Chapter

Scholefield, Paul; **Norton, Lisa**; **Rowland, Clare**; **Morton, Daniel**; **Henrys, Peter**. 2012 Developing a model of linear ecological networks for Great Britain. In: Dover, J.W., (ed.) *Hedgerow Futures: Proceedings of the 1st International Hedgelink Conference.* Published for Hedgelink by the Tree Council, 37-44

Developing a model of linear ecological networks for Great Britain

Scholefield, P.A., Norton, L.R., Rowland, C., Morton, D., Henrys, P.

Centre for Ecology and Hydrology, Lancaster Environment Centre, Library Avenue, Bailrigg, Lancs, UK, LA1 4AP.

lrn@ceh.ac.uk

ABSTRACT

The importance of woody linear features for; maintaining connectivity between habitats, providing landscape resilience and enhancing biodiversity in otherwise impoverished landscapes are widely recognised. Despite the acknowledged importance of hedges and lines of trees in the wider countryside they are often excluded from models of landscape function due to the difficulties of acquiring relevant data about their extents and locations. We present an approach currently being trialled to produce a linear product for Great Britain which uses a combination of datasets and a classification to predict the presence of hedges/lines of trees and 'other' categories of linear features. The datasets used include; a digital terrain model (Nextmap), the spatial framework for Land Cover Map 2007 and field survey data from Countryside Survey 2007. The ITE/CEH landclasses are used as a key classifier. Early outputs from the model are presented and discussed and the potential uses of a linear land cover map are explored.

INTRODUCTION

Data on the extent and condition of our natural resources is vital to ensure their effective management into the future (MEA, 2005). This can be done at any number of scales but to understand resource management at a national level it is important to have access to national data (NEA). One of the most effective ways to collect data relating to habitat types and their extents is through the use of satellite imagery (Kerr, 2003). Another method is to use stratified sampling which is representative at a national level, as is used in the Countryside Survey (Carey, 2008).

Land Cover Map is now in its third iteration and the data has been widely applied in, for example, urban planning, carbon accounting and flood risk modelling (Morton (2011). The pixel based habitat classification approach used in LCM does not however lend itself to assessing the presences and extents of linear landscape features such as hedges, lines of trees or walls due to their narrowness. In contrast, linear landscape features are an important part of the field mapping exercise included in the Countryside Survey, during which they are surveyed in detail in terms of length, type and to some extent, condition. CS uses a number of representative sample squares to provide national estimates of the lengths of these different types of linear features in the countryside (Petit et al., 2003). Whilst estimates based on the same approach over time provide useful indices of change for policy makers and essential information for reporting against BAP Priority Habitat targets[1], they do not provide valuable location specific information, except for in the actual squares in which CS takes place. A third national dataset, Nextmap provides digital terrain mapping for the entire UK surface, indicating the height of features and land parcels above ground height. Data is at relatively coarse resolution but coverage is comprehensive. These datasets all aim to provide national coverage, but have different strengths and weaknesses in relation to provision of information of linear landscape features. This work explores the potential benefits to be gained from their integration.

Linear landscape features are an integral part of the countryside in GB as well as in other temperate countries worldwide (Barr C.J., 2001). They may consist of a range of different feature types incorporating; walls, hedges, lines of trees, banks, dykes, fences, etc, either singly or in combination. Their management often depends as much on tradition (Antoine, 2001; Deckers et al., 2004; Tenbergen, 2001) as much as on practical issues (Staley et al., 2012) and may vary between locations from regional to local scales. In GB, woody features consisting of managed hedges or lines of trees are widespread and ecologically important landscape features. Countryside Survey includes tight definitions of such feature types in order to enable differentiation between them at reporting, and accurate assessment of the extent of change between feature types, e.g. from managed hedge to unmanaged line of trees. Both feature types, however, provide important contributions to landscape diversity and have, for the purposes of this work been grouped into one category of woody linear features (hereafter referred to as hedgerows).

