Ball, D. F.; Radford, G. L.; Williams, W. M.. 1983 Land characteristic data banks developed from map-derived material. In: Fuller, R. M., (ed.) Ecological mapping from ground, air and space. NERC/ITE, 28-38. (ITE Symposium, 10).

LAND CHARACTERISTIC DATA BANKS DEVELOPED FROM MAP-DERIVED MATERIAL

D F BALL, G L RADFORD AND W M WILLIAMS
Institute of Terrestrial Ecology, Bangor Research Station, Bangor

ABSTRACT

Ecological analysis and interpretation require the availability and accessibility of information about the physical environment. In Britain, such information is available nationally as maps or statistics but at a variety of scales and detail, and not in a convenient form for rapid quantitative and comparative reference. A common denominator between the map sources is the National Grid which provides recording cells appropriate to different levels of study. The grid cell approach to data extraction, storage, analysis and retrieval has been used in 2 projects at the Institute of Terrestrial Ecology, Bangor, which are dealt with in this paper. The first, a national land characteristic data set, covers Great Britain at the 10 x 10 km grid cell scale. The second used 0.5 x 0.5 km grid cells as the basis for land classification in a district study of upland land and its relationship to the distribution of moorland and grassland vegetation and to land use history.

INTRODUCTION

Ecological mapping could be defined in a limited sense as being concerned only with plant and animal species and communities. More correctly, it should also include land characteristics of the physical environment, because an understanding of ecosystems is impossible without environmental knowledge.

In many detailed ecosystem investigations, there is, of course, no substitute for the collation of new and existing land data specific to particular sites. No general-purpose data bank could provide all that is needed. Nevertheless, data banks of physical environmental information, with mapping facilities, can have important roles to play. They allow the initial selection of potential areas within which ecosystem studies should be located, and ensure comprehensive assessment of the possible options for such studies; they give, by direct correlation between recorded land characteristics and vegetation or fauna distribution, or by correlation of land classes and such distributions, a means of relating biological distributions to the physical environment; and they assist resource evaluation assessments at scales appropriate to the scale of the data set by giving statistical and map information on land character in a standard format. A natural resource information system is an 'integrated approach to collection, storage, manipulation, dissemination and use of resource data' (Mead 1981). Land characteristic data banks are a key element in such resource information systems.

There is an arguable case that the most desirable and effective support to ecological mapping that can be given by computer data-handling systems requires maximum accuracy in recording digitized data sets which, through map outputs, can allow retrieval of the exact locations of the input information. To supply such support requires sophisticated collection, storage and retrieval systems involving a major commitment of time, money and effort. However, it must be borne in mind that the original maps will always be available to return to when the exact locations of features are required, that features can change and hence maps become out of date, that some information sources are not of high precision, and that maps at any particular scale are selective in what they can include.

Our thinking with regard to land data sets for the Institute of Terrestrial Ecology (ITE), at this time, is that it is preferable to accept the use of less sophisticated methods, with less accurate but more rapid quantitative measurement of attributes from their source maps. Less sophisticated methods necessarily involve storage of data recorded for cells rather than for exact locations, the cells most conveniently being grid squares of the Ordnance Survey (OS) National Grid. This approach has allowed reasonably comprehensive land characteristic data sets for different purposes to be compiled relatively cheaply and quickly.

Two variants of the grid cell approach to land data collection and handling have been followed by ITE in recent years. The first (see Bunce *et al.* this symposium) uses a sampling procedure to record a wide range of attributes for a sample of grid squares. The second variant favours complete coverage, with data being recorded for all cells in the area of interest. Considered against the sampling approach, total coverage of all grid squares for any set of recorded attributes is obviously slower. However, what is recorded is factual for each square, within the limits of sources and the recording methods, rather than being probability assessments for a high proportion of them.

