset 1* (without Type A samples).

ANALYSIS

From Figure 4.3.112 we observe that most of the Statewide 2003 samples are clustered along the diagonal of the SOM map. Further we also see from Figure 4.3.113 that the Metro 2001 samples are all clustered on the top and the Statewide 2003 samples are clustered on the bottom and in the middle. This suggests that the Statewide 2003 samples are all clustered together and the Metro 2001 samples are scattered outward and display a similar soil characteristic.

[**Goal 7c**] To estimate how the Metro samples of Type D& E are related to the statewide samples and background samples based on the Subset 2.

DATASET USED: Same as Goal 6a except that in this case we remove all the samples of Type A. Moreover we use only the elements of Subset 2.

NAMING CONVENTION: Same as Goal 6a.

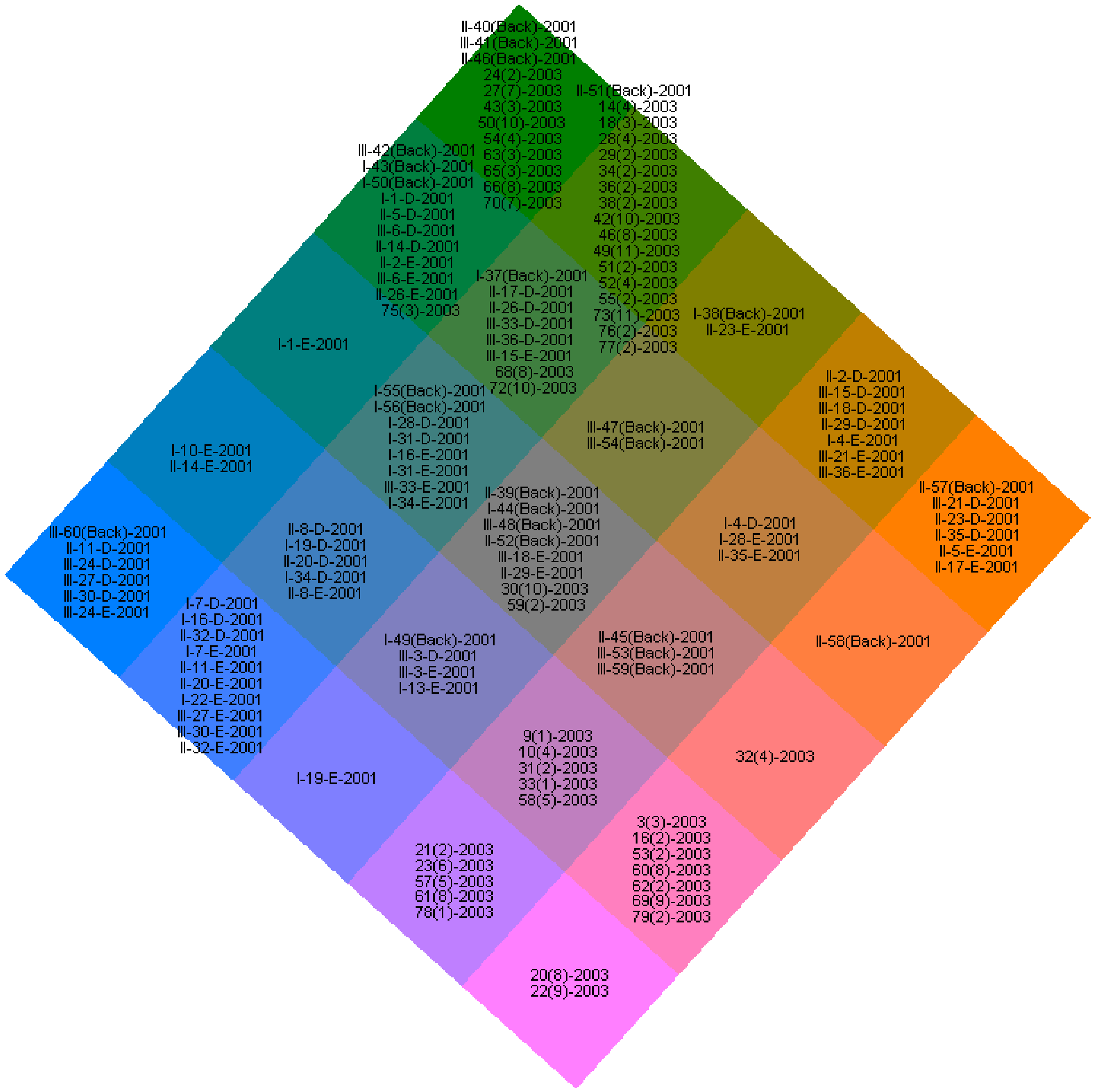


Figure 4.3.114. SOM model for both 2001 and 2003 Data using the *Subset 2* (without Type A samples).

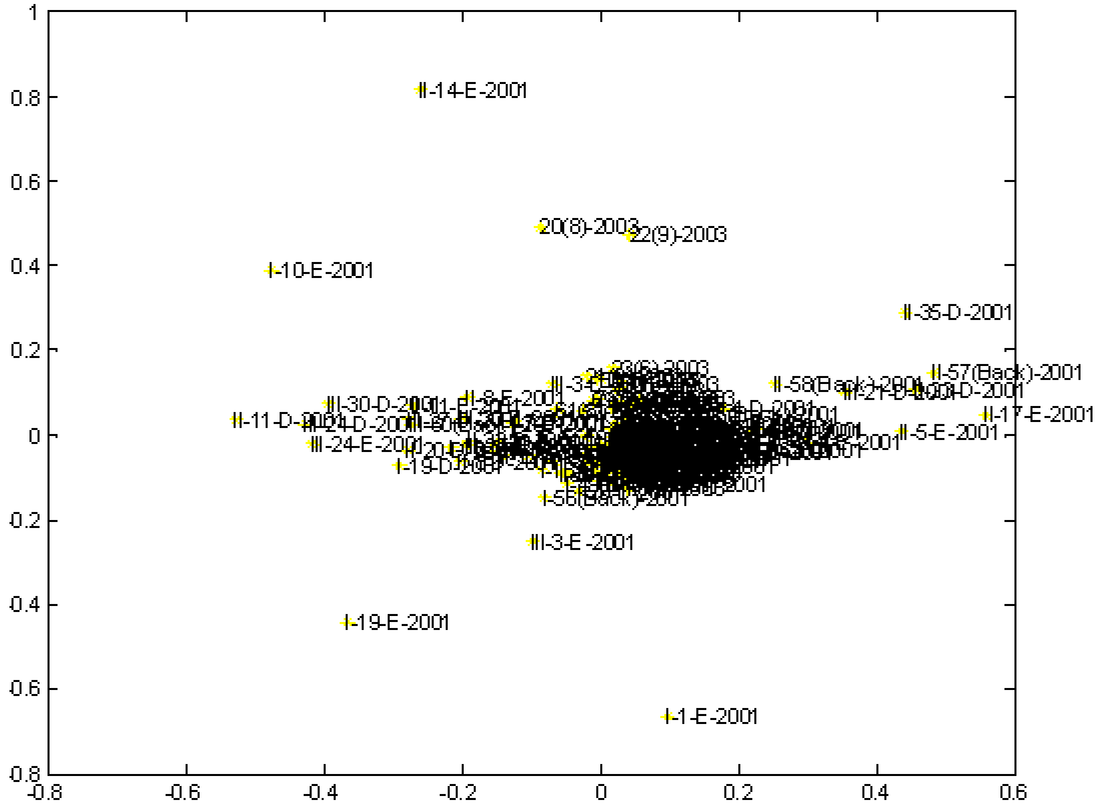


Figure 4.3.115. MDS output for both 2001 and 2003 Data using the *Subset 2* (without Type A samples).

ANALYSIS

From Figure 4.3.114 and Figure 4.3.115 we also observe that the Metro 2001 Type D&E samples are mostly clustered together and the Statewide 2003 samples form a different cluster. Although not very distinct however based on the *Subset 2* we do observe some pattern of clustering for both Metro 2001 samples and the Statewide 2003 samples.

Further as before we do the SOM modeling for each of the metals separately.

ELEMENT=Arsenic(As)

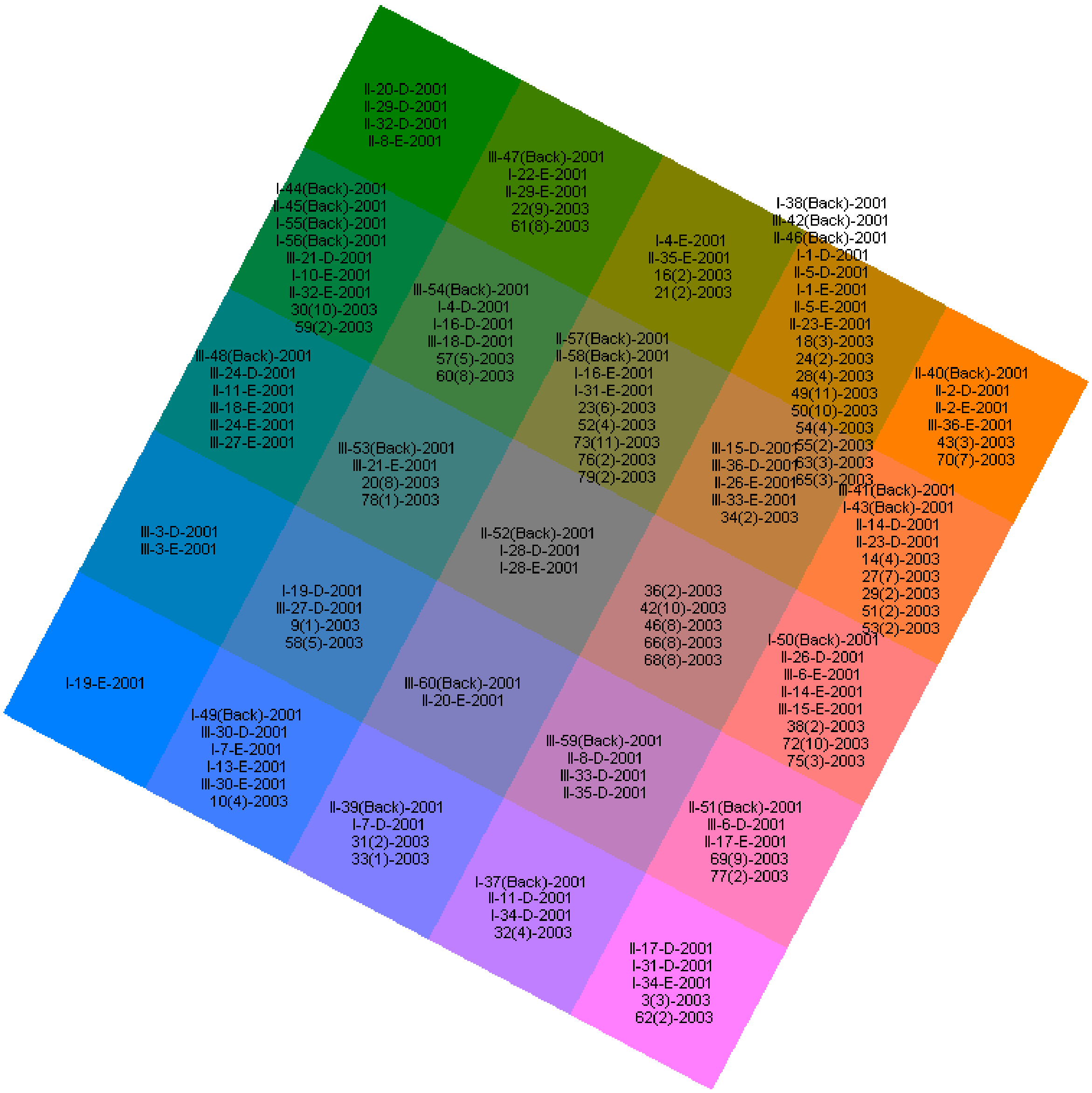


Figure 4.3.116. SOM model for both 2001 and 2003 Data using Arsenic (without Type A samples).

ANALYSIS

In this case we do not observe any good pattern in the clustering.

ELEMENT=Chromium (Cr)

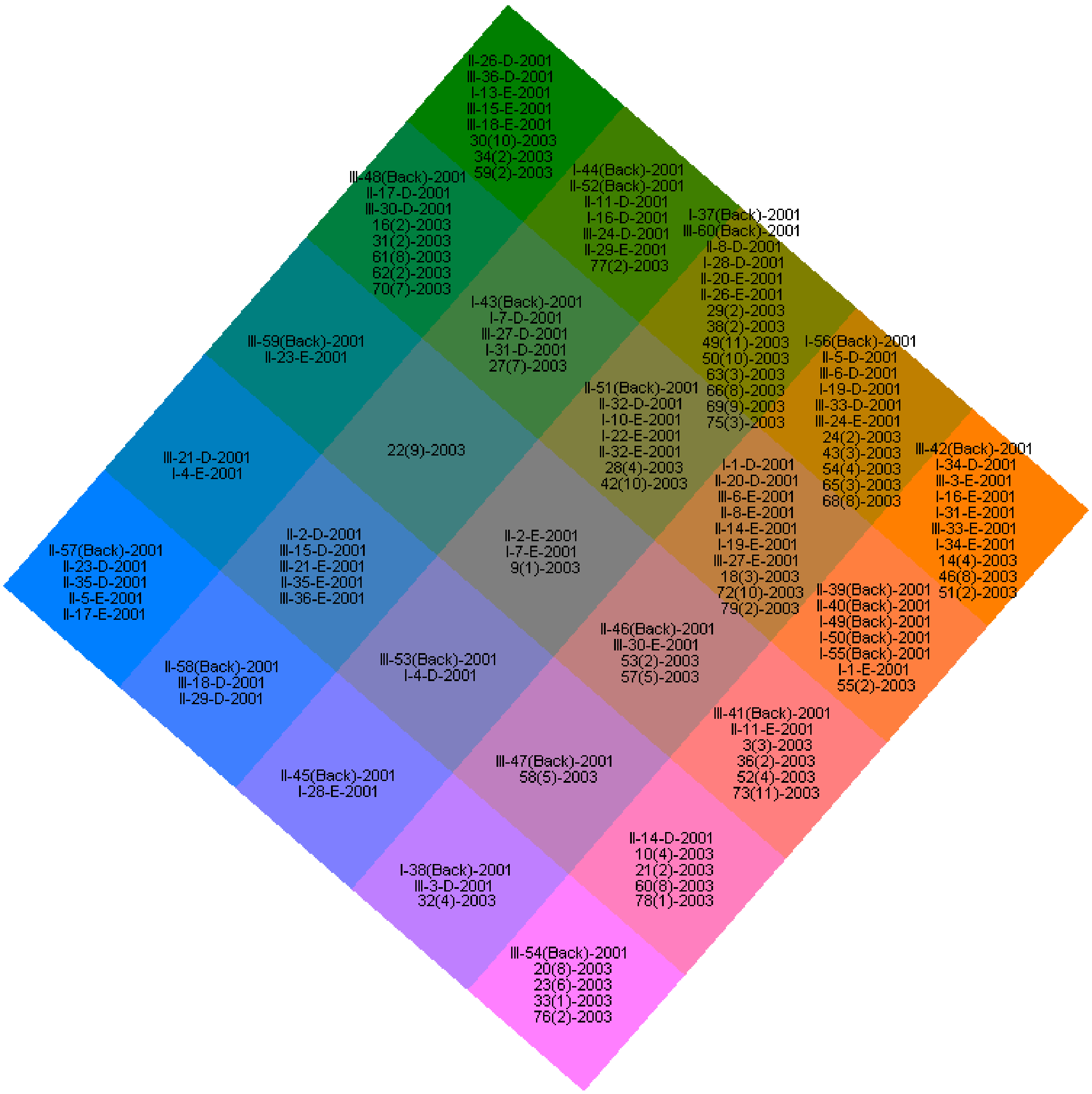


Figure 4.3.117. SOM model for both 2001 and 2003 Data using Chromium (without Type A samples).

ANALYSIS

In this case we do observe some weak pattern in the clustering. We observe that the Statewide 2003 samples form a cluster at the bottom of the SOM map. However, the clustering is not very prevalent throughout the SOM map. So we shall not make any conclusions from this.

ELEMENT=Copper(Cu)

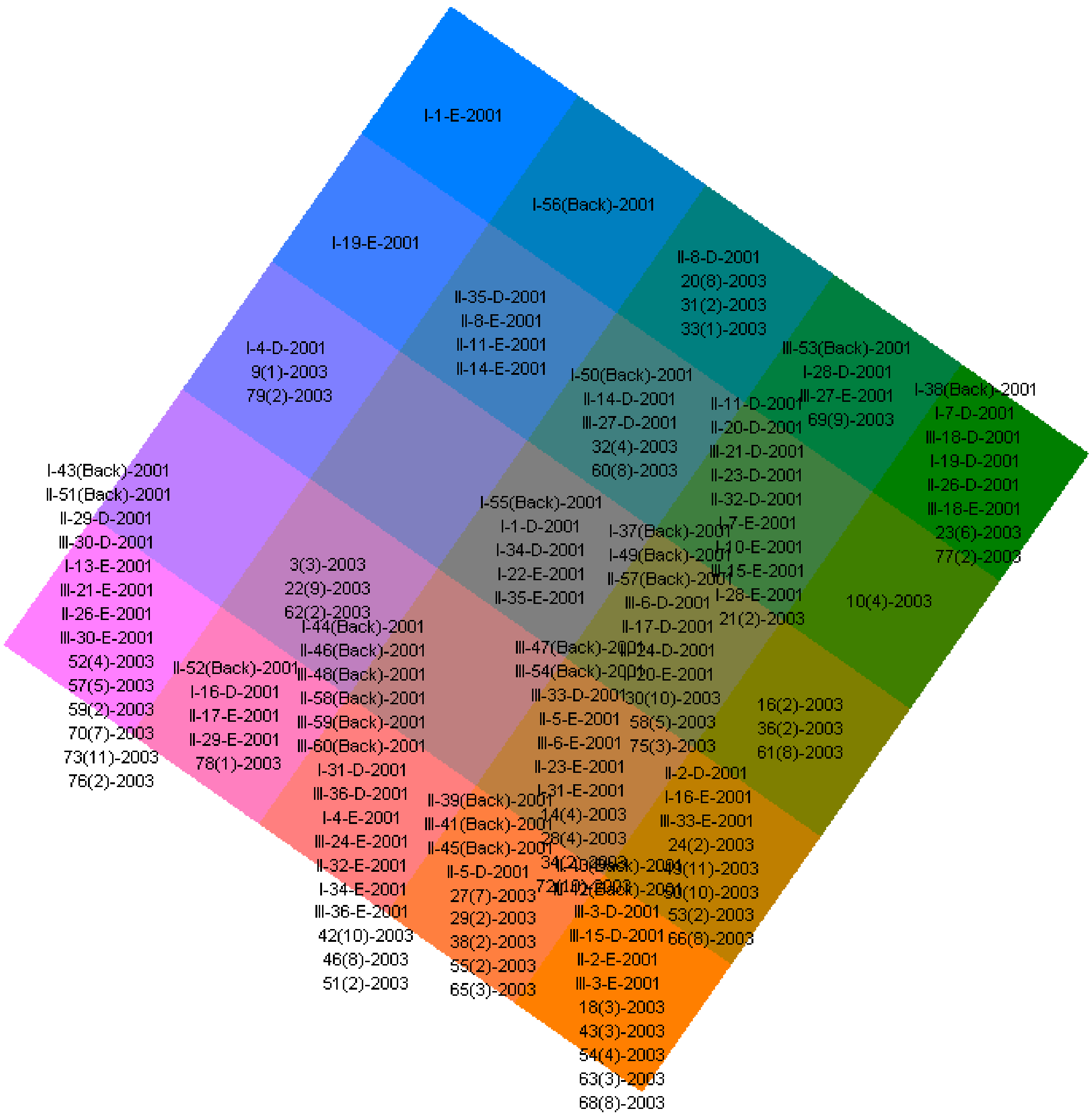


Figure 4.3.118. SOM model for both 2001 and 2003 Data using Copper (without Type A samples).

ANALYSIS

In this case there does not seem to be a very good pattern of clustering.

ELEMENT=Lead(Pb)

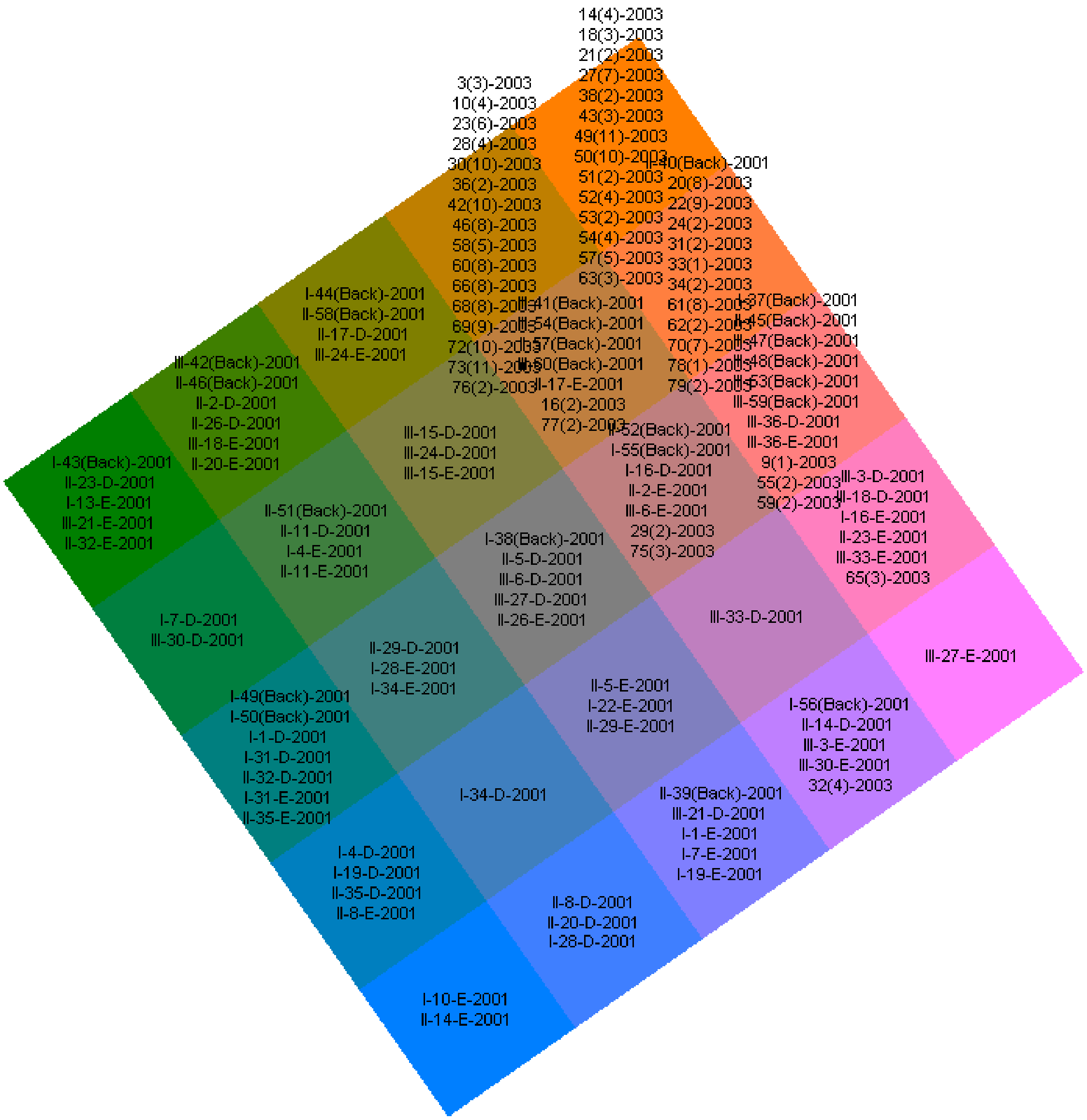


Figure 4.3.119. SOM model for both 2001 and 2003 Data using Lead (without Type A samples).

ANALYSIS

There seems to be a good pattern of clustering for this case. In this case we can observe that the Statewide 2003 samples cluster together and the Metro 2001 samples form a different cluster. Thus the Statewide 2003 soil samples seem to have a different concentration of Pb than the Metro 2001 soil samples of Type D and E. Moreover we have checked that the Metro 2001 samples have a relatively higher Pb concentration. Further, the background samples seem to be sandwiched between the Type D &E samples and the Statewide 2003 samples. Thus the background samples seem to have Pb concentration that is in between the Metro 2001 Type D &E samples and the Statewide 2003 samples.

ELEMENT=Nickel (Ni).

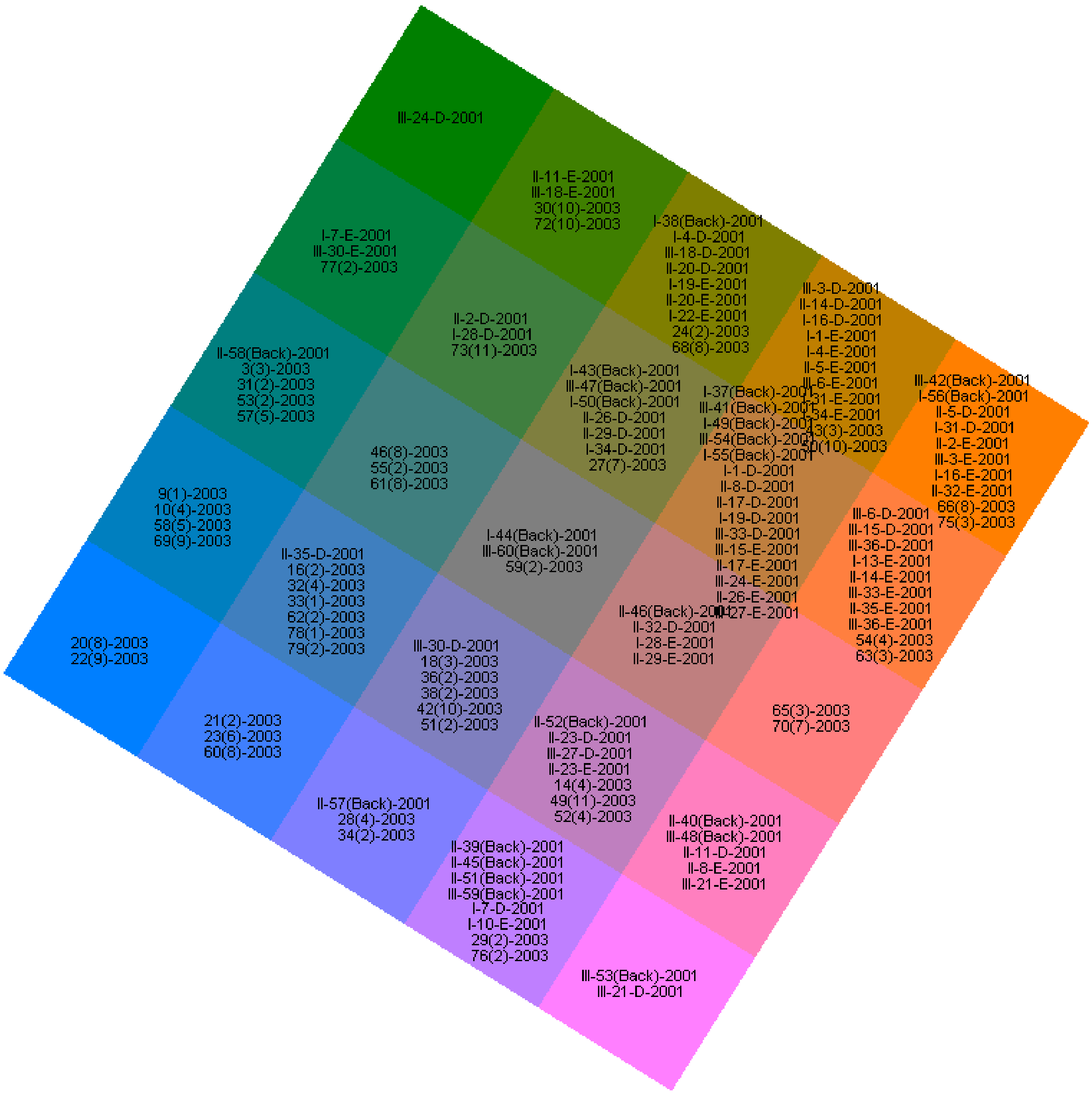


Figure 4.3.120. SOM model for both 2001 and 2003 Data using Nickel (without Type A samples).

ANALYSIS

There seems to be a good pattern of clustering for this case. In this case we can observe that the Statewide 2003 samples cluster together and the Metro 2001 samples form a different cluster. Thus the Statewide 2003 soil samples seem to have a different concentration of Ni than the Metro 2001 soil samples of Type D and E. Moreover we have checked that the Statewide 2003 samples have a relatively higher Ni concentration. Further, the background samples seem to bear a close relationship with the Metro 2001 Type D&E samples than the Statewide 2003 samples.

ELEMENT=Tungsten(W).

