

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/229043793>

# Visualizing Geographic Classifications Using Color

Article in *Cartographic Journal* · December 2000

DOI: 10.1179/0008704.37.2.p93

---

CITATIONS

5

---

READS

97

1 author:



[Bruce A. Maxwell](#)

Colby College

122 PUBLICATIONS 1,959 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Maine Concussion Management Initiative [View project](#)

# Visualizing Geographic Classifications Using Color

Bruce A. Maxwell  
Swarthmore College  
Swarthmore, PA 19081  
maxwell@swarthmore.edu  
(610)328-8081 (phone)  
(610)328-8082 (fax)

## Abstract

This paper presents the cartographic elements of a system for classifying and visualizing high-dimensional geographic datasets. The system has been developed as part of the Land Ocean Interactions in the Coastal Zone [LOICZ] project. The goal of the system is to develop regional and global typologies of coastal zones using large multi-variable datasets. Our implementation brings together statistical clustering algorithms with visualization capabilities to allow easy analysis and comprehension of the results. The two main tasks of the visualization are to allow for discrimination of multiple classes and to show relationships between those classes. These are accomplished in two different visual presentations. In both cases, the system selects colors appropriate to the purpose. In the latter case--showing relationships--the system uses a novel iterative refinement algorithm to select the colors. The results show that the system is successful at both generating the classes and visualizing the relationships between them.

**Keywords:** typology, clustering, visualization, color, color selection

# Visualizing Geographic Classifications Using Color

## 1 Introduction

The Land-Ocean Interaction in the Coastal Zone project [LOICZ] is a component of the International Geosphere-Biosphere Programme [IGBP] that focuses on the area of the earth's surface where land, ocean and atmosphere meet and interact. The overall goal of this project is to determine at regional and global scales: the nature of that dynamic interaction; how changes in various compartments of the Earth system are affecting coastal zones and altering their role in global cycles; to assess how future changes in these areas will affect their use by people; and, to provide a sound scientific basis for future integrated management of coastal areas on a sustainable basis [Pernetta, JC and Milliman, JD (Editors) (1995)].

A primary LOICZ objective is developing global scale-estimates of biogeochemical fluxes of carbon, nitrogen, and phosphorous [C, N and P] in and through the coastal zone [CZ]. The strategy adopted is to identify 'type-specimen' CNP budgets for well-characterized coastal regions, to further identify the coastal regions around the world of which such functional observations might be typical, and to use this typology relationship to upscale the limited local data to an estimate of global coastal zone function. Within this context, a typology is defined as a classification system that divides coastal zones into a set of classes according to one or more physical, geological, atmospheric, or human-related variables.

The development of an inventory of standard-format CZ budgets is in progress in the Biogeochemical Budgets task of LOICZ (<http://data.ecology.su.se/mnode/>). The Typology project (<http://www.nioz.nl/loicz/typo.htm>) is responsible for developing the coastal classification approach needed for budget upscaling. One of the major strategies adopted by the typology project is the development of clustering and visualization techniques suitable for classifying coastal areas in terms of their similarity with respect to environmental variables relevant to biogeochemical function.

Visualization of geographic classifications is a fundamental part of the systematic typology development process. The process is described in detail in [Maxwell, M and Buddemeier, B (in review)]. Appropriate visualization allows users to intuitively assess the typologies and make judgements about the similarity of geographic classes. This in turn can identify problems with typologies or build confidence in the results.

There are two major elements in the visualization of geographic classes. First, the user must be able to discriminate the classes in order to get a sense of the spatial distribution of individual class members. Second the user needs to be able to visualize class relationships such as their overall relative similarity.

The first issue is one of color selection for maximum discrimination. Appropriate user interaction can also add discriminability to a set of selected colors. Our solution to this problem, is to use an empirically selected set of colors for the initial ten classes, followed thereafter by randomly selected colors from the RGB spectrum. We augment the visualization by giving users the ability to turn the coloring for a particular class on or off.

The second issue involves mapping high-dimensional distances--e.g. between N-dimensional vectors, where N may be 20 to 100--to a low dimensional space--e.g. red, green, and blue [RGB].

This problem is quite different from the standard one of visualization of a single variable using a color spectrum. This paper presents a novel solution to this mapping based on an iterative refinement process. The resulting colors closely match the relative similarity relationships between the geographic classes, and provide an intuitive visual understanding of the class relationships.

We bring these tools together in the LOICZView application and demonstrate the results of these algorithms on a 17-variable subset of the global LOICZ data. The paper concludes with a discussion and directions for future work.

## 2 Theory and methodology

This sections provides a brief overview of the methods used in the typology development, and a more extensive description of the visualization methods.

