INT. J. GEOGRAPHICAL INFORMATION SCIENCE
VOL. 18, NO. 6, SEPTEMBER 2004, 611–626

Research Article

Spatial variability in classification accuracy of agricultural crops in the Dutch national land-cover database

P. A. J. VAN OORT, A. K. BREGT, S. DE BRUIN,
A. J. W. DE WIT

Wageningen University and Research Centre, Centre for Geo-Information,
PO Box 47, NL-6700 AA Wageningen, the Netherlands;
e-mail: pepijn.vanoort@wur.nl

A. STEIN

Wageningen University and Research Centre, Mathematical and Statistical
Methods Group, Wageningen, the Netherlands

(*Received 1 December 2003; accepted 11 December 2003*)

Abstract. Variability in per cell classification accuracy is predominantly modelled with land-cover class as the explanatory variable, i.e. with users' accuracies from the error matrix. Logistic regression models were developed to include other explanatory variables: heterogeneity in the 3×3 window around a cell, the size of the patch and the complexity of the landscape in which a cell is located. It was found that per cell, the probability of correct classification was significantly ($\alpha = 0.05$) higher for cells with a less heterogeneous neighbourhood, for cells part of larger patches and for cells in regions with a less heterogeneous landscape. To validate the models, a leave-one-out procedure was applied in which the absolute difference between the actual and the model-estimated number of cells correctly classified was summarized over 55 regions in the Netherlands. The sum of differences reduced from 60.9 to 48.1 after adding the variables 'patch size' and 'landscape dominance' to the land-cover class model. Spatial variability thus modelled therefore led to a substantial improvement in the estimation of the per cell classification accuracy.

1. Introduction

Land-cover data derived from classified satellite images are increasingly used in land-use planning and environmental management. Consequently, concern about the accuracy of these data has grown. Commonly, the classification accuracy is reported by the percentage correctly classified (PCC) and the error matrix (Congalton 1991, Janssen and Van der Wel 1994). Information on the spatial variability of these measures is rarely provided (Foody *et al.* 1992, Goodchild 1995, Smith *et al.* 2002). Lack of quantitative information on this spatial variability can be a serious problem: use of a PCC not representative for a region may lead to misleading outcomes in an error propagation analysis and to incorrect assessment of the fitness for use of the data for a specific region. Therefore, McGwire and

International Journal of Geographical Information Science
ISSN 1365-8816 print/ISSN 1362-3087 online © 2004 Taylor & Francis Ltd
http://www.tandf.co.uk/journals
DOI: 10.1080/13658810410001701969

Fisher (2001) recommend reporting the accuracy for the smallest region size where the producer expects the dataset to be used.

Responding to the demands from users, De Wit (2002) provided estimates of the accuracy of the Dutch national land-cover database (LGN) on a region (62.5 km^2) and province basis. These estimates were computed from reference data located within each of the regions concerned. The disadvantage of such an approach is that proper accuracy assessment may be impossible if there are no or too few reference points within a given region. Model-based approaches possibly do not suffer from this disadvantage.

For example, some researchers (Steele *et al.* 1998) used a spatial interpolator (kriging) for estimating probabilities of correct classification from sample points with known probability of misclassification and models of spatial continuity (variograms). Their method requires that the interpolated points are within the range of influence of sampled points. If this requirement is not met, models calibrated on exhaustively sampled explanatory variables can be considered. The most commonly applied model is based on users' accuracies derived from the error matrix, in which land-cover class is the explanatory variable. Smith *et al.* (2002, 2003) developed logistic regression models to assess the impact of other variables on per cell classification accuracy. They found significant ($\alpha = 0.05$) impacts of both patch size and focal heterogeneity (heterogeneity in the 3×3 window around a cell).

On the basis of visual analysis, De Wit *et al.* (1999) anticipated that regional differences in accuracy were also associated with complexity of the landscape and that differences in landscape complexity were not entirely captured by the variables land-cover class, patch size and focal heterogeneity. To the present authors' knowledge, such observation has never been substantiated by quantitative evidence.

The present paper aims to extend the work of Smith *et al.* (2002, 2003) by including landscape indices (Forman and Godron 1986, O'Neill *et al.* 1988, Li and Reynolds 1993, Riiters *et al.* 1995) as potential explanatory variables of per cell classification accuracy. These indices aggregate the distribution of land-cover classes and of focal heterogeneities within a region to one number representative for a whole region. The following sections describe an experiment with 55 regions taken from the Dutch national land-cover database. First, the datasets, variables and models are introduced and the procedure to validate the models is outlined. Next, the results are presented and the paper ends with a discussion of the results.

2. Methods

2.1. *Data*

Two datasets were used: (1) the national land-cover database (LGN) derived from satellite images and (2) a reference dataset derived from the agricultural census (REF). LGN has a resolution of 25 m and covers the whole of the Netherlands. It is produced every four years, and the most recent update of the database is based on images taken in 1999 and 2000 and was completed in 2001. The production process involves integration of multi-temporal satellite imagery (from Landsat TM and SPOT), ancillary data and expert knowledge (Thunnissen and Noordman 1996, Thunnissen and De Wit 2000). The classification system comprises 39 land-cover classes, seven of which are agricultural crops. In total, agricultural crops cover 52.9% of the country. The database is widely used by national and regional government agencies for water management, hydrological modelling, land-use planning and environmental management (Van Soest *et al.* 2001, De Wit 2002).

