PIXEL AND COLOUR MIXTURE: GIS TECHNIQUES FOR VISUALISATION OF FUZZINESS AND UNCERTAINTY OF NATURAL RESOURCE INVENTORIES

Tomislav Hengl[a], Dennis J.J. Walvoort[b], Allan Brown[c]

[a]Soil Science Division, International Institute for Geo-information Science and Earth Observation (ITC), P.O. Box 6, 7500 AA Enschede, the Netherlands, hengl@itc.nl;
[b]Laboratory of Soil Science & Geology - Wageningen University & Research Centre, Duivendaal 10, P.O. Box 37, 6700 AA Wageningen, The Netherlands, Dennis.Walvoort@oio.beng.wag-ur.nl;
[c]Division of Geoinformatics, Cartography and Visualisation, International Institute for Geo-information Science and Earth Observation (ITC), P.O. Box 6, 7500 AA Enschede, the Netherlands, brown@itc.nl;

Abstract

Natural resource inventories are more and more based on quantitative spatial methods. The discrete classes of vegetation, soil, land cover etc. can be quantified by using membership maps rather than only crisp delineations. GIS methods to visually explore fuzziness and uncertainty of the prediction maps are clearly needed. The paper presents two relatively new techniques for visualisation of multiple membership maps and uncertainty of spatial prediction methods: Pixel Mixture (PM) and Colour Mixture (CM). The PM assigns classes to sub-pixels relatively to the membership value, while CM calculates average colour taking the memberships as weights. The PM technique appears to be more simple, "user-friendly" and easier to interpret. The CM technique requires a special fuzzy legend that also reflects taxonomic relations and can be therefore considered a double-continuous approach to mapping. The derived saturation of the map calculated using the CM technique gives us insight into the spatial and taxonomic confusion of our map. Both PM and CM can be both applied to visualize the confusion index or relative error together with the maps of estimates. In this case we used brightness correction and white pixels to include the relative uncertainty and confusion. These techniques will be further on developed to produce operational maps to be used in decision-making.

Keywords: membership maps, Pixel Mixture, Colour Mixture, natural resources, confusion index

1. INTRODUCTION

Natural resources are natural objects that are usually mapped for management or conservation purposes. These can belong to the domains of botany, soil science, geology, ecology, hydrology, oceanology and other environmental and Earth sciences. Conventional natural resource inventories aim at producing maps and databases of natural objects grouped as types or classes and attributed with sets of descriptive and analytical variables. For example, forest vegetation is one type of natural resource. The forest vegetation resource inventory typically produces a map of forest vegetation communities. We will also expect to have descriptive data for each vegetation community class, structure, abundance of specific species and maybe also some measures important for management such as tree heights, canopy density and timber volume per hectare.

Some natural objects are rather discrete and static and therefore easier to map, while others are continuous, complex, hidden or rather dynamic. Discrete objects, such as lakes, rivers or geological strata are modelled in a GIS using polygon or line objects. Usually human

influence makes the semi-natural objects more discrete and therefore easier for inventorying. Different land cover types of an agricultural area will be easier to delineate on an aerial photo or satellite image compared to different forest types of a tropical forest region.

The discrete classes of natural resources are traditionally presented using sets of colours – the so called "choropleth", or more precisely "chorochromatic" map (Burrough and McDonnell, 1998). In GIS terms, this is a polygon map with defined topology and linked to attribute tables. This approach is also referred to as "double crisp", since the surveyor uses sharp boundaries to delineate soil bodies and crisp classes to classify natural objects (Burrough et al., 1997). Forest vegetation types, geomorphologic classes, soil types and similar classes are in fact not always so crisp as the map indicates. In fact, these are often complex and continuous in their distribution (Hootsmans, 1996). Especially soils can be considered as poorly defined classes of hidden objects and are suitable for fuzzy set or continuous approaches to model their distribution in a more continuous way (Burrough, 1993).

