Using geographical information systems (GIS) and large area databases to predict Yield Class: a study of Sitka spruce in Wales

IAN J. BATEMAN and ANDREW A. LOVETT

Both authors are at 1) Centre for Social and Economic Research on the Global Environment (CSERGE) and 2) School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, England

Summary

A geographic information system (GIS) is used to combine and analyse data from a variety of existing large area databases concerning tree growth, plantation management and the environmental characteristics of planted sites. Principal component analysis and regression techniques are employed to estimate a number of Yield Class models for Sitka spruce (*Picea sitchensis* (Bong). Carr.). The GIS is used to extrapolate results and generate maps of predicted yield for a large study area (the entirety of Wales). The resulting methodology produces well fitting models of timber yield which compare favourably with those reported in the literature while the GIS generated maps are readily compatible with those currently under construction by various UK forestry authorities in order to plan for a proposed increase in afforested land.

Introduction

Recent changes in UK forestry policy (UK Government, 1995) combined with ongoing reforms of the Common Agricultural Policy have strongly suggested that the area of the UK under forestry is likely to expand very significantly over the next half century. A recent joint paper from the Countryside and Forestry Commissions (1996) indicates that much of this planting will not be confined to upland areas but will encompass lowland sites which may well form the focus of future planting policy. Furthermore, these authorities have announced that planning will be guided by a series of forthcoming maps indicating optimal planting areas (Countryside and Forestry Commissions, 1996).[1] Identification of factors influencing spatial variations in tree productivity is therefore likely to be important in the implementation of such strategic policies. This paper describes several models of timber yield for Sitka spruce (*Picea sitchensis* (Bong.) Carr.), chosen as the most commonly planted UK softwood species (FICGB, 1992). The methodology developed in the research differs from previous approaches in that it uses a geographical information system (GIS) and databases covering a very large and diverse study area; the whole of Wales. When compared with previous approaches, use of a GIS permits a significant increase in both the range of variables and the number of observations used within the modelling process. These methodological improvements result in a noticeable improvement in the statistical predictive power of the yield models estimated here, a result which is particularly interesting given the extensive and diverse nature of the case study area under consideration. Furthermore, the cartographic facilities of the GIS are employed to produce maps of predicted timber yield for the entire study area,[2] these being readily compatible with the land use planning maps under developments by the Countryside and Forestry Commissions as mentioned previously.

In the second section of the paper we present a brief review of previous studies. These have been based upon relatively limited numbers of observations and have generally been confined to comparatively small areas and often to one topographic region, e.g. upland localities. The section concludes by reviewing the methodology developed for the present study and discusses potential limitations of this approach. The section 'Data and data manipulation' presents details of the various datasets used in our research and discusses how they were transformed and integrated for the purposes of subsequent regression analysis. Results from our models of Sitka spruce growth rates are presented in the next section while the section 'Mapping yield class' describes and interprets the maps of predicted yield class. The final section summarizes our findings, presents conclusions and considers the potential wider applicability of the methodology developed in this paper.

Literature review and methodological overview

Literature review

Tree growth rates depend upon a variety of species, environmental and silvicultural factors. Early work in this field relied on simple rules of thumb, with relatively little supporting data (Busby, 1974), or analyses of single factors. Reviews of this literature provide a number of clues regarding the specification of a Yield Class (YC) model. An early focus of interest was the impact of elevation upon productivity (Malcolm, 1970; Mayhead, 1973; Blyth, 1974). Subsequent papers considered the various mechanisms by which elevation affected YC including windiness (Grace, 1977), slope and aspect (Tranquillini, 1979). Other work examined the impact of factors such as soil type, soil moisture transport and droughtiness (Page, 1970; Blyth and Macleod, 1981; Jarvis and Mullins, 1987), rainfall and water deficits (Edwards, 1957; Berg, 1975) and crop age (Kilpatrick and Savill, 1981). However, while the principles of timber productivity analysis have been established for some time (see reviews by Carmean, 1975 and Hägglund, 1981), the estimation of statistical models across a full range of likely explanatory variables is a relatively recent innovation (Worrell, 1987a, b; Worrell and Malcolm, 1990a, b; Macmillan, 1991; Tyler *et al.*, 1996).

While there had been a number of earlier investigations of factors affecting the growth of Sitka spruce (Malcolm, 1970; Malcolm and Studholme, 1972; Mayhead, 1973; Blyth, 1974; Busby, 1974; Gale and Anderson, 1984), the work of Worrell (1987a, b) and Worrell and Malcolm (1990a, b) represented a notable advance as the first to adopt a multiple regression approach across an extensive range of explanatory variables. These were: elevation (including separate dummy variables for hilltop and valley bottom sites); windiness; temperature; aspect (measured as sine and cosine); and a full range of soil dummies. However, while this gives us vital pointers for our own modelling exercise, it is unclear as to what extent Worrell's results are transferable to other locations. Such concerns arise both because of the growing conditions prevailing in the case study area of upland northern Britain and because of the specific focus of Worrell's experiment on the influence of elevation upon YC in upland areas.[3] To this end Worrell selected 18 principal sample sites,[4] all of which had relatively steep slopes, and took measurements along a vertical transect at each site. By taking samples at sites ranging from 50 m to 600 m a.s.l. a very strong relationship with elevation was established. However, such a model is only applicable to similar, steeply sloping sites (strictly speaking, only the subset of those found within northern Britain), and it is not generalizable to the diversity of environmental conditions found in an area the size of Wales.

A similar, though less extreme, concern limits the applicability of the findings reported by Macmillan (1991). Here again the study area was geographically restricted, this time to lowland Scotland, although the 121 sites used were not selected to emphasize the influence of any particular explanatory variable and are therefore of somewhat wider relevance for lowland areas. However, with respect to our own research, the topographic variability of Wales means that a model based purely upon lowland data is insufficient for our needs. Nevertheless, the Macmillan study is interesting both because it uses multiple regression with a prior principal components analysis (PCA) of explanatory variables (reporting a final degree of explanation of R^2 = 36.8 per cent) and because the data collected has recently been re-analysed using GIS techniques (Elston *et al.*, 1997) to produce a somewhat improved model (R^2 = 43.9 per cent), a result which underlines the potential advantages of applying GIS methods within this field.

A short note regarding model fit is appropriate here. The YC of a plantation is its average annual growth rate assessed over an optimal rotation. YC is therefore given in m^3 ha^{-1} a^{-1}. However, YC values are rounded to the nearest even number so that while we have stands with YC 6 or 8 we do not have sites with YC 7. While this does not invalidate statistical analysis, as YC is the dependent variable such an approach to measurement does artificially structure the yield data and therefore make a high degree of statistical explanation difficult to attain. As such the absolute value for goodness of fit statistics such as R^2 should be treated with some caution and instead we should consider relative degrees of fit compared with those attained in other studies.

Overview of modelling approach

These previous studies provide very useful indications of the explanatory variables which should be considered in our analysis. The differences in modelling approach are also of interest and we consequently decided to investigate both PCA and standard multiple regression methodology. However, in other respects the methods of Worrell and Macmillan were not appropriate to the specific questions posed in our research. Our central aim was to identify areas over a large and diverse region which might be suitable for forestry. This necessitated the development of a methodology which was capable of producing estimates for both upland and lowland areas. In order to have input data across such a diversity of terrain it was decided to focus on the entire area of Wales, a country noted for its variation in landscapes.

Our methodology entailed the extensive use of GIS in the analysis of existing large forestry and land characteristic databases.[6] We took our basic YC data from the Forestry Commission (FC) Sub-Compartment Database (SCDB; described in detail subsequently) which holds information on each discrete stand (sub-compartment) in the FC estate. As this covers both upland and lowland sites, results from such a model are more generalizable than those described previously. Use of the SCDB has the added bonus of massively increasing our sample size relative to previous studies. However, rather than relate YC to the environmental variables reported in the SCDB, we extract these from a separate database, LandIS (described subsequently), which has complete national coverage (unlike the SCDB which only has data for forest areas). Our regression results could then be readily extrapolated to all other areas of Wales, including those not presently under forestry.

One possible disadvantage of such an approach is that, unlike the previous studies, here the data were not collected directly by the researchers but by many other individuals, often over an extended period. While this can be viewed as not entirely negative, our subsequent modelling indicated that allowances had to be made for variance introduced by apparently different approaches to measurement. In particular it should be noted that the spatial resolution of data used in this study varied between 1 km or better for SCDB and other FC derived data, but generally 5 km resolution for the LandIS data. While subsequent analysis reveals (not surprisingly) that better resolution data improves Yield Class prediction, nevertheless this does limit the applicability of our models to the prediction of tree growth across relatively small areas. However, such micro-scale prediction was not an objective of this study. Furthermore, this limitation is a product of data availability rather than the underlying methodology developed in this paper. Given this we felt justified in pursuing our approach once these limitations are noted.

