COMPARISON OF PIXEL- AND OBJECT-BASED SAMPLING STRATEGIES FOR THEMATIC ACCURACY ASSESSMENT

Julien Radoux, Pierre Defourny and Patrick Bogaert

Department of Environmetrics and Geomatics
Université catholique de Louvain
Croix du Sud, 2/16, B-1348 Belgium
(Julien.Radoux,Pierre.Defourny,Patrick.Bogaert)@uclouvain.be

KEY WORDS: Quality, object-based, pixel-based, confusion matrix, samples

ABSTRACT:

Confusion matrix and derived global indices (kappa, overall accuracy, producer accuracy) are widely accepted as a standard method for the accuracy assessment of land use/land cover maps. In order to build the confusion matrix, the ground truth labels of samples are crossed with the map labels. Most of the time, the sampling strategies are simply based on the spatial distribution of sample points (systematic, random or stratified random sampling). They do therefore not account for existing objects in the map. In this study, an object-based sampling strategy is compared with classical (pixel-based) sampling. Instead of selecting pixels on the basis of their coordinates, a random sampling was performed in the list of objects and the sampling probability was proportional to the object size. The central point of selected objects was used to create the confusion matrix, which differs from the standard confusion matrix by the fact that the weight of each object was on average proportional to its area.The performance of the two sampling strategies was quantified thanks to two sets of simulation : map alterations and sampling repetitions. An existing land cover map was regarded as ground truth and was spatially and/or thematically altered. Errors in the test maps were generated by blundering boundaries and/or changing object labels, respectively. The bias (difference between average estimate and the truth) and the variance of the overall accuracy estimates were then measured as an indicator of the robustness of the confusion matrix. Pixel- and object-based sampling did not lead to the same measure of accuracy. The former evaluated the global accuracy (influenced by boundary errors) while the latter measured the thematic accuracy (only influenced by labeling errors). The other big difference between the two sampling strategies was a smaller confidence interval on the accuracy estimates with the object-based strategy. For a given confidence interval, object-based strategy could thus reduce the sampling effort. The proposed object-based sampling strategy was easy to implement and could help to reduce the costs of map validation. Further work is needed to determine *a priori* the number of object samples necessary to fulfill a given level of confidence.

1 INTRODUCTION

Land use/land cover maps are of paramount importance in various applications such as change monitoring, land use planning, hydrological modelling or natural ressource management. Thanks to remote sensing, those maps are now produced from local to global scales. In order to use these products in an appropriate way, map users need reliable data quality information. Object-based image analyses yields two types of products : detection of a specific land cover class (for instance, crop field delineation) or classification of the whole image. In the first case, the quality assessment consists in comparing delineated objects with reference objects. In the second case, the thematic map content should be subject to a statistically robust thematic accuracy assessment. According to (Stehman and Czaplewski, 1998), the three basic components of a thematic accuracy assessment are: 1) the sampling scheme and sampling unit used to select reference samples; 2) the response design used to obtain the reference land-cover classification for each sampling unit and 3) the estimation and the analysis procedure.

The sampling unit is the link between a spatial location on the map and the corresponding spatial location on the reference. The commonly used sampling unit is the pixel, which is also the most appropriate when the pixels are classified independently (Franklin et al., 1991). Pixel blocks are more likely to be used when post-classification filters (smoothing or morphology) are applied on the classification results. Finally, polygons are used in studies where reference information was particularly difficult to obtain (George, 1986, Warren et al., 1990). Other strategies can also be used to reduce the sampling effort. Cluster sampling is a typical way to reduce the travel cost, but in this case the spatial correlation between samples of the same cluster must be accounted for.

The analysis of the validation results is usually based on the confusion matrix matching, for a large number of samples, classification result to reference information (Congalton, 1991). The confusion matrix is often summarized by global indices, such as the overall accuracy index or Cohen's Kappa (Foody, 2002, Congalton, 1991, Stehman, 1997). The former is the proportion of agreement between a map and the truth. The latter is the proportion of agreement corrected by chance. While kappa can be used to compare the performance of classification algorithms, overall accuracy provides more meaningful information to end users (Stehman, 1997).