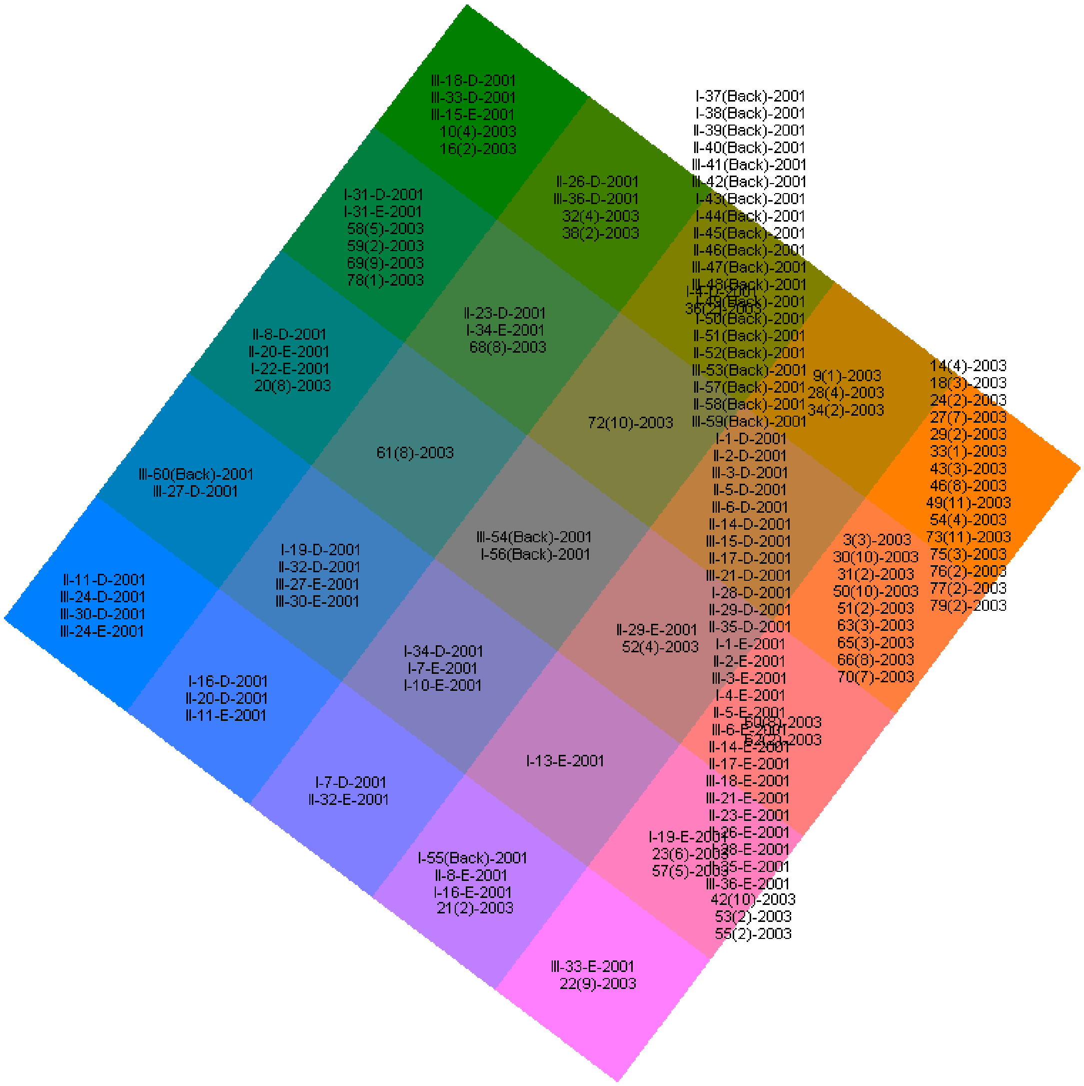


Figure 4.3.121. SOM model for both 2001 and 2003 Data using Tungsten (without Type A samples).

ANALYSIS

There seems to be a pattern of clustering for this case. In this case we can observe that the Statewide 2003 samples cluster together and the Metro 2001 samples form a different cluster. Thus the Statewide 2003 soil samples seem to have a different concentration of W than the Metro 2001 soil samples of Type D and E. Moreover we have checked that the Metro 2001 samples have a relatively higher W concentration. Further, the background samples seem to be sandwiched between the Metro 2001 Type D &E samples and the Statewide 2003 samples. Thus the background samples seem to have W concentration that is in between the Metro 2001 D &E samples and the Statewide 2003 samples.

ELEMENT=Zinc (Zn).

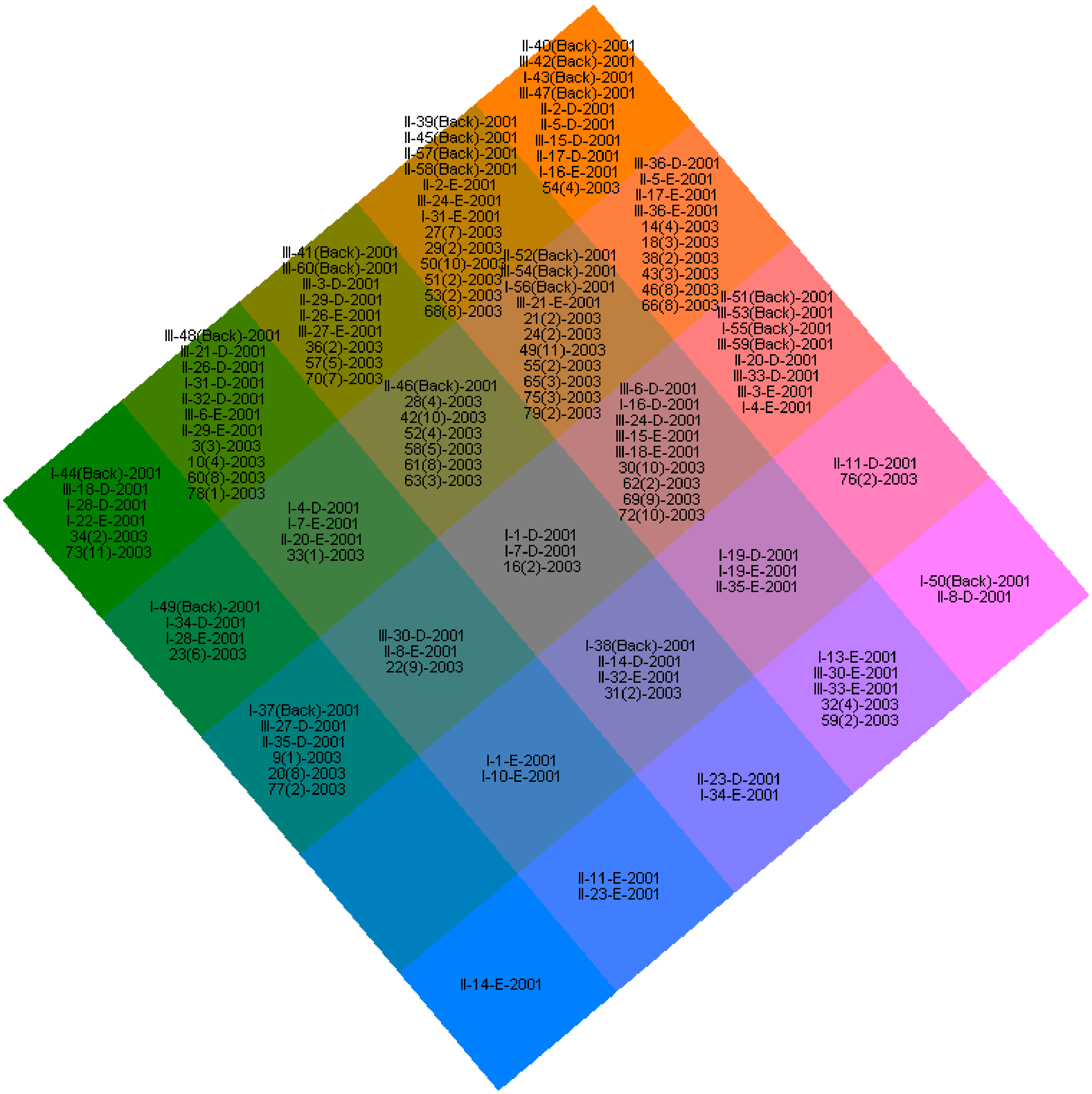


Figure 4.3.122. SOM model for both 2001 and 2003 Data using Zinc (without Type A samples).

ANALYSIS

There does not seem to be a very good pattern in this case.

4.3.6 Summary of results for the Metro 2001 Soil Survey Data and MN Statewide 2003 Soil Survey Data

Based on all elements

- The samples of Metro 2001 Type A form a distinct cluster and the samples of Statewide 2003 samples form another distinct cluster. The Metro 2001 Type D & E samples seem to be sandwiched between these two clusters. Further the background samples seem to be related more to the samples of Metro 2001 Type D & E and the Statewide 2003 samples.
- A detailed analysis of the Metro 2001 Type D & E samples with the Statewide 2003 samples reveal that the Metro 2001 Type D& E samples form a cluster and the Statewide 2003 samples form a different cluster. There is no pattern of clustering for the background samples.

Based on *Subset 1*.

- The samples of Metro 2001 samples form a distinct cluster and the samples of Statewide 2003 samples form another distinct cluster. There is no pattern in which the background samples are clustered.
- Further when using just the Metro 2001 Type D& E samples along with the Statewide 2003 samples we observe that the Metro 2001 Type D & E samples form a cluster and the Statewide 2003 samples form a different cluster. There is no pattern of clustering for the background samples.

Based on *Subset 2* (Metals of Regulatory interest).

- The samples of Metro 2001 samples form a distinct cluster and the samples of Statewide 2003 samples form another distinct cluster. There is no pattern in which the background samples are clustered.
- Further when using just the Metro 2001 Type D& E samples with the Statewide 2003 samples we observe that the Metro 2001 Type D & E samples form a cluster and the Statewide 2003 samples form a different cluster. There is no pattern in which the background samples are clustered.

Results pertaining to the metals of regulatory interests.

ELEMENT = Arsenic (As)

There is no pattern of clustering based on the concentration values of As.

ELEMENT = Chromium (Cr)

There seems to be a very weak pattern of clustering of the Metro 2001 samples against the Statewide 2003 samples. But we cannot make strong inferences from this weak pattern.

ELEMENT = Copper (Cu)

- There is a good pattern of clustering for the Metro 2001 Type A samples. The Metro 2001 Type A samples have relatively higher concentration values for Cu. (We have already observed a similar result in section 4.3.2).
- We do not observe any distinct pattern of clustering between the Metro 2001 Type D&E samples and the Statewide 2003 samples. Thus the Metro 2001 Type D&E samples have the same Cu concentration as the Statewide 2003 soil samples.
- Further there is no pattern of clustering for the background samples.

ELEMENT = Nickel (Ni)

- It seems like the Metro 2001 samples have similar Ni concentration values and cluster together, and the Statewide 2003 samples have similar Ni concentration values and form a different cluster. Further, the Statewide 2003 samples have a higher Ni concentration value.
- Moreover it seems like the Metro 2001 Type D &E samples form a different cluster than the Statewide 2003 samples. The background samples seem to be more related to the Metro 2001 Type D & E samples than to the Statewide 2003 samples based on the Ni concentration values.

ELEMENT = Lead (Pb)

- It seems like the Metro 2001 samples have the similar Pb concentration values and the Statewide 2003 samples have similar Pb concentration values. Further, the Metro 2001 samples have a higher Pb concentration value.
- Moreover it seems like the Metro 2001 Type D &E samples form a different cluster than the Statewide 2003 samples. The background samples seem to have a Pb concentration value in between the Metro 2001 Type D&E samples and the Statewide 2003 soil samples.

ELEMENT = Tungsten (W)

- It seems like the Metro 2001 samples of Type A have similar W concentration values which is different than other soil samples. Further, the Metro 2001 Type A samples has higher W concentration value.
- Moreover it seems like the Metro 2001 Type D &E samples form a different cluster than the Statewide 2003 samples. The background samples seem to have a W concentration value in between the Metro 2001 Type D&E samples and the Statewide 2003 soil samples.

ELEMENT = Zinc (Zn)

- There is a good pattern of clustering for the Metro 2001 Type A samples. Thus the Metro-2001 Type A samples has similar concentration values for Zinc.
- We do not observe any distinct pattern of clustering between the Metro 2001 Type D&E samples and the Statewide 2003 samples. Thus the Metro 2001 Type D&E samples have the same Zn concentration as the Statewide 2003 soil samples.
- Further there is no pattern of clustering for the background samples.

5 CONCLUSIONS

This report describes data-analytic modeling of the Minnesota soil chemical data produced by the 2001 metro soil survey and by the 2003 state-wide survey. Analysis of this data was performed using Self Organizing Map (SOM) method, which enables clustering of the high-dimensional soil data in the two-dimensional space suitable for understanding and visualization. The resulting SOM models enable understanding and analysis of the soil chemical concentration trends within the metro area and in the state of Minnesota. A brief summary of most significant results is provided below. More detailed inferences about the patterns and trends observed in the soil data can be found in sections 4.3.2, 4.3.4 and 4.3.6.

Metro 2001 Soil Survey Data

The purpose of analysis was to understand how the chemical characteristics of soil samples in the metro area change with the distance from major highways and location relative to the Minneapolis Main Post Office. SOM clustering based on the overall concentration of all chemical elements indicates that the Type A samples (collected close to the road) have similar characteristics, whereas the Type D&E samples (collected far away from the roads) have similar characteristics. Moreover, the background samples collected near parking lots or minor city roads have chemical characteristics similar to the Metro 2001 Type D&E samples. This analysis indicates that the enrichment of the concentration of chemical elements in the soil data is due to the proximity to the major highways in the metro area.

Further, our results suggest that the concentration levels of the elements of regulatory interests are higher for Type A soil samples (closest to the major highways). A detailed analysis shows that these samples have high concentration values for the elements Lead (Pb), Tungsten (W), Zinc (Zn), and Copper (Cu). These results could be helpful to determine the suitability of certain materials for usage as roadway bed or fill-in materials, at a particular location.

MN Statewide 2003 Soil Survey Data

The purpose of this analysis was to explore the pattern of clustering based on the geomorphologic units and the provenance of the glacial sediments. SOM modeling results suggest a good pattern of clustering based on the provenance. In particular, there is a good pattern of clustering based on the provenance when the soil characteristics are exclusively determined by the rock forming elements Al, Fe, Mg, Ca, Na, K, and Mn. This result can be of possible interest to the Minnesota Department of Agriculture to understand the geology of the top soil in Minnesota.

Further, our results indicate that samples with high concentration of metals of regulatory interest Chromium (Cr), Copper (Cu), Arsenic (As), Nickel (Ni), Tungsten (W) are located at the north-east of Minnesota, indicating possible pollution from mining activities.

REFERENCES

[1] Draft, Internal Mn/DOT Office of Environmental Services documents, (2001 Sampling plan.pdf), May 2001.

[2] Draft, Internal Mn/DOT Office of Environmental Services documents, (Sampling scheme.doc), May 2001.

[3] HDR Engineering, Inc., *Soil Sampling Plan*, Minneapolis: HDR Engineering, Inc., June 2003.

[4] V. Cherkassky, F. Mulier, *Learning from Data: Concepts, Theory and Methods*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2007.

[5] Save Our Sky Blue Waters (Internet), *Explore Minnesota: Minerals*, March 2008 (Accessed June 2010), http://www.sosbluewaters.org/businessminingminerals207%5b1%5d.pdf.

[6] Minnesota Geological Survey, *Quaternary Geology of Minnesota*, (Accessed June 2010), http://www.geo.umn.edu/mgs/surfexpl.html.

APPENDIX A: COMPARISON OF THE METRO 2001 SAMPLES WITH THE BACKGROUND SAMPLES

In this section we provide the list of sites that exceed the average value of the background samples within a particular Circle. All values are shown in parts per million (ppm).

Table A.1. List of sites within Circle I that exceed the average value of the background samples within Circle I

ELEMENT	AVG(BACK)	SITES
P10	98.8125	I-4-E I-31-E
Silt	44.3375	I-7-D I-7-E
Clay	3.95	I-7-E
SG	51.7125	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-1-D I-4-D I-16-D I-19-D I-28-D I-31-D I-34-D I-1-E I-4-E I-10-E I-13-E I-16-E I-19-E I-22-E I-28-E I-31-E I-34-E
OM	9.6875	I-10-E
PH	6.9625	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-4-D I-7-D I-16-D I-19-D I-28-D I-34-D I-4-E I-7-E I-10-E I-13-E I-16-E I-19-E I-22-E I-34-E
EP	86.125	I-1-D I-16-D I-19-D I-28-D I-34-D I-1-E I-22-E I-34-E
eK	438.75	
dAs	4.3875	I-16-A I-19-A I-25-A I-4-D I-7-D I-16-D I-19-D I-34-D I-7-E I-10-E I-13-E I-19-E I-22-E
dBr	9.025	I-19-A I-4-D I-19-D I-10-E
dCe	33.5	I-7-A I-10-A I-16-A I-22-A I-25-A I-31-A I-34-A I-7-D I-16-D I-4-E I-7-E I-10-E I-19-E I-22-E
dCo	6.55	I-7-A I-10-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-7-D I-16-D I-7-E I-10-E
dCr	38.25	I-1-A I-4-A I-7-A I-10-A I-13-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-4-D I-7-D I-16-D I-31-D I-4-E I-7-E I-13-E I-28-E
dCs	1.0125	I-25-A I-28-A I-7-D I-16-D I-28-D I-1-E I-7-E I-10-E I-22-E I-28-E
dFe	17409.75	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-7-D I-16-D I-28-D I-7-E I-10-E I-13-E I-28-E
dLa	19.425	I-22-A I-25-A I-34-A I-7-D I-16-D I-19-D I-4-E I-7-E I-10-E I-22-E
dMn	687	I-19-A I-34-A I-28-D
dMo	1.5125	I-1-A I-7-A I-10-A I-16-A I-19-A I-22-A I-25-A I-31-A I-34-A I-7-D I-16-D I-19-D I-31-D I-7-E I-10-E I-16-E I-19-E I-22-E
dNi	11.025	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-7-D I-28-D I-34-D I-7-E I-10-E I-28-E
dPb	31.625	I-1-A I-4-A I-7-A I-10-A I-13-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-1-D I-4-D I-7-D I-19-D I-28-D I-31-D I-34-D I-1-E I-7-E I-10-E I-13-E I-19-E I-22-E I-28-E I-31-E I-34-E
dRb	50.375	I-25-A I-7-D I-16-D I-7-E I-22-E
dSb	0.65	I-1-A I-4-A I-7-A I-10-A I-13-A I-19-A I-22-A I-25-A I-31-A I-34-A I-7-E I-10-E
dSc	5.35	I-7-A I-25-A I-7-D I-16-D I-7-E I-10-E I-22-E
dSe	1.382875	I-16-A I-19-A I-19-D I-34-D I-7-E I-19-E I-22-E I-34-E
dSm	3.1375	I-7-A I-10-A I-16-A I-31-A I-34-A I-4-D I-7-D I-16-D I-19-D I-4-E I-7-E I-10-E I-16-E I-19-E I-22-E I-34-E
dSr	141.75	I-1-A I-4-A I-7-A I-10-A I-13-A I-19-A I-25-A I-28-A I-31-A I-1-D I-4-D I-7-D I-16-D I-28-D I-31-D I-34-D I-1-E I-4-E I-7-E I-13-E I-19-E I-28-E I-34-E
dTh	5.75	I-16-A I-22-A I-25-A I-34-A I-7-D I-7-E I-10-E I-19-E I-22-E
dU	1.4875	I-16-A I-19-A I-25-A I-7-D I-16-D I-19-D I-7-E I-10-E I-22-E
dV	57.625	I-19-A I-7-D I-28-D I-7-E I-28-E
dW	1.2875	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-31-A I-34-A I-7-D I-16-D I-19-D I-31-D I-34-D I-7-E I-10-E I-13-E I-16-E I-19-E I-22-E I-31-E I-34-E
dY	11	I-1-A I-13-A I-4-E I-13-E I-28-E
dZn	70.25	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-1-D I-7-D I-19-D I-34-D I-1-E I-10-E I-13-E I-19-E I-28-E I-34-E
iCd	0.2875	I-16-A I-19-A I-25-A I-4-D I-7-D I-16-D I-19-D I-28-D I-10-E
iCu	16.625	I-1-A I-4-A I-7-A I-10-A I-13-A I-16-A I-19-A I-22-A I-25-A I-28-A I-31-A I-34-A I-4-D I-1-E I-19-E

Table A.2. List of sites within Circle II that exceed the average value of the background samples within Circle II.

ELEMENT	AVG(BACK)	SITES
P10	96.525	II-2-D II-5-D II-17-D II-23-D II-29-D II-32-D II-35-D II-2-E II-5-E II-17-E II-23-E II-29-E II-32-E II-35-E
Silt	37.825	II-14-D II-17-D II-29-D II-32-D II-11-E II-17-E II-29-E II-32-E
Clay	1.9875	II-14-D II-23-D II-26-D II-29-D II-32-D II-20-E II-26-E II-29-E II-32-E
SG	60.1875	II-2-A II-5-A II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-26-A II-29-A II-32-A II-35-A II-2-D II-5-D II-8-D II-11-D II-20-D II-23-D II-26-D II-35-D II-2-E II-5-E II-8-E II-14-E II-20-E II-23-E II-26-E II-35-E
OM	5.925	II-11-A II-29-A II-32-A II-8-D II-11-D II-23-D II-29-D II-32-D II-35-D II-5-E II-8-E II-11-E II-14-E II-17-E II-26-E II-29-E II-32-E II-35-E
PH	6.9	II-2-A II-5-A II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-26-A II-29-A II-32-A II-35-A II-14-D II-20-D II-26-D II-29-D II-32-D II-35-D II-2-E II-14-E II-20-E II-23-E II-26-E II-32-E
EP	101.875	II-14-A II-5-D II-5-E II-17-E II-32-E II-35-E
eK	337.75	II-23-A II-29-A II-8-D II-14-D II-29-D II-5-E II-8-E II-17-E II-26-E II-29-E
dAs	3.4375	II-11-A II-29-A II-8-D II-11-D II-17-D II-20-D II-29-D II-32-D II-35-D II-8-E II-11-E II-17-E II-20-E II-29-E II-32-E
dBr	5	II-11-A II-29-A II-32-A II-8-D II-17-D II-20-D II-29-D II-32-D II-35-D II-8-E II-11-E II-14-E II-20-E II-29-E II-32-E
dCe	30.5	II-8-A II-11-A II-29-A II-32-A II-8-D II-11-D II-14-D II-17-D II-20-D II-29-D II-32-D II-11-E II-17-E II-20-E II-29-E II-32-E
dCo	5.9875	II-8-A II-11-A II-14-A II-17-A II-23-A II-29-A II-32-A II-35-A II-11-D II-14-D II-17-D II-26-D II-29-D II-32-D II-35-D II-8-E II-11-E II-14-E II-17-E II-20-E II-32-E II-35-E
dCr	97	II-2-A II-14-A II-17-A II-23-A II-29-A II-35-A II-2-D II-23-D II-29-D II-35-D II-5-E II-17-E II-35-E
dCs	1.1125	II-17-A II-17-D II-23-D II-29-D II-32-D II-11-E II-14-E II-20-E II-29-E II-32-E
dFe	16225	II-2-A II-5-A II-8-A II-11-A II-14-A II-17-A II-23-A II-29-A II-32-A II-35-A II-8-D II-11-D II-14-D II-17-D II-23-D II-26-D II-29-D II-32-D II-8-E II-11-E II-14-E II-17-E II-26-E II-29-E II-32-E
dLa	14.6625	II-8-A II-11-A II-17-A II-20-A II-29-A II-32-A II-8-D II-11-D II-14-D II-17-D II-20-D II-26-D II-29-D II-32-D II-8-E II-11-E II-14-E II-17-E II-20-E II-26-E II-29-E II-32-E
dMn	630	II-14-A II-29-A II-23-D II-26-D II-29-D II-23-E II-26-E II-29-E
dMo	1.1375	II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-29-A II-32-A II-35-A II-8-D II-11-D II-17-D II-20-D II-29-D II-32-D II-8-E II-11-E II-17-E II-20-E II-23-E II-29-E II-32-E II-35-E
dNi	20.5	II-17-A II-26-A II-29-A II-35-A II-35-D
dPb	21	II-2-A II-5-A II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-29-A II-32-A II-35-A II-2-D II-5-D II-8-D II-11-D II-14-D II-20-D II-23-D II-26-D II-29-D II-32-D II-35-D II-5-E II-8-E II-11-E II-14-E II-20-E II-23-E II-26-E II-29-E II-32-E II-35-E
dRb	43.75	II-5-A II-8-A II-11-A II-17-A II-29-A II-32-A II-8-D II-11-D II-14-D II-26-D II-29-D II-32-D II-2-E II-8-E II-11-E II-14-E II-17-E II-20-E II-29-E II-32-E
dSb	0.3625	II-2-A II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-29-A II-32-A II-35-A II-5-D II-8-D II-11-D II-14-D II-26-D II-29-D II-32-D II-8-E II-11-E II-20-E II-29-E II-32-E
dSc	4.5625	II-8-A II-11-A II-17-A II-29-A II-32-A II-8-D II-11-D II-14-D II-17-D II-26-D II-29-D II-32-D II-8-E II-11-E II-14-E II-20-E II-29-E II-32-E
dSe	0.9125	II-2-A II-8-D II-11-D II-20-D II-29-D II-32-D II-8-E II-14-E II-20-E II-29-E
dSm	2.6	II-2-A II-8-A II-11-A II-17-A II-20-A II-29-A II-8-D II-11-D II-14-D II-17-D II-20-D II-23-D II-26-D II-29-D II-32-D II-35-D II-8-E II-11-E II-17-E II-20-E II-29-E II-32-E
dSr	167.5	II-5-A II-23-A II-26-A II-35-A II-2-D II-5-D II-8-D II-26-D II-35-D II-2-E II-5-E II-8-E II-23-E II-26-E II-35-E
dTh	5.65	II-8-A II-11-A II-32-A II-14-D II-17-D II-29-D II-8-E II-11-E II-17-E II-20-E II-29-E II-32-E
dU	1.325	II-8-A II-11-A II-14-A II-20-A II-32-A II-5-D II-8-D II-11-D II-14-D II-17-D II-20-D II-29-D II-32-D II-5-E II-8-E II-11-E II-14-E II-17-E II-20-E II-23-E II-29-E II-32-E
dV	69.375	II-14-A II-14-D II-17-D II-29-D II-29-E
dW	0.9	II-2-A II-5-A II-8-A II-11-A II-17-A II-20-A II-23-A II-26-A II-29-A II-32-A II-35-A II-8-D II-11-D II-20-D II-23-D II-26-D II-32-D II-8-E II-11-E II-20-E II-29-E II-32-E
dY	15.5375	II-14-D II-17-D II-29-D II-17-E II-29-E
dZn	37.875	II-2-A II-5-A II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-26-A II-29-A II-32-A II-35-A II-8-D II-11-D II-14-D II-20-D II-23-D II-26-D II-29-D II-32-D II-35-D II-8-E II-11-E II-14-E II-20-E II-23-E II-26-E II-29-E II-32-E II-35-E
iCd	0.275	II-29-A II-32-A II-8-D II-14-D II-17-D II-20-D II-23-D II-26-D II-29-D II-11-E II-20-E II-29-E
iCu	6.75	II-2-A II-5-A II-8-A II-11-A II-14-A II-17-A II-20-A II-23-A II-26-A II-29-A II-32-A II-35-A II-8-D II-11-D II-14-D II-17-D II-20-D II-23-D II-26-D II-29-D II-32-D II-35-D II-8-E II-11-E II-14-E II-17-E II-20-E II-26-E II-29-E II-32-E II-35-E

Table A.3. List of sites within Circle III that exceed the average value of the background samples within Circle III.

ELEMENT	AVG VAL	SITES
P10	98.125	III-24-A III-3-D III-6-D III-15-D III-27-D III-3-E III-6-E III-15-E III-24-E III-30-E
Silt	49.6875	III-3-D III-30-D III-24-E III-30-E
Clay	5.8875	III-24-D III-27-D III-30-D III-24-E III-27-E III-30-E
SG	44.425	III-3-A III-6-A III-15-A III-18-A III-21-A III-24-A III-27-A III-30-A III-33-A III-36-A III-6-D III-15-D III-18-D III-21-D III-24-D III-33-D III-36-D III-6-E III-15-E III-18-E III-21-E III-27-E III-33-E III-36-E
OM	6.7875	III-3-A III-3-D III-6-D III-24-D III-27-D III-30-D III-3-E III-6-E III-18-E III-24-E III-27-E III-30-E
PH	6.725	III-3-A III-6-A III-15-A III-18-A III-21-A III-24-A III-27-A III-30-A III-33-A III-36-A III-3-D III-6-D III-18-D III-21-D III-24-D III-27-D III-30-D III-33-D III-36-D III-3-E III-6-E III-18-E III-21-E III-27-E III-30-E III-33-E III-36-E
EP	80.875	III-33-A III-18-D III-36-D III-18-E III-24-E III-33-E III-36-E
eK	392.5	III-18-D III-21-D III-30-D III-18-E III-21-E III-24-E
dAs	4.3375	III-3-D III-18-D III-21-D III-24-D III-27-D III-30-D III-3-E III-18-E III-21-E III-24-E III-27-E III-30-E
dBr	6.0875	III-3-A III-3-D III-18-D III-24-D III-27-D III-30-D III-3-E III-6-E III-24-E III-27-E III-30-E
dCe	39.475	III-33-A III-27-D III-30-D III-30-E
dCo	7.9375	III-18-D III-27-D III-30-D III-3-E III-18-E III-30-E
dCr	57.75	III-18-A III-3-D III-15-D III-18-D III-21-D III-21-E III-36-E
dCs	1.575	III-27-D III-30-D III-30-E
dFe	17201.25	III-3-A III-6-A III-15-A III-18-A III-21-A III-33-A III-36-A III-3-D III-18-D III-21-D III-27-D III-30-D III-3-E III-18-E III-21-E III-30-E
dLa	19.3875	III-18-D III-24-D III-27-D III-30-D III-24-E III-27-E III-30-E
dMn	635.375	III-18-A III-24-A III-3-D III-18-D III-21-D III-24-D III-27-D III-30-D III-3-E III-18-E III-21-E III-24-E III-27-E III-30-E
dMo	1.5375	III-24-A III-27-A III-30-A III-33-A III-21-D III-24-D III-27-D III-30-D III-24-E III-27-E III-30-E
dNi	13.625	III-15-A III-18-A III-21-A III-36-A III-21-D III-24-D III-27-D III-30-D III-18-E III-21-E III-30-E
dPb	15.875	III-3-A III-6-A III-15-A III-18-A III-21-A III-24-A III-27-A III-30-A III-33-A III-36-A III-3-D III-6-D III-15-D III-18-D III-21-D III-24-D III-27-D III-30-D III-33-D III-36-D III-3-E III-6-E III-15-E III-18-E III-21-E III-27-E III-30-E III-33-E III-36-E
dRb	55.25	III-27-D III-30-D III-30-E
dSb	0.4375	III-6-A III-33-A III-36-A III-3-D III-6-D III-21-D III-27-D III-30-D III-21-E III-30-E
dSc	5.7375	III-18-D III-21-D III-27-D III-30-D III-18-E III-30-E
dSe	1.4245	III-27-A III-33-A III-27-D III-30-D III-6-E III-24-E III-27-E III-30-E
dSm	3.2375	III-33-A III-18-D III-21-D III-24-D III-27-D III-30-D III-18-E III-21-E III-24-E III-27-E III-30-E
dSr	155.25	III-3-A III-15-A III-18-A III-21-A III-33-A III-6-D III-18-D III-21-D III-27-D III-33-D III-36-D III-6-E III-15-E III-18-E III-21-E III-36-E
dTh	7.0375	III-30-D III-24-E III-30-E
dU	1.575	III-27-A III-33-A III-6-D III-15-D III-18-D III-21-D III-24-D III-27-D III-30-D III-6-E III-18-E III-21-E III-24-E III-27-E III-30-E III-36-E
dV	69.125	III-18-D III-21-D III-30-D III-18-E III-21-E III-30-E
dW	1.725	III-15-A III-18-A III-21-A III-24-A III-27-A III-30-A III-33-A III-36-A III-24-D III-27-D III-30-D III-24-E III-27-E III-30-E III-33-E
dY	13.6375	III-18-D III-21-D III-18-E III-21-E
dZn	38.875	III-3-A III-6-A III-15-A III-18-A III-21-A III-24-A III-27-A III-30-A III-33-A III-36-A III-3-D III-6-D III-18-D III-21-D III-24-D III-27-D III-30-D III-33-D III-3-E III-6-E III-15-E III-18-E III-21-E III-27-E III-30-E III-33-E
iCd	0.275	III-6-A III-18-A III-21-A III-3-D III-6-D III-18-D III-21-D III-24-D III-27-D III-30-D III-3-E III-6-E III-18-E III-21-E III-24-E III-27-E III-30-E
iCu	6.625	III-3-A III-6-A III-15-A III-18-A III-21-A III-24-A III-27-A III-30-A III-33-A III-36-A III-6-D III-18-D III-21-D III-24-D III-27-D III-30-D III-36-D III-15-E III-18-E III-21-E III-24-E III-27-E III-30-E III-36-E

APPENDIX B: MAP OF MN BASED ON BEDROCK GEOLOGY

BEDROCK GEOLOGY of MINNESOTA

This section has been reproduced from [5]. Minnesota's diverse bedrock geology ranges from early Archean to Cretaceous. The state's greatest mineral potential lies in its extensive Precambrian geology, of which four terrains are particularly promising.