Data and data manipulation

The research relied upon a diversity of data sources. In addition to the SCDB and LandIS databases, further environmental and topographic data were obtained from a variety of sources. In this section we describe all of these data and how they were manipulated before the subsequent statistical investigation of tree growth.

SCDB data

The SCDB is the FC's central forest inventory containing observations for all stands in the estate. As such it provides an invaluable source of high quality data. Some of this concerns internal administration and was not of interest to our investigation and so the final list of variables extracted for this study was as detailed in Table 1. This summary also shows how some of the data were manipulated to produce further (often dummy) variables. In doing this, one-way analyses of variance on the dependent variable

Table 1: Variables obtained from the Forestry Commission Sub-Compartment Database

Variable name	Values	Notes and recodings (in italics)
Grid reference	Easting Northing	100 m resolution OS grid references
Land use/crop type	PHF = plantation high forest PWB = uncleared windblown area PRP = research plantation	*Uncleared* = 1 if PWB = 0 otherwise *Research* = 1 if PRB = 0 otherwise
Storey	1 = single storey 2 = lower storey 3 = upper storey	*Single* = 1 if single storey 0 otherwise
Species	SS = Sitka spruce BE = beech, etc.	Used to identify target species
Planting year	Discrete variable	*Plantyr*: year in which stand was planted (1920 = 0; 1995 = 75)
Survey year	Discrete variable	*Survyr*: year in which stand was surveyed
Yield class	Even number	*YC*: tree growth rate: average $m^3\ ha^{-1}\ a^{-1}$ over an optimal rotation – the dependent variable
Productive forest area	ha	*Area*: stocked area of the sub-compartment
Unproductive forest area	ha	*Unprod*: the area within the sub-compartment which has a permanent effect upon the crop, e.g. rocky outcrops, etc.
Rotation	1 = 1st rotation on formerly non-forest land 2, 3 etc. = 2nd, 3rd rotation, etc. 9 = historical woodland sites S = ancient, semi-natural woodland	*1st Rot* = 1 for 1st rotation; = 0 otherwise *2nd Rot* = 1 for 2nd rotation; = 0 otherwise *Historic* = 1 if historic site; = 0 otherwise *Semi-nat* = 1 if ancient/semi-natural woodland, = 0 otherwise
Mixture	P = single species crop M = mixed species crop	*Mixed* = 1 if mixed crop; = 0 otherwise

Variable name	Values	Notes and recodings (in italics)
Legal status	P = purchased by FC L = leased E = extra land, managed by FC outside legal boundary	*Purchased* = 1 if purchased; = 0 otherwise *Leased* = 1 if leased; = 0 otherwise *Extra* = 1 if extra; = 0 otherwise
Landscape	1 = National Park 2 = AONB/National Scenic Area 3 = ESA (where not included in 1 or 2 above)	*NatPark* = 1 if National Park; = 0 otherwise *AONB/NSA* = 1 if AONB/National Scenic Area; = 0 otherwise *OthESA* = 1 if ESA area not included in above; = 0 otherwise
Forest park	1 = Forest Park	*FPark* = 1 if forest park; = 0 otherwise
Conservation	1 = SSSI (Site of Special Scientific Interest) 2 = NNR (National Nature Reserve) 3 = Non-FC Nature Reserve	*SSSI* = 1 if SSSI; = 0 otherwise *NNR* = 1 if NNR, = 0 otherwise *NonFCNR* = 1 if Non FC nature reserve; = 0 otherwise
FC conservation	1 = Forest Nature Reserve 2 = Other FC conservation	*FCNR* = 1 if Forest Nature Reserve; = 0 otherwise *FCcons* = 1 if other FC; = 0 otherwise
Ancient monument/ woodland	S = scheduled ancient monument U = unscheduled ancient monument W = ancient woodland	*Ancient* = 1 if S, U or W; = 0 otherwise *Monument* = 1 if S or U; = 0 otherwise
		Further recodes from above: *NpAonbSa* = 1 if any of Nat Park or AONB/ NSA; = 0 otherwise *Cons* = 1 if any of NNR, NonFCNR, FCNR, Fcons; = 0 otherwise *Reserve* = 1 if any of Cons, AOBN/NSA, OthESA; = 0 otherwise *Park* = 1 if any of Nat Park, F Park, SSSI; = 0 otherwise

AONB, area of outstanding natural beauty; ESA, environmentally sensitive area

(YC) were used to identify likely significant divisions in the data.

The SCDB also contains a variety of sub-compartment specific environmental variables such as soil type, altitude, terrain type and windblow hazard class. Normally these would be ideal for modelling purposes. However, as the FC only holds such data for those grid squares in which it has plantations, and since uninterrupted national coverage does not exist for many of these variables, findings based upon such data would not form a suitable basis for extrapolation to other, currently unforested areas. This is somewhat unfortunate as this site-specific data is almost certainly more accurate than that obtainable from more general databases such as LandIS. This means that the regression models produced using LandIS may not fit the YC data as well as those using the site characteristics given in the SCDB. However, for the purposes of this research, the advantage of being able to extrapolate out across the entire surface of Wales and consider currently unplanted areas easily outweighs such costs (which we subsequently argue, on the basis of our results, are likely to be small).

In all records for some 6082 Sitka spruce subcompartments were used in our regression analysis.[7] This represents a very significant increase over sample sizes used previously in the literature. These observations were distributed throughout upland and lowland Wales providing a good basis for extrapolation of results to other, presently unforested areas.

LandIS data

Background The first systematic attempt to analyse and record British soil information was the 'county series' of maps initiated by the Board of Agriculture in the late eighteenth and early nineteenth centuries. Until comparatively recently this remained the standard and unsurpassed source of soil data. During the 1940s the Soil Survey of England and Wales (SSEW) began a detailed mapping initiative. However, by the late 1970s, only one-fifth of the country had been covered. In 1979 the SSEW, which in the late 1980s became the Soil Survey and Land Research Centre (SSLRC), commenced a 5-year project to produce a soil map of the whole of England and Wales and to describe soil distribution and related land quality in appropriate detail.

The data collected in this exercise was digitized, spatially referenced, and subsequently expanded to include climate and other environmental information (Bradley and Knox, 1995). The resulting land information system (LandIS) database was initially commissioned by the Ministry of Agriculture, Fisheries and Food, with the stated aim of 'providing a systematic inventory capable of being used or interpreted for a wide range of purposes including agricultural advisory work, but also for the many facets of *land use planning and national resource use*' (Rudeforth *et al.*, 1984, emphasis added). However, although the maps and accompanying bulletins were completed in 1984 there has never been any major attempt since then to incorporate them into policy making. The present research therefore represents one of the first attempts to use LandIS for its originally intended purpose: national land use planning.

The data Detailed definitions, derivations and accuracy of the data extracted from LandIS are presented in Bateman (1996). These are summarized in Table 2. Further details of LandIS and the data therein are given in Jones and Thomasson (1985) with discussion of Welsh conditions given by Rudeforth *et al.* (1984). LandIS data is supplied at a 5-km resolution.

An immediate problem with LandIS data was the plethora of differing soil codes. These are taken from SSEW (1983) which lists many hundreds of separate soil types, a large number of which were present in our Welsh dataset. This level of detail far exceeds that used in previous YC studies such as Worrell (1987b) who uses seven soil type dummies derived from information given in the SCDB which in turn relies on the standard FC classification of soils. The large number of soil codes given in LandIS is a problem, both because of the implication for degrees of freedom in our subsequent regression analysis and because any such results would be of little practical use to the forester familiar with an alternative and simpler system. Furthermore, consultations with an expert in the field of soil science and forestry suggested that, for our purposes, many of the SSLRC soil codes could be merged with no effective loss of information and a substantial increase in clarity.[8] Details of the final categorization are given in the final row of Table 2.

Other data

Topex and wind hazard[9] Topex is a measure of the topographical shelter of a site. It is usually determined as the sum of the angle of inclination for the eight major compass points of a site (Hart, 1991). A low angle sum (low topex value) therefore represents a high degree of exposure. The resultant variable was labelled *Topex1km*.

Blakey-Smith *et al.* (1994) define wind hazard on the basis of four factors:[10] wind zone, elevation, topex and soil type. The resultant continuous variable (*Wind1km²*) is inversely associated with tree productivity and growth rates.