Pilot exercises at ITE Bangor used comprehensive data sets for a provisional land classification of the uplands of England and Wales at the 10 x 10 km scale, and for a district assessment in Snowdonia of land 'quality', in relation to agricultural and forestry potential, at the 0.5 x 0.5 km scale (ITE 1978). Subsequently, the 2 land data sets, which are discussed in this paper, have been produced. A land characteristic data bank for Great Britain has been assembled as a general-purpose source of information appropriate to national and regional scales of study. It is intended that this will remain valid and updatable over some years. At a larger scale, a land data set and a land classification based on this have been used in a study of vegetation and vegetation change in 12 upland areas.

NATIONAL LAND CHARACTERISTIC DATA SET

The National Land Characteristic and Classification (NLC) general-purpose data set at present stores 121 quantitatively determined attributes for each of the 2826 10 x 10 km grid squares recorded as containing land in Great Britain. The 7 land characteristic categories are summarized in Table 1. A full list of attributes, with details of their sources and methods of recording, will be included in a description of this data set it is planned to publish in 1982. They have been determined from maps of the most appropriate available scale, except for agricultural land use and population data which have been drawn from statistical sources.

Among the facilities offered by this data set are:

1. The presentation of standard statistical information on land characteristics for any 10 x 10 km grid square or combination of squares.

2. The identification, on maps and lists, of grid squares that meet a given specification. Because there is complete cover of the country, a search for grid squares with particular characteristics can be comprehensive rather than limited by personal knowledge. This may be particularly important for initial selection of areas justifying more detailed consideration as possible study locations.

3. The opportunity to relate recorded land characteristics, singly or in combination, to ecological distributions that have also been recorded at this scale.

4. The ability, through direct display and tabulation, supplemented by classification analyses, to interpret the physical environmental range in Great Britain or its regions in ways that may prove helpful to illustrating or evaluating land resources in relation to ecology, economics and planning at national and regional scales.

TABLE 1 Attribute categories included for each grid square in the ITE 10 x 10 km National Land Characteristic Data Set

Physiography:	Extent of land and sea Extent of altitude classes Altitude range Relative relief River frequency Extent of freshwater Length of coastline
Climate:	Extent of annual rainfall classes Short-term seasonal rainfall (from 3-year averages) Short-term seasonal air temperature Short-term seasonal sunshine Short-term seasonal windspeed
Geology:	Extent of stratigraphic units Extent of bedrock lithology categories Extent of surface geology categories
Soils:	Extent dominated by 8 major soil categories
Topography:	Settlement frequency Road frequency Railway frequency Total population
Land use:	Extent of farmland Extent of forest and woodland Extent of urban land Extent of individual agricultural crops Farm labour input
Agricultural land quality:	Extent in Agricultural Land Classification categories

As a small example of output statistics, Table 2 relates total values from the data set, drawn mainly from 1:250 000 OS or Ministry of Agriculture, Fisheries and Food (MAFF) maps, to other available measures of the same attributes. Total land areas recorded conform closely to official statistics, giving general confidence in the relatively rapid point-count method used for area assessment. Discrepancies between values for water body areas are attributable to the 1:250 000 source maps not showing small water bodies and not yet having included recent major reservoirs such as Kielder, Rutland and Brenig. The under-estimate of 'non-agricultural' land in the Agricultural Land Classification categories is because the MAFF area totals quoted (Agricultural Development and Advisory Service 1974) were taken from the original 1:63 360 maps produced, but the figures in the data set came from subsequently published 1:250 000 maps. In these, smaller areas of 'non-agricultural' land were, as a matter of policy, omitted and became part of the surrounding land of classes 1-5. In any land data set derived from maps, the nature and inevitable limitations of the source material must be clearly specified and understood.