The reference dataset used was derived from the agricultural census. For the census, the government annually sends to all farmers one or more 1:10 000 aerial photographs, with parcel boundaries derived from the national topographical map and the cadastral map printed on these photographs. Farmers return these maps, indicating the areas of crops produced within their parcels. Crop name(s) and area(s) are recorded without reference to their position within the parcel. The maps returned by all individual farmers are combined per region. Figure 1 shows 55 of these regions, their location corresponds with map sheets supplied by the national topographical database, which is often used in combination with LGN. The reference dataset was derived from this census dataset in four steps:

- Parcels with more than one crop and parcels with a reported crop area exceeding the geometrical area in the census dataset by more than 10% were

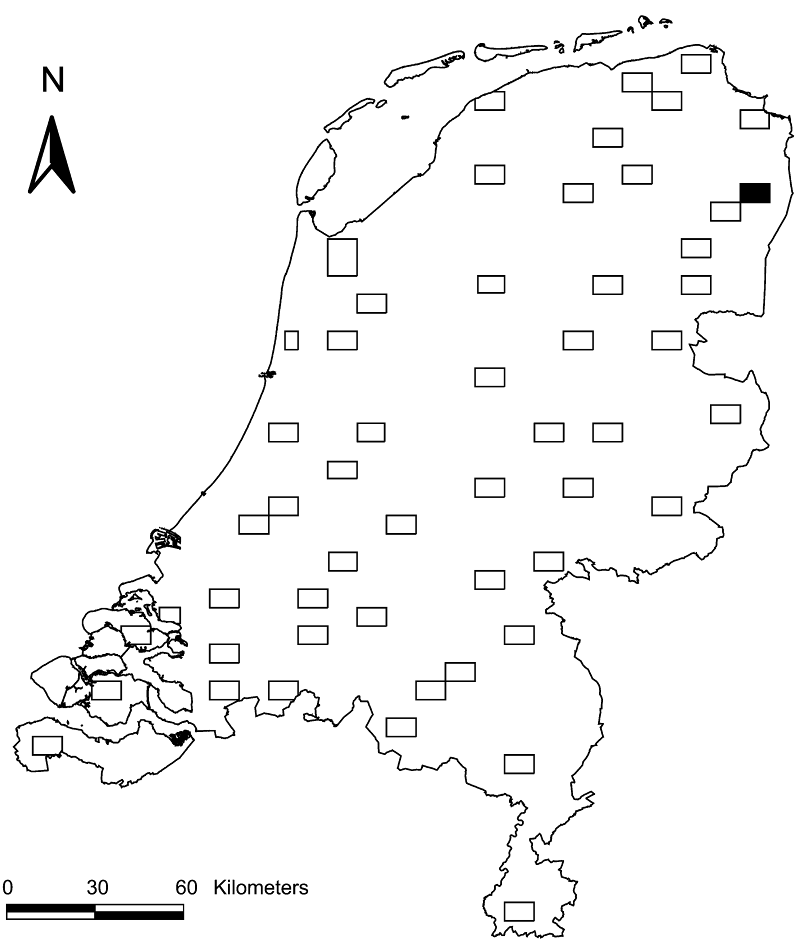

Figure 1. The Netherlands with the regions (rectangles) used. Models to estimate the percentage correctly classified of a validation region (black) were fitted on data of the other 54 regions (white). This procedure was repeated for all regions.

excluded. As a result, the coverage of census data per region ranged between 1 and 42%.

- The census dataset, originally in vector format, was converted to raster to enable overlay with LGN. As in LGN, only one class was assigned to each cell in the reference dataset using the majority rule. This rule assigns to a cell the land-cover class with the largest relative area within the cell.
- Fifty-five of the regions of the census dataset (figure 1) were randomly selected. Of the selected regions, 47 were sized 6.25×10.0 km, the other eight regions were slightly smaller. The resulting dataset contained 955 783 cells with reference land cover.
- Reference data were subsampled to obtain a more or less realistic data density. For that purpose, the dataset was overlaid with a point grid with spacing 30 cells (750 m) and a randomly drawn origin was made in each region (De Gruijter 1999). All points coinciding with census data cells were added to the reference dataset. The resulting reference dataset (REF) contained 1161 cells.

2.2. *Explanatory variables*

Four categories of explanatory variables were used (table 1). Category 1 contains a single variable CLASS, which specifies the land-cover class of the cell.

Category 2 variables quantify the heterogeneity of the focal ($=3 \times 3$) neighbourhood around each cell. Focal heterogeneity, HET equals the number of different land-cover classes in the neighbourhood. Focal homogeneity, HOM equals the number of neighbouring cells with the same land-cover class as the central cell.

The single category 3 variable L10P quantifies the size of the patch in which a cell is located. A patch was defined as a set of contiguous cells of the same land-cover class. Contiguity was defined as sharing a common boundary, thus each cell

Table 1. Explanatory variables.

Category	Variable	Key aspects of variable
1. Land cover	Land-cover class (CLASS)	six binary variables are used to indicate the presence of one of the seven land-cover classes*
2. Focal	Heterogeneity (HET)	number of different classes in the direct (eight-cell) neighbourhood of the cell
	Homogeneity (HOM)	number of cells in the direct (eight-cell) neighbourhood which have the same land-cover class as the centre cell
3. Patch	Patch size (L10P)	contiguous cells with the same land cover grouped into patches
4. Landscape	Heterogeneity (HTG)	independent of landscape texture, more sensitive to the number of land-cover classes
	Dominance (DMG)	independent of landscape texture, less sensitive to the number of land-cover classes
	Entropy (ENT)	dependent on landscape texture, more sensitive to the number of land-cover classes
	Contagion (CON)	dependent on landscape texture, less sensitive to the number of land-cover classes

*If all binaries are zero, then the class is LGN = 7.

had four contiguous cells. Patch shape indices express in one number the distribution of patch sizes occurring in a region (Forman and Godron 1986, p. 189, Baker and Cai 1992). They were not considered here because a preliminary data analysis showed little variation in the values of these indices for the 55 regions assessed in this study (O'Neill *et al.* 1988 obtained similar results for agricultural areas).

