4

Recreation: predicting visits

Introduction

In this chapter we utilise a geographical information system (GIS) to model the predicted number of visitors to a particular woodland site and test the efficiency of the resultant *arrivals function* in estimating visits to other sites. This is achieved through a zonal model which estimates visitor arrival rates from areas around a given site, and which is then applied to other sites through the definition of similar zones around them. Findings from our studies of the value of open-access woodland recreation (discussed in Chapter 3) are then applied to our predicted visits surface to obtain valuations of potential demand.[1]

Estimating an arrivals function

Previous studies

We are concerned with estimating overall visit rates which are applicable across populations, rather than being specific to individuals. By definition, conventional ITC valuation studies refer only to site visitors and say little about non-visitors. As a consequence they are unsuited to determining the absolute number of people who will visit a site. Therefore, our visitor arrivals model has to be composed of variables that have relevance across the population and can be readily transferred between sites.

To date there has been relatively little research regarding the level and determinants of visits to woodland in the UK. Furthermore, of those few studies which have examined this issue, most have looked at national recreational demand (Willis and Benson, 1989; Whiteman, 1991) rather than that at any particular forest site.

This chapter draws in part upon material presented in Bateman *et al*. (1999c) and we are grateful to the Regional Studies Association for permission to use this material.

[1] The GIS procedures employed here are presented in a non-technical descriptive manner. Further details of the commands used are presented in Bateman (1996).

One notable exception is provided by the work of Colenutt and Sidaway (1973) who model the demand for day-trip visits to the Forest of Dean. Here a combined on-site and household (postal) survey was used to collect information regarding trip origins and the factors determining visits. Analysis of these data revealed that by far the most important factor determining arrivals was travel time, to the effective exclusion of other explanatory variables.

The Colenutt and Sidaway result is important because it suggests that an arrivals function may be estimated relating travel time to the probability of a visit taking place. The analytical power provided by a GIS makes it possible to apply such a function to detailed population data, such as those provided in the UK Census, in order to predict arrivals at any existing or hypothetical site.[2] Obviously, in practice, the validity of taking an arrivals function estimated at one site and applying it to another needs to be carefully assessed in terms of the accuracy of the predictions made. Such a test is carried out and presented subsequently.

Recreation demand: the Thetford Forest study

The objective of this study was to estimate an arrivals function for a given forest which could then be applied across our Welsh study area. The base data for our initial investigation were obtained as part of the Thetford 2 study described in Chapter 3. Here individual journey distance and duration measures, adjusted for the availability and quality of the road network, were calculated for use within the ITC valuation study discussed previously. However, such individual-level variables were inappropriate for use within our arrivals function where travel times were required for all points across the study area rather than just those which were the outset origins of surveyed visitors. We therefore needed to convert our travel time road network data into complete coverage travel time zones which would have relevance to visitors and non-visitors alike. To obtain this continuity of coverage the vector (line) data derived for each individual segment of the road network had to be rasterised.

Rasterisation is a process of converting vector features (here roads) to cells on a regular grid, in this case of 500 m × 500 m squares,[3] covering the extended East Anglian area from which visitors originated. In this study the travel time values assigned to points along roads were reassigned to the grid cells which contained those points. A ‘majority filter’ was run recursively across the entire study area to smoothly fill in the gaps between roads, providing values for all grid cells and

[2] Kliskey (2000) also uses a GIS-based approach to generate models of recreation potential, reporting an empirical analysis of recreational snowmobiling in British Columbia.

[3] This produced a total of 161,195 cells for our entire study area, of which 58,364 were directly filled through the rasterisation process (i.e. they contained roads), the remainder being assigned values through the process described in the text.

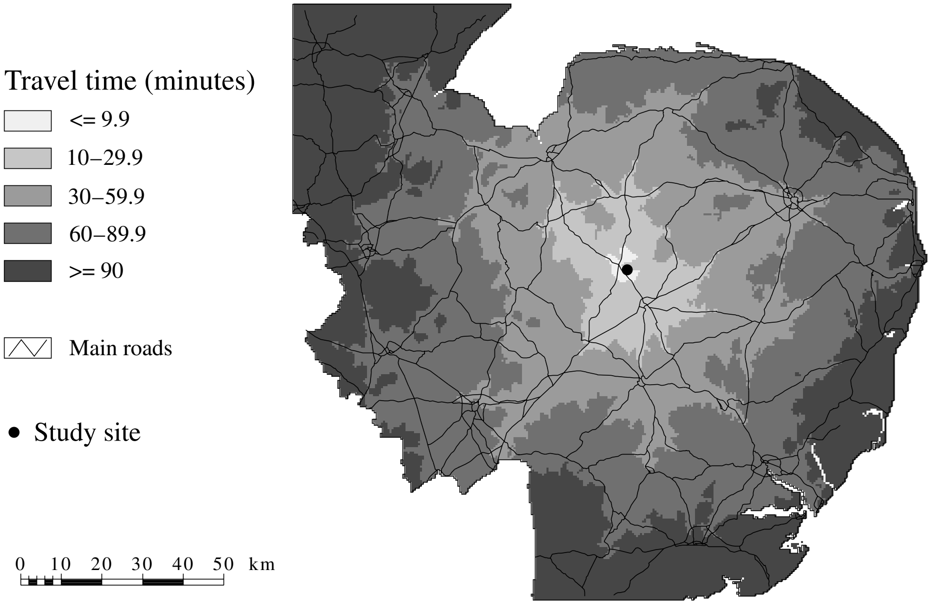

Figure 4.1. Travel time zones for the Thetford Forest study.

producing a continuous travel time surface centred upon the site and fanning out to fill the entire study area. The majority filter worked by means of a ‘moving window’ (usually eight by eight cells in extent),[4] where the centremost empty cell was assigned the value held by the majority of already assigned cells in the specified window. This approach worked well for the vast majority of cells. However, a few gaps remained in areas very remote from any roads where the filter window did not contain any cells filled directly by the rasterisation process. These grid cells were given the values of their nearest neighbours.[5] For further discussion of these procedures see Bateman *et al*. (1995b, 1999c) and Brainard *et al*. (1997).

