THE NATIONAL LAND COVER DATABASE OF THE NETHERLANDS

Herman Thunnissen, Allard de Wit
Wageningen University & Research centre
Alterra, Green World Research
Centre for Geo-information
H.A.M.Thunnissen@Alterra.wag-ur.nl
A.J.W.deWit@Alterra.wag-ur.nl

KEY WORDS: Land cover/use classification, Landsat TM, Ancillary data, commercial implementation

ABSTRACT

Because of the lack of suitable land cover/use data for environmental studies it is was decided in 1987 to produce a national land cover/use database of The Netherlands ("LGN data base") using satellite images (Landsat TM and SPOT) and ancillary data. Today, three versions of the database are available (LGN1, 2 and 3), based on satellite images from respectively 1986, 1992/94 and 1995/97. During the up-dating of the LGN database the classification method improved considerably, resulting in a sharp increase of classification accuracy and number of land cover/use classes. The classification procedure of the LGN3 database is based upon an integrated approach using multi-temporal satellite imagery, digital and analogue ancillary data, reference data and expert knowledge. A combination of automatic and interactive classification was used. The LGN3 database consists of 39 classes. For most non-agricultural LGN3 classes the minimum required classification result of 90% and 70% for respectively level I and level II classes is amply reached. In general, the classification results of most agricultural crops in the LGN3 database are high (> 70%), unless the total acreage of a crop in an ‘agricultural stratum’ is relatively low or the concerning crops show a large spectral variability.

The main users of the LGN database are national and regional governmental agencies. It has frequently been used for different purposes in the fields of environmental protection, water management, nature conservation and physical planning on regional and national scales. Mostly, the LGN data are combined with other geographical information. The LGN3 database has been implemented on a commercial basis and different innovative developments which will further improve the LGN database are foreseen.

1 INTRODUCTION

The Netherlands have a high population density. Because of the increasing concern about the impacts of man's intervention on the environment, timely and accurate information on land cover/use at regional and national scales is required by national and regional governmental agencies to support environmental policy and for physical planning purposes. In the middle '80 information on land cover/use could often only be obtained from land use statistics and topographical maps. However, land use statistics were only available for restricted areas (e.g. municipalities or provinces) and could not be derived for areas with deviating boundaries (e.g. river basins and groundwater protection areas). Topographical maps did often not contain all required land cover/use classes, were often outdated and were until recently not available in digital form. Therefore, it was decided in 1987 to produce a national land cover data base of The Netherlands (further to be mentioned 'LGN data base'), using satellite images (Thunnissen et al., 1992a and 1992b). The spatial resolution of images obtained with both the LANDSAT Thematic Mapper (TM) and the French satellite SPOT is in general sufficient for the recognition of the individual agricultural fields in The Netherlands. The LGN-database consists of raster cells of 25 m x 25 m which cover the entire Netherlands, and for each cell land cover is determined. The first version of the LGN database (further to be mentioned 'LGN1 database') was produced by automatic classification of manually stratified single-date satellite images from 1986. The classification accuracy showed a large variation over the country due to spectral confusion between different land cover classes. The extent of spectral confusion in a stratum was dependent on the occurring land cover, size and shape of the land cover units, spectral resolution and acquisition dates of the satellite images and crop development. Further, the applied stratification and the limited availability of field reference data influenced the classification result. Another disadvantage of the classification was that various land use classes could not be differentiated because they possessed similar spectral properties. For example, a short herbaceous cover may represent agricultural use, or recreational use, or residential use.

Although the classification results of the LGN1 database were disappointing in some areas users of the LGN1 database (i.e. national and regional governmental agencies) recognised the potential and were interested in an up-date of the
database. However, the classification results should be improved considerably. The LGN2 and LGN3 databases were produced, using satellite images from respectively 1992 and 1994 and 1995 and 1997 (Noordman et al., 1996 and Wit, et al., 1999). Compared with the LGN1 database the classification method improved considerably, resulting in a sharp increase of classification accuracy and number of land cover classes. In this paper the classification method, including the use of ancillary data and the availability of suitable satellite images, validation and applications of the LGN database and, especially, the LGN3 database will be presented. Finally, commercial implementation and future improvements of the LGN database are discussed.

