GEOMORPHOMETRY
Concepts, Software, Applications

Edited by

TOMISLAV HENGL
Institute for Biodiversity and Ecosystem Dynamics
University of Amsterdam
Amsterdam, The Netherlands

HANNES I. REUTER
Institute for Environment and Sustainability
DG Joint Research Centre
Land Management and Natural Hazards Unit – European Commission
Ispra, Italy
Vegetation Mapping Applications

S.D. Jelaska

Vegetation mapping and its importance · the role of geomorphometry in vegetation mapping · the spatial prediction of vegetation variables using land-surface parameters · statistical prediction models and their use · evaluating mapping accuracy · does data from remote sensing compete or cooperate with land-surface parameters and objects in predicting current vegetation cover? · spatial resolution and statistical methods for mapping vegetation

1. MAPPING VEGETATION

1.1 Why is it important?

Vegetation mapping started with the work of von Humboldt at the very beginning of the 18th century, but did not begin to develop into a profession until more than a century later. Although vegetation (i.e., plant cover of any kind) had been represented on maps for much longer than that, in those distant times, it was mainly shown in coarse thematic resolution, as supplementary information on maps of which the main topics were relief and/or settlements and roads. For example, on 18th and 19th centuries Austrian military maps, vegetation was mapped as forests, pastures, swamps, vineyards and crops. By the 20th century, the development of various hierarchical systems for vegetation classification boosted the creation of maps that focused mainly on vegetation. Especially after the Second World War, this trend gained further support with the development of aerial photography.

Another significant increase in vegetation mapping occurred during the last quarter of the 20th century due to:

• an increased need for spatially organised data about the living component of the world. This data was required to inform environmental and nature management, to predict scenarios, to identify and select important areas for nature protection and/or conservation, and to make environmental impact assessments, etc.;
• the development of GIS, as a very efficient way of storing, creating and analysing spatial data. An added attraction is that the capabilities of the system are constantly increasing, while the costs are decreasing;
• the development of remote-sensing techniques with ever richer in spatial and spectral detail (further insight into this topic can be found in Alexander and Millington, 2000).

An important, often previously neglected, attribute, that should accompany every vegetation map, is an assessment of the accuracy of the displayed data. This is very often carried out using Kappa statistics (Congalton and Green, 1999), although these have been criticised for being over-used, and that they are not always the best method available (Maclure and Willett, 1987; Feinstein and Cicchetti, 1990). For an overview of rater agreement methods, see e.g. Mun and Von Eye (2004). Besides providing information on the current type of biota\(^1\) at a given area, with the help of well-defined ecological indicator systems (Ellenberg et al., 1992), vegetation maps also provide plenty of information about the prevailing ecological conditions with respect to a number of environmental variables (such as soil acidity, soil-water content, mean air temperature, etc.).

A recent example of soil-parameter prediction using indicator values of current vegetation, mapped using remote-sensing techniques, can be found in Schmidtlein (2005). Furthermore, when mapped at community level, as defined in Braun-Blanquet (1928), vegetation maps provide a good basis for most habitat classifications (Antonić et al., 2005), and for land-cover mapping projects. Data on the spatial distribution of vascular plants can also be very valuable for estimating the overall biodiversity. This was shown by Sætersdal et al. (2003), who demonstrated that vascular plants are a good surrogate group of organisms in biodiversity analyses.

**Remark 1.** Knowing the spatial, and temporal, distribution of vegetation is important because vegetation acts as an identity card — it tells us about the environment and the potential biota under present conditions.

Nowadays, a thorough understanding of the global changes that are taking place in the environment is a necessity, as is the need to quantify the speed and amount of those changes. Under these circumstances, historical vegetation maps (of various thematic resolutions) have become a very valuable tool in these analyses and estimations. Consequently, there is increased pressure to produce baseline maps of the current situation, so that they can serve as reference for monitoring of future actions, especially in important nature-conservation areas.

There are also initiatives that include large areas, such as CORINE LAND COVER\(^2\) (CLC), serviced by the European Environment Agency (http://www.eea.europa.eu). Although some of the CLC’s 44 classes of 3-level nomenclature say very little, or nothing, about present vegetation (e.g. 1.1.1. Continuous urban fabric or 5.1.1. Water courses), some of them give more precise ‘green’ information

---

\(^1\) Biota — the animals, plants, fungi and microbes that live, or have lived, in a particular region, during a certain period.