Once all grid cells had been assigned a value they were grouped into convenient categories. Inspection of the calculated travel times showed that the extended road network encompassed values up to 120 minutes. Within this range, thirteen time zones were defined. Given the concentration of visit origins around the site, the innermost zones (between 0 and 30 minutes) were tightly defined at five-minute intervals, after which ten- and eventually fifteen-minute bands were used (between 30 and 60 minutes and 60 and 120 minutes, respectively). Figure 4.1 illustrates

[4] At the edge of the study area the window could feasibly reduce to as few as four cells (only filled cells are incorporated into the filter). The possibility of an edge distortion does exist but, given the very large number of cells used in the entire Thetford dataset, any such distortion would be extremely minor.

[5] Bateman *et al*. (1995b) undertook an analysis of vector and raster road speeds for this study. This shows that, within the Welsh study area considered subsequently, vector travel times were somewhat shorter than raster equivalents. Following this analysis, raster values were multiplied by a value of roughly 1.2 to ensure parity with the more accurate vector values.

Table 4.1. *Observed and predicted visitor rates*

Time zone[1] (1)	Actual visits[2] (2)	Zonal pop'n[3] (3)	Observed visit rate[4] (4)	Predicted visit rate[5] (5)	Predicted visits[6] (6)
5	13	954	0.0136268	0.0103972	9.9
10	31	21,596	0.0014355	0.0027285	58.9
15	8	13,326	0.0006003	0.0012476	16.6
20	10	14,377	0.0006956	0.0007160	10.3
25	26	26,811	0.0009698	0.0004655	12.5
30	38	58,416	0.0006505	0.0003274	19.1
40	46	191,009	0.0002408	0.0001879	35.9
50	65	405,831	0.0001602	0.0001222	49.6
60	17	375,134	0.0000453	0.0000859	32.2
75	48	776,817	0.0000618	0.0000559	43.4
90	15	562,508	0.0000267	0.0000393	22.1
105	7	253,762	0.0000276	0.0000292	7.4
120	—	—	—	0.0000225	—
150	—	—	—	0.0000147	—
180	—	—	—	0.0000103	—
210	—	—	—	0.0000077	—
240	—	—	—	0.0000059	—
300	—	—	—	0.0000038	—
360	—	—	—	0.0000027	—
500	—	—	—	0.0000014	—

Notes: [1] Upper limit of travel time zone measured in minutes of vehicle travel.
[2] Number of party visits recorded during survey (no repeat visits in sample).
[3] Number of households within each travel time zone as recorded in the 1991 Census.
[4] Column (2) divided by column (3).
[5] Visit rate predicted from the best-fitting arrival function (discussed subsequently).
[6] Predicted visit rate multiplied by zonal population (number of visiting parties).

resultant travel time zones, although for clarity of reproduction these have been amalgamated to five categories.

Once travel time zones were defined the relevant zone for each survey respondent was identified by matching the outset origin of each of the surveyed visitors to the travel time surface. Results from this exercise are presented in the first two columns of Table 4.1. Here column (1) shows the upper limit of each travel time zone (in minutes of vehicle travel to the site) and column (2) records the number of party visits to the site from each zone during the period of the survey[6] (other columns are discussed subsequently). Of the total sample of 351 parties, 324 (92.8 per cent)

[6] The possibility of repeat visits during the survey period was recognised. This was tested for and proved not to be a feature of the survey sample.

originated from time zones encompassed by our GIS road network. This provided a sufficient sample both to estimate an arrivals function and to extrapolate it beyond the limits of our road network.

The desired arrivals function would predict visits as a function of travel time. However, to achieve this it was necessary to account for varying population densities across our time zones (i.e. we needed to calculate a visit rate in terms of party visits per capita). Accordingly a population grid surface was interpolated which coincided in geographic extent with the travel time surface. Totals for persons usually resident in Enumeration Districts (EDs; the finest level of detail available) were extracted from 1991 Census[7] data using the SASPAC software (London Research Centre, 1992) and grid references for ED centroids were obtained from files held at Manchester Computing Centre.[8] Further discussion of the population surface concept is provided in Bracken and Martin (1995) and Martin (1996b).

Allocation of residential populations to the 500 m × 500 m grid cells composing the travel time zones was achieved through a volume-preserving algorithm, using a form of the SBUILD program described by Martin (1990). A mask image was used to prevent allocations outside the study area and initial input to the software consisted of 6,675 centroids with a population of 2,723,971. The surface produced by SBUILD (after cell totals were rounded to the nearest integer) contained a total population of 2,724,133 suggesting that the program produces accurate population estimates, at least at the aggregate level. Detailed inspection indicates that the characteristics of urban areas are well represented in the population surface and the only criticism which might be made is that some areas classed as 'unpopulated' undoubtedly contain isolated properties. This type of deficiency is, however, virtually inevitable given reliance upon data for areal aggregates such as Enumeration Districts and in the context of this research is not thought to represent a significant problem.

Population totals for our defined travel time zones were straightforward to calculate within the Grid module of Arc/Info. By allocating each of the surveyed parties to a travel time zone, summing to derive a total, and dividing by the resident population, a zonal visit rate was calculated. Results from this exercise are shown in Table 4.1. Here column (3) records the zonal population derived as discussed above. Column (4) divides visits from each zone, in column (2), by zonal population to give our observed visit rate. This represents the dependent variable in our arrivals function. The contents of columns (5) and (6) are described subsequently.

[7] Crown Copyright, ESRC/JISC purchase.

[8] A check on the accuracy of grid references was then conducted by calculating mean centres and standard distances for the EDs within each ward. This process revealed a few gross errors in grid references, which were corrected.