Figure B.1. Map of Minnesota based on the bedrock geology.

1. **The Superior Province**, which encompasses the northwestern third of Minnesota and represents a continuation of the mineral-rich Canadian Shield. Canadian gold discoveries have stimulated leasing in the Minnesota portion of the Superior Province. Most exploration in this area has been for gold, zinc-copper massive sulfides with various by-products, and magmatic sulfide deposits containing copper, nickel, and platinum group elements.

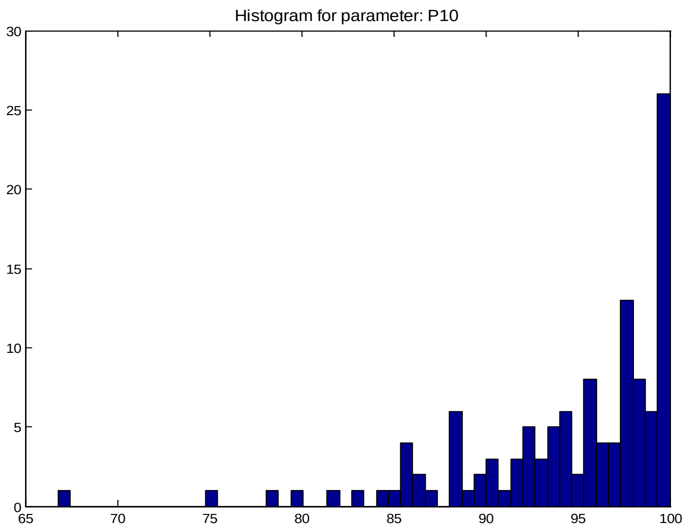
2. Variably metamorphosed **Middle Precambrian sediments and volcanic rocks**, including the Mesabi and Cuyuna iron ranges, with known reserves of iron and manganese. Recent exploration has focused on base metals.

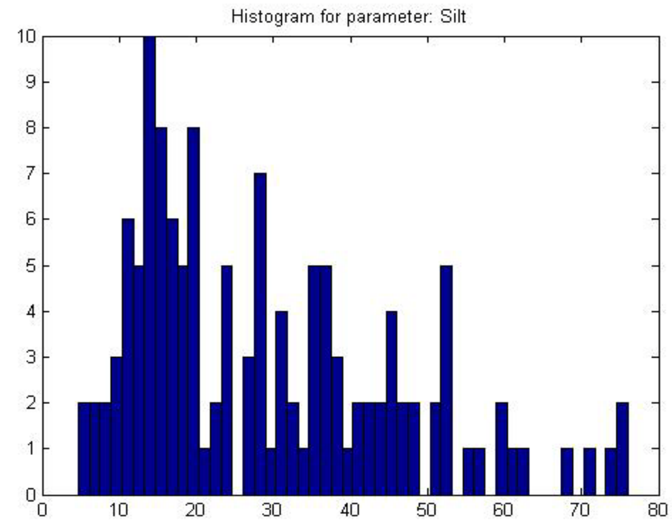
3. **The Duluth Complex and associated rocks**. This terrain contains most of the state's active non-ferrous mineral leases. The Minnesota Department of Natural Resources (DNR) estimates the identified copper-nickel resource of the area at about 4.4 billion tons averaging 0.66% copper and 0.2% nickel. This terrain also has significant titanium resources. Drilling has indicated the presence of other strategic minerals, such as chromium, vanadium, cobalt, and platinum-group elements.

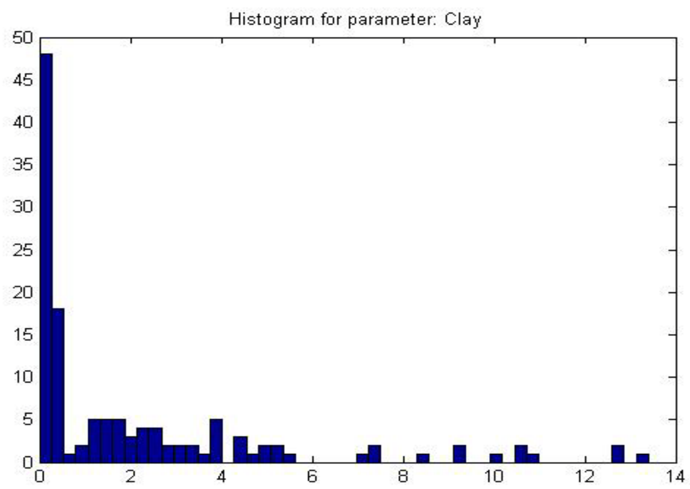
4. **Southern Minnesota's Archean migmatitic gneisses**, younger sedimentary rocks, batholithic granitic rocks and deformed volcanics. Limited exploration has focused on lead-zinc deposits in the southeast, base metals and precious metals in the Precambrian basement, and manganese in the southwest.

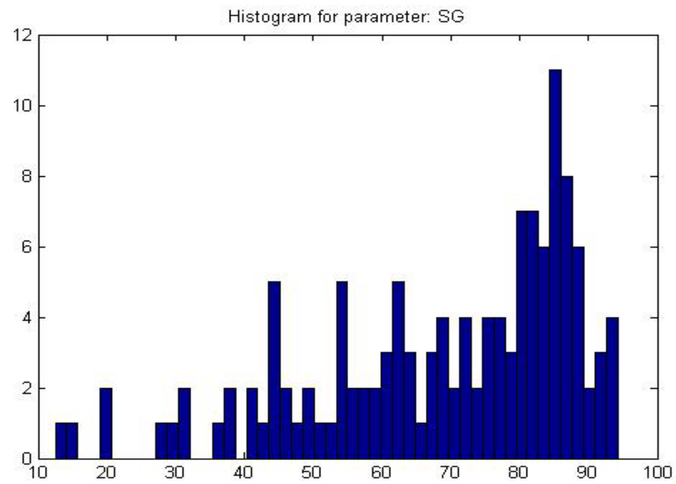
APPENDIX C: HISTOGRAM OF ALL ELEMENTS FOR METRO 2001 SOIL SURVEY DATA AND MN STATEWIDE 2003 SOIL SURVEY DATA

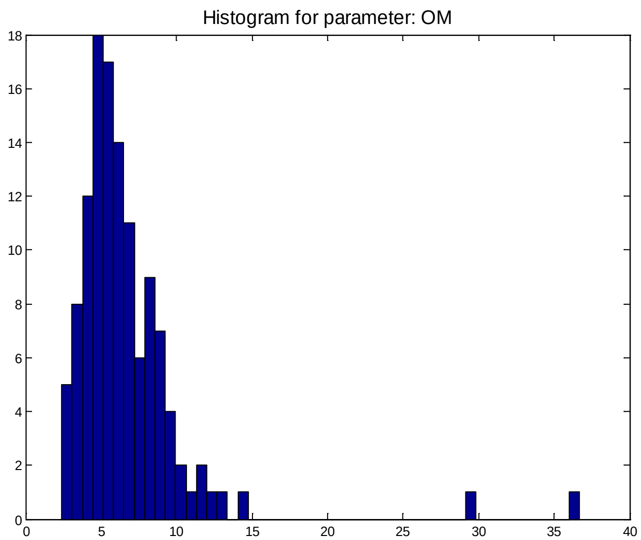
Metro 2001 Soil Survey Data

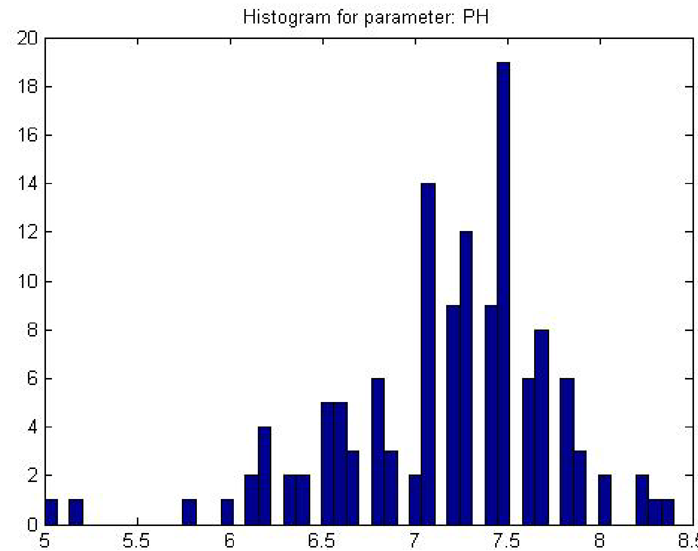


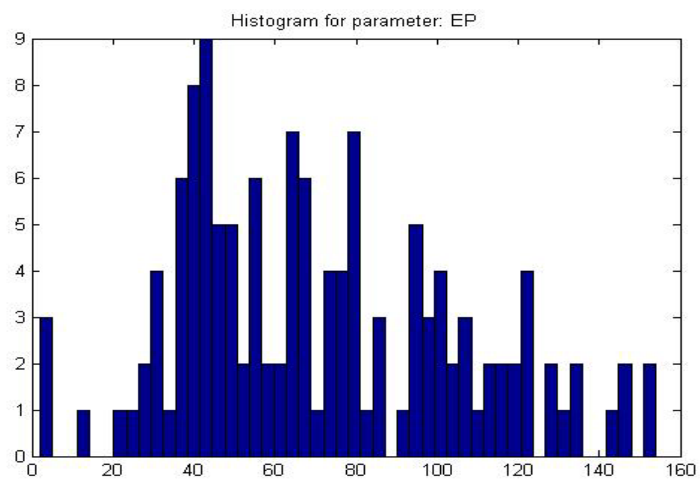


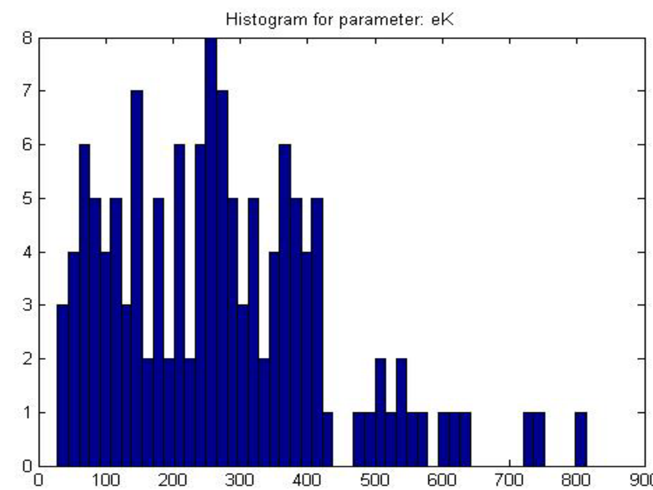


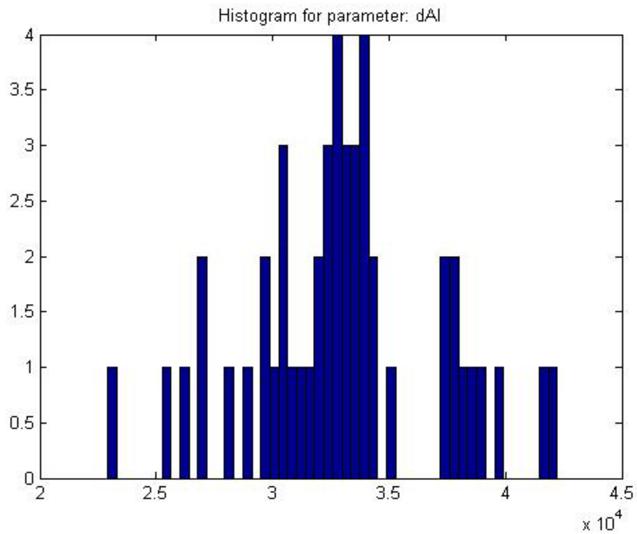


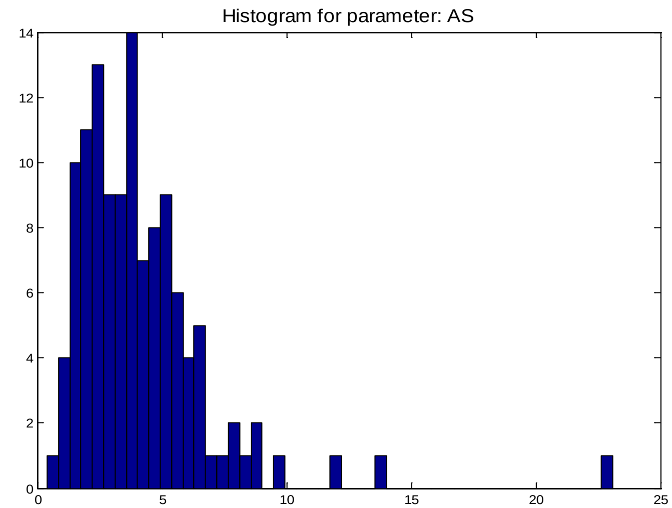

Histogram for parameter: OM

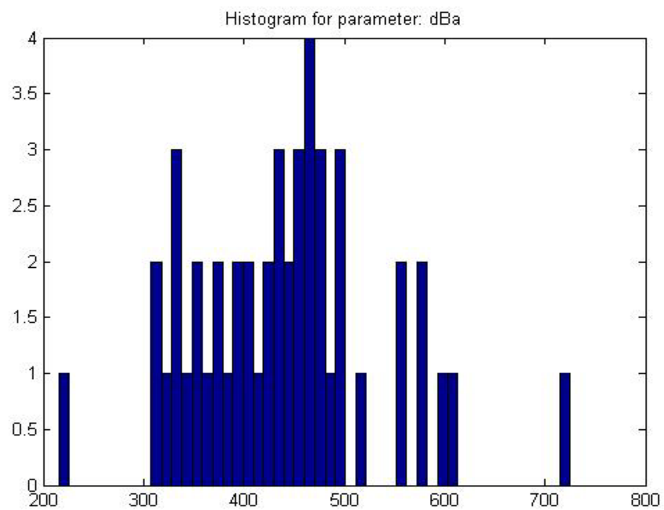

Histogram for parameter: PH

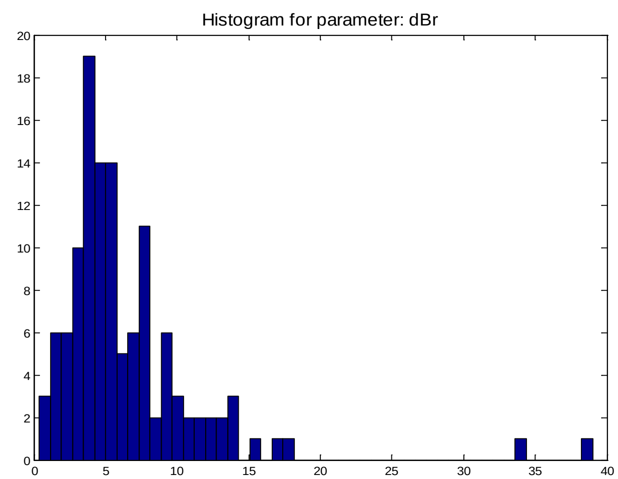

Histogram for parameter: EP

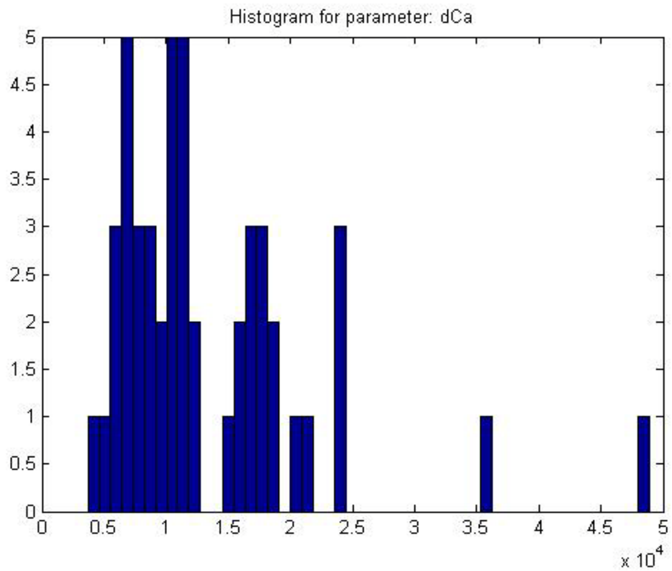

Histogram for parameter: eK

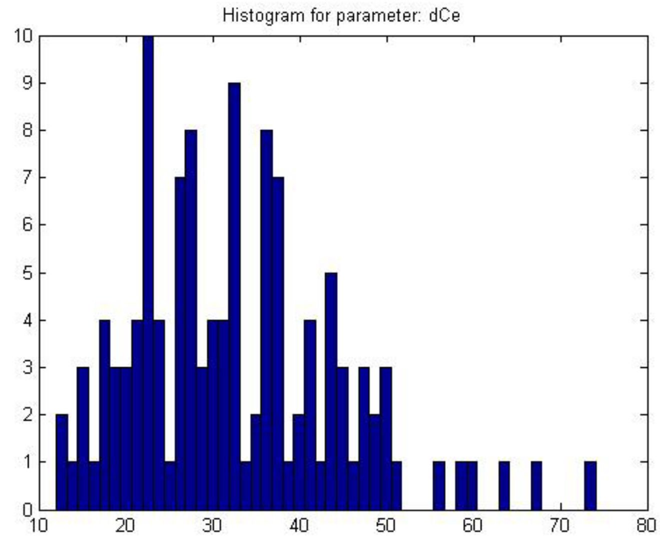

Histogram for parameter: dAl
0 0.5 1 1.5 2 2.5 3 3.5 4
2 2.5 3 3.5 4 4.5
x 10^4

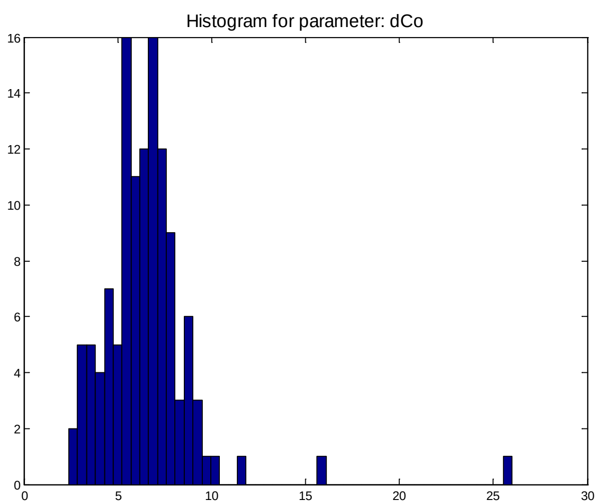

Histogram for parameter: AS
0 2 4 6 8 10 12 14
0 5 10 15 20 25

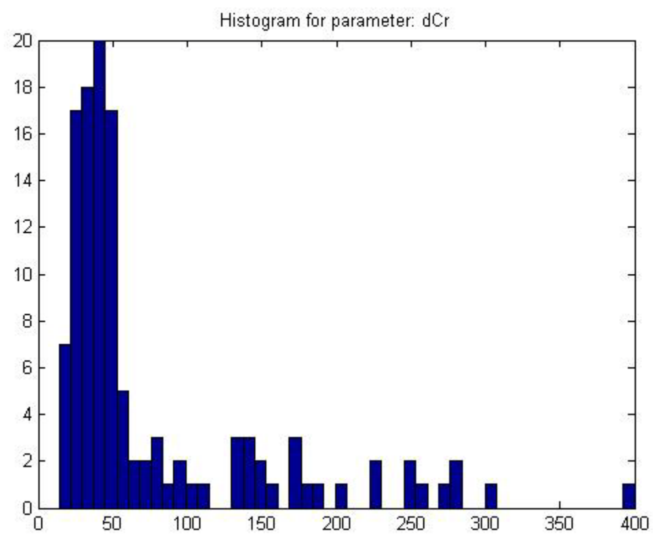

Histogram for parameter: dBa
0 0.5 1 1.5 2 2.5 3 3.5 4
200 300 400 500 600 700 800

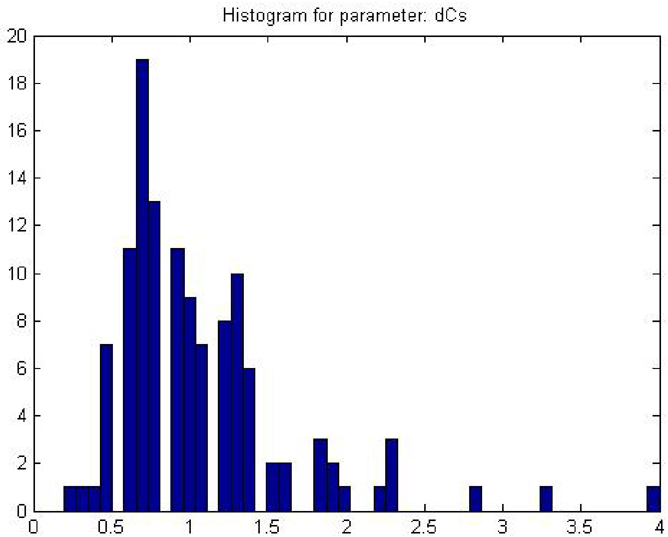

Histogram for parameter: dBr
0 2 4 6 8 10 12 14 16 18 20
0 5 10 15 20 25 30 35 40

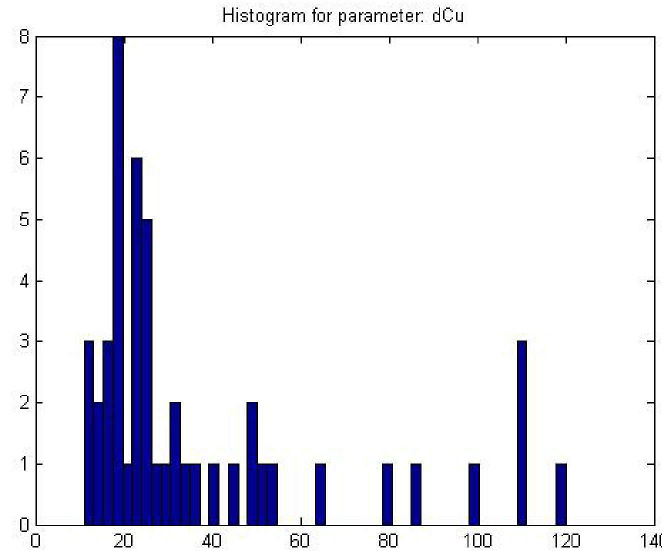

Histogram for parameter: dCa
x 10^4

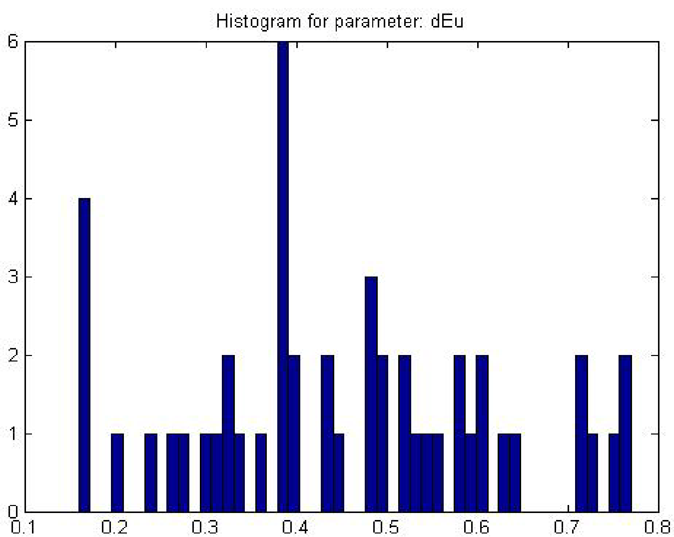

Histogram for parameter: dCe

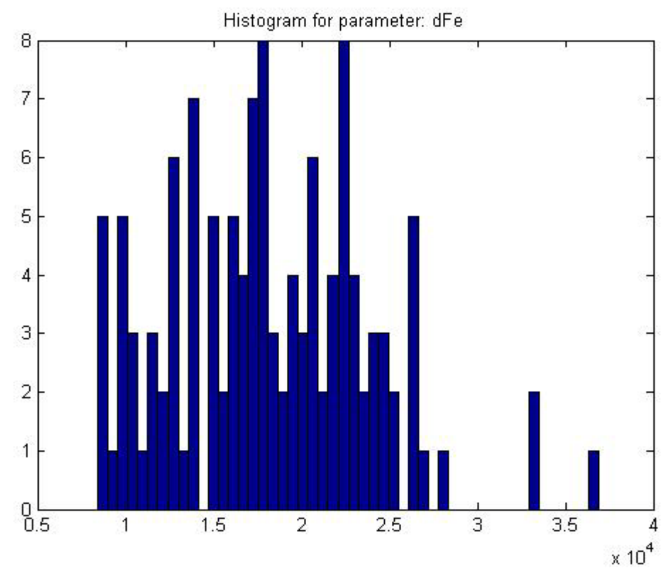

Histogram for parameter: dCo

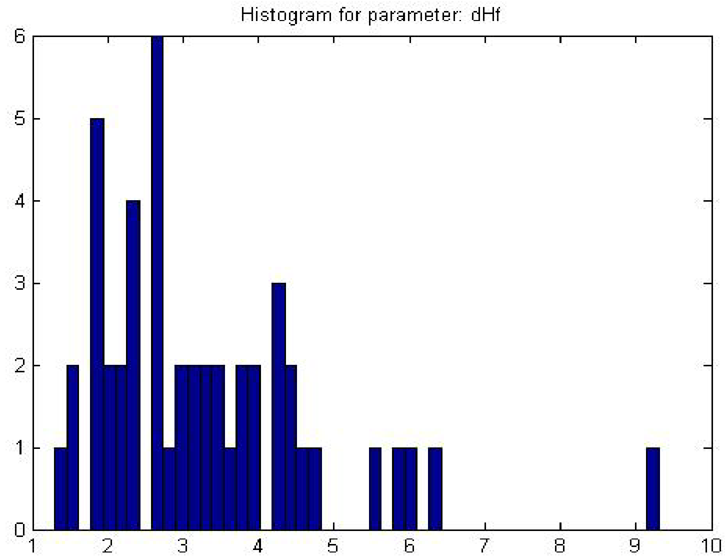

Histogram for parameter: dCr

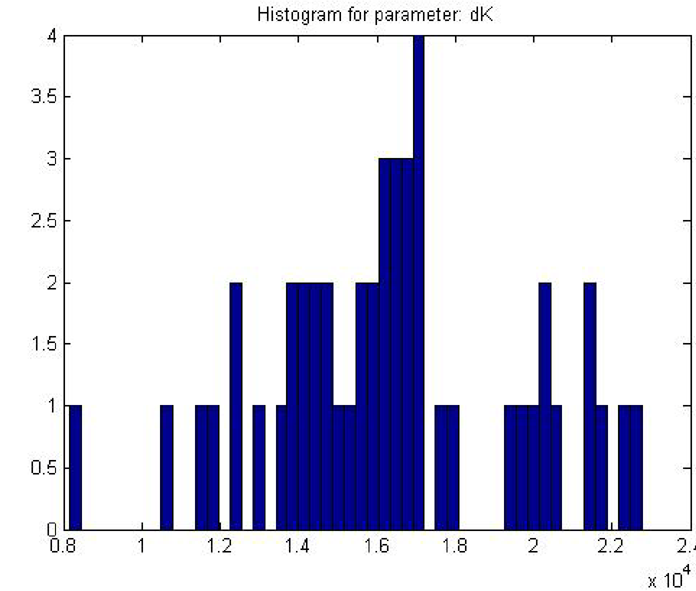

Histogram for parameter: dCs
0 2 4 6 8 10 12 14 16 18 20
0 0.5 1 1.5 2 2.5 3 3.5 4

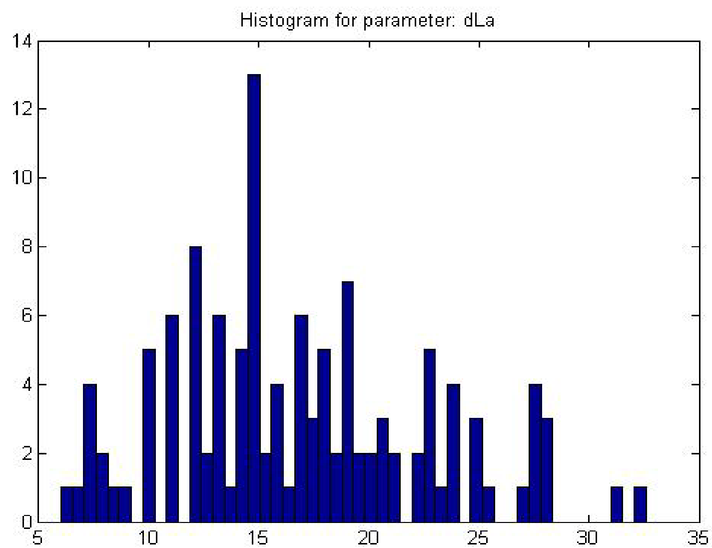

Histogram for parameter: dCu
0 1 2 3 4 5 6 7 8
0 20 40 60 80 100 120 140

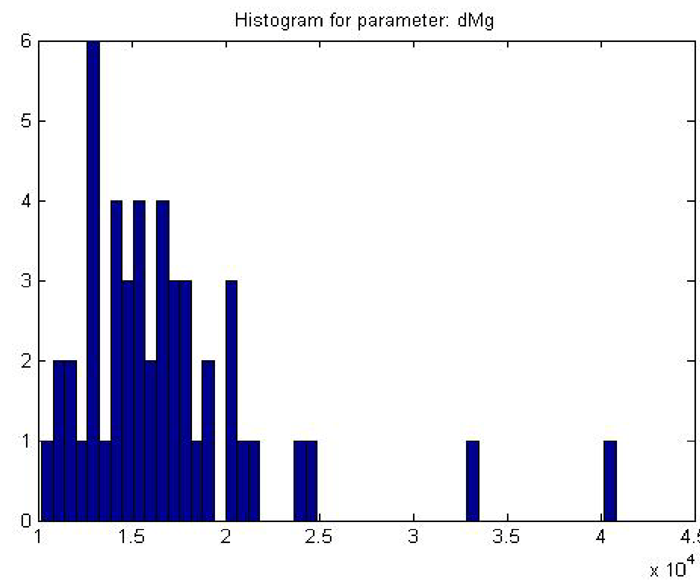

Histogram for parameter: dEu
0 1 2 3 4 5 6
0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8