### 2.1 Typology class development

There are two parts to our typology development process: 1) selecting a similarity measure for multivariate geographic data points, and 2) selecting a clustering method to generate classes of similar geographic points given the similarity measure.

#### 2.1.1 Distance measures for multivariate geographic data points

We have explored the use of several similarity measures for multivariate geographic data points. To clarify the meaning of this phrase, a data point is a physical location or area--in our dataset it constitutes a 1-degree by 1-degree area. Each data point has a set of measurements, or variables, associated with it. These include variables like air temperature, sea surface temperature, soil moisture, precipitation, etc. If we collect these variables into a single vector, then each data point is represented by an n-tuple, or n-dimensional vector, that describes the geographic location.

The significant issues we face when selecting an appropriate measure of similarity between data points--and thus between two multi-dimensional vectors--are that 1) the variables do not all have the same mean or standard deviation, and 2) not all of the variables are meaningful for all data points. For example, sea surface temperature is not meaningful for a data point that contains no ocean. Likewise, soil moisture is not meaningful for a data point that contains no land. Furthermore, because the variables are not necessarily normally distributed the covariance matrix of the data points is not necessarily invertible, invalidating traditional similarity measures such as the Mahalanobis distance [Harff, J and Davis, JC (1990)].

To deal with these problems, we have defined two different similarity measures, one based on the average scaled error, and one based on the maximum scaled error.

We define the average scaled Euclidean [ASE] distance between two points,  $D_A$ , as in (1),

$$D_A = \frac{\sum_{i \in Valid} \frac{(x_i - y_i)^2}{\sigma_i^2}}{card(Valid)} \quad (1)$$

where  $x$  and  $y$  are the two data point vectors,  $\sigma_i^2$  is the variance of variable  $i$ ,  $Valid$  is the set of dimensions that have valid data in both  $x$  and  $y$ , and  $card(Valid)$  is the number of valid dimensions.

The distance measure  $D_A$  can be interpreted in the following intuitive manner. If the value is less than one, then the average difference between  $x$  and  $y$  in any one dimension is less than a standard deviation. If the value is greater than one, then the average difference is greater than a standard deviation. Taking the square root of  $D_A$  would provide an exact measure in terms of standard deviations.

An alternative distance measure for geographic classification is to use the maximum scaled difference [MSD] between corresponding variables rather than the average scaled distance. In other words, two vectors that are identical except for a single variable  $x_i$ , will have the scaled difference in  $x_i$  as their distance. Compare this to a traditional measure, where the fact that most of the differences are zero drives the Euclidean or scaled Euclidean distance towards zero as the number of dimensions increases. A formal definition of the distance is given in (2).

$$MSD(A, B) = \max_{i \in I} \left( \frac{(A_i - B_i)^2}{\sigma_i^2} \right) \quad (2)$$

The MSD is a well-behaved distance measure since it obeys the properties of identity, symmetry, and the triangle inequality. In other words, two vectors that contain all variables can only have a distance of zero if they are equal to one another (identity property). Two vectors have the same distance no matter the order in which they are considered (symmetry property); and if  $MSD(A, B) \neq 0$  and  $MSD(A, C) = 0$ , then  $MSD(B, C) \neq 0$  (triangle inequality), which just states that if two points are not equal, they cannot both be equal to some third point. The MSD also behaves nicely both with respect to missing variables--it just considers variables that exist in both data points--and multiple variables that carry the same information--it considers only the maximum difference.

Another way of thinking about the MSD is that it lets the extremes rule judgements of similarity; two vectors cannot be similar if they have a single variable that is very different. In our implementation of MSD distance, we use the maximum normalized squared difference, where the normalization constants are the variances of the specific variables.

The MSD distance is inspired by the Hausdorff distance, which is a measure of similarity between *sets* that has been used successfully in image comparisons and object recognition tasks in the field of computer vision [Huttenlocher, DP, Klanderman, GA, and Rucklidge, WJ (1993)]. It has also recently been used in data mining applications to select variables and build decision trees [Pirumuthu, S (1999)]. The Hausdorff distance says that the distance between two sets  $A$  and  $B$  is the maximum of the minimum distances between all points in  $A$  and all points in  $B$ .

### 2.1.2 Clustering data points to generate classes

Given a definition of similarity, we can now start to look for natural groupings of similar points that may indicate the existence of a meaningful class. A standard method for clustering similar points is unsupervised k-means clustering, also called vector quantization [VQ] [Anderberg, M R (1973)][Rabiner, L, and Juang, B-H (1993)]. Overall, the algorithm takes as input a distance mea-

sure, a dataset, and a desired number of clusters. It then attempts to find a set of vectors that best represents the dataset. Each of these vectors is the mean vector of a unique subset of the data points. The output of the VQ algorithm is the set of mean class vectors and a tag for each data point, indicating its class membership.