Table 4.1 indicates the expected, strongly negative, relationship between travel time and visit rate.[9] An examination of this relationship revealed that a double-log model provided a good fit to the data.[10] Equation (4.1) summarises the resulting arrivals function.

$$\ln \mathrm{VR} = -1.46 - 1.93 \ln \mathrm{TZ} \qquad (4.1)$$
$$(-2.41) \quad (-11.39)$$

where:

VR = observed visit rate (number of party visits from zone i divided by zonal population)
TZ = travel time zone (minutes)
R^2 (adj.) = 92.1%. Figures in brackets are t-statistics.

Investigations into potential omitted variables and correlation of residuals failed to reveal any significant problems with Equation (4.1). Given the strength of this relationship we felt confident in extrapolating our arrivals function to more distant travel time zones. Columns (4) and (5) of Table 4.1 list observed and predicted visitor rates, while columns (2) and (6) report actual and predicted visitor numbers. The arrivals function predicted 317.8 party visits from the first twelve travel time zones during the sampling period. This compares with an actual figure of 324, an error of less than 2 per cent.

Our arrivals function refers to those visitors interviewed during the sampling period. One of the main reasons for conducting our survey at Thetford rather than at a Welsh site was that it is one of the very few forests for which accurate daily and weekly visitor records are available (weekly data being held for several years). This information enabled us to allow for those visitors to Thetford who were not interviewed during our sampling period and also to establish that a very stable relationship exists between annual and survey period visits (Bateman, 1996, gives full details of this analysis).[11] This allowed us to extrapolate our sample-period arrivals function to an annual basis. Comparison of predicted with actual annual visits showed a discrepancy of just over 1 per cent.

[9] Note that observations from the furthest time zone (120 minutes) were omitted from our analysis (full details of which are given in Bateman, 1996) as this zone was not completely encompassed by the road network (see Figure 4.1).

[10] The small number of observations means that we should exercise some caution here. The double-log form narrowly outperformed a semi-log (dependent) model, while other forms fitted the data poorly. This is similar to the findings of Colenutt and Sidaway (1973) who report results for both forms although it is not made clear which is superior.

[11] This analysis reveals a consistent pattern of visits over the year, in which arrivals were well predicted by seasonal factors, extreme weather events and national holidays.

Applying the arrivals function: predicting arrivals in Wales

Before considering the detail of this analysis it is worthwhile to remind ourselves of its place within the overall study. The objective of the research presented in this section is to yield a map of predicted arrivals to actual or potential woodland sites for a regular grid across our Welsh study area which can then be monetised using the values derived in Chapter 3. This money value map will provide the first element in the analysis of woodland benefits. Maps of the timber and carbon storage value of woodland, derived in Chapters 6 and 7, can then be readily added to this to yield our estimate of the total benefits of woodland. These results can then be compared to the map of agricultural values derived in Chapter 8 to allow us to conduct a spatial CBA of the net benefits of conversion of land from agriculture to woodland in Chapter 9.

Our first concern in the present analysis was to test the validity of our arrivals function against the actual number of visits made at a sample of Welsh woodland sites. A study area boundary was defined and coincident road network and population surfaces constructed in a manner similar to the Thetford analysis. In order to allow for distant travellers to potential woodland sites along the Welsh border, the study area was defined so as to reach deep into England.[12] Appropriate county boundaries were obtained from the Bartholomew database. Road data were extracted, clipped and corrected as described in Chapter 3. B-roads and minor roads outside Wales were deleted, except where their omission created significant gaps in road topology. Roads that were just outside the defined study area were also included (notably the M6 motorway outside Coventry) if their absence seemed likely to have a significant impact on calculations of population accessibility. The resulting road network is illustrated in Figure 4.2.

Roads were again rasterised onto a 500 m × 500 m regular grid. The value assigned to each cell was the class of the road segment (as recorded in the Bartholomew database) with the greatest cumulative length running through the grid square. As a consequence, a long section of road that just clipped the edge of a cell took precedence over a short segment of road that actually had the greatest length within the grid square. This was a feature of the rasterising algorithm and could not be readily circumvented. Urban boundaries were rasterised and overlaid onto the road network to allow separation of urban from rural roads.

Population data and centroids for Enumeration Districts were again obtained from Manchester Computing Centre. The study area encompassed 30,311 Enumeration Districts with a total resident population of 13,821,562. Once centroid grid

[12] The study area comprised the following counties and areas: Avon, Cheshire, Clwyd, Dyfed, Gloucester, Greater Manchester, Gwent, Gwynedd, Hereford & Worcester, Merseyside, Mid Glamorgan, Powys, Shropshire, South Glamorgan, Staffordshire, West Glamorgan, West Midlands and Anglesey & Holyhead. Minor islands off the coast of Britain were removed.