Elevation and associated variables The work of Worrell and Malcolm (1990a) shows that elevation and its associated characteristics are key predictors of YC. However, such a variable is not included in the LandIS database and the

Table 2: Variables obtained from LandIS

Variable name	Label	Definition
Accumulated temperature	*Acctemp*	Average annual accumulated temperature (in °C) above 0°C
Accumulated rainfall	*Rainfall*	Average annual accumulated rainfall (in mm)
Available water	*Avwatgra*	Amount of soil water available for a grass crop after allowing for gravity induced drainage
	Avwatcer	As per *Avwatgra* but adjusted for a cereal crop
	Avwatpot	As per *Avwatgra* but adjusted for potatoes
	Avwatsb	As per *Avwatgra* but adjusted for sugarbeet
Moisture deficit	*Mdefgra*	The difference between rainfall and the potential evapotranspiration of a grass crop
	Mdefcer	As per *Mdefgra* but adjusted for a cereal crop
	Mdefsbpt	As per *Mdefgra* but adjusted for a sugarbeet/potatoes crop
Field capacity	*Fcapdays*	Average annual number of days where the soil experiences a zero moisture deficit
Return to field capacity	*Retmed*	Median measure from a distribution of the number of days between the date on which a soil returns to field capacity and 31st December of that year
	Retwet	The upper quartile of the above distribution (measure of return to field capacity in wet years)
	Retdry	The lower quartile of the above distribution (measure of return to field capacity in dry years)
End of field capacity	*Endmed*	Median measure from a distribution of the number of days between 31st December and the subsequent date on which field capacity ends
	Endwet	The upper quartile of the above distribution (measure of the end of field capacity in wet years)
	Enddry	The lower quartile of the above distribution (measure of the end of field capacity in dry years)
Workability	*Workabil*	A categorical scale indicating the suitability of the land for heavy machinery work in spring and autumn
Spring machinery working days	*SprMWD*	The average number of days between 1st January and 30th April where land can be worked by machinery without soil damage
Autumn machinery working days	*AutMWD*	The average number of days between 1st September and 31st December when land can be worked by machinery without soil damage
Soil type	*SoilX*	SSLRC soil type classification code: *Soil1* = Lowland lithomorphic; *Soil2* = Brown earths; *Soil3* = Podzols; *Soil4* = Surface water gley; *Soil5* = Stagnogley (perched watertable); *Soil6* = Ground water gley; *Soil7* = Peats; *Soil8* = Upland lithomorphic. *Soil23* = areas with *Soil2* or *Soil3*

SCDB only gives heights for existing plantation sites. Clearly for extrapolation purposes this is inadequate and so an alternative source of data was required. This was provided in the form of a digital elevation model (DEM).[11] The DEM provided an image of the topography of Wales and was created within the GIS from three principal data sources: the Bartholomew 1:250000 database for the UK, summit points from Bartholomew's database and the spot height of plantations recorded in the SCDB. Subsequent testing indicated that the resulting elevation variable (*Wselvgr2*) had a root mean square error of 21 m which was considered acceptable accuracy for the purposes of the research. The elevation data were then used to generate two further GIS surface variables: slope angle (*Dsl2*) and aspect angle (*Wsaspgr2*). Data on these variables were produced at a 500-m resolution.

Creating GIS surfaces for explanatory variables

Before the regression analysis two fundamental issues had to be addressed regarding the definition of a common extent and common resolution for the environmental variables. When the geo-referenced data obtained from the LandIS and non-SCDB sources described above were compared within the GIS it soon became evident that the various sources differed in their geographical extents and spatial resolutions.

Data were supplied at a wide array of resolutions ranging from the (nominal) 100 m accuracy of the SCDB to the 5 km tiles of the LandIS variables. While the technical operation of interpolating from a coarse to a finer resolution is relatively straightforward within a GIS, it needs to be recognized that the apparent precision achieved may be rather higher than the underlying accuracy of the data (Goodchild, 1993), so deciding upon a common unit size[12] was a matter for some deliberation. Standardization upon the smallest unit (100 m) did not seem a sensible choice. For instance, the 100-m reference used in the SCDB is, the FC admit, spuriously precise. On the other hand, aggregation up to the 5-km scale of the coarsest data was thought likely to result in a loss of much relevant detail (e.g. for topographic features). As a compromise, a 1-km grid was settled upon and all the data were converted to this resolution.

The spatial extent of Wales was defined by converting a vector outline of the Welsh coast and border with England (from the Bartholomew 1:250000 scale database) to a raster grid representation consisting of 1 km^2 cells. This resulted in a layer within the GIS containing 20563 land cells and values of the variables in the LandIS and non-SCDB datasets described above were then estimated for each grid cell.[13] For characteristics such as topex or elevation this was done by aggregation and averaging, whereas with the LandIS variables each 1-km grid square was given the value of the 5-km cell it fell within.

With all data now at a common resolution and extent we now had the necessary complete surfaces of potential predictor variables for use in our regression model and from which extrapolation across all areas, whether currently planted or not, would be possible.

A final task concerned the extraction of values for all environmental variables for each YC observation in the SCDB. This was achieved by using point-in-polygon operations within the GIS to identify the 1-km grid cell corresponding to each sub-compartment grid reference.

Principal components analysis

As discussed in our literature review, two approaches have been adopted for the statistical modelling of YC data. While Worrell (1987a, b) and Worrell and Malcolm (1990a, b) used conventional regression analysis, Macmillan (1991) first subjected explanatory variables to a principal components analysis (PCA)[14] before entering the resultant factors into a regression analysis. It was decided that a comparison of these two approaches would be of interest and so a PCA was undertaken on our data.

In essence PCA attempts to identify patterns of covariance so that trends within a comparatively large number of variables are summarized by a smaller number of factors, i.e. it seeks to identify patterns of common variance. For example, in our literature review, we noted that the negative relation between YC and altitude was actually the product of a range of interrelated variables including elevation, slope, aspect, temperature, etc. A general 'height' factor which reflected these interrelations might therefore prove a strong predictor of tree growth. Standard PCA techniques (Norusis, 1985) were applied to all the environment variables identified in the preceding sections (i.e. excluding those management variables detailed in the SCDB). This analysis produced five significant PCA factors which together explained some 76.9 per cent of the total variance in the dataset.[15] Inspection of these factors indicated that they were relatively easy to interpret as follows:

Factor No.	*Interpretation*
1	Higher rainfall/soil wetness
2	Steeper slopes/low windiness
3	Waterlogging/poor workability/ high elevations
4	Cold/sine aspect
5	High elevation/cosine aspect

The 'communality' or proportion of variance in each input variable which is 'explained' by the five factors[16] was also calculated. This indicated that the only variable relatively poorly explained was *SprMWD* (communality = 0.39), all other variables having a reasonable proportion of variance explained by our five factors (mean communality = 0.80).

The factor score coefficient matrix was calculated via the regression method described by Norusis (1985).[17] Factor scores (which indicate the position of each observation (here each FC sub-compartment) on the PCA factors) were then calculated in the normal manner.[18] The site specific factor scores obtained by this process could then be entered directly into our YC regression model as the environmental explanatory variables.

Results: yield models for Sitka spruce[19]

Three types of model were fitted. These varied according to whether the environmental characteristics of a site were described by: (1) raw data; (2) factors for our PCA; (3) a mixture of these two (ensuring that raw variables retained in the model were not significantly correlated with retained factors). Clearly, interpretation of these latter mixed models is difficult if the site characteristic being described by a particular factor is also being explained by a raw data variable. For example *Factor 1*, which represents soil wetness and rainfall could not be included within the same model as the raw variable *Rainfall*. However, we wished to test whether some site characteristics might be better described by factors while, within the same model, other uncorrelated characteristics could be optimally described by raw variables. Our initial dataset for Sitka spruce contained a number of sites for which YC or other key data were missing and so these records were deleted to leave a dataset of 6082 sites. This is far larger than any of the studies considered in our literature review and demonstrates one of the principal advantages of our database approach compared with more common analyses based upon small site surveys.

Our regressions analyses followed the approach set out by Lewis-Beck (1980) and Achen (1982). An initial objective concerned the identification of an appropriate functional form for our models. Tests indicated that a linear model performed marginally better than other standard forms[20] and, given that such a form is both easily interpretable and typical of other studies, this seemed a sensible choice.

Initial comparison across the factor only, variable only and mixed model types suggested that there was little difference in the degree of explanation afforded by these various approaches but that the mixed model performed marginally better than the others and is reported as Model 1. Inspection of this shows that the large sample size has permitted the identification of a large number of highly significant predictors, many of which conform to prior expectations. With respect to the environmental characteristics of sites we can see that YC falls with increasing rainfall (*Rainfall*), elevation (*Wselvgr2*) and cosine aspect (*Factor 5*) and rises with low windiness (*Factor 2*).

Because of its categorical nature, soil type is considered as a series of dummy variables, two of which proved statistically significant. YC is significantly elevated by planting on brown earth or podzol soils (*Soil23*, which is a simple combination of *Soil2* and *Soil3*) and significantly depressed by planting on lowland lithomorphs (*Soil1*) which, despite their location, are poor soils. Both results conform to prior expectations.