Since there is a random element to the VQ algorithm, it is important to run it multiple times with the same inputs. The best set of class vectors  $V$  is the set that minimizes the overall representation error, which can be defined as the sum of the distances between each point and its nearest mean class vector.

We select the desired number of classes using a combination of expert judgement and the information theoretic description length of the classes and the resulting representational error [Ris-sanen, J (1989)]. This process is described in more detail in [Maxwell, M and Buddemeier, B (in review)].

## **2.2 Selection of colors and a user interface for maximum discrimination**

The two major issues in visualization are 1) how to enable users to discriminate between different classes, and 2) how to enable users to visualize relationships between different classes.

### **2.2.1 Selection of colors for maximum discriminability**

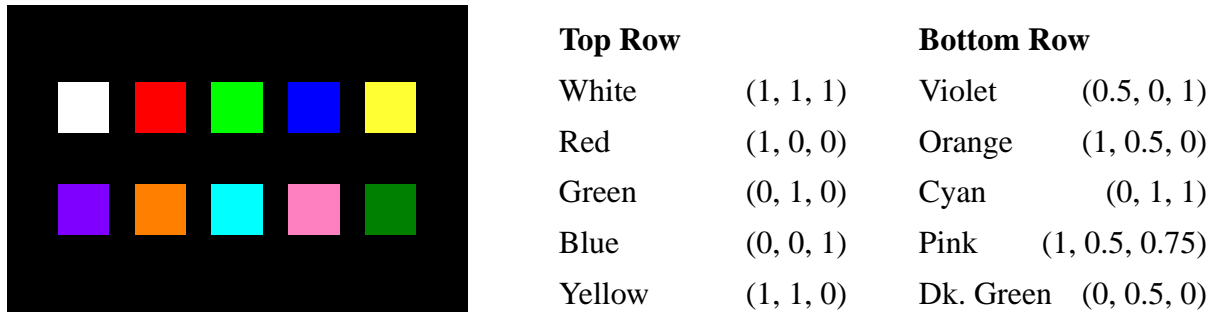
The selection of colors for maximum discriminability of the different classes is a much more difficult problem than map-coloring, which seeks to maximize local discriminability between adjacent neighbors. Since the class development method does not restrict the spatial characteristics of the classes--i.e. they do not have to be contiguous or located in a particular geographic area--it is possible for every class to be spatially located next to every other class. Therefore, each class must have its own unique color, and, if possible, each color must be discriminable from every other color.

We have taken an empirical approach to this problem based on our own observations and typical numbers of classes that users can usefully visualize. These empirical observations are as follows:

- It is difficult to select more than twenty colors that can be easily differentiated based on 4-6 pixel-wide dots on a black background (we have empirically found it better to use a black background than white when using a computer monitor),
- Users can easily visually identify points whose color is changing in response to some stimulus, like clicking a mouse, and
- For the LOICZ datasets, it will be rare for users to be looking at more than 20-30 classes, and more commonly they will be considering 10-20,

The second observation is also supported by the fact that people possess a strong ability to detect motion in their environment that is separate from their color detection capability [Sperling, G and Lu, Z-L (1996)].

This combination of observations led us to the following implementation. First, we have specified by hand the colors for the first 10 classes the user wants to visualize. These ten colors are shown in Figure 1. These colors are easily discriminated by a person with normal color vision, even when intermixed with points sizes of 2-4 pixels. Beyond the first ten, the colors are selected by generating random 24-bit values--creating 8-bit red, green, blue triples--and ensuring that each is



**Figure 1** Hand-picked colors for maximum discrimination between up to 10 classes.

bright enough to be visible. Second, to take advantage of the fact that people can easily see objects when they change, we give the user the ability to turn the coloring of classes on and off, switching the data points between a medium gray and their selected color. Thus, the points flash in response to a user's mouse clicks.