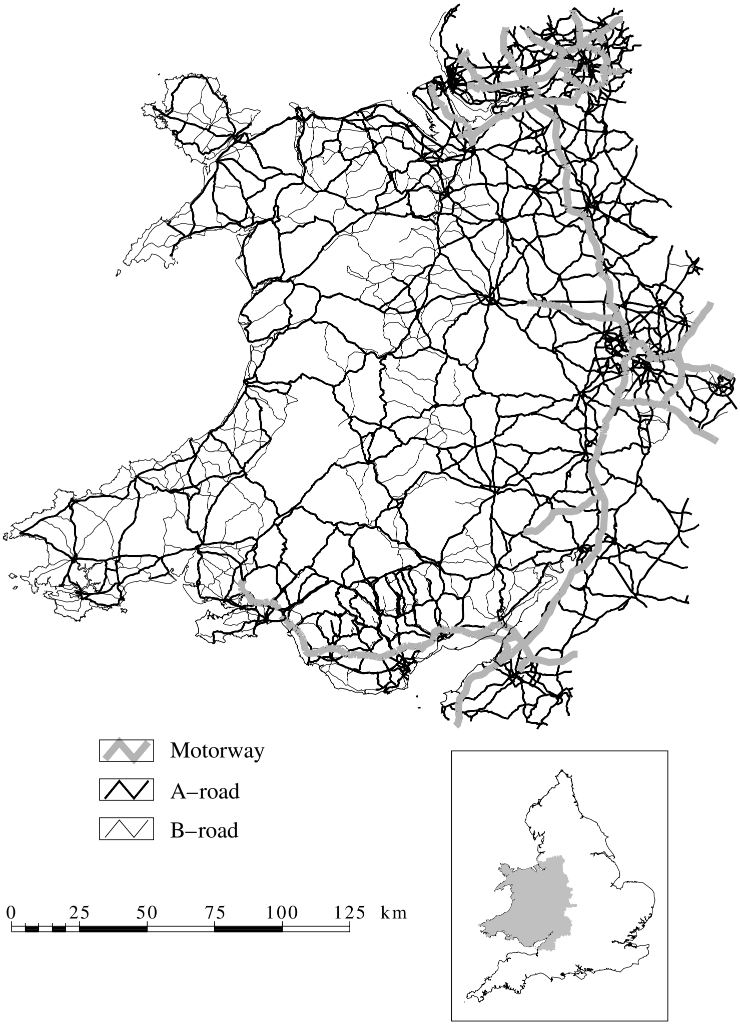

Figure 4.2. Digital road network for Wales and the English Midlands. For cartographic reasons English B-roads and all minor roads are omitted from the map.

references had been checked, the SBUILD program was again used to generate a population surface at 500 m × 500 m grid cell resolution. The program again performed well, yielding a total population estimate of 13,821,361 people. Figure 4.3 illustrates the resulting surface.

With the Welsh travel time zone algorithm and the relevant population surface defined, an actual versus predicted test of our arrivals function was possible. At the time of this analysis the Forestry Commission only held visitor data for five sites in Wales. Furthermore, in conversation with officials it became apparent that two of these were closed for unusually long periods during the year while a third contained several special attractions not normally found at forest sites which raised

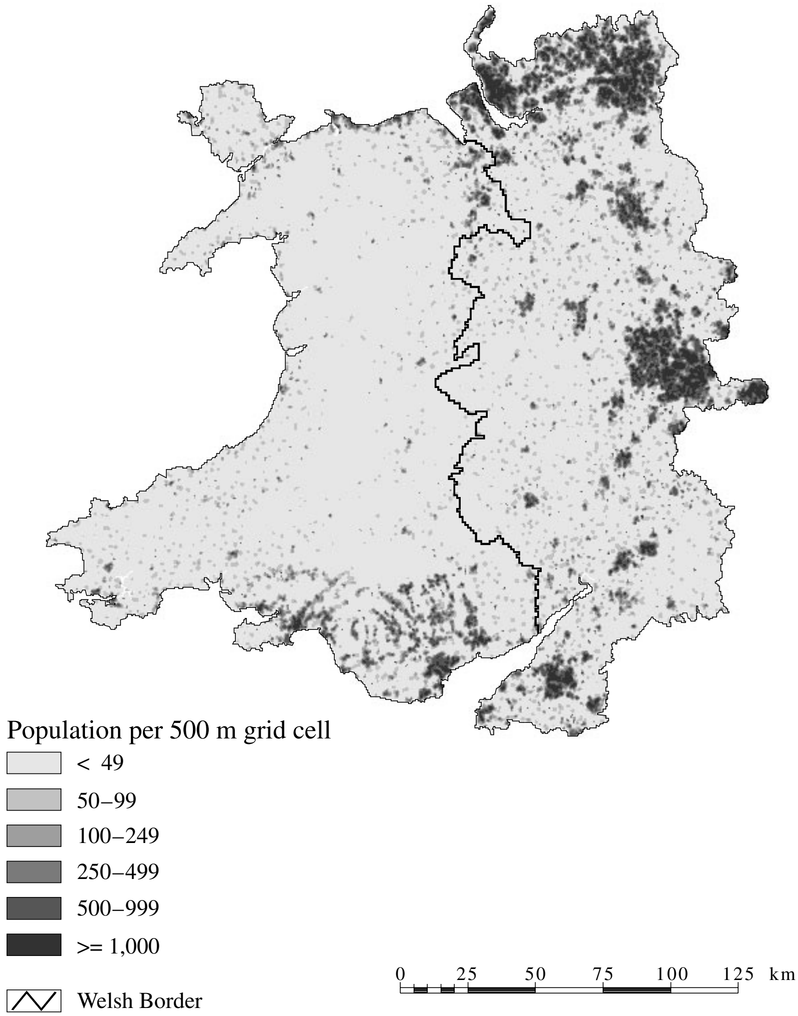

Figure 4.3. Population density surface for Wales and the English Midlands. (*Source of population data:* 1991 Census, Crown Copyright, ESRC/JISC purchase. The population density values were calculated using the SBUILD software with 1991 Census Enumeration District centroids.)

visitor numbers above those normally expected for such a location.[13] This sample size and associated complications meant that the desired standard of testing was not feasible (a problem which was not adequately addressed until additional sites subsequently became available and an extended test across more than thirty sites was carried out as described subsequently in this chapter). However, it was decided to undertake a simple comparison of predicted and actual arrivals at each of the five Welsh sites available. For each of these sites, arrivals were predicted by (i) using the

[13] These include a museum, catering facilities and a variety of organised recreational activities.

rastering algorithm to define zones and travel times; (ii) interrogating the SBUILD population surface to obtain an estimate of the population in each zone and; (iii) applying this information through the arrivals function given in Equation (4.1) in conjunction with the sample period/annual visitor conversion factor calculated during the Thetford survey. Equation (4.2) simply relates actual to predicted visits per annum.[14]

$$\underset{(4.420)}{\text{ACTUAL} = 0.903\ \text{PREDICTED}} \tag{4.2}$$

where:

ACTUAL = actual arrivals at site (party visits per annum)
PREDICTED = predicted arrivals at site (party visits per annum)
R^2 (adj.) = 83.0%. Figures in brackets are t-statistics.

Equation (4.2) indicates that, despite the limitations of this analysis, the arrivals function performs as expected with the slope coefficient for PREDICTED not being significantly different from 1. Given this result and the lack of data for further testing we concluded that the arrivals function did provide at least a defensible predictor of annual arrivals at a typical woodland site (i.e. one with similar basic facilities to that found at Thetford).