The model also highlights the importance of silvicultural factors. The positive relationship with the size of the plantation (*Area*) is interesting and accords with the recent findings of Maclaren *et al.* (1995). This would seem to indicate that trees which are part of large plantations are more likely to thrive than those in small areas. Maclaren *et al.* suggest that this may be due to decreased wind penetration within larger stands. However, alternative hypotheses are that such stands provide advantages in terms of the ease of adopting species

Model 1: Initial regression model for Sitka spruce (mixed model)

Predictor	Coef	Stdev	t-ratio	P
Constant	17.0792	0.2482	68.83	0.000
Rainfall	–0.00177733	0.00008489	–20.94	0.000
Wselvgr2	–0.0070769	0.0003906	–18.12	0.000
Factor 2	0.07469	0.03586	2.08	0.038
Factor 5	–0.16595	0.03365	–4.93	0.000
Soil23	0.89814	0.06729	13.35	0.000
Soil1	–4.9538	0.7437	–6.66	0.000
Area	0.0037050	0.0003260	11.36	0.000
Plantyr	0.030379	0.002682	11.33	0.000
1st Rot	–1.52753	0.08576	–17.81	0.000
MixCrop	–0.21314	0.06524	–3.27	0.001
Park	0.91121	0.07682	11.85	0.000
Ancient	1.1777	0.2783	4.23	0.000
Uncleared	2.4639	0.1808	13.63	0.000
Unprod	–0.076776	0.007079	–10.85	0.000
Reserve	–0.36615	0.07685	–4.76	0.000
Semi-nat	–4.5487	0.5983	–7.60	0.000

$s = 2.297$ $R^2 = 40.9\%$ R^2 (adj) = 40.7%

Analysis of variance

Source	DF	SS	MS	F	P
Regression	16	22122.7	1382.7	262.10	0.000
Error	6062	31978.7	5.3		
Total	6078	54101.4			

specific management regimes, or that large stands tend to condition their environment to their own advantage (for example by reducing competition from both flora and fauna).

The strong and positive influence of the time variable (*Plantyr*) is confirmed. This is usually explained as reflecting improvements in silvicultural methods such as the introduction of ploughing and fertilizers and/or improvements in the genetic stock.[21]

Two further silvicultural factors can be identified. Trees planted on ground which has not been previously used for afforestation (*1st Rot*) perform relatively worse than those planted in successive rotations. This may be because second rotation trees have on average been planted more recently than those in the first rotation (although the relatively low correlation with *Plantyr* indicates this may not be all of the story) or that second rotation trees inherit a nutrient enriched soil base from their forebears. Trees also seem to perform less well when grown in a mixed species plantation (*MixCrop*) than in monoculture, a finding which suggests that there may be a timber productivity benefit associated with the amenity cost of the latter.

Next, a number of site factors which arise from the interaction of environmental characteristics and management practice were identified. YC is significantly higher in parkland areas (*Park*), a result which may reflect more careful silvicultural management. The result that planting in areas which were previously ancient woodland (*Ancient*) boosts tree growth seems to be the corollary of the impact of *1st Rot*. A further and rather interesting boost to growth is implied by the variable *Uncleared* which identifies trees growing in areas which have been previously affected by windblow but have not yet been cleared. It seems that the surviving trees actually profit from windblow in that their immediate neighbours (and competitors) are removed thus boosting their access to nutrients. However, while growth rate may benefit from such events, the ensuing lack of cover raises the probability that the survivors will subsequently fall victim to windblow themselves.

Finally, three negative environmental/management factors are apparent. Plantations with higher amounts of unproductive land (*Unprod*) not surprisingly perform relatively worse than otherwise similar sites. Sub-compartments which fall within the boundaries of conservation areas (*Reserve*) also exhibit relatively lower YC, as do areas which are allowed to remain as semi-natural habitat (*Semi-nat*); results which may reflect the application of less intensive silvicultural techniques in such areas.

Conversations with a number of forestry experts[22] suggested that model fit might be improved by omitting those stands where YC measurements had been taken relatively soon after planting. The assessment of YC is particularly difficult in the early years of a rotation and our hypothesis is therefore that such observations are likely to have greater variability than those taken from more mature stands. To test this hypothesis a survey age variable (*Sage*) was calculated from the planting year (*Plantyr*) and YC survey year (*Survyr*) data previously described. Sub-compartments were iteratively removed from the dataset and on each iteration Model 1 was re-estimated. This analysis confirmed a small but noticeable increase in model fit as stands surveyed at a very early age were removed. This improvement was maximized by omitting all observations with a survey age of less than 10 years, a process which still leaves us with 5168 observations. All three model variants were re-estimated from scratch[23] and the no-factor model found to provide the most clearly interpretable results (see Model 2). We can also use this model to provide an interesting aside regarding the effect of aspect upon tree growth. This is achieved by including the variables *Sinasp* and *Cosasp* in the model.

Comparison of Model 2 with Model 1 shows that the exclusion of sites with *Sage<10* results in a small but noticeable improvement in the overall degree of explanation. The removal of all PCA factors has allowed some new environmental variables to enter the model and we can see that as topographic shelter (*Topex1km*) increases so does YC. As stated, we have deliberately included *Sinasp* and *Cosasp* in the model to assess aspect effects. As these variables are only interpretable as a pair it is likely that, as a result of how variables explain variation within a regression model, one of them may appear statistically significant.[24] However, if we adopt a conventional 5 per cent confidence test then neither of these aspect variables appears significant. Nevertheless, it is clear that we do not have to relax such a test by much before aspect does appear to have a significant effect.

If we temporarily accept that some weak aspect effect is occurring then we can use the coefficients given in Model 2 to see what this is. Figure 1 illustrates this predicted impact and compares our result with that of Worrell and Malcolm (1990b) from their study of Sitka spruce growing on upland sites in northern Britain. Comparing these results, the magnitude of the aspect effect is slightly higher in the latter study, a result which is not surprising given the more adverse conditions present in northern uplands. However, more striking is the subtle shift in the direction of aspect effects between these two studies. Worrell and Malcolm report that YC is most severely depressed on west facing sites and highest on eastern slopes. The complete negation of any effect due to increased solar radiation from the south might reflect the powerful impact which the prevailing westerly wind has upon such sites. Considering our own results we can see that here the aspect effect has shifted round to the south, so that in Wales it is south-east facing sites which appear to do best. It would seem that the less severe conditions of Wales mean that the southern solar energy effect is not completely cancelled out by the prevailing west wind. Nevertheless it is still the effect of that wind which makes a south-easterly facing site outperform one which faces south-west.

Returning to our analysis of models omitting stands of varying survey age; in undertaking this analysis it was observed that there was a strong

Model 2: Yield Class model for Sitka spruce after omitting stands with survey age <10 years: No PCA factors used

Predictor	Coef	Stdev	*t*-ratio	*P*
Constant	16.6333	0.2697	61.66	0.000
Rainfall	−0.00176521	0.00009584	−18.42	0.000
Wselvgr2	−0.0084288	0.0003633	−23.20	0.000
Topex1km	0.025931	0.006818	3.80	0.000
Sinasp	0.7872	0.4540	1.73	0.083
Cosasp	−0.6841	0.45792	−1.49	0.137
Soil23	0.82527	0.07273	11.35	0.000
Soil1	−4.8614	0.7504	−6.48	0.000
Area	0.0038847	0.0003639	10.67	0.000
Plantyr	0.050639	0.003230	15.68	0.000
1st Rot	−1.7636	0.1005	−17.56	0.000
MixCrop	−0.28948	0.06928	−4.18	0.000
Park	0.86170	0.08295	10.39	0.000
Ancient	0.9345	0.2985	3.13	0.002
Uncleared	2.4261	0.1821	13.32	0.000
Unprod	−0.086657	0.007912	−10.95	0.000
Reserve	−0.44077	0.08421	−5.23	0.000
Semi-nat	−4.6318	0.7299	−6.35	0.000

$s = 2.306$ $R^2 = 42.1\%$ R^2 (adj) = 41.9%

Analysis of variance

Source	DF	SS	MS	*F*	*P*
Regression	17	19921.2	1171.8	220.30	0.000
Error	5150	27394.0	5.3		
Total	5167	47315.1			