\(^2\) The CORINE (Coordination of Information on the Environment) Programme was established in 1985 by the European Commission, using three main CORINE Inventories (Biotopes, Corinair and Land Cover).
(e.g. 3.1.2. Coniferous forest or 3.2.2. Moors and peatland). A CLC map, with minimum mapping units of 25 ha, has been prepared, derived from interpretations of satellite images. It shows the land cover of part of Europe, between the 1990s and 2000, and includes a change analysis for the same period. This is a valuable tool and data set for environmental policy makers and for anyone else working in related fields.

1.2 Statistical models in vegetation mapping

Nowadays, statistical models are used in almost all vegetation mapping. Exceptions are local large-scale projects and, for example, in CLC projects for which a methodological prerequisite is that the boundaries of the RS images are delineated manually. In all other cases, the statistical approaches applied are almost as diverse as the vegetation itself. For example, the range of statistical models on disposition is huge, varying from simple univariate linear regressions to very complex models such as Neural Networks (Bishop, 1995), Support Vector Machines (Cristianini and Shawe-Taylor, 2000) or Naïve Bayesian Classifiers (Duda et al., 2000). Overviews of the techniques have been made by Franklin (1995) and Segurado and Araújo (2005), and some direct comparison can also be found in Oksanen and Minchin (2002), Jelaska et al. (2003). A valuable comparison of predictive models used to map distribution of species can be found in Latimer et al. (2004).

Numerous elements can determine which model would be the best to use. These can be objective elements, such as a certain type of variable scale (e.g. nominal, categorical, ordinal), or the size of the input sample and the number of predictors. At the other end of the range, the elements can be purely subjective, such as the researcher’s preference for particular methods. However, inevitably, the latter will be limited to those methods that satisfy the conditions dictated by the type and size of the input data. The only rule that can perhaps be pointed out here, is to try to use data sets that are sufficiently large to ensure that a stable model can be built, and that it can be tested on an independent data set. Obtaining a sufficiently large data set, especially when costly and time-consuming field sampling is involved, could be a critical factor.

**Remark 2.** Statistical methods used in vegetation mapping vary from simple univariate linear regression to neural networks and Bayesian classifiers. Generalised linear models (GLM), classification and regression trees (CART) and general additive models (GAM) are among the most frequently used methods.

The final combination of predictors and methods will be case-dependent and influenced by five main factors: (1) the density of field observations; (2) the size and character of the (support) data on input vegetation; (3) the availability and quality of auxiliary data, such as remote-sensing images and DEM derivatives; (4) the (thematic and spatial) resolution, i.e. the scale of predictor variables; (5) the capabilities of the GIS and the statistical software, etc.

Among the most frequently used methods in vegetation mapping are: generalized linear models (GLM), classification and regression trees (CART) and general additive models (GAM). These could be combined with ordination (e.g. cor-
respondence) and/or classification (e.g. cluster) analyses (Gottfried et al., 1998; Guisan et al., 1999; Pfeffer et al., 2003; Jelaska et al., 2006). See also Section 2.1 in Chapter 19 for additional information about statistical models.

Geostatistics is only occasionally included in vegetation-mapping projects and papers (e.g. Bolstad et al., 1998; Miller and Franklin, 2002; Pfeffer et al., 2003). Since various interpolation methods deal with continuous variables, when it comes to mapping vegetation classes, i.e. with discrete variables, it is only possible to use those methods indirectly, so it becomes even more complex to apply them. A good theoretical background to this problem can be found in a paper by Gotway and Stroup (1997). Another example can be found in Pfeffer et al. (2003) who employed universal kriging [see Equation (2.5) in Chapter 19] by correlating topographic variables and vegetation scores (specifically, abundance of 147 plant species on 223 plots).

Apart from the problem of nominal scale in vegetation data, Miller and Franklin (2002) found that output pattern is highly dependable on the spatial origin of the sample data set. However, with open-source, user-friendly software packages for calculating spatial statistics, and the more widely accessible they become, the more geostatistics is going to find its place in vegetation mapping.