TABLE 2 Areas for land use attributes (km^2) and human population, calculated from the National Land Characteristic (NLC) Data Set, and compared, as percentages, with equivalent values measured by other agencies

	Great Britain		England & Wales		Scotland		Comparable data - Reference
	NLC Value	% equiv	NLC Value	% equiv	NLC Value	% equiv	
POPULATION	54 030 700	100.1	48 747 900	100.0	5 282 800	101.0	COI 1980
LAND:							
Total land	231 410	100.4	152 083	100.2	79 327	100.7	COI 1980
Freshwater	2 026	82.4	499	58.4	1 527	95.1	COI 1980
Agricultural Class 1	-	-	3 701	104.2	-	-	ADAS 1974
Class 2	-	-	19 122	107.1	-	-	ADAS 1974
Class 3	-	-	65 106	108.3	-	-	ADAS 1974
Class 4	-	-	26 429	108.6	-	-	ADAS 1974
Class 5	-	-	18 012	105.1	-	-	ADAS 1974
Non-agricultural	-	-	18 193	63.2	-	-	ADAS 1974
Farmland	170 557	95.7	114 640	98.5	55 917	90.3	Coppock 1974 & 1976
Barley	22 222	99.4	18 950	99.5	3 272	98.5	Coppock 1974 & 1976
Total crops	45 913	99.1	40 517	99.9	5 396	93.2	Coppock 1974 & 1976

NOTES:

NLC Data sources

- 'Land' and 'freshwater' from 1:250 000 OS maps; ALC categories from MAFF 1:250 000 maps (total land in England and Wales as measured from these in the ITE data set = 150 563 km^2, 99.0% of the land area as measured from OS maps).
- Population from Office of Population Censuses and Surveys statistics.
- 'Farmland' and 'crops' from statistics provided by the Edinburgh Computing Centre from MAFF and Department of Agriculture and Fisheries for Scotland data used by Coppock to prepare his Agricultural Atlases.

Figures 1-4 are examples of output maps from the line printer at ITE Bangor (more sophisticated output facilities are available on-line from other centres). They show a possible type of search pattern: identification of upland areas which might have unexploited forestry potential. An approach to such a search could use the following assumptions:

1. that land between 244 and 427 m was particularly desirable (agriculture being assumed to have priority below 244 m, while land >427 m is assumed to be largely unsuitable for commercial forestry);

2. that a significant extent of podzol soils would be advantageous (such soils, in moorland terms, can be thought of as at least relatively well-drained, they are found typically on moderate, and hence readily plantable, slopes, and, because they are generally unimproved, they are of relatively low financial value);

3. that no significant forestry enterprise was already present.

At chosen levels of these specified characteristics, the altitude category (Figure 1) occurs in 16.5% of the total 10 x 10 km grid squares in Great Britain, while the soil category (Figure 2) occurs in 14.6% of them. Figure 3 shows that both specified characteristics are present in 6.1% of the squares (173 inland squares, occupying 17 300 km^2, or 7.5% of the land area of Great Britain). Of these 173 squares, 74 (Figure 4) presently have less than 10% forest. From such a search, accepting the assumptions made, upland areas of high potential for extensive forestry development, and hence of possible land use conflicts, have been located and found to occupy on a grid square basis 2.6% of Britain (the land in the squares being 3.2% of the land area of Britain). Further analysis within the main data set could explore other relevant factors for these squares, such as their present intensity of agricultural use assessed by stock density, or their climatic constraints assessed by average rainfalls and windspeeds. A supplementary data set will be able to identify possible administrative constraints, such as National Park or other designations. Surviving grid squares from such a reduction process could then be considered in more detail from source maps and in the field.

All the data that have been included in this 10 x 10 km land data bank already exist but it provides a co-ordinated, reasonably accurately quantitative information baseline that eliminates the need to search a range of sources on an *ad hoc* basis for each issue and study. The data set is intended to be open-ended and has a variety of uses. Particular applications may require additional data in existing or new categories which may either be retained as part of the main set, or discarded after their specific use. The permanent data can be upgraded by substituting more accurately measured attributes or by adding attributes from newly available sources.

