Remote Sensing Letters

Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies

Alexis J. Comber [a]

[a] Department of Geography, University of Leicester, Leicester, LE1 7RH, UK

PLEASE SCROLL DOWN FOR ARTICLE

Geographically weighted methods for estimating local surfaces of overall, user and producer accuracies

ALEXIS J. COMBER*

Department of Geography, University of Leicester, Leicester LE1 7RH, UK

The confusion matrix is the standard way for reporting the accuracy of land cover and other information classified from remote-sensing imagery. This letter describes a geographically weighted method for generating spatially distributed measures of accuracy (overall, user and producer accuracies) from a logistic geographically weighted regression. A kernel-based approach defines the data and weights that are used to calculate the accuracies at each location in the study area. The results compare the global accuracy measures from a standard confusion matrix with those that have been allowed to vary locally. Maps of spatially varying user and producer accuracies describe the spatial autocorrelation of error. The use of geographically weighted models in the context of land-cover accuracy is discussed and suggested as a generic approach for examining how and where error processes vary.

1. Introduction

The confusion matrix and the measures it supports, such as user accuracy and producer accuracy, have become the accepted paradigm for reporting accuracy in remote sensing and land cover. However, the information provided by the confusion matrix is aspatial – it does not tell potential users where the data are better or worse and does not describe the spatial distribution of errors (Canters 1997, Steele *et al*. 1998, McGwire and Fisher 2001, Foody 2002, 2005, 2007). The tendency for errors in land-cover data to be spatially autocorrelated is well known (Campbell 1981, Congalton 1988, Canters 1997, Foody 2002, 2005, 2007, Fotheringham *et al*. 2002, Friedl *et al*. 2002, Chen and Wei 2009, Gonzalez *et al*. 2010, Comber *et al*. 2012). To overcome this, some research has developed methods for estimating and describing the spatial distribution of land-cover accuracy. Of particular relevance to the methods proposed here are the works of Foody (2005) and Comber *et al*. (2012). Foody developed geographically distributed measures of accuracy, calculating confusion matrices at a series of locations and then interpolating between them to create surfaces of error. Comber *et al*. (2012) extended Foody's work and proposed a geographically weighted measure of portmanteau accuracy for Boolean land cover and a geographically weighted difference measures for fuzzy land cover.

This letter outlines geographically weighted methods for generating spatially varying measures of land-cover accuracies. Geographically weighted approaches use a kernel or moving window to make local calculations using the data points that are

---

*Email: ajc36@le.ac.uk

under the kernel, but whose contribution to the calculation is weighted by their distance to the kernel centre. The methods extend Foody's (2005) work by incorporating distance weighting and extend Comber *et al*.'s (2012) work by developing geographically weighted estimates of the spatial distribution of user, producer and overall accuracies.