### 2.2.2 Iterative refinement for visualization of cluster relationships

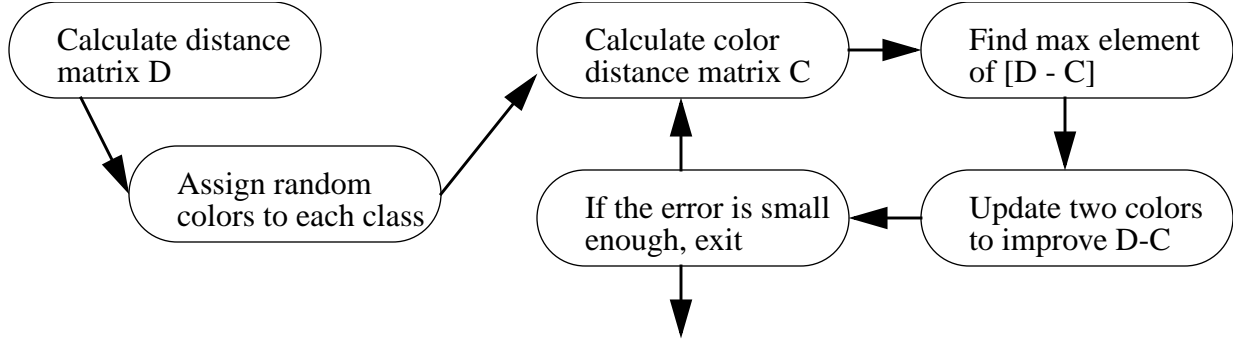
Throughout the process of cluster or region development and merging it is important to be able to visualize the process and the results. The LoiczView program provides an intuitive graphical user interface to the set of tools that implement the typology development and visualization methods described above. In particular, it allows the user to visualize both the spatial distribution of classes and, through color relationships, the similarity of classes.

The program uses a novel iterative refinement technique for selecting the display colors to represent distances between color vectors. This is a hard problem because the distances calculated between classes reside in a high-dimensional space--up to 100 dimensions--while color resides in a three dimensional space. Therefore, in most cases we cannot select a set of colors whose distances exactly mirrors the true distances between the mean class vectors.

As a simple example of this problem, consider five points in a five dimensional space that are all equidistant from one another. One set of points that meets this criteria is the set  $\{(1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 1, 0), (0, 0, 0, 0, 1)\}$ . In this case, each 5-D point is  $\sqrt{2}$  away from every other point. In a 3-dimensional space, it is only possible to have four points equidistant from one another--a tetragon. It is not possible to generate five points that are equidistant from one another in a 3-D space. Therefore, the best we can do when selecting colors is to approximate the true distances in color space.

The problem can be set up as follows. First, calculate the matrix of distances between each class vector. Normalize this matrix by dividing each element by the largest element of the matrix. Now all of the distances are in the range  $[0, 1]$ .

Second, generate a set of random colors and assign one color to each class. Now calculate the matrix of distances between the colors in color space. In this description of the technique, we will use the RGB color space, letting each axis ranges from  $[0, 1]$ . Now we have two matrices whose elements are in the range  $[0, 1]$ . The following algorithm will iteratively modify the class colors so that it reduces the difference between the two matrices.



**Figure 2** Graphical representation of iterative algorithm

```

Calculate the normalized class distance matrix D
Assign a random color to each class
Set the adjustment rate A (e.g. 20%)
Loop
    Calculate the color distance matrix C
    Let  $E_{ij}$  be the largest magnitude element of D-C
    Let I and J be the classes whose error is  $E_{ij}$ 
    Let  $C_{ij}$  be the color vector from  $color_j$  to  $color_i$ 
    Adjust the color values of I and J to reduce E
Until the matrices are close enough or we've looped enough
  
```

The update rule for the class colors is given in (3).

$$\begin{aligned}
 color_i &= color_i + C_{ij}AE_{ij} \\
 color_j &= color_j - C_{ij}AE_{ij}
 \end{aligned} \tag{3}$$

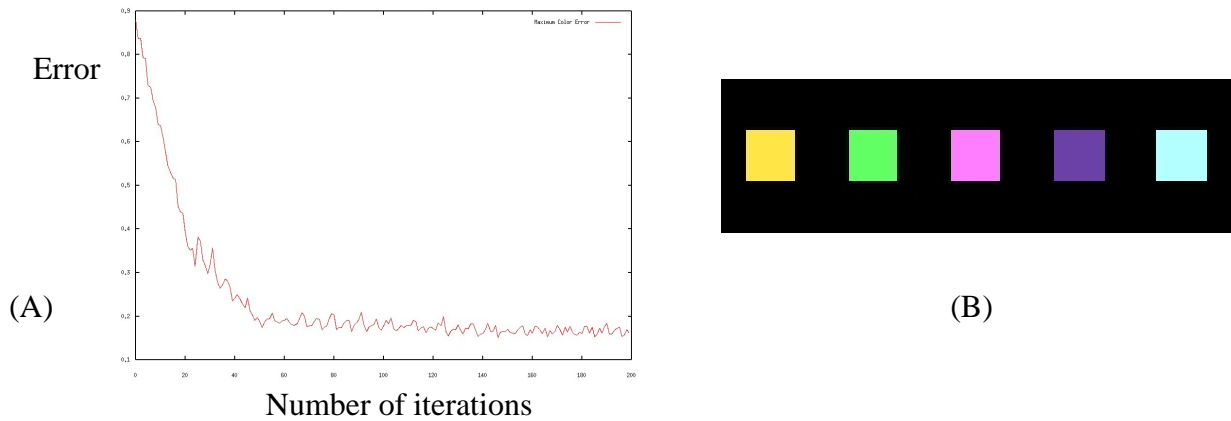
Figure 2 shows a graphical representation of the color selection process.

