Modelling and Mapping Timber Yield and Its Value Using Geographical Information Systems: *A Study of Sitka Spruce and Beech*

Ian J Bateman
Andrew A Lovett

Ian Bateman and Andrew Lovett are at
the Centre for Social and Economic Research on the Global Environment (CSERGE)
and the School of Environmental Sciences,
University of East Anglia, Norwich, NR4 7TJ, UK.
Email address: I.BATEMAN@UEA.AC.UK.

Department of Resource Management/Lincoln Environmental
Lincoln University

September 1997

Studies in Resource Management No 1

Lincoln Environmental
PO Box 84
Lincoln University
Canterbury
New Zealand

ISSN 1174-2437
ISBN 1-86931-038-1

Lincoln Environmental and the Department of Resource Management are involved in research, consulting and teaching, and are based at Lincoln University. Research is focused on the development of conceptually sound methods for resource use that may lead to a sustainable future. Lincoln Environmental and the Department of Resource Management offer staff the freedom of enquiry. Therefore, the views expressed in this Paper are those of the authors and do not necessarily reflect those of the organisations.

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. We are also grateful to Piers Maclaren of the Forestry Research Institute (FRI), New Zealand, for helpful comments on an earlier draft of this paper, the usual disclaimer applies.

This research was completed while Ian Bateman was on study leave at the Department of Resource Management, Lincoln University, New Zealand, to whom he is very grateful.

CONTENTS

Page

LIST OF FIGURES:

Page

LIST OF TABLES:

Page

1 INTRODUCTION

Recent changes in UK forestry policy (Rural White Paper, 1995) Countryside Commission, (1996) 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. Furthermore, a recent joint paper from the Countryside and Forestry Commissions (1996) indicates that much of this planting will occur in lowland areas and is likely to be guided by a series of forthcoming maps indicating optimal planting areas[1]. Clearly the incorporation of information concerning spatial variations in tree productivity is a desirable feature for such decisionmaking. This paper describes various models of the production of timber for two tree species: Sitka spruce and beech, chosen as representative softwood and broadleaf species. The methodology developed as part of this study differs from previous approaches in that it uses a geographical information system (GIS) to utilise large scale existing databases covering a very large and diverse study area; the whole of Wales. Furthermore, the cartographic facilities of the GIS are employed to produce maps of predicted timber yield for the entire study area. This maps are presented both in terms of quantitative timber production and the value of that yield thus facilitating ready incorporation within the declared decisionmaking system.

In Section 2 we present a brief review of previous studies. These have been based upon relatively small scale surveys of tree growth, furthermore, they have also generally been confined to comparatively small areas and often to one topographic region, e.g. upland areas. Section 3 presents details regarding the various datasets used in this study and discusses how these data were transformed for the purposes of subsequent regression analysis. Results from our models of Sitka spruce and beech growth rates are presented in Sections 4 and 5 respectively while Section 6 presents and analyses GIS created map images of predicted yield class. Finally Section 7 applies the findings of previous research regarding the value of timber yield to produce monetised equivalents of these results.

[1] This work will be carried out in collaboration between the Countryside Commission, Forest Authority and English Nature (Countryside and Forestry Commission, 1996).

2 LITERATURE REVIEW AND METHODOLOGICAL OVERVIEW

2.1 Literature Review

Clearly tree growth rates will depend upon a variety of species, environmental and silvicultural factors. Early work in this field relied on simple rules of thumb reliant upon relatively little supporting data (Busby, 1974) or analyses of single factors. Reviews across 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 routes 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 H④gglund, 1981), the estimation of statistical models across the 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). Amongst such investigations we could find no examples concerning the productivity of beech and believe the model presented subsequently to be the first such investigation of this species. However, there has been more attention paid to the other species under analysis; Sitka spruce, which has been separately analysed both by Richard Worrell (then of the University of Edinburgh) and Douglas Macmillan and colleagues (Macauley Land Use Research Institute, MLURI)[2].

While there had been a number of earlier considerations 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) is notable as being the first to adopt a multiple regression approach across a highly 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 areas. Such concerns arise both because of the growing conditions prevailing in the case study area of upland Scottish and because of the specific focus of Worrell's experiment. Worrell was mainly interested in detecting the influence of elevation upon YC in upland areas[3]. To this end he selected 18 principal sample sites[4], all of which had relatively steep slopes, and took measurements along a vertical transect at each site. By locating samples at sites ranging from 50 m to 600 m above sea level a very strong, central tendency relationship with elevation could be established. However, such a model is only applicable to similar, steeply sloping sites (strictly speaking, only the subset of those found within Scotland), and is not generalisable to the plethora of environmental conditions found in an area the size of Wales.

---

2 We are grateful to both Richard and Douglas for extensive discussions of their work.

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.

A similar, though less extreme, consideration prevents us applying the findings of Macmillan (1991). Here again the study is geographically confined, this time to lowland Scotland, although the 121 sites used are not selected to emphasise the influence of any particular explanatory variable and are therefore somewhat more generalisable within lowland areas. However, while this would, in many cases, be adequate, with respect to our study area the topographic variability of Wales means that a model based purely upon lowland data is insufficient for our needs. Nevertheless, the Macmillan paper is interesting for another reason in that it comprises multiple regression with a prior principal components analysis (PCA) of explanatory variables, reporting a final degree of explanation of R2 = 36.8% [5].

A short note regarding model fit is justified here. The YC of a plantation is its average annual growth rate assessed over an optimal rotation (that is the rotation length which maximises net present value (NPV) at a given discount rate)[6]. YC is therefore given in m^3 ha^{-1} yr^{-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 this approach to measurement does induce variance into the dataset and therefore makes high degrees of explanation difficult to attain. As such the absolute value of fit statistics such as R2 should be treated with some caution and instead we should consider relative degrees of fit compared to those attained in other studies. To our knowledge the Macmillan (1991) result represents the best fitting previous model for Sitka spruce assessed over diverse terrain. Given this any yield model which gives a degree of explanation in excess of about 40% would represent a substantial improvement over prior UK studies[7].

2.2 Overview of Modelling Approach

These prior studies provide very useful indications regarding the likely 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 a PCA and standard multiple regression methodology. However, in other respects the methods of Worrell and Macmillan were not appropriate to the specific types of question asked in our research. Our central aim is to identify areas over a large and diverse area which might be suitable for conversion out of agriculture and into forestry. This necessitated the development of a methodology which was capable of producing estimates for both upland and lowland areas. In order to provide input observations across such a diversity of terrain it was decided to focus on the entire area of Wales, a country noted for its variation in land types.

---

5 Although not specified this appears to be an unadjusted R2 statistic.

6 An analysis of the relationship between discount rate and optimal rotation length across yield class and species is given in Bateman (1996). The definition of NPV used here and underpinning subsequent results is conventional (in particular it conforms to that of the FC FIAP investment appraisal model) in that it considers both costs and revenues but costs are not sensitive to opportunity costs, or spatial variations in felling and transportation costs (although a GIS analysis of this aspect is planned) while revenues are not adjusted for risk and quality variations are only incorporated to the extent which they affect the price-size curve. For full details see (Bateman, 1996).

7 In conversations with Piers Maclaren (FRI, New Zealand) it seems likely that application of the methodology developed in this paper to data which was not subject to this measurement problem would result in a very substantial increase in model fit.

Our methodology entailed the development of a GIS based approach to the analysis of existing large forestry and land characteristic databases[8]. This takes our base 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's estate[9]. As this covers both upland and lowland sites, results from such a model is more generalisable 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[10] (described subsequently), which has complete national coverage (unlike the SCDB which only has data for forest areas). Our regression results can then be readily extrapolated to all other areas of Wales, including those not presently under forestry. The one disadvantage of such an approach is that, unlike the previous studies, here the data is not collected directly by the researcher but by many others, often over an extended period. While this can be viewed as not entirely negative, subsequent modelling indicated that allowances had to be taken for variance induced by such an approach to measurement.

---

8 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 utilising multiple data sources and variables. An alternative approach using Landsat Thematic Mapper data is presented by Gemmell (1995).

9 We are greatly obliged to Adrian Whiteman, Chris Quine and the Forestry Commission for use of the SCDB.

10 We are greatly obliged to Arthur Thomasson, Ian Bradley and the Soil Survey and Land Research Centre (Cranfield) for use of LandIS.

3 DATA AND DATA MANIPULATION

This research relies upon a diversity of data sources. In addition to the SCDB and LandIS databases, further environmental and topographic data was obtained from a variety of sources. In this section we describe these data and how they were manipulated prior to consideration within the subsequent statistical investigation of tree growth. It is important to remember that, while the SCDB holds detailed data regarding individual plantation sites, it does not extend to the majority of Wales which is unplanted. Therefore the environmental variables given in the SCDB are, for our purposes, unsuitable predictions of YC as complete land surface coverages for these variables are not available and therefore cannot be used for extrapolation of predictions to presently unplanted areas. The complete coverages of variables held in LandIS and the other data described subsequently are therefore needed to allow for this extrapolation of regression results.

3.1 SCDB Data

The SCDB is the FC's central forest inventory detailing observations for all stands in the Estate[11]. 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 also shows how certain of this data was manipulated to produce further (often dummy) variables. In doing this, one-way analyses of variance on the dependent variable (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 these are not (with the exception of altitude) variables for which uninterrupted national coverages exist, 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 will not fit the YC data as well as those using the site factor information 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 and 766 beech sub-compartments were used in our regression analysis[12]. 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.

---

[11] The FC were, as always, most willing to allow access to their data, for which we are most grateful.

[12] Bateman (1996) details observation locations and descriptive statistics for variables used in the best fitting Sitka spruce and beech YC models discussed subsequently.

Table 1: Variables obtained from the SCDB (except where shown otherwise). Ordered as per the 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 PRP; = 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	Used to identify target species
Planting year	Discrete variable	*plantyr*: year in which stand was planted
Survey year	Discrete variable	*survyr*: year in which stand was surveyed[1]
Yield class	Even number	*YC*: tree growth rate: average m^3/ha/year over an optimal rotation - the dependent variable (see previous definition of terms)
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 affect 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 (Note for BE this includes some subsequent rotations.) *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
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, AONB/NSA, OthESA; = 0 otherwise *Park* = 1 if any of Nat Park, F Park, SSSI; = 0 otherwise

Note:

1. Supplied by Chris Quine at the FC Northern Research Station, Roslin, to whom we are very grateful.

3.2 LANDIS Data

3.2.1 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 18th and early 19th 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 1980's become the Soil Survey and Land Research Centre (SSLRC), commenced a five-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 digitised, 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[13].

3.2.2 The Data

Detailed definitions, derivations and accuracy of the data extracted from LandIS are presented in Bateman (1996). These are summarised 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 the 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 are a problem both because of their 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[14]. Details of the final categorisation are given in Table 3.

---

[13] Agreement to use the data was obtained from Arthur Thomasson in 1987. However, at the time the SSEW was undergoing the trauma of being privatised, 'downsizing', and becoming part of what is now Cranfield University. We are grateful to Ian Bradley and R.J.A. Jones of the SSLRC for subsequently honouring this commitment and to the School of Environmental Sciences/UEA for funding the entailed data transfer costs.

[14] 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.

Table 2: Variables obtained from LandIS

Variable name	Label	Definition
Accumulated temperature	*Acctemp*	Average annual accumulated temperature (in °C) above 0°C[1]
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 the 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	See Table 3	SSLRC soil type classification code

Note:

1. Tree grow is zero at 0°C and conversations with Piers Maclaren (FRI, New Zealand) showed that there was some uncertainty regarding the minimum temperature at which growth commences. However, here the variable is acting merely as a temperature and its absolute base is of no consequence.

Table 3: Soil type codes

Soil type	Variable label	Subsumed SSLRC soil codes[1]
Lowland lithomorphic	*soil 1*	361
Brown earths	*soil 2*	514, 541, 551, 561, 571,
Podzols	*soil 3*	572
Surface water gley	*soil 4*	611, 631
Stagnogley (perched watertable)	*soil 5*	651, 654, 711, 712, 713, 721
Ground water gley	*soil 6*	813
Peats	*soil 7*	1011, 1013
Upland lithomorphic	*soil 8*	311
Urban	*n/a*	n/a

Note:

1. Here we have only listed categorisations down to the subgroup level (as defined in Avery, 1980). LandIS further subdivides these into numerous soil associations as detailed in SSEW (1983).

Subsequent statistical analysis suggested that, if anything, merging of soil codes could have been taken even further and some combinations of the variables given in Table 3 are considered later.

3.3 OTHER Data

3.3.1 Topex and Wind Hazard [15]

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). Here then a low angle sum (low topex value) represents a high degree of exposure. The resultant variable was labelled *Topex 1 km*.

Blakey-Smith *et al.* (1994) define wind hazard on the basis of four factors[16]:

i) Wind zone - delimited on a GB map;

ii) Elevation - high values increasing wind hazard;

iii) Topex - as defined above;

iv) Soil type - those which relatively speaking promote growth (brown earths, podzols, etc.) being low wind hazard while those which restrict growth (gleys, peats, etc.) are higher wind hazard.

The resultant continuous variable (*Wind1km*^{2}: literally the wind speed calculated for a given 1 km square) is inversely linked with tree productivity and growth rates.

[15] One 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.

[16] Blakey-Smith *et al.* (1994) also discuss a funnelling variable which tends to have higher values in valley bottoms.

3.3.2 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 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 GIS digital elevation model (DEM)[17]. The DEM is a GIS-based digital image of the topography of Wales. This was created from three principal data sources:

i) The Bartholomew 1:250,000 database for the UK. This gives 50 m contours up to 1000 m after which 100 m intervals are reported;

ii) Spot heights from Bartholomew's paper maps. These were particularly useful for assessing the predictive accuracy of the DEM and for addressing the problems associated with identifying mountain tops;

iii) Spot heights of plantations from the SCDB. This provided additional information used in the interpolation of heights between contours.

After exhaustive accuracy testing of the resulting elevation variable (*Wselvgr2*), the authors of the DEM also used it to provide two further GIS surface variables: slope angle (*Dsl2*) and aspect angle (*Wsaspgr2*). Data on all these variables was supplied at a 500 m resolution.

3.4 Creating GIS Surfaces for Explanatory Variables

Prior to the regression analysis two fundamental issues had to be addressed regarding the definition of a common extent and common resolution for the environmental variables. While the geo-referenced data obtained from the LandIS and non-SCDB sources detailed above were converted into GIS surfaces, inspection of these showed that the various data obtained differed both in its geographical extent and spatial resolution.

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 interpolation facilities available within the GIS made conversion to a common scale relatively straightforward[18], choice of that scale was a matter for some deliberation. While standardisation upon the smallest unit (100 m) is given the interpolation capabilities of a GIS, perfectly feasible, it did not seem a sensible choice. The 100 m reference used in the SCDB is, the FC admit, spuriously precise. Furthermore, use of a larger scale would, in the case of say the DEM entail an averaging out of predictions which was likely to avoid problems associated with single point estimates. However, aggregation up to the 5 km scale of the coarsest data was felt to be likely to result in a loss of valid and interesting detail. Consequently a 1 km grid was settled upon and all data were interpolated to this resolution.

---

[17] The DEM was custom-created for this research for which the assistance of Julii Brainard of the School of Environmental Sciences, UEA, is gratefully acknowledged.