In last decades, methods for quantifying taxonomic classes, such as fuzzy classification, have been developed (De Gruijter and McBratney, 1988). However, there is still a need for GIS methods to visually explore results of fuzzy classification (Burrough et al., 1997, p. 284). Moreover, visualisation techniques are needed to allow users to explore uncertainty in spatial data visually and to investigate the effects of different decisions in the classification process (MacEachren and Kraak, 1997). Some of the new interactive data exploration techniques such as the use of slide bars, point and click events, blinking and animations have proven to give a good impression of data uncertainty. In this paper we give insight into two relatively new visualisation techniques.

2. QUANTIFICATION OF DISCRETE INFORMATION

2.1 Multiple membership maps

Memberships can be derived by means of different methods. Most commonly, memberships are derived using some continuous classification algorithm such as fuzzy k-means (De Gruijter and McBratney, 1988). Multiple possibilities are generally termed membership grades or membership values (MV). The membership values from the point samples can be interpolated using compositional kriging to produce membership maps (Walvoort and De Gruijter, 2001). Having multiple memberships for each pixel of a map, we can also make a conclusion about the *ambiguity* i.e. *indistinctness* of a specific class and overall *confusion* among all classes. We can also inspect the *fuzziness* defined as the overall deviation of all possible values (Hootsmans, 1996).

2.2 Continuous maps of discrete variables

Geoinformation on natural resources is today needed in a quantified manner and in fine grain of detail. This counts also for discrete variables, traditionally presented using chorochromatic maps. In many applications today, crisp maps are replaced by so called "continuous maps" and the resulting continuous approach to management of natural resources (McBratney and Whelan, 1995). Irvin et al. (1997) produced membership maps of landform classes using fuzzy *k*-means classification of landform parameters: elevation, slope, curvature, wetness index and incident solar radiation. This work was then further developed by Burrough et al. (2000). Similar, continuous maps can be found in vegetation (Brown, 1998) and land cover mapping (Hansen and Reed, 2000). De Gruijter and Boogaard (2001) give an application of fuzzy logic in quantitative land evaluation. A good overview of developed applications can be found in a special issue of Fuzzy sets and systems Journal called "Uncertainty in Geographic

Information Systems and Spatial Data" (Vol. 113) and a special issue of Geoderma Journal called "Fuzzy sets in soil systems" (Vol. 77).

3. VISUALISATION OF MULTIPLE MEMBERSHIP MAPS

Multiple membership maps are more detailed and can provide better ground for GIS modelling and decision-making than crisp polygon maps. The membership maps themselves can be used separately or reclassified to create chorochromatic maps. Membership maps for each class are the preferred source of information especially for advanced GIS applications since they contain more information than the crisp maps and can be used separately. However, in some cases we require to visualise the multiple memberships maps in order to get a compound picture of overall fuzziness in the map. Let us imagine the product of 8, 10 or more membership maps of vegetation classes. This large amount of membership maps will be hard to explore visually. The most common procedure to display the multiple membership maps is so called "defuzzification" where we construct the colour map taking the class with the highest membership in an individual pixel (Burrough et al., 2000). Normally, this is done only if the CI does not exceed some threshold value, e.g. Burrough et al. (2000) suggest 0.6 as the threshold value. The CI is usually not visualised together with the final map, although there have been some investigations in this direction (Fisher, 1993; Hootsmans, 1996, §7). Here we give an insight into two relatively new, somewhat different GIS techniques: Pixel Mixture (PM) and Colour Mixture (CM). We applied these two techniques using existing data sets from different projects. The CM technique is still under development.
The full colour visualisations and comparison of the PM and CM are available on-line at http://www.itc.nl/personal/hengl/VMM/.