Increasingly the original purposes of hedgerows as either stockproof or ownership boundaries have lessened in importance with alternative, relatively cheap to maintain, fencing options available to farmers. Whilst the provision of shelter to stock almost certainly remains an important (although diminishing) function of hedgerows in the landscape, recognition of their importance as semi-natural habitat spanning increasingly sterile agricultural landscapes is growing (Natural Environment White Paper 2012). By providing a refuge for a wide range of taxa effectively eliminated from the majority of fields as a result of agricultural improvement (Smart et al., 2006) hedgerows help to maintain functioning agroecosystems in which predators of crop pests, pollinators and pollen producing species all play their roles (Pocock et al., 2012). Their presence in the landscape forms part of our cultural heritage with many National Character Areas[2] including descriptions of hedgerows and field shape and pattern as defining characteristics. Their importance to game species is also a measure of their cultural importance as is their use for

[1] http://ukbars.defra.gov.uk/archive/plans/national_plan.asp?HAP=%7BBC11363F%2D7C31%2D4CC6%2DA5F3%2D497601778845%7D (Ancient or species rich hedgerows)

[2] http://www.naturalengland.org.uk/publications/nca/default.aspx

collecting produce such as fruit or fungi in the autumn. Their contributions to other ecosystem services is perhaps less well documented due to difficulties of measurement or relative ease of measuring the more obvious ecosystem components contributing to those services. For example, carbon or water storage is generally measured/estimated on the basis of blocks of land cover, e.g. grassland, woodland, bog. Hedgerows, as lines, rather than blocks of habitat tend to be considered as less important. However, their position within habitats may make them particularly important at local scales and their prevalence may mean that their impact is more substantial than expected.

Here we present the approaches taken for development of a spatially explicit national map of GB linear features which differentiates between the presence of hedgerows and all other features using a binary classification.

METHODS

Datasets

The Countryside Survey data provided us with an ideal dataset which could be used to train and test a linear model for GB 1km squares (as described below).

Countryside Survey 2007

Countryside Survey methodologies are well documented and complex (Firbank et al., 2003). Full methodologies for the most recent survey (2007) are available at www.countrysidesurvey.org.uk. Relevant methods are summarised in brief, below. In CS2007 data on habitat extent was collected using a digital field mapping system based on ESRI ArcGIS 9.2 (ESRI, 2006). Using the digital mapping system, field surveyors delineated and mapped areas of different habitat types, effectively converting the 1 km square to a mapped format, at a resolution of 20 × 20 m for areal features and recording all linear (>20 m in length) and point features present within sample 1 km squares. Surveyors were provided with a field handbook containing detailed definitions of linear features, including fences, walls, hedges and lines of trees. Data collected using the digital mapping system was entered into a database containing both spatial and attribute information for all linear features recorded in the sample squares.

The following datasets were used to provide variables for the linear model enabling extrapolation of linear data outside of CS sites:

NEXTMAP Great Britain™

NEXTMAP Great Britain™ is a digital terrain model (DTM) produced by Intermap Technologies for Norwich Union Insurance to assist with flood risk mapping. NEXTMAP data were generated by airborne survey using synthetic aperture radar (SAR), and single-pass interferometry (IfSAR) (Chiverrell et al., 2008). NEXTMAP digital elevation data were collected at a flight height c. 6500 m and the data are supplied at a 5m resolution. The NEXTMAP data was chosen because it is cheap and there is comprehensive coverage for the United Kingdom, whereas in 2007 the Environment Agency LiDAR dataset (which has higher resolution data) was available for less than 33% of the country.

Ordnance Survey Land-Form PANORAMA®

Land-Form PANORAMA® is a height dataset available as a set of contours with spot heights, breaklines, coastline, lakes, ridges and formlines with a 10 m contour interval. We used the gridded DTM with 50 m post spacing to provide us with altitude information.