[18] This is somewhat misleading. In reality careful interpolation is a highly time consuming exercise involving the iterative reassessment of a range of interpolation decay weights with actual versus predicted verification. Whilst advances in processor speed have considerably improved the time which such analyses take, they are still somewhat arduous to undertake. This issue is addressed at length in Bateman, Lovett and Brainard (forthcoming).

The spatial extent of Wales was defined by rasterising on to a 1 km grid the Bartholomew's vector data for the coast and border with England. This resulted in a GIS surface consisting of 20,563 land cells which was used as a mask file to extract 1 km values for each of the variables in the LandIS and non-SCDB datasets described above. However, in undertaking this exercise it was found that, with the exception of the custom written DEM and associated variables, virtually all variables were missing at least some observations. Given our principal aim of extrapolating regression results across the whole of Wales, this situation had to be rectified.

In some cases the problem of missing data was relatively minor. With respect to the topex and wind hazard data, which was supplied in a 1 km rasterised form, just 103 of the required 20,563 cells were missing, all of these being located at the tips of various peninsula. Here interpolation from surrounding cells provided a ready solution to this problem.[19]

The missing data problem was more serious in the LandIS database both because more data tiles were missing and because of their larger, 5 km, resolution. Using the OS grid, Wales extends to some 942 of these tiles[20]. Only three of the variables described in Table 2 had data for all of these tiles. Table 4 lists omissions from this database.

Table 4: Omissions from the LandIS database

Variable label[1]	No. of 5 km land tiles[2] supplied	% of all Welsh 5 km land tiles
Acctemp, Growseas, Grazseas	942	100.0
Rainfall, Retmed, Retwet, Retdry, Endmed, Endwet, Enddry, Fcapdays	898	95.3
Mdefgra, Mdefce,	858	91.1
Avwatgra, Avwatcer, Avwatpot, Avwatsb	812	86.2
Workabil, SprMWD, AutMWD, Soils	780	82.8

Notes:

1. From Table 2.
2. This includes any 5 km OS grid square containing any Welsh land (some may be mainly in England or in the sea).

As before the majority of omissions were clustered around the Welsh coast. However, to allow our extrapolation analysis to proceed, such empty squares had to be filled. Inspection of nearby cells for which data was available showed strong spatial trends in all variables with the exception of soil type. Consequently empty cells for all non-soil variables were filled using

[19] This and subsequent interpolation operations were conducted by Andrew Lovett, to whom I am very grateful.

[20] Note that coastal and border tiles will not be fully filled. This accounts for the implicit difference in the extent of this coverage as opposed to the 1 km mask.

interpolation techniques. These were clearly inappropriate for soil type which tended to change abruptly and was consequently not interpretable from other data points.

Interrogation of the Bartholomew's digital database identified 19 of the 162 5-km grid squares missing soil values as being urban areas in which soil surveys had not been undertaken. The remaining missing values were filled by consulting the SSEW 1:250,000 Soils of Wales paper map.

All the LandIS data was the interpolated on to a 1 km grid and our coast/border outline used to delete squares which fell outside this extent.

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 via a GIS macro command.

3.5 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) use conventional regression analysis, Macmillan (1991) first subjects explanatory variables to a principal components analysis (PCA) before entering the resultant factors within a regression analysis. It was decided that a comparison of these two approaches would be of interest and so our data was made the subject of a PCA.

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) and we shall use the terms 'factor' and 'component' interchangeably in the following discussions.

In essence PCA attempts to identify patterns of covariance so that trends within a comparatively large number of variables are summarised 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. Norusis (1985) identifies four steps conducted in PCA:

i) A correlation matrix is prepared so that variables which do not appear to be related to others within the dataset can be identified (suppression-type problems can also be identified at this stage). The appropriateness of PCA can also be assessed at this point;

ii) The number of factors necessary to adequately represent the dataset is identified. Clearly unless this is substantially less than the number of variables then the exercise is of little value;

iii) The factors may be transformed (rotated) to make them more interpretable;

iv) Factor scores are computed to indicate how individual observations perform on each factor. These may then be used as predictors within a regression model.

However, before we could start our PCA study we were concerned to first consider whether a single analysis might be appropriate for both Sitka spruce and beech sites or not. By dividing our data into two sets according to whether sites were planted with Sitka spruce or beech it was noted that the former generally faced more adverse environmental conditions to the latter. Table 5 details summary statistics for certain environmental variables divided according to site species[21].

Table 5 indicates that, on average, Sitka spruce sites are at higher elevation, colder, wetter and less workable than their beech counterparts. This is not surprising as we would expect broadleaf plantations to be generally confined to relatively lowland areas while hardy species such as Sitka spruce have been planted over a wide variety of sites. This substantial difference in site characteristics suggested that separate rather than common PCA investigations of explanatory variables should be conducted.

Table 5: Description of environmental variables for forestry sub-compartments by species (SS = Sitka spruce; BE = beech)

Variable	Species	Mean	Median	St. dev.	Coef. of Variation
Wselvgr2	SS	323.70	333	102.72	31.7
	BE	196.83	183	99.90	50.8
Windlkm2	SS	14.890	14.96	2.36	15.9
	BE	12.009	11.89	2.25	18.8
Acctemp	SS	1401.2	1385.0	243.70	17.4
	BE	1591.8	1600.0	240.90	15.1
Rainfall	SS	1713.6	1705.0	433.80	25.3
	BE	1386.5	1284.0	423.50	30.5
Fcapdays	SS	313.39	322.0	48.27	15.4
	BE	267.29	258.0	56.19	21.0
MdefGra	SS	25.30	20.0	25.54	100.9
	BE	57.00	53.0	38.20	67.0
AutMWD	SS	2.122	0.0	9.66	455.2
	BE	16.623	0.0	24.23	145.8

[21] Descriptive statistics for the full range of environmental variables as used in our best fitting YC models for Sitka spruce and beech are detailed in Bateman (1996).

3.5.2 Defining Input Variables

While most of our environmental variables were in a form amenable to initial consideration within a PCA, this was not true of our aspect variable (*Wsaspgr2*) which was recorded in terms of compass direction. This is unsuitable for PCA which simply focuses on linear correlations so that values of 1° and 359° would be interpreted as very different rather than virtually identical. The solution adopted was to calculate both the sine and cosine of aspect (*Sinasp* and *Cosasp* respectively) and include these variables in the PCA instead. The combination of these two transformations allows aspect to be interpreted in linear terms.

When an initial attempt was made to undertake PCA using the FACTOR command of SPSS-X, a warning message of the form 'ill conditioned data matrix' was encountered (though results were generated). Further investigation suggested that this situation might reflect either:

i) variables with a very small coefficient of variation (e.g. <0.002%); or
ii) high correlations between a number of the input variables.

Subsequent calculations suggested that the former was unlikely to be a problem (see Table 5) but that the latter might well be. It is almost ironic that while PCA searches out for relationships between variables, if some of these are extremely strong then calculation problems can arise. To investigate this possibility Pearson correlations matrices were calculated for both Sitka spruce and beech datasets of environmental variables (see Bateman, 1996, for details of these analyses). Inspection of these results identified five groupings of correlated variables as follows:

Group 1: *Acctemp; Growseas; Grazseas
Group 2: *Rainfall; RetWet; *RetMed; RetDry; *FcapDays; EndWet; *EndMed; EndDry
Group 3: *MdefGra; MdefCer; MdefSbpt
Group 4: *AvwatGra; AvwatCer; AvwatPot
Group 5: AutMWD; *SprMWD

Within each of these groups, one or more (depending upon the degree of correlation) variables were chosen to be entered into the PCA (marked * above). Choice of 'input' variable depended upon the biological plausibility of a relationship with YC, the degree of correlation with other variables and the consequent requirement that the resultant data matrix should not be ill-conditioned. All these conditions were satisfied. In addition to the above, seven other less correlated input variables were also identified for inclusion within the PCA (*Workabil, Wselvgr2, Dsl2, Topex1km, Wind1km2, Cosasp, Sinasp*).

This analysis resulted in a consistent list of predictor variables for both our Sitka spruce and beech datasets with the single exception of AutMWD and SprMWD, both of which could be included for spruce but not beech. As it was considered important to use the same set of variables for each species, the weaker AutMWD variable was deleted from both PCA studies.

3.5.3 PCA for Sitka Spruce Environmental Variables

i) Examining the correlation matrix

The first task was to calculate the degree of sampling adequacy for both individual variables and the entire sample. This shows the extent to which individual variables can be explained by other variables and the extent to which factors describing the variation of the overall dataset can be created. With respect to the entire sample this is given by the Kaiser-Mayer-Olkin (KMO) measure of sampling adequacy. KMO compares the magnitude of observed correlation coefficients to partial correlation coefficients. If partial correlations are relatively high then KMO will be low suggesting that correlations between pairs of variables cannot be explained by other variables. Conversely when partial correlation coefficients are low, KMO is high and communality is high. KMO ranges from 0 (totally inadequate) to 1 (perfectly adequate) with values below 0.5 indicating samples for which PCA is inappropriate. Calculating KMO for the Sitka spruce dataset gave a value of 0.76 which Kaiser (1974) describes as middling to meritorious. Sampling adequacy for individual variables was confirmed through inspection of the anti-image correlation (AIC) matrix (see Bateman, 1996, for details).

ii) Component extraction

Here linear combinations of the variables are formed. The first principal component (or factor) will be that which accounts for the largest amount of variance in the data. The second factor accounts for a lesser amount of variation and is uncorrelated with the first. We can carry on defining factors up to the number of variables in the sample but this would be a rather pointless exercise. Therefore we need to consider the amount of variation explained by each factor and devise some rule to determine where we will draw the line with respect to the minimum number of factors to which we can reduce our input variables. The most common approach is to standardise all variables and factors with a mean of zero and variance of one. This will mean that the total standardised variance of the sample will be equal to the number of input variables, here 15. The total amount of standardised variance explained by any one factor (known as its eigenvalue) can then be compared to the total standardised variance of the sample and the percentage variance explained calculated.

Factors which have eigenvalues of less than 1 perform less well than simple variables (which are constrained to have a standardised variance of 1) and so this is commonly used as a cut-off point below which factors are discarded. In this case the first five factors all satisfy this criteria and account for 76.9% of the total variance in the sample.

iii) Improving interpretability: factor rotation

Interpretation of the factors may be achieved by calculating a 'factor matrix' detailing the correlation coefficient or 'component loading' between each factor and each variable. This is then 'rotated' using the 'varimax' method of Kaiser (1958) to minimise the number of variables having a high loading on each factor thereby enhancing the interpretability of each factor. Table 6 details component loadings for our rotated factor matrix.

Table 6: Rotated factor matrix: Sitka spruce sub-compartments

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Acctemp	-0.34	0.15	-0.28	-0.59	-0.38
Rainfall	0.92	0.20	0.13	0.08	-0.06
RetMed	-0.94	-0.14	-0.18	-0.09	0.05
EndMed	0.91	0.20	0.17	0.13	-0.05
FcapDays	0.94	0.15	0.17	0.06	0.05
MdefGra	-0.77	0.11	-0.14	-0.19	-0.22
AvwatGra	0.16	-0.04	0.88	-0.04	-0.04
Workabil	0.19	-0.10	0.87	-0.03	-0.06
SprMWD	-0.51	0.27	0.12	0.15	-0.16
Wselvgr2	0.16	-0.38	0.51	0.41	0.38
Dsl2	0.10	0.73	0.06	0.06	0.31
Topex1km	0.21	0.81	-0.07	0.02	0.04
Wind1km2	0.00	-0.78	0.36	0.18	0.23
Cosasp	-0.03	0.13	-0.09	-0.10	0.81
Sinasp	0.07	0.05	-0.16	0.84	-0.22

Inspection of the PCA factors detailed in Table 6 indicated that they were relatively easy to interpret as follows:

Factor No.	Label
1	Soil wetness/rainfall
2	Steeper slopes/low windiness
3	Waterlogging/workability/high elevations
4	Cold/sine aspect
5	Cosine aspect/elevation

The 'communality' or proportion of variance in each input variable which is 'explained' by the five factors[22] was also calculated. This indicated that the only variable which is relatively poorly explained is *sprMWD* (communality = 0.39), all other variables having a reasonable proportion of variance explained by our five factors (mean communality = 0.80).

iv) Calculating factor scores

The factor score coefficient matrix was calculated via the regression method described by Norusis (1985)[23]. Factor scores (which indicate the position of each observation (here each sub-compartment) on the extracted, rotated factors) were then calculated in the normal manner (Bateman, 1996, gives examples for both our Sitka spruce and beech factor matrices). The site specific factor scores obtained by this process can then be entered directly into our YC regression model as the environmental explanatory variables.

3.5.4 PCA for Beech Environmental Variables

The PCA procedure applied to the beech sub-compartments was identical to that used for the Sitka spruce sites and so results will be presented in brief.

---

[22] The communality is the sum of the squared factor loadings.

[23] This is the default method in SPSS-X.

i) Examining the correlation matrix
The KMO measure of sampling adequacy was calculated to be 0.77, a figure similar to that for Sitka spruce. Inspection of the AIC matrix for beech. Generally these are indicated that sampling adequacy for individual variables was generally as desired for a successful PCA although the individual values for *Avwatgra* and *Workabil* were rather lower than for Sitka spruce (see Bateman, 1996, for details).

ii) Component extraction
As before five factors satisfied our criteria for extraction.

iii) Improving interpretability: factor rotation
A rotated factor matrix was calculated as before and is detailed in Table 7.

Table 7: Rotated factor matrix - beech sub-compartments

	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
Acctemp	-0.44	-0.60	0.06	-0.38	-0.01
Rainfall	0.94	0.03	0.02	0.19	0.03
RetMed	-0.96	-0.09	-0.03	-0.15	-0.02
EndMed	0.94	0.08	0.04	0.29	0.04
FcapDays	0.96	0.10	0.05	0.16	0.02
MdefGra	-0.85	-0.35	0.05	-0.14	0.12
AvwatGra	-0.03	0.02	0.95	0.03	0.02
Workabil	0.10	0.01	0.92	-0.07	-0.14
SprMWD	-0.74	-0.22	-0.06	0.14	0.17
Wselvgr2	0.16	0.89	0.04	0.16	0.06
Dsl1	0.19	0.07	-0.00	0.77	0.04
Topex1km	0.51	-0.14	-0.03	0.65	-0.05
Wind1km2	0.11	0.83	0.04	-0.42	-0.05
Cosasp	-0.15	0.17	-0.13	0.39	-0.65
Sinasp	-0.19	0.14	-0.11	0.24	0.77

We can interpret these rotated factors as follows:

Factor No.	Label
1	Soil wetness/rainfall
2	High elevation/cold/windiness
3	Waterlogging/workability
4	Steep slopes/low windiness
5	Aspect

Communality coefficients were calculated. These were relatively high for all input variables, none having values under 0.60 (mean communality = 0.81).

iv) Calculating factor scores
Factor scores were calculated as discussed previously.

4 YIELD MODELS FOR SITKA SPRUCE

In this section we present details for the various regression models estimated for prediction of Sitka spruce YC. Further details regarding the regression models estimated as well as accompanying correlation matrices and descriptions of the base data are given in Bateman (1996).