2 DATA

2.1 Nomenclature

The nomenclature of the LGN3 database consists of 39 classes, which describe the land use for each cell of 25 x 25 meter. In this context it is important to distinguish between 'land use' and 'land cover'. Whereas land use refers to human activity of a certain kind for a certain area, land cover refers to the vegetational and artificial construction occupying the land surface. The LGN3 database consists of land cover as well as land use classes. The LGN3 legend has been grouped in a two-level hierarchy. On the highest level five main land use types are considered: ‘agricultural area’, ‘forest’, ‘nature reserve’, ‘water’ and ‘built-up area’. These level I classes consist of respectively 11, 2, 17, 2 and 8 level II classes (Figure 1).

2.2 Satellite imagery and optimal acquisition periods

The classification of agricultural crops in the LGN3 database is based on a multi-temporal approach. Optical satellite imagery obtained by Landsat TM and SPOT XS have been used during the LGN project. In general, Landsat TM imagery is the preferred type of imagery due to the low cost per unit area and the presence of bands in the middle-infrared part of the electromagnetic spectrum. With the launch of SPOT-4 also this sensor can make observations in the middle-infrared wavelength range, which makes this sensor more suitable in the framework of the LGN Project. The higher cost of SPOT imagery per unit area however hampers its use.

In general, mono-temporal classification using Landsat TM images, obtained during the period mid-May to late September, provides good classification results (> 90%) for most non-agricultural classes. For an accurate classification of most agricultural crops the use of multi-temporal satellite data is required (Thunnissen, 1999 and Thunnissen and Noordman, 1997). The optimal acquisition periods of optical satellite imagery are determined by the phenological characteristics of the main crops (crop emergence and ripening) and cultivation practices (harvesting, after growth for green manuring, mowing, conversion of grassland to arable land and vice versa). A minimum Landsat TM image dataset should include at least one image from spring (April, May) and one image from summer (July, August). The optimal image dataset should include an additional image (June or July) obtained between the previous two periods. Rather optimal TM images were available for the production of the LGN3 database. For the classification of TM images bands 3 (red), 4 (near-infrared) and 5 (middle infra-red) were used. The use of the Normalised Difference Vegetation Index (NDVI) proved to be useful for the discrimination between bare and vegetated fields, especially in spring. The classification of agricultural crops was based on images from 1995 or 1997.

2.3 Ancillary data

The (nation wide) ancillary data used for the production of the LGN3 database are presented below:

The BARS and CBS land use databases have much in common and are therefore considered in the same section. The CBS (Central Bureau of Statistics) land use database provides information on land use of the total area of the Netherlands. The nomenclature consists of 33 land use classes, mainly in artificial areas. The classification is largely based on functional land use like for example ‘Parks and public gardens’. In general, areas smaller than 1 ha are not included in the CBS land use database. Up to 1986 the main sources of information were municipal administrations, and the data were stored on analogue maps and published as land use statistics for grid cells of 500 x 500 meter. From 1989 onwards information on land use is mainly obtained by interpretation of aerial photographs (scale 1:50.000). The actual land use data are stored as digitised maps in a GIS. The digital land use database is available for the entire Netherlands based on aerial photographs of 1989 and revised versions based on photographs of 1993 and 1995. The BARS land use database consists of a digitised land use map and was produced by the State Department for Physical Planning. The nomenclature consists of 33 land use classes. Since the availability of the digital CBS land use database the BARS database was not updated anymore.

The digital 1:10.000 topographic map of the Netherlands (TOP10-vector). The Netherlands Topographic Service (TDN) produces the 1:10.000 digital topographic map of The Netherlands (TOP10-vector). The nomenclature of the
TOP10-vector consists of a few hundred entities which are related to polygon, line and point features. Since 1998, the entire Netherlands are covered by around 1350 map sheets, which cover an area of 5 km to 6.25 km each. The 1:10,000 digital topographic map is being produced from a cartographic point of view and therefore contains mainly information on land cover. The functional use of many classes can only be determined by contextual information, e.g. grassland and forest located in urban areas (parks or sport grounds).