There are, of course, inevitable constraints arising from the cell approach and the recording methods used. All that is possible from a grid cell output map is to know that some features (eg a particular altitude range, woodland, a soil class, a population level) co-exist within a grid square, not whether they coincide at a particular location. Such locations must be sought by following up the pointer provided by a grid cell search with a return to the source maps. Common combinations of factors cannot be so substantially reduced in their possible locations as can scarcer ones, but even in such uses it can be helpful to determine the possible geographic range over which these features occur. A second constraint is that grid square boundaries do not match administrative or natural geographic boundaries. Statistics and maps provided for administrative areas such as Counties or National Parks, or for geographic regions, must be

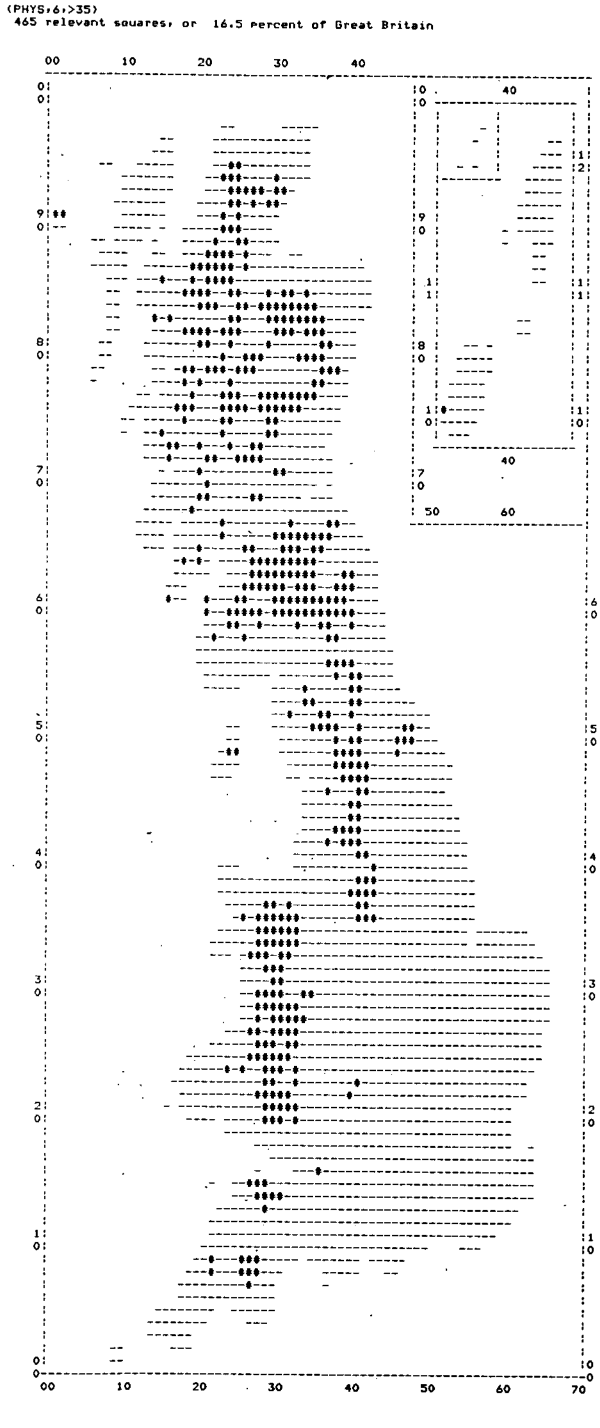


Figure 1 *Line printer maps of grid squares meeeting an altitude specification*

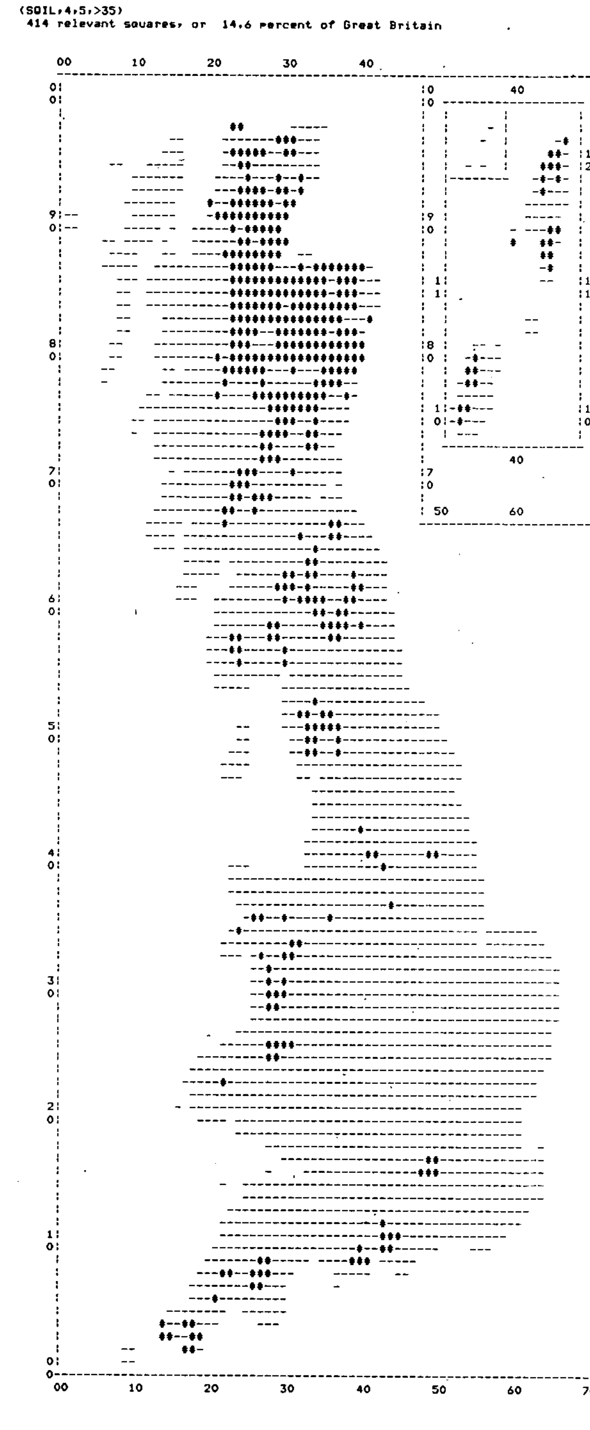
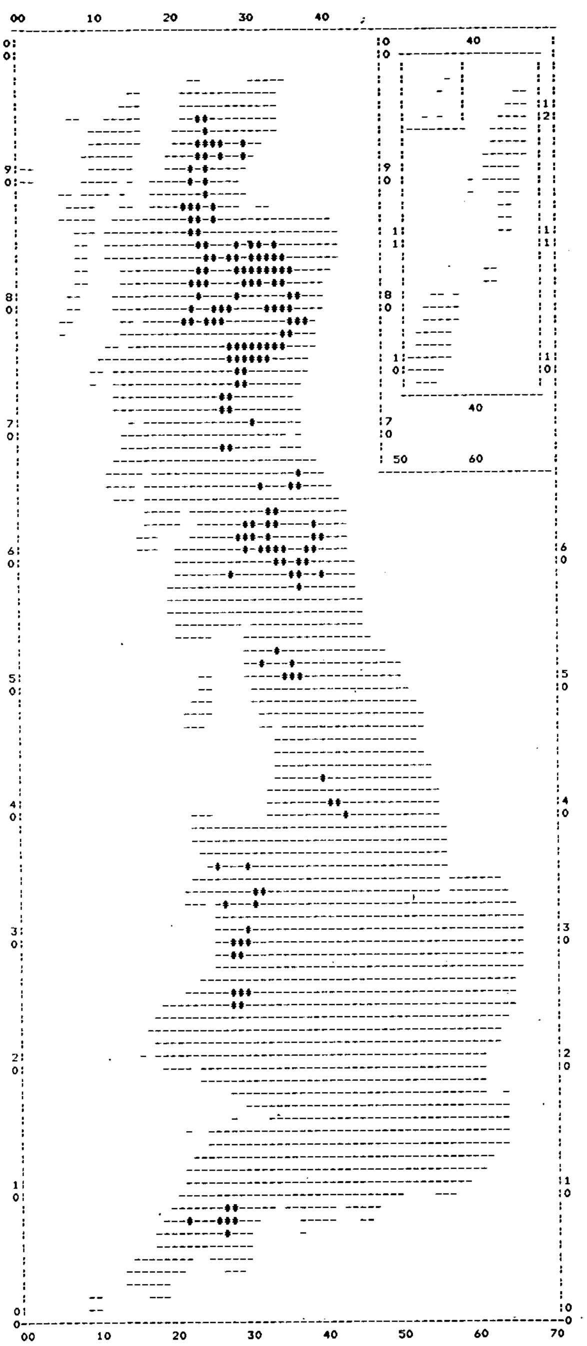


