Estimating arrival numbers and values for informal recreational use of British woodlands

Final report to the Forestry Commission

Andy Jones, Ian Bateman, Jan Wright

CSERGE
School of Environmental Sciences
University of East Anglia
Norwich
NR4 7TJ

February 2003

Overview

This report encompasses two interlinked research projects. The first of these investigates the potential for generating transferable models for predicting visitor arrival numbers at woodland recreation sites across Great Britain. The second project sets out to estimate transferable monetary assessments of the value of such woodland visits through a meta-analysis of previous valuation studies.

Both of these projects develop novel methodologies which demonstrably improve upon previous approaches to the issues addressed. The study of arrival numbers takes as its underlying approach a function transfer method whereby models relating the number of visitors to a sample of sites are estimated and then applied to predict arrivals at other sites. Here the basic assumption is that the functional relationship between the number of arrivals and a well specified set of predictors (such as measures of population distribution and socio-economic profile, accessibility and travel time, substitute availability, site characteristics, etc.), as described by the coefficients on those predictors, will hold between sites. Note that only the coefficients are assumed to be constant, not the value of the predictors themselves. In this manner we assume that the *relationship* between say substitute availability and visits is constant, but allow for the fact that the *level* of substitute availability will vary between sites.

The function transfer model developed in this research explicitly addresses one of the major empirical problems facing successful function transfer; spatial complexity between sites. The assumption of coefficient stability is only valid for well specified models and this is unlikely to be the case unless those models can incorporate the complex interplay between and variation within the rich set of predictors outlined above. Conventional analyses face severe problems in addressing this issue. For example, even the most fundamental determinants for site visits, such as site accessibility and consequent travel time, will vary enormously between sites. Each site has a different part of the road network serving it resulting in very different accessibility and hence arrivals. As we have shown elsewhere, simple assumptions designed to bypass the modelling of such spatial complexity can result in very substantial errors (Bateman et al., 1999a). Therefore the methodology developed here directly addresses the spatial dimension of functions transfer through application of a Geographical Information System (GIS).

GIS provides a ready route for obtaining measures of the underlying determinants of recreational visits including travel time and distance, travel cost, population distribution and outset origins for potential visitors, the socio-economic characteristics of those populations, and the spatial availability of substitutes and complements. Furthermore, these measures can be obtained in a consistent manner for both surveyed ‘study’ sites and unsurveyed ‘policy’ or ‘target’ sites. It is this consistency, compatibility, availability and richness of measures which provides the quantitative cornerstone which is a vital prerequisite for successful function transfer.

The models developed in the first part of this report are innovative not only for their use of GIS to incorporate the spatial complexity of function transfer, but also because they combine these methods with advanced statistical analysis techniques. Specifically models utilising the Poisson distribution (which are highly applicable to the data which consists of counts of visits from differing outset areas to a given set of sites) are implemented via multilevel modelling techniques. These permit explicit incorporation of natural hierarchies in the data within the modelling exercise. So, for example the clustering (or ‘nesting’) of visits within sites inherent in a multi-site transferable model can be converted from a source of unexplained error to be minimised to a source of variation providing insight into observed patterns of visitation. The resultant GIS-based, multilevel, function transfer model allows the superior prediction of woodland recreation arrivals numbers for sites distributed across Great Britain and, we would contend, constitutes a substantial advance in function transfer research.

The second part of the research undertakes a meta-analysis of previous studies examining the monetary value of woodland recreation visits. Here again a novel methodology is developed utilising multilevel modelling techniques. These permit explicit incorporation of hierarchical effects within the modelling exercise. This approach is employed to investigate a number of natural hierarchies in estimates obtained from the valuation literature. For example, tests of the theoretical consistency of recreational values examine whether value estimates are significantly higher in studies conducted by differing research authors. This example underlines the importance of this methodological development as conventional analyses suggest significant differences in value estimates may exist between authors, whereas explicit incorporation of natural data hierarchies within the multilevel analysis we have undertaken shows that such differences are, in fact, not significant.

We propose that, taken together, these models of visitor arrival numbers and recreational values represent a substantial advance upon previous research in this area. They outline a methodology which has ready potential to be developed into a practical tool for forest planning and management allowing the identification of optimal sites for the provision or development of new woodland recreational opportunities.