The response design is often the main source of error in quality control (Congalton and Green, 1993). Most of the time, it is either based on field survey or on more precise remote sensing data, the latter being less accurate due to interpreters errors but also less expensive. The error risk can be reduced using appropriate protocols for the validation crew or by repeating the interpretation and merging the results. On the other hand, it was shown that the sampling strategy could strongly influence the results of the accuracy assessment. Simple probabilistic sampling was indeed shown to be sensitive to the planimetric precision of the map and to the fragmentation of the landscape. Planimetric errors can be reduced by considering sample location away from object boundaries (Warren et al., 1990, Wulder et al., 2006). With pixel-based classification, the use of homogeneity contraints on sampling unit location (Plourde and Congalton, 2003, e.g.) leads to optimistic bias (Hammond and Verbyla, 1996) because the probability to sample edge pixels becomes null while they are proned to more frequent errors (Stehman and Czaplewski, 1998, Powell et al., 2004). However, this is not the case with objects because they are intrinsically homogeneous so that the location inside each object can be arbitraly chosen (George, 1986).

In this study, an object-based validation for land-use/land cover (LULC) maps produced by object-based classification is proposed and discussed. The method is compared with point-based validation with regard to the true error measured from simulated maps. First, we look at the effect of systematic boundary errors on the overall thematic accuracy estimates. Second, we assess the efficiency of the estimates in a case without planimetric errors.

2 METHOD

The diversity of existing accuracy assessment methods clearly shows that there is no single best method due to different user needs and producer constraints. The proposed method meets the two following requirements:

- To provide a reliable and cost-effective estimate of the overall accuracy. In other words, the validation scheme should provide an unbiased estimate of the overall accuracy with a small variance of prediction.
- To evaluate the thematic accuracy of the map. In other words, the overall accuracy should not be sensitive to registration and delineation errors on the map nor on unprecise sample positions.

2.1 Sampling unit and response design

Response design and sampling units should be tightly related to a consistent LULC typology and a few concepts need to be clarified beforehand.

For the multinational Africover project, Food and Agriculture Organization (FAO) has developed a conceptual framework to define in a flexible but standard way any land cover typology suited to the local, national or global needs. This Land Cover Classification System (LCCS) (Di Gregorio and Jansen, 2000) was also selected for the Global Land Cover 2000 initiative (Fritz et al., 2003) and for the ESA 2005 Globcover product (Defourny et al., 2006). It becomes more and more popular because of its efficiency for class definition and is well adapted to object-based classification and multi-scale processing (Gamanya et al., 2007). The visual interpretation is naturally a qualitative multiscale approach. Object-based classifications are convenient for a multiscale approach but in a quantitative and fully-documented way.

The combination of the multiscale segmentation and object-based classification process actually requires the distinction of three concepts: the Elementary Processing Unit (EPU), the Smallest Legend Unit (SLU) and the well-known Minimum Mapping Unit (MMU).

The EPU corresponds to the smallest delineated object that can be classified. Theoretically, it can be as small as a pixel, but it is most often made of several pixels. Its size depends on the segmentation algorithm and its parameters, as well as on the image structure and local contrasts. Though, parametric segmentation algorithm allows the user to infer the mean object size but seldom the EPU. Post-processing algorithms, such as morphological filters, are therefore needed when the classification process requires a minimum number of pixels for statistical consistency.

The SLU is the minimum size of object for labeling. It is a thematic constraint that should be defined prior to the classification process depending on the LULC typology. The SLU can be as small as the EPU in the case of simple legend but is often significantly larger. It then corresponds to objects produced either by an upper level of segmentation or by the a posteriori aggregation of smaller objects. These larger objects can consist in i) parent classes in a hierachical classification, ii) composite classes arising from contextually meaningful object combination or iii) mosaic classes including a mixture of unrelated classes. For this last case, the respective proportion of the smaller objects discriminated by the classification process enables the user to precisely label the SLU according the LULC mosaic. Furthermore, an *a posteriori* distribution analysis of the elements belonging to a given class could document the composition and the internal spatial pattern of this class.

The well-known MMU is the smallest element to be represented on the mapping output and must be set according to cartographic standards. The MMU is of course larger or equal to the SLU. However, the increasing use of numeric vector databases reduces the interest in a crisp MMU. MMU should be viewed more as a display constraint for paper maps or WebGIS application while EPU and SLU are directly linked to the database.