Three types of model were fitted. These varied according to whether the environmental characteristics of a site were described by: (i) raw data; (ii) factors for our PCA; (iii) a mixture of these two (ensuring that raw variables retained in the model were not significantly correlated with retained factors). Clearly these latter mixed models are invalid 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 (for our Sitka spruce sub-compartment) 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 was missing and so these sites were deleted to leave an initial complete 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 large database approach compared to 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. These indicated that a linear model performed marginally better than other standard forms 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*).

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	0.07692	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-sq = 40.9% R-sq(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			

Because of its discrete nature, soil type is considered as a series of dummy variables, two of which proved statistically significant. YC is significantly elevated by planting on relatively good brown earth or podzol soils (*soil23*, which is a simple combination of *soil2* and *soil3*) and significantly depressed by planting on lowland lithomorphs[24] (*soil1*). 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 conforms with the 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. This might be because large stands provide advantages in terms of the ease of adopting species specific management regimes, or because such stands tend to condition their environment to their own advantage (for example, by reducing competition from both flora and fauna). Conversely this latter factor may be one of the pressures mitigating against smaller stands. 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 fertilisers and/or improvements

[24] Despite occurring in lowland areas these are relatively poor soils for plant growth.

in the genetic stock[25]. However, a further explanation might be that increased atmospheric concentrations of carbon dioxide has resulted in elevated growth rates[26].

Two further silvicultural factors are 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 forbears. 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 identified. Plantations with higher amounts of unproductive land (*unprod*) not surprisingly perform relatively worse than otherwise similar others. 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[27] 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 higher variance 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. Figure 1 illustrates the resulting impact upon the fit of the model (R^2-adj) of this progressive truncation of survey age (Bateman, 1996, reports precise values).

Close inspection of Figure 1 confirms the expected (although small) increase in model fit as stands surveyed at a very early age are removed. Omitting all observations with a survey age of less than ten years seems a reasonable assumption which still leaves us with 5168

---

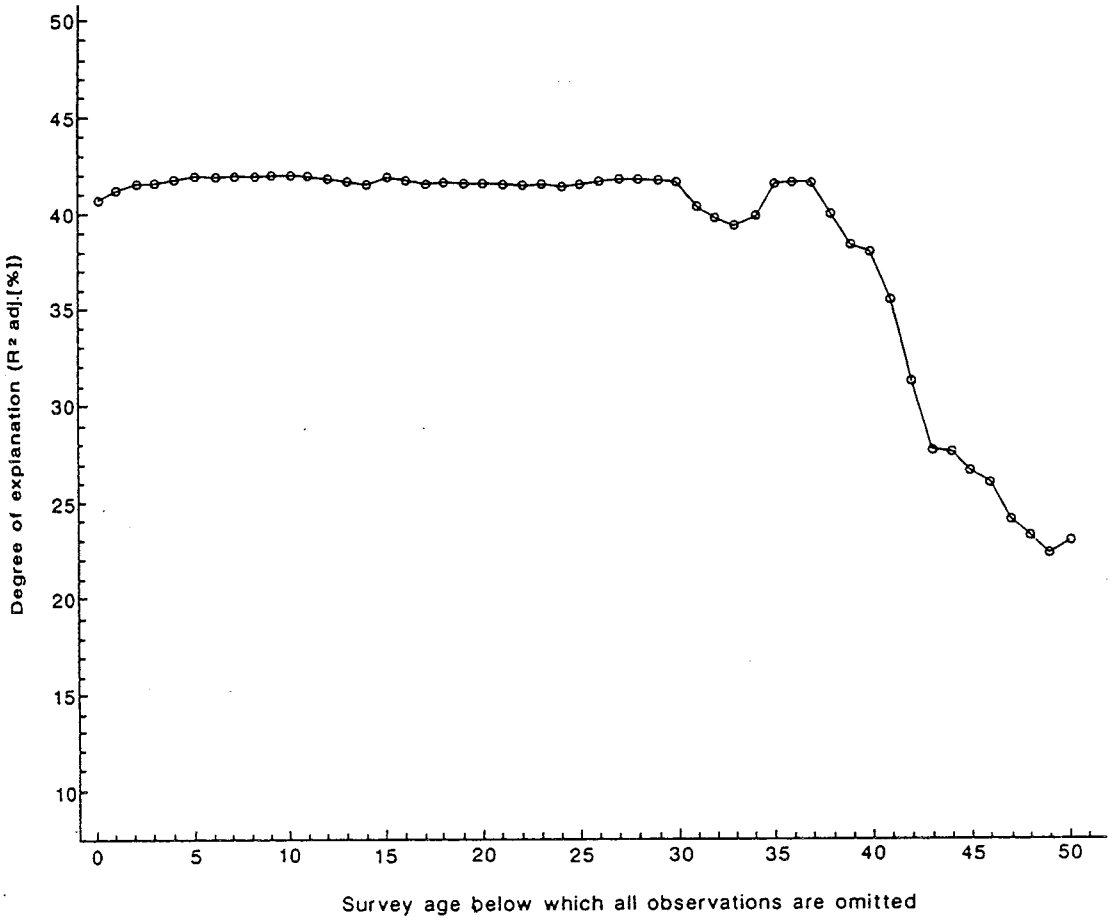
[25] A counter explanation, given by a senior FC official who shall be nameless, is that this effect may also arise out of errors in the YC tables.

[26] We are grateful to Piers Maclaren of the Forestry Research Institute (FRI), New Zealand, for this suggestion.

[27] These included Chris Quine and Adrian Whiteman of the FC and Douglas Macmillan of the MLURI.

observations. All three model variants were re-estimated from scratch[28] and the no factor model found to provide the most clearly interpretable results as reported in Model 2. We also use this model to provide an interesting aside regarding the effect of aspect upon tree growth. This is achieved by including the variable *Sinasp* and *Cosasp* in the model.

Figure 1: The impact of omitting stands surveyed at different ages



Comparison of Model 2 with Model 1 shows that the omission 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 geomorphic 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[29]. However, if we adopt a conventional 5% confidence test then neither of these aspect variables appear significant. Nevertheless, it is clear that we do not have to relax such a test by too much before aspect does appear to be having a significant effect.

---

[28] 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 ten years are omitted.

[29] 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.

Model 2: YC model for Sitka spruce after omitting stands with survey age <10 years

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

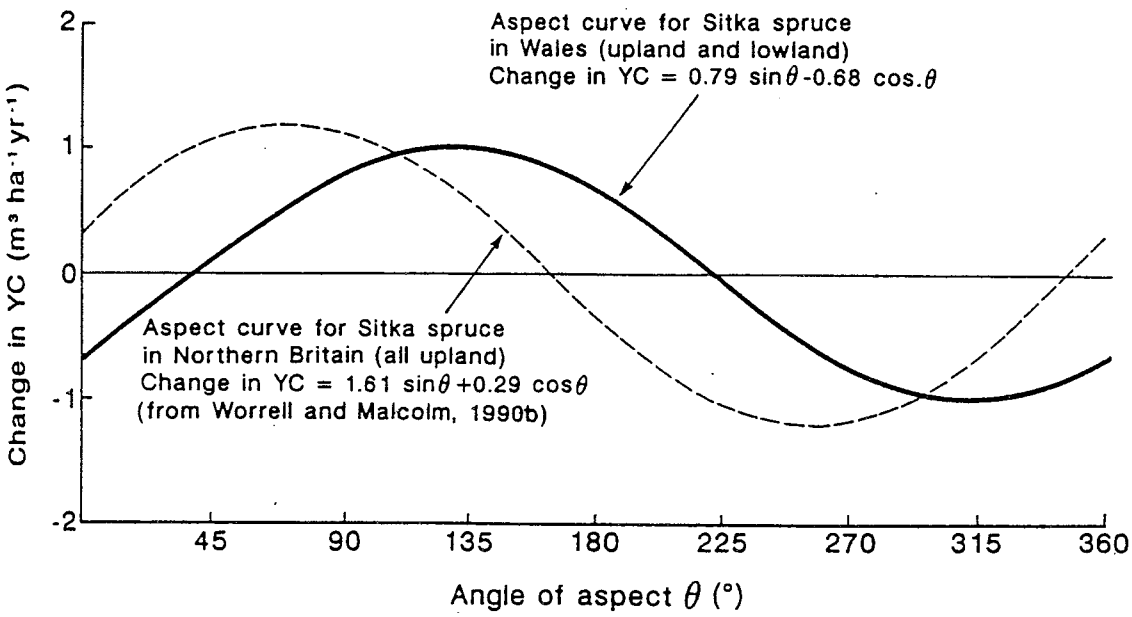
s = 2.306 R-sq = 42.1% R-sq(adj) = 41.9%

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			

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 2 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.

Figure 2: The effect of aspect upon YC



The comparison of our results with those of Worrell and Malcolm (1990b) proves interesting. The magnitude of aspect impacts is slightly higher in the latter study, a result which is not surprising given the relatively more adverse conditions of upland areas in northern Britain. However, the most striking feature 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. This complete negation of any effect which increased solar radiation from the south might seem to be due to the clearly 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 somewhat so that in Wales it is south east facing sites which appear to do best. It would seem that the relatively less adverse 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 consider Figure 1, while there does appear to be an increase of fit from omitting site surveyed at a young age that sub-compartments surveyed in their prime are relatively well predicted, there is a comparatively dramatic fall in fit which occurs when we confine 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)[30]. 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. Secondly, 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

[30] Indeed in Bateman (1996) the series of truncations is extended until this effect starts to increase R2 statistics.

of YC estimates will increase. Comparison with our subsequent analysis of beech sub-compartments suggests that there may be some merit in this argument to which we shall return.

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 site 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. As before 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 as well as each other. 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, in his more recent work, Macmillan (Tyler, Macmillan and Dutch, 1995, 1996)[31].

Model 3: Optimal PCA 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 = 2.3726 R-sq = 40.4% R-sq(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			

[31] This study concerns species other than those under investigation and is consequently omitted from our literature review.

Given the very similar performance of our no-factor and mixed models, the former is preferred for ease of interpretation and is reported as 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.

Model 4: Best fit YC model for Sitka spruce: no PCA 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-sq = 43.0% R-sq(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			

For the purposes of extrapolation Bateman (1996) gives descriptive statistics for all the explanatory variables in all models. The appropriateness of using our best fit model for such 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 8 which shows that 76.5% of YC predictions are within one division of actual YC[32].

Table 8: Comparing actual with predicted YC for our 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 to 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

[32] This is a higher degree of accuracy than that achieved by the thematic mapper approach of Gemmell (1995) who reports that roughly 75% of predictions were within 25% of actual growth rate. Here we have over 75% of predictions within 20% of actual, with no predictions in excess of 40% of actual.

5 YIELD MODELS FOR BEECH

The analysis of YC for beech sub-compartments followed the same methodology adopted in our investigation of Sitka spruce sites. Consequently only brief discussions of methodology are presented here with detailed results again being presented in Bateman (1996).

Following the deletion of sites for which key data was missing (giving us a dataset of 766 observations), initial investigations again confirmed the suitability of a linear functional form for our model. However, now a no-factor model provided the best initial fit to the data as reported in Model 5.

Model 5: Initial regression model: beech

Predictor	Coef	Stdev	t-ratio	p
Constant	5.5089	0.5600	9.84	0.000
Rainfall	-0.0002490	0.0001686	-1.48	0.140
Wselvgr2	-0.0043064	0.0005302	-8.12	0.000
Avwatgra	0.003182	0.002302	1.38	0.167
Plantyr	0.008443	0.002452	3.44	0.001
Historic	0.5229	0.1067	4.90	0.000
Monument	-0.9295	0.6180	-1.50	0.133
NpAonbSa	0.4978	0.1444	3.45	0.001
OthESA	-0.4987	0.2998	-1.66	0.097
ForPark	-0.3877	0.1894	-2.05	0.041
National	1.0305	0.3223	3.20	0.001
FCconst	-0.6026	0.1468	-4.10	0.000
Soil2	0.2423	0.1323	1.83	0.067

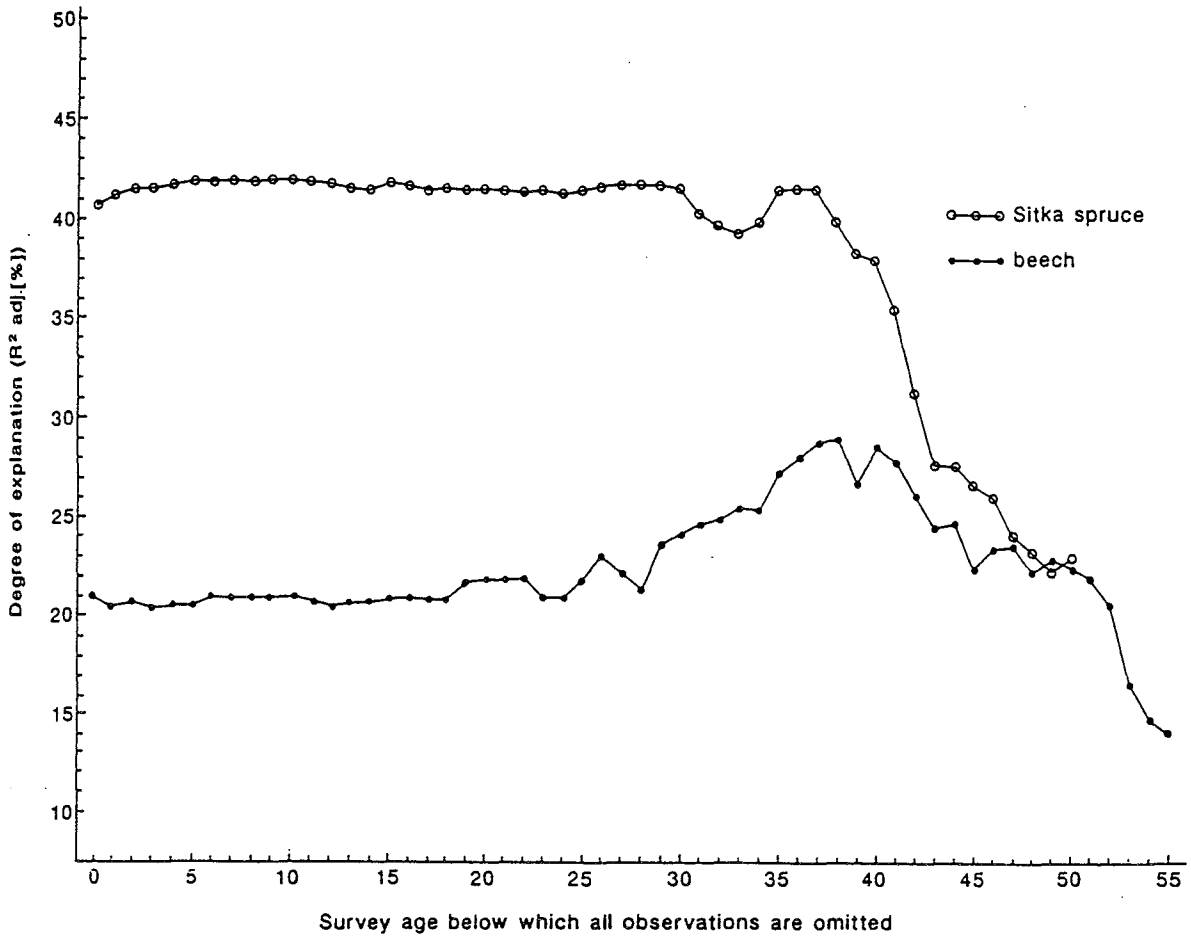
s = 1.363 R-sq = 22.2% R-sq(adj) = 21.0%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	12	399.763	33.314	17.94	0.000
Error	753	1398.070	1.857		
Total	765	1797.833			

The explanatory variables included in Model 5 are similar to those considered within our Sitka spruce models and so their interpretation is as before. While some of these variables are clearly rather weak, it was felt that this model provided an adequate base to analyse the impact of omitting sub-compartments on the basis of increasing survey age. This analysis was undertaken as before and results are illustrated in Figure 3 which for comparative purposes reproduces results from our analysis of Sitka spruce sub-compartments.