Category 4 includes four landscape indices that differ in sensitivity to the number of land-cover classes and in their ability to distinguish between different landscape textures. The indices are known as landscape heterogeneity, landscape dominance, landscape entropy and landscape contagion (O'Neill *et al.* 1988, Li and Reynolds 1993, Riiters *et al.* 1995). We will refer to the indices as landscape variables.

Landscape heterogeneity and landscape dominance are derived from marginal probabilities $p(i)$, $i=1,\ldots, I$, where I is the number of land-cover classes in the region. Let $A(i)$ be the area of land-cover class i in a region. Then $p(i)$, landscape heterogeneity (HTG) and landscape dominance (DMG) are as follows:

$$p(i)=\frac{A(i)}{\sum_{i=1}^{I} A(i)} \tag{1}$$

$$\mathrm{HTG}=-\sum_{i=1}^{I} p(i)\cdot \ln(p(i)) \tag{2}$$

$$\mathrm{DMG}=\ln(I)-\mathrm{HTG}. \tag{3}$$

Landscape entropy (ENT) and landscape contagion (CON) are obtained from the probabilities $p_{\mathrm{adj}}(i, j)$ that a randomly chosen cell is classified as i and that at least one adjacent cell (in a 3×3 window) is classified as j. Let $N_{\mathrm{adj}}(i, j)$ be the total number of adjacencies between cells with class i and class j within a region. Then:

$$p_{\mathrm{adj}}(i,j)=\frac{N_{\mathrm{adj}}(i,j)}{\sum_{i=1}^{I}\sum_{j=1}^{I} N_{\mathrm{adj}}(i,j)} \tag{4}$$

$$\mathrm{ENT}=-\sum_{i=1}^{I}\sum_{j=1}^{I} p_{\mathrm{adj}}(i,j)\cdot \ln\left(p_{\mathrm{adj}}(i,j)\right) \tag{5}$$

$$\mathrm{CON}=2\cdot \ln(I)-\mathrm{ENT}. \tag{6}$$

Table 2 shows the range of values of the landscape variables. In distinguishing between relatively heterogeneous and finely textured landscapes, the variables HTG and ENT are more sensitive to I than the variables DMG and CON.

Table 2. Minimum and maximum values of landscape variables.

Landscape	HTG	DMG	Landscape	ENT	CON
Heterogeneous	$\ln(I)$	0	fine texture	$2\cdot\ln(I)$	0
One class dominates	0	$\ln(I)$	coarse texture	0	$2\cdot\ln(I)$

I, number of land-cover classes in a region.

2.3. *Statistical analysis*

2.3.1. *Logistic regression.*

Logistic regression (Agresti 1990, Collet 1991) was used to calculate the probability of correct classification $p_{corr}(c)$ of a cell c as a function of the explanatory variables introduced above. A logistic regression model with intercept β_0 and with $k = 1, \ldots, K$ explanatory variables x_k equals:

$$p_{\mathrm{corr}}(c) = \frac{\exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k \cdot x_k(c)\right)}{1 + \exp\left(\beta_0 + \sum_{k=1}^{K} \beta_k \cdot x_k(c)\right)}. \tag{7}$$

A linear function $\mathrm{logit}(p_{corr}(c)) = \beta_0 + \sum \beta_k \cdot x_k(c)$ is fitted through the data. In some cases, the linearity of this function can be improved by transforming a variable x. Preliminary data analysis showed that this was the case for patch size, which was transformed to a logarithmic scale.

Regression coefficients are obtained by minimizing the -2 log likelihood (also known as the deviance) of the model. The difference between the deviances of two models follows a χ_l^2 distribution, where l is the number of explanatory variables additional to those shared by the two models. A χ^2-test can then be used to test if adding these l variables to the model significantly improves the fit of the model. Ignorance of spatial dependence in the residuals of the fitted regression models will result in models seeming more significant than they actually are (Bio 2000). We visually checked for spatial dependence in the residuals by plotting their variograms.

An exhaustive model selection procedure was applied that would result in finding the model containing the highest number of significant (at $\alpha = 0.05$) explanatory variables. Table 3 lists the evaluated models; table 6 lists the tests. At each step in the procedure, the significance of the addition of a variable to a model was tested at three significance levels (0.05, 0.01 and 0.001):

Table 3. Models evaluated.

Model number (m)	Model
0	β_0
1	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS}$
2a	$\beta_0 + \beta_1 \cdot \mathrm{HET}$
⋮	⋮
2g	$\beta_0 + \beta_1 \cdot \mathrm{CON}$
3a	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{HET}$
⋮	⋮
3g	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{CON}$
4a	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{HET} + \beta_8 \cdot \mathrm{L10P}$
4b	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{HET} + \beta_8 \cdot \mathrm{DMG}$
4c	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{L10P} + \beta_8 \cdot \mathrm{DMG}$
5a	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{L10P} + \beta_8 \cdot \mathrm{DMG} + \beta_9 \cdot \mathrm{HET}$
5b	$\beta_0 + \beta_{1\text{–}6} \cdot \mathrm{CLASS} + \beta_7 \cdot \mathrm{L10P} + \beta_8 \cdot \mathrm{DMG} + \beta_9 \cdot \mathrm{L10P} \cdot \mathrm{DMG}$

For each of the six binary variables in CLASS, a regression coefficient is estimated: $\beta_1 \ldots \beta_6$.