These concepts defined in the context of multiscale segmentation and object-classification clearly fit in the LCCS framework. Indeed, the LCCS drives the user to combine well-defined elementary descriptors to explicitly define each land cover class from several elements mixed in a given proportion. For instance, broadleaved deciduous trees higher than 5 m with their canopy covering at least 40 % of the ground surface typically describes a "Closed deciduous broadleaved forest". Besides the local name given by the user to the class, a comprehensive LCCS code documents the composition of elements included in the class and their respective quantitative proportion. LCCS rules can also define a "cartographic mixed unit" where two specific elements are present, for instance a Mosaic class of cropland (50-70 %) and natural and semi-natural vegetation (grassland) (20-50 %). In both cases, the class is quantitatively described by its components thus allowing aggregating them in a different way to make them compatible with other land cover typology.

This kind of precise typology is a first step toward a robust response design. As the Legend Unit may include complex and heterogeneous classes which are difficult to assess on the field, the Processing Unit is more appropriate because these elementary objects can be assumed to be homogeneous at this level. With this assumption in mind, any location within the object can be arbitrarly chosen, and the object center should be preferred as it helps avoiding edge effects. Moreover, the validation crew should be aware of the EPU in order to ignore too small gaps in the object.

2.2 Sample selection and result analysis

As a matter of fact, a sample of objects (i.e. groups of pixels) covers a larger area than a sample of pixels with the same number of elements. However, the contribution of each sample in the confusion matrix must be adjusted in order to avoid an inaccurate estimate of the overall accuracy. To our knowledge, none of the previous studies using polygons as sample unit addressed this issue.

The analytical solution for adjusting sample weights is complex because it depends on the distribution of the size of the objects. While it is trivial if all objects have the same area (in which case there is no adjustment needed), it is necessary to account for the fact that misclassifying a large object has more impact on the overal accuracy than misclassifying a small object. On the other hand, weigthing each sampled object based on its area yields an overall accuracy estimate that depends on the distribution of size of the misclassified objects. On average, this method may thus over- or underestimate the true overall accuracy, contrary to the point-based simple probabilistic sampling, which is always unbiased.

The proposed method is to build an object-based sampling in accordance with the simple point-based probabilistic sampling under the hypothesis that objects are homogeneous at the scale of the map. The sample of objects is created by iteratively selecting objects with a drawing probability directly proportional to their area. The iteration ends when the user-defined sample size, S, is reached. The number of times each object i is selected, n_i, is kept in memory and is used as the weight for each corresponding object in the confusion matrix. In this sense, the object-based validation is thus close to a point-based simple random sampling of $\sum_{i=1}^{S} n_i \, (\geq S)$ sampling units.

3 CASE STUDY

Test maps were simulated for the quantitative assessment of the different sampling strategies. In order to make these simulations as realistic as possible, all these maps were derived from a real LULC map which was taken as reference. The map template is a subset of the CORINE Land Cover map in Southern Belgium, which was produced by photointerpretation. Figure 1 shows the actual map, which includes 34 different LULC classes, with a simplified symbology. Despite this variety of LULCtypes, it is worth noting that more than 80 percents of the total map area are covered by only 7 land cover classes.