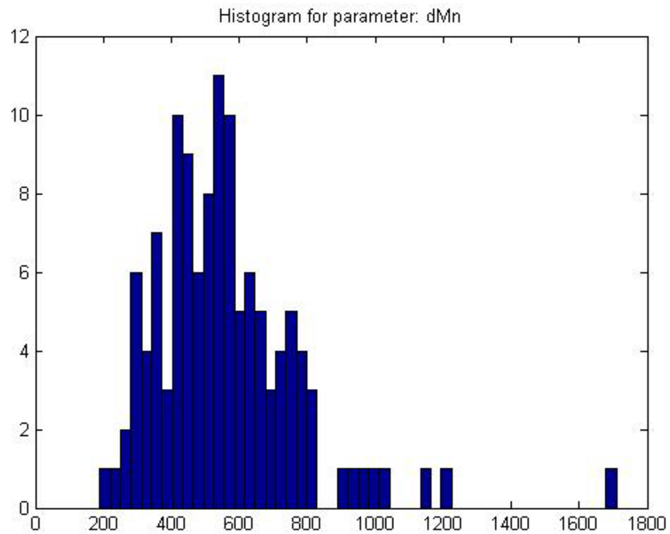

Histogram for parameter: dFe
0 1 2 3 4 5 6 7 8
0.5 1 1.5 2 2.5 3 3.5 4
x 10^4

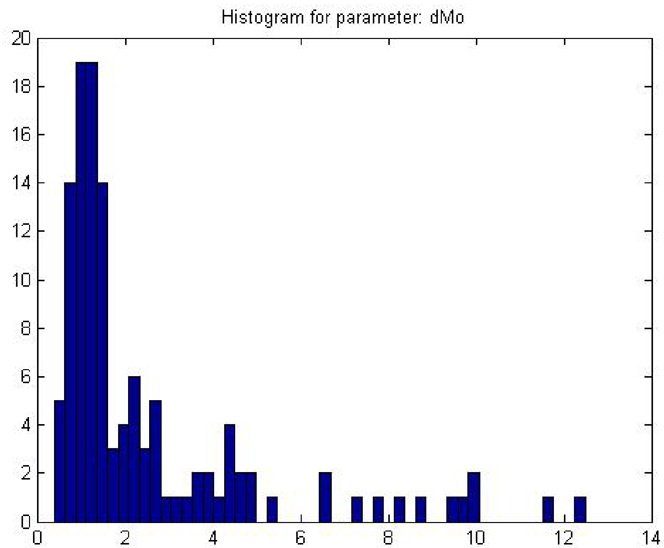

Histogram for parameter: dHf
0
1
2
3
4
5
6
1 2 3 4 5 6 7 8 9 10

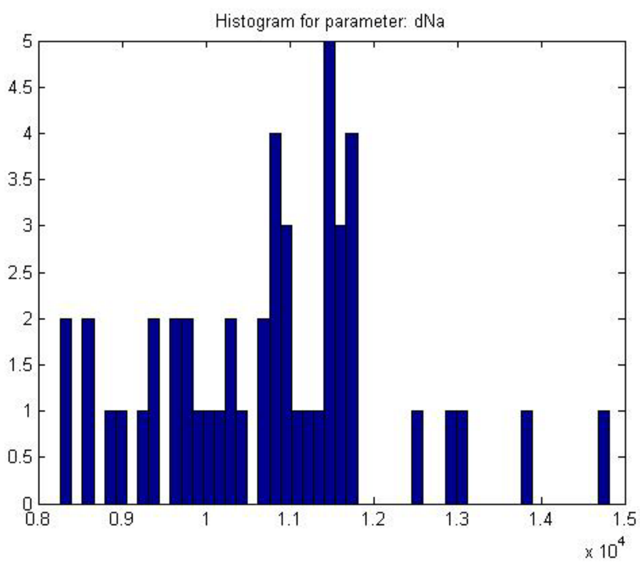

Histogram for parameter: dK
0
0.5
1
1.5
2
2.5
3
3.5
4
0.8 1 1.2 1.4 1.6 1.8 2 2.2 2.4
x 10^4

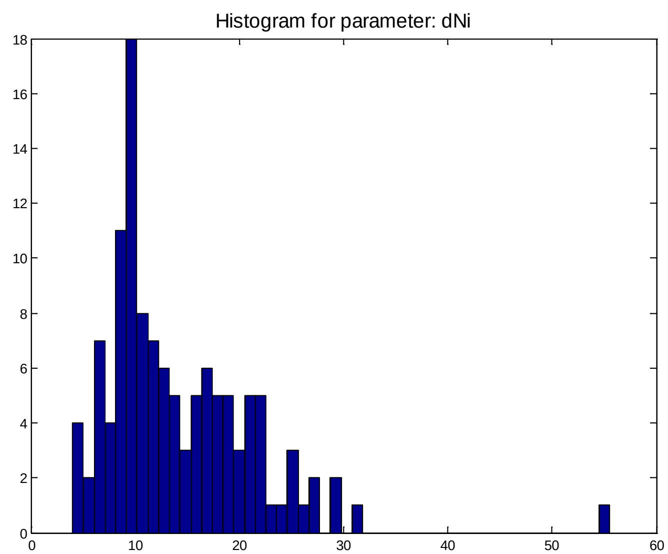

Histogram for parameter: dLa
0
2
4
6
8
10
12
14
5 10 15 20 25 30 35

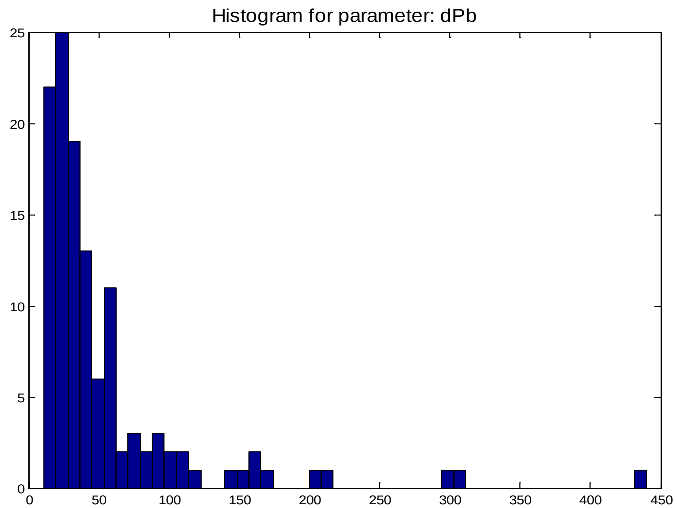

Histogram for parameter: dMg
0
1
2
3
4
5
6
1 1.5 2 2.5 3 3.5 4 4.5
x 10^4

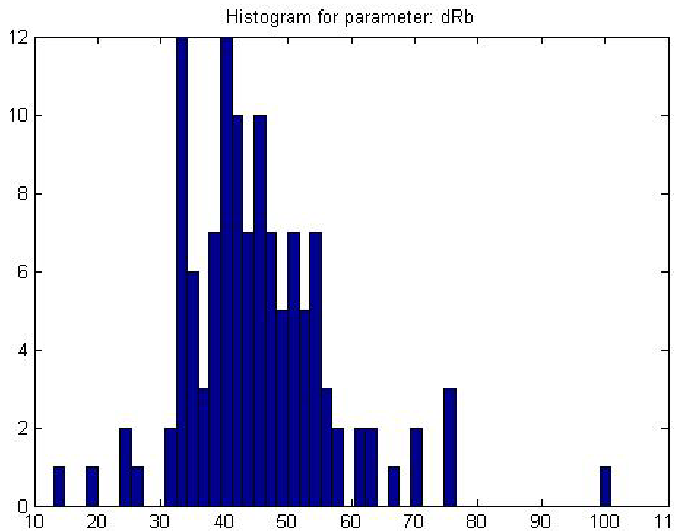

Histogram for parameter: dMn
0 2 4 6 8 10 12
0 200 400 600 800 1000 1200 1400 1600 1800

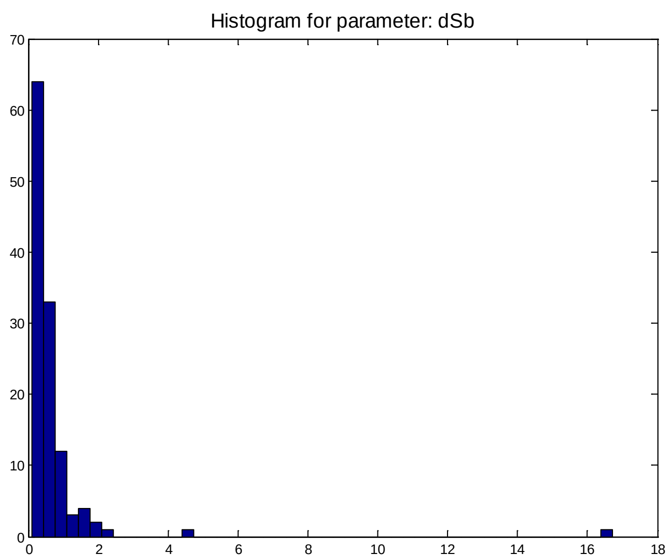

Histogram for parameter: dMo
0 2 4 6 8 10 12 14 16 18 20
0 2 4 6 8 10 12 14

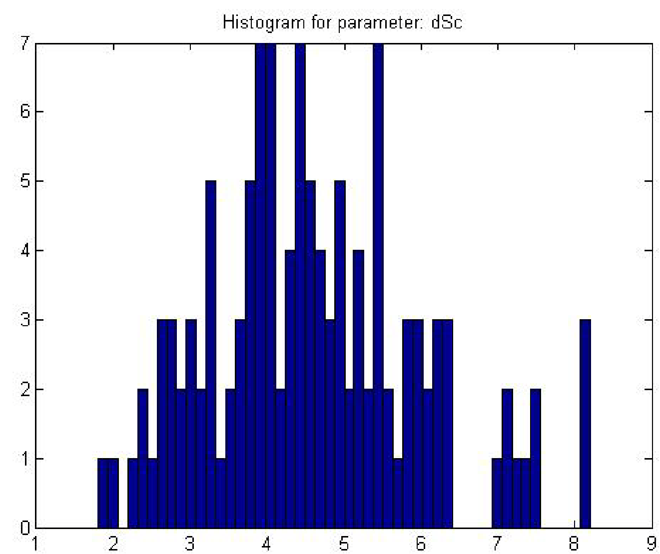

Histogram for parameter: dNa
0 0.5 1 1.5 2 2.5 3 3.5 4 4.5 5
0.8 0.9 1 1.1 1.2 1.3 1.4 1.5
x 10^4

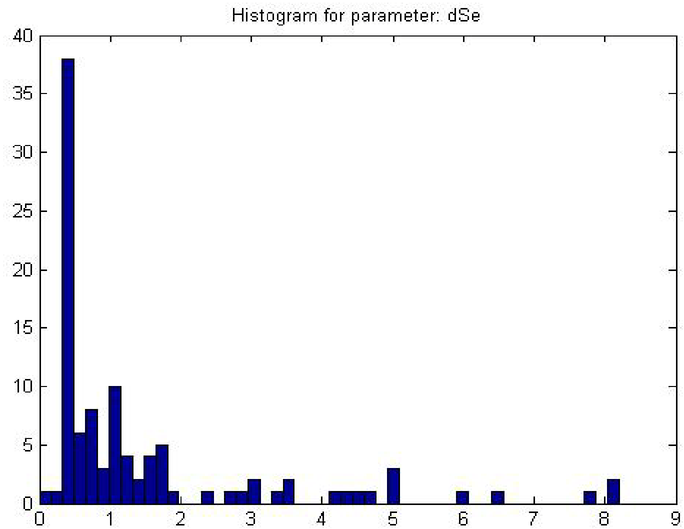

Histogram for parameter: dNi
0 2 4 6 8 10 12 14 16 18
0 10 20 30 40 50 60

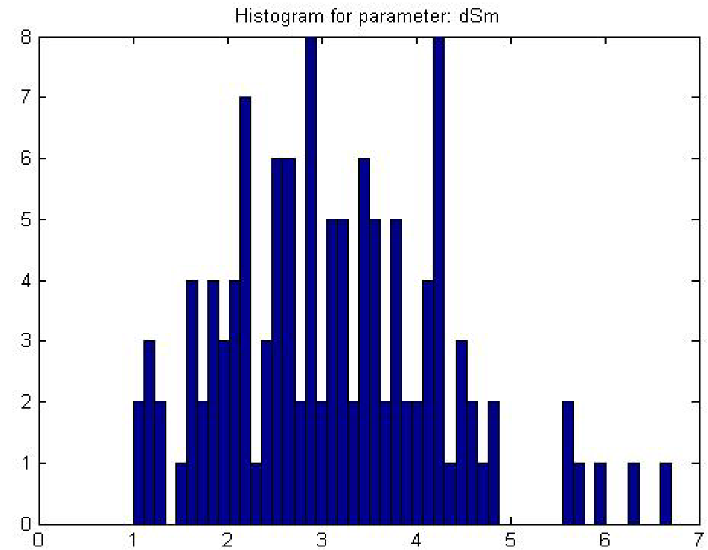

Histogram for parameter: dPb
0 5 10 15 20 25
0 50 100 150 200 250 300 350 400 450

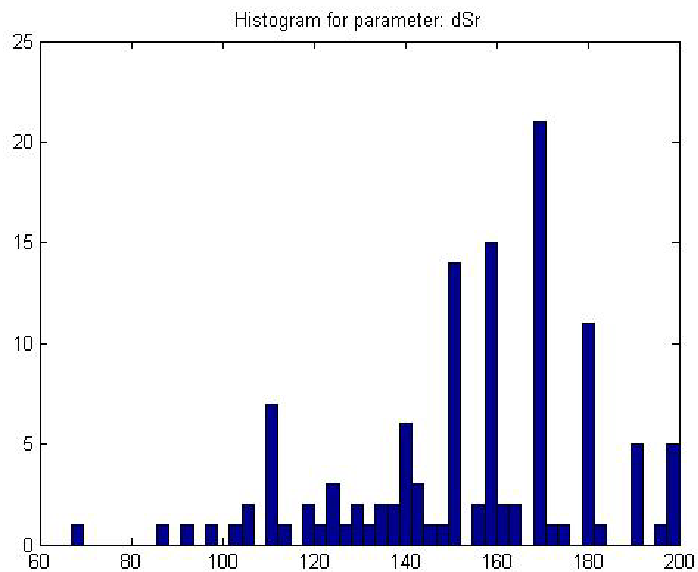

Histogram for parameter: dRb
0 2 4 6 8 10 12
10 20 30 40 50 60 70 80 90 100 110

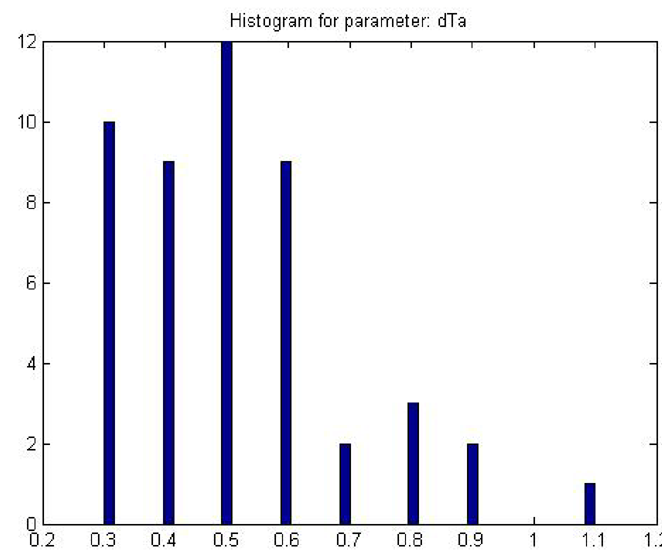

Histogram for parameter: dSb
0 10 20 30 40 50 60 70
0 2 4 6 8 10 12 14 16 18

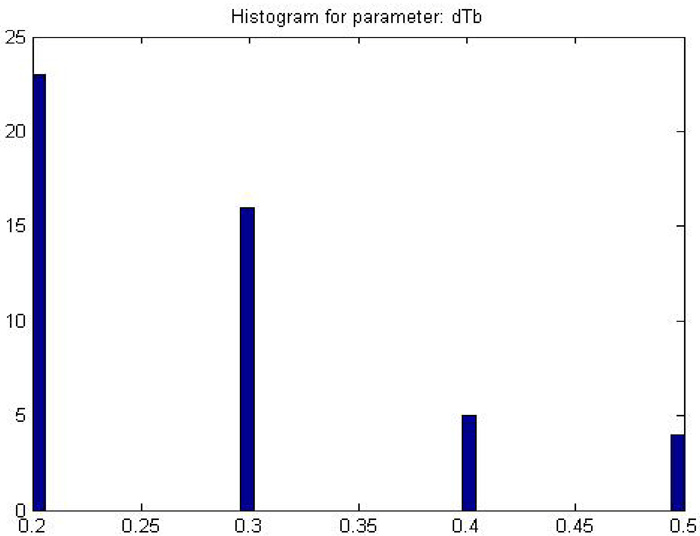

Histogram for parameter: dSc
0 1 2 3 4 5 6 7
1 2 3 4 5 6 7 8 9

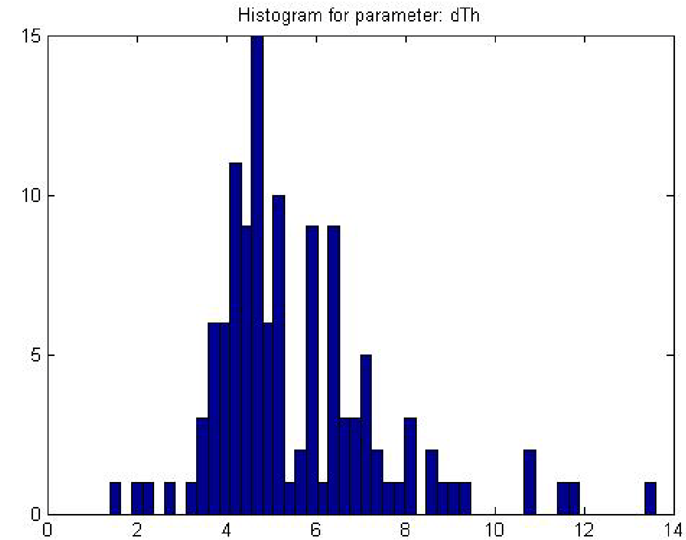

Histogram for parameter: dSe

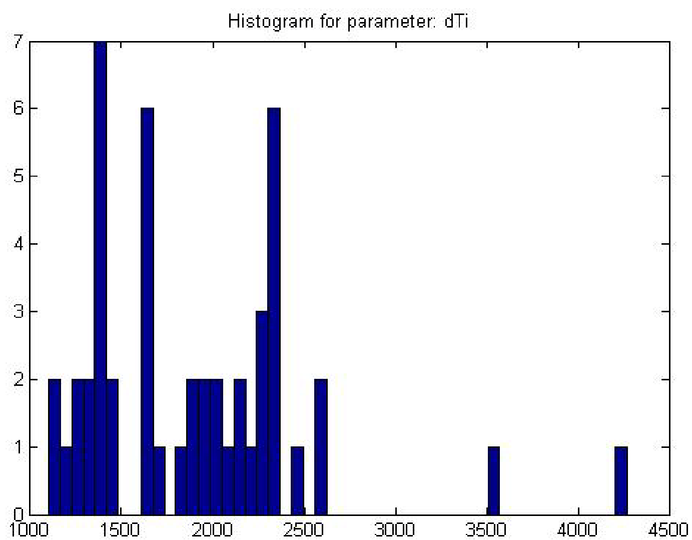

Histogram for parameter: dSm

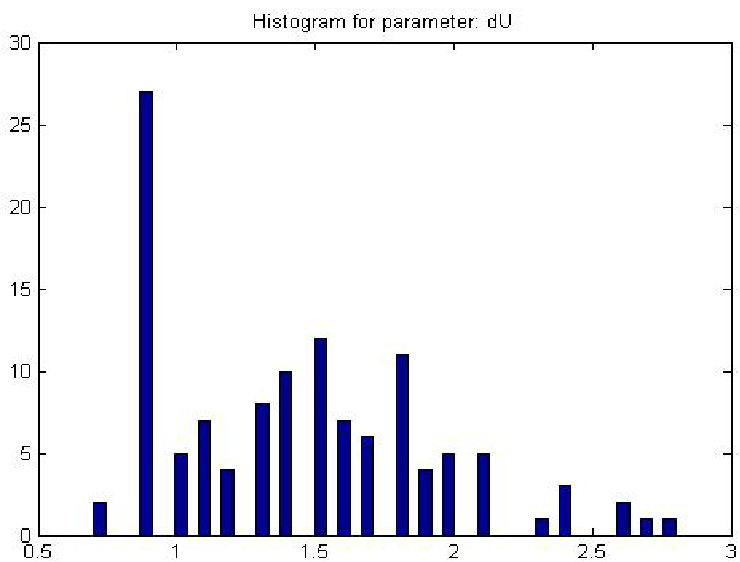

Histogram for parameter: dSr

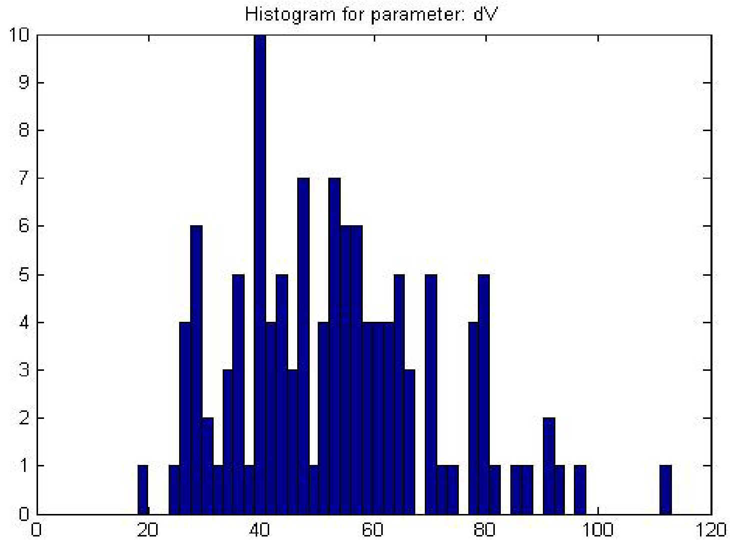

Histogram for parameter: dTa

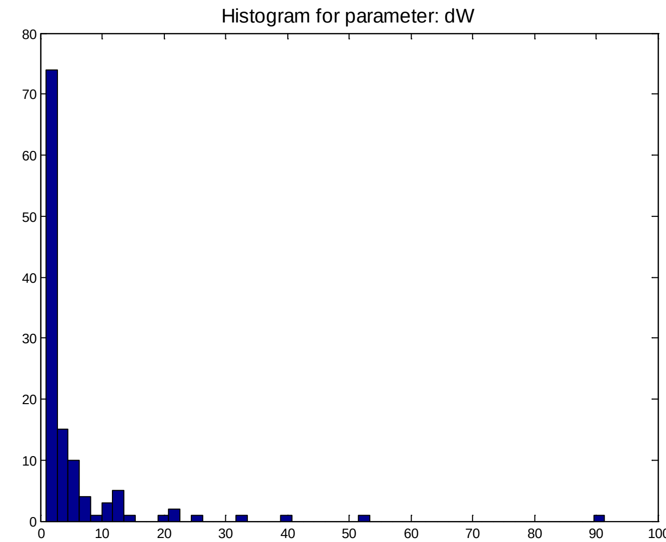

Histogram for parameter: dTb
25
20
15
10
5
0
0.2
0.25
0.3
0.35
0.4
0.45
0.5

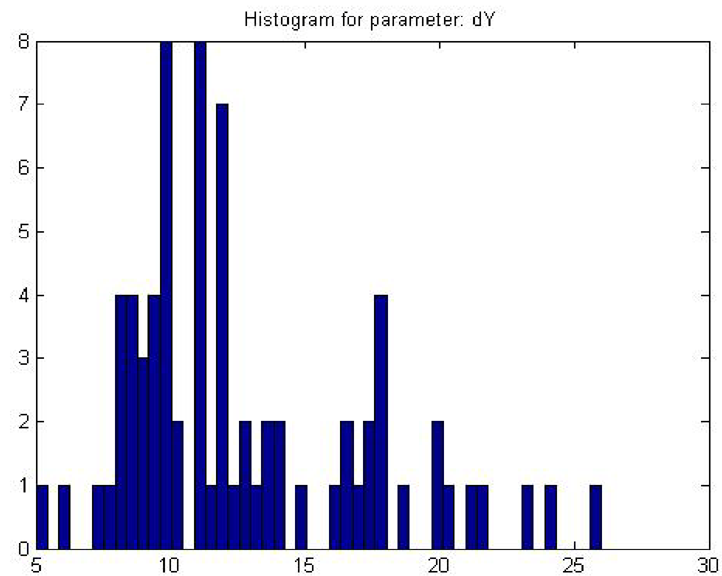

Histogram for parameter: dTh
15
10
5
0
0
2
4
6
8
10
12
14

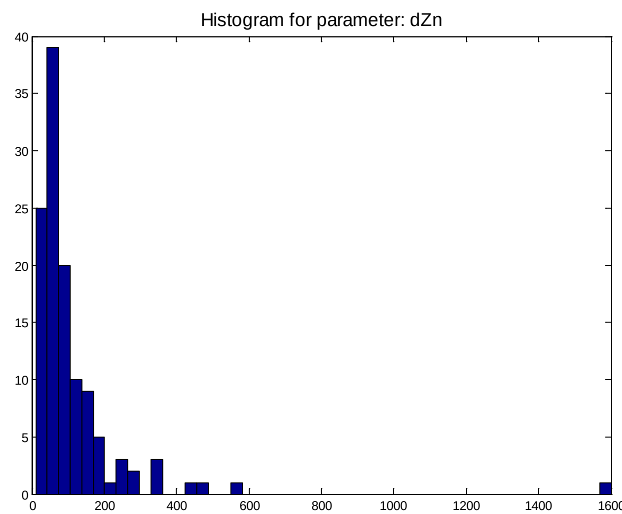

Histogram for parameter: dTi
7
6
5
4
3
2
1
0
1000
1500
2000
2500
3000
3500
4000
4500