Figure 3: Impact upon model fit of omitting sites at successive survey age: beech and Sitka spruce



In assessing Figure 3 an immediate point is the relatively lower degree of fit exhibited by our models of beech growth. This is very likely to be a product of the relatively restricted range of the beech (as opposed to Sitka spruce) dependent variable discussed in Section 2.1. However, both curves initially rise (albeit slowly), peak and then eventually decline. Considering the curve for beech, the increase in fit from about *sage*=20 is probably due to the exclusion of stands surveyed at an early age. Note that this upward trend is much longer lasting than for our Sitka spruce analysis indicating, as expected, that it is much more difficult to assess the YC of a beech stand at say *sage*=10 than a Sitka spruce stand. Here the optimal fit excluding only low *sage* observations is achieved by omitting all sites with *sage*<38 (this compares with an optimal lower truncation at *sage*<10 for Sitka spruce). This gave a dataset of 359 observations for which model 6 provided the best fit.

Figure 3 also shows (as observed in our Sitka spruce data) that the degree of explanation afforded by models falls as we consider stands with relatively high *sage*, here values in excess of about 50 years seem to raise variance substantially. As previously postulated this seems likely to be connected to such stands being consequently quite old at the time of surveying. Uneven introduction of advances in silviculture may in part account for the increase in variance here. Furthermore it may be that planting date is less certain in these stands. This is

more likely to be a problem with beech sub-compartments than with Sitka spruce as the latter were almost all originally planted by the FC, who generally keep good records (and may apply new silvicultural techniques in a more uniform manner), while older beech stands may have been planted by a variety of private agents for which complete and accurate planting records may not be available. Given the importance of accurate age measurements in calculating YC such uncertainty may well translate into higher variance within such stands.

Model 6: Optimal (no-factor) model for beech: sites with sage*<38 omitted*

Predictor	Coef	Stdev	t-ratio	p
Constant	4.7663	0.7357	6.48	0.000
Rainfall	-0.0001754	0.0002479	-0.71	0.480
Wselvgr2	-0.0043157	0.0007218	-5.98	0.000
Avwatgra	0.003301	0.003648	0.90	0.366
Plantyr	0.013391	0.003044	4.40	0.000
Historic	0.4699	0.1535	3.06	0.002
Monument	-0.0937	0.9340	-0.10	0.920
NpAonbSa	0.6353	0.2317	2.74	0.006
OthESAt	-0.0556	0.4753	-2.22	0.027
ForPark	-0.4153	0.2602	-1.60	0.111
National	0.4156	0.5096	0.82	0.415
FCcons	-0.3452	0.2238	-1.54	0.124
Soil2	0.2145	0.1863	1.15	0.250

s = 1.258 R-sq = 27.9% R-sq(adj) = 25.4%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	12	211.712	17.643	11.14	0.000
Error	346	547.798	1.583		
Total	358	759.510			

Given this we felt justified in additionally omitting those stands with high *sage*. A sensitivity analysis suggested that omission of *sage* >49 would optimise the fit of our model. This gave an effective dataset of some 205 observations. Given the extent of the omission of observations, regression analysis was begun again afresh so as to redefine an appropriate set of explanatory variables. Here many variables failed to enter the model. When using our PCA approach to describing the environmental characteristics of sites only *Factor 2* proved adequately significant to enter our model which is reported as Model 7.

Model 7: Best factor-only model for beech: sites with sage*<38 and* sage*>49 omitted*

Predictor	Coef	Stdev	t-ratio	p
Constant	-5.227	1.854	-2.82	0.005
Factor 2	-0.35371	0.08458	-4.18	0.000
Plantyr2	0.08038	0.01278	6.29	0.000
AONB/NSA	0.4614	0.2719	1.70	0.091
OthESA	-1.5826	0.4941	-3.20	0.002

s = 1.266 R-sq = 35.6% R-sq(adj) = 34.3%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	4	177.140	44.285	27.65	0.000
Error	200	320.303	1.602		
Total	204	497.444			

A no-factor alternative was also estimated and is reported as Model 8.

Model 8: Optimal (no-factor) model for beech: sites with sage*<38 and* sage*>49 omitted*

Predictor	Coef	Stdev	t-ratio	p
Constant	-4.428	1.923	-2.30	0.022
Wselvgr2	-0.0038638	0.0009149	-4.22	0.000
Plantyr	0.07995	0.01279	6.25	0.000
AONB/NSA	0.4751	0.2710	1.75	0.081
OthESA	-1.4812	0.4969	-2.98	0.003

s = 1.265 R-sq = 35.7% R-sq(adj) = 34.4%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	4	177.649	44.412	27.78	0.000
Error	200	319.794	1.599		
Total	204	497.444			

Models 7 and 8 are extremely similar both in terms of their degree of explanation and their choice of explanatory variables; *Factor 1* in Model 7 is essentially the effect of elevation which is the raw data environmental variable *Wselvgr2* used in Model 8. Consequently we cannot have a mixed model for beech. Given its ease of interpretation we prefer Model 8 as our optimal model for predicting YC in beech sub-compartments.

An interesting supplementary analysis concerns the consideration of aspect effects. In building up our best fit model these had been investigated and rejected as insignificant. Nevertheless it is interesting to see if the logical relationship between aspect effects for Sitka spruce in northern Britain and Wales noted previously had any implications for aspect effects upon

beech in Wales. The aspect variables Sinasp and Cosasp were therefore added into our best fit model which was then re-instated to produce Model 9.

Model 9: Including aspect effects within our preferred beech model

Predictor	Coef	Stdev	t-ratio	p
Constant	-4.375	1.921	-2.28	0.024
Wselvgr2	-0.0037821	0.0009141	-4.14	0.000
Sinasp	0.1203	0.1274	0.94	0.346
Cosasp	-0.1905	0.1242	-1.53	0.127
Plantyr	0.07952	0.01278	6.22	0.000
AONB/NSA	0.4856	0.2703	1.80	0.074
OthESA	-1.4455	0.5007	-2.89	0.004

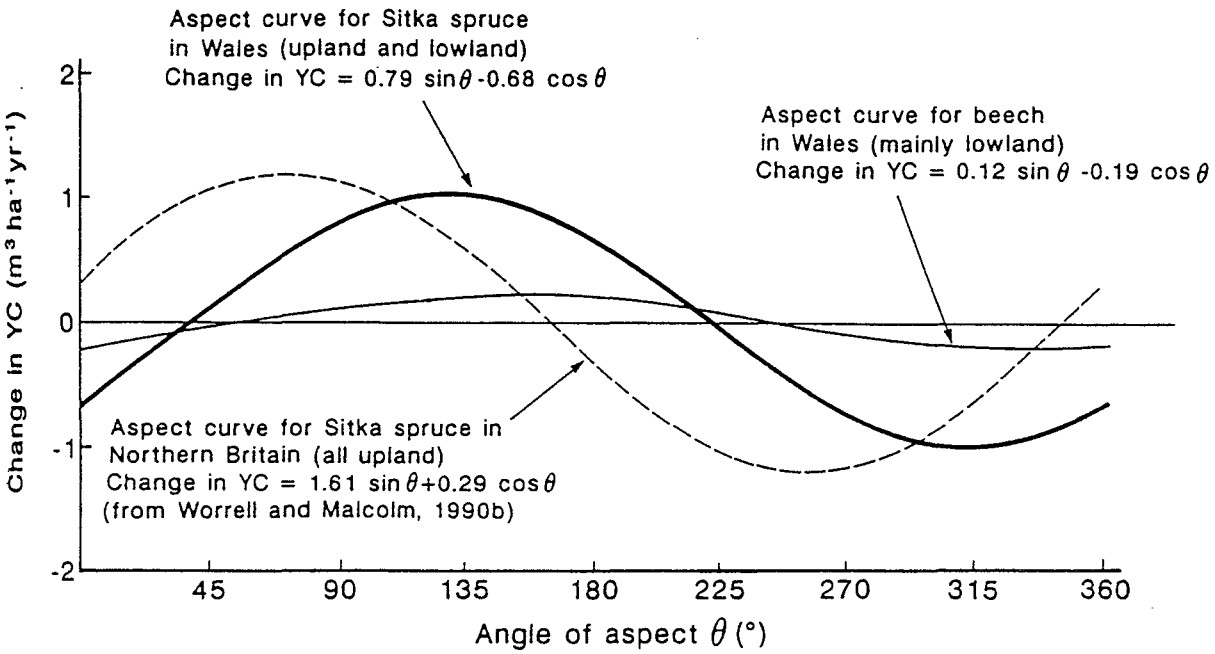
s = 1.261 R-sq = 36.7% R-sq(adj) = 34.8%

Analysis of Variance

Source	DF	SS	MS	F	p
Regression	6	182.734	30.456	19.16	0.000
Error	198	314.710	1.589		
Total	204	497.444			

As can be seen, both of the aspect variables are of very low significance. This of itself is interesting as aspect was clearly significant in the study conducted by Worrell and Malcolm (1990b) and on the edge of statistical significance in our Sitka spruce study. Similarly, consideration of coefficient estimates shows that the absolute magnitude of predicted effects was largest in the Worrell and Malcolm study, less sizeable in our Sitka spruce study and smallest here. Inspection of summary statistics given at the end of this section gives us a consistent explanation of all these results. While the Worrell and Malcolm study considered only sites in upland areas of northern Britain, or Sitka spruce analysis considers both upland and lowland sites in the less harsh climate of Wales. Furthermore comparison of descriptive statistics for our Sitka spruce and beech studies shows that beech is generally planted at significantly lower altitudes than those of Sitka spruce sites. So it seems that the impact of aspect upon tree growth depends upon altitude such that on lowland sites this may be insignificant while on upland sites aspect can have a major effect upon tree growth. Figure 4 superimposes the aspect curve implied by the results of Model 9 on to those previously described for Sitka spruce in the uplands of northern Britain (from Worrell and Malcolm, 1990b) and in the uplands and lowlands of Wales (from our models).

Figure 4: Aspect effects for Sitka spruce and beech in differing locations

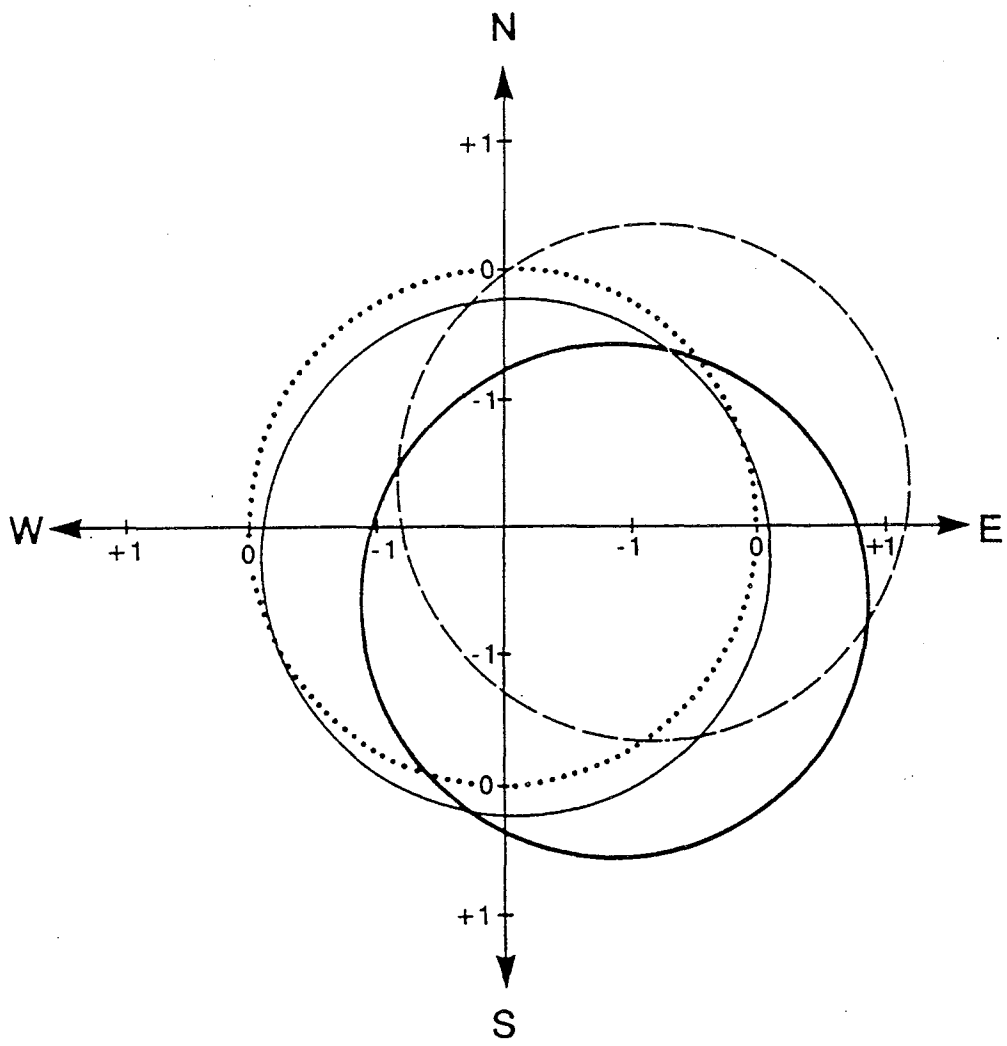


Inspection of Figure 4 tells a clear and coherent story. In the upland areas of northern Britain the intensity of the prevailing westerly wind causes aspect to be a major factor determining tree growth such that trees in relatively sheltered east facing sites perform significantly better than those facing west. The radiative energy advantage of south facing slopes is completely negated by the impact of the prevailing wind. In our Welsh study of Sitka spruce we consider both upland and lowland sites. Here both the magnitude and statistical significance of the impact of aspect is reduced. Furthermore, the reduction in the power of the prevailing wind (induced both because we are considering sites at lower altitude and the less arduous conditions of Wales relative to northern Britain) means that the solar energy advantage of southerly sites can now be detected as our aspect effect is now maximised at south east (rather than east) facing sites. This trend is continued when we consider our beech sub-compartments. Here altitude is again substantially reduced such that the absolute magnitude and statistical significance of the aspect effect is markedly reduced. Furthermore, the reduction in the impact of the prevailing westerly wind means that the solar energy advantage of south facing is further boosted such that we find that the aspect curve for beech sites now peaks for sites facing south-south-east.