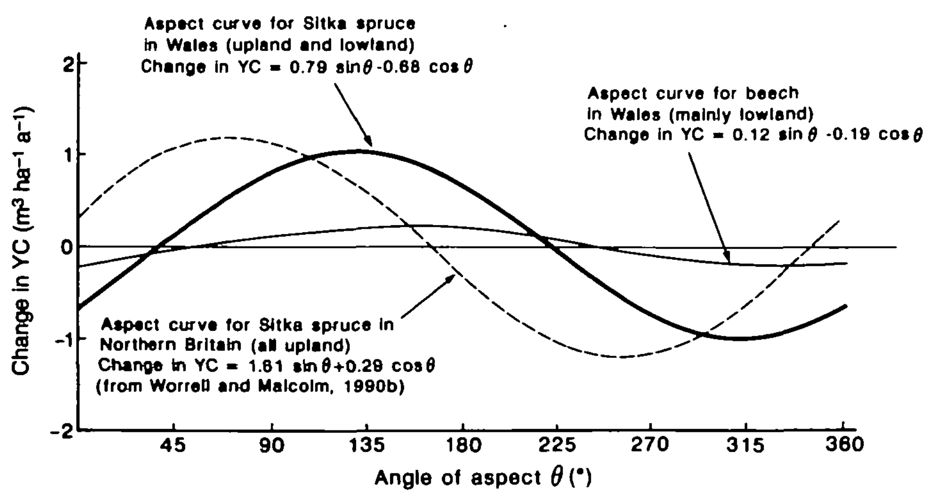

Figure 1. The effect of aspect upon predicted Yield Class (YC).

decline in model fit when we confined ourselves to only examining sub-compartments in which YC surveying occurred very many years after planting. This does not seem to be a product of the smaller sample size of such analyses as we are still considering many hundreds of sites (indeed, as sample size falls, the relatively large number of predictors in the model would tend to inflate goodness of fit statistics).[25] Two reasons may in part account for this effect, both of which arise from the observation that, as we restrict ourselves to older survey age, we are in turn restricting ourselves to older stands. First, improved silvicultural methods, now applied to virtually all new stands, may well have been applied in a less uniform manner to such older stands. New techniques may not have been simultaneously adopted for all plantations but rather tried on a subset of these. The result would be, as observed, that these older stands are more variable than younger ones. Second, it may be that records regarding planting age are relatively less reliable for older stands. As YC is a function of plantation age then if this becomes uncertain so the variability of YC estimates will increase.

Whatever the reason, it seems that omission of those stands with relatively old survey ages is likely to further improve the fit of our model. A sensitivity analysis suggested that omission of sites with survey age above 36 years resulted in an optimal fit for our models while still leaving us with some 4307 sub-compartments in our sample. As before models were rebuilt afresh to allow for the possibility of new explanatory variables better describing this revised dataset. In this analysis the aspect variables exhibited somewhat suspect levels of significance and were accordingly omitted from these final models.

All three model types were estimated. Model 3 reports results from our model which describes site environmental characteristics via PCA factors. While this is of interest and all relationships conform to prior expectations it is outperformed by both our no-factor and mixed models which performed equally well. This is an interesting finding suggesting that the PCA approach used by Macmillan (1991) may not provide any significant improvement over the more widespread conventional regression models used by Worrell (1987a, b), Worrell and Malcolm (1990a, b) and also by Macmillan in his more recent work (Tyler *et al.*, 1995 unpublished, 1996).[26]

Given the very similar performance of our no-factor and mixed models, the former is preferred for ease of interpretation and is reported as our 'best-fit' Model 4. The optimal list of predictor variables was found to be as before and this lack of change in model specification between truncation options gives some added weight to overall validity. Note that the degree of explanation exceeds that of previous non-GIS studies concerning Sitka spruce. Interestingly the GIS based study of Elston *et al.* (1997) yields an almost identical degree of explanation although this latter study applies to a smaller and more homogeneous study area.

The appropriateness of using our best fit model for extrapolation was assessed by comparing predicted with actual YC for the 4307 observations in our revised dataset. Results of this analysis are presented in Table 3 which shows that 76.5 per cent of YC predictions were within one division of actual YC.[27]

Mapping Yield Class

The regression analyses ultimately resulted in two YC models, one including PCA factor explanatory variables (Model 3) and the other without (Model 4), the latter providing a marginally better fit to the data. Both models were subsequently employed to produce estimates of YC in the form of raster grid maps covering the whole of Wales.

Producing predicted Yield Class maps within a GIS

To generate a YC map (or image) the GIS requires data on all the predictor variables for all the grid cells in the area for which we wish to predict, in this case 20 563 1-km squares representing the entire land area of Wales. If we take our best fitting model of Sitka spruce yield (Model 4) as an example, we can see that this is predicted by a constant and a number of explanatory variables. The constant is in essence a data layer in its own right which has identical values (here 16.709) for all grid cells. The first explanatory variable in this

Model 3: Optimal principal component analysis factor model for Sitka spruce: observations with *sage*<10 or *sage*>36 omitted

Predictor	Coef	Stdev	t-ratio	P
Constant	11.8800	0.3090	38.45	0.000
Factor 1	−0.70932	0.04135	−17.15	0.000
Factor 2	0.29481	0.04177	7.06	0.000
Factor 3	−0.92229	0.06664	−13.84	0.000
Factor 4	−0.23857	0.03667	−6.51	0.000
Factor 5	−0.40778	0.03685	−11.07	0.000
Soil23	0.0441	0.1366	0.32	0.747
Soil1	−4.2384	0.9869	−4.29	0.000
Area	0.0036537	0.0003872	9.44	0.000
Plantyr	0.049234	0.004954	9.94	0.000
1st Rot	−2.0853	0.1117	−18.67	0.000
MixCrop	−0.26907	0.07848	−3.43	0.001
Park	0.80303	0.09635	8.33	0.000
Ancient	0.8805	0.3171	2.78	0.006
Uncleared	2.7353	0.2329	11.75	0.000
Unprod	−0.086739	0.008315	−10.43	0.000
Reserve	−0.42987	0.09636	−4.46	0.000
Semi-nat	−4.3591	0.7831	−5.57	0.000

$s = 2.372$ $R^2 = 40.4\%$ R^2 (adj) = 40.1%

Analysis of variance

Source	DF	SS	MS	F	P
Regression	17	16342.51	961.32	170.86	0.000
Error	4289	24131.05	5.63		
Total	4306	40473.56			

model is the predictor *Rainfall* for which we have estimates from the LandIS database. We can therefore begin to build up our predicted YC map by employing the GIS software to calculate a new raster map which contains the values from multiplying the values in the *Rainfall* grid by the relevant coefficient (−0.00167). With the Idrisi GIS software (Eastman, 1993) this operation is performed as an example of map algebra (Berry, 1993) by using the Scalar command. The values in the resultant image can then be added to those for the constant by use of the Overlay command, which as its name suggests, can combine these two raster maps to produce a third which represents YC as predicted by these first two elements in the model. Subsequent predictors can be incorporated in a similar manner with separate intermediate layers being created by multiplying values of variables by their coefficients using the Scalar command and then adding the output onto the YC map using the Overlay command.

When working with the PCA based models we first needed to construct component score images covering the whole of Wales. This was achieved by first creating z-score images of each of the variables considered in the PCA[28] and then using the calculated component score coefficients to produce raster maps of each factor. These were then treated in the same way as the explanatory variables discussed above.

In all the models a number of the predictor variables were related to management (e.g.

Model 4: Best fit Yield Class model for Sitka spruce: no principal component analysis factors used, observations with *sage*<10 or *sage*>36 omitted

Predictor	Coef	Stdev	*t*-ratio	*P*
Constant	16.7097	0.3487	47.92	0.000
Rainfall	−0.0016700	0.0001067	−15.65	0.000
Wselvgr2	−0.0087750	0.0003933	−22.31	0.000
Topex1km	0.024262	0.007592	3.20	0.001
Soil23	0.80489	0.08046	10.00	0.000
Soil1	−4.8827	0.9660	−5.05	0.000
Area	0.0039518	0.0003788	10.43	0.000
Plantyr	0.049890	0.004838	10.31	0.000
1st Rot	−1.9280	0.1093	−17.64	0.000
MixCrop	−0.30832	0.07670	−4.02	0.000
Park	0.94769	0.09385	10.10	0.000
Ancient	0.9266	0.3089	3.00	0.003
Uncleared	2.6411	0.2276	11.61	0.000
Unprod	−0.085426	0.008143	−10.49	0.000
Reserve	−0.43395	0.09452	−4.59	0.000
Semi-nat	−5.1415	0.7644	−6.73	0.000

s = 2.319 R^2 = 43.0% R^2 (adj) = 42.8%

Analysis of variance

Source	DF	SS	MS	*F*	*P*
Regression	15	17391.3	1159.4	215.54	0.000
Error	4291	23082.2	5.4		
Total	4306	40473.6			

Area), policy (e.g. *Reserve*) or when the site was planted (e.g. *Plantyr*). These are not specifically spatial variables so were treated by holding them at certain fixed values (i.e. as per the constant) and varying some of them in a sensitivity analysis. The variables *MixCrop*, *Ancient*, *Unprod*, *Reserve*, *Park*, *Uncleard* and *Semi-nat* are all dummies for infrequently occurring, unusual sites and were consequently held at zero (their median value) for all images. Similarly the variable *Area* was held at its median value of 33 ha. Given the very low value of the coefficient on this variable and its relatively small range (see the descriptive statistics given in Bateman, 1996) sensitivity analysis did not seem justified here. However, this was not the case for the variables *Plantyr* and *1st Rot* and full sensitivity analyses were conducted for these.