1.3 The role of geomorphometry in vegetation mapping

Because geomorphometry can be used to describe (and define) the physical environment, the expectation is that it will be possible to use it to explain and model vegetation that is directly dependent on environmental conditions and its spatial characteristics [see also Equation (1.2) in Chapter 19]. In fact, the physical environment has always been used for this purpose, since, only occasionally, entire areas have been completely field surveyed and mapped for their vegetation at that point in time. Depending on thematic and spatial mapping resolutions, and the diversity of the terrain at the time of mapping, mappers use land-surface parameters (elevation belts, aspect, slope, etc.) combined with field observations to create polygons that covered the entire area of interest. When these estimators are not sufficient for estimating the occurrence of a particular type of vegetation, they use land-surface parameters in combination with estimators, such as geology, annual rainfall, and mean temperature.

These conditioned rules can be viewed as simple spatial inference systems, where conditions can be rather trivial: e.g. if the elevation is 350–500 m, then map in mixed oak–beech forest. However, conditions can also be complex: e.g. everywhere in an elevation belt where the soil acidity (pH) is lower than 4, acid beech forest is present, otherwise there is mixed oak-beech forest. In the majority of cases, the mapper has to deal with a combination of conditions. From the schematic distribution of six different vegetation types, represented in Figure 1, several facts can be observed. Vegetation types follow the temperature gradient in both horizontal (i.e. geographical latitude) and vertical directions (i.e. in elevation belts). However, if we use elevation as the sole estimator, we might make a wrong prediction, depending on whether we have input data from, for example, the northern or southern slopes of a mountain. This is because vegetation belts are lower on the
FIGURE 1  A schematic distribution of six types of vegetation, each represented by a different symbol. The base of the temperature affinity triangle denotes an affinity for higher temperatures, and the apex for the lower ones. (The direction of North is shown by the letter “N” and an arrow.)

northern slopes than they are on the southern ones (in the northern hemisphere, that is, and the opposite applies in the southern hemisphere).

Furthermore, vegetation belts occur at higher elevations on larger mountains, therefore we should be careful when extrapolating our models outside the sampled area. Special geomorphometric features such as sinkholes (shown somewhat exaggeratedly between two peaks in Figure 1) could cause temperature inversion that would lead to an inversion of the vegetation belts, and these will differ on the northern and southern sides of the sinkhole. The intensity of a slope can be critical for the development of a distinct type of vegetation within the same elevation belt. This is illustrated in Figure 1 by two vegetation types that have the same temperature affinity. Besides the basic land-surface parameters (i.e. DEM, SLOPE, ASPECT) shown and discussed here, other land-surface parameters (e.g. WTI and/or SPI) could also be crucial for certain types of vegetation.

REMARK 3. Importance of land-surface parameters in vegetation mapping is case dependent. Thematic resolution of vegetation determines whether elevation, flow accumulation potential or some other parameter will play a crucial role in spatial distribution of a given vegetation type.

Nowadays, land-surface parameters and objects are not just a set of GIS layers used for shaping and transferring polygons of present vegetation onto a paper map. They are used for constructing very complex statistical models to predict the spatial distribution of vegetation. Table 1 lists several examples of how geomorphometry is applied in mapping vegetation.
TABLE 1  Examples from the literature on the use of geomorphometry in vegetation mapping

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of input LSPs</th>
<th>Other predictors</th>
<th>Resolution (DEM/thematic)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gottfried et al. (1998)</td>
<td>18</td>
<td>No</td>
<td>1 m / species &amp; vegetation</td>
</tr>
<tr>
<td>del Barrio et al. (1997)</td>
<td>6</td>
<td>Yes (2)</td>
<td>10 m / landscape units</td>
</tr>
<tr>
<td>Beck et al. (2005)</td>
<td>11</td>
<td>Yes (12)</td>
<td>20 m / species</td>
</tr>
<tr>
<td>Davis and Goetz (1990)</td>
<td>5</td>
<td>Yes (2)</td>
<td>30 m / vegetation types</td>
</tr>
<tr>
<td>Sperduto and Congalton (1996)</td>
<td>2</td>
<td>Yes (2)</td>
<td>30 m / vegetation types</td>
</tr>
<tr>
<td>Franklin (1998)</td>
<td>3</td>
<td>Yes (5)</td>
<td>30 m / species</td>
</tr>
<tr>
<td>Guisan et al. (1999)</td>
<td>10</td>
<td>No</td>
<td>30 m / species</td>
</tr>
<tr>
<td>Jelaska et al. (2006)</td>
<td>3</td>
<td>Yes (1)</td>
<td>30 m / species</td>
</tr>
<tr>
<td>Fischer (1990)</td>
<td>3</td>
<td>Yes (4)</td>
<td>50 m / plant communities</td>
</tr>
</tbody>
</table>