Figure 5 shows an alternative approach to illustrating these aspect effects. Here the basis for comparison is given by the dotted circle which is centred directly upon the compass axes. This illustrates the situation in the absence of any aspect effect with points around the perimeter of this circle showing a zero impact of aspect upon YC. The results of Worrell and Malcolm (1990b) are represented by the dashed line circle which is centred a considerable way off towards the east showing the relatively positive aspect effect of east facing sites and the negative impact of westerly sites. The extent of this displacement shows the magnitude of this aspect effect which in this case raises tree growth by a maximum of just over 1 m^3 ha^{-1} yr^{-1}. The thick solid line circle represents our results for Sitka spruce in upland and lowland Wales. Here the displacement is a little less extreme, being most positive in the south east quadrant and most negative in the north west. Finally the thinner solid line circle shows results from our

analysis of beech growing in mainly lowland areas of Wales. Here the circle is only slightly displaced and shows the most positive aspect effect to be on sites facing south-south-east.

Figure 5: Comparison of aspect effects between Wales and upland northern Britain



Finally we can attempt to assess the validity of our best fit model (8) by comparing actual YC at all sub-compartments in our final dataset with YC as predicted by our model. Table 9 details results from this comparison.

Table 9: Comparing actual with predicted YC for our best fit YC model of beech (cell contents are counts)

	Predicted YC			
Actual YC	**4**	**6**	**8**	**ALL**
2	0	1	0	1
4	9	29	2	40
6	7	66	20	93
8	0	29	37	66
10	0	0	5	5
ALL	16	125	64	205

Predicted YC compared to actual YC	**Percentage of total sample (%)**
Prediction is two classes too high	1.5
Prediction is one class too high	23.9
Predicted YC equals actual YC	54.6
Prediction is one class too low	20.0
Prediction is two classes too low	0.0

Consideration of Table 9 shows that 98.5% of YC predictions are within one division of actual YC. This is a considerably higher rate of correct prediction than that achieved by our Sitka spruce model although given the restricted range for the dependent variable for beech this is perhaps not surprising and should therefore be treated with a little caution. Nevertheless, even accepting this warning, the apparent validity of the model is encouraging.

6 MAPPING YIELD CLASS

We have now estimated, for both of the tree species considered, two YC models, one including PCA factor explanatory variables and the other without. For our Sitka spruce dataset model 3 provides the best fitting PCA based model while 4 gives a slightly better fit without using PCA factors. Similarly for our beech dataset, model 7 gives the best PCA based predictions of yield while model 8 provides a slightly better fit without using PCA factors. These four models are used to provide estimates for the GIS images of YC presented and analysed below.

6.1 Producing GIS Images of Yield Class

In this context, an image is simply a spatially referenced depiction of a dataset produced by the GIS which can then be displayed upon a screen or printed as required. To produce a YC image the GIS requires data on all the predictor variables for all the grid points (the 'coverage') for which we want to predict, in this case the entire land area of Wales. If we take the best fitting Sitka spruce VAR 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 coverage in its own right which has identical values (here 16.709) for all grid points. The first explanatory variable in this model is the predictor *Rainfall* for which we have a full coverage from the LandIS database. We can therefore build up our GIS predicted YC map by telling the software to calculate a new image being the coverage *Rainfall* multiplied by its coefficient (-0.00167). Using the Idrisi GIS this operation is performed by use of the *Scalar* command. The resultant image can then be combined with that for the constant by use of the Overlay command, which as its name suggests, in effect overlays these two images to produce a third being YC as predicted by these first two elements in the model. Subsequent predictors are incorporated in a similar manner with separate images being created by multiplying the variables coverage values by its coefficient using the *Scalar* command and then incorporating the resultant image into the YC map using the *Overlay* command.

When using the PCA based models we need to first 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[33] and then using the component score coefficients calculated for Sitka spruce and beech to produce images of each factor. These were then treated as were the explanatory variables discussed above.

In all the models a number of the predictor variables are related to management (e.g. *Area*), policy (e.g. *reserve*) or when the site was planted (e.g. *plantyr*). These are not specifically spatial variables are so where treated by holding them at certain fixed values (i.e. as per the constant) and varying certain of these 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 for Sitka spruce sites and 10 ha for beech sites. Given the very low value of the coefficient on this variable and its

[33] The means and standard deviations necessary for this operation were taken from the variable values for all the forestry sub-compartments (both species). 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.

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 here.

6.2 GIS Timber Yield Images for Sitka Spruce

We produced images based on both our best non-PCA and PCA based yield models. Further to this 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[34]. For both of these analyses we initially hold 1st Rot = 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 test the effect of letting 1st Rot = 0 (i.e. second rotation) when plantyr = 75. This combination of differing models and assumptions resulted in 6 images being created. Table 10 details these images and provides a simple labelling system which we adopt subsequently.

Table 10: Sitka spruce GIS timber yield class images created: image labels

Model type	plantyr=0 1st Rot=1	plantyr=75 1st Rot=1	plantyr=75 1st Rot=0
No PCA factors used (model 4)	SS1VAR	SS2VAR	SS3VAR
PCA factors used (model 3)	SS1FAC	SS2FAC	SS3FAC

Images were produced using the procedure outlined in section 6.1. Figure 6 illustrates the predicted YC image created from model 4 (no PCA factors used) with plantyr = 75 (present day) and 1st Rot = 0 (replanting on a previously planted site) i.e. image SS3VAR.

Inspection of Figure 6 clearly shows the very strong influence which environmental characteristics have upon our predictions of YC. The influences of lower altitude, better soil and less-excessive rainfall combine to produce high YC. The pattern of lower YC produced by 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 and the Preseli Mountains produce YC values which lie between these extremes. Also clearly noticeable 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[35].

---

[34] See our previous discussion of possible interpretations of this effect.

[35] 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.

Figure 6: Image SS3VAR: predicted yield class from our optimal (no factor) model of Sitka spruce growth (assuming plantyr = 75; 1st Rot = 0).

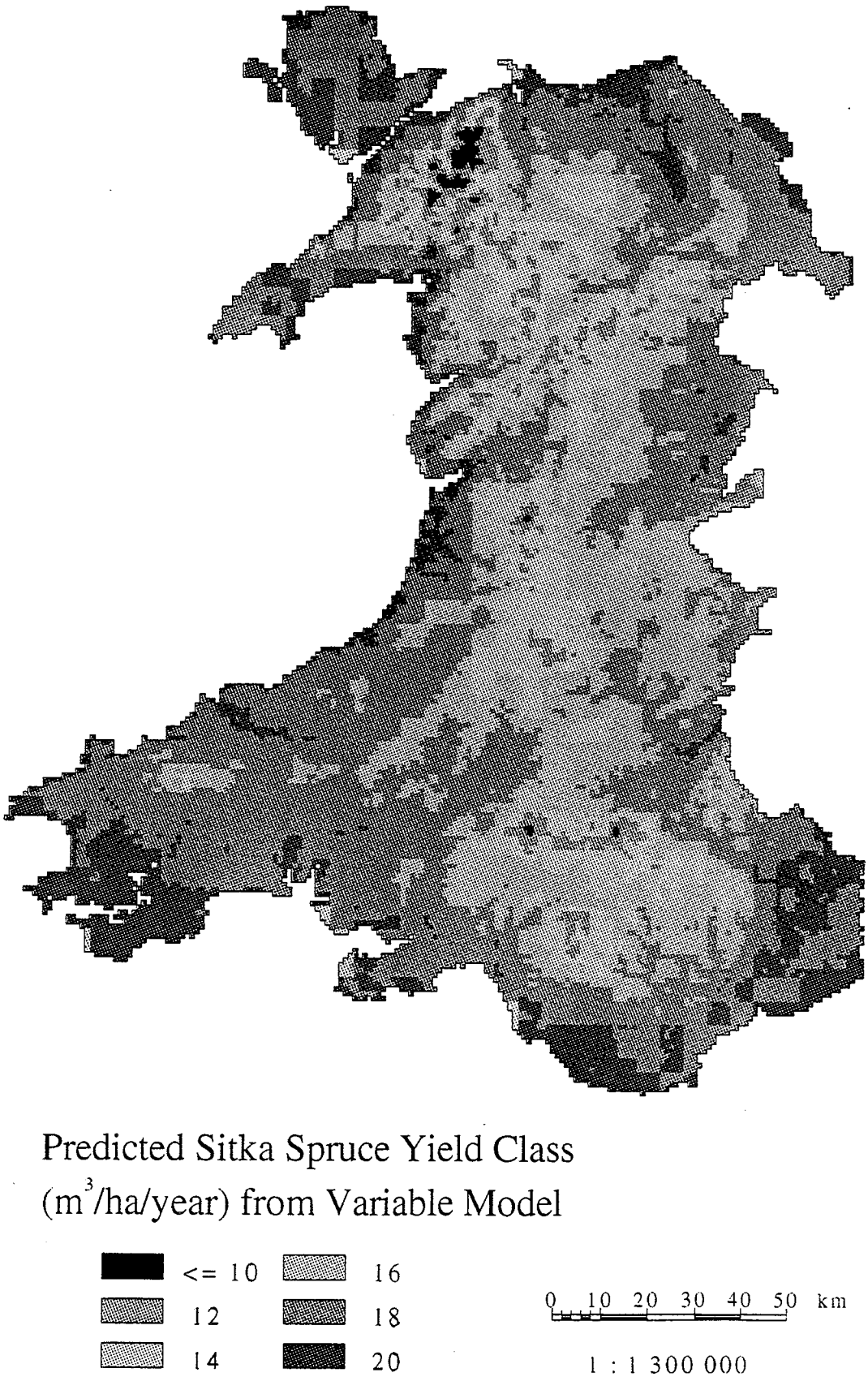


Figure 7 reproduces image SS3FAC, which uses the same assumptions regarding Plantyr and 1st Rot as Figure 6, but employs our best fitting factor based model (3) of YC. While the general pattern of YC predictions is similar between our factor-based (Figure 7) and no-factor models (Figure 6), some interesting differences can be detected. Figure 7 illustrates a smaller range of YC values than does Figure 6 (compare estimates for Pembroke, the Lleyn Peninsula, Anglesey and the North Wales coast where Figure 6 records many more high values than Figure 7; also compare upland areas such as Snowdonia and the Brecon Beacon where Figure 6 records lower values). Another noticeable difference is that Figure 7 is considerably more "blocky" than is Figure 6. This arises because of the formers reliance upon PCA factors dominated by $5km^2$ resolution variables such as those linked to water availability, while the latter is driven by variables such as elevation which is recorded on a $1km^2$ grid.

These difference excepted, images SS3VAR (Figure 6) and SS3FAC (Figure 7) give reasonably similar YC predictions. However, predicted YC systematically falls when we alter our assumptions regarding *plantyr* and *1st Rot*. Table 16 details predicted YC for all our Sitka spruce images showing the extent of this decline.

While our YC images seem highly plausible (and we would defend them as such for the majority of Wales), Table 11 and Figures 6 and 7 do indicate a weakness in our models with regard to their ability to predict YC for extreme environmental conditions such as, for example, mountain tops. Our best fitting model (SS3VAR) fails to predict any sites of less than YC6. However, clearly if trees were planted at the very tops of mountains they might well fail to survive or would at best produce only very low YC. Similarly our model does not predict any cells to have YC in excess of 20, yet our dataset indicated a few cases of YC being as high as 24. We therefore appear to be overpredicting YC at the lower extreme and underpredicting at the upper tail.

Figure 7: Image SS3FAC: predicted yield class from our best fitting factor based model of Sitka spruce growth (assuming plantyr = 75; 1st Rot = 0).

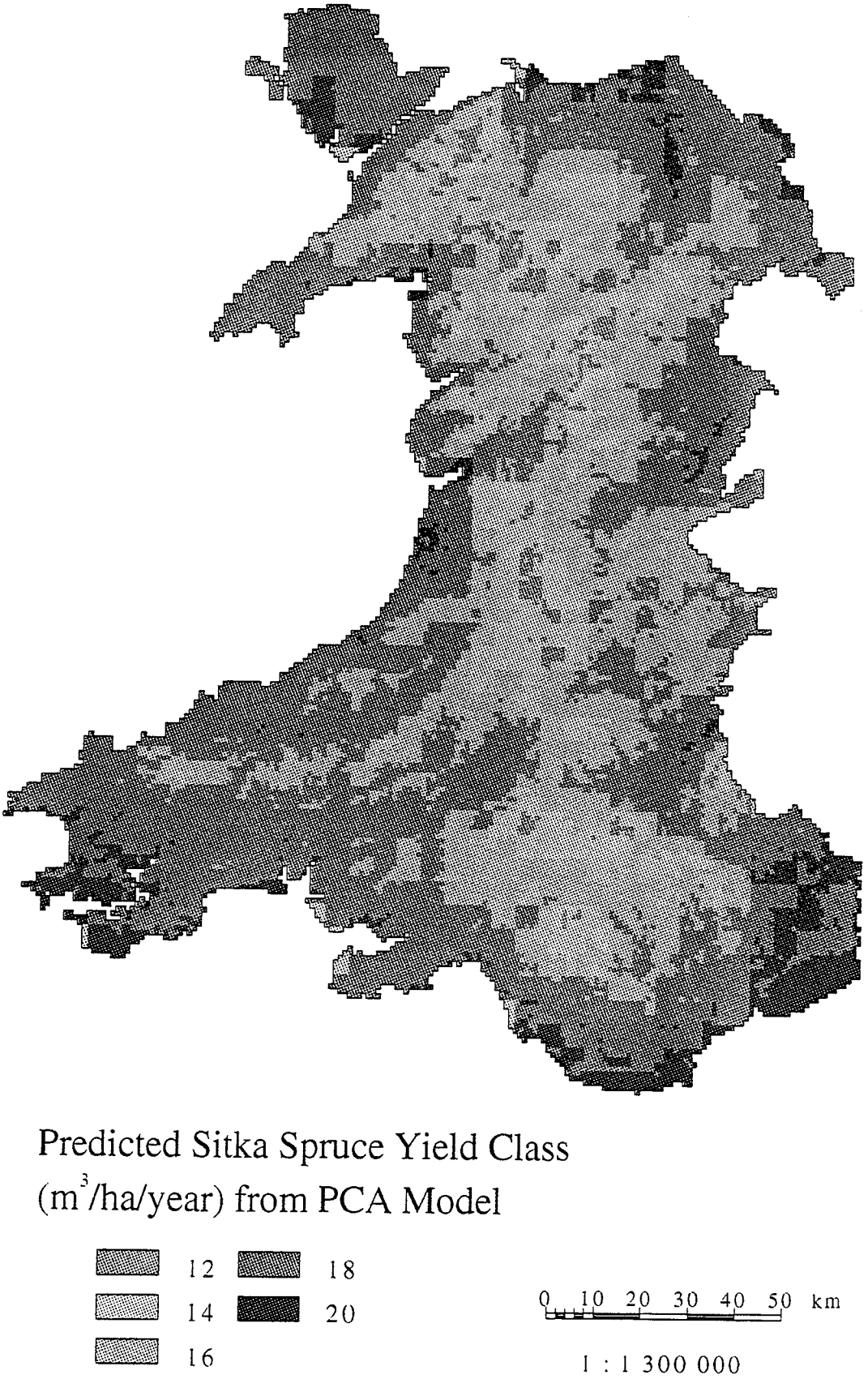


Table 11: Predicted timber yield class from various Sitka spruce maps[1]

YC	SS1VAR		SS2VAR		SS3VAR		SS1FAC		SS2FAC		SS3FAC	
	Freq.[2]	%	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 s.d.	11.38 2.81		15.12 2.81		17.05 2.81		11.21 2.65		14.90 2.65		16.98 2.65	

Notes:

1. For key to images see Table 10
2. Each map consists of 20563 1km^2 land cells

Three factors seem pertinent in explaining this. Firstly, we are predicting average YC over a 1 km^2 grid square. This will tend to remove any extremes and therefore gives some support to our findings. Secondly, and less positively, in creating our DEM we were unable to fully capture the upper extremes of altitude (see detailed discussion in Bateman, 1996). This means that we are under-representing elevation at the tops of mountains and therefore over-estimating YC at these points. Thirdly, as there is relatively little planting at the extremes of altitude so resultant low YC observations are relatively under-represented in the FCs sub-compartment database resulting in a lesser ability of statistical models based on such data to estimate accurately for such locations. However, while these are problems, the actual versus predicted comparison reported in Table 8 suggests that the degree of over and underprediction at the tails is not overly serious.

