

USING PURE AND MIXED OBJECTS IN THE TRAINING OF OBJECT-BASED IMAGE CLASSIFICATIONS

H. Costa ^{a*}, G. M. Foody ^a, D. S. Boyd ^a

^a School of Geography, University of Nottingham, Nottingham NG7 2RD, UK - (lgxhag, giles.foody, doreen.boyd)@nottingham.ac.uk

KEY WORDS: Artificial neural networks, OBIA, Over-segmentation, Sample size, Under-segmentation, WorldView.

ABSTRACT:

Training of object-based land cover classifications is often performed with objects generated via image segmentation. The objects are commonly assumed to be thematically pure or excluded from training if a mixture of classes is associated with them. However, excluding mixed objects has several consequences such as reducing the size of the training data sets. In this study, it is hypothesized that mixed objects may be used in the training stage of a classification to increase the accuracy with which land cover may be mapped from remotely sensed data, with outputs evaluated in relation to a conventional analysis using only pure objects in training. WorldView-2 data covering the University Park campus of the University of Nottingham were submitted to a series of segmentation analyses in which a range of under- to over-segmentation outputs were intentionally produced. Training objects representing four classes (bare soil, impervious surfaces, vegetation, and water) were selected from the segmentation outputs, resulting in training samples of varying size and proportion of mixed objects. A single-layer artificial neural network equivalent to multinomial logistic regression and able to use both pure and mixed training units was adopted as the classifier. A visual inspection of the results shows that using mixed training objects produced land cover maps of higher quality. Furthermore, the overall and class-specific accuracy of the classifications was systematically higher when mixed training was used (e.g. up to 48% in overall accuracy). The advantage of using mixed objects in training was beneficial even when the size of the mixed training samples was equivalent to that of the pure training samples.

1. INTRODUCTION

In object-based image analysis it is common for some of the objects generated via image segmentation to be of mixed thematic composition as a result of under-segmentation error (Clinton et al. 2010). This violates the commonly made assumption of object purity that is implicit in a conventional classification analysis, and thus can be a problem throughout the analysis. For instance, mixed objects can result in degraded training statistics and act to reduce mapping accuracy (Wang et al. 2004; Gao et al. 2011; Hirata and Takahashi 2011). Therefore, mixed objects are often excluded from the training stage so that they are not used in the derivation of training statistics (Dean and Smith 2003; Smith 2010; Dronova et al. 2011; Cai and Liu 2013).

Excluding mixed objects has, however, several consequences. Using only pure objects has the drawback of representing only homogenous areas in training whereas afterwards the classifier may be confronted with some mixed objects. Thus pure objects may be not totally representative of the characteristics of the objects that a classifier has to classify. Furthermore, excluding mixed objects has the consequence that the size of the training samples will inevitably be reduced.

Deviation from the object purity assumption can, however, sometimes be made throughout a classification analysis (e.g. Foody 1999a). Specifically, impure units can be accounted for in training (Foody 1997; Zhang and Foody 2001; Eastman and Laney 2002; Matthew 2012), class allocation (Wang 1990; Foody 1996; Dronova et al. 2011), and testing stages of a supervised image classification (Foody 1995; Binaghi et al.

1999; Stehman et al. 2007). However, little research has been undertaken on the use of mixed units in training object-based image classifications.

This paper sets out to test the hypothesis that mixed objects may be used in the training stage of a classification to increase the accuracy with which land cover may be mapped from remotely sensed data, with outputs evaluated in relation to a conventional analysis using only pure objects in training.

2. MATERIALS

2.1 Study area

A square area of 2502 m per 2502 m over the University Park campus of the University of Nottingham in the United Kingdom was used to undertake the analyses (Figure 1). The central coordinates of the study area are 52°56'28.212"N and 1°11'44.3217"W. A total of four land cover classes were defined: bare soil, impervious surface, vegetation, and water.