The use of geomorphometry for vegetation mapping applications can be summarised around three points:

- there is no ideal DEM resolution for a thematic resolution of specific types of vegetation, however, most vegetation mapping projects utilise 10–50 m DEMs;
- there are no preferred land-surface parameters that can be used to map vegetation, however, ecological land-surface parameters (climatic and hydrological modelling) are more efficient, in general, for making predictions;
- in most cases, land-surface parameters are used in combination with other parameters — ranging from regolith thickness, substratum characteristics, and parameters derived from remote sensing such as snow cover, water cover, normalised difference vegetative index (NDVI), climatic variables, land-use, and leaf-area index.

Another very important role of land-surface parameters in vegetation mapping applications, even if they are not used directly as vegetation predictors, is for the topographic correction of RS images (Riaño et al., 2003; Shepherd and Dymond, 2003; Svoray and Carmel, 2005), especially in hilly and mountainous areas.

The importance of particular land-surface parameters in vegetation mapping is case-dependent. Whether the elevation, flow accumulation potential or another parameter will play a crucial role in the spatial distribution of a given type of vegetation, will depend on the thematic resolution of the vegetation map and with the current diversity of the land-surface parameters.

Land-surface parameters can also be very useful in mapping vegetation that is influenced by human activities, since man adjusts his activities according to existing ecological conditions. For instance, after clear-cutting the forest vegetation from an area, it is more likely that crops will be grown on the flatter terrain, and vineyards (in the case of the Baranja Hill area) on steeper terrain. Similarly, crops will be grown on lower elevations, and the higher elevations will be reserved for pastures.
2. CASE STUDY

In the following section, using the land-surface parameters of the Baranja Hill case study, we will demonstrate how to map the distribution of the presence of a particular plant species (in this case *Robinia pseudoacacia* L. — Black Locust) and the CORINE land-cover categories. Quantitative data of the presence of the Black Locust (step value 0.2) was obtained by field observations. This represents a coverage percentage ranging from 0 (species absent) to 1 (species completely covering the area — i.e. a pure stand of Black Locust plants). We will compare two sets of predictors: (a) land-surface parameters and (b) LANDSAT image bands.

We will use seven land-surface parameters: elevation (DEM), slope (SLOPE), cosine of aspect (NORTHNESS), sine of aspect (EASTNESS), natural logarithm of flow accumulation potential increased by 1 (LNFLOW), profile curvature (PROFC) and plan curvature (PLANC). All these are prepared in an *ArcInfo* GRID module (see Chapter 11) using the 25 m DEM. The second set consists of eight spectral channels of LANDSAT ETM+ and NDVI (Normalized Difference Vegetative Index).

Both sets were used first separately, and then in combination, which finally gave three sets of predictor variables. We used the STATISTICA program (http://www.statsoft.com) to build predictive models, although operations resembling them are available in *R* (http://r-project.org) and in similar open-source statistical packages.

2.1 Mapping the distribution of a plant species

*Multiple (linear) regression models* (MR) can be calculated making exclusive use of independent variables. To select variables that significantly \( p < 0.05 \) contribute to explaining the variability of Black Locust data, a stepwise regression was used. The MR models follow the general form:

\[
\text{Robinia} = \beta_0 + \beta_1 \cdot q_1 + \beta_2 \cdot q_2 + \cdots + \beta_p \cdot q_p
\]

or in matrix format:

\[
\text{Robinia} = \beta^T \cdot q
\]

where ‘Robinia’ is the coverage of Black Locust, \( b_0 \) the intercept and \( \beta_1, \beta_2, \ldots, \beta_p \) or \( q \) the coefficients of corresponding predictors \( q_1, q_2, \ldots, q_p \), or \( \beta \), included in the model. After the estimation of the regression coefficient, the spatial predictions can be calculated in the *ArcInfo* GRID module to produce the coverage of Black Locust over the whole Baranja Hill area (Figure 2).