Land Cover Map Spatial Framework

The OS MasterMap topography layer (OSMM) provides a highly detailed view of Great Britain's landscape including individual buildings, roads and areas of land. In total it contains over 400 million individual features. From this, the data relating to polygon objects (100 million) were used to create the spatial framework for the GB Land Cover Map (LCM) 2007. Since the resolution of OSMM is greater than that used for LCM (which uses 20x20m pixel satellite data) the OSMM was spatially generalised, removing unnecessary detail whilst retaining relevant details on location of boundaries (Morton et al. 2011).

ITE Land Classification 2007

The 2007 version of the ITE Land Classification was developed from earlier classifications which used environmental data from all 1km squares in GB to create a stratification based on underlying physical variables (Bunce et al., 1996). The stratification contains 46 classes or strata, distributed across Great Britain. Each stratum consists of areas with similar environmental characteristics.

Model variables

Linear data from Countryside Survey (see above) were attributed with variables from the above four datasets. Prior to use of the NEXTMAP 5m canopy dataset the data was prepared by masking out the Urban, Forest, Littoral and Sub-littoral Broad Habitats and all areas above 450m altitude (i.e. all canopy height data in these areas was set at zero). Raw data from the resulting 5m resolution dataset formed one of the variables used in the models. The remaining data were treated in two ways;

1) Aggregated canopy heights were calculated within a 15m buffer around each 5m resolution linear (total linear width 35m) – 35m resolution

2) Data were resampled at 25m resolution for maximum height associated with linear features

A number of variables were calculated using the altitude data from OS Land Form data. Curvature (a measure of how convex or concave a feature is) was derived using Arc GIS 3D analyst at 50m resolution from this data. All variables used in the final models are detailed in Table 1.

Model training and validation

A boosted binary classification tree (described in more detail below) enabling a mix of both classification and continuous data was used for the modelling approach. The CS dataset classified into: 1) hedgerows and 2) other linear features, was separated into two datasets which constituted an initial training dataset (70% of the data) and a testing dataset (30% of the data). The model was then asked to predict the output values for the data in the testing set. Model performance was assessed using % accuracy statistics for individual CS squares, but also the model produced mean square error statistics for each individual prediction**.**

Bagged decision tree classifiers

The model architecture employs bagged decision trees as a binary classifier based on its performance on CS2007 linear feature classes; 1) hedgerows and 2) other linear features. The decision tree is a machine learning classifier with a tree data structure. Classification decisions are the result of traversing from the tree root to a leaf node. Each non-leaf node employs a split variable and split value to determine the path of traversal to a leaf node, where each tree makes a final class assignment. Each leaf bases the assignment on prior probabilities from the remaining sample subpopulation at that leaf established during training. During training, the selection of *n* samples from the training set (of size *n)* with replacement constitutes a bootstrap sampling. The resulting classifier uses a majority vote of the individual trees in the ensemble. The model tested uses an ensemble of 30 trees with default parameters. By default, TreeBagger considers random features for each cut variable, allows a minimum of one observation per leaf node and employs the Gini's diversity index as an impurity measure. Other defaults include no pruning and using equal misclassification costs (Dube et al., 2012).

Model selection

Since the model's ability to predict the type of feature present should depend more on the important variables in the model and less on unimportant variables, this can be used to identify the important variables and test whether a smaller subset of variables can be substituted for the full set. For each linear feature, it is possible to permute the values of this feature across all of the observations in the data set and measure how much worse the mean-squared error (MSE) becomes after the permutation. This can be repeated for each feature, so that it is then possible to identify the importance of each different model variable for prediction.

This process was carried out with the original set of 13 variables input to the model and used to identify a subset of 6 (Table 1, **in bold**). Land class did not feature in the top 6 (or even the top 10) variables after testing of the 13 + landclass. However, when subsequently included with the 6 most important variables it proved to be more than twice as important as all other variables.

Hence, four models were implemented, two with and two without the land classification (table 2). After testing with the CS data, the 'best' model was then used to predict linear feature cover and type in areas outside of CS squares using the LCM spatial framework. Due to the large amounts of linear features which would need to be modelled at a UK scale and the considerable amount of data and computer processing power required to do so, initially this was trialled for specific areas of the UK with known high hedge densities. Indicative results are presented here.