Timber yield maps for Sitka spruce

We produced 1-km grid images based on both our best non-PCA and PCA-based yield models. In addition, we also considered the impact of changing the *Plantyr* variable from 0 (being the base year in which the Forestry Commission started to plant Sitka spruce) to 75 (being the present day, i.e. Sitka spruce planting commenced about 75 years ago) thereby arguably reflecting technological progress over that period.[29] For both of these analyses we initially held *1st Rot* at 1, i.e. examining first rotation trees at both of these time periods. However, many present day Sitka spruce plantations are now in their second rotation. Therefore we also tested the effect of letting *1st Rot* = 0 (i.e. second rotation) when *Plantyr* = 75. This

Table 3: Comparing actual with predicted Yield Class (YC) for the best fit YC model of Sitka spruce (cell contents are counts)

	Predicted YC									
Actual YC	4	6	8	10	12	14	16	18	20	All
4	0	0	1	0	0	0	0	0	0	1
6	0	0	7	63	0	0	0	0	0	70
8	1	3	12	161	220	0	0	0	0	397
10	0	0	9	169	395	141	0	0	0	714
12	0	0	4	176	516	285	63	0	0	1044
14	0	0	0	90	415	276	124	33	1	939
16	0	0	0	0	201	313	179	33	1	727
18	0	0	0	0	0	152	144	45	3	344
20	0	0	0	0	0	0	41	26	3	70
22	0	0	0	0	0	0	0	1	0	1
All	1	3	33	659	1747	1167	551	138	8	4307

Predicted YC compared with actual YC	Percentage of total sample (%)
Prediction is two classes too high	12.8
Prediction is one class too high	23.4
Predicted YC equals actual YC	27.9
Prediction is one class too low	25.2
Prediction is two classes too low	11.4

combination of differing models and assumptions resulted in six maps being created.

Maps were produced using the procedure outlined in the previous section. Figure 2 illustrates the predicted YC raster image created from our best fitting Model 4 (no PCA factors used) with *Plantyr* = 75 (i.e. planting in the present day) and *1st Rot* = 0 (i.e. replanting on a previously planted site).

Inspection of Figure 2 clearly shows the very strong influence which environmental characteristics have upon our prediction of of YC. The influences of lower altitude, better soil and less-excessive rainfall combine to produce high YC. The pattern of lower YC associated with higher elevations is particularly noticeable with the mountain ranges of Snowdonia, the mid Cambrians and the Brecon Beacons clearly picked out. Less extreme upland areas such as the Preseli Mountains produce YC values which lie between these extremes. Also clearly apparent is the adverse excess rain-shadow lying to the east of the Cambrians which results in large areas of relatively depressed YC values stretching in some cases up to (and across) the English border. The adverse effect of sandy and estuarine soils upon growth can also be seen in the small but significantly depressed areas of low yield at places such as the tip of the Gower Peninsula and nearby Pembrey, the southernmost part of Anglesey and the Landudno peninsula.[30]

Table 4 summarizes the distribution of predicted yield for the entire area of Wales as estimated from each of the six model permutations described at the start of this section. While the afforestation of the entire country is clearly hypothetical, such an exercise permits inspection of the responsiveness of predictions to changes in the underlying model and assumptions regarding when, and under what conditions, planting occurs. Notice that as we increase *Plantyr* (i.e. consider contemporary rather than historical planting) and/or move

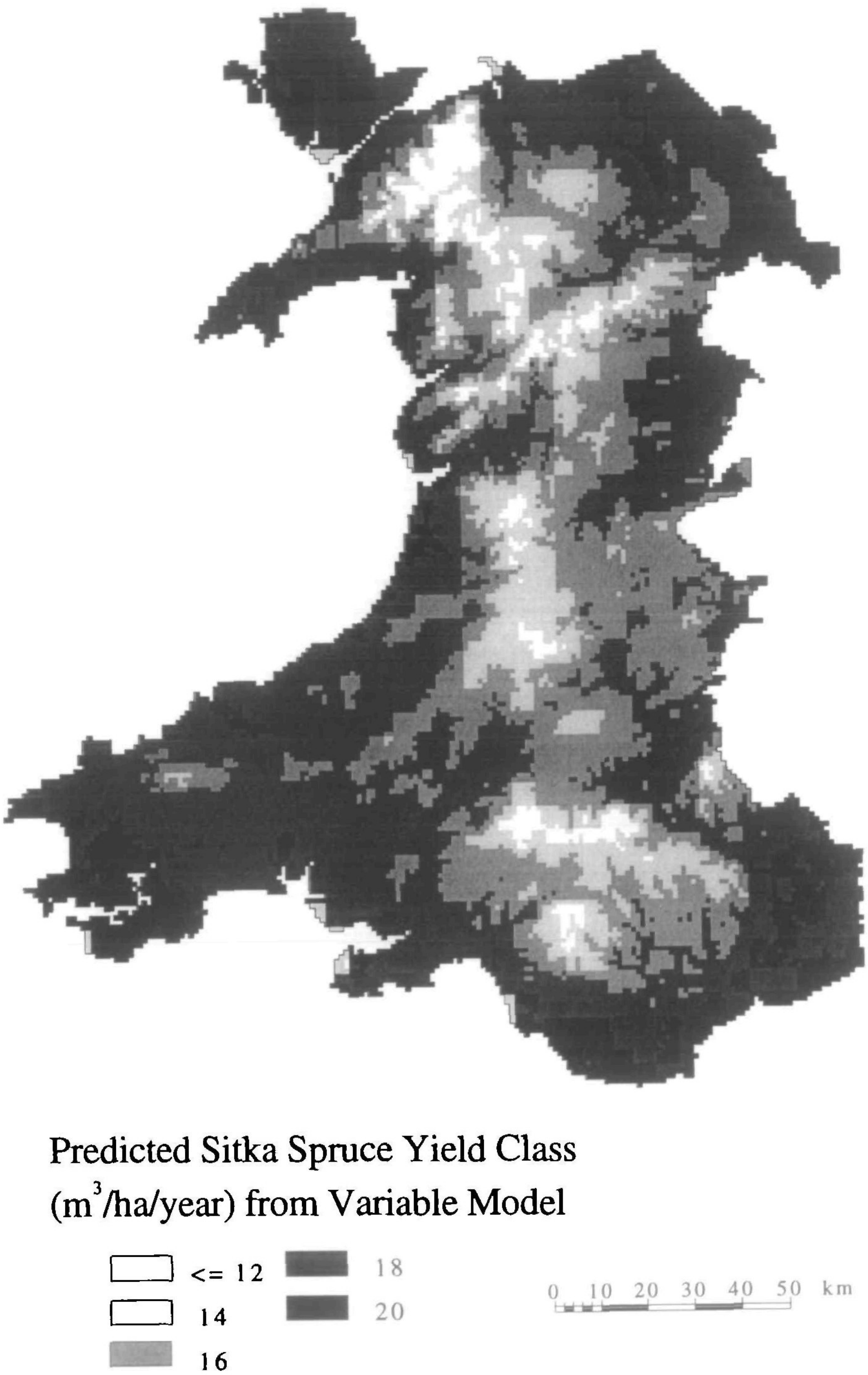

Figure 2. Map of predicted Yield Class (YC) for Sitka spruce from best fitting model not using principal components analysis (PCA) factors (Model 4) assuming *plantyr* = 75; *1st Rot* = 0.