Histogram for parameter: dU
30
25
20
15
10
5
0
0.5
1
1.5
2
2.5
3

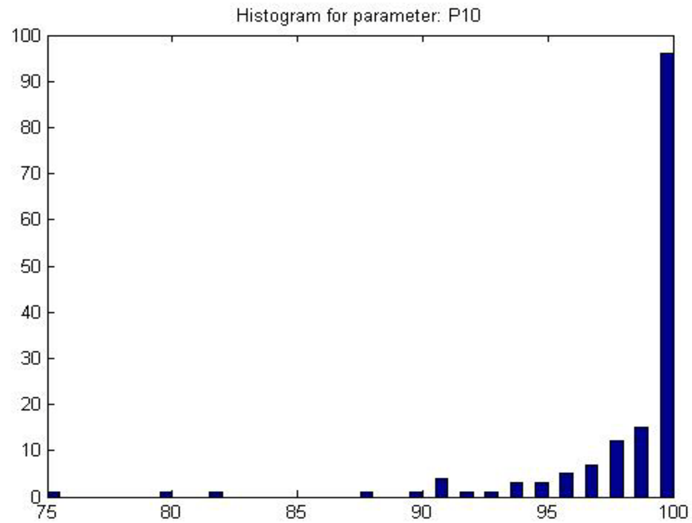

Histogram for parameter: dV

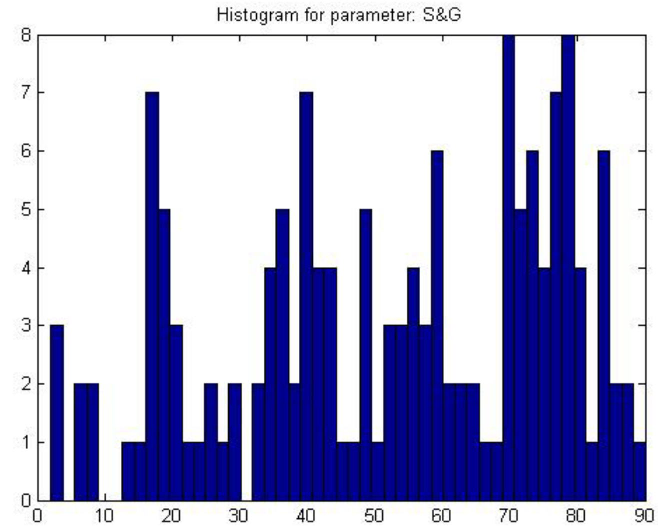

Histogram for parameter: dW

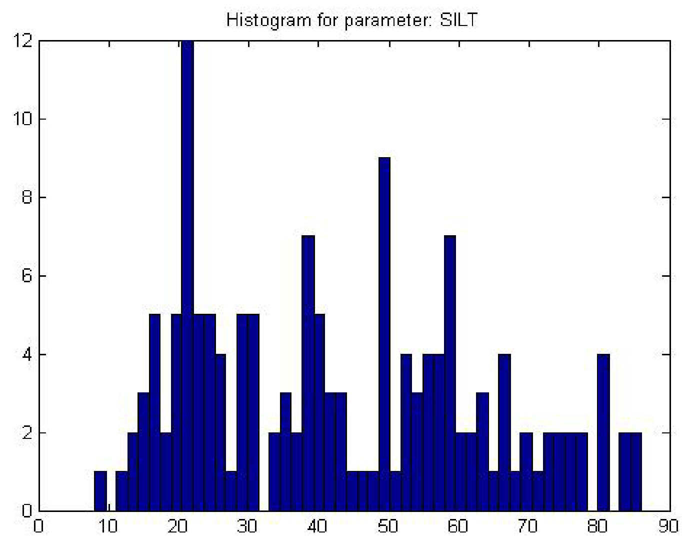

Histogram for parameter: dY

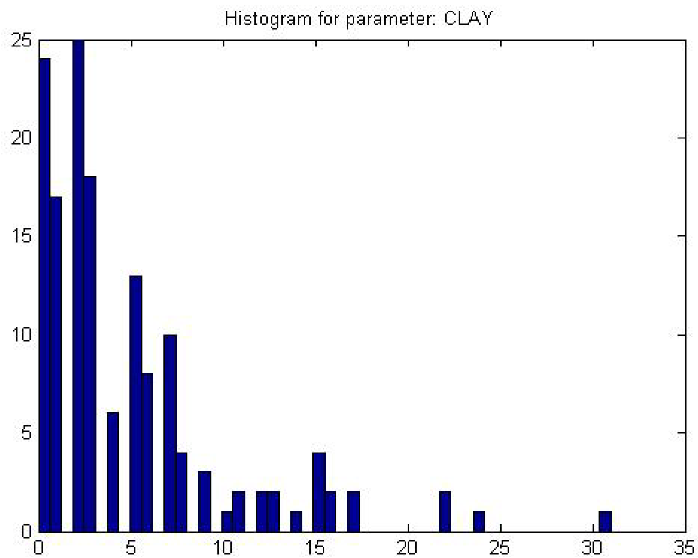

Histogram for parameter: dZn

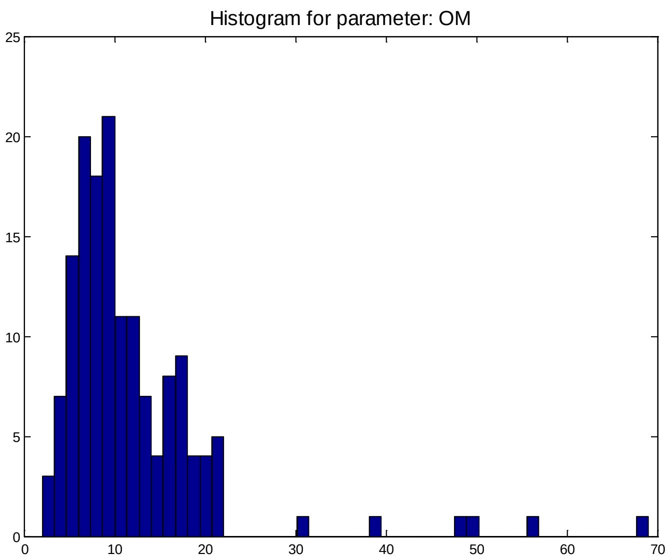
MN Statewide 2003 Soil Survey Data

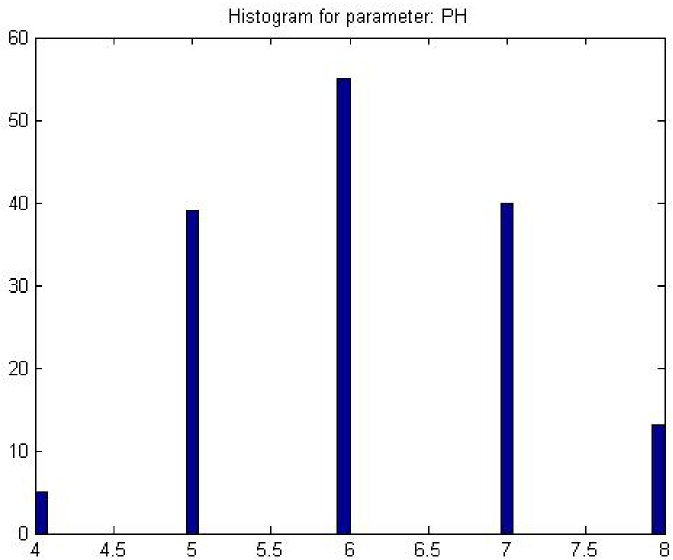


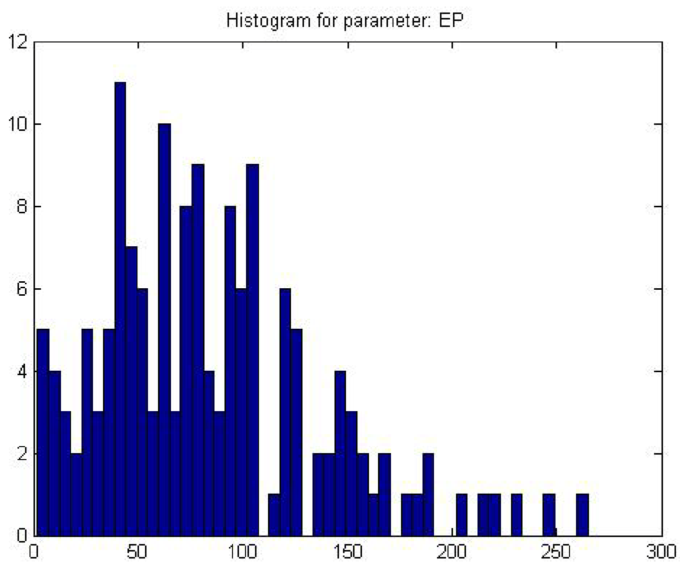


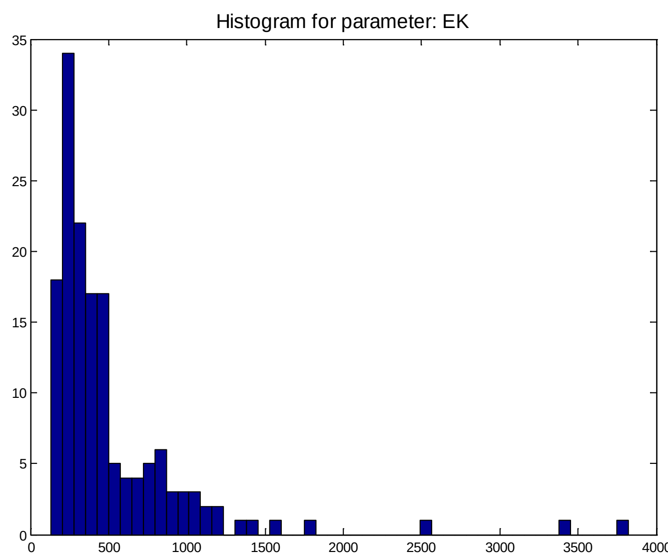


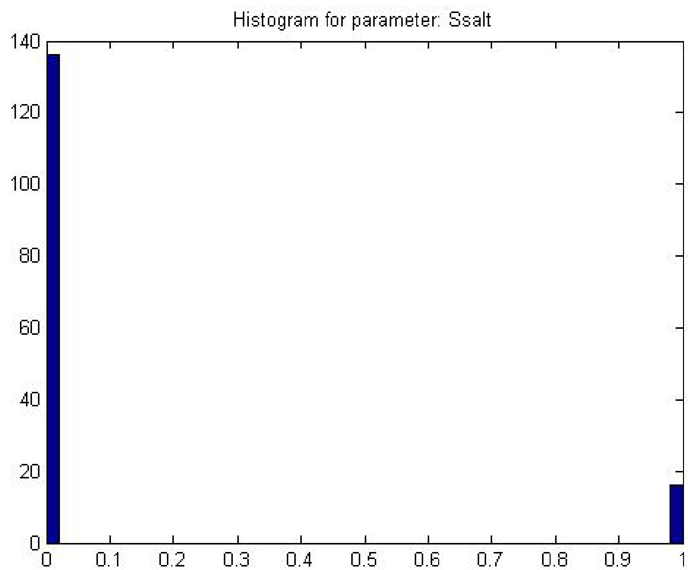


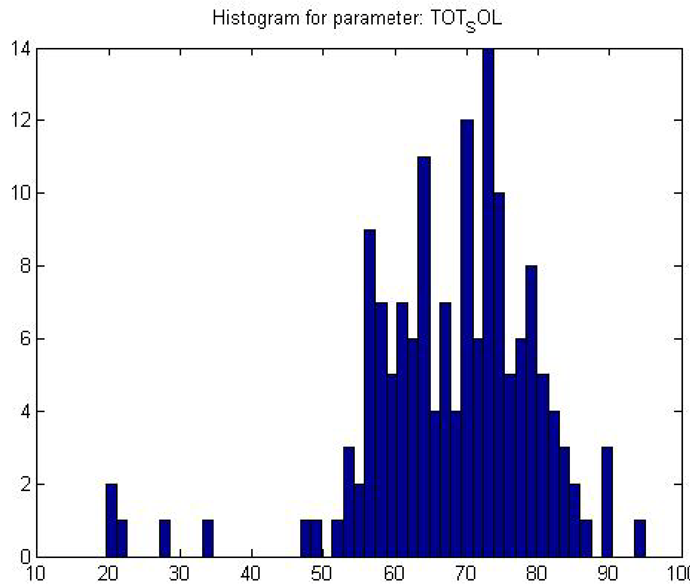

Histogram for parameter: OM
0
5
10
15
20
25
0
10
20
30
40
50
60
70

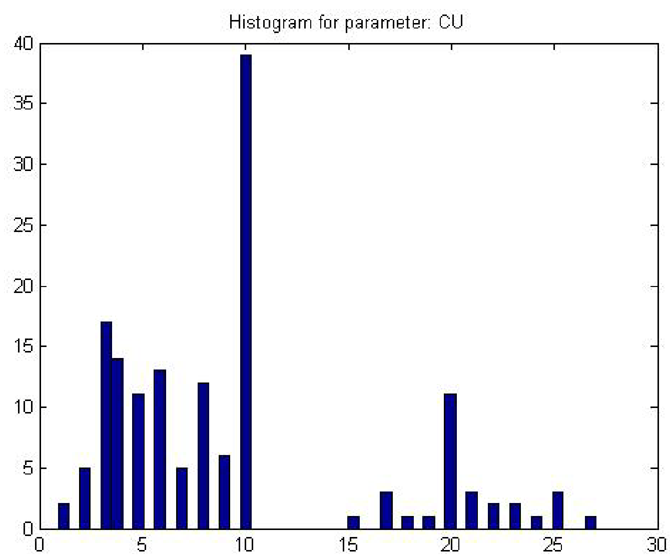

Histogram for parameter: PH
0
10
20
30
40
50
60
4
4.5
5
5.5
6
6.5
7
7.5
8

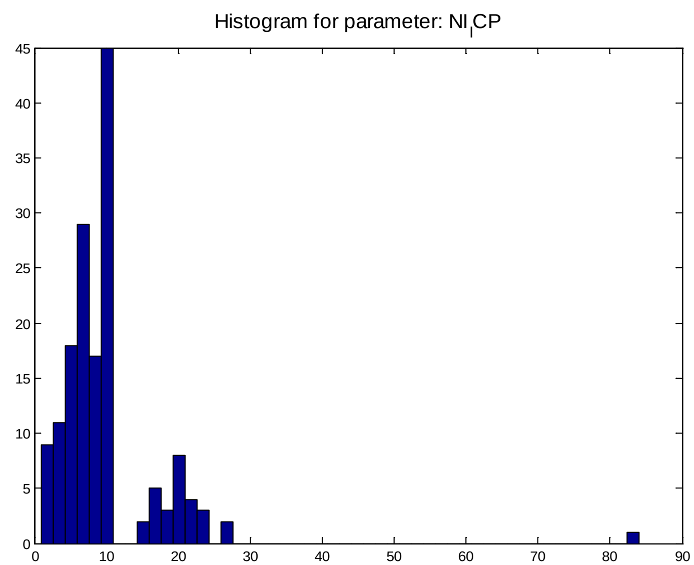

Histogram for parameter: EP
0
2
4
6
8
10
12
0
50
100
150
200
250
300

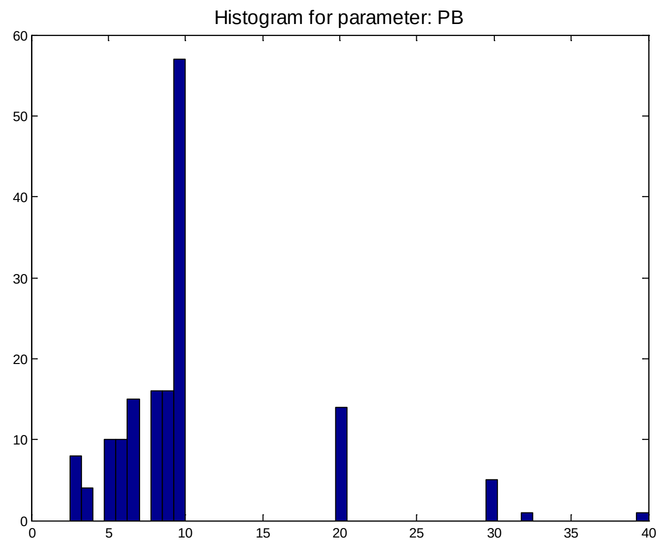

Histogram for parameter: EK
0
5
10
15
20
25
30
35
0
500
1000
1500
2000
2500
3000
3500
4000

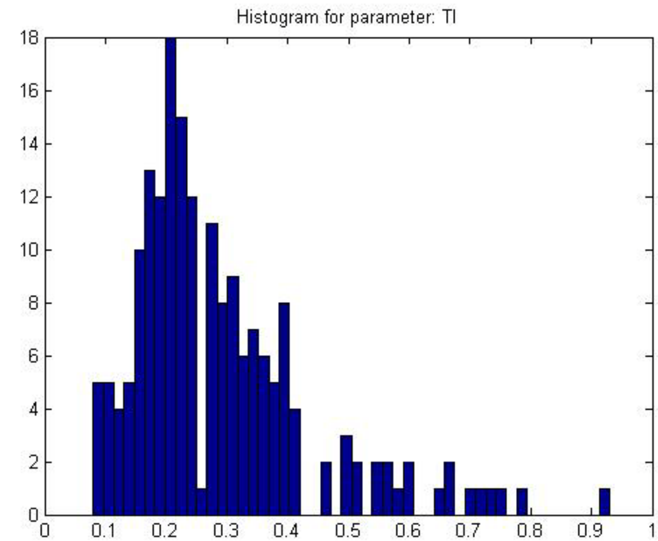

Histogram for parameter: Ssalt

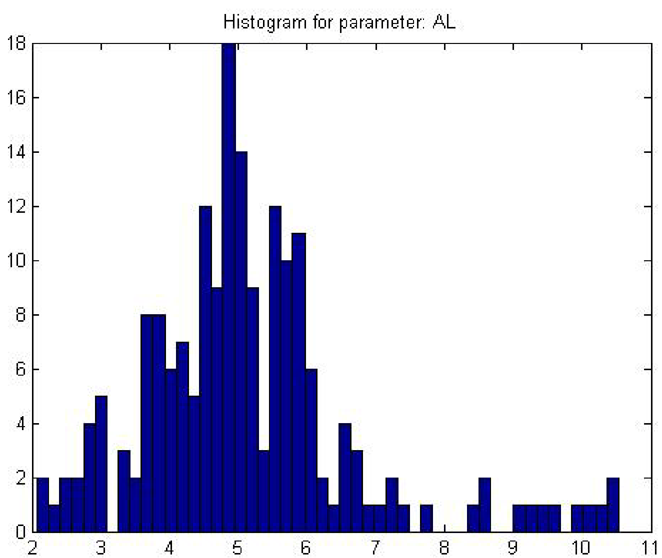

Histogram for parameter: TOTSOL

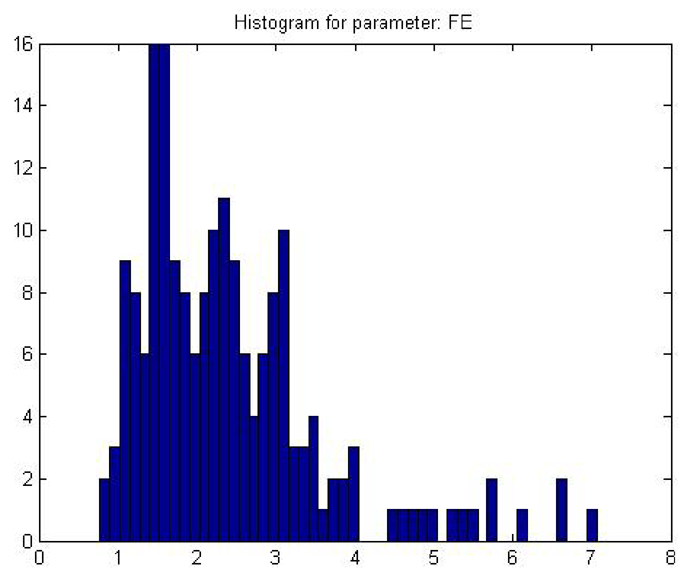

Histogram for parameter: CU

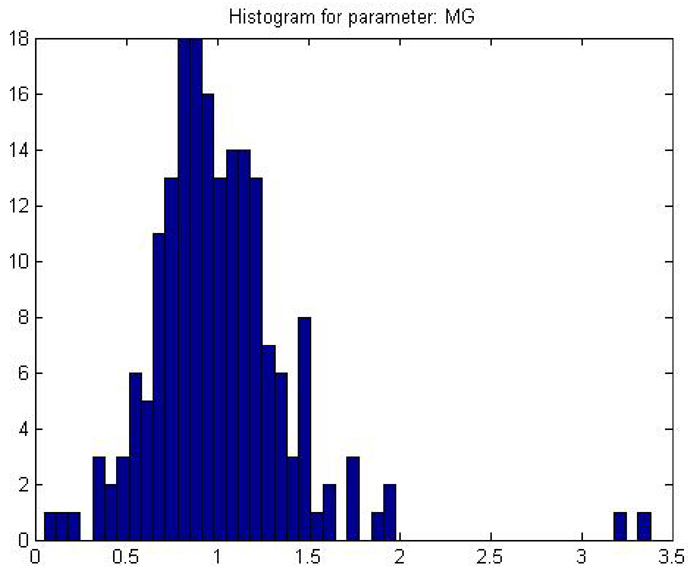

Histogram for parameter: NICP

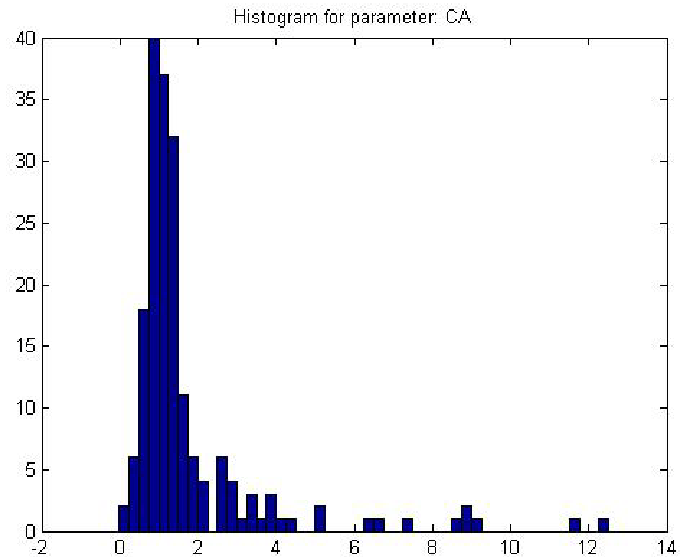

Histogram for parameter: PB
0
5
10
15
20
25
30
35
40
60
50
40
30
20
10

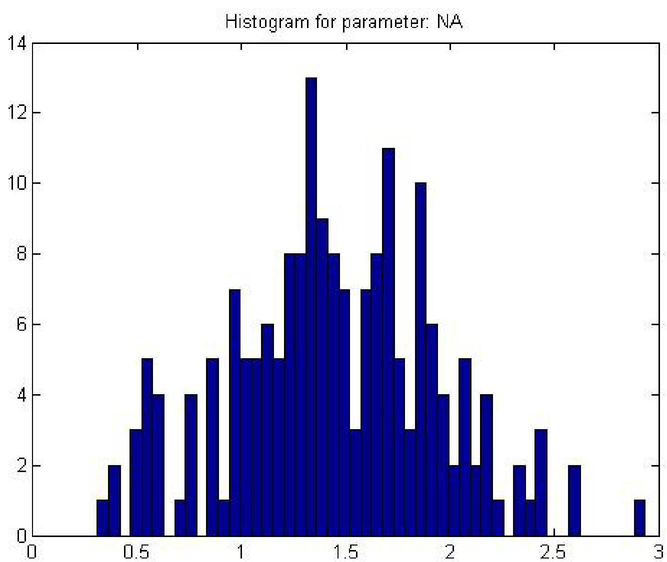

Histogram for parameter: TI
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
18
16
14
12
10
8
6
4
2