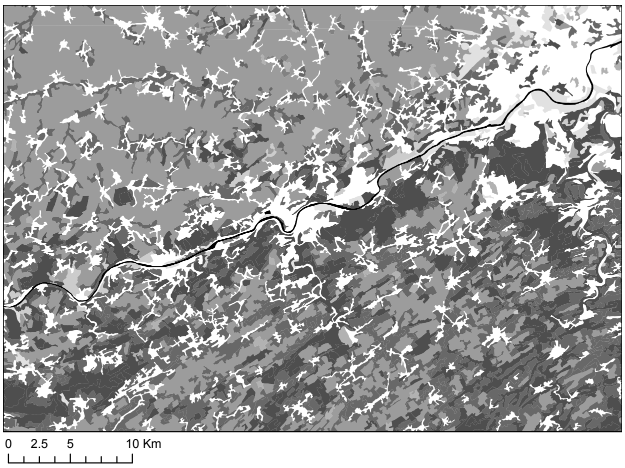

Figure 1: Subset of CORINE Land Cover used as reference

Let us assume that the test maps to be validated were produced by classification of homogeneous object. In other words, the land cover enclosed by each object is unique. Test maps were simulated by introducing thematic (bad object labeling) and planimetric errors (bad object delineation) into the reference template. On one side, thematic errors were simulated by changing the label of a random selection of objects so that a given percentage of the total map area received bad labels. These selections were repeated 10 times for each category (around 5 and around 25 percent of thematic errors) and the exact thematic accuracy was measured by dividing the area of correctly labeled object by the total map area. On the other side, the simulation of planimetric errors were based on combination of accuracy and precision errors. A systematic planimetric error was obtained after shifting the database by 0, 10 and 20 m to mimic bad registration. The edge quality was reduced by converting the map into a raster of lower resolution (1, 5, 10, 20, 30 and 40 m) to mimic image spatial resolution and object delineation errors, that is the precision of the edges. The combination of these planimetric disturbances yielded 18 test maps which were crossed with the reference map to measure the exact overall accuracy. These maps were also divided into 16 smaller maps to assess the effect of landscape structure. The combined effect of landscape structure and edge precision on the thematic accuracy was evaluated thanks to the proportion of the map covered by a buffer on the map boundaries with a buffer distance proportional to their planimetric precision.

The number of samples was chosen based on equation 1(Plourde and Congalton, 2003) for a simple point-based probabilistic sampling. For the sake of comparison, the sample size of the object-based validation was identical to the point based validation. Each sampling was reproduced 10 times so as to measure the bias and standard deviation of the overall accuracy estimate. In the present case study, a confidence of 95 % ($b = 0.05$) required 778 points in the worst case (maximum number of points). These points were selected on the map thanks to "Hawth's tools" (Beyer, 2004).

$$n = max_i \left\{ B \, P_i \, (1 - P_i) \, / b^2 \right\}, \tag{1}$$

where P_i is the proportion of the class i.

4 RESULTS

In this section, pixel- and object-based validations are compared in terms of conservative bias (average difference between overall accuracy estimates and true thematic accuracy) due to planimetric errors and variance of prediction computed *a posteriori*.

The effect of planimetric errors was negligible for object-based validation due to the size of objects with respect to the magnitude of the planimetric errors,. For the pixel-based validation, however, the conservative bias was larger than 5 percents in the worst case. Figure 2 shows that even a small systematic position shift between the reference data and the test map could lead to underestimate the thematic accuracy. Moreover, in the absence of shift, there is a strong linear relationship between the alteration of the edge planimetric precision and the bias. Eventually, the effects of the shift stacked with the effect of the edge blundering, but only when the amplitude of the latter was larger than the former.

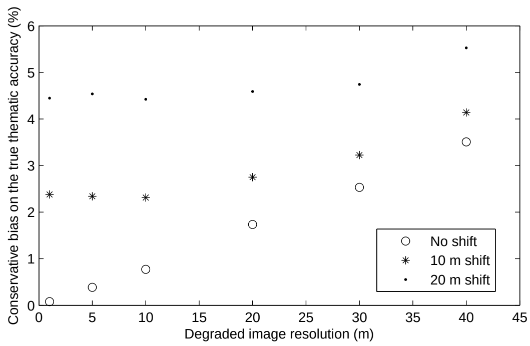

Figure 2: Combined effects of map shift and boundary alteration on the bias of the point-based thematic accuracy estimate

It is further shown that the planimetric errors due to the precision of edge delineation depend on the fragmentation of the landscape. It is indeed directly proportional ($R^2 = 0.99$) to the percentage of the map covered with the buffer area around the boundaries, and independent of the number of sampling points. When the buffer distance is equal to the edge blundering amplitude (the resolution of the degraded raster in our case), the sampling errors can be estimated with a 0.02 % RMSE by taking $1/8$ of the buffer area percentage as a rule of thumb (fig 3).

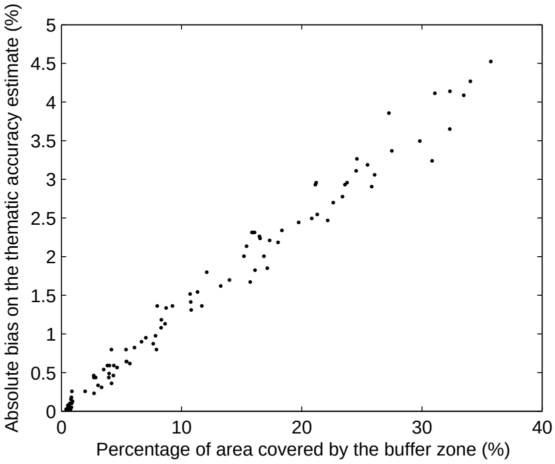

Figure 3: Relationship between conservative bias due to planimetric errors and map fragmentation

On the test maps without planimetric errors, both point-based and object-based validation provided unbiased estimates of the thematic accuracy. However, there was a better confidence interval on the errors with object-based validation scheme than with the point-based sampling, as shown by the smaller standard deviation on figure 4.