3.1 Pixel Mixture (PM)
The pixel mixture technique was introduced by De Gruijter et al. (1997, Fig. 6) to visualise six membership maps of soil classes for a case study in the Netherlands. The algorithm works as follows. First we have to generate a grid within each cell, e.g. 3x3, 4x4, 5x5, 6x6 etc. We call these "sub-pixels". Second, we assign to each of the new sub-pixels a class so that the total number of sub-pixels assigned to each class in a cell is proportional to the membership values of each class. Finally, we randomise the position of the sub-pixels to avoid creating patterns. To understand the PM principle better, we give a simple example with 4 classes in a single pixel map (see Figure 1). Here the given membership values in a pixel *i* are 0.6, 0.2, 0.1 and 0.1. Since the new sub-pixel grid is 3x3, we assign five pixels to class A, two pixels to class B, one to class C and one to class D. The resulting realisation of PM can then be presented using the original colours for the classes.
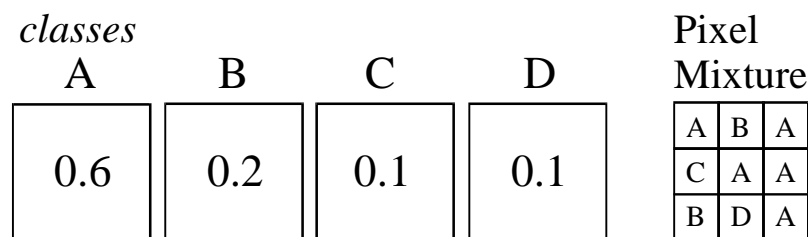
Figure 1. Schematic example of Pixel mixture technique for four MVs (left) and onepresentation of PM in 3x3 window (right).

We can notice that the MVs can be back-derived by simply counting the pixels for a specific class in the 3x3 window. This is just an approximation of original MVs that depends on the
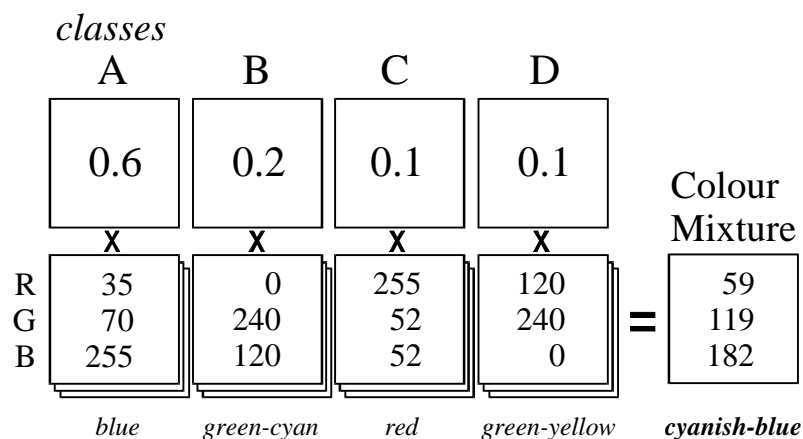
discretisation level of the PM.

3.2 Colour Mixture (CM)

In the case we have only three membership maps, we can easily display them using a colour composite command in image processing software. For example: the classes A, B and C can be presented using the Red Green and Blue bands (RGB), which is a standard colour coding system (Brown and Feringa, 1999). If the number of classes is higher than three, we require some other technique. Here we introduce the CM technique, where we calculate mixed colours as an averaged brightness of RGB bands:

$$[R,G,B] = \frac{\sum_{i=1}^{n}\sum_{C=1}^{k}(MV_i \cdot [R_C, G_C, B_C])}{\sum_{i=1}^{n} MV_i}$$

(1.)

where the $[R,G,B]_i$ are the new colours in pixel i, $[R_C, G_C, B_C]$ are the selected class colours or representations and k is the number of classes. In the case of standardised data, the denominator in the Eq. (1), i.e. the sum of MVs equals unity. The mixed RGB are then again combined in a colour composite. The simplified example can be seen in Figure 2.

classes

	A	B	C	D	
	0.6	0.2	0.1	0.1	Colour Mixture
	X	X	X	X	
R	35	0	255	120	59
G	70	240	52	240	119
B	255	120	52	0	182
	blue	*green-cyan*	*red*	*green-yellow*	***cyanish-blue***

Figure 2. *Schematic example of Colour mixture technique: the new colour is calculated as the weighted average of class representations.*