2. Methods

2.1 *Data*

Système Pour l'Observation de la Terre (SPOT) 5 imagery from 2009 for area in the northwestern part of Libya around Tripoli was resampled to 30 m × 30 m as part of a wider study. It was classified into six classes using the maximum-likelihood classifier in the IDRISI Maxlike module (Eastman 2009): Urban, Woodland, Vegetation, Grazing Land and Bare. A field survey in 2010 collected validation data at 210 locations selected by stratified random sample that divided the study areas into blocks, from which an average of 10 locations per block was selected randomly. The data set used in the analysis contained data from the field survey ('reference' data) and from the remote-sensing analysis ('classified data') at each of the 210 locations. The study area and the locations of the sample points are shown in figure 1 along with the grid used in the geographically weighted analyses (described below). Table 1 shows the confusion matrix arising from the validation. Overall accuracy is calculated from the diagonal and off-diagonal elements in the confusion matrix. User and producer accuracies are calculated using the row and column totals. Overall accuracy describes the proportion of the total number of pixels that have the same class in the reference and
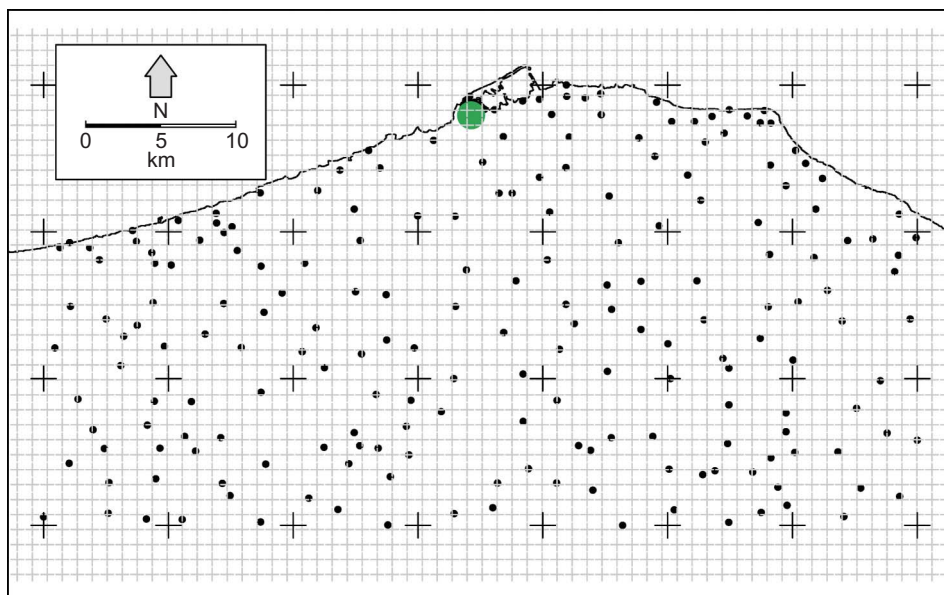

Figure 1. The study area in Libya with the validation locations (●), the Tripoli area (shaded), the 1 km grid in grey over which the kernel was passed and the larger grid was used to stratify the random sample of validation locations.

Table 1. The confusion matrix comparing reference and classified land-cover classes, showing 60% overall agreement.

Classified	Reference					Total	User
	Bare	Grazing land	Urban	Vegetation	Woodland		
Bare	18	8	7	2	4	39	0.46
Grazing land	3	23	3	8	6	43	0.53
Urban	0	0	27	1	2	30	0.90
Vegetation	0	4	7	31	5	47	0.66
Woodland	0	4	2	18	27	51	0.53
Total	21	39	46	60	44	210	
Producer	0.86	0.59	0.59	0.52	0.61		

classified data for all classes. User and producer accuracies describe the errors related to individual classes.

2.2 *Aspatial analysis*

It is informative to examine how the probabilities related to these accuracy measures can also be estimated using ordinary logistic regressions. First, a logit function needs to be defined to transform any value, q:

$$\text{logit}(q) = \frac{\exp(q)}{1 + \exp(q)} \tag{1}$$

Overall accuracy, A_O, can be estimated using a reduced logistic regression model that only contains an intercept term, b (i.e. it seeks to estimate the overall accuracy using an explanatory variable of 1):

$$A_O = \text{logit}(b) \tag{2}$$

which returns an estimate of the probability of A_O being equal to 1. When equation (2) is used to analyse the confusion matrix data in binary form, where the 126 data points summarized by the diagonal elements are indicated as 1 and the 84 data points summarized by the off-diagonal elements are indicated as 0, the result is an estimate of the overall accuracy of 0.60.

Now consider a subset of the data for the class of Grazing Land. It is represented by two similarly ordered binary vectors of 210 elements, with the reference vector having 39 elements scored as by 1, the classified vector having 43, with 23 data elements scored as 1 in common. User accuracy can be estimated using a logistic regression to analyse the reference data against the classified data in the following way:

$$P(y = 1) = \text{logit}(b_0 + b_1 x_1) \tag{3}$$

where $P(y = 1)$ is the probability that the reference land-cover class, y, is correctly predicted by the classified data, x_1, b_0 is the intercept term and b_1 is the slope. The result describes the probability that the reference data equal 1, given that the classified data equal 1 (i.e. the user accuracy), which in this case is equal to 0.53.