- Addition of CLASS to model 0 (creates model 1), addition of HET to model 0 (creates model 2a), …, addition of CON to model 0 (creates model 2g).
- Addition of HET to model 1 (creates model 3a), …, addition of CON to model 1 (creates model 3g).
- Addition of a second explanatory variable to a model containing CLASS and one variable of the same category (table 1). For example, addition of HOM to model 3a. Because none of these additions was significant (at $\alpha = 0.05$), the analysis was continued with one variable out of each category: HET, L10P and DMG.
- Addition of one of the variables HET, L10P and DMG to a model containing CLASS and one of these three. For example, addition of L10P to model 3a (creates model 4a).
- Addition of HET to model 4c and addition of L10P·DMG to model 4c.

2.3.2. *Validation.*

A leave-one-out procedure was applied to cross-validate each model. Each time the model was fit on data from 54 regions and one region r was used as the test dataset (figure 1). The resulting 55 parameter sets are referred to as versions of a model. For each of the cells $c = 1, \ldots, n(r)$, where $n(r)$ is the number of cells in region r, the probability of correct classification $p_{\mathrm{corr},m}(c)$ was calculated with model m. The binary variable $y(c)$ obtained from the test dataset indicates if c was actually correctly classified or misclassified. The measure SM_m (equation 8) summarizes over all cells the absolute difference between model estimated and actual correctness of classification. $R0_m$ (equation 9) is the relative improvement of model m to the model assuming the same probability of correct classification for all cells ($m = 0$), $R1_m$ (equation 10) is the relative improvement to the model with CLASS as the explanatory variable ($m = 1$). SM_m, $R0_m$ and $R1_m$ are calculated as:

$$SM_m = \sum_{r=1}^{55} \sum_{c=1}^{n(r)} |p_{\mathrm{corr},m}(c) - y(c)| \tag{8}$$

$$R0_m = 100\% \cdot (SM_0 - SM_m)/SM_0 \tag{9}$$

$$R1_m = 100\% \cdot (SM_1 - SM_m)/SM_1. \tag{10}$$

3. Results

3.1. *Error matrix*

The error matrix is shown in table 4. Overall, LGN has a high classification accuracy: 90.2%. There are large differences in user accuracies between different crops (ranging between 33.3% and 95.8%). An implication of the large differences is that in, for example, a region with a relatively large area of bulb cultivation, the assumption of a PCC of 90.2% could overestimate the PCC of this region.

3.2. *Model selection*

Table 5 shows the estimated regression coefficients for a selection of models. Model 1 shows the effect of CLASS on classification accuracy, models 2a–3g show the effect of the explanatory variables HET to CON. Models 3a–g show the effect

Table 4. Error matrix.

LGN—classified data		REF—reference data							Total	Users' accuracy (%)
		Grass	Maize	Potatoes	Beets	Cereals	Other crops	Bulb cultivation		
1	Grass	566	12	4	1	4	13	1	601	94.2
2	Maize	1	182	4	2	0	1	0	190	95.8
3	Potatoes	7	5	99	0	3	2	0	116	85.3
4	Beets	1	3	1	56	0	0	0	61	91.8
5	Cereals	4	0	0	0	122	9	0	135	90.4
6	Other crops	4	4	16	2	6	20	0	52	38.5
7	Bulb cultivation	3	1	0	0	0	0	2	6	33.3
	Total	586	207	124	61	135	45	3	1 161	
	Producers' accuracy (%)	96.6	87.9	79.8	91.8	90.4	44.4	66.7		PCC (%) 90.2

Table 5. Estimated regression coefficients.

Model number (m)	Regression coefficients β_0	β_1	β_2	β_3	β_4	β_5	β_6	β_7	β_8
0	2.2174								
1	−0.6931	3.4764	3.8177	2.4551	3.1091	2.9322	0.2231		
2a	2.8899	−0.4751							
2b	1.3078	0.1277							
2c	−0.4525	0.9934							
2d	5.7134	−1.8275							
2e	−0.1922	2.0886							
2f	5.8240	−1.3646							
2g	−3.3758	1.6022							
3a	0.0316	3.3821	3.9384	2.4439	3.1197	2.9139	0.2028	−0.4899	
3b	−1.4315	3.3831	3.8174	2.4018	3.0770	2.8734	0.1666	0.1125	
3c	−2.7223	2.1251	3.5502	2.0543	3.0299	2.4440	−0.1030	1.0404	
3d	2.0440	3.4337	4.0300	2.9414	3.5135	3.2936	0.6484	−1.5253	
3e	−2.5261	3.2388	3.7770	2.8400	3.4262	3.2043	0.5814	1.5559	
3f	2.2878	3.4579	4.0815	2.9336	3.5296	3.2677	0.6347	−1.1993	
3g	−4.7307	3.1681	3.6906	2.7595	3.3866	3.1146	0.5129	1.1663	
4c	−3.5752	2.1731	3.4820	2.2875	3.1923	2.6263	0.1395	0.9248	0.9309

See table 3 for model descriptions. For example, in model 2c, β_1 is multiplied by L10P; in model 3c, β_1 is multiplied by the binary indicating if CLASS = 1 and β_7 is multiplied by L10P.

of these variables when CLASS is accounted for. Model 4c is the model that contained the highest number of significant ($\alpha = 0.05$) explanatory variables. If $\beta_k < 0$, the probability of correct classification decreases with an increase of k; if $\beta_k > 0$ the probability increases. The impact of the variable k decreases as β_k