2.2 Spectral data

A WorldView-2 (WV2) image of acquired on 26 May 2012 was used (Figure 1). This image is comprised of eight spectral wavebands (coastal, blue, green, yellow, red, red edge, NIR1, and NIR2) with a spatial and radiometric resolution of 2 m and 11 bit respectively.

* Corresponding author



Figure 1. Subset of the WorldView-2 image acquired on 26 May 2012 over the study area (University Park campus of the University of Nottingham and its environs). The circular areas of 100 m radius outlined orange are clusters randomly located used for selecting training objects. The red points locate the pixels of the random testing sample used to assess the accuracy of the classification produced using segmentation scale=50 and pure training.

2.3 Reference data

Reference data were produced to assist the production of training and testing samples. Visual interpretation of the WV2 image (including the spectral wavebands and a panchromatic band of 0.5 m spatial resolution) was the basis for populating training and testing samples with land cover class labels. Imagery available through Google Earth was also inspected.

The training samples were produced via cluster sampling (Whiteside et al., 2014) in which the primary sampling unit, the clusters, was defined to be ten circular areas of 100 m radius randomly located (Figure 1). Thus, a total of ~5% of the study area was allocated to training purposes. The land cover in the clusters was delimited by visual interpretation as explained above. The secondary sampling unit was the objects generated via image segmentation that intersected the clusters (see 3.2).

Testing samples were produced to assess the accuracy with which WV2 data was classified based on each of the segmentation outputs produced. Therefore, a total of eight samples were produced. They comprised 50 randomly selected pixels per mapped class (Figure 1 shows an example). Although the classifications were object-based, the pixel is a legitimate and practical option for accuracy assessment (Stehman and Wickham, 2011). Land cover in the testing pixels was determined by visual interpretation. Although it may be beneficial to address the potential thematic mixed nature of the pixels, only the dominant class was considered for simplicity.

3. METHODS

A series of analyses were undertaken to explore the potential of using mixed objects in training object-based classifications. The

methods used included segmentation of the WV2 image, production of training samples, classification of the segmentation outputs, and assessment of the classification accuracy.

3.1 Image segmentation

The eight spectral bands of the WV2 image were segmented to generate objects using the multiresolution algorithm implemented in GeoDMA software (Körting et al., 2013), version 0.2.1, which is based on the popular algorithm of Baatz and Schäpe (2000). This is a region-based algorithm that uses spectral and shape properties of the objects being generated, and the most influential parameter of the algorithm, and hence most often manipulated, is scale.

A series of segmentation analyses were undertaken in which the parameter scale was manipulated in order to produce a range of under- to over-segmented outputs. The value of the scale parameter was set at 30, 50, 70, and 90 while the remaining parameters were set at 50. As a result, four segmentation outputs were obtained, ranging from over-segmented results (mostly composed of small and possibly pure objects) to under-segmented results (mostly composed of large and possibly mixed objects). In this study an object was taken to be pure if the dominant class covered more than 90% of the object's area, similar to Cai and Liu (2013).

3.2 Training

Training samples were produced to classify each of the segmentation outputs generated. The training samples were formed by the objects generated with each of the parameter settings that intersected the primary sampling unit defined, the 10 random clusters (section 2.3). Therefore, the set of training objects used varied between the segmentation outputs (ranging from over-segmented to under-segmented training objects) while the same geographical area was used in training each classification.

The eight spectral bands were used to calculate training statistics, which were the mean and standard deviation of the pixel values associated with the training objects. The mean digital number provides a value of central tendency whereas standard deviation provides a value of variability (texture). As a result, 16 bands were used as discriminating variables in classifications.

The training objects were assigned reference class labels accordingly to those visually delimited in the clusters. The proportion of the area that each class occupied in a training object was calculated. The proportions calculated were 0.0 if the class was absent and 1.0 if the object was pure. Intermediate values for at least two classes were calculated when the object was of mixed class composition.

Two training strategies were followed. First, the traditional procedure of using only pure objects at the training stage was tested (i.e. the mixed objects, whose dominant class covered <90% of the object, were excluded). Second, all of the training objects were used even if the dominant class covered <90% of the area. The fractional coverage of the classes found in the objects was used as a measure of class membership, and objects were allowed multiple and partial membership.