The number of iterations required to produce a good result is dependent upon the size of the matrix and the number of classes. For a 10x10 matrix--i.e. 10 classes--200 iterations achieves a result that no longer changes significantly in terms of the largest error between the two matrices. For a much larger matrix, more iterations may be required. As each iteration requires a search for the maximum different between two matrices, each iteration takes  $O(n^2)$  time, where  $n$  is the number of classes. Since this is a process that only needs to be done once per visualization, the algorithm is sufficiently fast.

The adjustment rate is an important parameter of the problem. The adjustment rate needs to be fast enough to allow improvement, but not so large that the system overshoots good solutions. A rate of 20% appears to work well for the LOICZ dataset.

To demonstrate the process, we can show the results of the algorithm on a set of mean class vectors identical to the example given above:  $\{(1, 0, 0, 0, 0), (0, 1, 0, 0, 0), (0, 0, 1, 0, 0), (0, 0, 0, 1, 0), (0, 0, 0, 0, 1)\}$ . Using 200 iterations and a 20% learning rate, the resulting colors and the training graph are shown in Figure 3. From this example we can see that the resulting colors are saturated and easily discriminable from one another.





**Figure 3** A) Training curve showing maximum error between distance and color matrices, B) Resulting colors for simple equidistant example.

### 3 Experiments and results

The methods described above allows us to analyze and visualize large heterogeneous datasets such as the LOICZ dataset. To test and refine these methods we have applied them to a subset of the LOICZ dataset and compared the results with expert judgements.

Our process for developing and validating a horizontal typology (not hierarchical) is as follows.

1. Select the variables to use
2. Select how many classes (clusters) to create
3. Apply the VQ algorithm using an appropriate distance measure
4. Apply semantic labels to each cluster
5. Compare with expert judgement or pre-existing typologies

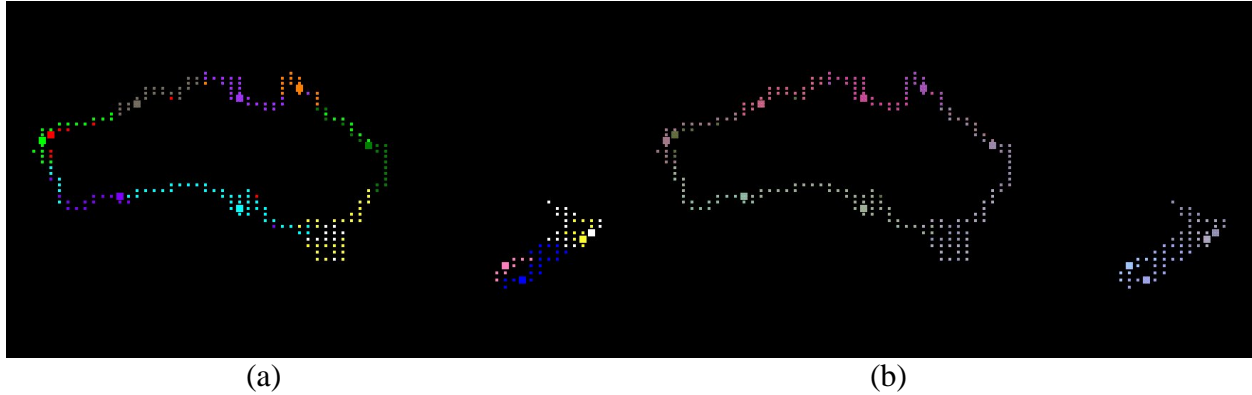
For our prototype typology development we use a subset of the LOICZ dataset corresponding to the Australia/New Zealand coastline. This dataset has a spatial resolution of 1 degree.

For this paper, which focuses on the cartographic aspects of this process, we will follow through steps 1 and 3 and show how the visualization helps to analyze and validate the results for a 12-class subdivision.

#### 3.1 Variables

In this experiment the variable selection was based on two factors. First, did the variable provide good coverage of the area (<10% missing data). Second, did the variable actually provide useful information (vary in a reasonable way over the dataset). Beyond these two considerations, the primary concern was not to give too strong a weight to any one aspect of the environment. The end result was a set of 17 variables.