The PIPO system. LASER (Landelijke Service bij Regelingen) is an implementing body of the Dutch Ministry of Agriculture, Nature management and Fisheries which has the task of supervising the allocation of agricultural subsidies in the framework of the Common Agricultural Policy (CAP) of the European Union. An administrative system (PIPO) has been developed which uses GIS technology to check all acreage-based applications for subsidies. In the PIPO system, all subsidised crops, their acreages and the concerning farmers are linked with topographical parcels obtained from TOP10-vector. One parcel from TOP10-vector can contain multiple crops and/or different farmers. PIPO is operational since 1997 and samples of the information stored in PIPO are checked for fraud by farmers using remote sensing.

Agricultural Statistics. The General Census of Agriculture, organised by the CBS, is taken annually in May. Among other things, the General Census of Agriculture provides information on the acreage of crops grown. The CBS Agricultural Statistics contain cultivated areas, not including roads, ditches and hedges less than 4 m wide. The CBS Agricultural Statistics are published per municipality, per province and per ‘agricultural region’. Agricultural regions are more or less homogeneous areas as far as soil type and agricultural land use are concerned. The Netherlands are subdivided in 66 agricultural regions.

3 METHOD

3.1 Stratification and pre-processing of satellite data

The satellite images were stratified prior to the classification based on ancillary data. Stratification involves a division of the area into smaller, more homogeneous areas or strata. Each stratum is classified separately, which may improve classification accuracy considerably (Hutchinson, 1982). Stratification of the LGN database took place at different levels. First of all a stratification has been applied separating the main land use classes: agricultural area, built-up area, natural area, forest and water. Different nation-wide digital geographical databases of the Netherlands enable the discrimination of the main land use types. However, at the time of the production of the LGN2 database the BARS Land Use Database was the only available nation-wide digital geographical database. Depending on the source data used for the production of the BARS database the actuality of different map sheets of the BARS database varied strongly. The main land use classes were up-dated by visual interpretation of satellite images, supported by simultaneous consultation of topographic maps and aerial photographs. The second level of stratification used the provincial boundaries to separate the area in what can be regarded as ‘production regions’ as the database is produced province by province. This is merely a reduction of the amount of data, making it easier to handle. The third level of stratification used the CBS agricultural regions to separate the agricultural areas in the provinces into smaller, more homogeneous areas with known agricultural statistics.

All Landsat TM imagery was geometrically corrected to the Dutch Reference system using ground control points obtained from TOP10-vector. The RMS error can be as large as 1 à 2 pixel. Moreover, due to the ageing of the Landsat TM 5 sensor, locally, geometrical shifts of a few pixels may occur in the imagery which are hard to correct. Cubic convolution resampling (25 m pixels) was used during the georeferencing procedures. This type of resampling improves the image for visual interpretation.

3.2 Classification approach

The classification procedure of the LGN3 database is based upon an integrated approach using multi-temporal satellite imagery, digital and analogue ancillary data, reference data and expert knowledge. Each stratum is classified separately and the land use per stratum is known beforehand from stratification units and statistics. A combination of automatic and interactive classification was used. In general, automatic classification of mono-temporal Landsat TM images, obtained during the period mid-May to late September, provided good results for most non-agricultural classes in the LGN-database. However, some of the concerning LGN3 classes could not be distinguished because they possessed similar spectral reflectances. For example, densely forested residential areas could not be distinguished from (patches of) forest situated in urban area. Discrimination of these LGN3 classes could be achieved by recoding the concerning spectral classes on the basis of the BARS database. For the production of the LGN3 database a large part of the non-agricultural classes could be copied from the LGN2 database, while changes in the non-agricultural classes were interactively updated by combined use of the LGN2-database, recent satellite images and 1 : 25,000 topographic maps.
The classes greenhouses, orchards, tree nurseries and buildings in agricultural areas, which are in general difficult to classify accurately, using satellite imagery alone (Thumissen and Noordman, 1996), were after some preprocessing (e.g. buffering and vector to raster conversion) copied from the Top10-vector. The often outdated information from Top10-vector was updated by visual interpretation of the satellite imagery. The railways and the main roads and water-courses in the LGN3 database were for the greater part copied out of the CBS Land Use database and up-dated by visual interpretation of satellite images.