Table 4: Predicted timber Yield Class (YC) from various maps of Sitka spruce YC

	Predictions from Model 4 (not using PCA factors)						Predictions from Model 3 (using PCA factors)					
	Plantyr = 0; *1stRot* = 1*		*Plantyr* = 75 *1stRot* = 1		*Plantyr* = 75; *1stRot* = 0		*Plantyr* = 0; *1stRot* = 1		*Plantyr* = 75; *1stRot* = 1		*Plantyr* = 75; *1stRot* = 0	
YC	Freq.†	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%	Freq.	%
2	10	0.049	–	–	–	–	–	–	–	–	–	–
4	46	0.224	1	0.005	–	–	–	–	–	–	–	–
6	367	1.785	15	0.073	1	0.005	225	1.094	–	–	–	–
8	2255	10.966	54	0.263	16	0.079	2253	10.957	1	0.005	–	–
10	4691	22.813	504	2.451	56	0.272	5332	25.930	418	2.033	–	–
12	8747	42.538	2524	12.274	554	2.694	10431	50.727	2628	12.780	359	1.746
14	4447	21.626	5106	24.831	2609	12.688	2322	11.292	6187	30.088	2524	12.274
16	–	–	9287	45.164	5209	25.332	–	–	10182	49.516	5915	28.765
18	–	–	3072	14.939	9416	45.791	–	–	1147	5.578	10329	50.230
20	–	–	–	–	2702	13.140	–	–	–	–	1436	6.983
Mean YC	11.38		15.12		17.05		11.21		14.90		16.98	
s.d.	2.81		2.81		2.81		2.65		2.65		2.65	

* The variables row defines the values of the variables *Plantyr* and *1stRot* used in each map, where *Plantyr* = year in which stand was planted (0 = 1920; 75 = 1995); *1stRot* = 1 if stand is the first planted in that compartment, = 0 otherwise (i.e. compartment is in second or subsequent rotation).
† The frequency columns refer to 1-km squares. Each map consists of 20563 such squares.
PCA, principal component analysis.

from considering first rotation to subsequent rotation plots so predicted YC increases. This trend holds for predictions based on either Model 3 or 4. Indeed the maps derived from the PCA-based model were, as expected, very similar to those derived from the no-factor model and are therefore not reproduced here.[31]

While the relationships described in our YC map conform to prior expectations, one possible weakness in these predictions deserves mention at this point. Note that in extrapolating from our models to the grid maps we are predicting beyond the range of available observations. Specifically the base data obtained from the SCDB does not contain observations for the most extreme upland locations (e.g. the top of Snowdon) as there are no sub-compartments in such areas. Such missing data points cannot play a part in the underlying statistical model which is producing estimates for these areas. Similarly there are relatively few Sitka spruce sub-compartments located on prime lowland sites (these being generally occupied by higher market priced agricultural activities). While the estimated relationships could hold for such extremes it may also be that our models tend to overpredict YC in highly adverse areas and underpredict for the most productive locations. Such a criticism can be levelled at all such statistically based models, not just those estimated in this paper. Furthermore, it is difficult to either prove or disprove such a charge. However, as we are predicting average YC over a minimum 1-km grid square this will tend to reduce any extremes in actual growth rate and therefore improve the validity of our predictions. Given this and the overall strength of the models produced (which, as noted, achieve a substantially higher degree of overall explanation than achieved previously in the non-GIS literature), we feel that these maps provide useful and credible predictions of YC over the majority of their extent.

Summary and conclusions

We have used the integrative and analytical capabilities provided by a GIS to estimate timber yield models for Sitka spruce based upon variables drawn from existing databases covering the entire extent of Wales. The methodology developed here has permitted a very substantial increase in both the area appraised and the number of observations and explanatory variables incorporated within the modelling process. The resulting models achieve a substantially higher degree of explanation of the data than that achieved by conventional, non-GIS based, approaches.[32] The relationships identified in our models strongly accord with prior expectations and, accepting the desirability of caution in all analyses and extrapolations, the present models permit prediction of yield for a wider range of terrain than has been previously considered within a single analysis. Additionally, the cartographic functions of the GIS permit the researcher to produce maps of predicted yield which accord directly with the approach currently being developed collaboratively between the various authorities concerned with forestry in the UK (Countryside Commission and Forestry Commission, 1996). Accordingly we feel that the methodology developed and presented in this paper may be of interest to academics, professional foresters and policy makers alike.

Future work will examine the issue of uncertainty within the data and its implications for yield predictions. We also aim to extend the methodology developed in this paper to the modelling of timber yield for further species (particularly broadleaves) and the mapping of associated financial values. Research will also be undertaken to assess the potential for using GIS techniques in the modelling of wider forestry values (including recreation[33]) and competing land use values, in particular those generated by agriculture, so permitting the identification of optimal areas for land use conservation.

Acknowledgements

The authors gratefully acknowledge the support of the Forestry Commission and the Soil Survey and Land Research Centre (Cranfield) in supplying data for this study. In particular we thank Chris Quine, Adrian Whiteman and Roger Oakes at the FC and Ian Bradley and Arthur Thomasson of the SSLRC. The authors are also grateful to Piers Maclaren of the New Zealand Forest Research Institute and to an anonymous referee for helpful comments on an earlier draft of this paper. This research was undertaken in part while Ian Bateman was on study leave at the Department of Resource Management, Lincoln University, New Zealand, to whom he is very grateful.

Notes

1. This work will be carried out in collaboration between the Countryside Commission, Forest Authority and English Nature (Countryside Commission and Forestry Commission, 1996).
2. Bateman and Lovett (1997) combine this output with a net present value optimizing model (in which discount rate, Yield Class and rotation length are allowed to vary) to produce maps of the timber value corresponding to predicted yield.
3. An important question given that this is the location of much of the existing stock of Sitka spruce.
4. The number of individual tree measurements is not reported.
5. Although not specified in this or the Macmillan paper this appears to be an unadjusted R^2 statistic.
6. While there has been recent interest in the application of GIS to agricultural modelling (Moxey, 1996) this is the first GIS based application to timber production utilizing multiple data sources and variables. An alternative approach using Landsat Thematic Mapper data is presented by Gemmell (1995).
7. Bateman (1996) details observation locations and descriptive statistics for variables used in the best fitting timber yield models discussed subsequently.
8. Dr Bill Corbett of the School of Environmental Sciences, UEA, and formerly of the SSEW, kindly advised in the merging of soil codes to produce a simple eight-category system which groups together similar soils. Further details are given in Bateman (1996).
9. 1-km referenced data on topex and wind hazard were kindly supplied by Chris Quine at the Forestry Commission's Northern Research Station, Roslin, to whom we are very grateful. For further details see Quine and White (1994).
10. Blakey-Smith *et al.* (1994) also discuss a funnelling variable which tends to have higher values in valley bottoms.
11. The authors are grateful to Julii Brainard of the School of Environmental Sciences, UEA, for assistance in creating the DEM.
12. This issue is addressed at length in Bateman *et al.* (in press).
13. This exercise revealed some relatively minor missing observations in the LandIS database. Measurements for these cells were proxied using interpolation and related techniques. For details see Bateman and Lovett (1997).
14. Discussion of the PCA approach is given in Johnston (1978), Norusis (1985) and Dunteman (1994). PCA is in fact a special case of factor analysis (Lewis-Beck, 1994).

15. Full details of this analysis are given in Bateman and Lovett (1997).
16. The communality is the sum of the squared factor loadings.
17. This is the default method in SPSS-X.
18. Bateman (1996) gives a worked example of this procedure.
19. Further details regarding the models estimated are given in Bateman and Lovett (1997).
20. Semi-log (dependent and independent), double log and quadratic forms were also tested and cross product terms investigated.
21. A counter explanation, given by a senior FC official who shall be nameless, is that this effect may also arise out of errors in the YC tables.
22. These included Chris Quine and Adrian Whiteman of the FC and Douglas Macmillan of the Macaulay Land Use Research Institute.
23. By which we mean the full procedure for entering variables into the model was repeated. This was necessary as we cannot be sure that the set of variables which best describes the untruncated dataset will also be optimal when all stands with a survey age of less than 10 years are omitted.
24. Intuitively one of these two may absorb the variation due to aspect so that it appears that there is little for the other to explain. However, entered separately the variables would be meaningless.
25. Indeed in Bateman (1996) the series of truncations is extended until this effect starts to increase R^2 statistics.
26. This study concerns species other than those under investigation and is consequently omitted from our literature review.
27. This is a higher degree of accuracy than that achieved by the thematic mapper approach of Gemmell (1995) who reports that roughly 75 per cent of predictions were within 25 per cent of actual growth rate. Here we have over 75 per cent of predictions within 2YC of actual, with no predictions in excess of 4YC of actual.
28. The means and standard deviations necessary for this operation were taken from the variable values for all the forestry sub-compartments. These will be somewhat different from those for the entirety of Wales but given the size of the forestry dataset, any discrepancy is liable to be minor.
29. See our previous discussion of possible interpretations of this effect.
30. Interestingly both Pembrey and Newborough (Anglesey) are the sites of large forests, underlining the point that forests are often confined to the most marginal land.
31. See Bateman and Lovett (1997) for details.
32. The improvement in fit between the 'conventional' analysis reported by Macmillan (1991) and the GIS-based reworking of the same data reported by Elston *et al.* (1997) reinforces this conclusion.
33. Some work has already been completed to this end, see Bateman *et al.* (1996).