RESULTS

The results for the four different models are shown in Figure 1. All models showed between 78% and 82% accuracy for predicting whether a linear feature in the CS testing dataset was either a type 1) hedgerows or type 2) other linear feature. The lowest % accuracy (78%) was for Model 1 with the 6 variables and no classifier, with model accuracy increasing with the inclusion of the classifier and increasing numbers of variables, so that model 4 had the highest % accuracy.

Predictions for linear features outside of CS squares using the LCM spatial framework are shown in figure 2. Figure 2 shows the model results for hedgerows for an area in Cornwall alongside LCM data for forestry in the same area

DISCUSSION

Early results presented here indicate that the approach developed may provide an excellent tool for mapping UK linear features. The high % accuracy for the CS testing dataset provides some confidence that extrapolations outside of CS squares are likely to be reasonably accurate, even with the use of a relatively coarse resolution dataset (NEXTMAP).

Next steps are to continue work to validate the model and understand where it performs well and where it performs poorly. The sheer size of the dataset means that it can be rather unwieldy to deal with and uses up a lot of computation power and time. A UK linear data set would involve modelling over 20 million individual features. Even with a simple binary classification this is a very extensive dataset.

A first step is to investigate, in detail, predictions for CS squares and examples from other 1km square datasets which have used the CS linear mapping methodology. The availability of detailed 'ground-truthing' datasets will allow for a much better understanding of any issues decreasing the accuracy of the model.

In the future it may be possible to refine the model using different higher resolution datasets. LiDAR (Light Detection And Ranging) is an optical remote sensing technology that can measure the distance to, or other properties of a target by illuminating the target with light, often using pulses from a laser. It has been used for a range of purposes relating to ecosystem studies (Lefsky et al., 2002) including predicting biodiversity from observed habitat variables (Muller and Brandl, 2009). LiDAR data is available for a number of the CS squares and can be used to test the potential for not only identifying the type of linear feature present but also some of its more detailed characteristics including height, width and species composition. It is also likely to be a more appropriate dataset than the NEXTMAP for identifying the presence of other important linear features such as stone walls or banks. However, the extent to which LiDAR data may be used for any future national linear product will depend on the extent of coverage of the UK and on access to the considerable computing power required to process such detailed data at a national scale.

Further work is required to ensure that any model of UK linears produced using this approach is robust and that uncertainties in the model are quantified. In addition, improvements in the availability of earth observation data like LiDAR will broaden the potential resolution of any linear product. However, it is clear that both the datasets used here and the approach developed show great potential for enhancing our ability to quantify the natural resources of the UK. Even a basic model of hedgerow extent and locations, such as developed here, could help to fill in key gaps in our understanding about landscape function in relation to ecosystem services production through their important role in, for example, carbon storage, the movement of water and diffuse pollutants or habitat provision for biodiversity and aesthetic enjoyment. Increases in spatial resolution of satellite derived data such as LiDAR point to potential future products which would enable application at local levels for use in strategic planning for sustainable landscapes.

REFERENCES

Antoine, A., 2001. From past to present: the changing functions of hedges in England and the West of France - Bocages, woodlands and enclosures, in: Barr C.J., Petit, S (Ed.), Hedgerows of the world: their ecological functions in different landscapes. IALE (UK), University of Birmingham.
Barr C.J., P., S, 2001. Hedgerows of the World, in: Barr C.J., Petit, S (Ed.), Hedgerows of the world: their ecological functions in different landscapes. IALE (UK), University of Birmingham.
Bunce, R.G.H., Barr, C.J., Gillespie, M.K., Howard, D.C., 1996. The ITE Land Classification: Providing an environmental stratification of Great Britain. Environmental Monitoring and Assessment 39, 39-46.
Carey, P.D.W., S.; Chamberlain, P.M.; Cooper, A.; Emmett, B.A.; Maskell, L.C.; McCann, T.; Murphy, J.; Norton, L.R.; Reynolds, B.; Scott, W.A.; Simpson, I.C.; Smart, S.M.; Ullyett, J.M., 2008. Countryside Survey: UK Results from 2007. http://countrysidesurvey.org.uk/