The variables we selected included: seasonal precipitation (max and min), seasonal air temperature (max and min), seasonal sea surface temperature (max and min), seasonal soil moisture (max and min), seasonal salinity (max and min), seasonal Coastal Zone Color Scanner [CZCS] (max and min), average annual runoff, an annual evaporation proxy, average wave height, standard deviation of elevation, and a tidal mixing proxy. Precipitation and air temperature information are from [IPCC Data Distribution Centre for Climate Change and Related Scenarios for Impacts



**Figure 4** (a) 12-class clustering result for Australasia using average scaled Euclidean distance and randomly selected colors. (b) Same clustering result but with colors selected to reflect the similarity of the classes

Assessment, CD-ROM, Version 1.0, April 1999.], the remaining variables are from the LOICZ typology dataset [LOICZ typology dataset, <http://www.kellia.nioz.nl/loicz/typo.htm>, 1999.]. For the Australasia coast we modified the LOICZ typology data by interpolating it to cover locations with no data. For the most part this meant taking land cell variables and interpolating them onto adjacent coastal cells, and taking sea cell variables and interpolating them onto adjacent coastal cells.

The evaporation proxy is a combination of wind speed and vapor pressure. The proxy variable is the product of the two multiplied by 10 (vapor pressure is water vapor pressure multiplied by 10). The vapor pressure variable came from [IPCC Data Distribution Centre for Climate Change and Related Scenarios for Impacts Assessment, CD-ROM, Version 1.0, April 1999.] and the wind speed from [LOICZ typology dataset, <http://www.kellia.nioz.nl/loicz/typo.htm>, 1999.].

The tidal mixing proxy is a combination of a tidal form variable [semidiurnal, mixed, diurnal] and tidal range. The tidal mixing proxy is tidal range multiplied by tidal frequency, where tidal frequency is [semidiurnal = 2, mixed = 1.5, and diurnal = 1]. The two base variables came from [LOICZ typology dataset, <http://www.kellia.nioz.nl/loicz/typo.htm>, 1999.].

### 3.2 Cluster the data

We used the VQ algorithm using the average scaled Euclidean distance measure to generate a set of representative classes. To get a good set of classes we ran it ten times and took the lowest error result. This provided us with a reasonable set of representative classes for the data.

Figure 4(a) shows a visualization of the resulting classes by mapping them into an image using latitude, longitude, and using color to identify the class of each data point. Figure 4(b) shows a visualization of the same clustering result, but with the class colors selected using the iterative selection algorithm to show the relationships between classes. Note that three distinct classes exist, while the others merge into more of a continuum in the color similarity presentation.

This observation is supported if we look at the matrix of distances between clusters. Table 1 shows this matrix for the clusters in Figure 4. In each column, the most similar cluster is in bold. Note the pink and red clusters are both dissimilar to all of the other clusters (distances > 1.0), whereas the remaining clusters are all fairly similar to at least one other cluster (distances < 1.0).

**Table 1 Distance matrix between clusters in Figure 4**

Color	white	red	green	blue	yellow	dark purple	orange	cyan	pink	dark green	purple	brown
white	0	3.4	0.98	0.85	<b>0.32</b>	0.67	2.2	0.50	2.8	0.62	2.3	2.0
red	3.4	0	2.4	4.2	3.7	4.1	3.8	3.3	6.7	3.4	3.1	2.6
green	0.98	<b>2.4</b>	0	2.5	1.3	1.1	1.1	0.7	5.2	<b>0.49</b>	0.93	0.72
blue	0.85	4.2	2.5	0	0.54	1.6	2.9	2.0	<b>1.1</b>	1.5	3.4	3.3
yellow	<b>0.32</b>	3.7	1.3	<b>0.54</b>	0	0.62	2.3	0.70	2.1	0.58	2.7	2.4
dk. purple	0.67	4.1	1.1	1.6	0.62	0	2.3	<b>0.31</b>	3.9	0.68	2.8	2.9
orange	2.2	3.8	1.1	2.9	2.3	2.3	0	2.5	5.0	0.92	<b>0.45</b>	0.93
cyan	0.5	3.3	0.7	2.0	0.70	<b>0.31</b>	2.5	0	4.5	0.82	2.4	2.3
pink	2.8	6.7	5.2	1.1	2.1	3.9	5.0	4.5	0	3.4	5.8	5.7
dk. green	0.62	3.4	<b>0.49</b>	1.5	0.58	0.68	0.92	0.82	3.4	0	1.4	1.2
purple	2.3	3.1	0.93	3.4	2.7	2.8	<b>0.45</b>	2.4	5.8	1.4	0	<b>0.57</b>
brown	2.0	2.6	0.72	3.3	2.4	2.9	0.93	2.3	5.7	1.2	0.57	0