3.3 Classification

A multinomial log-linear classifier was applied via a neural network with no hidden layer (R Core Team, 2014; Venables and Ripley, 2002) to produce land cover maps. This neural network allows the objects of mixed class composition to be used in training in a form similar to that explained in Foody (1997) for per-pixel classification. The mean and standard deviation of the objects across the WV2 spectral bands were used as discriminating variables.

Although the classifier used produces soft classifications, traditional hard land cover maps were obtained by allocating each object the label of the class with which it had the greatest membership. Each segmented output generated was thus used to produce hard land cover maps based on different training strategies: pure and mixed.

3.4 Accuracy assessment

The testing samples were used to assess the accuracy of the classifications. Confusion matrices comparing the reference labels of the testing pixels and the classification labels of the corresponding objects were constructed. Overall accuracy and per-class estimates of accuracy (user's and producer's accuracy) were calculated in terms of proportion of area correctly classified as described in Olofsson et al. (2014), including confidence intervals at the 95% confidence level.

The accuracy of the four segmentation outputs produced was also assessed. In this case, the accuracy assessment essentially aimed at determining whether the training data sets used were over-segmented, under-segmented, or balanced. The magnitude of under- and over-segmentation errors is associated with the presence of mixed objects. An empirical discrepancy method proposed by Möller et al. (2013) and slightly refined by Costa et al. (2015) was used in this study to assess segmentation accuracy. This method essentially compares the objects generated to a reference data set to measure the geometric match between them. The land cover delimited in the clusters by visual interpretation (previously used for producing the training data sets) was reused as reference data. The outcome of the method is a metric, M^g , that measures the strength and type of error. Negative M^g values indicate that under-segmentation error dominates while positive M^g values represent the opposite case in which over-segmentation error dominates. So, $M^g \sim 0$ is deemed to be indicative of optimal segmentation accuracy as the two types of error are balanced (Möller et al., 2013).

4. RESULTS

The segmentation outputs were, as expected, notably over-segmented when the scale parameter was small as the objects generated were noticeably smaller than when the scale parameter was set at a large value. Figure 2 shows the level of under/over-segmentation error of the training samples. Negative M^g values indicate that under-segmentation error dominates while positive M^g values represent the opposite case in which over-segmentation error dominates. The M^g value closest to 0 was that of parameter scale 50.

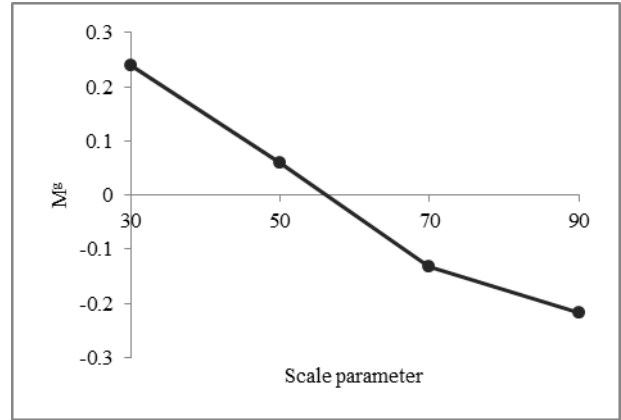


Figure 2. Image segmentation accuracy.

The difference between the classifications produced with pure and mixed training was evident with mixed training affording classification higher accuracy. Figure 3 shows the classifications of the segmentation output produced with scale=50. The estimated overall accuracy of the maps is $43.0 \pm 10.4\%$ (pure training, Figure 3a) and $91.5 \pm 5.3\%$ (mixed training, Figure 3b) respectively. Figure 4 shows the overall accuracy of all the classifications. Mixed training enabled classification to achieve higher accuracy values than pure training for all of the segmentation settings used.

The per-class estimators of accuracy were also higher when the training stage of the classifications was mixed (Figure 5). For example, the user's accuracy of class vegetation of the map of Figure 3b (produced with mixed training) was larger than that of the map shown in Figure 3a (produced with pure training), specifically $94.0 \pm 6.3\%$ and $48.0 \pm 14.0\%$ respectively.