6.3 GIS Timber Yield Images for Beech

As before, we produced images based on both our best non-PCA and PCA based yield models. Further to this we again considered the impact of changing the *plantyr* and *1st Rot* variables. In the case of the plantyr variable, unlike our Sitka spruce analysis there was no distinct year in which beech planting commenced. Thus although we have a date at which plantyr = 0 this corresponds purely to the oldest record in the dataset (some 162 years ago) rather than to some actual initial planting date. Accordingly it was decided to adopt a somewhat different strategy here and our sensitivity analysis examined two values: plantyr = 144 (which equalled both the mean and median planting date); and plantyr = 162 (the present day). The dataset showed comparatively few beech sub-compartments were not in their first rotation and so this analysis was not performed, 1st Rot being held at a value of 1 for all beech images. The combination of factor and non-factor models and differing plantyr values yielded four different beech YC images. Table 12 details these images and provides labels as before.

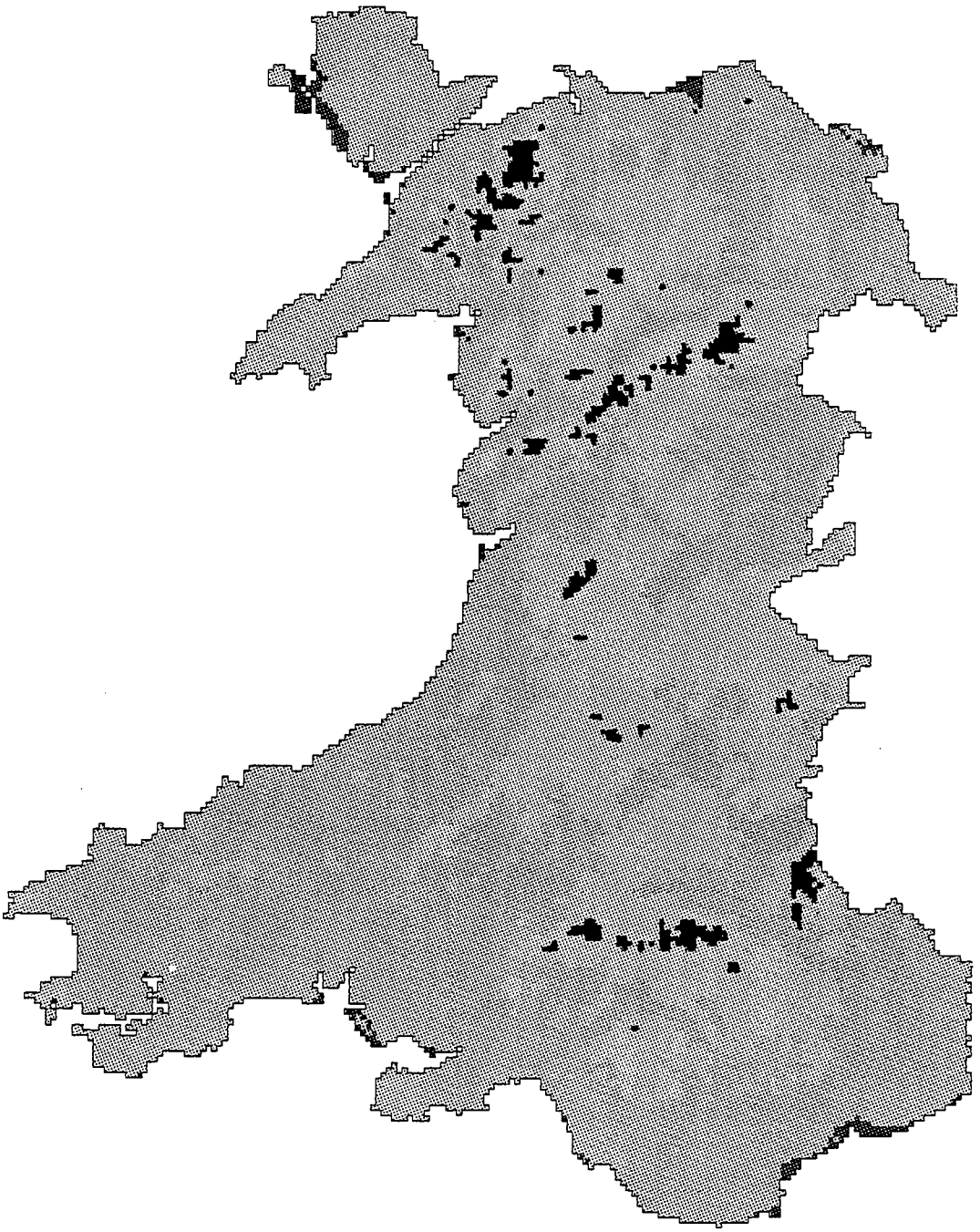
Table 12: Beech GIS timber yield class images created: image labels

Model type	plantyr=144 1st Rot=1	plantyr=162 1st Rot=1
No PCA factors used (model 8)	BE1VAR	BE2VAR
PCA factors used (model 7)	BE1FAC	BE2FAC

Images were produced using the procedure outlined in section 6.1. Figure 8 illustrates the predicted YC image created from our best fit beech model 8 (no PCA factors used) with plantyr = 162 (present day) and 1st Rot = 1 (first rotation) i.e. image BE2VAR.

As expected the general pattern of YC predictions observed for our Sitka spruce images is repeated in our beech images with high elevation and poor soils being associated with lower YC. However, both the absolute values of YC and its range are much lower than before. This is again as expected and reflects the restricted range of beech YC values recorded in the sub-compartment database. Our comments regarding these and other limitations to these predictions are as for our discussion of the Sitka spruce images.

Figure 8: Image BE2VAR: predicted yield class from our optimal (no factor) model of beech growth (assuming plantyr = 162; 1st Rot = 1).



Predicted Beech Yield Class
(m^3/ha/year) from Variable Model

<= 6 8
7 9

0 10 20 30 40 50 km

1 : 1 300 000

As for our Sitka spruce analysis, the general pattern of predicted YC for beech is reasonably consistent between images (with FAC images again being somewhat more blocky than their VAR equivalents) and so no further maps are reproduced here. However, Table 13 presents YC results from the four images detailed in Table 12.

Table 13: Predicted timber yield class from various beech maps[1].

Class	BE1VAR		BE2VAR		BE1FAC		BE2FAC	
	Freq[2]	%	Freq	%	Freq	%	Freq	%
3	1	0.005	-	-	-	-	-	-
4	84	0.409	-	-	14	0.068	-	-
5	1970	9.580	17	0.083	1725	8.389	-	-
6	10437	50.756	421	2.047	13251	64.440	208	1.012
7	8071	39.250	7003	34.056	5573	27.102	6775	32.948
8	-	-	12925	62.856	-	-	13580	66.041
9	-	-	197	0.958	-	-	-	-
Mean	6.25		7.69		6.19		7.63	
s.d.	0.80		0.78		0.76		0.70	

Notes:

1. For key to images see Table 12
2. Each map consists of 20563 1km^2 land cells.

With our predicted YC images for Sitka spruce and beech defined we can take the optimal of these (SS3VAR and BE2VAR respectively) and use them to produce images of timber value.

Bateman (1996) details tables of NPV and annuity equivalents for Sitka spruce and beech timber values across a full range of YC and at various discount rates. These results are used here to convert our optimal predicted YC images into maps detailing the monetary equivalent of those yields.

7.1 Maps of Timber Value: Sitka Spruce

We have two measures of timber value, NPV and its annuity equivalent. Each of these have been calculated at various discount rate and in the following analysis we shall concentrate on four of these: the exponential discount rates 1.5%, 3% and 6%; and a 6% hyperbolic discount rate. We therefore have 8 Sitka spruce timber value images which we wish to create. Table 14 details these and provides labels for subsequent referral.

Table 14: Sitka spruce GIS timber value images created: image labels

Value measure	Discount rate[1]			
	1.5%	3%	6%	6% hyperbolic
NPV	SS1tNPV	SS3tNPV	SS6tNPV	SS6HtNPV
Annuity	SS1tANN	SS3tANN	SS6tANN	SS6HtANN

Note:

1. All discount rates are exponential unless otherwise stated.

7.1.1 Estimating Equations to Convert from Yield Class to Values

A simple method to relate the YC images to their value equivalents was to use the tables given in Bateman (1996) as a source of data to estimate linear equations relating NPV and annuity values to YC for the various discount rates considered.

All timber values are considerably influenced by the numerous planting grants and subsidy schemes applicable (see Bateman, 1996, for review). Consideration of all permutations would make the following analysis impractically cumbersome and complex. Accordingly in the following we have taken the case which is most general for our study area, namely planting upon unimproved grassland without the benefit of Community Woodland Supplement. Deviations from the resulting financial measures can be calculated from the tables reported in Bateman (1996).

Within this general case we have two rates of grant payable depending upon whether grants are paid at the rate for disadvantaged/specially disadvantaged areas (DA/SDA) or otherwise. Table 15 details linear equations linking Sitka spruce NPV sums for DA/SDA areas to YC across various discount rates while Table 16 details results for an equivalent non-disadvantaged area.

Table 15: NPV of timber from an optimal rotation[1] of Sitka spruce: linear predictive equations with YC as the single explanatory variable (various discount rates). For disadvantaged and severely disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-3645.4 (-31.96)	996.621 (140.34)	100.0
3%	-3013.7 (-16.80)	570.20 (51.06)	99.7
6%	-1540.2 (-9.12)	209.02 (19.88)	97.8
6% hyperbolic	-2037.6 (-12.57)	558.78 (55.37)	99.7

Note:

1. See previous definitions regarding NPV and optimal rotations.

Table 16: NPV of timber from an optimal rotation of Sitka spruce: linear predictive equations with YC as the single explanatory variable (various discount rates). For non-disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-4204.9 (-36.88)	996.670 (140.39)	100.0
3%	-3540.9 (-19.74)	570.20 (51.07)	99.7
6%	-2008.0 (-11.89)	209.01 (19.87)	97.8
6% hyperbolic	-2518.6 (-15.53)	558.74 (55.34)	99.7

The discount rates used in tables 15 and 16 are chosen to cover a variety of analyses. The exponential 1.5%, 3% and 6% rates represent the diversity of real discount rates implicit in the decisions of farmers (the principle land users) in the study area. Furthermore the 6% rate is also that used by the UK Treasury for assessing projects which are considered to yield public benefits. A justification for the possible use of hyperbolic discount rates for the assessment of social preferences is given in Henderson and Bateman (1995).

A similar analysis was also conducted to link Sitka spruce annuity values to YC estimates. Table 17 details linear equations linking Sitka spruce annuity equivalents for DA/SDA areas to YC across the same discount rates as used before, while Table 18 details results for non-disadvantaged areas.

Table 17: Timber annuity equivalent of a perpetual series of optimal rotations of Sitka spruce: linear predictive equations with YC as the single explanatory variable (various discount rates). For disadvantaged and severely disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-104.183 (-34.55)	25.3951 (135.28)	100.0
3%	-119.003 (-15.90)	21.4204 (45.97)	99.6
6%	-104.24 (-8.37)	13.8902 (17.91)	97.3
6% hyperbolic	-172.51 (-14.98)	44.0728 (61.45)	99.8

Table 18: Timber annuity equivalent of a perpetual series of optimal rotation of Sitka spruce: linear predictive equations with YC as the single explanatory variable (various discount rates). For non-disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-116.398 (-38.55)	25.3135 (134.67)	100.0
3%	-136.324 (-17.88)	21.3151 (44.90)	99.6
6%	-132.22 (-10.67)	13.7573 (17.83)	97.2
6% hyperbolic	-207.35 (-17.74)	43.9472 (60.38)	99.8

7.1.2 Maps of Timber NPV: Sitka Spruce

Given that the majority of Wales qualifies for DA/SDA rates of subsidy we shall use these rates in the following images[36]. NPV maps for Sitka spruce timber value were produced by multiplying our optimal YC image (SS3VAR) by the relevant linear equation as detailed in Table 15. This was achieved using the *Scalar* command discussed previously. This operation was repeated for each of the four discount rates considered to produce the images defined in the upper row of Table 14. Table 19 details results from this analysis.

[36] An obvious extension, which we hope to address in future work, is to prepare a DA/SDA boundary image and use this to define a single map applicable to all areas of Wales. However, at the time of writing, permission to use such an image (which is Crown Copyright) had been requested but not granted.