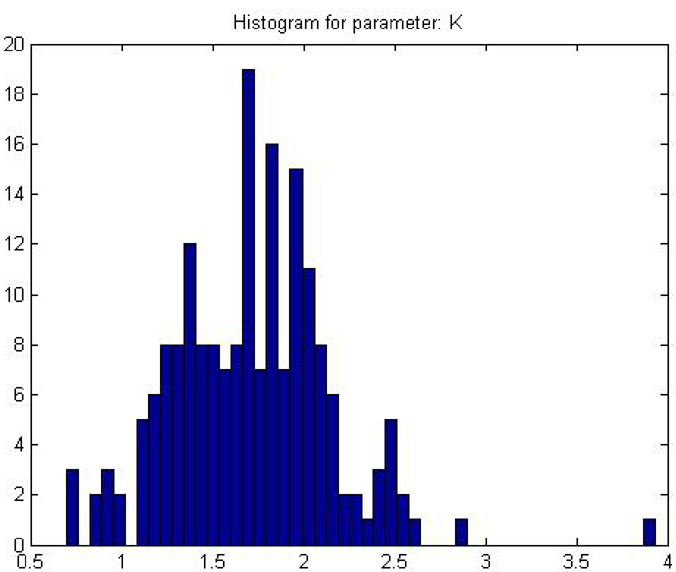

Histogram for parameter: AL
2
3
4
5
6
7
8
9
10
11
18
16
14
12
10
8
6
4
2
0

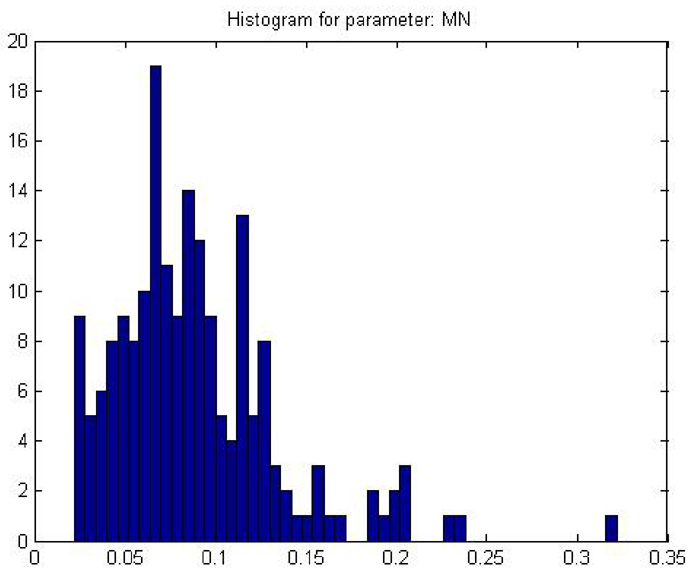

Histogram for parameter: FE
0
1
2
3
4
5
6
7
8
16
14
12
10
8
6
4
2

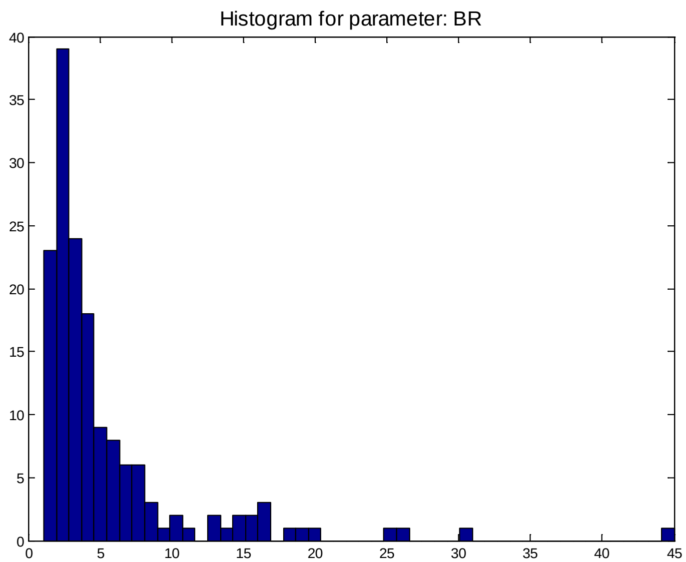

Histogram for parameter: MG
0 2 4 6 8 10 12 14 16 18
0 0.5 1 1.5 2 2.5 3 3.5

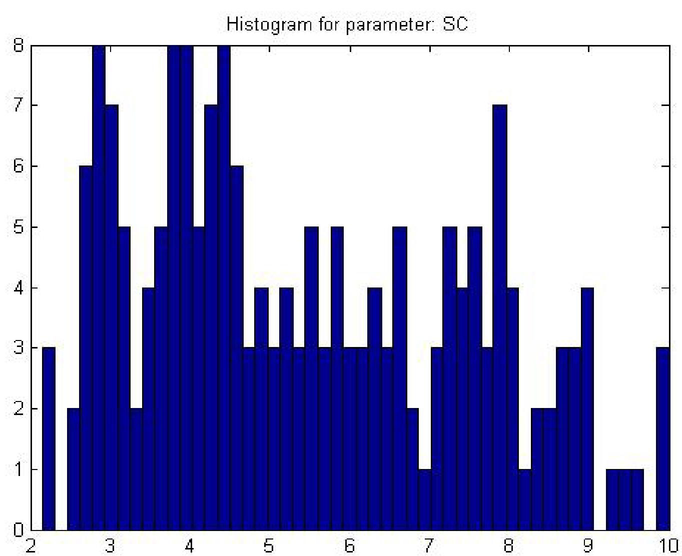

Histogram for parameter: CA
0 5 10 15 20 25 30 35 40
-2 0 2 4 6 8 10 12 14

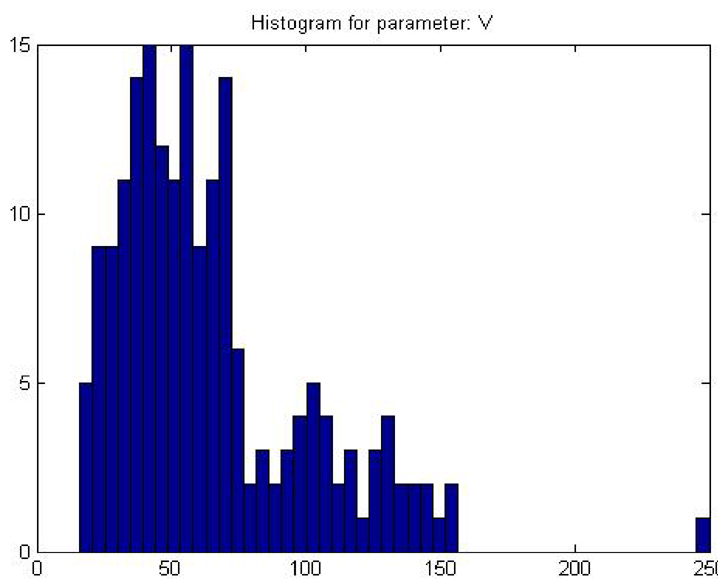

Histogram for parameter: NA
0 2 4 6 8 10 12 14
0 0.5 1 1.5 2 2.5 3

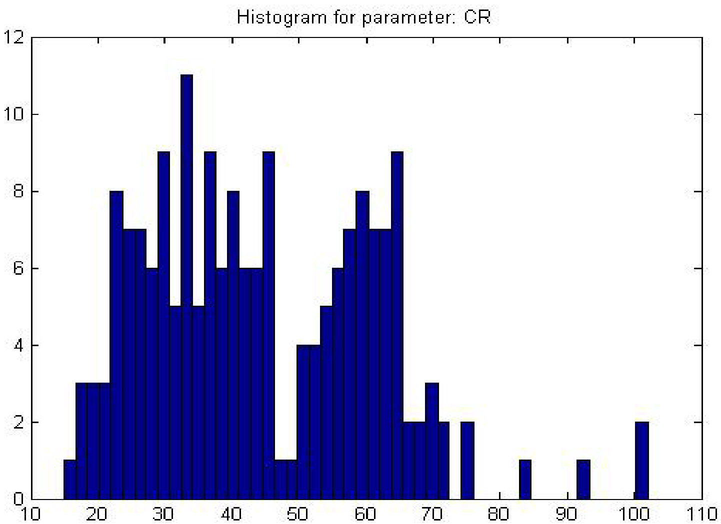

Histogram for parameter: K
0 2 4 6 8 10 12 14 16 18 20
0.5 1 1.5 2 2.5 3 3.5 4

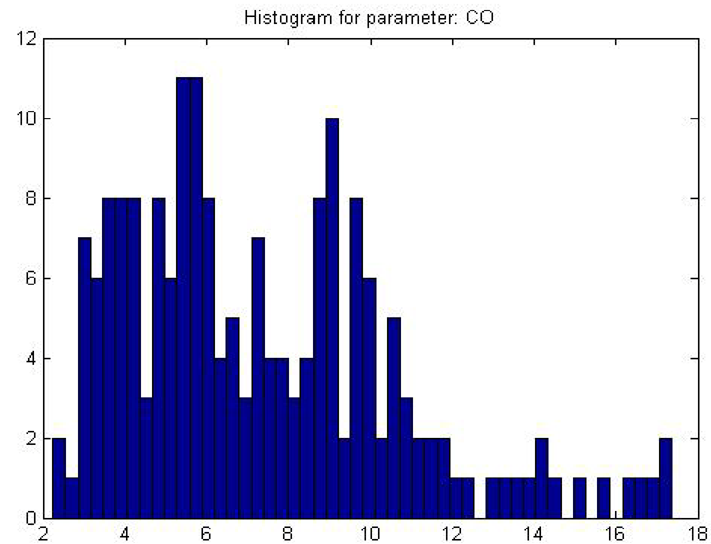

Histogram for parameter: MN

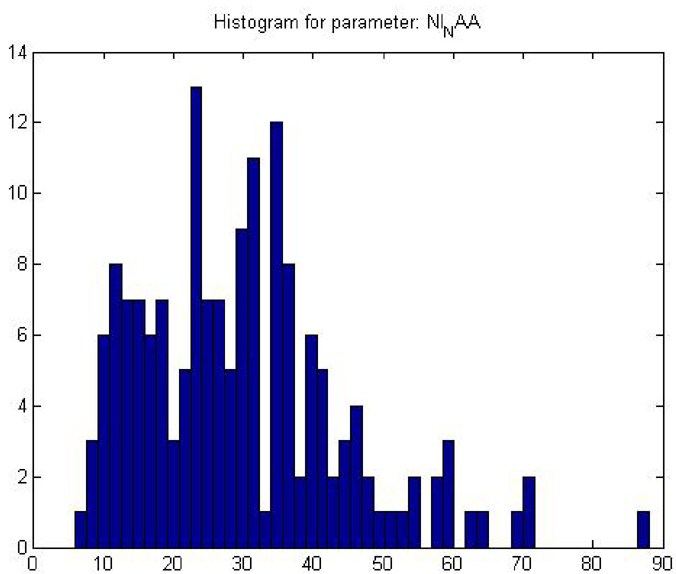

Histogram for parameter: BR

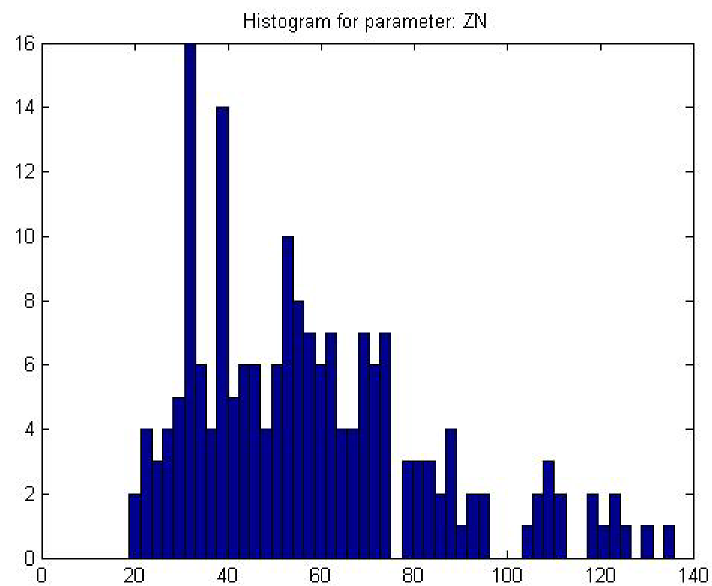

Histogram for parameter: SC

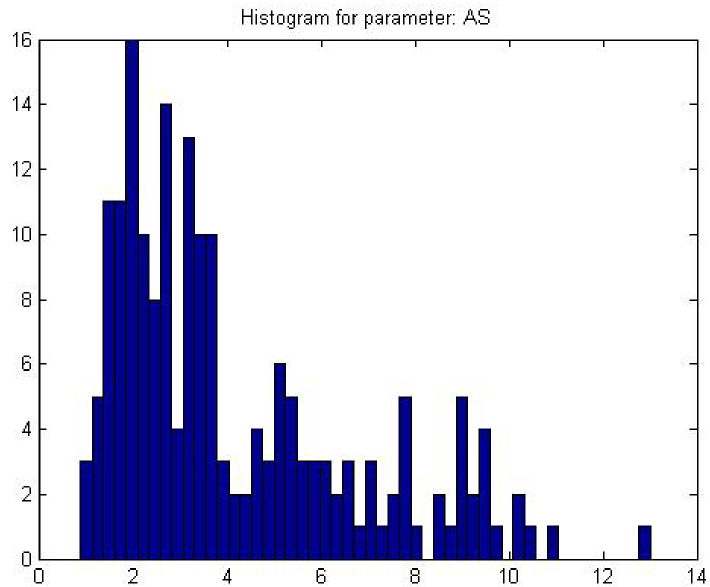

Histogram for parameter: V

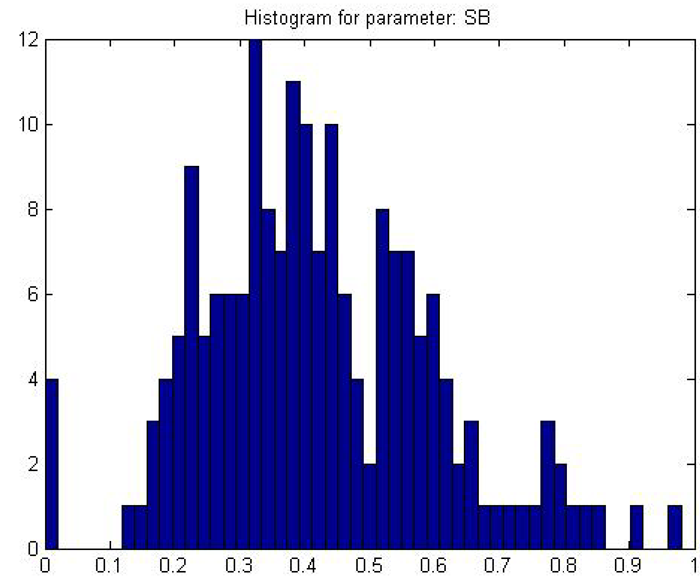

Histogram for parameter: CR

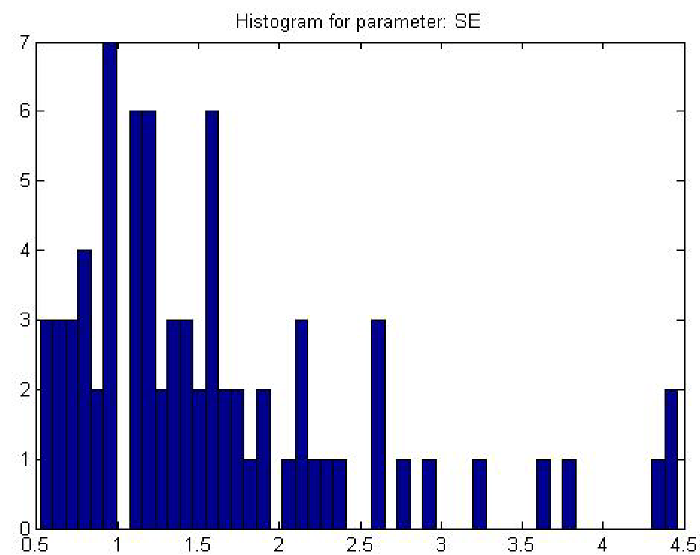

Histogram for parameter: CO

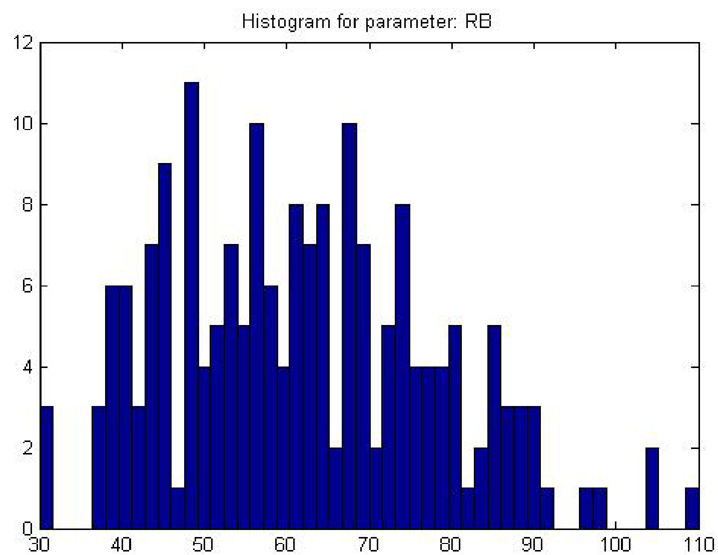

Histogram for parameter: NI_NAA

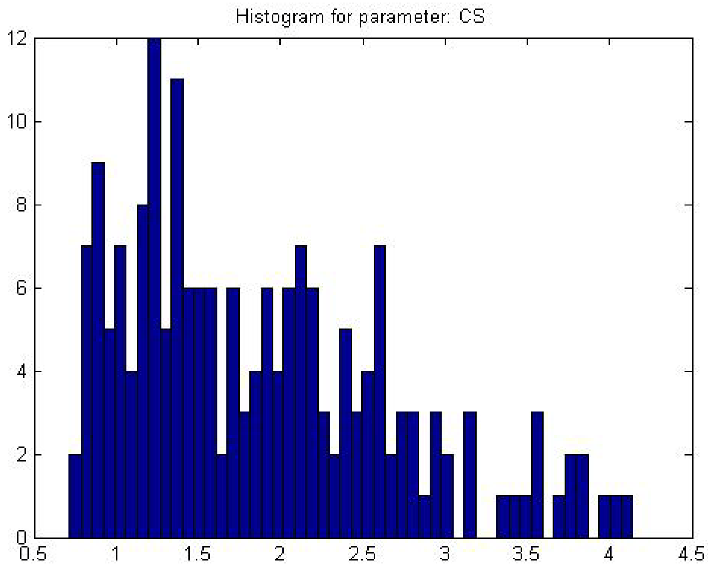

Histogram for parameter: ZN

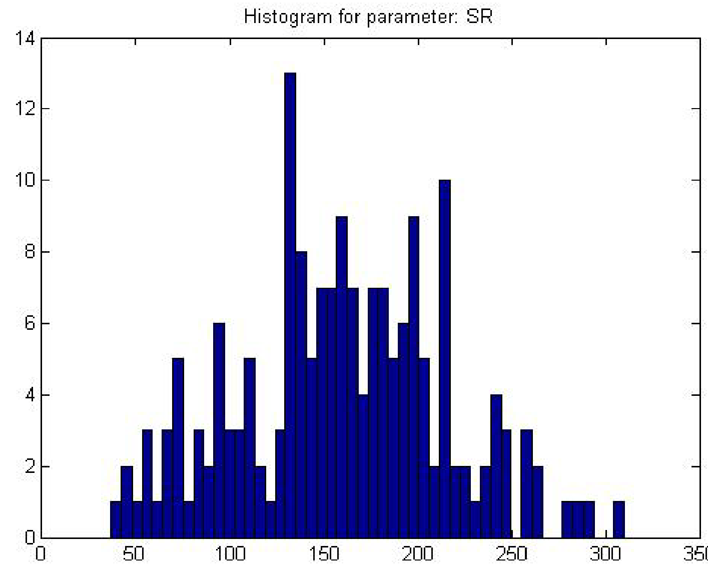

Histogram for parameter: AS
0
2
4
6
8
10
12
14
16

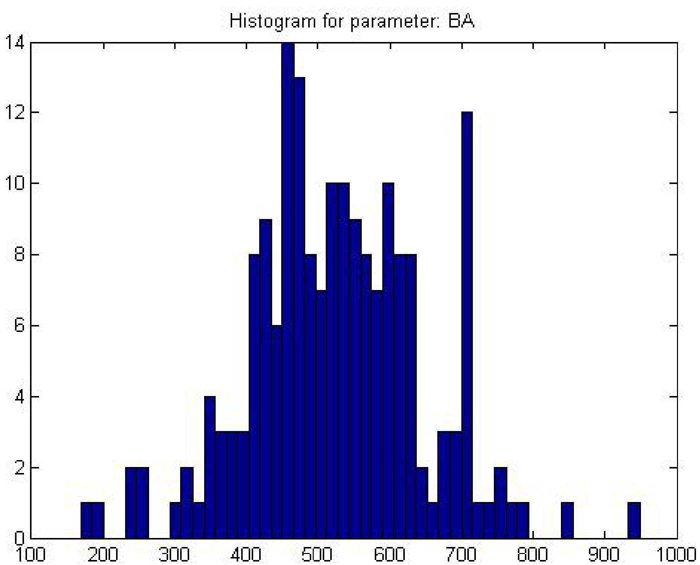

Histogram for parameter: SB
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
2
4
6
8
10
12

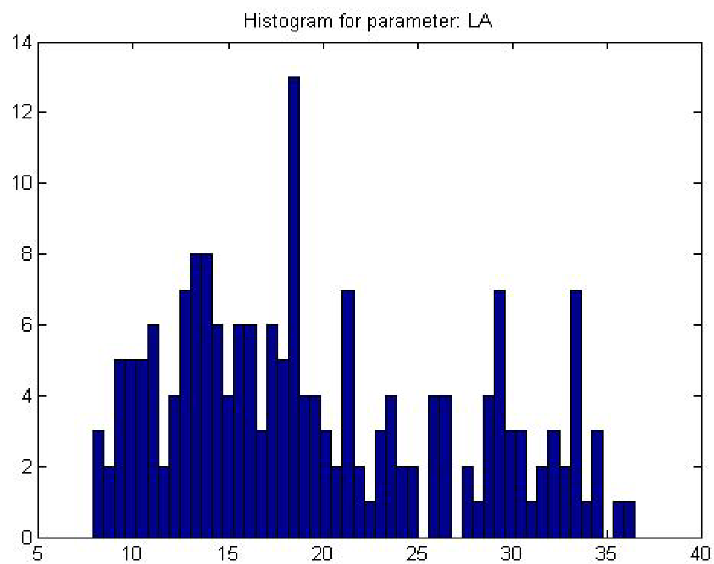

Histogram for parameter: SE
0.5
1
1.5
2
2.5
3
3.5
4
4.5
0
5
6
7

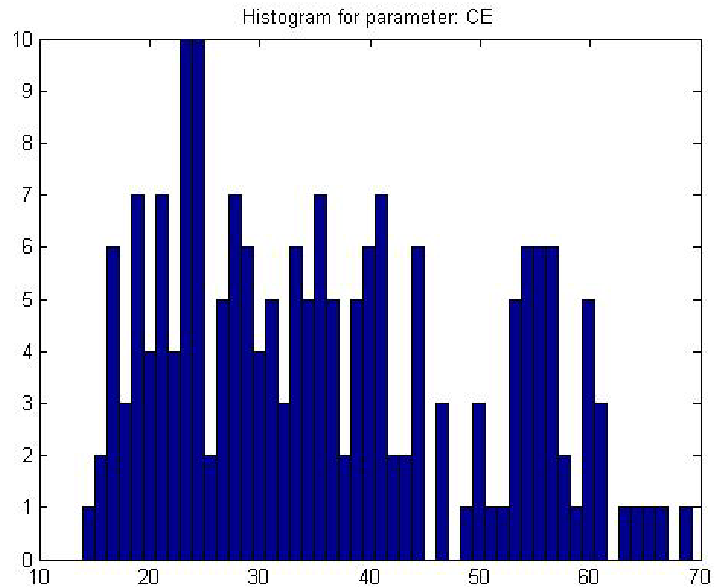

Histogram for parameter: RB
30
40
50
60
70
80
90
100
110
0
2
4
6
8
10
12

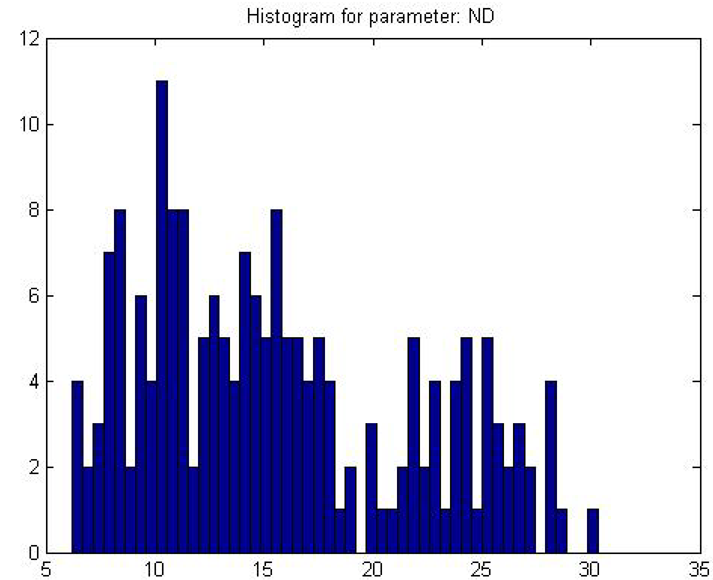

Histogram for parameter: CS
0
2
4
6
8
10
12
0.5
1
1.5
2
2.5
3
3.5
4
4.5

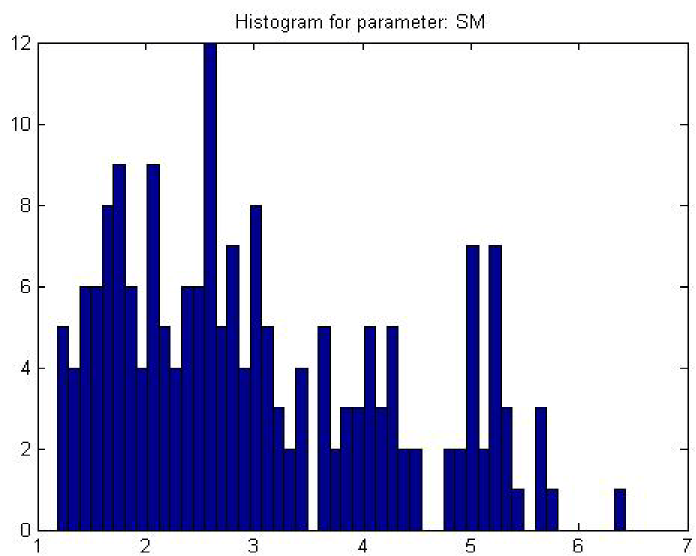

Histogram for parameter: SR
0
2
4
6
8
10
12
14
0
50
100
150
200
250
300
350

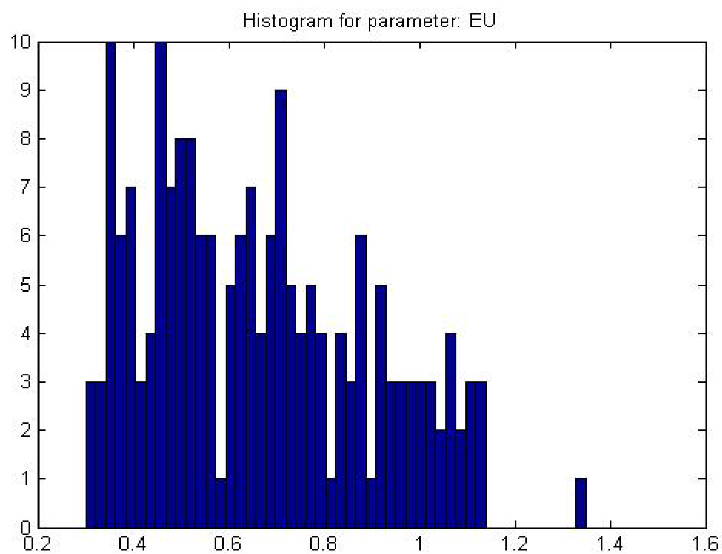

Histogram for parameter: BA
0
2
4
6
8
10
12
14
100
200
300
400
500
600
700
800
900
1000

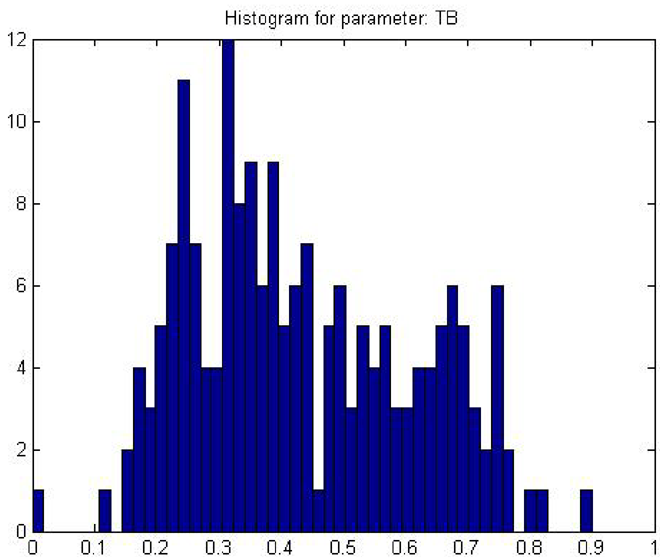

Histogram for parameter: LA
0
2
4
6
8
10
12
14
5
10
15
20
25
30
35
40

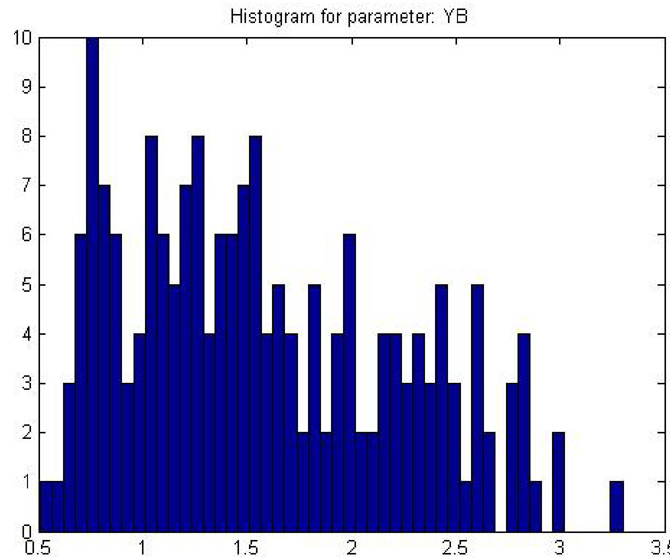

Histogram for parameter: CE

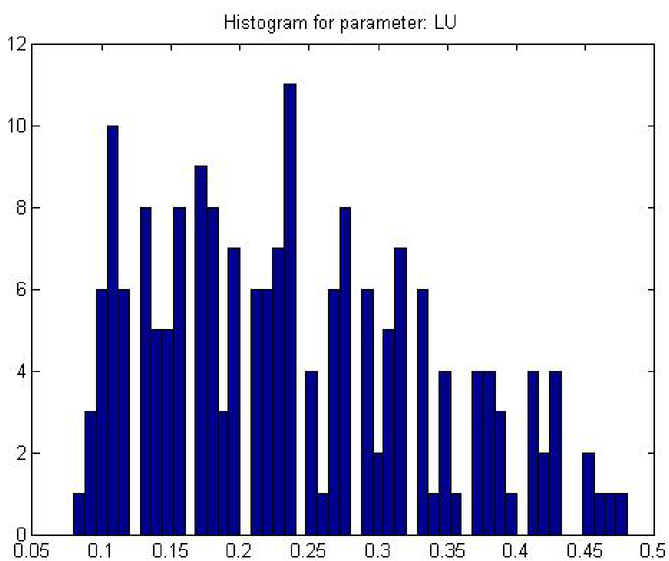

Histogram for parameter: ND

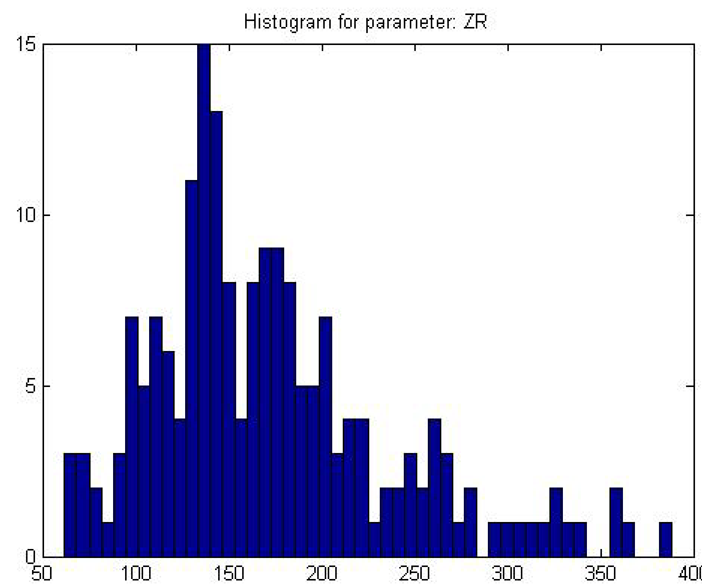

Histogram for parameter: SM

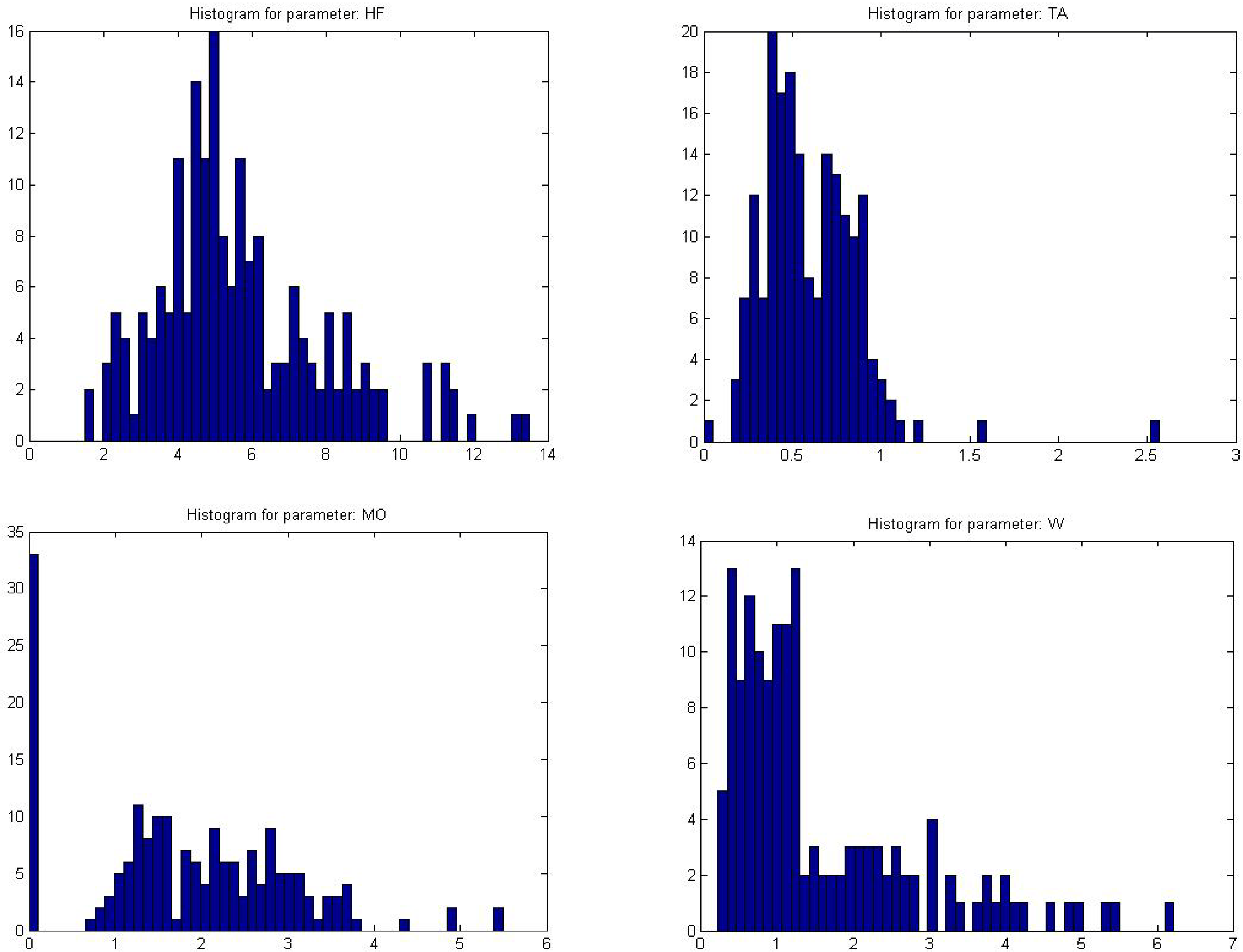

Histogram for parameter: EU

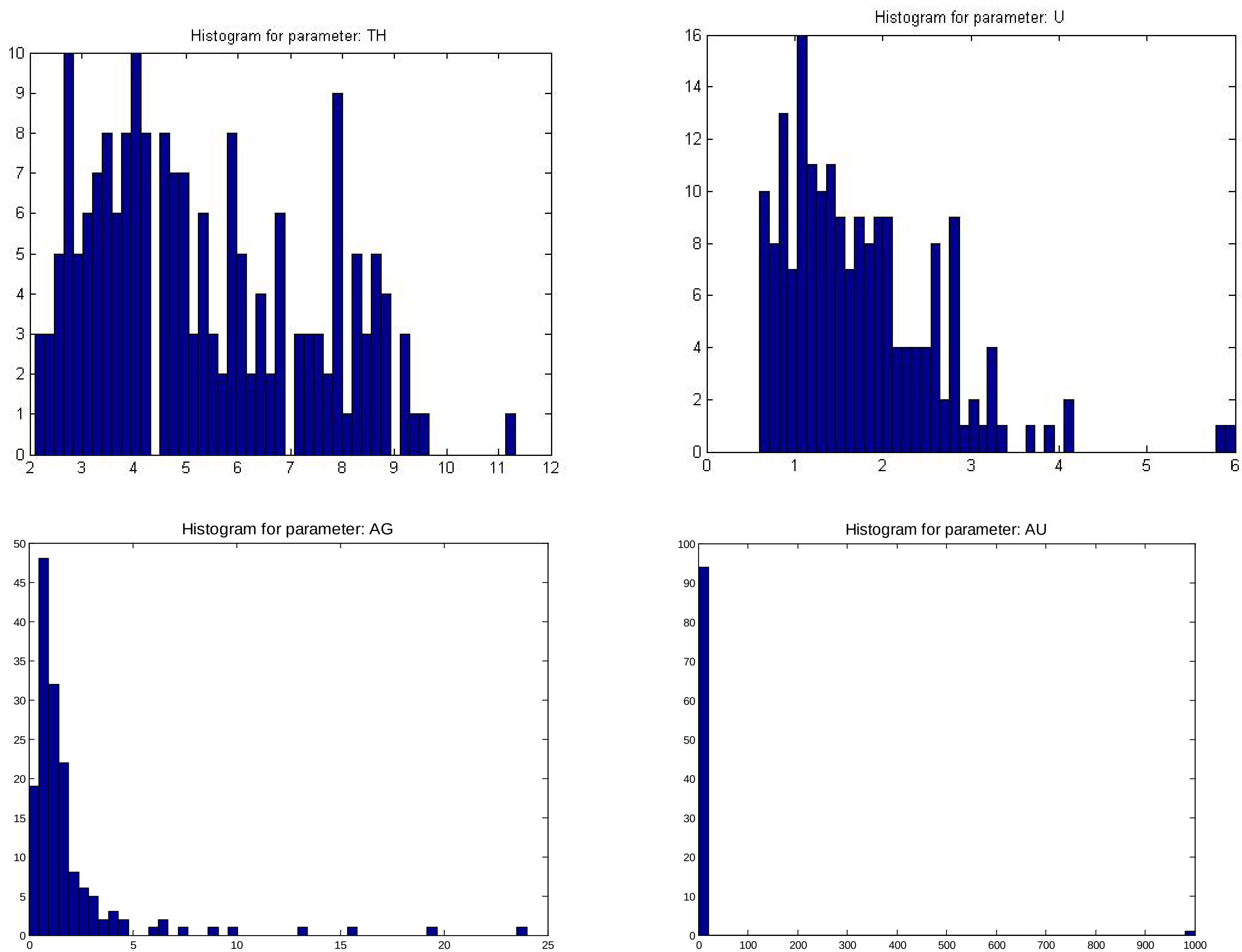

Histogram for parameter: TB
0 2 4 6 8 10 12
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1