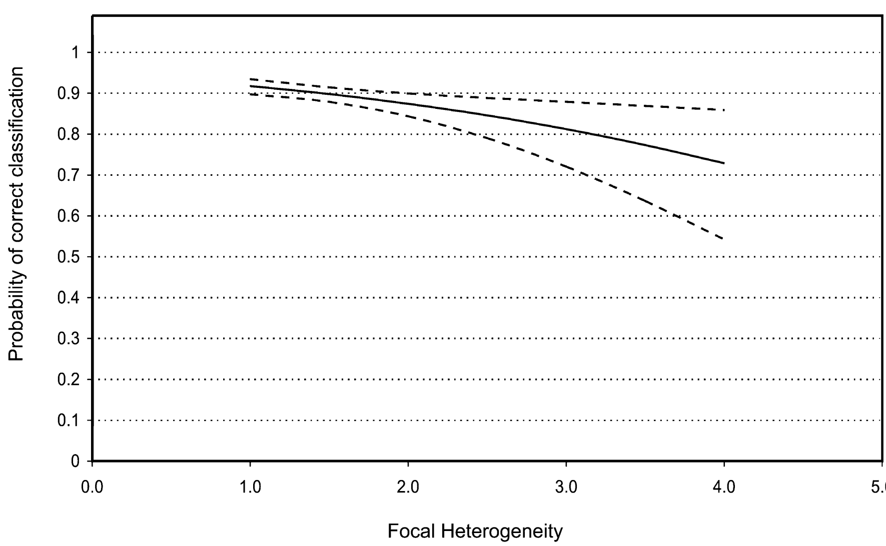

Figure 2. Impact of focal heterogeneity (HET) on the probability of correct classification: model 2a (solid line) and 95% confidence interval (dotted line).

approaches 0. Of interest are the effect of our explanatory variables on classification accuracy and the influence of accounting for CLASS. We see that classification accuracy is higher for higher values of focal homogeneity, patch size, in regions with landscapes with one dominant class and in regions with a coarser texture. As an illustration, figures 3–5 show for the models 2a, 2c, and 2e, the expected probabilities of correct classification and the 95% confidence intervals. The broad

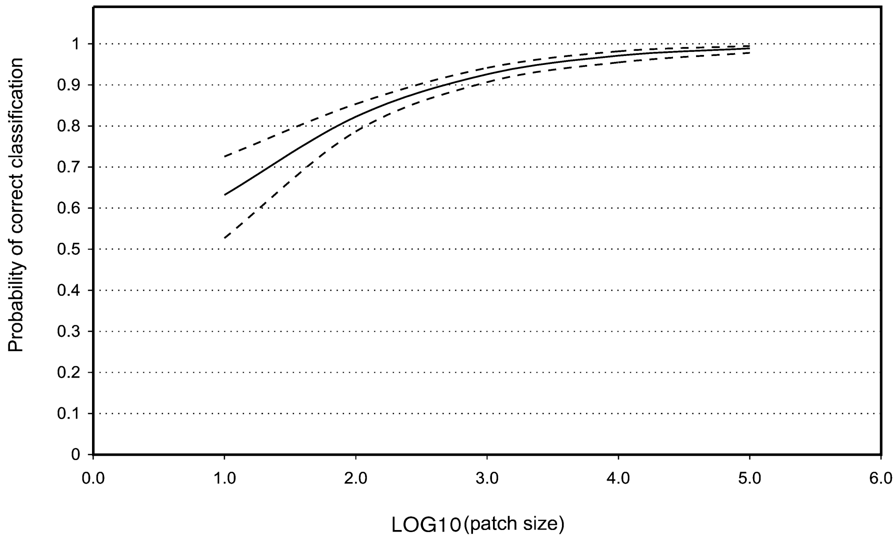

Figure 3. Impact of patch size (L10P) on the probability of correct classification: model 2c (solid line) and 95% confidence interval (dotted line).

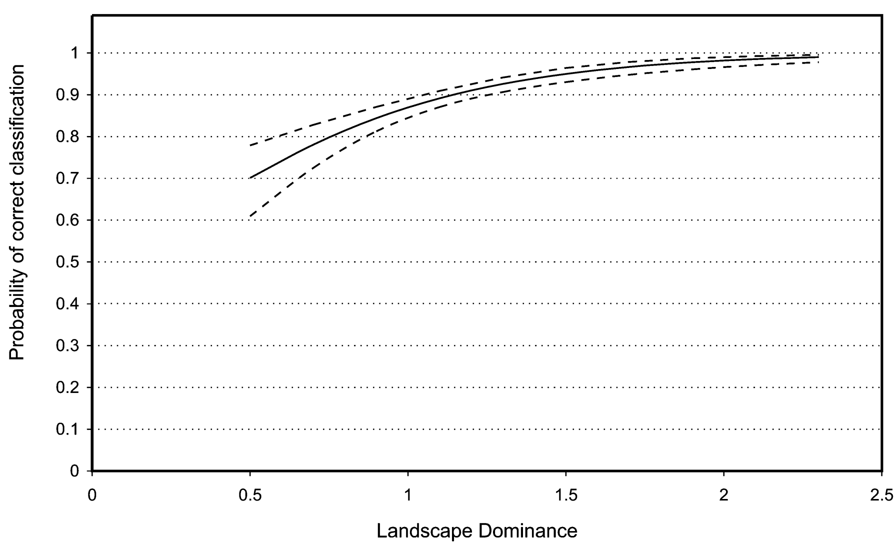

Figure 4. Impact of landscape dominance (DMG) on the probability of correct classification: model 2e (solid line) and 95% confidence interval (dotted line).