The producer accuracy is estimated by inverting the response and explanatory variables in equation (3):

$$P(x = 1) = \text{logit}(b_0 + b_1 y_1) \tag{4}$$

where $P(x = 1)$ is the probability that the classified land-cover class is correctly predicted by the reference data, y_1, b_0 is the intercept term and b_1 is the slope. Using the ordered binary vectors described above, the producer accuracy for Grazing Land is equal to 0.59.

2.3 *Geographically weighted analysis*

Having outlined the probabilities related to the principal accuracies arising from the confusion matrix in terms of logistic regression, it is now possible to consider the geographically weighted extensions to these models (Brunsdon *et al*. 1996). They are similar in form to ordinary regression, but geographically weighted approaches use a moving window or kernel under which local models such as logistic regression are computed at locations throughout the study area, in this case the 1 km grid shown in figure 1.

The aspatial models described in equations (2)–(4) were extended to geographically weighted models in the following way:

$$P(A_O = 1) = \text{logit}\left(b_{0(u_i, v_i)}\right) \tag{5}$$

$$P(y_i = 1) = \text{logit}\left(b_{0(u_i, v_i)} + b_1 x_{1(u_i, v_i)}\right) \tag{6}$$

$$P(x_i = 1) = \text{logit}\left(b_{0(u_i, v_i)} + b_1 y_{1(u_i, v_i)}\right) \tag{7}$$

where $P(A_O = 1)$ is the overall accuracy probability at location i, x_1 and y_1 are the classified and reference data, respectively, and (u_i, v_i) is a vector of two-dimensional coordinates describing the location of i over which the coefficient estimates are assumed to vary.

In contrast to ordinary regression models where the processes under investigation are assumed to be stationary (i.e. location invariant), local models are spatial disaggregations of global models, the results of which are location specific. A moving window allows a local regression to be calculated at each location and data points that are further away from the location under consideration contribute less to the solution. Thus, the kernel defines the data and weights that are used to calibrate the model at each location. The weight, w_i, associated with each location (u_i, v_i) is a decreasing function of d_i, the distance from the centre of the kernel to (u_i, v_i):

$$w_i = \begin{cases} \left(1 - \frac{d_i^2}{h^2}\right)^2 & \text{if } d_i < h, \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

where h is the *bandwidth* and the weights for each data point, i, change for each location. Here, the bandwidth was selected to include the nearest 15% of the data points.

3. Results

The geographically weighted logistic regressions described in equations (5)–(7) were applied to the land-cover class of Grazing Land over the 1-km grid illustrated in figure 1. Table 2 summarizes the distribution of overall accuracy and user and producer accuracies for Grazing Land, as calculated at each grid cell. Overall accuracy varies from 0.57 to 0.65 in different parts of the study area compared to a global figure of 0.60. User accuracy varies from 0.13 to 0.71 compared to a global figure of 0.53 and producer accuracy varies from 0.52 to 0.72 across the study area compared to a global figure of 0.59. It is interesting to note that there is much greater variation in the distribution of user accuracy estimates, moderate variation in the distribution of producer accuracy estimates and less variation in the distribution of estimates of overall accuracy.

The spatial variations in user and producer accuracies are shown in figures 2 and 3. Figure 2 indicates that the user accuracy for Grazing Land is generally higher to the south of the study area and lower towards the north (around Tripoli). User accuracy provides an estimate of the probability that a classified pixel correctly represents the features on the ground. It indicates the errors of commission (inclusion) and for the

Table 2. A summary of the variation in geographically weighted overall accuracy and Grazing Land user and producer accuracies.