In the above example, we derive a new colour (cyanish-blue) as a weighted average of four colours. The derived colour reflects the highest MVs: blue and green-cyan colour. However, the CM technique based on the Eq. (1) is more complex than it appears on the first sight. The main problems are how to interpret the new colours and how to create a legend showing all possible colours for all combinations of MVs? We can easily conclude that, if we choose the legend class colours for each individual class entirely freely, it may happen that a new colour derived using the weighted means method will not appear in the original legend (accidental colour). Second more serious problem can be illustrated by means of following example: let us imagine a case with three classes (A, B, C), where the representation of class B lies midway along the line connecting the A and C classes in the colour cube (RGB). So a pixel colour which appears the same as the colour of class B in the legend could indeed indicate class B, but it could also result from the maximum confusion between the A and C classes (i.e. $MV_A = MV_C = 0.5$). The possibility of such situations increases with the addition of more class colours in the colour cube, although the number of possible outcomes can be reduced if

all original class colours are placed on a plane within the colour cube. At the end, it makes sense to choose colours so that classes that are close together taxonomically are also close together in the colour cube. Therefore, we will need to design our representation colours in such a way that:

- all legend colours lie in the same colour plane;
- none of the colours in the legend can be derived accidentally from the different combinations of class representations;
- mixed colours will represent relative relations both in the geographical and taxonomic space;

Based on these assumptions we concluded that the optimal solution is to design the legend as a cross section of the colour cube, so that all derived colours lie on the same plane. In order to avoid some colours being too dominant perceptually, we decided to keep the intensity/brightness constant. The intensity (I), here calculated using the formula:

$$I = \frac{R+G+B}{3}$$

(2.)

will be constant if we take a colour plane perpendicular to the diagonal of the RGB colour cube. The Hue, Saturation, Intensity (HSI) colour model has been already recognised as promising for the visualisation of uncertainty (Jiang et al., 1995). The diagram in Figure 3 shows colours used for five classes, in which the angular distance between any two classes is roughly proportional to their "taxonomic" distance in the classification. This means, for example, that the colour lying on the intersection of the lines c1-c2 and c3-c4 is more likely to be a mixture of c1 with c2 than of c3 with c4. The mixed colour itself will appear very similar to colour c4, however. The most desirable situation occurs if class c4 is itself taxonomically between c1 and c2. Towards the centre of the circle we are less sure about the components of any mixture, but here the colours are approaching grey and the hues become less distinct.
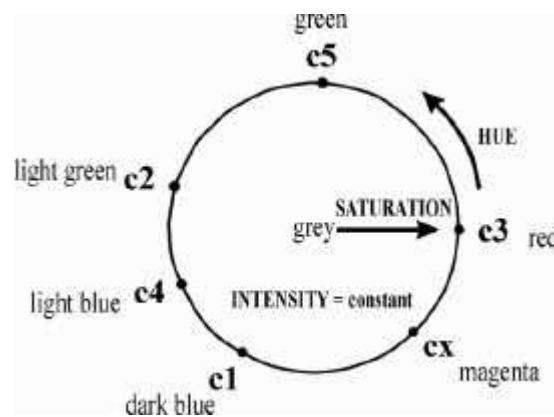
Figure 3. Fuzzy legend for the CM technique. The colour wheel is has constant intensity while saturation changes radially.

4. VISUALISATION OF UNCERTAINTY

A quantitative GIS method will often provide both the estimates and the uncertainty or error related to the specific model. This error can often be mapped over the entire area to show spatial distribution of uncertainty. Similarly, we can map the fuzziness of the continuous classification using the Confusion Index. This is our classification uncertainty. The PM and CM can also be applied to visualize the estimation error or confusion index together with the estimates of the target variables in the same map. Then we will be able to investigate how

does the error changes with the change of the target variable. If the uncertainty or fuzziness are visualised for different mapping methods, we can also make direct comparison of the mapping quality.