Chiverrell, R.C., Thomas, G.S.P., Foster, G.C., 2008. Sediment–landform assemblages and digital elevation data: Testing an improved methodology for the assessment of sand and gravel aggregate resources in north-western Britain. Engineering Geology 99, 40-50.
Deckers, B., Hermy, M., Muys, B., 2004. Factors affecting plant species composition of hedgerows: relative importance and hierarchy. Acta Oecologica-International Journal of Ecology 26, 23-37.
Dube, T., Raines, R., Peterson, G., Bauer, K., Grimaila, M., Rogers, S., 2012. Malware target recognition via static heuristics. Computers & Security 31, 137-147.
ESRI, 2006. ArcGIS Version 9.2. , Redlands, California, USA.
Firbank, L.G., Barr, C.J., Bunce, R.G.H., Furse, M.T., Haines-Young, R., Hornung, M., Howard, D.C., Sheail, J., Sier, A., Smart, S.M., 2003. Assessing stock and change in land cover and biodiversity in GB: an introduction to Countryside Survey 2000. Journal of Environmental Management 67, 207-218. Kerr, J.T., & Ostrovsky, M., , 2003. From space to species: ecological applications for remote sensing. TRENDS in Ecology and Evolution 18, 299-305.
Lefsky, M.A., Cohen, W.B., Parker, G.G., Harding, D.J., 2002. Lidar remote sensing for ecosystem studies. BioScience 52 (1), 19-23.
MEA, 2005. Millennium Ecosystem Assessment, Ecosystems and Human Well-being: Synthesis.
Morton, D., Rowland, C., Wood, C. Meek, L., Marston, C., Smith, G., Wadsworth, R., Simpson, I.C., 2011. Final Report for LCM2007 - the new UK land cover map., Countryside Survey Technical Report No 11/07, NERC/Centre for Ecology & Hydrology, p. 112pp.
Muller, J., Brandl, R., 2009. Assessing biodiversity by remote sensing in mountainous terrain: the potential of LiDAR to predict forest beetle assemblages. Journal of Applied Ecology 46, 897-905.
Petit, S., Stuart, R.C., Gillespie, M.K., Barr, C.J., 2003. Field boundaries in Great Britain: stock and change between 1984, 1990 and 1998. J. Environ. Manage. 67, 229-238.
Pocock, M.J.O., Evans, D.M., Memmott, J., 2012. The Robustness and Restoration of a Network of Ecological Networks. Science 335, 973-977.
Smart, S.M., Marrs, R.H., Le Duc, M.G., Thompson, K., Bunce, R.G.H., Firbank, L.G., Rossall, M.J., 2006. Spatial relationships between intensive land cover and residual plant species diversity in temperate farmed landscapes. Journal of Applied Ecology 43, 1128-1137.
Staley, J.T., Sparks, T.H., Croxton, P.J., Baldock, K.C.R., Heard, M.S., Hulmes, S., Hulmes, L., Peyton, J., Amy, S.R., Pywell, R.F., 2012. Long-term effects of hedgerow management policies on resource provision for wildlife. Biological Conservation 145, 24-29.
Tenbergen, B., 2001. Hedgerows in North-western Germany: History, ecological functions, management and future of hedges in the cultural landscape of Westphalia, Germany, in: Barr C.J., Petit, S (Ed.), Hedgerows of the world: their ecological functions in different landscapes. IALE (UK), University of Birmingham.