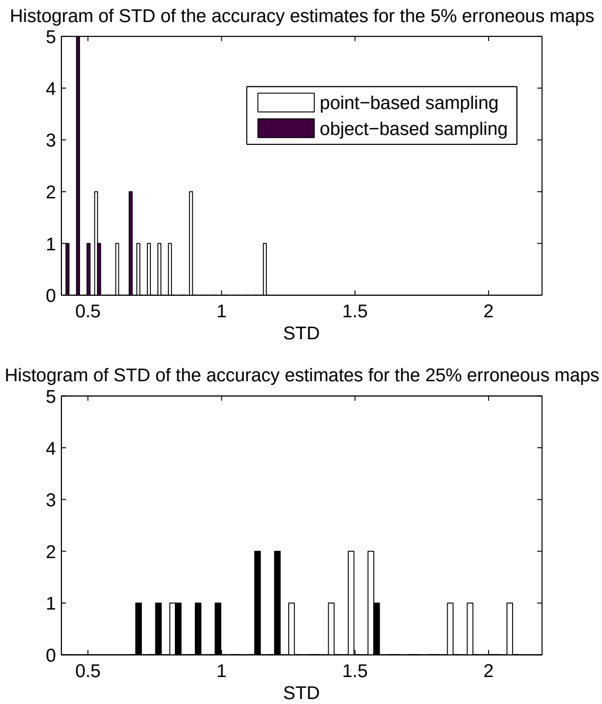

Figure 4: Standard deviation of pixel-based and object-based samplings (n=10)

5 DISCUSSION

As a matter of fact, there is no single definition of map quality because maps can be used for different purposes. In this study, map accuracy was decomposed in planimetric and thematic components. It was shown that probabilistic point-based sampling evaluates the global map accuracy while object-based sampling evaluates the thematic accuracy. Planimetric errors for the edges are linked to many scale related effects such as pixel size, orthorectification accuracy, segmentation algorithm, edge post-processing and edge generalisation e.g. (Congalton and Green, 1993). In other cases, the line representation of interface between different LULC types is inappropriate e.g. (Ranson et al., 2004) because it corresponds to a gradual land-cover change (ecotone). On the other hand, the precision of the ground truth samples can also be an issue as poorly located sample points lead to the same error than poorly delineated boundaries. These kind of global errors do not affect the topology of the map nor the average area estimates. Nevertheless, systematic class specific edge errors do influence the area estimates, so that edge quality assessment can bring useful complementary information in some cases (Radoux and Defourny, 2007). Unfortunately, information on edge quality is costly so that a global quality control could be advantageous when edge quality is difficult to evaluate.

The lower variance obtained with object-based sampling demonstrates its usefulness as it means that the sampling effort (number of samples) can be reduced. However, object-based validation has its drawbacks which could compensate the gain from sampling effort reduction. First, the extent of image-objects may not correspond to the extent of real world objects. This is not a problem in the case of over-segmentation as the land cover can then be unambiguously identified, but it is a real issue in the case of under-segmentation. When under-segmentation cannot be avoided, a fuzzy validation scheme is necessary, which increases the cost of the analysis and the cost per sample point. Second, object-based validation is more sensitive to labelling errors than point-based validation. A clear and well understood definition of each land cover class is therefore of paramount importance and the labelling of each sample point could suffer from this. Finally, there is no simple analytical solution to calculate the number of samples necessary to fullfill a given precision of accuracy assesment. This may increase the cost of the planning as map simulations are needed to estimate the variance.

6 CONCLUSIONS

Object-based validation can provide accurate and precise estimate of the confusion matrix in order to assess the thematic quality of a land use/land cover maps. Contrary to pixel-based validation, boundary errors can be consequently reduced without affecting the sample representativity. However, object-based quality assessment is more sensitive to labelling errors than pixel-based accuracy. The increased cost for the validation of a single object may thus compensate the gain from the reduced number of objects needed because of the lower prediction variance.