5. DISCUSSION

The common practice of using only pure objects in training may compromise the accuracy of object-based classifications. Excluding mixed objects from training reduces the value of a training data set and thus should not be adopted. One of the advantages of allowing mixed objects to derive training statistics is that the size of the training data sets is larger than when only pure objects are used. Because in this study the mixed training strategy did not exclude mixed objects from training, the size of the mixed training data sets was larger than that of the pure training data sets (Table 1). Therefore, the difference between the accuracy values of the classifications trained with pure and mixed data sets are partly caused by different sizes of the training samples.

Scale	Pure	Mixed
30	544	767
50	278	420
70	171	262
90	130	207

Table 1. Size (number of objects) of the training samples

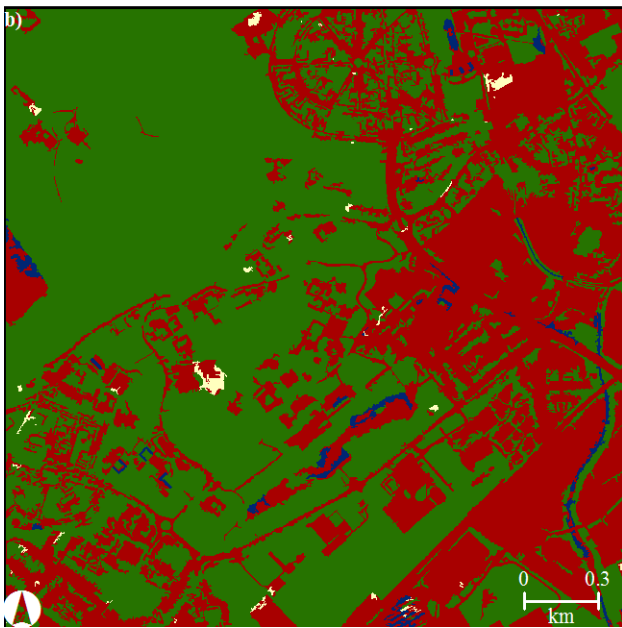
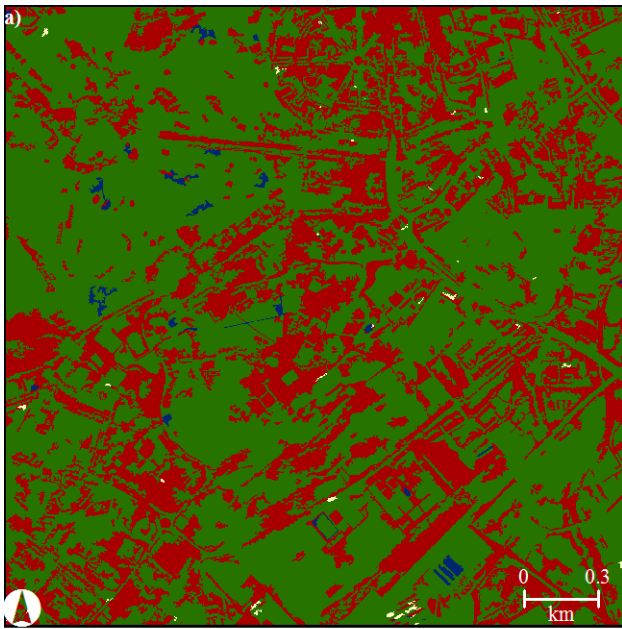


Figure 3. Segmentation output produced using scale=50 and classified using a) pure and b) mixed training. Yellow, red, green, and blue represent bare soil, impervious, vegetation, and water, respectively.