Figure 1% accuracy of linear models 1 (6 variables), 2 (6 variables+ land class), 3 (13 variables) and 4 (13 variables+ land class)

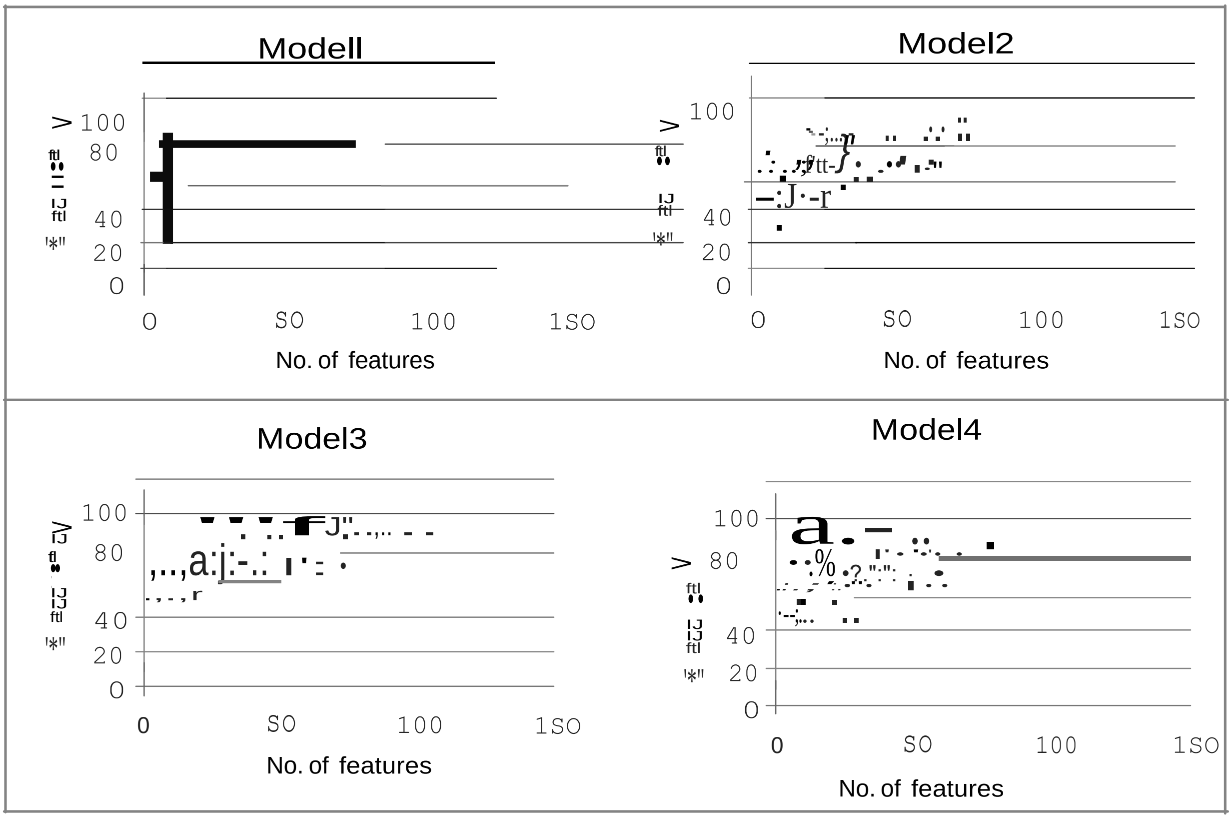
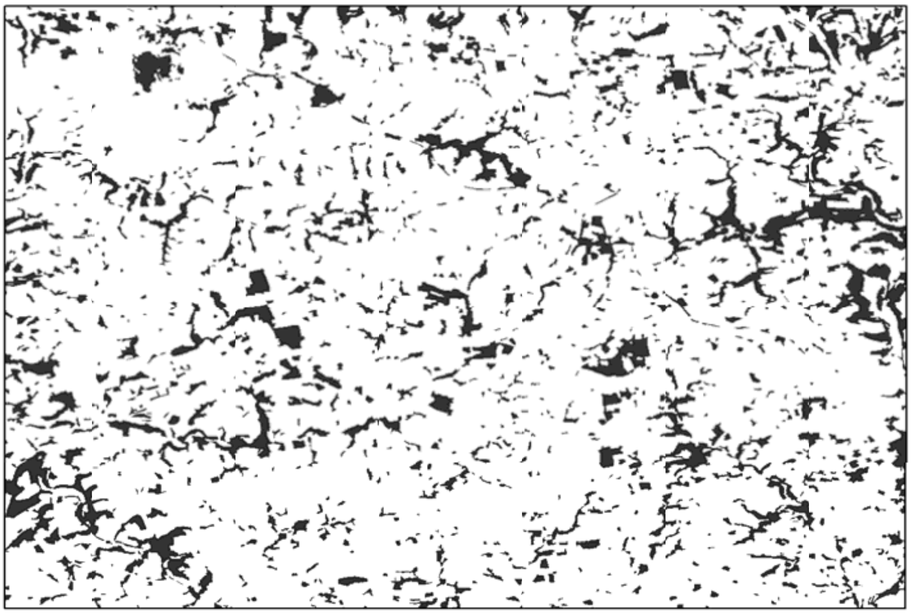