4.1   Confusion index

The measure most commonly used to inspect interference among multiple membership maps is the confusion index (CI) calculated as the difference between the first and second highest membership class per pixel:

$$CI = 1 - (MV_{max} - MV_{2nd\,max}) \text{ or } CI = \frac{MV_{2nd\,max}}{MV_{max}} \tag{3.}$$

where MV_{max} is the highest MV and the MV_{2ndmax} is the second highest MV (Burrough et al., 1997). The CI map calculated from one of the equations in Eq. (3) will show higher values in the areas of transition between different classes and therefore is suitable to detect soft boundaries. If the CI map does not show distinct line pattern, it is more probable that we have higher overall confusion.. The brightness or intensity has been considered the preferred parameter to visualise the model fuzziness (Hootsmans, 1996; Jiang, 1996). Here we code the confusion index using the white colour. The final colours are corrected with the brightness factor: the brighter the pixel the higher is the fuzziness. Ultimately, white pixels indicate maximum confusion and can be considered as the undefind class. The colours are corrected using the linear stretch:

$$cor[R, G, B] = [R, G, B] + CI_i \cdot (255 - [R, G, B]) \tag{4.}$$

The corrected brightness value can be interpreted as follows: colours are equal to the original RGB bands for the CI=0, and completely white for maximum confusion index (CI=1).

4.2   Prediction error

Uncertainty gives an insight into the risks involved with the specific choice of the decision-making. Information on uncertainty of data, or data quality is typically not visualised in thematic maps (Burrough and McDonnell, 1998, §9). In the case of continuous variables, the uncertainty is directly related to the prediction error. We can visualise the uncertainty of estimation (or prediction) by using the brightness correction of the original colour map. Here we propose to use the brightness correction from Eq. (4) also to visualise the relative error. A map of a continuous variable is first displayed using the gradual representation from blue (low values) to red (high values). This is used as the default representation for system domain "Value" and system domain "Distance" in the ILWIS package (Unit Geo Software Development, 2001). The colours used in this legend are also suitable since they all have more or less the same brightness. The relative uncertainty is than used to correct the original RGB values of the map of estimation as in Eq. (4). Here instead of the CI we use relative error (u_r) which can be calculated as the ratio between the model variance and the total variability s_T, or the total range of target variable (s). The map resulting form this procedure looks very similar to the map resulting from the confusion index procedure described in the previous section: in both cases, colours fade towards white as uncertainty increases.

5. CONCLUSIONS

Both Pixel and Colour Mixture are relatively new GIS techniques that can be used to visualise multiple membership maps and their fuzziness in a composite map. Our experiences with these two methods have shown that both have advantages and disadvantages (Table 1). The PM technique appears to be more simple, "user-friendly" and easier to interpret. It also gives a better impression of the fuzziness and continuity in geographic space. On the other hand, the CM technique requires us to think also about the taxonomic classification space and therefore can be termed "double-continuous" or double-fuzzy approach. It also requires a predefined fuzzy-metric legend. However, doing calculations with colours is more complex since we derive colours as numbers but they can create rather different psychological impression. In this case we have decided to locate our representation classes on the circumference of the colour wheel with constant intensity (see Figure 3). The derived saturation of the mixed-colour map calculated using the CM technique can give us insight into both the spatial and the taxonomic confusion of our mapping system.

Table 1. Summary comparison of Pixel Mixture and Colour Mixture techniques.

Pixel Mixture	Colour Mixture
Crisp classes, continuous model of spatial variation	Continuous classes, continuous model of spatial variation
Becomes crisp if there is no confusion or if the size of the sub-pixel is 1x1 (defuzzification)	Becomes crisp if there is no confusion
Requires selection of the degree of discretisation (sub-pixel window)	Requires design of the legend
MVs are easy to estimate by counting the number of pixels inside the sub-pixel window	MVs are difficult to estimate; we have to do GIS calculation on colours
No need to design colours, no accidental colours;	Gives insight into both confusion in geographical and taxonomic space

Operational continuous maps are yet to be developed. The examples shown here are based on a few classes and for small case study areas. Very often, the surveyors will produce tens of classes in real cases. For example: a single soil map sheet can consist from few up to several dozens of mapping units. The next step will be to test the methodology on real data sets developed as a part of large inventory projects. We imagine that users require user-friendly maps that will also show topography in the background. Some parts of the PM and CM principles could be also used in a more interactive way and they offer better insight into the classification uncertainty of discrete data. This will satisfy more advanced users.