We were therefore able to make a case for applying the arrivals function to a regular grid of points across the study area and so predict expected annual recreational visits to actual and hypothetical woodland sites across Wales.[15] An important practical issue, however, is the appropriate grid size for such an analysis. Even with the use of a raster structure and other efforts to shorten processing, determination of travel time zones for a representative grid covering the whole of Wales represented a significant computational exercise. Using available computing facilities each site took between fifteen and thirty minutes to process (depending on workload). Assuming the former time, calculation of a 1 km grid surface for the entire area of Wales (some 20,500 cells) would take over 200 days of continuous processing; clearly a coarser sampling scheme was required.

The issue of grid size was investigated by defining two transects across Wales. The first of these ran due east from the coast near Aberystwyth to the English border and was composed of thirteen sites, each separated by 2.5 km, and another five sites at 5 km spacing. The second transect ran from a similar origin due south to a point just outside Swansea and was composed of sites all at 5 km intervals. Travel

[14] Analysis confirmed that any constant was not significantly different from zero.
[15] Such estimates do not take into account the substitution effects which would arise in any specific area if a number of woodlands were planted in that locality. The object of the current exercise is to identify those areas where the establishment of a wood would be beneficial. The impact of supply-side changes is considered subsequently.

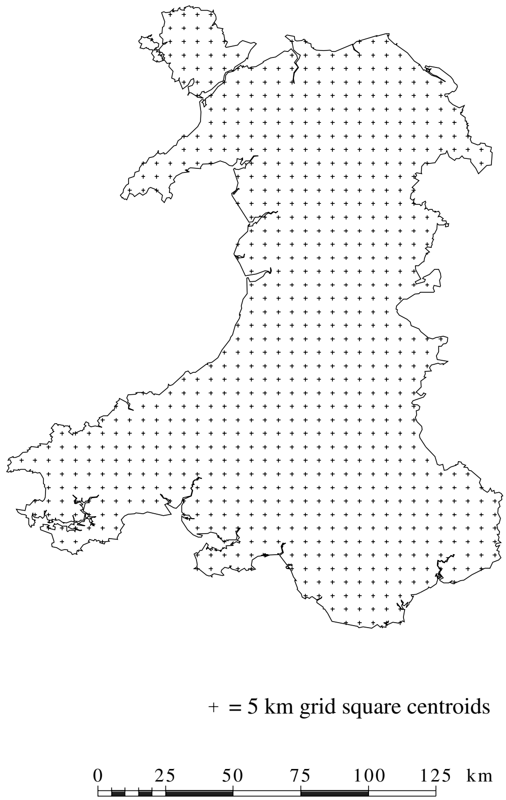

Figure 4.4. 5 km grid points used to generate the predicted woodland visitors surface.

time zones and zonal populations were defined for all of these sites and predicted visits estimated using the arrivals function. Inspection of these predictions showed that both the 2.5 km and 5 km resolution sites were sensitive to changes in local population density and the quality of surrounding road infrastructure (details in Bateman *et al*., 1995b). The detail afforded by the 5 km grid system indicated that such a resolution was adequate in reflecting the major contrasts in predicted visitor numbers engendered by population density and road availability/quality. Clearly a 2.5 km grid would give greater information regarding rates of change. However, given the very considerable processing demands of such a grid, and the acceptability of results from the 5 km resolution sites, such an approach seemed unnecessary. Accordingly travel time zones were calculated for a 5 km grid for the whole of Wales. The base map of grid points used to generate subsequent visitor potential surfaces is illustrated in Figure 4.4.

Regardless of the chosen resolution, certain sampling problems are difficult to alleviate. Inconsistencies arise from the interaction of the road network with the sampling pattern. Cell values depend upon how far a sampling point falls from any kind of road. Two areas equally far from population and with comparable road infrastructure might have different estimated travel times (and therefore predicted visit numbers) if in one of the areas the sampling point falls right on a road and in the other the sampling point is far from any road. There is no straightforward way around this arbitrariness. However, the findings for the two transects (and subsequently the entire area of Wales) were reassuringly sensible and predictable, suggesting that these inconsistencies had not had any significant impact.

Travel times were calculated for each of the 5 km grid sites as follows. A window was defined around each site and the site rasterised. An allocation process, using a cost impedance grid based on road characteristics (see Brainard *et al*., 1997), was run to find the shortest path linking the site and each other cell in the raster surface. The impedance necessary to reach each of these locations was assigned to corresponding cells in an output grid. This provided, in minutes of travel, a time-surface output which was then classified into time zones. Information on total residents for each of these areas was subsequently extracted from the rasterised population surface and recorded in a separate file. This process was then iterated across all sample sites in the 5 km grid.

Once time zones and zonal populations had been calculated for all grid points, woodland recreation demand (in terms of total party visits per annum) was predicted using the arrivals function. Figure 4.5 illustrates the resulting predicted woodland visitors surface.

Figure 4.5 strongly reflects the influence of population distribution upon the prediction of recreational woodland visits. In southern Wales the influence of cities such as Swansea and Cardiff and the densely populated 'valleys' area results in relatively high visitor predictions. Similarly, in the north-east, the influence of nearby English cities such as Manchester and Liverpool is very clear. Conversely, in mid Wales and western coastal areas, the sparse population results in very low visitor arrival estimates. Population impacts tend to be compounded by the distribution of higher quality transport infrastructure. This inflates the already high arrival numbers generated by the proximity of large centres of population. However, infrastructure effects are perhaps best demonstrated in areas of relatively low population density such as coastal, mid and north Wales. Figure 4.6 shows this area in detail, superimposing the relevant major road network. Here we can see that the presence of a major road creates a heightened potential visitor corridor as it facilitates visits by individuals from relatively distant travel time zones.