A disadvantage of the MR is that the predictions may be outside the physical range of the values (in this case <0 and >1). This is obviously erroneous. A better alternative for interpolating the indicator data is to use *Multiple logistic regression* models (Neter et al., 1996):

\[
\text{Robinia} = \left[ 1 + \exp(-\beta^T \cdot q) \right]^{-1}
\]
FIGURE 2  Multiple regression models of the percentage of cover of Black Locust on Baranja Hill: on the left, full regression models constructed with predictor variables consisting of: (a) RS data only; (c) land-surface parameters plus RS data; (e) land-surface parameters only; and, on the right, stepwise regression models constructed using: (b) RS data only, (d) land-surface parameters plus RS data, and (f) land-surface parameters only.
which can easily be linearised if the target variable ‘Robinia’ is transformed to the logit variable:

\[ \text{Robinia}^+ = \ln \left( \frac{\text{Robinia}}{1 - \text{Robinia}} \right) \]  

(2.4)

where \(0 < \text{Robinia} < 1\). To select just those predictors that contribute significantly (\(p < 0.05\)) towards explaining the variability of Black Locust data, a stepwise multiple logistic regression can be run, contrary to the full multiple regression that will use all seven land-surface parameters. The two logistic predictive models were also applied to the land-surface parameter grids in the ArcInfo GRID module.

Six MR predictive models of the percentage of coverage (Figure 2) give a similar general pattern for the distribution of Black Locust, with some differences in the north-western corner, whereas models with LANDSAT bands as predictors tend to over-estimate the presence of Black Locust. Over-estimation is also evident in the south-eastern corner, except for those models that use land-surface parameters as predictors. This co-mission is probably due to field sampling that did not cover forests present in that section, as can be seen in the orthophoto of the area (Figure 5).

Models using land-surface parameters seem to have a higher local variability, i.e. a more structured output. The proportions of explained variability for all three sets of predictor variables are similar in the models that use a full set of predictors, to those obtained by stepwise regression. For the Black Locust cover on Baranja Hill, the highest adjusted R-squares were those of models using both LANDSAT channels and land-surface parameters. The value for the full model was 0.57 and for the stepwise model (including SLOPE, PROFC and SC2) was 0.60. The value for the other models was 0.50, with the exception of the stepwise Landsat model that had a value of 0.48 for the predictors SC2, SC3, SC5, and also included NDVI.

The predictors selected from the stepwise regression, using land-surface parameters only, were SLOPE and PROFC. Analysis from the regression model that includes SLOPE, PROFC and SC2, does not reveal the spatial autocorrelation of residuals. Hence geostatistical prediction techniques (e.g. regression-kriging, as used in Chapter 20) are not suitable.

Estimating the accuracy of logistic predictive models (Figure 3) is highly dependent upon the chosen threshold value, since logistic models return values of between 0 and 1 to represent the probability of occurrence of a particular species. Whether we choose 0.2 or 0.8 as the threshold \(^3\) value, it will dramatically affect the outcome of the predicted occurrence on a binary presence/absence level. For the Black Locust distribution, we calculated accuracies of input data for threshold values of 0.4 and 0.6.

A full multiple logistic regression model [see Figure 3(a)] shows a high omission error, or under-estimation, of the occurrence of Black Locust. It only predicted its presence accurately, at just one field point. The stepwise model, that included

---

\(^3\) A threshold value is a distinct, calculated, probability of the presence or absence of a species. For some very rare species, a smaller threshold value (e.g. 0.3) will produce a more realistic map of the occurrence of that species, but for more dominant species, to prevent over-estimation, higher values (e.g. 0.7 or 0.8) need to be used.
Logistic regression models showing the probability of occurrence of Black Locust on Baranja Hill: (a) the full model and (b) stepwise-regression model. Constructed by using land-surface parameters only.

DEM and SLOPE as predictors, has an overall accuracy of 78% for a threshold set at a value of 0.4, and 80% at 0.6, and the latter value gave a higher omission error.

REMARK 4. Remote sensing data could be better predictor of main land-cover classes, while land-surface parameters of finer thematic resolution. However this depends upon their spatial resolution.