Accuracy	Min.	First quartile	Median	Mean	Third quartile	Max.
Overall	0.572	0.596	0.606	0.605	0.610	0.645
User	0.134	0.479	0.575	0.535	0.621	0.709
Producer	0.520	0.573	0.608	0.608	0.639	0.720

Figure 2. Spatial variation in user accuracy for the class of Grazing Land. Crosses indicate locations where Grazing Land was recorded in the reference data and circles indicate locations where Grazing Land was recorded in the classified data.
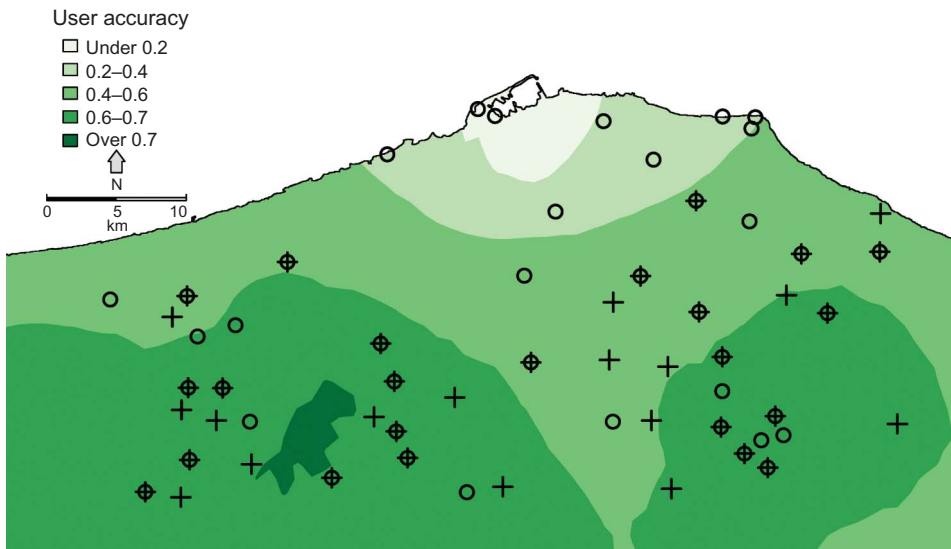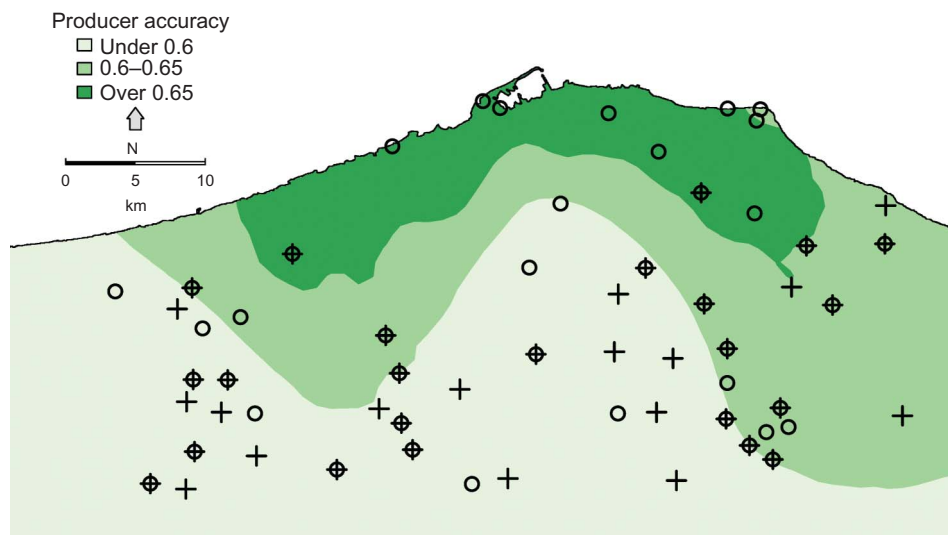
Figure 3.   Spatial variation in producer accuracy for the class of Grazing Land. Crosses indicate locations where Grazing Land was recorded in the reference data and circles indicate locations where Grazing Land was recorded in the classified data.

potential user of the map it indicates the probability of correctly finding that class on the ground. In this case, the results in figure 2 would suggest less confidence in areas mapped as Grazing Land around Tripoli actually being that class on the ground. Figure 3 shows that there is much less variation in the estimates of Grazing Land producer accuracy but that there is a trend of marginally higher accuracy towards the north of the study area. Producer accuracy provides an estimate of the probability that a reference pixel is correctly labelled in the classified data. It indicates the errors of omission (exclusion) and for the producer of the map it indicates the probability that the features of interest are omitted from the classified data. Thus, in this case, we can be confident that most of the grazing land that exists has been mapped (low omission) but that there are also high levels of commission errors in the north of the study area.