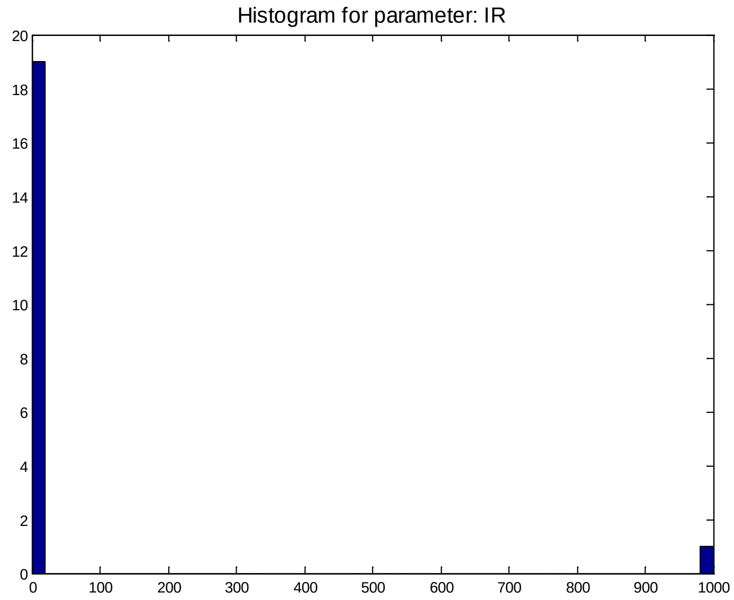

Histogram for parameter: YB
0 1 2 3 4 5 6 7 8 9 10
0.5 1 1.5 2 2.5 3 3.5


Histogram for parameter: LU
0 2 4 6 8 10 12
0.05 0.1 0.15 0.2 0.25 0.3 0.35 0.4 0.45 0.5


Histogram for parameter: ZR
0 5 10 15
50 100 150 200 250 300 350 400


Histogram for parameter: HF
0
2
4
6
8
10
12
14
16
Histogram for parameter: TA
0
0.5
1
1.5
2
2.5
3
18
20
Histogram for parameter: MO
5
15
20
25
30
35
Histogram for parameter: W
7


Histogram for parameter: TH
Histogram for parameter: U
Histogram for parameter: AG
Histogram for parameter: AU


Histogram for parameter: IR
20
18
16
14
12
10
8
6
4
2
0
0
100
200
300
400
500
600
700
800
900
1000

APPENDIX D: SPEARMAN CORRELATION COEFFICIENT FOR THE SCALED DATA

Metro 2001 Spearman Correlation

	'P10'	'Silt'	'Clay'	'SG'	'OM'	'PH'	'EP'	'eK'	'dAs'	'dBr'	'dCe'	'dCo'	'dCr'	'dCs'	'dFe'	'dLa'	'dMn'	'dMo'
'P10'	**1.0000**	0.3336	0.1491	-0.3930	0.2808	-0.4307	0.0970	0.2526	0.1451	0.2229	0.0104	-0.1825	-0.0253	0.2048	-0.3642	0.0404	-0.0389	-0.0289
'Silt'	0.3336	**1.0000**	0.6381	-0.9304	0.5584	-0.1088	0.2191	0.6721	0.4902	0.4443	0.4833	0.1629	0.0282	0.6798	-0.0560	0.5772	0.3386	0.0926
'Clay'	0.1491	0.6381	**1.0000**	-0.6691	0.4112	0.0545	0.1330	0.5059	0.5081	0.3846	0.5750	0.3161	-0.0771	0.5630	0.0313	0.6976	0.3713	0.2866
'SG'	-0.3930	-0.9304	-0.6691	**1.0000**	-0.6191	0.1218	-0.1673	-0.6392	-0.5699	-0.5109	-0.4441	-0.2257	0.0555	-0.6477	0.0130	-0.5461	-0.4041	-0.0913
'OM'	0.2808	0.5584	0.4112	-0.6191	**1.0000**	-0.1737	0.1213	0.5071	0.5096	0.7128	0.2828	0.2867	-0.0313	0.6271	0.1309	0.3685	0.3411	0.1623
'PH'	-0.4307	-0.1088	0.0545	0.1218	-0.1737	**1.0000**	-0.4251	-0.3263	0.1558	0.0941	0.1105	0.2733	0.0423	-0.1975	0.3933	0.1444	0.1820	0.2873
'EP'	0.0970	0.2191	0.1330	-0.1673	0.1213	-0.4251	**1.0000**	0.4023	-0.1412	-0.1056	0.0638	-0.0117	0.0742	0.2287	-0.1889	0.0593	-0.0215	-0.1306
'eK'	0.2526	0.6721	0.5059	-0.6392	0.5071	-0.3263	0.4023	**1.0000**	0.4209	0.2541	0.3254	0.0549	-0.0505	0.5844	-0.2472	0.4394	0.2111	-0.0721
'dAs'	0.1451	0.4902	0.5081	-0.5699	0.5096	0.1558	-0.1412	0.4209	**1.0000**	0.6842	0.5796	0.4103	-0.1934	0.4772	0.2102	0.6373	0.3110	0.5055
'dBr'	0.2229	0.4443	0.3846	-0.5109	0.7128	0.0941	-0.1056	0.2541	0.6842	**1.0000**	0.4602	0.3606	-0.1464	0.5268	0.2413	0.5332	0.3396	0.4426
'dCe'	0.0104	0.4833	0.5750	-0.4441	0.2828	0.1105	0.0638	0.3254	0.5796	0.4602	**1.0000**	0.5623	0.0099	0.5233	0.2850	0.8941	0.2129	0.5651
'dCo'	-0.1825	0.1629	0.3161	-0.2257	0.2867	0.2733	-0.0117	0.0549	0.4103	0.3606	0.5623	**1.0000**	0.2876	0.3528	0.7586	0.4808	0.4537	0.3712
'dCr'	-0.0253	0.0282	-0.0771	0.0555	-0.0313	0.0423	0.0742	-0.0505	-0.1934	-0.1464	0.0099	0.2876	**1.0000**	-0.0217	0.4084	-0.1221	0.2074	-0.1409
'dCs'	0.2048	0.6798	0.5630	-0.6477	0.6271	-0.1975	0.2287	0.5844	0.4772	0.5268	0.5233	0.3528	-0.0217	**1.0000**	0.0550	0.5373	0.3643	0.2133
'dFe'	-0.3642	-0.0560	0.0313	0.0130	0.1309	0.3933	-0.1889	-0.2472	0.2102	0.2413	0.2850	0.7586	0.4084	0.0550	**1.0000**	0.1456	0.3687	0.2557
'dLa'	0.0404	0.5772	0.6976	-0.5461	0.3685	0.1444	0.0593	0.4394	0.6373	0.5332	0.8941	0.4808	-0.1221	0.5373	0.1456	**1.0000**	0.2713	0.5344
'dMn'	-0.0389	0.3386	0.3713	-0.4041	0.3411	0.1820	-0.0215	0.2111	0.3110	0.3396	0.2129	0.4537	0.2074	0.3643	0.3687	0.2713	**1.0000**	0.0707
'dMo'	-0.0289	0.0926	0.2866	-0.0913	0.1623	0.2873	-0.1306	-0.0721	0.5055	0.4426	0.5651	0.3712	-0.1409	0.2133	0.2557	0.5344	0.0707	**1.0000**
'dNi'	-0.3457	-0.0506	0.0442	0.1142	-0.0470	0.2909	-0.1177	-0.1122	0.0687	0.0458	0.2402	0.5301	0.4530	0.0965	0.5377	0.1239	0.2894	0.2689
'dPb'	-0.2600	-0.3179	-0.2176	0.3440	-0.0125	0.1725	-0.0001	-0.2569	0.0463	0.2465	0.1068	0.1641	0.0274	-0.1004	0.2961	0.0092	-0.0772	0.3073
'dRb'	0.1910	0.5696	0.5003	-0.5122	0.4587	-0.0616	0.2084	0.4738	0.3738	0.3785	0.6109	0.4561	0.1109	0.6378	0.1330	0.5685	0.1214	0.1635
'dSb'	-0.2705	-0.0003	0.1336	0.0308	0.1168	0.3455	-0.1534	-0.1417	0.3367	0.2576	0.4502	0.5452	0.1356	0.0480	0.6233	0.3356	0.0361	0.5528
'dSc'	-0.1168	0.4803	0.5041	-0.4177	0.3678	0.1799	0.0660	0.3204	0.4159	0.3794	0.7280	0.7591	0.2634	0.5786	0.5689	0.6589	0.2905	0.3040
'dSe'	0.1367	0.3363	0.2786	-0.3355	0.3086	-0.2262	0.2435	0.4483	0.2735	0.3078	0.1675	-0.1309	-0.2015	0.3540	-0.2194	0.2394	0.0363	0.0150
'dSm'	0.0301	0.5420	0.6303	-0.5180	0.2918	0.1361	0.0372	0.3540	0.6130	0.4824	0.8279	0.4460	-0.0698	0.4670	0.1444	0.8450	0.1760	0.4869
'dSr'	-0.0275	-0.1629	-0.2985	0.2102	-0.0941	-0.2585	0.2671	0.0367	-0.4078	-0.2773	-0.3367	-0.1899	0.3434	-0.0647	-0.0464	-0.4726	-0.1012	-0.4637
'dTh'	-0.0734	0.4304	0.4577	-0.3687	0.2747	0.0772	0.0949	0.2394	0.3584	0.3487	0.6760	0.4958	0.1706	0.4869	0.3258	0.6085	0.3304	0.4080
'dU'	0.2322	0.5999	0.5482	-0.5658	0.3331	0.0058	0.2003	0.4634	0.4207	0.3291	0.5561	0.3147	-0.0765	0.5591	0.0339	0.6509	0.1940	0.3944
'dV'	-0.0644	0.3998	0.3593	-0.3259	0.2576	0.1458	0.0962	0.3447	0.0792	0.1278	0.2525	0.4791	0.4249	0.4142	0.3725	0.3193	0.5210	-0.0867
'dW'	-0.2894	-0.0854	0.1817	0.1137	-0.0639	0.4190	-0.2854	-0.2133	0.3041	0.2436	0.5003	0.3631	-0.1635	-0.0180	0.2580	0.4657	-0.0249	0.7177
'dY'	-0.0671	0.0719	-0.1632	-0.0430	-0.0118	-0.0819	0.2055	0.1256	-0.3890	-0.3010	-0.3371	-0.0646	0.5063	-0.0375	0.1312	-0.3421	0.1951	-0.6717
'dZn'	-0.3398	-0.3040	-0.1238	0.3482	-0.0662	0.2717	-0.0223	-0.3441	-0.0396	0.1112	0.1735	0.3988	0.1551	-0.0373	0.4637	0.0328	0.0178	0.3334
'iCd'	0.0951	0.4937	0.3793	-0.5701	0.5599	0.1506	-0.1509	0.3481	0.4877	0.5654	0.2736	0.3058	0.0038	0.4199	0.2458	0.3557	0.4803	0.1247
'iCu'	-0.5744	-0.3111	-0.1935	0.3822	-0.0998	0.3863	-0.0621	-0.3659	-0.0479	0.0690	0.1577	0.4032	0.2306	-0.1057	0.5945	0.0537	-0.0027	0.2961

Metro 2001 Spearman Correlation

	'dNi'	'dPb'	'dRb'	'dSb'	'dSc'	'dSe'	'dSm'	'dSr'	'dTh'	'dU'	'dV'	'dW'	'dY'	'dZn'	'iCd'	'iCu'
'P10'	-0.3457	-0.2600	0.1910	-0.2705	-0.1168	0.1367	0.0301	-0.0275	-0.0734	0.2322	-0.0644	-0.2894	-0.0671	-0.3398	0.0951	-0.5744
'Silt'	-0.0506	-0.3179	0.5696	-0.0003	0.4803	0.3363	0.5420	-0.1629	0.4304	0.5999	0.3998	-0.0854	0.0719	-0.3040	0.4937	-0.3111
'Clay'	0.0442	-0.2176	0.5003	0.1336	0.5041	0.2786	0.6303	-0.2985	0.4577	0.5482	0.3593	0.1817	-0.1632	-0.1238	0.3793	-0.1935
'SG'	0.1142	0.3440	-0.5122	0.0308	-0.4177	-0.3355	-0.5180	0.2102	-0.3687	-0.5658	-0.3259	0.1137	-0.0430	0.3482	-0.5701	0.3822
'OM'	-0.0470	-0.0125	0.4587	0.1168	0.3678	0.3086	0.2918	-0.0941	0.2747	0.3331	0.2576	-0.0639	-0.0118	-0.0662	0.5599	-0.0998
'PH'	0.2909	0.1725	-0.0616	0.3455	0.1799	-0.2262	0.1361	-0.2585	0.0772	0.0058	0.1458	0.4190	-0.0819	0.2717	0.1506	0.3863
'EP'	-0.1177	-0.0001	0.2084	-0.1534	0.0660	0.2435	0.0372	0.2671	0.0949	0.2003	0.0962	-0.2854	0.2055	-0.0223	-0.1509	-0.0621
'eK'	-0.1122	-0.2569	0.4738	-0.1417	0.3204	0.4483	0.3540	0.0367	0.2394	0.4634	0.3447	-0.2133	0.1256	-0.3441	0.3481	-0.3659
'dAs'	0.0687	0.0463	0.3738	0.3367	0.4159	0.2735	0.6130	-0.4078	0.3584	0.4207	0.0792	0.3041	-0.3890	-0.0396	0.4877	-0.0479
'dBr'	0.0458	0.2465	0.3785	0.2576	0.3794	0.3078	0.4824	-0.2773	0.3487	0.3291	0.1278	0.2436	-0.3010	0.1112	0.5654	0.0690
'dCe'	0.2402	0.1068	0.6109	0.4502	0.7280	0.1675	0.8279	-0.3367	0.6760	0.5561	0.2525	0.5003	-0.3371	0.1735	0.2736	0.1577
'dCo'	0.5301	0.1641	0.4561	0.5452	0.7591	-0.1309	0.4460	-0.1899	0.4958	0.3147	0.4791	0.3631	-0.0646	0.3988	0.3058	0.4032
'dCr'	0.4530	0.0274	0.1109	0.1356	0.2634	-0.2015	-0.0698	0.3434	0.1706	-0.0765	0.4249	-0.1635	0.5063	0.1551	0.0038	0.2306
'dCs'	0.0965	-0.1004	0.6378	0.0480	0.5786	0.3540	0.4670	-0.0647	0.4869	0.5591	0.4142	-0.0180	-0.0375	-0.0373	0.4199	-0.1057
'dFe'	0.5377	0.2961	0.1330	0.6233	0.5689	-0.2194	0.1444	-0.0464	0.3258	0.0339	0.3725	0.2580	0.1312	0.4637	0.2458	0.5945
'dLa'	0.1239	0.0092	0.5685	0.3356	0.6589	0.2394	0.8450	-0.4726	0.6085	0.6509	0.3193	0.4657	-0.3421	0.0328	0.3557	0.0537
'dMn'	0.2894	-0.0772	0.1214	0.0361	0.2905	0.0363	0.1760	-0.1012	0.3304	0.1940	0.5210	-0.0249	0.1951	0.0178	0.4803	-0.0027
'dMo'	0.2689	0.3073	0.1635	0.5528	0.3040	0.0150	0.4869	-0.4637	0.4080	0.3944	-0.0867	0.7177	-0.6717	0.3334	0.1247	0.2961
'dNi'	**1.0000**	0.3558	0.0792	0.4613	0.3942	-0.1372	0.1207	0.0578	0.3918	0.0295	0.3571	0.3832	0.0131	0.4690	0.0609	0.5960
'dPb'	0.3558	**1.0000**	-0.0815	0.4295	0.0438	-0.0313	-0.0233	0.0022	0.0486	-0.1957	-0.1274	0.3328	-0.2413	0.5915	-0.1761	0.5922
'dRb'	0.0792	-0.0815	**1.0000**	0.2195	0.7255	0.2232	0.4979	0.0355	0.4576	0.5995	0.4208	0.0686	-0.0562	0.0477	0.3065	-0.0272
'dSb'	0.4613	0.4295	0.2195	**1.0000**	0.4721	-0.0638	0.3564	-0.2250	0.3612	0.1963	0.0668	0.5391	-0.3022	0.5327	0.1457	0.5895
'dSc'	0.3942	0.0438	0.7255	0.4721	**1.0000**	0.0533	0.6391	-0.1544	0.6255	0.5218	0.6164	0.2422	0.0566	0.2232	0.3691	0.2752
'dSe'	-0.1372	-0.0313	0.2232	-0.0638	0.0533	**1.0000**	0.2021	0.0172	0.0733	0.2110	0.0448	-0.0961	-0.0339	-0.0842	0.1104	-0.0964
'dSm'	0.1207	-0.0233	0.4979	0.3564	0.6391	0.2021	**1.0000**	-0.4286	0.5584	0.5237	0.2490	0.4043	-0.2971	0.0234	0.3245	0.0537
'dSr'	0.0578	0.0022	0.0355	-0.2250	-0.1544	0.0172	-0.4286	**1.0000**	-0.1910	-0.3043	0.0257	-0.5357	0.4818	-0.0887	-0.1349	-0.0045
'dTh'	0.3918	0.0486	0.4576	0.3612	0.6255	0.0733	0.5584	-0.1910	**1.0000**	0.4523	0.4081	0.3235	-0.0821	0.1870	0.2813	0.2246
'dU'	0.0295	-0.1957	0.5995	0.1963	0.5218	0.2110	0.5237	-0.3043	0.4523	**1.0000**	0.3412	0.1646	-0.1993	-0.0847	0.3497	-0.1202
'dV'	0.3571	-0.1274	0.4208	0.0668	0.6164	0.0448	0.2490	0.0257	0.4081	0.3412	**1.0000**	-0.1268	0.4833	0.1024	0.3368	0.1373
'dW'	0.3832	0.3328	0.0686	0.5391	0.2422	-0.0961	0.4043	-0.5357	0.3235	0.1646	-0.1268	**1.0000**	-0.6883	0.4193	-0.0191	0.4081
'dY'	0.0131	-0.2413	-0.0562	-0.3022	0.0566	-0.0339	-0.2971	0.4818	-0.0821	-0.1993	0.4833	-0.6883	**1.0000**	-0.1987	0.1002	-0.0493
'dZn'	0.4690	0.5915	0.0477	0.5327	0.2232	-0.0842	0.0234	-0.0887	0.1870	-0.0847	0.1024	0.4193	-0.1987	**1.0000**	-0.1159	0.7406
'iCd'	0.0609	-0.1761	0.3065	0.1457	0.3691	0.1104	0.3245	-0.1349	0.2813	0.3497	0.3368	-0.0191	0.1002	-0.1159	**1.0000**	-0.0886
'iCu'	0.5960	0.5922	-0.0272	0.5895	0.2752	-0.0964	0.0537	-0.0045	0.2246	-0.1202	0.1373	0.4081	-0.0493	0.7406	-0.0886	**1.0000**

State 2003 Spearman Correlation

	'P10'	'S&G'	'SILT'	'CLAY'	'OM'	'PH'	'EP'	'EK'	'Ssalt'	'TOT_SOL'	'CU'	'NI_ICP'	'PB'	'TI'	'AL'	'FE'	'MG'
'P10'	1.0000	-0.2431	0.2226	0.1798	0.1724	0.0657	0.0551	0.0895	0.2107	-0.2995	-0.0906	-0.1057	0.1251	-0.0333	0.0286	-0.2365	-0.1147
'S&G'	-0.2431	1.0000	-0.9924	-0.7892	-0.3676	-0.5101	0.0463	-0.5980	-0.3610	0.4470	-0.5339	-0.6552	-0.3106	-0.2231	-0.1782	-0.3973	-0.1525
'SILT'	0.2226	-0.9924	1.0000	0.7286	0.3442	0.4666	-0.0251	0.5639	0.3193	-0.4360	0.5132	0.6334	0.3067	0.2695	0.2032	0.3903	0.1775
'CLAY'	0.1798	-0.7892	0.7286	1.0000	0.3936	0.5940	-0.0832	0.6239	0.4457	-0.4160	0.5921	0.6836	0.2223	0.0388	0.1245	0.3612	0.1493
'OM'	0.1724	-0.3676	0.3442	0.3936	1.0000	-0.0034	-0.0933	0.6005	0.2870	-0.8493	0.3456	0.3541	0.5846	-0.1941	-0.3418	0.0441	-0.0601
'PH'	0.0657	-0.5101	0.4666	0.5940	-0.0034	1.0000	-0.1875	0.4488	0.5028	-0.0067	0.2883	0.3658	-0.2279	0.0558	0.1121	0.1141	0.1995
'EP'	0.0551	0.0463	-0.0251	-0.0832	-0.0933	-0.1875	1.0000	0.0967	-0.1796	0.1576	-0.0187	-0.0869	0.1382	0.2205	0.0414	-0.0008	0.1425
'EK'	0.0895	-0.5980	0.5639	0.6239	0.6005	0.4488	0.0967	1.0000	0.3237	-0.4370	0.4612	0.6167	0.4072	-0.0963	-0.2725	0.2704	0.0462
'Ssalt'	0.2107	-0.3610	0.3193	0.4457	0.2870	0.5028	-0.1796	0.3237	1.0000	-0.3175	0.2411	0.2555	-0.0225	-0.0980	-0.1009	0.0828	-0.1133
'TOT_SOL'	-0.2995	0.4470	-0.4360	-0.4160	-0.8493	-0.0067	0.1576	-0.4370	-0.3175	1.0000	-0.1988	-0.2363	-0.3987	0.0540	0.2301	0.0656	0.0318
'CU'	-0.0906	-0.5339	0.5132	0.5921	0.3456	0.2883	-0.0187	0.4612	0.2411	-0.1988	1.0000	0.8517	0.4487	-0.0076	0.1072	0.7702	0.1754
'NI_ICP'	-0.1057	-0.6552	0.6334	0.6836	0.3541	0.3658	-0.0869	0.6167	0.2555	-0.2363	0.8517	1.0000	0.4215	-0.0234	0.0886	0.7471	0.1318
'PB'	0.1251	-0.3106	0.3067	0.2223	0.5846	-0.2279	0.1382	0.4072	-0.0225	-0.3987	0.4487	0.4215	1.0000	-0.2000	-0.2373	0.2912	-0.1711
'TI'	-0.0333	-0.2231	0.2695	0.0388	-0.1941	0.0558	0.2205	-0.0963	-0.0980	0.0540	-0.0076	-0.0234	-0.2000	1.0000	0.5415	0.1264	0.4369
'AL'	0.0286	-0.1782	0.2032	0.1245	-0.3418	0.1121	0.0414	-0.2725	-0.1009	0.2301	0.1072	0.0886	-0.2373	0.5415	1.0000	0.2352	0.5760
'FE'	-0.2365	-0.3973	0.3903	0.3612	0.0441	0.1141	-0.0008	0.2704	0.0828	0.0656	0.7702	0.7471	0.2912	0.1264	0.2352	1.0000	0.2552
'MG'	-0.1147	-0.1525	0.1775	0.1493	-0.0601	0.1995	0.1425	0.0462	-0.1133	0.0318	0.1754	0.1318	-0.1711	0.4369	0.5760	0.2552	1.0000
'CA'	-0.0124	-0.1886	0.2004	0.1648	0.0106	0.3647	0.0430	0.1339	0.1626	0.0251	0.3032	0.2692	-0.0373	-0.0119	0.3266	0.1962	0.4922
'NA'	-0.0793	0.1705	-0.1205	-0.3256	0.1087	-0.1279	-0.1982	-0.0954	-0.1529	-0.2819	-0.4411	-0.3415	-0.2132	-0.0230	-0.1577	-0.4416	0.0608
'K'	0.0551	-0.3844	0.3791	0.3827	0.2928	0.2163	-0.0479	0.4324	0.2365	-0.2763	0.1262	0.3875	0.2140	0.0771	0.0845	0.1514	0.2457
'MN'	-0.0812	-0.1113	0.1397	-0.0049	-0.3206	0.2080	-0.1852	-0.2257	0.0225	0.2985	-0.0060	0.0358	-0.1828	0.4165	0.4915	0.0516	0.2852
'BR'	0.2741	-0.7413	0.6941	0.8003	0.5415	0.5898	-0.1559	0.5984	0.4850	-0.5344	0.5577	0.5886	0.2744	-0.0471	-0.1056	0.2359	0.0310
'SC'	-0.0178	-0.6881	0.6917	0.5357	0.3168	0.1846	0.0325	0.5136	0.1662	-0.2751	0.8159	0.8413	0.4616	0.1344	0.1485	0.8642	0.2208
'V'	-0.0763	-0.2798	0.3102	0.1198	-0.2442	0.1950	0.0731	-0.0988	-0.0849	0.1142	0.0712	0.0812	-0.1686	0.8470	0.7340	0.2612	0.5586
'CR'	0.0140	-0.6396	0.6505	0.4257	0.2464	0.1240	0.0515	0.4752	0.0912	-0.2476	0.6604	0.7116	0.3639	0.1718	0.1351	0.7962	0.1630
'CO'	-0.0833	-0.5975	0.6021	0.4929	0.2375	0.1563	0.0762	0.4746	0.0658	-0.2195	0.7091	0.8292	0.4199	0.0984	0.0883	0.8110	0.2254
'NI_NAA'	-0.1960	-0.4234	0.4325	0.3626	0.0506	0.2703	-0.1369	0.2782	0.1750	-0.0310	0.6480	0.6471	0.2700	0.0825	0.2645	0.7259	0.2011
'ZN'	0.0270	-0.6626	0.6520	0.5961	0.5032	0.2644	0.1504	0.6686	0.2555	-0.3694	0.7336	0.8205	0.5431	-0.0174	-0.0334	0.6873	0.2087
'AS'	0.0123	-0.6107	0.5738	0.6202	0.1175	0.3935	0.0402	0.4891	0.2463	0.0292	0.7567	0.7520	0.3574	0.0461	0.1569	0.7815	0.1684
'SB'	0.0415	-0.6190	0.6000	0.5864	0.2287	0.2796	0.1091	0.5105	0.3567	-0.1471	0.4496	0.6000	0.3687	-0.0185	-0.0148	0.4172	-0.0244
'SE'	0.0683	-0.1095	0.0524	0.2714	0.3054	0.2656	-0.2644	0.3602	0.2232	-0.1360	0.1165	0.1738	-0.0062	-0.3996	-0.1499	0.0002	-0.0308
'RB'	0.0599	-0.6355	0.6372	0.5053	0.2754	0.1514	0.1791	0.4019	0.1745	-0.3660	0.3406	0.5653	0.4175	0.1844	0.0969	0.4284	0.1801
'CS'	0.0606	-0.8292	0.8150	0.7143	0.4752	0.2969	0.0826	0.6795	0.2701	-0.4585	0.5924	0.7926	0.4957	0.1528	0.0499	0.5809	0.1704
'SR'	-0.0595	0.1458	-0.1220	-0.2533	0.1550	0.0482	-0.2794	-0.0663	0.0290	-0.2745	-0.4317	-0.3281	-0.2995	-0.1165	-0.1930	-0.5093	0.0099

State 2003 Spearman Correlation

	'P10'	'S&G'	'SILT'	'CLAY'	'OM'	'PH'	'EP'	'EK'	'Ssalt'	'TOT_SOL'	'CU'	'NI_ICP'	'PB'	'TI'	'AL'	'FE'	'MG'
'BA'	0.0397	-0.5335	0.5583	0.2424	0.2080	0.1418	-0.0275	0.3931	-0.0570	-0.2768	0.0327	0.3105	0.2662	0.0895	-0.0984	0.1009	0.0342
'LA'	0.0786	-0.8558	0.8389	0.7296	0.2678	0.4657	0.0754	0.6224	0.2626	-0.2209	0.6948	0.7864	0.4553	0.1576	0.1378	0.6466	0.1780
'CE'	0.0535	-0.8345	0.8177	0.7160	0.2568	0.5061	0.0546	0.6435	0.2936	-0.2128	0.6852	0.8109	0.3914	0.1473	0.1262	0.6326	0.1947
'ND'	0.0874	-0.8218	0.8035	0.7326	0.2263	0.5010	0.0561	0.5870	0.3073	-0.2014	0.7585	0.8195	0.3812	0.2061	0.1931	0.6979	0.2190
'SM'	0.0382	-0.8128	0.7975	0.7161	0.1799	0.4606	0.0519	0.5536	0.2597	-0.1590	0.7559	0.8224	0.4113	0.1746	0.2223	0.7490	0.2036
'EU'	0.0329	-0.8206	0.8130	0.6867	0.2007	0.4690	0.0131	0.5596	0.2659	-0.1842	0.7838	0.8550	0.3624	0.2083	0.2506	0.7793	0.2541
'TB'	-0.0160	-0.7230	0.7174	0.6332	0.0470	0.4352	0.1175	0.4720	0.2460	-0.0328	0.6865	0.7716	0.3425	0.2280	0.2202	0.6986	0.2061
'YB'	0.0073	-0.6739	0.6661	0.5780	-0.0038	0.2938	0.1308	0.4218	0.1537	0.0261	0.6885	0.7340	0.3704	0.2252	0.2873	0.7667	0.1971
'LU'	0.0240	-0.6414	0.6343	0.5280	-0.0482	0.2655	0.1296	0.3784	0.1258	0.0479	0.6561	0.6816	0.3509	0.2236	0.3015	0.7787	0.1698
'ZR'	0.0613	-0.7446	0.7689	0.4598	-0.0548	0.3578	0.0539	0.3541	0.1268	-0.0623	0.3385	0.4768	0.1597	0.3183	0.3221	0.4118	0.1695
'HF'	0.0493	-0.6648	0.6926	0.3978	-0.1609	0.2628	0.1925	0.3241	0.0110	0.0714	0.3371	0.4491	0.1342	0.3350	0.3674	0.4427	0.2184
'TA'	0.1326	-0.7240	0.7284	0.5420	0.1311	0.2340	0.1125	0.4889	0.0976	-0.1461	0.5961	0.6987	0.3377	0.2442	0.1720	0.7001	0.1798
'MO'	0.1482	-0.3772	0.3575	0.2792	0.4316	0.2242	0.0249	0.4236	0.3781	-0.4554	0.0468	0.1035	0.1405	-0.0163	-0.4750	-0.1259	-0.2621
'W'	-0.1781	-0.4004	0.4263	0.1934	-0.0791	0.0825	-0.1326	0.1592	0.0602	0.0620	0.3425	0.4668	0.1805	0.1360	0.1855	0.5726	0.2262
'TH'	0.0954	-0.7826	0.7710	0.6489	0.2809	0.3927	0.0626	0.5994	0.2998	-0.2568	0.5283	0.7098	0.3854	0.0546	0.0450	0.4969	0.1434
'U'	0.0944	-0.9028	0.8885	0.7367	0.3209	0.4592	0.0857	0.6165	0.3123	-0.3284	0.5618	0.6648	0.3765	0.1794	0.0995	0.5286	0.1738
'AG'	0.0039	-0.1766	0.2000	0.0525	0.0893	0.0691	0.0005	0.0054	0.0737	-0.0833	0.0765	0.1512	0.2387	-0.0132	0.2416	0.1028	0.0630