2.2 Mapping land-cover classes

Three classification trees were constructed, one for each set of predictors, using an exhaustive CART-style search for univariate splits, as the split selection method, with a Gini measure of goodness of fit, and, as a stopping rule, FACT-style direct stopping, using the stopping parameter Fraction of Object set at 0.35 in the STATISTICA package. Due to space constraints, only the classification tree constructed using land-surface parameters and Landsat spectral channels is shown in Figure 4. Kappa statistics for all three models are shown in Table 3. The classification tree model that was developed was then run in the GRID module of the ArcInfo software, using a series of nested IF statements, on grids containing data about the predictors that had been used:

Grid: IF (sc3<=39.5) map = 311
:: else if (sc3>39.5 & sc2>62.5 & dem>220.5) map = 22
:: else if (sc3>39.5 & sc2>62.5 & dem <=220.5) map = 211
:: else if (sc2<=62.5 & lnflow>1.354 & dem>107) map = 22
:: else if (sc2<=62.5 & lnflow>1.354 & dem<=107) map = 32
:: else if (sc2<=62.5 & lnflow<=1.354 & eastn<-0.8962)
    map = 311
:: else if (sc2<=62.5 & lnflow<=1.354 & eastn>-0.8962 &
    curvp > 0.10585) map = 32
:: else map = 24
:: endif
## TABLE 2  Reclassification scheme of land cover for the Baranja Hill field data, prior to development of the predictive models

<table>
<thead>
<tr>
<th>Code</th>
<th>Code description</th>
<th>Cases</th>
<th>New code</th>
<th>Reclassification description</th>
</tr>
</thead>
<tbody>
<tr>
<td>211</td>
<td>agricultural land</td>
<td>19</td>
<td>211</td>
<td>agricultural land</td>
</tr>
<tr>
<td>221</td>
<td>vineyards</td>
<td>3</td>
<td>22</td>
<td>permanent crops</td>
</tr>
<tr>
<td>222</td>
<td>fruit trees and berry plantations</td>
<td>3</td>
<td>22</td>
<td>permanent crops</td>
</tr>
<tr>
<td>231</td>
<td>pastures</td>
<td>2</td>
<td>32</td>
<td>non-forest vegetation</td>
</tr>
<tr>
<td>242</td>
<td>complex cultivation pattern</td>
<td>2</td>
<td>24</td>
<td>heterogeneous agricultural land</td>
</tr>
<tr>
<td>243</td>
<td>land principally occupied by agriculture</td>
<td>4</td>
<td>24</td>
<td>heterogeneous agricultural land</td>
</tr>
<tr>
<td>311</td>
<td>broad-leaved forest</td>
<td>21</td>
<td>311</td>
<td>broad-leaved forest</td>
</tr>
<tr>
<td>321</td>
<td>natural grasslands</td>
<td>2</td>
<td>32</td>
<td>non-forest vegetation</td>
</tr>
<tr>
<td>322</td>
<td>moors and peatland</td>
<td>3</td>
<td>32</td>
<td>non-forest vegetation</td>
</tr>
</tbody>
</table>

where $\text{sc2, sc3, dem, lnflow, eastn}$ and $\text{curvp}$ represents the predictor grid: second and third spectral channels, elevation, natural logarithm of flow accumulation potential, eastness and planform curvature, respectively. The map represents the output grid, i.e. the predicted land-cover map (Figure 4).

The overall kappa value is the highest for the model that combines Landsat channels and land-surface parameters, and the lowest for the classification tree (land-surface parameters). Landsat predictors, based on Kappa statistics, predicted all the land-cover classes better than the land-surface parameters, with the exception of non-forest vegetation, where the kappa values were equal. Combining Landsat channels with land-surface parameters enhanced the predictions of permanent crops, heterogeneous agricultural land and broad-leaved forest compared with the Landsat model, while for non-irrigated arable land and non-forest vegetation, slightly lower predictions were given.

It is interesting to observe in the classification tree built with land-surface parameters and Landsat channels (Figure 4), that the first two splits were carried out, based on values of the spectral channels that classified the majority of cases in forest (code 311) and agricultural areas (codes 211 and 22). Further on, land-surface parameters were used to classify most of the occurrences of other types of vegetation cover, representing, in fact, combinations of various types of cover. This suggests that once spectral channels have separated the main classes (e.g. of forest vs. agriculture), land-surface parameters would be more useful for higher thematic resolutions (e.g. pasture vs. meadows).