4.   Discussion

The confusion matrix provides a convenient summary of land-cover errors, but it provides only global measures of error and accuracy and is aspatial in nature. However, descriptions of the local spatial variation in land-cover accuracy are useful as land-cover (and land-use) data are used as inputs to many environmental and socio-economic models. Determining the local variations in accuracy and error can help users assess the suitability of land-cover data for their application and assess the implications of those errors for their analyses. Geographical analyses are fundamentally concerned with how and why processes vary spatially. They explicitly use the locational attributes of data to identify local variations in relationships. This is not the same as analysing spatial data *per se* under the assumption that because the data are spatial, then their analysis somehow represents a 'spatial analysis'. This sometimes appears to be the case in the remote-sensing community, where the failure to adopt spatially explicit approaches is puzzling, especially as it is well recognized that phenomena such as land-cover errors cluster for a variety of well-documented reasons

and that there is a need for local error assessments which are often by informal visual and qualitative methods (Friedl *et al*. 2002).

This letter shows how logistic regression can be used to generate probabilities such as user and producer accuracies and that geographically weighted extensions are able to generate spatial distributions that describe the variation of these probabilities. Other research has suggested geographically weighted approaches and this research is very much in the spirit of Foody (2005) and Comber *et al*. (2012). Foody (2005) developed a method for describing the geographical distribution of error. Comber *et al*. (2012) proposed and developed a geographically weighted portmanteau measure of accuracy. Portmanteau measures combine both the specificity and the sensitivity of the data by including measures of the probability that either the presence or the absence of a particular land-cover class is correctly classified. However, the remote-sensing community are probably more familiar with the concepts of overall, user and producer accuracies and the methods suggested here may better reflect their needs.

This analysis used data collected as part of a standard remote-sensing validation exercise. While there is an extensive body of research describing the key issues in relation to validation data – see, for example, Stehman (2000), Stehman *et al*. (2003, 2005) and Stehman (2009) – this is not the subject of this letter. Geographically weighted approaches compute a local statistic from data points under the kernel. The critical considerations for their implementation and the effective operation of the kernel include having a spatially distributed sample of data points over which to develop geographically weighted analyses and selecting an appropriate bandwidth.

Determining the bandwidth or the number of data points to include in the local model under the kernel is a trade-off between working with a data set that is too small to calibrate the model reliably and one that is too big such that it averages out local effects. Thus, the aim is to select a bandwidth to ensure that enough data points are used in the calibration to minimize the cross-validation prediction error. The kernel bandwidth can be specified as a fixed distance or to include a proportion of the data points as in this research. Additionally, it can be selected automatically by leave-one-out cross-validation, which identifies a bandwidth for a given geographically weighted analysis that optimizes the ability of the model to predict each individual response variable when it is removed from the data set. Investigating the effects of applying different bandwidths offers a heuristic way of assessing the sensitivity of the data and the model with regard to cross-validation prediction errors. Further details and discussion on bandwidth selection can be found in Fotheringham *et al*. (2002).

As well as developing geographically weighted user and producer accuracies, this research illustrates how geographically weighted methods provide a generic framework: the 'geographically weighted' approach is one that uses a kernel under which any statistic can be calculated. By allowing bandwidths to vary, by using spatially dependent weights and by computing the statistic under the kernel using the distance weighted data, geographically weighted approaches address one of the fundamental tenets of geographical analyses: they operate under the assumption that the process under investigation varies continuously over space. In so doing, they explicitly address Waldo Tobler's first law of geography: '. . . but near things are more related than distant things' (Tobler 1970, p. 236).