Table 19: NPV sums for Sitka spruce timber GIS images at various discount rates (£/ha, 1990)

(£/ha)	SS1tNPV		SS3tNPV		SS6tNPV		SS6HtNPV	
	Freq[1]	%	Freq	%	Freq	%	Freq	%
-500:-1	-	-	-	-	1	0.005	-	-
0:499	-	-	-	-	31	0.151	-	-
500:999	-	-	1	-	187	0.909	-	-
1000:1499	-	-	2	0.005	2232	10.854	-	-
1500:1999	-	-	8	0.010	5786	28.138	1	0.005
2000:2499	-	-	20	0.039	11208	54.506	4	0.019
2500:2999	-	-	24	0.097	1118	5.437	13	0.063
3000:3499	1	0.005	48	0.117	-	-	16	0.078
3500:3999	-	-	163	0.233	-	-	30	0.146
4000:4499	4	0.019	514	0.793	-	-	81	0.394
4500:4999	5	0.024	1019	2.500	-	-	239	1.162
5000:5499	10	0.048	1307	4.956	-	-	711	3.458
5500:5999	11	0.053	1757	6.356	-	-	1139	5.539
6000:6499	8	0.039	2556	8.544	-	-	1480	7.197
6500:7000	17	0.083	3380	12.430	-	-	2073	10.081
7000:7499	23	0.112	4055	16.437	-	-	2927	14.234
7500:7999	62	0.302	4534	19.720	-	-	3919	19.059
8000:8499	80	0.389	1173	22.049	-	-	4447	21.626
8500:8999	207	1.007	2	5.704	-	-	3358	16.330
9000:9499	352	1.712	-	0.010	-	-	125	0.608
9500:9999	525	2.553	-	-	-	-	-	-
10000:10499	649	3.156	-	-	-	-	-	-
10500:10999	739	3.594	-	-	-	-	-	-
11000:11499	826	4.017	-	-	-	-	-	-
11500:11999	1112	5.408	-	-	-	-	-	-
12000:12499	1194	5.807	-	-	-	-	-	-
12500:12999	1595	7.757	-	-	-	-	-	-
13000:13499	1820	8.851	-	-	-	-	-	-
13500:13999	2162	10.514	-	-	-	-	-	-
14000:14499	2225	10.820	-	-	-	-	-	-
14500:15000	2605	12.668	-	-	-	-	-	-
15000:15499	2600	12.644	-	-	-	-	-	-
15500:15999	1561	7.591	-	-	-	-	-	-
16000:16499	168	0.817	-	-	-	-	-	-
16500:16999	2	0.010	-	-	-	-	-	-
Mean	13362.45		6707.30		2023.25		7488.72	
s.d.	1938.29		1189.19		438.32[2]		1167.57	

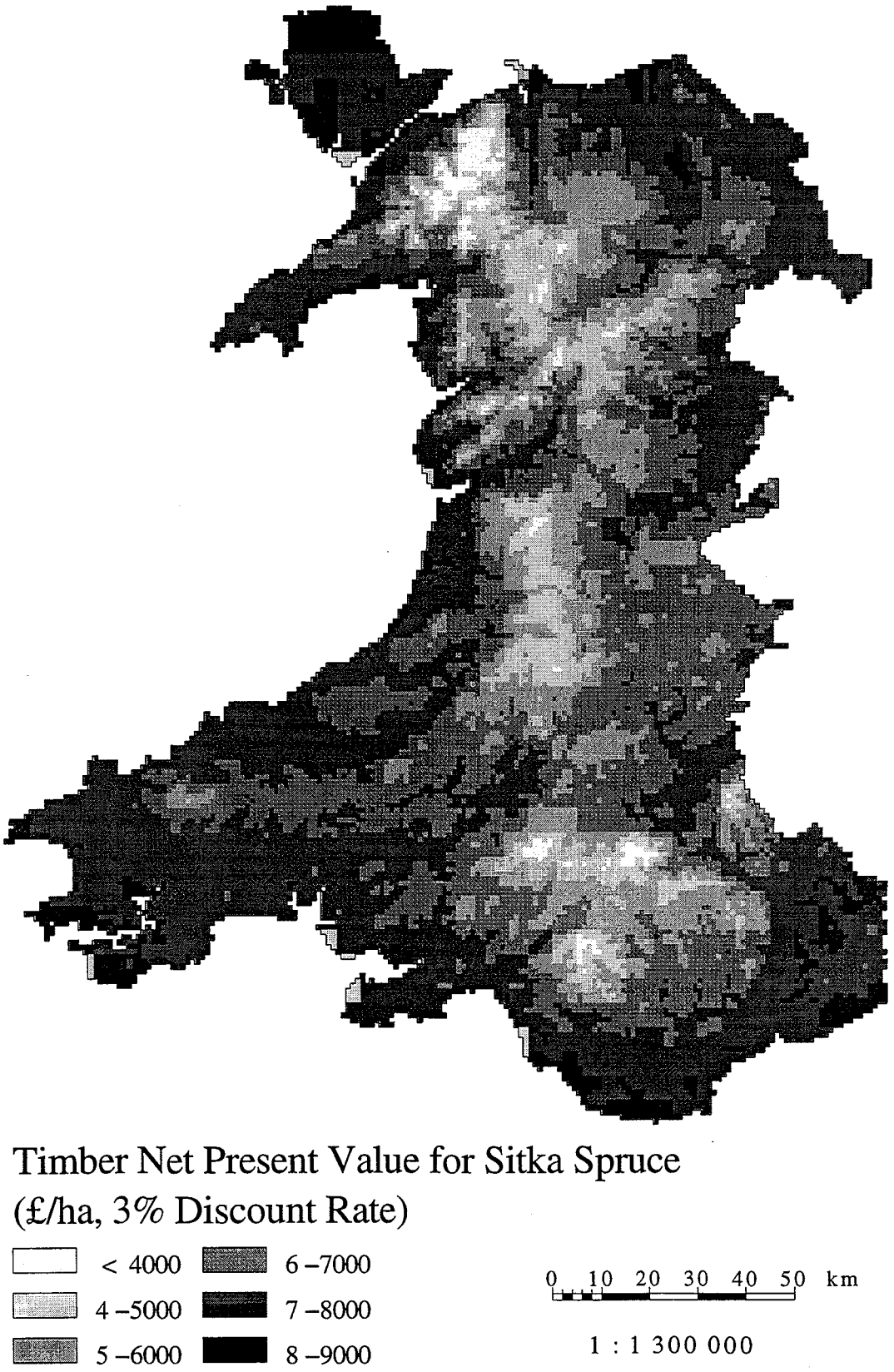
Notes:

1. From a total of 20563 1km^2 land cells.
2. Estimated (not calculated due to the GIS assigning zero values to non-land cells; this problem is adjusted for in the calculation of the mean).

Table 19 clearly shows both the range of NPV sums which are implied by our YC predictions and the impact of varying discount rate upon these. As exponential discount rates increase so the absolute value of NPV, its range and consequently variance, decline markedly. Switching to hyperbolic discounting increases these measures of NPV substantially as shown. Figure 9 illustrates the distribution of NPV sums estimated by the above analysis using the 3% discount rate (the middle of the three representing farmers decision horizon).

The distribution of NPV sums shown in Figure 9 strongly reflects that of the YC image upon which it is based (Figure 6). Consequently our comments are as before.

Figure 9: Image SS3tNPV: predicted timber NPV sums for Sitka spruce (based on yield class image SS3VAR; optimal no-factor model 4). Discount rate = 3% (£/ha, 1990)



7.1.3 Maps of Timber Annuity: Sitka Spruce

Annuity equivalents of the NPV sums detailed in Table 19 were prepared. This was again achieved via the Scalar command now relating our optimal Sitka spruce YC model (4) through the linear equations given in Table 17 (DA/SDA areas), to produce the four annuity images described in the lower row of Table 14. Results from this exercise are detailed in Table 20.

Table 20: Annuity values for Sitka spruce timber at various discount rates (£/ha, 1990)

Annuity value (£/ha)	SS1tANN		SS3tANN		SS6tANN		SS6HtANN	
	Freq[1]	%	Freq	%	Freq	%	Freq	%
-25:-1	-	-	-	-	1	0.005	-	-
0:24	-	-	-	-	21	0.102	-	-
25:49	-	-	3	0.015	53	0.258	-	-
50:74	1	0.005	16	0.079	479	2.329	-	-
75:99	2	0.010	22	0.107	2183	10.616	-	-
100:124	15	0.073	60	0.292	4068	19.783	-	-
125:149	18	0.088	263	1.279	7318	35.588	1	0.005
150:174	34	0.165	993	4.829	6434	31.289	2	0.010
175:199	115	0.559	1682	8.180	6	0.029	5	0.024
200:224	411	2.000	2413	11.735	-	-	10	0.048
225:249	1044	5.077	3962	19.268	-	-	13	0.063
250:274	1460	7.100	5175	25.167	-	-	8	0.039
275:299	1994	9.697	5626	27.360	-	-	22	0.107
300:324	3010	14.638	348	1.692	-	-	29	0.141
325:349	4172	20.289	-	-	-	-	78	0.379
350:374	4837	23.523	-	-	-	-	136	0.661
375:399	3380	16.437	-	-	-	-	312	1.517
400:424	70	0.340	-	-	-	-	546	2.655
425:449	-	-	-	-	-	-	730	3.550
450:474	-	-	-	-	-	-	812	3.949
475:499	-	-	-	-	-	-	966	4.698
500:524	-	-	-	-	-	-	1230	5.982
525:549	-	-	-	-	-	-	1551	7.543
550:574	-	-	-	-	-	-	1865	9.070
575:599	-	-	-	-	-	-	2326	11.312
600:624	-	-	-	-	-	-	2539	12.347
625:649	-	-	-	-	-	-	2897	14.088
650:674	-	-	-	-	-	-	2946	14.327
675:699	-	-	-	-	-	-	1447	7.037
700:724	-	-	-	-	-	-	92	0.447
Mean		328.84		246.18		132.57		578.86
s.d.		54.17		47.61		30.10[2]		86.44

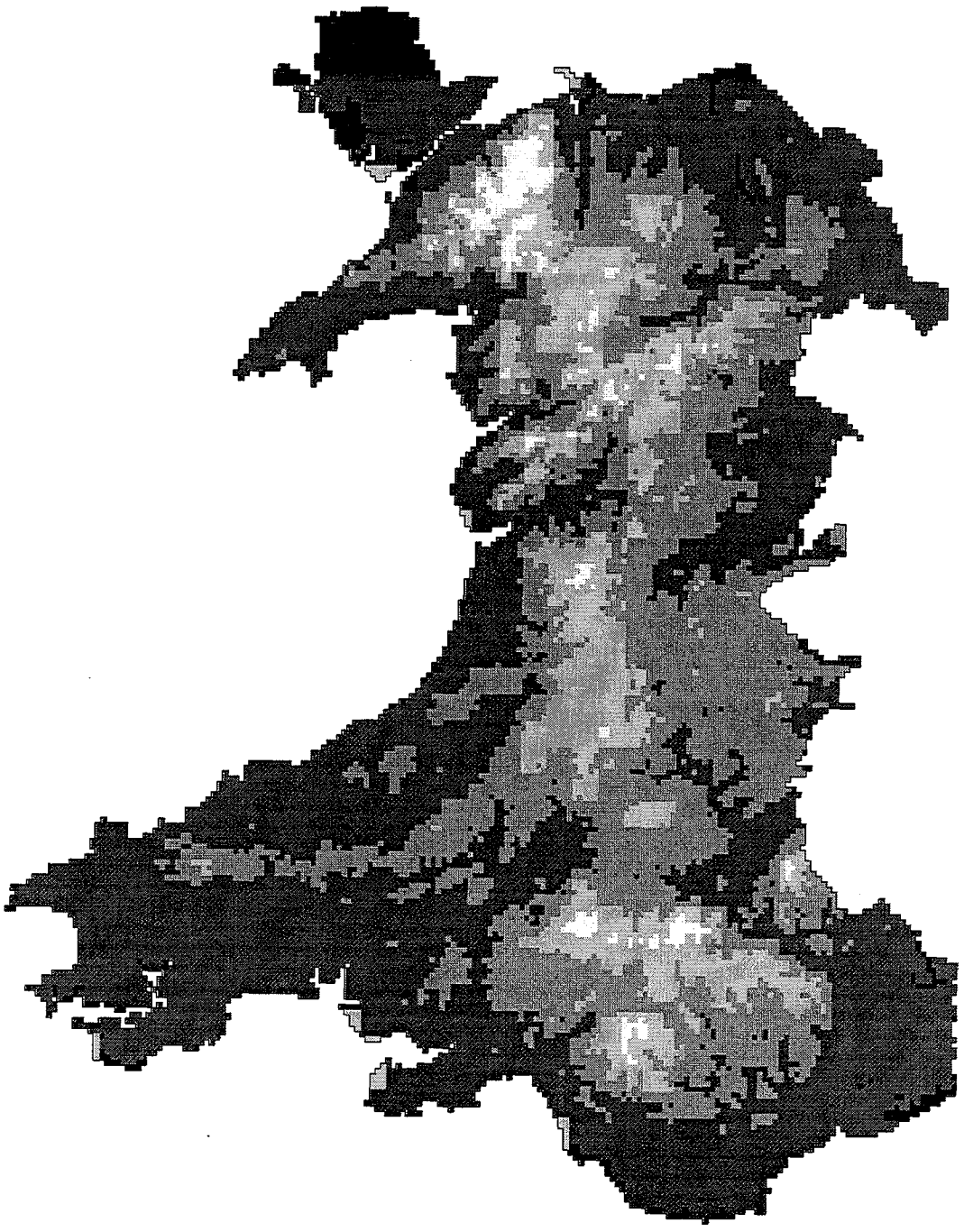
Notes:

1. From a total of 20563 1km^2 land cells.
2. Estimated (not calculated due to the GIS assigning zero values to non-land cells; this problem is adjusted for in the calculation of the mean).

As with our NPV analysis, Table 20 clearly shows that increasing the discount rate reduces the absolute value, range and variance of annuity sums. For comparative purposes Figure 10 reproduces image SS3tANN.

Figure 10 again reflects the broad distribution pattern observed in previous images and underscores the relationship between NPV sums and their annuity equivalents.

Figure 10: Image SS3tANN: predicted timber annuity equivalents for Sitka spruce (based on yield class image SS3VAR; optimal no-factor model 4). Discount rate = 3% (£/ha, 1990)



Timber Annuity Value for Sitka Spruce
(£/ha, 3% Discount Rate)

< 150
150 – 199
200 – 249
250 – 299
>= 300

0 10 20 30 40 50 km

1 : 1 300 000

7.2 Maps of Timber Value: Beech

As before we calculate NPV and annuity equivalents for our four discount rates. Table 21 details the 8 beech timber value images created from such an analysis.

Table 21: Beech GIS timber value images created: image labels.

	Discount rate[1]			
Value measure	**1.5%**	**3%**	**6%**	**6% hyperbolic**
NPV	BE1tNPV	BE3tNPV	BE6tNPV	BE6HtNPV
Annuity	BE1tANN	BE3tANN	BE6tANN	BE6HtANN

Note:

1. All discount rates are exponential unless otherwise stated.

7.2.1 Estimating Equations to Convert from Yield Class to Values

As before linear equations were estimated to related our Beech YC images to their value equivalents. Data was taken from Bateman (1996) assuming planting on unimproved grassland without Community Woodland supplement. Table 22 details equations linking beech NPV sums for DA/SDA areas to YC across various discount rates while Table 23 reports results for non-disadvantaged areas.

Table 22: NPV of timber from an optimal rotation of beech: linear predictive equations with YC as the single explanatory variable (various discount rates). For disadvantaged and severely disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-1513.9 (-3.09)	749.95 (11.26)	97.7
3%	-260.0 (-1.35)	349.50 (9.63)	96.8
6%	455.90 (5.89)	63.30 (6.01)	92.1
6% hyperbolic	-1024.8 (-2.65)	624.90 (11.89)	97.9

Table 23: NPV of timber from an optimal rotation of beech: linear predictive equations with YC as the single explanatory variable (various discount rates). For non-disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-2299.9 (-4.70)	749.95 (11.26)	97.7
3%	-1096.7 (-4.10)	349.35 (9.60)	96.8
6%	-160.20 (-2.07)	63.10 (5.98)	92.1
6% hyperbolic	-1679.4 (-4.36)	624.70 (11.92)	97.9

A similar analysis was also conducted to link beech annuity values to YC estimates. Table 24 details linear equations linking annuities to YC across various discount rates for DA/SDA areas, while Table 25 details results for non-disadvantaged areas.

Table 24: Timber annuity equivalent of a perpetual series of optimal rotation of beech: linear predictive equations with YC as the single explanatory variable (various discount rates). For disadvantaged and severely disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-29.832 (-3.20)	14.416 (11.36)	97.7
3%	-12.813 (-1.48)	11.327 (9.64)	96.8
6%	27.032 (5.63)	4.009 (6.13)	92.4
6% hyperbolic	-76.02 (-2.76)	44.553 (11.88)	97.9

Table 25: Timber annuity equivalent of a perpetual series of optimal rotation of beech: linear predictive equations with YC as the single explanatory variable (various discount rates). For non-disadvantaged areas.