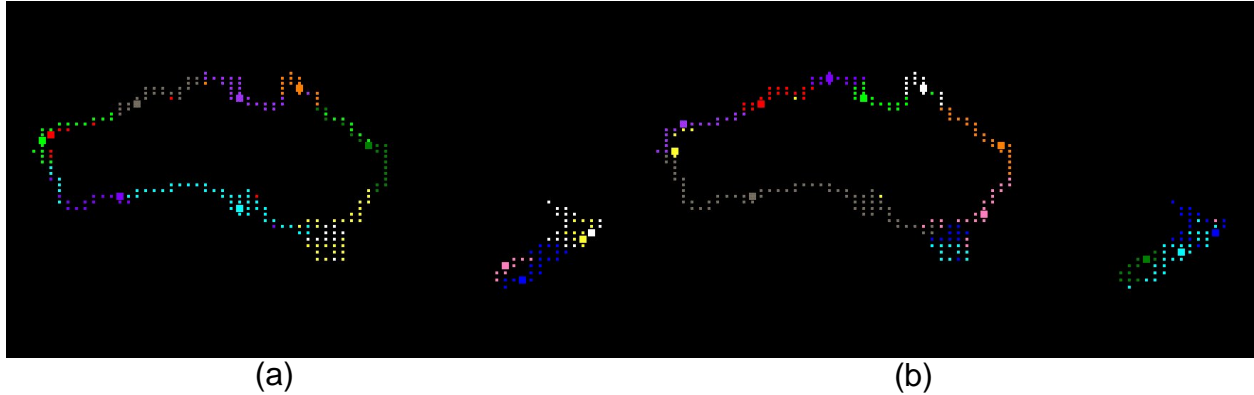
This relationship is echoed in the iterative color scheme, where both the pink and red clusters in Figure 4(a) stand out in comparison to their neighbors in Figure 4(b). What is also indicated in the similarity color scheme is that the northern clusters are fairly similar to one another but quite different than the southern clusters. This is supported by examining the distance matrix. The brown, purple, and orange clusters are all fairly similar to one another, but are distant from the yellow, cyan, dark purple, white, and blue clusters that make up the southern area.

### 3.3 Comparison of average scaled Euclidean distance to the MSD distance

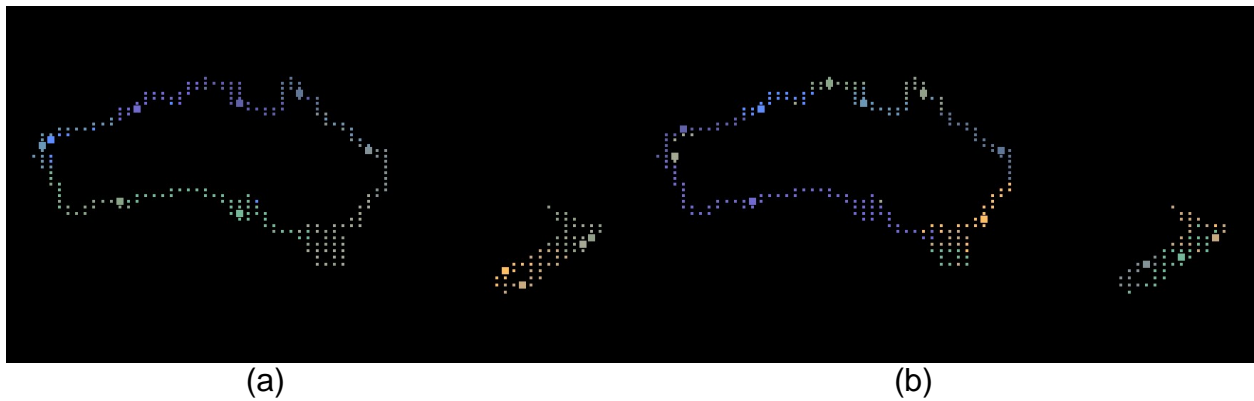
We can undertake the same process of typology development using the alternative MSD distance measure. Figure 5 compares 12-class clusterings using the average scaled Euclidean distance and the Hausdorff distance.

Note the similarities and differences between the two results. The biggest differences occurs on the southern and northern coasts of Australia where the southern coast apparently has fewer extreme differences (but higher average differences) than the northern coast. Thus, the MSD distance does not divide the southern coast into two sections in a 12-class clustering, but the average scaled Euclidean distance does.

Figure 6 shows the same clusterings but uses associates color with the similarity of the classes. In both cases the south coast of New Zealand shows up as a unique location. Likewise, the south east coast of Australia shows up as being similar to north New Zealand. The different between the ASE and MSD results are that the North-South Australia similarity is more pronounced in the ASE division than it is in the MSD division. This likely reflects the fact that many variables are different between the north and south classes, but the maximum differences may be more similar to the differences between the other classes.



**Figure 5** (a) 12-class clustering using average scaled Euclidean distance.  
(b) 12-class clustering using MSD distance.



**Figure 6** (a) 12-class ASE clustering with colors related to similarity, (b) 12-class MSD cluster with colors related to similarity.