Figure 2 *Squares meeting a soil specification*

(PHYS,6,>35)AND(SOIL,4,5,>35)
173 relevant squares, or 6.1 percent of Great Britain



♦ SQUARES WITH MORE THAN 35% LAND OVER 800 FT, AND MORE THAN 35% DOMINATED BY PODZOLS

Figure 3 Squares meeting both an altitude and a soil specification

(PHYS,6,>35)AND(SOIL,4,5,>35)AND(USE,5,<10)
74 relevant squares, or 2.6 percent of Great Britain

♦ SQUARES WITH MORE THAN 35% LAND OVER 800 FT, MORE THAN 35% DOMINATED BY PODZOLS, AND WITH LESS THAN 10% WOODLAND COVER

Figure 4 Squares meeting altitude, soil and forestry use specifications for the closest match of grid squares to the administrative or physiographic boundary. A third constraint is that information which can be retrieved and specifications which can be framed must be confined to the ranges recorded. For example, as the altitude ranges used in this data set include 0-61 m and 62-122 m (from contours transformed from feet), it is not possible to ask how much land below 100 m is in a square, or which squares have more than 80% of their area below 100 m. The data set cannot realistically cover all possible requirements, so these must be tailored to what is available.

UPLAND STUDY AREA DATA SET

The approach of measuring land characteristics of grid squares from existing map data and storing them in a computer in a form suitable for statistical and map outputs and for classification analyses has been applied at a district scale in a contract study, 'Ecology of Vegetation Change in Upland Landscapes', carried out by ITE for the Department of the Environment (Ball *et al.* 1981a, b). A data set, drawn mainly from 1:25 000 scale OS maps and covering 31 physiographic, topographic and rainfall attributes for 2977 0.5 x 0.5 km grid cells, has been used to analyse the land characteristics and permit a land classification of 12 parish areas spread through the uplands of England and Wales. The land analysis was part of an ecological study of grassland-moorland vegetation and its relationship to environment, recent land use history, and management.

The computer-stored, quantitative, land characteristic data set allows the retrieval of schematic grid cell maps to illustrate aspects of the environment of the study areas, and enables statistical comparisons to be made between areas. To give a simpler framework for discussion of contrasts and similarities between the scattered upland parishes and for interpretation of the environmental influence on vegetation distribution, a land classification was carried out using Indicator Species Analysis (ISA) (Hill *et al.* 1975) on a transformed set of land characteristics in a presence/absence form. ISA is a computer analytical method which successively divides a population of individuals to give 2, 4, 8, 16, etc, classes. It assesses the balance, in each individual, of a range of measured factors, in order to group together those individuals with the most similarity in selected significant properties (the 'Indicator Species' of the title, chosen in the analysis to give the clearest distribution between the 2 classes into which the whole or part of the population is divided at each stage of the analysis). The 'populations', in the cases referred to here, are a number of grid squares for which a series of land attributes are the measured factors, or they can be a series of plant records from sites, in which lists of plant species present are the measured factors, as in the original applications of the method.