**Remark 5.** Although DEM has proven to be a powerful input for mapping vegetation, accuracy-assessment should accompany every vegetation map.
Using split conditions from constructed classification trees through a series of nested if-then-else statements, three new grids were calculated in the ArcInfo GRID module representing predictive land-cover maps of the Baranja Hill area (Figure 5). Water bodies and inland marshes could not be predicted with these models, since these two categories were not sampled during the field work. It can be seen from Figure 5 that there are obvious discrepancies between point and polygon data with respect to their land-cover classification, i.e. compared with the polygons, a significant number of point localities have been classified differently. Since point data originate from direct field observations, they are obviously more accurate than the CLC classification at any given point. From that perspective, some information is inevitably lost due to generalisation.

When visually comparing modelled land-cover maps with CLC and point (field) data, the usage of different levels of classification in reference (3rd level) and modelled (2nd and 3rd level) maps should be taken into account. The two classes with the largest number of cases in the field data — forest and agricultural land —
FIGURE 5 An automated extraction of land-cover classes: (a) an orthophoto of the Baranja Hill area, overlaid with manually digitised land-cover areas; (b) land-cover classes from the CLC 2000 Croatia (www.azo.hr) and field observations; (c) the land-cover of the study area, predicted using land-surface parameters only; (d) the land-cover of the study area predicted using land-surface parameters plus RS data; (e) the land-cover of the study area, predicted using RS data only. (See page 746 in Colour Plate Section at the back of the book.)
TABLE 3  Kappa statistics for predicting the land cover classes (CORINE codes) derived using the Baranja Hill field data

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Landsat only</th>
<th>LSPs only</th>
<th>Landsat &amp; LSPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>211</td>
<td>agricultural land</td>
<td>0.87</td>
<td>0.70</td>
<td>0.81</td>
</tr>
<tr>
<td>22</td>
<td>permanent crops</td>
<td>0.47</td>
<td>0.35</td>
<td>0.91</td>
</tr>
<tr>
<td>24</td>
<td>heterogeneous agricultural land</td>
<td>0.51</td>
<td>0.47</td>
<td>0.57</td>
</tr>
<tr>
<td>311</td>
<td>broad-leaved forest</td>
<td>0.89</td>
<td>0.67</td>
<td>0.96</td>
</tr>
<tr>
<td>32</td>
<td>non-forest vegetation</td>
<td>0.84</td>
<td>0.84</td>
<td>0.80</td>
</tr>
<tr>
<td></td>
<td>overall</td>
<td>0.78</td>
<td>0.66</td>
<td>0.84</td>
</tr>
</tbody>
</table>

Also had the best predicted resemblance to the reference map. The predicted distribution of heterogeneous agricultural land (a second-level CORINE class), mainly follows the pattern of agricultural areas in the reference map, in other words, the first-level CORINE class with which it corresponds. Non-forest vegetation shows a very different pattern in the predicted models, compared with the reference map, although it has equally high kappa values in all the models (Table 3).

None of the three predicted models gave a satisfactory visual distribution of this class, compared with the reference map. This could be due either to an overestimation of non-forest vegetation on the reference map, or because the models did not explain the variability of this class sufficiently, or perhaps it was a combination of both. This would not be surprising, bearing in mind that non-forest vegetation was created by merging three classes that had insufficient data sets (Table 2). In conclusion, it can be said that, to achieve a sensible predictive model, input data has to be collated according to thematic resolution and the expected frequency of the mapped classes.

3. SUMMARY POINTS

The role of vegetation maps nowadays is increasing. In practice, there is a serious imbalance between end-users (i.e. policy makers in fulfilling various international conventions, park managements, etc.) and map producers (i.e. experts). Vegetation mappers are currently having to face a vegetation-data paradox, which means that there is an increasing need for information about vegetation, but there is not enough field data to support this. On the one hand, there is an instant daily need for updated, spatially organised, vegetation data, but, on the other hand, there are fewer trained biologist-ecologists able to recognise plant entities in the field.

In addition to this, end-users often do not understand, or deliberately neglect to acknowledge that, even though it consumes a lot of time and money, fieldwork is very important for producing a good, i.e. accurate, usable, vegetation map. Fortunately, as two side of the coin, development of GIS and RS did increase pressure on vegetation experts, but also helps them to satisfy the rapidly increasing needs of end-users. Statistical models and land-surface parameters have a very important
role to play in vegetation mapping; a role that has at least been partly discussed in this chapter.