For an accurate classification of most agricultural crops the use of multi-temporal satellite data is required. Original spectral bands were used (TM bands 3, 4 and 5) for image classification. Mostly, a NDVI image from spring was used to mask the satellite images from summer prior to classification. Spectral confusion between different agricultural crops may have different reasons. An important reason is a deviating crop development due to, for example, shortage of water and/or nutrients or water logging. Agricultural fields may also show (large) spatial variability in spectral reflectance due to ripening, withering or management practices. Based on field shape, context, patterns and/or spatial variation in reflectance (texture) these fields can often be visually recognised as separate fields with a particular crop. A second reason to prefer visual interpretation was the existence of geometrical problems with the satellite imagery. When using automatic multi-temporal classification these distortions will strongly influence classification accuracy. In practice, visual interpretation often appears to be a valuable tool, complementary to automatic classification. Advanced hardware and software enable the simultaneous interpretation of different satellite images, while the interpretation result can directly be stored in digital form by on screen digitising. Therefore, automatic classification of agricultural crops was only used when the relevant class could be discriminated on the basis of a single satellite image. To remove noise and to improve the overall classification accuracy a 3 x 3 pixel majority filter was applied on the output from the automatic classifier.

The stratum 'nature reserve' in the LGN3 database was classified using a combination of visual interpretation of multi-temporal imagery and mono-temporal unsupervised classification (Wit et al., 1999). The classification was heavily supported by analogue ancillary data, which was generally obtained on a much larger scale (1:5000). Unsupervised classification was preferred to supervised classification because many classes are diffuse and training samples are difficult to assign. The class labels were assigned to the resulting clusters after classification.

To enable applications of the LGN3 database on a regional scale, minimum classification accuracies and reliabilities at level I and level III of respectively 90% and 70% are required. However, poor spectral separability, due to non-optimal acquisition dates, may result in classification accuracies and reliabilities below 70%. Therefore, it was decided to update the LGN database only in those years with sufficient suitable satellite images.

### 3.3 Validation approach

With a view to the validity of the decisions based on the information in the LGN database it is of great importance that sufficient reference data are obtained for assessing the accuracy of the results obtained. Congalton (1991) distinguishes 'site-specific' and 'non-site-specific accuracy assessment'. Site-specific accuracy assessment means that the locational and classification accuracy are both assessed together. Non-site-specific accuracy assessment applies only to the classification accuracy while ignoring locational accuracy.

**Site-specific accuracy assessment.** For the assessment of the classification accuracy of the LGN3 database a distinction was made between the non-agricultural classes and the agricultural classes. The accuracy assessment of most non-agricultural classes was based on a systematic sampling scheme. The (potential) sampling points coincide with the points of intersection of the 1 km grid lines on the Dutch topographic maps. To avoid oversampling a stratified systematic sampling was performed. The stratification is based on (groups of) the actual thematic classes themselves. The sampling is concentrated on the most important (groups of) non-agricultural LGN3 classes, namely: deciduous forest, coniferous forest, open vegetated nature area, built-up area and green urban area. From all the points of intersection of the 1 km grid lines situated in the different (groups of) LGN3 classes 100 samples were selected randomly. The reference classes for the selected sampling points were derived from topographic maps and, if necessary, from aerial photographs. The accuracy assessment of the land cover classes which are less important and/or show little variability (i.e. bare soil in nature area, water and bare soil in rural built-up areas) and the land cover classes that are (partly) copied out of other high quality databases (i.e. greenhouses, orchards, buildings in agricultural area, heath land and main roads and railways) was only performed qualitatively and/or in a non-site-specific way.

The agricultural classes of the LGN3 database were validated using data from the PIPO system, which contains a huge amount of reference data. On the average for about 50% of the agricultural area the crops grown are stored in the PIPO system. About 75% of this area concerns plots with only one crop. However, before 1997 PIPO data were not yet digitally stored. Therefore, only the LGN3 database of the eastern part of The Netherlands, that has been classified
using satellite imagery from 1997, could be validated. For the validation 10 Top10-vector map sheets, situated in areas with representative agricultural land use for The Netherlands, were selected. From these maps all plots that were fully covered by a single crop were selected for validation and rasterised to the 25 meter LGN grid.