Mapping predicted recreation values

In Chapter 3 we derived various estimates for the unit value of a party visit to a recreational woodland. In particular we emphasised a lower-bound value of £1.82 per

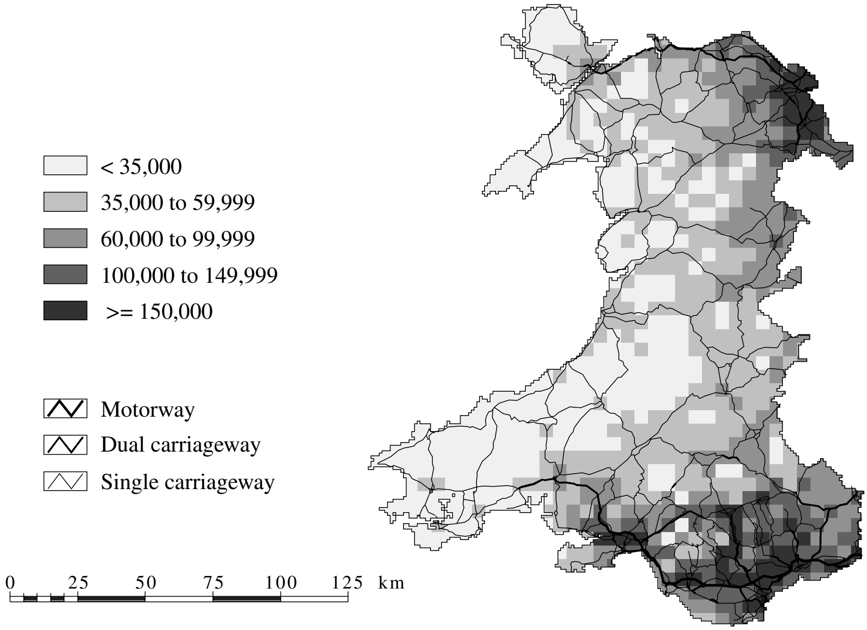

Figure 4.5. Woodland recreation demand in Wales: predicted annual total party visits per site. (*Source:* Bateman *et al*., 1999c.)

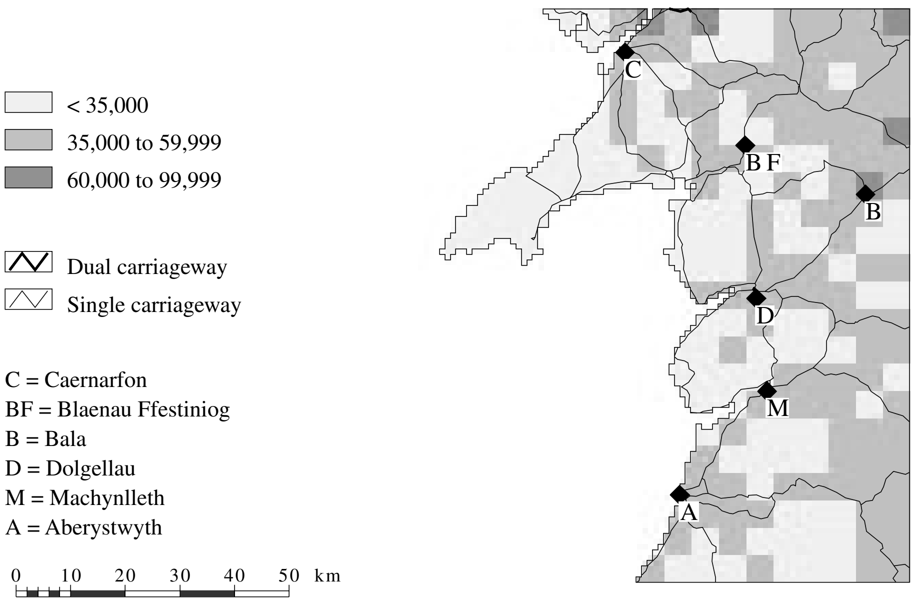

Figure 4.6. Woodland recreation demand in north-western Wales: predicted annual total party visits per site. (*Source:* Bateman *et al*., 1999c.)

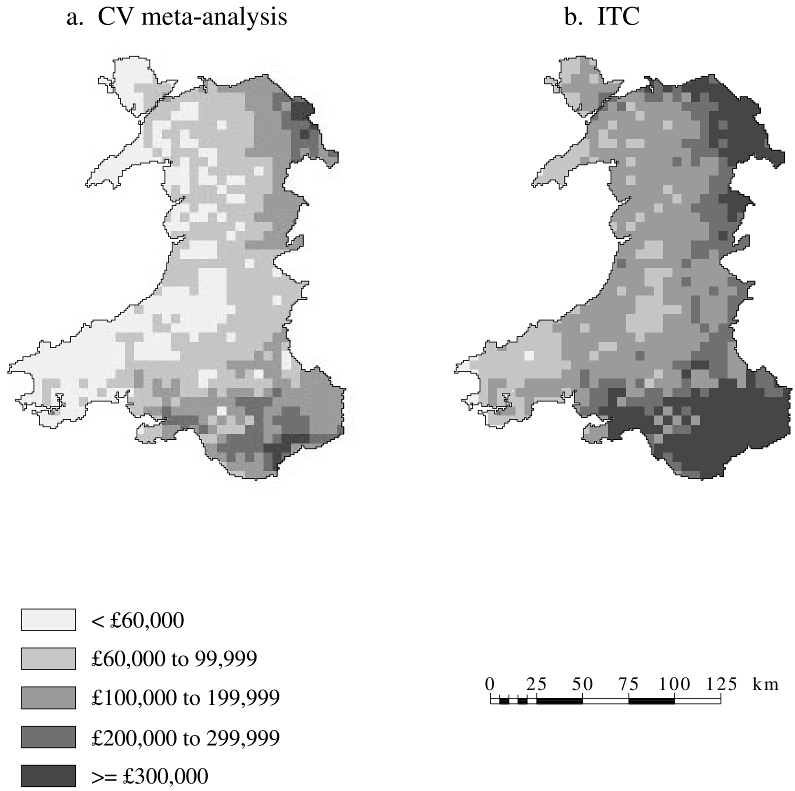

Figure 4.7. Predicted value of total annual woodland recreation demand per site using two valuation estimates: (a) lower-bound values based on cross-study analysis of CV values; (b) upper-bound values based on ITC study.

party per visit derived from our cross-study analysis of CV results and an upper-bound estimate of £3.59 per party per visit obtained from our ITC analysis. GIS capabilities were used to apply these values to our estimates of the number of annual party visits to a given (real or hypothetical) woodland to yield predictions of the total annual recreational value of sites. Figure 4.7 illustrates the maps of recreational value produced by this exercise.