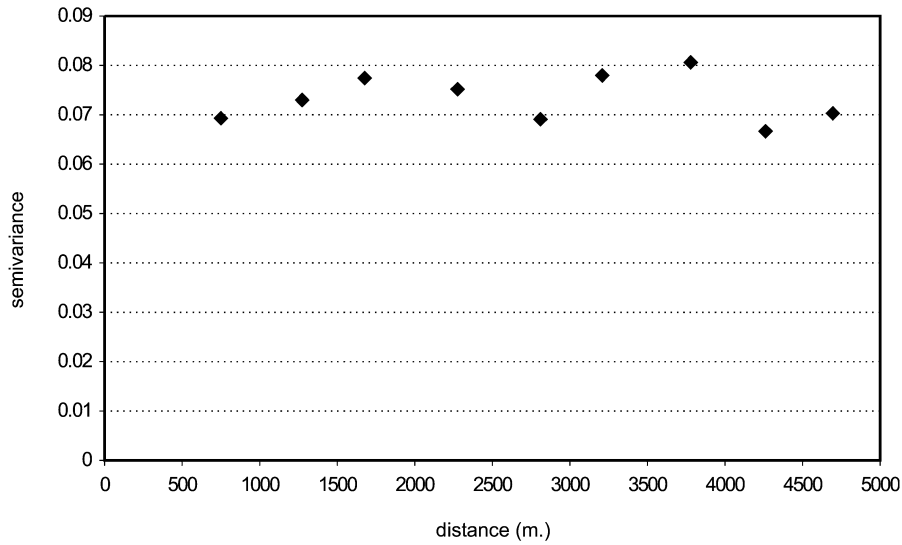

Figure 5. Semivariogram of residuals of model 2c (explanatory variable: L10P).

confidence interval for HET < 2 is due to the distribution of the values of HET in the observations: HET = 1: 68%, HET = 2: 29% and HET < 2: 3%. If we compare the β_1s of models 2a–3g with the β_7s of models 3a–g, we see that the effect of the focal and patch variables does not change when CLASS is accounted for. The impact of the landscape variables (models 2d–3g and 3d–g) is lower in a model containing variable CLASS, indicating that a part of the impact of these variables is accounted for by CLASS.

Figure 5 shows for model 2c that there is no risk of overestimation of significance levels due to spatial dependence in the residuals. Variograms of residuals of the other models also revealed no spatial dependence. Table 6 shows the significance testing of improvement of fit of models by adding an additional explanatory variable x_l. Table 6 shows that of the explanatory variables only HOM was not significant at $\alpha = 0.05$. The patch and landscape variables were highly significant (all at $\alpha = 0.001$) also when added to the model already containing CLASS. HET was significant at $\alpha = 0.01$ in fewer versions if CLASS was already in the model (54 versus 24 versions). Adding L10P or DMG to a model containing CLASS and HET was in all 55 versions significant at $\alpha = 0.001$. Adding HET to, respectively, CLASS & L10P and CLASS & DMG was significant at the $\alpha = 0.05$ level in 0 and 54 of the 55 versions. It is concluded that a model should at least contain CLASS and either L10P or DMG and that it need not contain HET. However, would a model containing CLASS and both L10P and DMG still be better? Adding DMG was significant at $\alpha = 0.05$ in 55 versions and adding L10P was significant at $\alpha = 0.001$ in 55 versions. Finally, it was tested if adding to model 4c the variable HET or the interaction L10P·DMG would significantly (at $\alpha = 0.05$) improve the fit. This was never the case, so model 4c was the model containing the highest number of significant explanatory variables.

Table 6. Chi-square tests for selected models.

Chi-square test (d.f.)	Description: significance of an additional explanatory variable x_l to a model already containing variables x_k		Frequencies of significance at α in the 55 versions* of each model		
	x_k	x_l	$\alpha=0.05$	$\alpha=0.01$	$\alpha=0.001$
D_0–D_1 (6)	1	CLASS	55	55	55
D_0–D_{2a} (1)	1	HET	55	54	0
D_0–D_{2b} (1)	1	HOM	5	0	0
D_0–D_{2c} (1)	1	L10P	55	55	55
D_0–D_{2d} (1)	1	HTG	55	55	55
D_0–D_{2e} (1)	1	DMG	55	55	55
D_0–D_{2f} (1)	1	ENT	55	55	55
D_0–D_{2g} (1)	1	CON	55	55	55
D_1–D_{3a} (1)	1 & CLASS	HET	54	24	0
D_1–D_{3b} (1)	1 & CLASS	HOM	0	0	0
D_1–D_{3c} (1)	1 & CLASS	L10P	55	55	55
D_1–D_{3d} (1)	1 & CLASS	HTG	55	55	55
D_1–D_{3e} (1)	1 & CLASS	DMG	55	55	55
D_1–D_{3f} (1)	1 & CLASS	ENT	55	55	55
D_1–D_{3g} (1)	1 & CLASS	CON	55	55	55
D_{3a}–D_{4a} (1)	1 & CLASS & HET	L10P	55	55	55
D_{3a}–D_{4b} (1)	1 & CLASS & HET	DMG	55	55	55
D_{3c}–D_{4a} (1)	1 & CLASS & L10P	HET	0	0	0
D_{3c}–D_{4c} (1)	1 & CLASS & L10P	DMG	55	7	0
D_{3e}–D_{4a} (1)	1 & CLASS & DMG	HET	54	4	0
D_{3e}–D_{4c} (1)	1 & CLASS & DMG	L10P	55	55	55
D_{4c}–D_{5a} (1)	1 & CLASS & L10P & DMG	HET	0	0	0
D_{4c}–D_{5b} (1)	1 & CLASS & L10P & DMG	L10P·DMG	0	0	0

*Version is a model fitted on data from 54 regions, with cells of one region left out for cross-validation.