Seven land types in 3 groups resulted from ISA, the average land character of which can be quantitatively defined in relation to the study areas as a group, by reference to the original map-derived data. The *'hill'* land group consists of 3 types, named as steep hill, hill and high plateau. All are dominated by altitudes above 427 m combined with low levels of settlement as shown by buildings, roads and field boundaries, but they are characterized by differences in their average relief and slopes. The *'upland'* land group, generally dominated by altitudes between 244 and 427 m, also contains 3 land types: steep upland, upland and upland plateau, again distinguished by their relief and slopes. All have relatively intensive settlement, particularly so in the upland land type. The seventh land type *'upland margin'* is dominated by low to moderate altitudes, moderate relief and low to moderate slopes, while having a relatively high settlement level. Within these classes determined by physiography and settlement pattern, rainfall phases can be separated at another level of classification.

Vegetation records were obtained from approximately 100 sites in grassland and moorland situations in the study areas, from which a vegetation classification provided 16 classes in 4 groups: improved pastures, rough pastures, grassy heaths and shrubby heaths. By treating the sites as a set of points which sampled grid squares classified in the 7 land types, the association between land type, as a simplification of a range of environmental factors, and vegetation class and group distribution has been used to interpret the present situation and to support predictions of possible changes in identified sectors of the study areas. These associations are discussed at length in the reports referred to (Ball *et al.* 1981a, b). Table 3 shows the broad association between vegetation groups at recorded sites and land types of the grid cells in which the sites fall. Grassy heaths are a particular feature of the landscape in steep upland and steep hill land types, while shrubby heaths are most prominent in the hill and high plateau land. Improved pastures play virtually no role in the hill land, while rough pastures, though not dominant in any land type, are particularly represented in upland margin, upland and steep upland land. Historical evidence of land use change between farmland and moorland over the past 200 years shows these 3 land types to have been the main focus of such change, 75% of all grid squares in which such change has occurred falling in 3 land types which occupy 41% of the total parish areas.

TABLE 3 The proportion (%) of different vegetation groups associated with 7 land types in 12 upland parishes, as recorded at *c.* 1000 sites

Land type	Vegetation group			
	Improved pastures	Rough pastures	Grassy heaths	Shrubby heaths
Steep hill		13	43	44
Hill		3	23	74
High plateau	2	5	21	72
Steep upland	19	23	42	16
Upland	44	25	9	22
Upland plateau	51	10	12	27
Upland margin	48	28	13	11

DEVELOPMENTS

Ecological mapping, as we have broadly defined it, involves both physical and biological data. At the 10 x 10 km scale, the biological component derives from the species distributions held at the Biological Records Centre (BRC), ITE Monks Wood. The availability of physical data for the same cell size as that of the BRC records provides an opportunity to re-address a number of questions left unanswered during early attempts to analyse the species distributions. The physical data set can be used to illustrate the existence, strength and geographical consistency of expected field correlations, for example between soil or geology categories and calcicole or calcifuge species. Further investigations can be made of less obviously related distributions with a view to identifying ecologically valid groupings of species at this broad scale. These groupings may be retained alongside the land characteristic data bank as associated material for subsequent retrieval and analyses.

A land characteristic data set on a grid cell basis at the 10 x 10 km scale has both a scientific value and a potential for application in strategic planning. Interest is consistently being expressed in the availability of data, and guidelines for their use, as aids to decision making in the planning process. The provision of 'ecological maps' of biological and environmental characteristics, and of land statistics, is an essential major component in resource assessment and development planning. Raster data, derived from gridded maps and satellite imagery, and vector data, derived from conventional maps or aerial photographs, are both likely to make contributions in this field.