There is a wide range of applications of vegetation mapping. At one end, there is, for example, the UNEP/GRID\(^4\) Global Vegetation map and at the other end, the paper by Gottfried et al. (1998). The former was derived for the globe as a whole. This Global Vegetation map is based exclusively upon remote-sensing data from the Advanced Very High Resolution Radiometer (AVHRR). It has a spatial resolution of approximately 1 km\(^2\) and a thematic resolution of eight general vegetation types, namely: desert, semi-desert, alpine desert, tundra, grassland, deciduous forest, evergreen forest and tropical forest.

In contrast to the Global Vegetation map, Gottfried et al. (1998) mapped 1.7 km\(^2\) of a single Alpine summit in Austria for particular plant species and vegetation communities. The level of association was set at a spatial resolution of 1 m\(^2\), using land-surface parameters only.

**Remark 6.** Vegetation mapping applications range from maps of whole globe at 1 km\(^2\) resolution, to projects that cover couple of square kilometres with 1 m\(^2\) spatial resolution.

Between these two ‘extreme’ examples, there are dozens of papers dealing with vegetation mapping (e.g. Fischer, 1990; Marshall and Lee, 1994; Krishnaswamy et al., 2004; Jelaska et al., 2005) that differ in method, thematic and spatial resolution, and the size of the area of interest. Vegetation maps can also be used for gaining information about e.g. water supply, soil pH and soil fertility, as in Schmidtlein (2005). This has already been mentioned in Section 1.1, and in e.g. research dealing with creating habitat suitability models for various animal species (Ball et al., 2005).

Nowadays, land-surface parameters and RS data, as well as the software and hardware to manipulate them, are far more readily available and accessible than they were 10–15 years ago. Hence, the main driving factors that will determine how many land-surface parameters and how much RS data will be incorporated into applications of vegetation mapping will be dictated by the thematic and spatial resolution requirements.

Although, today, satellites are equipped with ever more sophisticated sensors, with respect to spatial and thematic (in the sense of the number of spectral channels) resolution, it is expected that land-surface parameters will remain very important predictors at the higher thematic resolutions of vegetation types in small and medium-scale projects (e.g. on particular habitats, protected areas, county levels, etc.). Support for this claim can be found in the work of Jensen et al. (2001), De Colstoun et al. (2003), Dirnbock et al. (2003), Jelaska et al. (2005). In these papers, the more the thematic resolution increases, the more the accuracy achieved in mapping vegetation decreases. By using more complex methods, researchers are continuously searching for ways of improving the accuracy of RS data classification (Carpenter et al., 1999; Krishnaswamy et al., 2004).

---

\(^4\) UNEP/GRID — United Nations Environment Programme/Global Resource Information Database.
The biggest bottleneck for successful vegetation mapping still seems to be making accurate ground-truth vegetation observations. Remote-sensing data is tending to replace field data. However, at finer thematic resolutions, these are not always sufficiently accurate. The increasing need for data about vegetation, as mentioned in Section 1.1, is not supported enough by real data observed in the field. Future research will inevitably continue to use land-surface parameters and RS data as predictors, but with a focus on achieving better, and more accurate, predictive vegetation maps.

With respect to mapped thematic resolution, and the nature of mapped vegetation, in some cases the classes that are present will have discrete boundaries as for (e.g. crops, pastures and forest), whereas, elsewhere, the boundaries may be continuously changing, as in (e.g. Mediterranean rocky pastures and garigue).

An indisputably good method for determining current vegetation types is to use climatic factors. However, the extent of this correlation is largely dependent upon the scale, where land-surface parameters can explain significant degrees of local variability, from the micro-climatic conditions. Land-surface parameters, therefore, can replace climatic data in certain circumstances, and vice-versa. Actually, those two data sets are partly redundant, or mutually predictable in terms of the temperature characteristics, and, when applied to extremely small-scale problems, have certain limitations.

**Remark 7.** The future of vegetation mapping is in use of advanced statistical methods and new sources of land-surface parameters and remote sensing images. However, ground-truth observations should always remain an irreplaceable data source.

Where there is a present spatial trend, using geostatistics such as regression-kriging, can enhance the predictive powers of models, in cases where the variables being predicted are represented by real numbers (e.g. the abundance of a particular plant species or vegetation type, or the probability of their occurrence). There is no doubt that we can expect significant improvement in obtaining ever more accurate and precise vegetation maps in various scales and classifications. One of the main sources of data for all phases of vegetation mapping, from optimising the field sampling or running an analysis, to making final predictions, will be land-surface parameters.

**Important Sources**