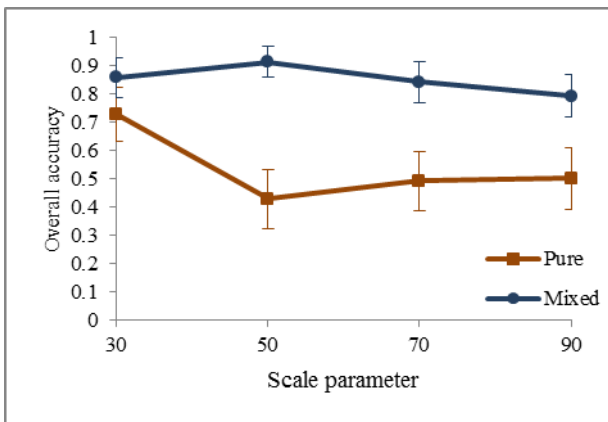


Figure 4. Accuracy of the classifications

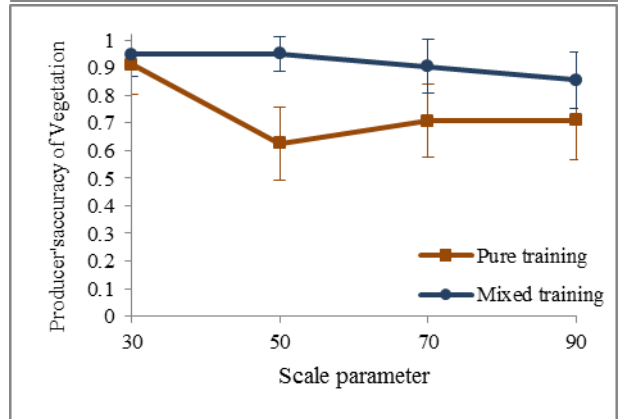
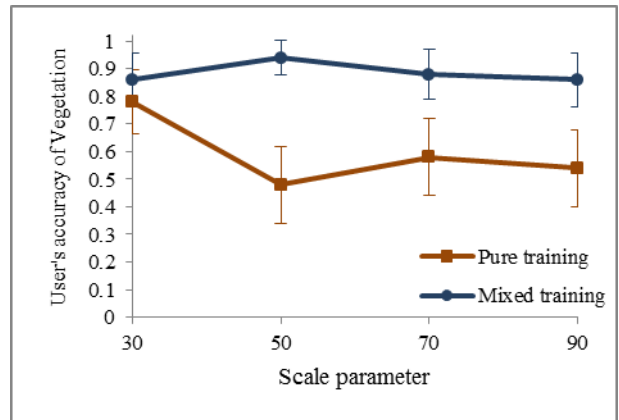


Figure 5. User's and producer's accuracy of class vegetation.

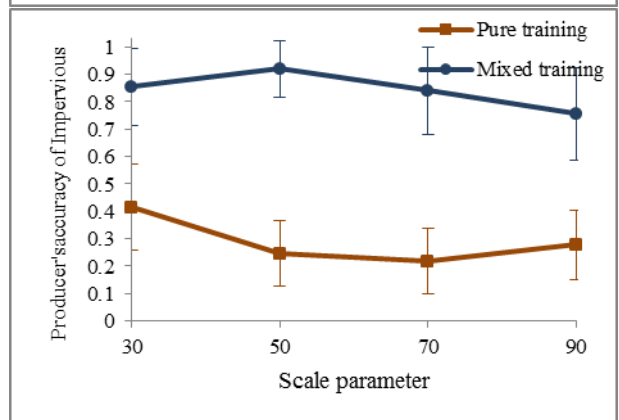
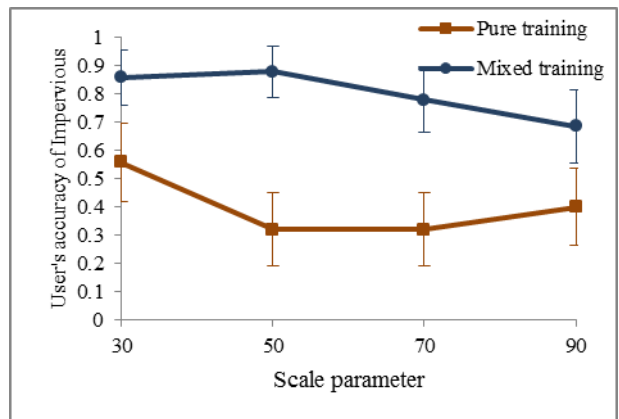


Figure 6. User's and producer's accuracy of class impervious surface.