Acknowledgements

the statistical analyses and mapping were implemented in R version 2.15.1, the open source statistical software (http://cran.r-project.org). The data and code used in this analysis will be provided to interested researchers on request.

References

BRUNSDON, C.F., FOTHERINGHAM, A.S. and CHARLTON, M., 1996, Geographically weighted regression – a method for exploring spatial non-stationarity. *Geographical Analysis*, **28**, pp. 281–298.

CAMPBELL, J., 1981, Spatial correlation effects upon accuracy of supervised classification of land cover. *Photogrammetric Engineering of Remote Sensing*, **47**, pp. 355–364.

CANTERS, F., 1997, Evaluating the uncertainty of area estimates derived from fuzzy land-cover classification. *Photogrammetric Engineering and Remote Sensing*, **63**, pp. 403–414.

COMBER, A., FISHER, P.F., BRUNSDON, C. and KHMAG, A., 2012, Spatial analysis of remote sensing image classification accuracy. *Remote Sensing of Environment*, **127**, pp. 237–246.

EASTMAN, J.R., 2009, IDRISI Taiga (Worcester, MA: Clark University).

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

FOODY, G.M., 2005, Local characterization of thematic classification accuracy through spatially constrained confusion matrices. *International Journal of Remote Sensing*, **26**, pp. 1217–1228.

FOODY, G.M., 2007, Map comparison in GIS. *Progress in Physical Geography*, **31**, pp. 439–445.

FOTHERINGHAM, A.S., BRUNSDON, C. and CHARLTON, M.E., 2002, *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships* (Chichester: Wiley).

FRIEDL, A.F., MCIVER, D.K., HODGES, J.C.F., ZHANG, X.Y., MUCHONEY, D., STRAHLER, A.H., WOODCOCK, C.E., GOPAL, S., SCHNEIDER, A., COOPER, A., BACCINI, A., GAO, F. and SCHAAF, C., 2002, Global land cover mapping from MODIS: algorithms and early results. *Remote Sensing of Environment*, **83**, pp. 287–302.

GONZALEZ, P., ASNER, G.P., BATTLES, J.J., LEFSKY, M.A., WARING, K.M. and PALACE, M., 2010, Forest carbon densities and uncertainties from Lidar, QuickBird, and field measurements in California. *Remote Sensing of Environment*, **114**, pp. 1561–1575.

MCGWIRE, K.C. and FISHER, P., 2001, Spatially variable thematic accuracy: beyond the confusion matrix. In *Spatial Uncertainty in Ecology: Implications for Remote Sensing and GIS Applications*, C.T. Hunsaker, M.F. Goodchild, M.A. Friedl and T.J. Case (Eds.), pp. 308–329 (New York: Springer).

STEELE, B.M., WINNE, J.C. and REDMOND, R.L., 1998, Estimation and mapping of misclassification probabilities for thematic land cover maps. *Remote Sensing of Environment*, **66**, pp. 192–202.

STEHMAN, S.V., 2000, Practical implications of design-based sampling inference for thematic map accuracy assessment. *Remote Sensing of Environment*, **72**, pp. 35–45.

STEHMAN, S.V., 2009, Sampling designs for accuracy assessment of land cover. *International Journal of Remote Sensing*, **30**, pp. 5243–5272.

STEHMAN, S.V., SOHL, T.L. and LOVELAND, T.R., 2003, Statistical sampling to characterize recent United States land-cover change. *Remote Sensing of Environment*, **86**, pp. 517–529.

STEHMAN, S.V., SOHL, T.L. and LOVELAND, T.R., 2005, An evaluation of sampling strategies to improve precision of estimates of gross change in land use and land cover. *International Journal of Remote Sensing*, **26**, pp. 4941–4957.

TOBLER, W.R., 1970, A computer movie simulating urban growth in the Detroit region. *Economic Geography*, **46**, pp. 234–240.