D_m, deviance of model m; variables are described in table 1, models in table 3; $x_k=1$ is multiplied by the intercept β_0.

3.3. *Model validation*

Table 7 shows the absolute differences between actual classification correctness and estimated probabilities of correct classification, summarized over all cells. It shows that all model estimates were better than model 0 ($R0_m>0$ for all m). The models containing focal variables were slightly better than model 0 ($R0_{2a}=1.3\%$, $R0_{2b}=1.5\%$), though they produced worse estimates than model 1 ($R1_{2a}=-34.4\%$, $R1_{2b}=-34.1\%$). Their contribution to the model containing variable CLASS was marginal: $R1_{3a}=0.6\%$, $R1_{3b}=0.2\%$. The patch and landscape variables provided better estimates than models 0 and 1. Adding a patch or landscape variable to a model containing intercept only (models 2c–g) improved model 1 estimates by $R1_{2c}=3.5\%$ to $R1_{2g}=19.2\%$. Adding a patch or landscape variable to a model containing CLASS (models 3c–g) improved model 1 estimates by 10.6–15.7%. Model 4c, the model with the highest number of significant variables, also had the

Table 7. Validation: comparison between the actual and the estimated number of cells correctly classified.

			Relative to models 0 and 1	
Model number (m)	Model description	SM_m	$R0_m$ (%)	$R1_m$ (%)
0	same PCC in all regions	83.0		
1	CLASS (=error matrix)	60.9	26.6	
2a	HET	81.8	1.3	−34.4
2b	HOM	81.7	1.5	−34.1
2c	L10P	58.8	29.1	3.5
2d	HTG	58.3	29.8	4.3
2e	DMG	51.0	38.5	16.2
2f	ENT	58.5	29.5	3.9
2g	CON	49.2	40.7	19.2
3a	CLASS & HET	60.5	27.0	0.6
3b	CLASS & HOM	60.8	26.8	0.2
3c	CLASS & L10P	52.4	36.8	13.9
3d	CLASS & HTG	54.4	34.4	10.6
3e	CLASS & DMG	51.5	37.9	15.4
3f	CLASS & ENT	54.0	35.0	11.4
3g	CLASS & CON	51.4	38.1	15.7
4c	CLASS & L10P & DMG	48.1	42.0	21.0

highest $R0_m$ and $R1_m$. Relative to models 0 and 1, model 4c improved estimates of probabilities of correct classification by 42.0% and 21.0%, respectively.

3.4. *Model validation versus model selection*

In general, the ranking of models according to significance levels of contributing variables in the model selection corresponded well with the ranking of models according to $R0_m$ and $R1_m$ in the validation. Landscape variables yielded the greatest improvement in estimates of per cell probabilities of correct classification, the patch variable yielded a greater improvement than the focal variables. Model selection showed that adding CLASS to a model containing DMG or CON was significant at $\alpha=0.001$ (testing $\chi^2_{l=6}=D_{2e}-D_{3e}$ and $\chi^2_{l=6}=D_{2g}-D_{3g}$), yet validation showed a deterioration when CLASS was added ($R1_{3e}<R1_{2e}$ and $R1_{3g}<R1_{2g}$). Finally, the big differences in significance between HET and HOM observed during model selection were absent in the model validation.

3.5. *Comparison between landscape variables*

Table 7 shows that $R1_{3e}-R1_{3d}>R1_{3f}-R1_{3d}$, indicating a stronger increase in $R1$ values when HTG was replaced by DMG than when HTG was replaced by ENT. Similar comparisons between $R0_m$ and $R1_m$ of other models, e.g. $R1_{2g}-R1_{2f}>R1_{2g}-R1_{2e}$, confirmed that PCC estimates improved more strongly when a landscape variable less sensitive to the number of land-cover classes was included in the model (DMG or CON) than when a variable sensitive to texture was included (ENT or CON). The table further shows that the impact of replacing HTG by DMG or ENT by CON is smaller in models 3d–g than in models 2d–g. This indicates that a part of the effect of accounting for sensitivity to the number of land-cover classes is accounted for by the inclusion of the variable CLASS in the model.

4. Conclusion and discussion

The aim was to extend the work by Smith *et al.* (2002, 2003) by evaluating landscape indices (Forman and Godron 1986, O'Neill *et al.* 1988, Li and Reynolds 1993, Riiters *et al.* 1995) as potential explanatory variables of variability in the classification accuracy of cells. Four landscape indices were applied, differing in sensitivity to the number of land-cover classes and in ability to distinguish between different landscape textures. This section will discuss the outcomes of the model selection and validation, and then it will discuss the impact of the landscape variables.

4.1. *Model selection and validation*

The model selection procedure showed the significance of variables in a series of models with increasing number of variables. The model with the highest number of significant ($\alpha = 0.05$) explanatory variables included the variables 'land-cover class', 'patch size' and 'landscape dominance'. It was found that per-cell classification accuracy was significantly higher for cells with more heterogeneous focal ($= 3 \times 3$) neighbourhoods, cells located in larger patches, cells located in regions with a less heterogeneous landscape and cells located in regions with a more coarsely textured landscape.

To assess the validity of our statistical analysis, spatial dependence was checked in the model residuals (Bio 2000). Variograms of the residuals did not reveal such dependence. A leave-one-out procedure was applied to validate the models. Improvement was measured relative to the model assuming the same probability of correct classification for all cells and relative to the model with 'land-cover class' as the single explanatory variable. The model with the highest number of significant explanatory variables also yielded the largest relative improvements: 21% and 42%, respectively. In general, the ranking of models according to significance levels during model selection corresponded well with the ranking according to relative improvements during the model validation.