References

Achen, C.H. 1982 Interpreting and using regression. *Quantitative Applications in the Social Sciences No 29*. Sage, London.

Bateman, I.J. 1996 An economic comparison of forest recreation, timber and carbon fixing values with agriculture in Wales: a geographical information systems approach. *PhD Thesis*, Department of Economics, University of Nottingham.

Bateman, I.J., Garrod, G.D., Brainard, J.S. and Lovett, A.A. 1996 Measurement, valuation and estimation issues in the travel cost method: A geographical information systems approach. *J. Agric. Econ.* 47(2), 191–205.

Bateman, I.J. and Lovett, A.A. 1997 Modelling and mapping timber yield and its value using geographical information systems: a study of Sitka spruce and beech. *Studies in Resource Management* No 1, Department of Resource Management, Lincoln University, Canterbury, New Zealand, 70pp.

Bateman, I.J., Lovett, A.A. and Brainard, J.S. in press *Applying Environmental Economics through Geographical Information Systems: Cost–Benefit Analysis of Converting Agricultural Land to Forestry*. Manchester University Press, Manchester.

Berg, J. van D. 1975 Moisture availability in forest soils and the growth of Japanese larch on moist podzolic soils. *Ned. Bosbouw Tijdschr.* 47, 159–163.

Berry, J.K. 1993 Cartographic modelling: the analytical capabilities of GIS. In *Environmental Modelling with GIS*. M.F. Goodchild, B.O. Parks and L.T. Steyaert (eds). Oxford University Press, Oxford, 58–74.

Blakey-Smith, J., Miller, D. and Quine, C. 1994 An appraisal of the susceptibility of forestry to windblow using GIS. *Proceedings of the GIS Research UK 1994 Conference*, 11–13 April 1994. Department of Geography, University of Leicester, Association of Geographical Information, 47–56.

Blyth, J.F. 1974 Land capability assessment for forestry in north-east Scotland. *PhD Thesis*, University of Aberdeen.

Blyth, J.F. and MacLeod, D.A. 1981 Sitka spruce (*Picea sitchensis*) in north-east Scotland: I. Relationships between site factors and growth. *Forestry* 54, 41–62.

Bradley, R.I. and Knox, J.W. 1995 The prediction of irrigation demands for England and Wales. *Proceedings of the GIS Research UK 1995*

Conference, 5–7 April 1995. Department of Surveying, University of Newcastle upon Tyne, Association for Geographical Information, 17–20.

Busby, R.J.B. 1974 Forest site yield guide to upland Britain. *Forestry Commission Forest Record No 97*. HMSO, London.

Carmean, W.H. 1975 Forest site quality evaluation in the United States. *Adv. Agron.* 4, 329–398.

Countryside Commission and Forestry Commission 1996 *Woodland Creation: Needs and Opportunities in the English Countryside*. Countryside Commission, Cheltenham, UK.

Dunteman, G.H. 1994 Principal components analysis. In *Factor Analysis and Related Techniques (International Handbooks of Quantitative Applications in the Social Sciences, Volume 5)*. M.S. Lewis-Beck (ed.). Sage Publications and Toppan Publishing, London, 105–146.

Eastman, J.R. *IDRISI Version 4.1*. Clark University, Worcester, Massachusetts.

Edwards, M.V. 1957 Larix leptolepis. In *Exotic Forest Trees in Great Britain. Forestry Commission Bulletin* No 30. HMSO, London, 69–73.

Elston, D.A., Jayasinghe, G., Buckland, S.T., Macmillan, D.C. and Aspinall, R.J. 1997 Adapting regression equations to minimise the mean squared error of predictions made using covariate data from a GIS. *Int. J. Geograph. Inf. Sc.* 11(3), 265–280.

Forestry Industry Committee of Great Britain (FICGB) 1992 *The Forestry Industry Yearbook 1991–1992*. FICGB, London.

Gale, M.F. and Anderson, A.B. 1984 High elevation planting in Galloway. *Scott. For.* 38(1), 3–15.

Gemmell, F.M. 1995 Effects of forest cover, terrain, and scale on timber volume estimation with thematic mapper data in a Rocky Mountain site. *Remote Sensing Environment* 51, 291–305.

Goodchild, M.F. 1993 Data models and data quality: problems and prospects. In *Environmental Modelling with GIS*. M.F. Goodchild, B.O. Parks and L.T. Steyaert (eds). Oxford University Press, Oxford, 94–103.

Grace, J. 1977 *Plant Responses to Wind*. Academic Press, London, 204pp.

Hägglund, B. 1981 Evaluation of forest site productivity. *Forest Abstracts* 42, 515–527.

Hart, C.E. 1991 *Practical Forestry*, 3rd edn. Alan Sutton Publishing, Stroud.

Jarvis, N.J. and Mullins, C.E. 1987 Modelling the effects of drought on the growth of Sitka spruce in Scotland. *Forestry* 60, 13–30.

Johnston, R.J. 1978 *Multivariate Statistical Analysis in Geography*. Longman, London.

Jones, R.J.A. and Thomasson, A.J. 1985 An agroclimatic databank for England and Wales. *Technical Monograph No 16*. Soil Survey of England and Wales, Harpenden.

Kilpatrick, D.J. and Savill, P.S. 1981 Top height growth curves for Sitka spruce in Northern Ireland. *Forestry* 54, 63–73.

Lewis-Beck, M.S. 1980 Applied regression: an introduction. *Quantitative Applications in the Social Sciences* No 22, Sage, London.

Lewis-Beck, M.S. 1994 Editor's introduction. In *Factor Analysis and Related Techniques (International Handbooks of Quantitative Applications in the Social Sciences, Volume 5)*. M.S. Lewis-Beck (ed.). Sage Publications and Toppan Publishing, London, 1–20.

Maclaren, J.P., Grace, J.C., Kimberley, M.O., Knowles, R.L. and West, G.G. 1995 Height growth of *Pinus radiata* as affected by stocking. *N. Z. J. For. Sc.* 25(1), 73–90.

Macmillan, D.C. 1991 Predicting the general yield class of Sitka spruce on better quality land in Scotland. *Forestry* 64(4), 359–372.

Malcolm, D.C. 1970 Site factors and the growth of Sitka spruce. *PhD Thesis*, University of Edinburgh.

Malcolm, D.C. and Studholme, W.P. 1972 Yield and form in high elevation stands of Sitka spruce and European larch in Scotland. *Scott. For.* 26(4), 296–308.

Mayhead, G.J. 1973 The effect of altitude above sea level on the Yield Class of Sitka spruce. *Scott. For.* 27, 231–237.

Moxey, A. 1996 Geographical information systems and agricultural economics. *J. Agric. Econ.* 47(1), 115–116.

Norusis, M.J. 1985 *SPSS-X: Advanced Statistics Guide*. McGraw-Hill, New York.

Page, G. 1970 Quantitative site assessment: some practical applications in British forestry. *Forestry* 43, 45–56.

Quine, C.P. and White, I.M.S. 1994 Using the relationship between rate of tatter and topographic variables to predict site windiness in upland Britain. *Forestry* 67(3), 245–256.

Rudeforth, C.C., Hartnup, R., Lea, J.W., Thompson, T.R.E. and Wright, P.S. 1984 Soils and their use in Wales. *Bulletin No. 11*, Soil Survey of England and Wales, Harpenden.

Soil Survey of England and Wales 1983 *Legend for the 1:250,000 Soil Map of England and Wales*. Soil Survey of England and Wales, Rothamsted, Harpenden, Herts.

Tranquillini, W. 1979 *Physiological Ecology of the Alpine Treeline—Tree Existence at High Altitudes with Special Reference to the European Alps*. Springer-Verlag, Berlin.

Tyler, A.L., Macmillan, D.C. and Dutch, J. 1966 Models to predict the General Yield Class of

Douglas fir, Japanese larch and Scots pine on better quality land in Scotland. *Forestry* 69(1), 13–24.

UK Government 1995 *This Common Inheritance, UK Annual Report 1995*, Cm 2822, HMSO, London.

Worrell, R. 1987a Geographical variation in Sitka spruce productivity and its dependence on environmental factors. *PhD Thesis*, Department of Forestry and Natural Resources, University of Edinburgh.

Worrell, R. 1987a Predicting the productivity of Sitka spruce on upland sites in northern Britain. *Forestry Commission Bulletin No 72*. HMSO, London.

Worrell, R. and Malcolm D.C. 1990a Productivity of Sitka spruce in northern Britain: 1. The effects of elevation and climate. *Forestry* 63(2), 105–118.

Worrell, R. and Malcolm, D.C. 1990b Productivity of Sitka spruce in northern Britain: 2. Prediction from site factors. *Forestry* 63(2), 119–128.

Received 26 March 1997