The distribution of values within each of the maps shown in Figure 4.7 mirrors that of the base demand map (Figure 4.5). However, the fact that our upper-bound valuation is nearly twice that of our lower-bound estimate is well illustrated here. The degree to which this variability constitutes a cause for concern is uncertain. If we are confident of these bounds then, in a cost-benefit context, if the lower-bound value is sufficient to justify a switch from other land uses into woodland, further precision may be unnecessary. Similarly, if even upper-bound values are not large enough to justify such conversion, then again these estimates are sufficient for decision analysis. Only if the cost-benefit balance lies within these bounds is further precision required. Given this, then, at least as an exercise in methodological development, use of these estimates seems justified.

Extensions

The work described above details the extent of our research to date on the case study area of Wales and is used as the basis of the cost-benefit analysis presented in Chapter 9 of this volume. It also represents our only attempt to date to generate maps of arrivals and recreation values for large areas, embracing both existing and potential recreation sites. However, our recent and ongoing research concerning other areas of Britain extends our methodology for modelling visits and values. In this section we briefly review this work to provide the reader with a flavour of the directions in which this research is developing.

In work described in Lovett *et al*. (1997), Bateman *et al*. (1998) and Brainard *et al*. (1999) we examine how both the number of arrivals and the value of those visits to woodland sites alters according to a range of attribute characteristics. These include:

(i) travel costs, described by the accessibility of the site to the potential visiting population (i.e. taking into account the spatial distribution of the whole of the British population in relation to the study site)
(ii) the socio-economic and demographic characteristics of that potential visiting population (allowing for the possibility that, say, richer households or those with more children may visit such sites more often)
(iii) the availability of substitute woodlands described by an inverse weighted distance to all other British woodlands from all possible visitor outset origins
(iv) site quality characteristics (for example, presence and size of a car park, length of woodland walks, etc.).

These models represent a substantial extension to those described previously, both because of the additional explanatory variable considered and because they permit the estimation of site-specific coefficients, yielding estimates of consumer surplus for each individual site. This allows for the possibility that the value of the recreational experience varies between sites.

The first stage of this analysis involved calculation of a variety of variables describing items (i) to (iv) above for Thetford Forest. These variables were obtained from a variety of sources. Travel distances and times were calculated and population distribution obtained using the GIS as described previously. Data on the

Table 4.2. *Official recreational visit numbers, predictions of arrivals and consumer surplus estimates for twenty-seven English woodlands*

Site name	Official estimate of visits (per annum)	Predicted visits (per annum)	Per party consumer surplus (£ per visit)	Site consumer surplus (£ per annum)
Dunwich	18,980	15,957**	1.56	24,828
Two Mile Bottom	22,636	22,678**	2.72	61,676
Kielder Castle	24,243	56,747*	3.57	202,767
Forest Drive	31,641	26,200**	3.57	93,616
Warksburn	3,794	5,351*	7.42	39,706
Bogle Crag	14,924	47,475	5.38	255,408
Grizedale	85,181	81,015**	3.48	281,824
Noble Knott	7,543	35,407	3.51	124,149
Whinlatter	55,797	60,838**	3.36	204,571
Blackwater	39,338	37,518**	5.19	147,813
Bolderwood	22,963	28,503**	4.86	182,318
Moors Valley	165,552	157,561**	4.14	652,149
Bucknell	21,360	45,526	1.63	74,117
Salcey	77,650	75,644**	2.23	168,735
Wakerley	51,490	42,354**	2.06	87,456
Dalby	130,151	77,804*	3.31	257,260
Chopwell	42,298	54,251*	6.36	344,846
Hamsterley	76,796	71,770**	3.50	251,462
Simonside	12,430	32,526	2.94	95,462
Blidworth Bottom	54,547	41,844**	3.15	131,776
Blidworth Lane	52,754	45,103**	3.16	142,394
Blidworth Tower	37,596	45,288**	2.91	131,660
Chambers Farm	23,605	22,808**	1.92	43,836
Goyt, The Street	84,279	73,400**	2.63	193,058
Normans Hill	30,936	35,975**	2.66	95,748
Thieves Wood	72,276	45,617*	2.66	121,474
Sherwood Centre	38,919	42,325**	1.78	75,430

Notes: * = predictions within 50% of official estimates;
** = predictions within 25% of official estimates.

socio-economic and demographic characteristics of the population were obtained from the UK Census and this information was spatially assigned using the GIS. Distances from each possible outset origin on a regular grid across Britain to each potential woodland recreation substitute site[16] were calculated and an inverse weighting scheme applied (with weights being empirically derived by analysis of the outset origin of visitors to Thetford in relation to substitute availability from those origins) to give prominence to those nearer to each potential outset origin. Finally, site quality characteristics were obtained from the Forestry Commission.

These data were then used to estimate a model to predict visits (and values) for Thetford Forest. This was then transferred to predict arrivals and recreation values at twenty-seven English woodlands[17] for which official estimates of visits were available (although the Forestry Commission freely admitted that these estimates were somewhat approximate). Results from the transfer exercise are detailed in Table 4.2, contrasting official estimates with predictions derived from our transfer function from which estimates of per party and per annum consumer surplus are obtained and detailed.