State 2003 Spearman Correlation

	'CA'	'NA'	'K'	'MN'	'BR'	'SC'	'V'	'CR'	'CO'	'NI_NAA'	'ZN'	'AS'	'SB'	'SE'	'RB'	'CS'	'SR'
'P10'	-0.0124	-0.0793	0.0551	-0.0812	0.2741	-0.0178	-0.0763	0.0140	-0.0833	-0.1960	0.0270	0.0123	0.0415	0.0683	0.0599	0.0606	-0.0595
'S&G'	-0.1886	0.1705	-0.3844	-0.1113	-0.7413	-0.6881	-0.2798	-0.6396	-0.5975	-0.4234	-0.6626	-0.6107	-0.6190	-0.1095	-0.6355	-0.8292	0.1458
'SILT'	0.2004	-0.1205	0.3791	0.1397	0.6941	0.6917	0.3102	0.6505	0.6021	0.4325	0.6520	0.5738	0.6000	0.0524	0.6372	0.8150	-0.1220
'CLAY'	0.1648	-0.3256	0.3827	-0.0049	0.8003	0.5357	0.1198	0.4257	0.4929	0.3626	0.5961	0.6202	0.5864	0.2714	0.5053	0.7143	-0.2533
'OM'	0.0106	0.1087	0.2928	-0.3206	0.5415	0.3168	-0.2442	0.2464	0.2375	0.0506	0.5032	0.1175	0.2287	0.3054	0.2754	0.4752	0.1550
'PH'	0.3647	-0.1279	0.2163	0.2080	0.5898	0.1846	0.1950	0.1240	0.1563	0.2703	0.2644	0.3935	0.2796	0.2656	0.1514	0.2969	0.0482
'EP'	0.0430	-0.1982	-0.0479	-0.1852	-0.1559	0.0325	0.0731	0.0515	0.0762	-0.1369	0.1504	0.0402	0.1091	-0.2644	0.1791	0.0826	-0.2794
'EK'	0.1339	-0.0954	0.4324	-0.2257	0.5984	0.5136	-0.0988	0.4752	0.4746	0.2782	0.6686	0.4891	0.5105	0.3602	0.4019	0.6795	-0.0663
'Ssalt'	0.1626	-0.1529	0.2365	0.0225	0.4850	0.1662	-0.0849	0.0912	0.0658	0.1750	0.2555	0.2463	0.3567	0.2232	0.1745	0.2701	0.0290
'TOT_SOL'	0.0251	-0.2819	-0.2763	0.2985	-0.5344	-0.2751	0.1142	-0.2476	-0.2195	-0.0310	-0.3694	0.0292	-0.1471	-0.1360	-0.3660	-0.4585	-0.2745
'CU'	0.3032	-0.4411	0.1262	-0.0060	0.5577	0.8159	0.0712	0.6604	0.7091	0.6480	0.7336	0.7567	0.4496	0.1165	0.3406	0.5924	-0.4317
'NI_ICP'	0.2692	-0.3415	0.3875	0.0358	0.5886	0.8413	0.0812	0.7116	0.8292	0.6471	0.8205	0.7520	0.6000	0.1738	0.5653	0.7926	-0.3281
'PB'	-0.0373	-0.2132	0.2140	-0.1828	0.2744	0.4616	-0.1686	0.3639	0.4199	0.2700	0.5431	0.3574	0.3687	-0.0062	0.4175	0.4957	-0.2995
'TI'	-0.0119	-0.0230	0.0771	0.4165	-0.0471	0.1344	0.8470	0.1718	0.0984	0.0825	-0.0174	0.0461	-0.0185	-0.3996	0.1844	0.1528	-0.1165
'AL'	0.3266	-0.1577	0.0845	0.4915	-0.1056	0.1485	0.7340	0.1351	0.0883	0.2645	-0.0334	0.1569	-0.0148	-0.1499	0.0969	0.0499	-0.1930
'FE'	0.1962	-0.4416	0.1514	0.0516	0.2359	0.8642	0.2612	0.7962	0.8110	0.7259	0.6873	0.7815	0.4172	0.0002	0.4284	0.5809	-0.5093
'MG'	0.4922	0.0608	0.2457	0.2852	0.0310	0.2208	0.5586	0.1630	0.2254	0.2011	0.2087	0.1684	-0.0244	-0.0308	0.1801	0.1704	0.0099
'CA'	1.0000	-0.0659	0.2995	0.1784	0.1485	0.2129	0.1786	0.1081	0.2159	0.3857	0.2609	0.2627	0.2771	0.0239	0.0743	0.1081	-0.0191
'NA'	-0.0659	1.0000	0.1774	-0.0617	-0.2890	-0.2517	-0.0252	-0.0967	-0.1678	-0.1985	-0.2085	-0.6548	-0.4100	-0.1422	-0.0563	-0.2052	0.8539
'K'	0.2995	0.1774	1.0000	0.1718	0.1788	0.2466	0.1473	0.2341	0.3617	0.2305	0.3813	0.1948	0.4628	0.0108	0.5878	0.5536	0.1080
'MN'	0.1784	-0.0617	0.1718	1.0000	-0.0443	-0.0239	0.5260	-0.0593	-0.0458	0.1481	-0.1577	0.0641	0.0816	-0.0860	0.0553	0.0109	-0.0111
'BR'	0.1485	-0.2890	0.1788	-0.0443	1.0000	0.4954	-0.0612	0.3623	0.3804	0.2536	0.5533	0.5717	0.4716	0.2414	0.3722	0.6304	-0.1004
'SC'	0.2129	-0.2517	0.2466	-0.0239	0.4954	1.0000	0.2300	0.9341	0.9125	0.7144	0.8661	0.7654	0.5091	-0.0165	0.5937	0.7923	-0.3342
'V'	0.1786	-0.0252	0.1473	0.5260	-0.0612	0.2300	1.0000	0.2551	0.1883	0.2880	0.0457	0.1527	-0.0082	-0.2777	0.2505	0.1923	-0.0713
'CR'	0.1081	-0.0967	0.2341	-0.0593	0.3623	0.9341	0.2551	1.0000	0.8651	0.6453	0.7670	0.6679	0.4483	-0.0724	0.5602	0.7306	-0.2168
'CO'	0.2159	-0.1678	0.3617	-0.0458	0.3804	0.9125	0.1883	0.8651	1.0000	0.7050	0.8604	0.6863	0.5209	-0.0797	0.7146	0.8005	-0.2889
'NI_NAA'	0.3857	-0.1985	0.2305	0.1481	0.2536	0.7144	0.2880	0.6453	0.7050	1.0000	0.5588	0.6257	0.4133	-0.0411	0.3438	0.4646	-0.3245
'ZN'	0.2609	-0.2085	0.3813	-0.1577	0.5533	0.8661	0.0457	0.7670	0.8604	0.5588	1.0000	0.6918	0.5533	0.1042	0.6639	0.8295	-0.2559
'AS'	0.2627	-0.6548	0.1948	0.0641	0.5717	0.7654	0.1527	0.6679	0.6863	0.6257	0.6918	1.0000	0.6931	0.1544	0.4108	0.6536	-0.6313
'SB'	0.2771	-0.4100	0.4628	0.0816	0.4716	0.5091	-0.0082	0.4483	0.5209	0.4133	0.5533	0.6931	1.0000	0.0871	0.5517	0.6502	-0.4229
'SE'	0.0239	-0.1422	0.0108	-0.0860	0.2414	-0.0165	-0.2777	-0.0724	-0.0797	-0.0411	0.1042	0.1544	0.0871	1.0000	-0.1761	0.0474	-0.0325
'RB'	0.0743	-0.0563	0.5878	0.0553	0.3722	0.5937	0.2505	0.5602	0.7146	0.3438	0.6639	0.4108	0.5517	-0.1761	1.0000	0.8673	-0.0844
'CS'	0.1081	-0.2052	0.5536	0.0109	0.6304	0.7923	0.1923	0.7306	0.8005	0.4646	0.8295	0.6536	0.6502	0.0474	0.8673	1.0000	-0.1893
'SR'	-0.0191	0.8539	0.1080	-0.0111	-0.1004	-0.3342	-0.0713	-0.2168	-0.2889	-0.3245	-0.2559	-0.6313	-0.4229	-0.0325	-0.0844	-0.1893	1.0000

State 2003 Spearman Correlation

	'CA'	'NA'	'K'	'MN'	'BR'	'SC'	'V'	'CR'	'CO'	'NI_NAA'	'ZN'	'AS'	'SB'	'SE'	'RB'	'CS'	'SR'
'BA'	0.1027	0.3538	0.4889	-0.0246	0.2023	0.3698	0.1483	0.4213	0.4664	0.1731	0.4529	0.1126	0.3270	-0.1763	0.6167	0.5764	0.3090
'LA'	0.1662	-0.3887	0.3657	0.0702	0.6584	0.8002	0.2324	0.7268	0.7502	0.5326	0.7734	0.8276	0.7086	0.0982	0.6915	0.8655	-0.3776
'CE'	0.2161	-0.3477	0.4159	0.0958	0.6601	0.7868	0.2236	0.7036	0.7697	0.5326	0.7794	0.8014	0.6942	0.0718	0.7072	0.8686	-0.3063
'ND'	0.2323	-0.4219	0.3771	0.1430	0.6669	0.8196	0.2872	0.7307	0.7697	0.6062	0.7648	0.8450	0.6806	0.0403	0.6740	0.8411	-0.3828
'SM'	0.2216	-0.4346	0.3518	0.1307	0.6210	0.8521	0.2884	0.7677	0.7892	0.6368	0.7594	0.8615	0.6902	0.0318	0.6742	0.8365	-0.4329
'EU'	0.2789	-0.3769	0.3358	0.1220	0.6296	0.9083	0.3279	0.8212	0.8346	0.6751	0.8067	0.8439	0.6300	0.0078	0.6538	0.8399	-0.3608
'TB'	0.2658	-0.4476	0.4030	0.1765	0.5112	0.7680	0.2959	0.6764	0.7469	0.6115	0.6836	0.8039	0.6974	-0.0673	0.6800	0.7737	-0.4742
'YB'	0.2482	-0.5996	0.2591	0.0958	0.4296	0.7873	0.2803	0.7043	0.7537	0.6288	0.6760	0.8658	0.6907	-0.0445	0.5863	0.7191	-0.6440
'LU'	0.2270	-0.6046	0.2038	0.0558	0.3860	0.7769	0.2951	0.7099	0.7336	0.6300	0.6417	0.8552	0.6389	-0.0525	0.5497	0.6736	-0.6608
'ZR'	0.2807	-0.1213	0.4254	0.1711	0.2794	0.5561	0.3545	0.5860	0.5496	0.4838	0.4787	0.5225	0.6214	-0.1210	0.5500	0.6084	-0.2402
'HF'	0.2776	-0.2615	0.2576	0.1495	0.2299	0.5572	0.3380	0.5778	0.5532	0.4455	0.4630	0.5810	0.5994	-0.1238	0.4574	0.5364	-0.4031
'TA'	0.1526	-0.3209	0.3009	-0.0214	0.4355	0.8212	0.2297	0.8324	0.7956	0.5394	0.7562	0.7421	0.6094	-0.0029	0.6056	0.7636	-0.4489
'MO'	-0.3603	0.0777	0.0924	-0.0815	0.4762	0.1116	-0.1968	0.0829	0.0667	-0.1503	0.2268	0.0377	0.1472	0.1847	0.3277	0.4052	0.2307
'W'	0.1936	-0.1965	0.2195	0.3167	0.1561	0.5515	0.2988	0.5372	0.5549	0.6292	0.4155	0.5178	0.4248	-0.0580	0.3424	0.4324	-0.3155
'TH'	0.1263	-0.2368	0.4359	0.0789	0.6061	0.6870	0.1151	0.6285	0.6456	0.4108	0.6850	0.6893	0.7013	0.0983	0.6741	0.8144	-0.1946
'U'	0.1288	-0.3166	0.4152	0.1012	0.6766	0.7094	0.2446	0.6587	0.6362	0.4530	0.6810	0.7296	0.7098	0.1385	0.7003	0.8688	-0.2990
'AG'	0.0727	0.0661	0.1738	0.1236	0.0736	0.1969	0.2248	0.2192	0.1526	0.2414	0.1879	0.0687	0.1068	-0.1607	0.1988	0.1442	0.0773

State 2003 Spearman Correlation

	'BA'	'LA'	'CE'	'ND'	'SM'	'EU'	'TB'	'YB'	'LU'	'ZR'	'HF'	'TA'	'MO'	'W'	'TH'	'U'	'AG'
'P10'	0.0397	0.0786	0.0535	0.0874	0.0382	0.0329	-0.0160	0.0073	0.0240	0.0613	0.0493	0.1326	0.1482	-0.1781	0.0954	0.0944	0.0039
'S&G'	-0.5335	-0.8558	-0.8345	-0.8218	-0.8128	-0.8206	-0.7230	-0.6739	-0.6414	-0.7446	-0.6648	-0.7240	-0.3772	-0.4004	-0.7826	-0.9028	-0.1766
'SILT'	0.5583	0.8389	0.8177	0.8035	0.7975	0.8130	0.7174	0.6661	0.6343	0.7689	0.6926	0.7284	0.3575	0.4263	0.7710	0.8885	0.2000
'CLAY'	0.2424	0.7296	0.7160	0.7326	0.7161	0.6867	0.6332	0.5780	0.5280	0.4598	0.3978	0.5420	0.2792	0.1934	0.6489	0.7367	0.0525
'OM'	0.2080	0.2678	0.2568	0.2263	0.1799	0.2007	0.0470	-0.0038	-0.0482	-0.0548	-0.1609	0.1311	0.4316	-0.0791	0.2809	0.3209	0.0893
'PH'	0.1418	0.4657	0.5061	0.5010	0.4606	0.4690	0.4352	0.2938	0.2655	0.3578	0.2628	0.2340	0.2242	0.0825	0.3927	0.4592	0.0691
'EP'	-0.0275	0.0754	0.0546	0.0561	0.0519	0.0131	0.1175	0.1308	0.1296	0.0539	0.1925	0.1125	0.0249	-0.1326	0.0626	0.0857	0.0005
'EK'	0.3931	0.6224	0.6435	0.5870	0.5536	0.5596	0.4720	0.4218	0.3784	0.3541	0.3241	0.4889	0.4236	0.1592	0.5994	0.6165	0.0054
'Ssalt'	-0.0570	0.2626	0.2936	0.3073	0.2597	0.2659	0.2460	0.1537	0.1258	0.1268	0.0110	0.0976	0.3781	0.0602	0.2998	0.3123	0.0737
'TOT_SOL'	-0.2768	-0.2209	-0.2128	-0.2014	-0.1590	-0.1842	-0.0328	0.0261	0.0479	-0.0623	0.0714	-0.1461	-0.4554	0.0620	-0.2568	-0.3284	-0.0833
'CU'	0.0327	0.6948	0.6852	0.7585	0.7559	0.7838	0.6865	0.6885	0.6561	0.3385	0.3371	0.5961	0.0468	0.3425	0.5283	0.5618	0.0765
'NI_ICP'	0.3105	0.7864	0.8109	0.8195	0.8224	0.8550	0.7716	0.7340	0.6816	0.4768	0.4491	0.6987	0.1035	0.4668	0.7098	0.6648	0.1512
'PB'	0.2662	0.4553	0.3914	0.3812	0.4113	0.3624	0.3425	0.3704	0.3509	0.1597	0.1342	0.3377	0.1405	0.1805	0.3854	0.3765	0.2387
'TI'	0.0895	0.1576	0.1473	0.2061	0.1746	0.2083	0.2280	0.2252	0.2236	0.3183	0.3350	0.2442	-0.0163	0.1360	0.0546	0.1794	-0.0132
'AL'	-0.0984	0.1378	0.1262	0.1931	0.2223	0.2506	0.2202	0.2873	0.3015	0.3221	0.3674	0.1720	-0.4750	0.1855	0.0450	0.0995	0.2416
'FE'	0.1009	0.6466	0.6326	0.6979	0.7490	0.7793	0.6986	0.7667	0.7787	0.4118	0.4427	0.7001	-0.1259	0.5726	0.4969	0.5286	0.1028
'MG'	0.0342	0.1780	0.1947	0.2190	0.2036	0.2541	0.2061	0.1971	0.1698	0.1695	0.2184	0.1798	-0.2621	0.2262	0.1434	0.1738	0.0630
'CA'	0.1027	0.1662	0.2161	0.2323	0.2216	0.2789	0.2658	0.2482	0.2270	0.2807	0.2776	0.1526	-0.3603	0.1936	0.1263	0.1288	0.0727
'NA'	0.3538	-0.3887	-0.3477	-0.4219	-0.4346	-0.3769	-0.4476	-0.5996	-0.6046	-0.1213	-0.2615	-0.3209	0.0777	-0.1965	-0.2368	-0.3166	0.0661
'K'	0.4889	0.3657	0.4159	0.3771	0.3518	0.3358	0.4030	0.2591	0.2038	0.4254	0.2576	0.3009	0.0924	0.2195	0.4359	0.4152	0.1738
'MN'	-0.0246	0.0702	0.0958	0.1430	0.1307	0.1220	0.1765	0.0958	0.0558	0.1711	0.1495	-0.0214	-0.0815	0.3167	0.0789	0.1012	0.1236
'BR'	0.2023	0.6584	0.6601	0.6669	0.6210	0.6296	0.5112	0.4296	0.3860	0.2794	0.2299	0.4355	0.4762	0.1561	0.6061	0.6766	0.0736
'SC'	0.3698	0.8002	0.7868	0.8196	0.8521	0.9083	0.7680	0.7873	0.7769	0.5561	0.5572	0.8212	0.1116	0.5515	0.6870	0.7094	0.1969
'V'	0.1483	0.2324	0.2236	0.2872	0.2884	0.3279	0.2959	0.2803	0.2951	0.3545	0.3380	0.2297	-0.1968	0.2988	0.1151	0.2446	0.2248
'CR'	0.4213	0.7268	0.7036	0.7307	0.7677	0.8212	0.6764	0.7043	0.7099	0.5860	0.5778	0.8324	0.0829	0.5372	0.6285	0.6587	0.2192
'CO'	0.4664	0.7502	0.7697	0.7697	0.7892	0.8346	0.7469	0.7537	0.7336	0.5496	0.5532	0.7956	0.0667	0.5549	0.6456	0.6362	0.1526
'NI_NAA'	0.1731	0.5326	0.5326	0.6062	0.6368	0.6751	0.6115	0.6288	0.6300	0.4838	0.4455	0.5394	-0.1503	0.6292	0.4108	0.4530	0.2414
'ZN'	0.4529	0.7734	0.7794	0.7648	0.7594	0.8067	0.6836	0.6760	0.6417	0.4787	0.4630	0.7562	0.2268	0.4155	0.6850	0.6810	0.1879
'AS'	0.1126	0.8276	0.8014	0.8450	0.8615	0.8439	0.8039	0.8658	0.8552	0.5225	0.5810	0.7421	0.0377	0.5178	0.6893	0.7296	0.0687
'SB'	0.3270	0.7086	0.6942	0.6806	0.6902	0.6300	0.6974	0.6907	0.6389	0.6214	0.5994	0.6094	0.1472	0.4248	0.7013	0.7098	0.1068
'SE'	-0.1763	0.0982	0.0718	0.0403	0.0318	0.0078	-0.0673	-0.0445	-0.0525	-0.1210	-0.1238	-0.0029	0.1847	-0.0580	0.0983	0.1385	-0.1607
'RB'	0.6167	0.6915	0.7072	0.6740	0.6742	0.6538	0.6800	0.5863	0.5497	0.5500	0.4574	0.6056	0.3277	0.3424	0.6741	0.7003	0.1988
'CS'	0.5764	0.8655	0.8686	0.8411	0.8365	0.8399	0.7737	0.7191	0.6736	0.6084	0.5364	0.7636	0.4052	0.4324	0.8144	0.8688	0.1442
'SR'	0.3090	-0.3776	-0.3063	-0.3828	-0.4329	-0.3608	-0.4742	-0.6440	-0.6608	-0.2402	-0.4031	-0.4489	0.2307	-0.3155	-0.1946	-0.2990	0.0773

State 2003 Spearman Correlation

	'BA'	'LA'	'CE'	'ND'	'SM'	'EU'	'TB'	'YB'	'LU'	'ZR'	'HF'	'TA'	'MO'	'W'	'TH'	'U'	'AG'
'BA'	1.0000	0.4303	0.4738	0.3772	0.3812	0.4091	0.3756	0.2747	0.2562	0.5409	0.4334	0.4379	0.2884	0.2760	0.4383	0.4412	0.1976
'LA'	0.4303	1.0000	0.9776	0.9694	0.9666	0.9289	0.9009	0.8672	0.8371	0.7291	0.6890	0.8518	0.3064	0.4458	0.9017	0.9235	0.1116
'CE'	0.4738	0.9776	1.0000	0.9745	0.9456	0.9211	0.8924	0.8414	0.8055	0.7112	0.6642	0.8231	0.3359	0.4375	0.9090	0.8959	0.0826
'ND'	0.3772	0.9694	0.9745	1.0000	0.9728	0.9508	0.9145	0.8736	0.8451	0.6981	0.6510	0.8293	0.2818	0.4492	0.8667	0.8856	0.1154
'SM'	0.3812	0.9666	0.9456	0.9728	1.0000	0.9736	0.9471	0.9125	0.8905	0.7290	0.6982	0.8454	0.2021	0.5135	0.8302	0.8901	0.1622
'EU'	0.4091	0.9289	0.9211	0.9508	0.9736	1.0000	0.9123	0.8893	0.8670	0.7089	0.6830	0.8439	0.1613	0.5394	0.7883	0.8477	0.2194
'TB'	0.3756	0.9009	0.8924	0.9145	0.9471	0.9123	1.0000	0.9196	0.8787	0.7485	0.7192	0.8028	0.1642	0.5142	0.7924	0.8140	0.0990
'YB'	0.2747	0.8672	0.8414	0.8736	0.9125	0.8893	0.9196	1.0000	0.9850	0.7414	0.7988	0.8571	0.0120	0.5609	0.7121	0.7629	0.0710
'LU'	0.2562	0.8371	0.8055	0.8451	0.8905	0.8670	0.8787	0.9850	1.0000	0.7259	0.7872	0.8540	-0.0223	0.5599	0.6602	0.7343	0.0716
'ZR'	0.5409	0.7291	0.7112	0.6981	0.7290	0.7089	0.7485	0.7414	0.7259	1.0000	0.9248	0.7542	0.0488	0.4728	0.6524	0.7166	0.1058
'HF'	0.4334	0.6890	0.6642	0.6510	0.6982	0.6830	0.7192	0.7988	0.7872	0.9248	1.0000	0.7723	-0.0559	0.5132	0.5904	0.6535	0.0692
'TA'	0.4379	0.8518	0.8231	0.8293	0.8454	0.8439	0.8028	0.8571	0.8540	0.7542	0.7723	1.0000	0.1359	0.5121	0.7189	0.7678	0.0371
'MO'	0.2884	0.3064	0.3359	0.2818	0.2021	0.1613	0.1642	0.0120	-0.0223	0.0488	-0.0559	0.1359	1.0000	-0.1590	0.3938	0.4130	-0.2538
'W'	0.2760	0.4458	0.4375	0.4492	0.5135	0.5394	0.5142	0.5609	0.5599	0.4728	0.5132	0.5121	-0.1590	1.0000	0.3831	0.4854	0.2128
'TH'	0.4383	0.9017	0.9090	0.8667	0.8302	0.7883	0.7924	0.7121	0.6602	0.6524	0.5904	0.7189	0.3938	0.3831	1.0000	0.8410	0.0915
'U'	0.4412	0.9235	0.8959	0.8856	0.8901	0.8477	0.8140	0.7629	0.7343	0.7166	0.6535	0.7678	0.4130	0.4854	0.8410	1.0000	0.1426
'AG'	0.1976	0.1116	0.0826	0.1154	0.1622	0.2194	0.0990	0.0710	0.0716	0.1058	0.0692	0.0371	-0.2538	0.2128	0.0915	0.1426	1.0000

APPENDIX E: ABBREVIATIONS USED FOR THE DIFFERENT ELEMENTS FOR THE DIFFERENT DATASETS

Metro 2001 Soil Survey Data

Element	Abbreviation
% Sample Passing USGS Sieve # 10	P10
% Silt	Silt
% Clay	Clay
% Sand and Gravel	SG
% Organic Matter	OM
Soil pH	PH
Extractable Phosphorous	EP
Exchangeable Potassium	eK
DUBNAAluminum	dAl
DUBNAAntimony	dSb
DUBNAArsenic	dAs
DUBNABarium	dBa
DUBNABromine	dBr
DUBNACalcium	dCa
DUBNACerium	dCe
DUBNACobalt	dCo
DUBNAChromium	dCr
DUBNACesium	dCs
DUBNACopper	dCu
DUBNAEuropium	dEu
DUBNAIron	dFe
DUBNAHafnium	dHf
DUBNAPotassium	dK
DUBNALanthanum	dLa
DUBNAMagnesium	dMg
DUBNAManganese	dMn
DUBNAMolybdenum	dMo
DUBNASodium	dNa
DUBNANickel	dNi
DUBNALead	dPb
DUBNARubidium	dRb
DUBNAScandium	dSc
DUBNASelenium	dSe
DUBNASamarium	dSm
DUBNAStrontium	dSr
DUBNATantalum	dTa
DUBNATerbium	dTb
DUBNAThorium	dTh
DUBNATitanium	dTi
DUBNAUranium	dU
DUBNAVanadium	dV
DUBNATungsten	dW
DUBNAYttrium	dY
DUBNAZinc	dZn
InterpollCadmium	iCd
InterpollCopper	iCu
InterpollLead	iPb

MN Statewide 2003 Soil Survey data

Element	Abbreviation
% Sample Passing USGS Sieve # 10	P10
% Sand and Gravel	S&G
% Silt	SILT
% Clay	CLAY
% Organic Matter	OM
Soil pH	PH
Extractable Phosphorous	EP
Exchangeable Potassium	EK
Soluble Salts	Ssalt
Total Solids	TOT_SOL
Copper	CU
Nickel (Inductively Coupled Plasma)	NI_ICP
Lead	PB
Titanium	TI
Aluminium	AL
Iron	FE
Magnesium	MG
Calcium	CA
Sodium	NA
Potassium	K
Manganese	MN
Bromine	BR
Scandium	SC
Vanadium	V
Chromium	CR
Cobalt	CO

Element	Abbreviation
Nickel(Neutron Activation analysis)	NI_NAA
Zinc	ZN
Arsenic	AS
Antimony	SB
Selenium	SE
Rubidium	RB
Cesium	CS
Strontium	SR
Barium	BA
Lanthanum	LA
Cerium	CE
Neodymium	ND
Samarium	SM
Europium	EU
Terbium	TB
Ytterbium	YB
Lutetium	LU
Zirconium	ZR
Hafnium	HF
Tantalum	TA
Molybdenum	MO
Tungsten	W
Thorium	TH
Uranium	U
Silver	AG

APPENDIX F: MULTI DIMENSIONAL SCALING (MDS)

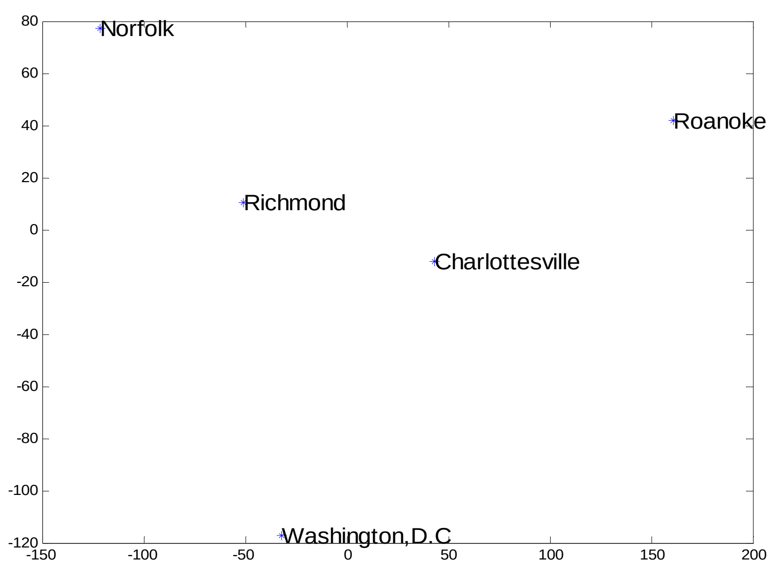
Multi Dimensional Scaling (MDS)

The goal of multidimensional scaling (MDS) is to produce a low-dimensional coordinate representation of distance information. For each data samples, a corresponding location in a low dimensional space is determined that preserves (as much as possible) the inter-point distances between the input samples. For instance, let us consider that we are provided with the Table F.1 which has the Traveling distance information between different cities in miles.

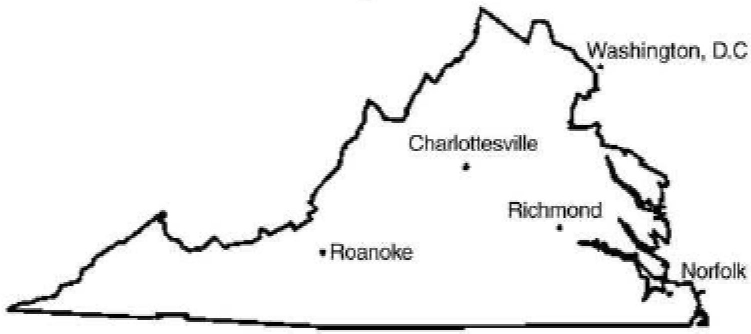
Table F.1. Pair wise Distances between Data Points (cities) used as Input for MDS.

Traveling Distance (in miles)	Washington,D.C	Charlottesville	Norfolk	Richmond	Roanoke
Washington,D.C	0	118	196	108	245
Charlottesville	118	0	164	71	123
Norfolk	196	164	0	24	285
Richmond	108	71	24	0	192
Roanoke	245	123	285	192	0

MDS will use these pair wise distances between different cities and construct a 2-D map that will preserve the inter-point distance measures between the different samples (cities). A typical MDS output for this data is provided in Figure F.1 (a)



(a)



(b)

Figure F.1 Coordinate reconstruction using Multi Dimensional scaling (MDS). (a) This plot shows the output produced by MDS for pair wise distance data in Table F.1. (b) For comparison, the actual location of the cities on a map of Virginia. A comparison shows that the relative distances between the cities are preserved, but a rotation of the MDS Map is required to match the actual Map.

As seen from Figure F.1 (a) it is apparent that the MDS preserves the distance relation between the different cities, however it needs to be noted that the pair wise distances are invariant to translations and rotations, MDS cannot reconstruct these aspects of the input data. This is seen from a comparison of Figure F.1 (a) and (b) where we see that the actual Map (as in Figure F.1 b) can be obtained by rotating the MDS Map in Figure F.1 (a) by 180^0.

Note: This example has been reproduced from [4].