4.2. *Impact of landscape variables*

Model selection showed in all stages a significant contribution of the landscape variables HTG, DMG, ENT and CON. The impact was smaller but still significant if the model contained variable 'land-cover class', which indicates that a part of the impact of these variables is already accounted for by this variable. Validation results showed that estimates of per cell probabilities of correct classification were better in models with landscape variables less sensitive to the number of land-cover classes (DMG and CON) and that estimates were slightly better in models with variables sensitive to landscape texture (ENT and CON).

Acknowledgements

The authors thank the Dutch Ministry of Agriculture, Fisheries and Nature Management for making the data available for the validation of the crop classification. The work presented in this study is part of a PhD project funded by the Netherlands Geodetic Commission (NCG), the Survey Department of the

Ministry of Transport, Public Works and Water Management, Geo-Information and ICT Department (RWS-AGI), the Netherlands Topographical Service (TDN), the Netherlands Council for Geographical Information (Ravi), the Netherlands Institute of Applied Geoscience—National Geological Survey (TNO-NITG), and the Netherlands Green World Research Institute (Alterra).

References

AGRESTI, A., 1990, *Categorical Data Analysis* (New York: Wiley).

BAKER, W. L., and CAI, Y., 1992, The r.le programs for multiscale analysis of landscape structure using the GRASS geographical information system. *Landscape Ecology*, **7**, 291–302.

BIO, A. F. M., 2000, *Does Vegetation Suit Our Models?: Data and Model Assumptions and the Assessment of Species Distribution in Space* (Utrecht: Nederlandse Geografische Studies, The Royal Dutch Geographical Society).

COLLET, D., 1991, *Modelling Binary Data* (London: Chapman & Hall).

CONGALTON, R. G., 1991, A review of assessing the accuracy of classifications of remotely sensed data. *Remote Sensing and Environment*, **37**, 35–46.

DE GRUIJTER, J., 1999, Spatial sampling schemes for remote sensing. In *Spatial Statistics for Remote Sensing*, edited by Stein, A., van der Meer, F., and Gorte, B. (Dordrecht: Kluwer), pp. 211–284.

DE WIT, A. J. W., 2002, Homepage of the Dutch land use database (available at: http://www.cgi.girs.wageningen-ur.nl/cgi/projects/lgn/index_uk.htm).

DE WIT, A. J. W., VAN DER HEIJDEN, T. G. C., and THUNNISSEN, H. A. M., 1999, *Vervaardiging en nauwkeurigheid van het LGN3-grondgebruiksbestand* (Wageningen: DLO-Staring Centre).

FOODY, G. M., CAMPBELL, N. A., TRODD, N. M., and WOOD, T. F., 1992, Derivation and applications of probabilistic measures of class membership from the maximum-likelihood classification. *Photogrammetric Engineering and Remote Sensing*, **58**, 1335–1341.

FORMAN, R. T. T., and GODRON, M., 1986, *Landscape Ecology* (New York: Wiley).

GOODCHILD, M. F., 1995, Attribute accuracy. In *Elements of Spatial Data Quality*, edited by Guptill, S. C., and Morrison, J. L. (Amsterdam: Elsevier), pp. 59–79.

JANSSEN, L. L. F., and VAN DER WEL, F. J. M., 1994, Accuracy assessment of satellite derived land-cover data: a review. *Photogrammetric Engineering and Remote Sensing*, **60**, 419–426.

LI, H., and REYNOLDS, J. F., 1993, A new contagion index to quantify spatial patterns of landscapes. *Landscape Ecology*, **8**, 155–162.

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*, edited by C. Hunsaker and M. Goodchild (New York: Springer), pp. 308–329.

O'NEILL, R. V., KRUMMEL, J. R., GARDNER, R. H., SUGIHARA, G., JACKSON, B., DEANGELIS, D. L., MILNE, B. T., TURNER, M. G., ZYGMUNT, B., CHRISTENSEN, S. W., DALE, V. H., and GRAHAM, R. L., 1988, Indices of landscape pattern. *Landscape Ecology*, **1**, 153–162.

RIITERS, K. H., O'NEILL, R. V., HUNSAKER, C. T., WICKHAM, J. D., YANKEE, D. H., TIMMINS, S. P., JONES, K. B., and JACKSON, B. L., 1995, A factor analysis of landscape pattern and structure metrics. *Landscape Ecology*, **10**, 23–39.

SMITH, J. H., WICKHAM, J. D., STEHMAN, S. V., and YANG, L., 2002, Impacts of patch size and land-cover heterogeneity on thematic image classification accuracy. *Photogrammetric Engineering and Remote Sensing*, **68**, 65–70.

SMITH, J. H., WICKHAM, J. D., STEHMAN, S. V., and YANG, L., 2003, Effects of landscape characteristics on land-cover class accuracy. *Remote Sensing of Environment*, **84**, 342–349.

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

THUNNISSEN, H. A. M., and DE WIT, A. J. W., 2000, The national land-cover database of the Netherlands. In *Geoinformation For All; XIX Congress of the International Society for Photogrammetry and Remote Sensing* (Lemmer: GITC), pp. 223–230.

THUNNISSEN, H. A. M., and NOORDMAN, E., 1996, *National Land-cover Database of The Netherlands: Classification Methodology and Operational Implementation* (Delft: BCRS).

VAN SOEST, F., STEIN, A., DEKKERS, A. L. M., and VAN DUIJVENBOODEN, W., 2001, A quantitative evaluation of monitoring networks for region-specific nitrate reduction policies. *Journal of Environmental Management*, **61**, 215–225.