The size of the training data sets is not, however, the only factor explaining the results. Note that the accuracy of the classification of scale=30 that used pure training (544 training objects) was lower than that of classification of scale=90 with mixed training (207 training objects). The results, thus, suggest that using pure training data sets is a drawback in that only homogenous areas are represented at the training stage whereas afterwards the classifier may be confronted with some mixed objects. Mixed training, on the contrary, informs on the occurrence of mixed objects.

To highlight the advantage of using mixed training over pure training, an additional analysis was performed without having to deal with different sizes of the training samples. The size of the mixed training data set for the segmentation output produced with scale=50 was reduced from 420 to 228 objects, the latter being the size of the corresponding pure training data set (Table 1). This was achieved by excluding randomly selected training objects, with all of the objects, pure and mixed, having the same probability of being excluded. This allowed the size of the training data set to be reduced without changing substantially the inherent ratio of pure to mixed objects. The reduced mixed training data set enabled classification accuracy to reach $84.2 \pm 7.2\%$, which is lower than the accuracy of the classification that used the entire mixed training data set, but still significantly larger than when using the pure training data set (Figure 4).

6. CONCLUSIONS

Image segmentation is a necessary stage for object-based image classification. Commonly segmentation errors produce objects of mixed class composition, which commonly are excluded from the derivation of statistics at the training stage. However, including mixed objects in training is advantageous because the size of a training data set is not reduced and they better represent the characteristics of the objects that a classifier will be confronted with for producing a map via image classification.

ACKNOWLEDGEMENTS

The WorldView-2 data used was provided by the Earth Observation Technology Cluster, a knowledge transfer initiative funded by the Natural Environment Research Council (grant NE/H003347/1).

REFERENCES

- Baatz, M., Schäpe, A., 2000. Multiresolution Segmentation: an optimization approach for high quality multi-scale image segmentation, in: Strobl, J., Blaschke, T., Griesebner, G. (Eds.), *Angewandte Geographische Informationsverarbeitung XII. Beiträge zum AGIT-Symposium Salzburg 2000*. Herbert Wichmann Verlag, Heidelberg, Germany, pp. 12-23.
- Costa, H., Foody, G.M., Boyd, D.S., 2015. Integrating user needs on misclassification error sensitivity into image segmentation quality assessment. *Photogrammetric Engineering and Remote Sensing*, 81(6), pp. 451-459.
- Foody, G.M., 1997. Fully fuzzy supervised classification of land cover from remotely sensed imagery with an artificial neural network. *Neural Computing & Applications*, 5(4), pp. 238-247.
- Körting, T.S., Fonseca, L.M.G., Câmara, G., 2013. GeoDMA—Geographic Data Mining Analyst. *Computers & Geosciences*, 57, pp. 133-145.
- Möller, M., Birger, J., Gidudu, A., Gläßer, C., 2013. A framework for the geometric accuracy assessment of classified objects. *International Journal of Remote Sensing*, 34(24), pp. 8685-8698.
- Olofsson, P., Foody, G.M., Herold, M., Stehman, S.V., Woodcock, C.E., Wulder, M.A., 2014. Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, 148, pp. 42-57.
- R Core Team, 2014. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Stehman, S.V., Wickham, J.D., 2011. Pixels, blocks of pixels, and polygons: Choosing a spatial unit for thematic accuracy assessment. *Remote Sensing of Environment*, 115, pp. 3044-3055.
- Venables, W.N., Ripley, B.D., 2002. *Modern applied statistics with S*. Fourth Edition ed. Springer, New York.
- Whiteside, T.G., Maier, S.W., Boggs, G.S., 2014. Area-based and location-based validation of classified image objects. *International Journal of Applied Earth Observation and Geoinformation*, 28, pp. 117-130.