Considering Table 4.2, our extended transferable model provides estimates which are highly correlated with those of the Forestry Commission ($p < 0.001$).[18] Given the lack of a gold standard for determining the accuracy of either set of estimates, this seems an adequate basis for future research and arguably provides an acceptable planning tool. Certainly this was the opinion of the Forestry Commission which recently asked the authors, together with their colleague Andy Jones (also at the University of East Anglia), to apply this methodology to a larger dataset of nearly 11,000 interviews conducted at forty sites across Britain. When completed, this analysis will be combined with a second, recently finished study (again with Andy Jones) commissioned by British Waterways, examining over 5,000 interviews conducted at fifty-three inland waterway sites across Britain. These studies further extend the methodology set out above by incorporating wider sets of socio-economic, site quality and substitute availability variables (for example, non-woodland substitutes such as waterways, beaches, built heritage and urban attractions are considered). At the time of writing, results from these studies were being prepared for publication. However, in both cases similar messages were clearly given by the data, which are of particular relevance to the work described in this volume. While

[16] Potential substitute sites were taken from a variety of sources including satellite imagery, the Institute of Terrestrial Ecology land use map and a joint Countryside Commission and Forestry Commission large-sample household survey of outdoor recreation.

[17] Brainard *et al*. (1999) consider a further six subsites, that is sites within a larger forest with multiple sites. However, our transfer model was unreliable for such applications, i.e. it only predicts for visitors to a distinct forest rather than for areas within a given forest.

[18] A regression test relating official estimates to our transfer predictions showed a coefficient which was not significantly different from 1 with a constant which was not significantly different from zero.

issues such as the socio-economic characteristics of potential visiting populations, the availability of substitutes and site characteristics are all significant predictors of visits and values, all of these variables are dwarfed by the significance of travel costs. It seems that the business world mantra of 'location, location, location' being the vital determinant of demand applies equally well to the demand for open-access recreational public goods. Indeed omission of all other explanatory variables yields relatively small estimation errors. Given the strength of this result we feel that the analysis presented in preceding sections remains a valid input to the CBA conducted in Chapter 9 of this volume.

Limitations of the predicted recreation values

We now return to our analysis of the case study area of Wales. While we feel that the recreation value maps illustrate the methodological potential of applying GIS techniques in this context, it is important to conclude this chapter with a brief discussion of a number of potential limitations and further issues which would have to be addressed before the full decision-making potential of this approach could be realised.

The supply side

Our analysis only considers the demand side of the woodland recreation 'market'. The recreation value maps indicate the recreation demand for a typical woodland established at any of the 5 km grid intersections of the base map (Figure 4.4). They do not tell us about the supply side of this market. There are two major ways in which the supply side interacts with demand to determine actual visits. First, the existing distribution of woodland will already have soaked up some of our predicted demand. Second, as new forests are planted and (with some time lag) recreational services become available, so demand becomes satisfied. If supply exceeds demand in any one area such that non-congested recreation sites already exist, then the demand for new sites will be lower than that predicted in Figure 4.7 which ignores the distribution of existing sites.

To a substantial degree these concerns are incorporated within the extension work described earlier through the addition of substitute availability variables. However, this work also shows that it is travel time and cost which remain by far the strongest determinant of visits and values. Therefore, while there is clearly scope for using this research to refine our visit prediction maps, the same research suggests that the results summarised in Figures 4.5 to 4.7 remain valid approximations of underlying relationships and are acceptable as an element within our subsequent CBA assessment.

Applicability of the Thetford Forest period to annual conversion factor

As part of our arrivals function calculations we had to convert from the survey period to an annual basis. One concern here is whether the conversion factor used is valid for other sites or unique to Thetford Forest. In order to test this fully we would ideally need data regarding the annual distribution of visits both at Thetford and at any site to which we wish to extrapolate. Unfortunately, as described in relation to Table 4.2, official estimates are still only rough approximations and robust values are currently unavailable for our Welsh study area. Gillam (pers. comm.)[19] suggests that seasonality patterns are likely to be roughly similar across England and Wales and only differ in very remote areas such as the north of Scotland where seasonal peaks are likely to be relatively more pronounced. On the basis of this information, and in the absence of any contrary evidence, we feel that we have adopted a defensible approach to this issue.

Comparability of recreation in Thetford Forest with that in Wales

The major demographic and infrastructure differences which separate Wales from our East Anglian survey site are explicitly accounted for in our arrivals function which allows for both population density and road network quality. Two remaining issues are pertinent here. First, does our survey site provide similar recreational services to those of our visitor potential map? By definition, the answer here is yes, because we are looking at the creation of similar service sites where the major recreational attraction is open-access walking and its associated activities. However, in the absence of data concerning site quality and facilities this approach will not be appropriate for predicting arrivals to non-standard real or hypothetical sites. Second, does the psychological perception of woodland recreation differ between East Anglia and Wales? In considering this we must separate it from the supply-side problem commented upon above. Once such a distinction is made we see no reason to suspect any inconsistency here (although it cannot be ruled out), an assertion reinforced by the earlier work of Colenutt and Sidaway (1973) in the Forest of Dean (on the Welsh border) which reports similar visitation patterns to those observed in our own analyses.

Conclusions

The analysis presented in this chapter has used a variety of GIS techniques to model visits to a specific woodland and then apply the resultant arrivals function to

[19] Simon Gillam (Chief Statistician, Forestry Commission) noted that the Thetford Forest estimates were believed to be among the most reliable available and therefore provided a reasonable basis for this analysis.

produce estimates of visitation to similar woodlands across our Welsh study area. The estimates have then been converted into money values using the valuation studies presented previously. Results conform well to prior expectations showing predicted demand to be linked to population distribution and site accessibility.

A number of problems have been identified in the course of this analysis. Both the per visit values and visit number estimates were not sensitive to certain site characteristics. However, the extensions described above provide a methodology for addressing these problems and the results of this recent work suggest that the errors created by such omissions are acceptably small, travel costs being the overriding determinant of visits and values.

Given this we can defend our analysis both on methodological and empirical grounds. Furthermore, the adoption of a sensitivity analysis approach, using upper- and lower-bound valuation assumptions to create an envelope of recreational values, represents a substantial improvement over the common omission of such values from land use planning. In subsequent chapters we augment these with further forest values before making a comparison of aggregate values with those from conventional agriculture in the Welsh study area.