Discount rate	Intercept (t-value)	Slope (t-value)	R^2 (adj)
1.5%	-44.445 (-4.73)	14.373 (11.23)	97.7
3%	-35.687 (-4.10)	11.246 (9.50)	96.7
6%	-10.143 (-2.10)	3.9165 (5.97)	92.0
6% hyperbolic	-121.30 (-4.36)	44.444 (11.74)	97.9

7.2.2 Maps of Timber NPV: Beech

As before we assume DA/SDA rates for the following analysis. NPV images were produced as per our Sitka spruce analysis. Table 26 details results for the four beech timber NPV images defined in the upper row of Table 21.

Table 26: NPV sums for beech timber GIS images at various discount rates (£/ha, 1990)

NPV (£/ha)	BE1tNPV		BE3tNPV		BE6tNPV		BE6HtNPV	
	Freq[1]	%	Freq	%	Freq	%	Freq	%
500: 999	-	-	-	-	20563	100.000	-	-
1000:1499	-	-	10	0.049	-	-	-	-
1500:1999	-	-	1281	6.229	-	-	-	-
2000:2499	10	0.049	14524	70.626	-	-	27	0.131
2500:2999	97	0.472	4748	23.088	-	-	332	1.615
3000:3999	5410	26.307	-	-	-	-	13440	65.355
4000:4999	15046	73.165	-	-	-	-	6764	32.891
mean	4250.78		2326.53		942.49		3778.66	
s.d.	494.83		331.31		317.49		426.95	

Notes:

1. From a total of 20563 1km^2 land cells.

Analysis of Table 26 shows a similar pattern of NPV to those observed for Sitka spruce. However, as a result of the longer delay in returns and their lower growth rate, the absolute level of timber NPVs for beech are considerably below those observed for Sitka spruce. To allow comparison with the SS3tNPV image printed above (Figure 9), Figure 11. reproduces image BE3tNPV.

Figure 11 shows the now familiar pattern of values corresponding closely to the environmental characteristics of sites. Comments are therefore as before.

7.2.3 Maps of Timber Annuity: Beech

Annuity equivalents were prepared as before. Results for all four of the images defined in the lower row of Table 26 are given in Table 32.

Table 27: Annuity equivalents for beech timber GIS images various discount rates (£/ha, 1990)

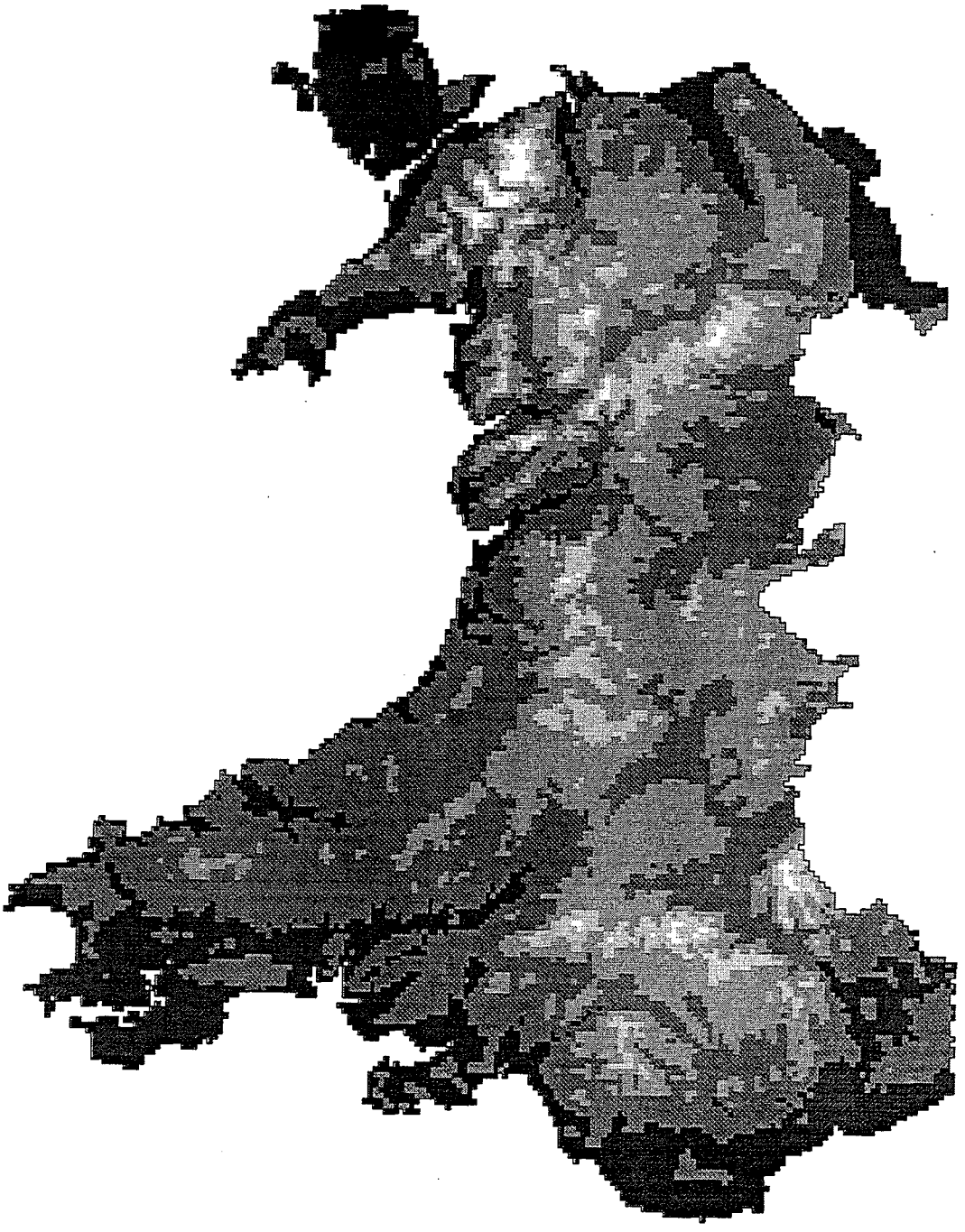
Annuity (£/ha)	BE1tANN		BE3tANN		BE6tANN		BE6HtANN	
	Freq[1]	%	Freq	%	Freq	%	Freq	%
40:49	20	0.097	20	0.097	37	0.180	-	-
50:59	179	0.870	327	1.590	16203	78.797	-	-
60:69	1798	8.744	4756	23.129	4323	21.023	-	-
70:79	6253	30.409	10841	52.721	-	-	-	-
80:89	8960	43.573	4619	22.463	-	-	-	-
90:99	3353	16.306	-	-	-	-	-	-
100:149	-	-	-	-	-	-	1	0.005
150:199	-	-	-	-	-	-	173	0.841
200:249	-	-	-	-	-	-	4962	24.131
250:310	-	-	-	-	-	-	15427	75.023
mean		80.98		74.25		57.85		266.45
s.d.		12.97		12.09		11.52		26.97

Notes:

1. From a total of 20563 1km^2 land cells.

For comparative purposes, Figure 12 reproduces image BE3tANN. The shows clearly the expected pattern of values. Other comments are as before.

Figure 11: Image BE3tNPV: predicted timber NPV sums for beech (based on yield class image BE2VAR; optimal no-factor model 8). Discount rate = 3% (£/ha, 1990)



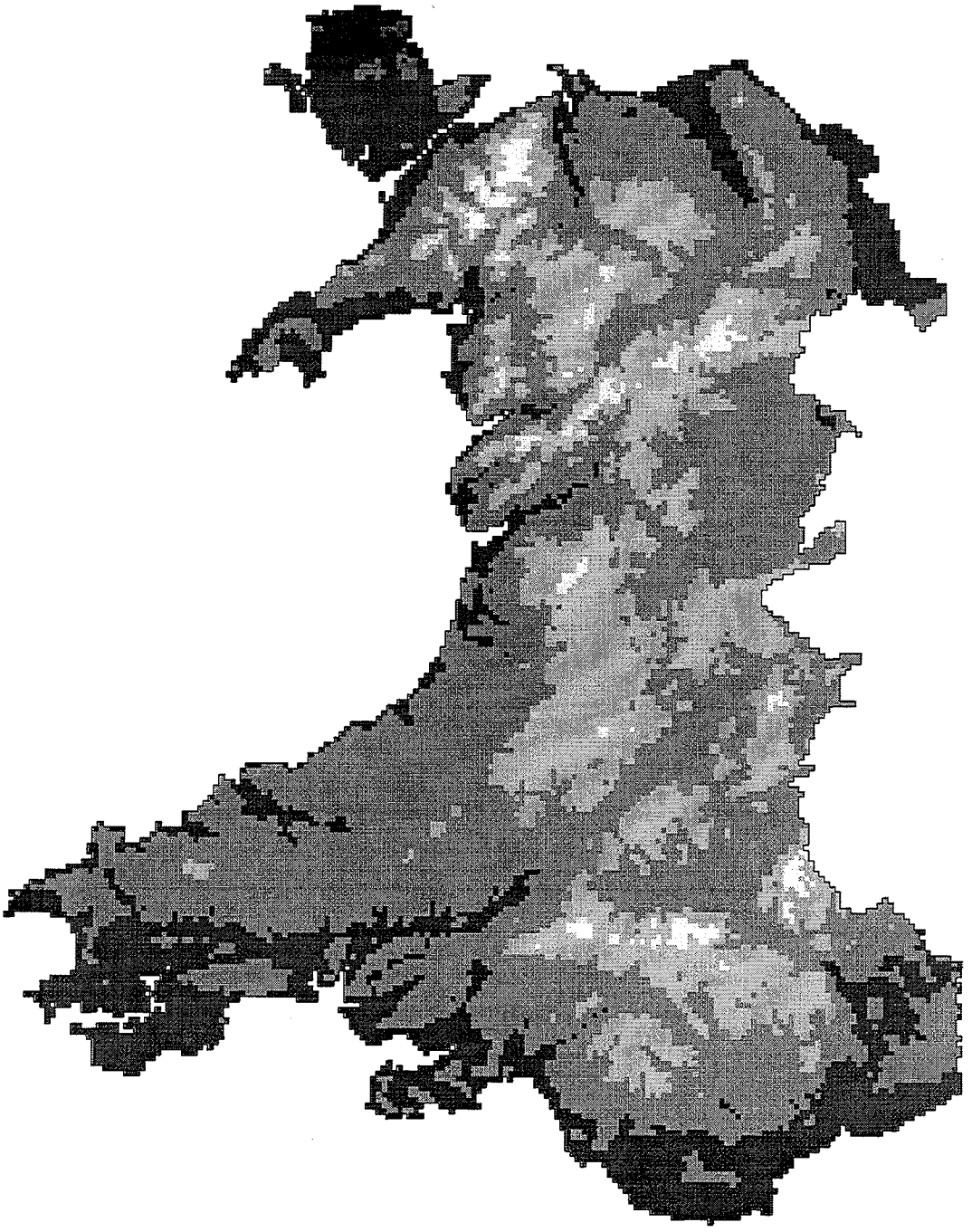
Timber Net Present Value for Beech
(£/ha, 3% Discount Rate)

< 1750
1750 – 1999
2000 – 2249
2250 – 2499
2500 – 2749

0 10 20 30 40 50 km

1 : 1 300 000

Figure 12: Image BE3tANN: predicted timber annuity values for beech (based on yield class image BE2VAR; optimal no-factor model 8). Discount rate = 3% (£/ha, 1990)



Timber Annuity Value for Beech
(£/ha, 3% Discount Rate)

< 59
60 – 69
70 – 79
80 – 89

0 10 20 30 40 50 km

1 : 1 300 000

8 CONCLUSIONS

We have estimated yield class models for Sitka spruce and beech based in part upon variables drawn from GIS datasets covering the entire extent of Wales. This has allowed us to use those models to produce predicted yield maps for both species for the entire Principality. We have then used these maps in conjunction with our previous work on timber values to produce NPV and annuity equivalent maps. In general we are reasonably happy with this analysis. However, we would mention at least one point of caution regarding the methodology developed in this study. The YC models fit the data quite well by the standards of models reported in the literature. Furthermore, the equations linking YC to NPV and annuity equivalents clearly also fit well. If this were not the case the possibility exists that errors in the first of these models might multiply with those at the second. This is a point to be wary of in any wider application of such a methodology.

Accepting the desirability of caution in all analyses and extrapolations, the methodology developed in this paper does appear to have advantages over previous studies in that the large scale databases employed permit the estimation of yield for a wide and diverse area. While we would not expect this analysis to outperform models developed for small scale specific areas when applied to those same small areas, the large area productivity and value maps produced are readily and more usefully incorporated within the cartographic decision making process currently being developed by UK forestry authorities and may, we hope, provide a significant input to that process.

REFERENCES

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

Avery, B.W. (1980): Soil classification for England and Wales (higher categories). *Technical Monograph No. 14*, Soil Survey of England and Wales, Rothamsted, Harpenden, Herts.

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

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

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.

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.

Blyth, J.F. and MacLeod, D.A. (1981): Sitka spruce (*Picea sitchensis*) in north-east Scotland: 1. Relationships between site factors and growth. *Forestry*, 54:41-62.

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

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.

Busby, R.J.N. (1974): Forest site yield guide to upland Britain. *Forestry Commission Forest Record 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: Lewis-Beck, M.S. (ed) *Factor Analysis and Related Techniques (International Handbooks of Quantitative Applications in the Social Sciences, volume 5).* Sage Publications and Toppan Publishing, London.

Edwards, M.V. (1957): Larix leptolepis, in *Exotic Forest Trees in Great Britain, Forestry Commission Bulletin No. 30*, HMSO, London, pp. 69-73.

Gale, M.F. and Anderson, A.B. (1984): High elevation planting in Galloway. *Scottish Forestry*, 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.

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

H④gglund, B. (1981): Evaluation of forest site productivity, *Forestry Abstracts*, 42: 515-527

Hart, C.E. (1991): *Practical Forestry*, 3rd ed. Alan Sutton Publishing, Stroud.

Henderson, N. and Bateman, I.J. (1995): Empirical and public choice evidence for hyperbolic social discount rates and the implications for intergenerational discounting, *Environmental and Resource Economics*, 5: 413-423.

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 Monography No.16*, Soil Survey of England and Wales, Harpenden.

Kaiser, H.F. (1958): The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23: 187-200.

Kaiser, H.F. (1974): Little Jiffy: Mark IV. *Educational and Psychological Measurement*, 34: 111-117.

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: Lewis-Beck, M.S. (ed.) *Factor Analysis and Related Techniques (International Handbooks of Quantitative Applications in the Social Sciences: Volume 5*, Sage Publications and Toppan Publishing, London.

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, *New Zealand Journal of Forestry Science,* 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. *Unpublished 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. *Scottish Forestry*, 26(4):296-308.

Mayhead, G.J. (1973): The effect of altitude above sea level on the Yield Class of Sitka spruce. *Scottish Forestry*, 27:231-237.

Moxey, A. (1996): Geographical information systems and agricultural economics, *Journal of Agricultural Economics*, 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.

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. (1995): Models to predict the general yield class of Douglas fir, Japanese larch and Scots pine on better quality land in Scotland, *mimeo*. Macauley Land Use Research Institute.

Tyler, A.L., Macmillan, D.C. and Dutch, J. (1996): 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.

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. (1987b): Predicting the productivity of Sitka spruce on upland sites in northern Britain. *Forestry Commission Bulletin 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.