Figure 2. Woody cover aerial habitat maps for an area in Cornwall 1) without woody linear features and 2) with woody linear features

1)



2)

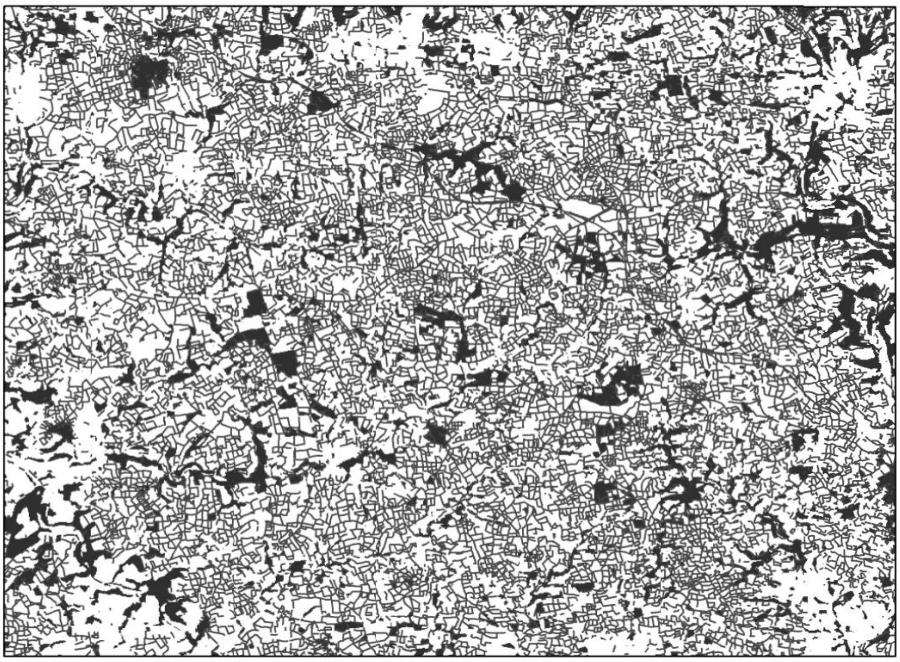


Table 1. Variables used in the models (those **in bold** were used in all four models), variable 14 was the classifier variable.

	Variables	Derived from
1	Altitude – weighted mean	OS Land Form altitude data
2	Curvature – linear weighted mean	OS Land Form altitude data
3	**Curvature –minimum**	OS Land Form altitude data
4	**Curvature – maximum**	OS Land Form altitude data
5	**Maximum height (25m) – linear weighted mean**	NEXTMAP
6	Maximum height (25m) - minimum	NEXTMAP
7	Maximum height (25m)- maximum	NEXTMAP
8	**5m canopy data - linear weighted mean**	NEXTMAP
9	5m canopy data - minimum	NEXTMAP
10	**5m canopy data - maximum**	NEXTMAP
11	Aggregated 35m - linear weighted mean	NEXTMAP
12	**Aggregated 35m - minimum**	NEXTMAP
13	Aggregated 35m - maximum	NEXTMAP
14	Land class	ITE Land Classification

Table 2. Model design

	With Classifier		Without Classifier	
Number of variables	6 **(in bold** Table 1)	13	6 **(in bold** Table 1)	13