## 4 Discussion and Future Directions

From the Australasia example, we can see that the process appears to produce a reasonable set of classes. The results show broad agreement with previous expert typologies [Smith, SV, and Crossland, C J (1999)]. Furthermore, they highlight localized phenomena that do not show up in the expert version, but nevertheless exist in the data. Note that we obtained these results despite heterogeneous variables with some missing data, indicating that the distance measures we used are appropriate for the task.

With respect to the visualization, our strategy for selecting colors for maximum discrimination appears to work well. It allows for easy discrimination of the different classes, as shown in Figure 5. In cases where colors may be confusing, the inclusion of the ability to turn clusters on and off makes it possible to easily distinguish class membership for each point in the visualization.

The iterative color selection for matching distances also appears to provide an intuitive visualization of similarity. The most common criticism of users and viewers is that the color selections are not always aesthetically pleasing or easy to see. However, the color relationships do appear to be intuitively correct. One area of future work is to objectively evaluate how well people interpret the results of the color selection program.

A second area of future work is to look at more limited color spaces such as pure intensity. The reason for pursuing this is to have an algorithm that anyone who can perceive intensity--but perhaps not certain colors--can use to visualize relationships. A mapping like this must be used in any situation where universal access to the information is required.

Overall this paper presents a set of methods that permit clustering, classification, and visual comparison of environments at regional and global scales. Clustering of high-dimensionality datasets can be based on scaled Euclidian distances in ways that permit the use of datasets that are incomplete, not normally distributed, or otherwise unsuitable for more traditional statistical analysis. Two different distance criteria -- the average scaled Euclidean distance and the maximum scaled distance -- provide alternative ways to explore the nature of environmental similarities and differences.

The paper also presents ancillary techniques that expand the applicability and ease of use of these methods. One of these is a strategy of using color and motion to visualize classes. The second is the use of a novel color-similarity approach that permits visualization of the similarities of spatially distributed clusters.

These techniques have been applied to a 17-variable coastal dataset for Australia and neighboring regions. The results are highly consistent with an independent expert-judgement coastal typology, and the differences and similarities between the various approaches to cluster definition are intuitively understandable in terms of the variables and techniques used.

The methods provide a novel and potentially powerful set of tools for classifying and visualizing relationships between environments. Initial applications will be regionalization and globalization of coastal C, N, P budgets as part of the LOICZ projects. However, the techniques are further applicable to different datasets, and to issues of global and regional change. In addition, the color selection techniques are applicable to any application that wants to visualize high dimensional relationships using color.

## **Acknowledgements**

The author would like to thank LOICZ for their support of the typology project and the work described herein. We would also like to thank AAAS for the Earth Systems Science Conference in 1997 in South Dakota which brought together scientists from a variety of disciplines and launched this approach to coastline typology development.

## References

- Anderberg, M R (1973) *Cluster Analysis for Applications*, Academic Press, New York.
- Harff, J and Davis, JC (1990) "Regionalization in Geology by Multivariate Classification", *Mathematical Geology*, Vol. 22, No. 5, pp.573-588.
- Huttenlocher, DP, Klanderman, GA, and Rucklidge, WJ (1993) "Comparing Images Using the Hausdorff Distance", *PAMI*(15), No. 9, pp. 850-863.
- IPCC Data Distribution Centre for Climate Change and Related Scenarios for Impacts Assessment, CD-ROM, Version 1.0, April 1999.
- LOICZ typology dataset, <http://www.kellia.nioz.nl/loicz/typo.htm>, 1999.
- Maxwell, M and Buddemeier, B (in review) "Coastal Typology Development with Heterogeneous Data Sets", *Journal of Regional Environmental Change*.
- Pernetta, JC and Milliman, JD (Editors) (1995) *Land-Ocean Interactions in the Coastal Zone: Implementation Plan*. IGBP Report No. 33. IGBP, Stockholm, 215 pp.
- Piramuthu, S (1999) "The Hausdorff distance measure for feature selection in learning applications", *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences*. IEEE Computing Society.
- Rabiner, L, and Juang, B-H (1993) *Fundamentals of Speech Recognition*, Prentice-Hall, Englewood Cliffs, NJ.
- Rissanen, J (1989) *Stochastic Complexity in Statistical Inquiry*, World Scientific Publishing Co. Ptc. Ltd., Singapore.
- Smith, SV, and Crossland, C J (1999) *Austalasian Estuarine Systems: Carbon, Nitrogen and Phosphorus Fluxes*, LOICZ Reports & Studies No. 12, ii + 182 pp. LOICZ, Texel, The Netherlands.
- Sperling, G and Lu, Z-L (1996) "The functional architecture of visual motion perception", *International Journal of Psychology*, 1996, 31, (3/4), 362.