The proposed method proved to be better than pixel-based sampling in terms of variance of the estimate. However, further studies are necessary to provide a generic analytical solution of the prediction variance as it already exists for pixel-based validation.

ACKNOWLEDGEMENTS

Authors thank the Belgian Scientific Policy ORFEO program for funding.

REFERENCES

Beyer, H., 2004. Hawth's analysis tools for arcgis. Available at http://www.spatialecology.com/htools.

Congalton, R., 1991. A review of assessing the accuracy of classification of remotely sensed data. Remote Sensing of Environment 37(1), pp. 35–46.

Congalton, R. and Green, K., 1993. A practical look at the sources of confusion in error matrix generation. Photogrammetric engineering and remote sensing 59(5), pp. 641–644.

Defourny, P., Vancutsem, C., Bicheron, P., Brockmann, C., Nino, F., Schouten, L. and Leroy, M., 2006. Globcover : a 300 m global land cover product for 2005 using envisat meris time series. Proceedings of ISPRS Commission VII Symposium: Remote Sensing from Pixels to Processes, Enschede, The Netherlands pp. 8–11.

Foody, G., 2002. Status of land cover classification accuracy assessment. Remote Sensing of Environment 80, pp. 185–201.

Franklin, S., Peddle, D., Wilson, B. and Blodgett, C., 1991. Pixel sampling of remotely sensed digital imagery. Comput. Geosci. 17(6), pp. 759–775.

Fritz, S., Bartholomé, E., Belward, A. and et al, 2003. Harmonisation, mosaicing and production of the global land cover 2000 database. Official Publication of European Communities EUR20848, Joint Research Center (Ispra, Italy).

Gamanya, R., De Maeyer, P. and De Dapper, M., 2007. An automated satellite image classification design using object-oriented segmentation algorithms: A move towards standardization. Expert Systems with Applications 32, pp. 616–624.

George, T., 1986. Aerial verification of polygonal ressource maps: a low-cost approach to accuracy assessment. Photogrammetric engineering and remote sensing 52(6), pp. 839–848.

Hammond, T. and Verbyla, D., 1996. Optimistic bias in classification accuracy assessment. International Journal of Remote Sensing 7(6), pp. 1261–1266.

Di Gregorio, A. and Jansen, L., 2000. Land cover classification system (lccs): Classification concepts and user manual. GCP/RAF/287/ITA Africover-East Africa Project and Soil Resources, Management and Conservation Service, Food and Agriculture Organization.

Plourde, L. and Congalton, R., 2003. Sampling method and sample placement: how do they affect the accuracy of remotely sensed maps? Photogrammetric Engineering and Remote Sensing 69, pp. 289–297.

Powell, R., Matzke, N., de Souza Jr., C., Clark, M., Numata, I., Hess, L. and Roberts, D., 2004. Sources of error in accuracy assessment of thematic land-cover maps in the brazilian amazon. Remote sensing of Environment 90, pp. 221–234.

Radoux, J. and Defourny, P., 2007. A quantitative assessment of boundaries in automated forest stand delineation using very high resolution imagery. Remote Sensing of Environment 110, pp. 468–475.

Ranson, K., Sun, G., Kharuk, V. and Kovacs, K., 2004. Assessing tundrataiga boundary with multi-sensor satellite data. Remote Sensing of Environment 93, pp. 283–295.

Stehman, S., 1997. Selecting and interpreting measures of thematic classification accuracy. Remote Sensing of Environment 62, pp. 77–89.

Stehman, S. and Czaplewski, R., 1998. Design and analysis for thematic map accuracy assessment: fundamental principles. Remote Sensing of Environment 64, pp. 331–344.

Warren, S., Johnson, M., Goran, W. and Victor, E., 1990. An automated, objective procedure for selecting selecting representative field sample size. Photogrammetric engineering and remote sensing 56(3), pp. 333–336.

Wulder, M., White, J., Luther, J., Strickland, G., Remmel, T. and Mitchell, S., 2006. Use of vector polygon for the accuracy assessment of pixel-based land cover maps. Canadian journal of remote sensing 32(3), pp. 268–279.