**Non-site-specific accuracy assessment.** The CBS Agricultural Statistics contain information on the acreage of the agricultural crops, orchards and greenhouses for each of the 66 CBS ‘agricultural’ regions in The Netherlands. By comparing the classified areas in the LGN3 database with the areas provided by the CBS Agricultural Statistics, a non-site-specific accuracy assessment was performed.

4 **LGN3 DATABASE AND VALIDATION RESULTS**

Figure 1 shows the LGN3 database of a part of The Netherlands and the LGN3 classes. The LGN3 database contains 39 classes in grid cells of 25 x 25 meter. The site specific accuracy assessment showed that for most non-agricultural LGN3 classes the minimum required classification result of 90% and 70% for respectively level I and level III classes is amply reached. Visual comparison with topographic maps and aerial photographs showed that the classification results of the land cover classes which are less important and/or show little spectral variability and the land cover classes that are (partly) copied out of other high quality databases, are generally good. However, locally large deviations may occur.

In general, the site specific classification accuracies and reliabilities of most agricultural crops in the LGN3 database are high (> 70%), unless the total acreage of a crop in an ‘agricultural region’ or the number of reference pixels is relatively low. In the former case small patches of a crop scattered over the region are easily overlooked during the classification, while in the latter case the reference pixels do probably not represent a representative sample. The classification accuracy and reliability of ‘Other agricultural crops’ are, generally, low. This is caused by the relatively small plots and the large spectral and temporal variability of the crops.

Comparison of the acreages in the LGN3 database with the acreages in the CBS Agricultural Statistics (non-site specific accuracy assessment) for all ‘agricultural regions’ shows, in general, a high agreement (i.e. deviations between corresponding acreages in both databases are less than 15%) for most agricultural crops (Wit et al., 1999). However, for the class ‘Other agricultural crops’ large deviations between the corresponding acreages in both databases were found. This confirms the results of the site-specific validation, indicating that ‘Other agricultural crops’ are difficult to classify because of their large spectral variability.

5 **APPLICATIONS**

The main users of the LGN database are national and regional governmental agencies and it has frequently been used for different purposes in the fields of environmental protection, water management, nature conservation and physical planning on regional and national scales. Mostly, the LGN data are combined with other digital geographical information, such as soil type, water-table, the occurrence of seepage, meteorological data, and the application of animal manure, fertilisers and pesticides. Underneath some main applications of the LGN database are summarised.

**Physical planning.** The LGN database can provide an overview of the land cover in a specific area, in contrast to the CBS Agricultural Statistics, which are available for administrative units only and therefore have a low spatial resolution. The LGN database has been used to allocate soil and groundwater protection areas, to design monitoring networks for soil and groundwater quality and to determine the relation between land use/cover and soil and groundwater quality (e.g. TAUW, 1994 and Van Drechte et al., 1996). LGN has also been used for an inventory of land cover/use in nature development and groundwater extraction areas.

In the early stages of planning of new highways and railways very detailed information is generally not necessary. Alternative routes need only to be evaluated roughly with regard to different environmental aspects. The LGN database is often used to provide land use information in such early assessment studies.

**Environmental protection.** The effect of (proposed) measures on the quantity and quality of groundwater and surface water are studied regularly, using regional water and solute transport models. In these models, evapotranspiration, groundwater recharge and irrigation are often strongly related to land cover, just like nutrient loads and uptake, and the
Figure 1. LGN3 database and legend

Legend:
- Grass
- Maize
- Potatoes
- Beets
- Cereals
- Other agricultural crops
- Greenhouses
- Orchards
- Bulb cultivation
- Deciduous forest
- Coniferous forest
- Fresh water
- Salt water
- Continuous urban area
- Built-up in rural area
- Deciduous forest in urban area
- Coniferous forest in urban area
- Built-up area with dense forest
- Grass in built-up area
- Bare soil in built-up area
- Main roads and railways
- Buildings in rural area
- Salt marshes
- Beaches and dunes
- Sparsely vegetated dunes
- Vegetated dunes
- Heathlands in dune areas
- Heathlands
- Raised bogs
- Forest in raised bogs
- Miscellaneous swamp vegetation
- Reed swamp
- Forest in swamp areas
- Swampy pastures in peat areas
- Herbaceous vegetation
- Bare soil in natural areas
- Heathlands with major grass influence
- Heathlands with minor grass influence