Another intended development of the cell approach to mapping is to determine how classifications, to be derived from the National Land Characteristic and BRC data sets at the 10 x 10 km cell size, may be related to larger scales. The 0.5 x 0.5 km grid square data from upland areas could be extended to cover a sample of 10 km squares. It would then be possible to see whether 10 km squares, of a given national land class, consistently have a particular pattern of land types classified at the more detailed scale. There is also the opportunity to relate classes derived from the comprehensive 10 x 10 km land data set to those recognized in the 1 x 1 km data set of the Ecological Survey of Britain (see Bunce *et al.* this symposium).

From the alternative classification options which will be explored in this national data set, it is hoped that useful stable classes will result, giving the possibility of using stratified sampling to provide predicted values for additional ancillary attributes. This may be justified because predicted values are sufficient for the purposes in hand, or because data can only be obtained from a sample of squares. It must be borne in mind that sampling would not help in giving factual data for non-sampled cells, and that the sample population itself will not be appropriate for relating to other versions of a land classification.

In addition to the recorded quantitative data, and to grid square classifications that will simplify the range of land variation, and which will be stored as a second group of associated data, supplementary information likely to be used in conjunction with the main data set will be recorded for each 10 x 10 km cell. This third data group will provide an index to relevant scientific material, listing, for example, whether climatological stations, Forestry Commission census plots, National Nature Reserves, ITE or other Natural Environment Research Council study sites occur within a cell. This level of information will also include major conservation or landscape designations that place land use constraints on the square.

The view that a grid square approach to mapping simply displays statistics and cannot produce a 'real map' can be argued. However, grid square mapping has much to commend it as a means of presenting data in a readily appreciable form. The factual basis for conventional mapping at a level of accuracy appropriate to even medium scales for Britain does not yet exist over the full range of factors summarized in the national land characterization project. As such data become available they will increase the impetus to develop large digitized data bases, but the cell approach will retain a continuing, positive role in ecological investigation and mapping. At large scales the cell approach has a potentially direct link with remote sensing activities. Grid squares of 100 m or less, corresponding to squared pixel transformations of satellite data, may be used as sample units for ground truth, and as recording units for supplementary field survey data. The potential value of this combination in ecological survey and mapping has considerable appeal.

REFERENCES

AGRICULTURAL DEVELOPMENT AND ADVISORY SERVICE. 1974. *Agricultural land classification of England and Wales: presentation and exhibition.* Pinner: MAFF/ADAS Land Service.

BALL, D.F., DALE, J., SHEAIL, J., DICKSON, K.E. & WILLIAMS, W.M. 1981a. *Ecology of vegetation change in upland landscapes. Part I: General synthesis.* (Bangor occasional paper no. 2). Bangor: Institute of Terrestrial Ecology.

BALL, D.F., DALE, J., SHEAIL, J. & WILLIAMS, W.M. 1981b. *Ecology of vegetation change in upland landscapes. Part II: Study areas.* (Bangor occasional paper no. 3). Bangor: Institute of Terrestrial Ecology.

CENTRAL OFFICE OF INFORMATION. 1980. *Britain 1980: an official handbook.* London: HMSO.

COPPOCK, J.T. 1974. *An agricultural atlas of England and Wales.* London: Faber.

COPPOCK, J.T. 1976. *An agricultural atlas of Scotland.* Edinburgh: J. Donald.

HILL, M.O., BUNCE, R.G.H. & SHAW, M.W. 1975. Indicator species analysis, a divisive polythetic method of classification, and its application to a survey of native pinewoods in Scotland. *J. Ecol.*, **63**, 597-613.

INSTITUTE OF TERRESTRIAL ECOLOGY. 1978. *Upland land use in England and Wales.* (CCP 111). Cheltenham: Countryside Commission.

MEAD, D.A. 1981. Statewise natural-resource information systems - A status report. *J. For.*, **79**, 369-372.