6. REFERENCES

Brown, A. and Feringa, W., 1999. *A colour handbook for GIS users and cartographers*. International Institute for Geo-information science and Earth Observation, Enschede, pp. 93.

Brown, D.G., 1998. "Mapping historical forest types in Baraga County Michigan", USA as fuzzy sets. *Plant Ecology*, 134(1): 97-111.

Burrough, P.A., 1993. "Soil variability: a late 20th century view". *Soils and fertilizers*, 56 (May): 529-562.

Burrough, P.A., Gaans, P.F.M. and van Hootsmans, R., 1997. "Continuous classification in soil survey: spatial correlation, confusion and boundaries". *Geoderma*, 77 (Fuzzy sets in Soil systems): 115-135.

Burrough, P.A. and McDonnell, R.A., 1998. *Principles of geographical information systems*. Oxford University Press, Oxford, 327 pp.

Burrough, P.A., van Gaans, P.F.M. and MacMillan, R.A., 2000. "High-resolution landform classification using fuzzy k-means". *Fuzzy Sets and Systems*, 113: 37-52.

De Gruijter, J.J. and Boogaard, H.L., 2001. *Fuzzy sets voor zachte klassegrenzen. Toepassing op het landevaluatiesysteem BODEGA*. 346, ALTERRA, Wageningen.

De Gruijter, J.J. and McBratney, A.B., 1988. "A modified fuzzy k-means method for predictive classification". In: H.H. Bock (Editor), *Classification and Related Methods of Data Analysis*. Elsevier Science Publishers B.V., pp. 97-105.

De Gruijter, J.J. et al., 1997. "Continuous soil maps - a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models". *Geoderma* (Fuzzy sets in soil science), 77(2-4): 169-195.

Fisher, P.F., 1993. "Visualizing uncertainty in soil maps by animation". *Cartographica*, 30(2-3): 20-27.

Hansen, M.C. and Reed, B., 2000. "A comparison of the IGBP DISCover and University of Maryland 1 km global land cover products". *International Journal of Remote Sensing*, 21(6-7): 1365-1373.

Hootsmans, R.M., 1996. *Fuzzy sets and series analysis for visual decision support in spatial data exploration*. PhD Thesis, University of Utrecht, Utrecht, 167 pp.

Irvin, B.J., Ventura, S.J. and Slater, B.K., 1997. "Fuzzy and isodata classification of landform elements from digital terrain data in Pleasant Valley, Winsconsin". *Geoderma*, 77: 137-154.

Jiang, B., 1996. *Fuzzy Overlay Analysis and Visualization in Geographic Information Systems*. PhD Thesis, University of Utrecht, Utrecht, 167 pp.

Jiang, B., Brown, A. and Ormeling, F., 1995. "Visualisation of uncertainty with fuzzy color system", Cartography crossing borders: *17th ICA Conference*, Barcelona, pp. 507.

MacEachren, A.M. and Kraak, M.J., 1997. "Exploratory cartographic visualization: advancing the agenda". *Computers & Geosciences*, 23(4): 335-344.

McBratney, A.B. and Whelan, B.M., 1995. "Continuous models of soil variation for continuous soil management", In: Site specific management for agricultural systems. *Proc. conference, ASA-CSSA-SSSA*, Minneapolis, pp. 325-338.

Unit Geo Software Development, 2001. *ILWIS 3.0 Academic user's guide*, ITC, Enschede, 520 pp. http://www.itc.nl/ilwis/.

Walvoort, D.J.J. and De Gruijter, J.J., 2001. "Compositional Kriging: A Spatial Interpolation Method for Compositional Data". *Mathematical Geology*, 33(8): 951-966.