Kilometers
1:500000
distribution of nutrient applications and uptake over the year. Atmospheric deposition and the application of pesticides are also often related to land cover. The required information on land cover for these models has often been derived from the LGN database (e.g. Schouwmans et al., 1990; Thunnissen et al., 1992a, Reijerink and Breeuwsma, 1992; Vermulst, 1993; Querner, 1993; Van Walsum et al., 1996; Van Walsum and Veldhuizen, 1996; Van der Bolt et al., 1994; Meinardi, 1994 and Gehrels, 1995). In The Netherlands pollution from diffuse sources (e.g. application of pesticides and manure surpluses) threatens soil and groundwater quality. Regional water and solute transport models which can predict changes in soil and groundwater quality as functions of different pollution scenarios have been developed (e.g. Schoumans et al., 1990). The integrated effects of land cover, climate, soil, groundwater fluctuation, and pollution input can be calculated by automatic integration of these regional transport models with databases on land cover (LGN), soil, climate and input of pollutant., using a GIS. Using this system, different strategies aiming to control pollution intensity can be quickly evaluated.

In the early eighties, pesticides were found in drinking water extracted in the Drentsche Aa catchment area (province of Drenthe). The main causes of these contaminations were spray drift, surface run-off, and spilling when water was being pumped for spraying. To prevent this, alternative filling places were constructed so that farmers need not use streams and ditches for their water supply. Because pesticides are mainly used on arable land, the sites of these filling places were selected, using the LGN database in order to minimise the distance between the fields and the filling places.

**Water management.** During the winter periods of 1993 and 1995 extremely high water levels were reached in the large rivers in the Netherlands, due to excessive rainfall and thaw in the upper parts of the catchment. In 1993 the high water levels caused, particularly in the central river area ‘Betuwe’, inhabitants to be evacuated due to the risk of flooding. As a result of these near flooding events there was a renewed political interest for river flow modelling, including the occurrence of inundation. A key aspect in modelling the flow of water is the roughness of the surface, which effects stream velocity. Different land use types have different roughness coefficients The LGN3 land cover database has been used to define the spatial distribution of the surface roughness coefficient.

6 COMMERCIAL IMPLEMENTATION AND FUTURE DEVELOPMENTS

The LGN3 database has been implemented on a commercial basis. The cost of production of the LGN3 database amounted to Dfl 650.000,- (i.e. circa Dfl 15,50 per km²), while the earnings from sale of the LGN3 database amount to circa Dfl 900.000,- (situation 2000). In practice, the LGN3 database, once being available, appears often to be used for all kinds of unintended applications. An accurate estimation of the benefits, however, is troublesome. Information from the LGN3 database is mainly used by governmental agencies which are responsible for policies in the fields of environmental protection, water management, nature conservation and physical planning on regional and national scales. The effects of using the LGN3 database on the quality of the pursued policy are difficult to assess.

The chance of maintaining commercial implementation of the LGN database in (near) future is determined by the need of land cover/use data and the cost of gathering these data. In this framework the LGN database has to compete with other available nation-wide digital land cover/use databases. Evaluation of different databases showed that the LGN3 database distinguishes itself favourably from other databases, especially concerning cost, thematic classes, timeliness and data processing. In practice it proves that the different databases do not replace each other. Although there is some overlap, the databases largely supplement each other. In practice, combined use of different databases may give a surplusvalue to the separate databases.

The classification result of agricultural crops could be improved considerably by performing a field-based classification (Thunnissen, 1999). A large part of the field boundaries can be obtained from the Top10-vector. Because automatic segmentation techniques are not yet operational, the remaining field boundaries have to be obtained by interactive interpretation and classification and on screen digitising. By retaining the added field boundaries, up-dating of the LGN database can be accelerated considerably in future. Other future developments that are being studied are raster to vector conversion of the LGN database, linking of the LGN database with the Top10-vector and adapting the different LGN databases (concerning legend and geometric and thematic accuracy) to enable monitoring.

REFERENCES