This report is divided into two sections. In Part One we discuss the development of the transferable models for predicting visitor arrival numbers at woodland recreation sites across Great Britain. Part Two sets out the work we undertook to estimate transferable monetary assessments of the value of such woodland visits through a meta-analysis of previous valuation studies.

Part One:
Estimating recreational arrival numbers at British woodlands

1. INTRODUCTION

1.1 General overview

The valuation of environmental recreational resources such as woodlands, has become an important issue in assisting policy makers to make decisions concerning the allocation of resources (Liston-Heyes and Heyes, 1999). A common requirement for bodies charged with managing a range of recreational sites is an understanding of the factors that influence the choices made by the public when deciding where to visit for their recreational activities. Typically gaining this information has involved undertaking large scale visitor surveys (Boxall et al, 1996). However, this form of data collection is a very time consuming and expensive process. A particular problem arises in situations where there are few visits made to a site, either because of its remote nature or the fact that it is undeveloped. In these cases there is little potential to gain information from surveys.

The potential to estimate models of arrivals from data gleaned from a set of surveyed sites and transfer them to estimate visits to unsurveyed sites provides an attractive alternative to repeated surveys. Consequently a large literature has developed in recent years to examine the potential for such transfers (see, for example, Bergh et al., 1997; Bergland et al., 1995; Boyle (n.d.); Boyle and Bergstrom, 1992; Brouwer et al., 1999; Desvousges et al., 1992; Downing and Ozuna, 1996; Kirchhoff et al., 1997; Krupnick, 1993; Loomis, 1992; Loomis et al., 1995; Smith and Kaoru, 1990; Walsh et al., 1992; Willis and Garrod, 1994; and review in Bateman et al., 2001). However, the principal focus of this research has been upon the transferral of value estimates rather than visitor arrival numbers. Bateman et al., (2002) characterise this situation as a case of 'horse and rabbit stew' whereby the focus of research has been upon the value of a unit of recreation (the 'rabbit') whereas the greater influence upon estimates of the total value of recreational demand is exerted by the number of visitors arriving at sites (the 'horse'). As demonstrated throughout the two parts to this research, estimates of the value of a day's recreation vary far less across woodlands than do estimates of the numbers of visitors to different woodlands. Given this disparity, it seems somehow odd that research has focussed so intensely upon transferral of values as opposed to transferral of visitor numbers and it is our opinion that this reflects both the allure of the former task and the spatial complexity of the latter.

Ourselves and colleagues have conducted a number of recent investigations into the estimation of transferable functions for predicting arrivals at woodland sites (Lovett et al., 1997; Bateman et al., 1999b; Brainard et al., 1999; 2001). The premise behind such function transfer analyses is that the availability of data from surveys of a set of sites allows statistical models to be developed to quantify the factors which determine visitation levels. If successful, these models may then be applied to predict potential visitor arrivals at other, unsurveyed sites.

As noted above, the majority of research in this area has focussed upon the estimation of functions for transferring recreation benefit values rather than arrival numbers. Indeed for the purposes of clarification we can separate out benefit function transfers from arrival function transfers (although, as discussed subsequently, both benefit values and arrival numbers may be estimated from certain types of transferable functions).

Considering for a moment the estimation of benefit transfer functions alone, data for such exercises may be obtained from two distinct types of valuation exercise; expressed preference and revealed preference techniques. Expressed preference methods, such as contingent valuation (Mitchell and Carson, 1989; Bateman and Willis, 1999), rely upon individuals' assessments of both market and non-market site values elicited through direct survey questioning concerning, for example, their willingness to pay for environmental goods or resources at a site. Conversely revealed preference methods, such as the travel cost method (Freeman, 1993), infer demand via observed behaviour. Travel cost studies elicit such observations via surveys of visitors to recreational sites, relating visit numbers to a variety of factors including characteristics of visitors, visits and the site.

Both expressed and revealed preference approaches have a number of advantages and disadvantages. Advantages of expressed preference methods, such as contingent valuation, include their ability to incorporate measures of both the use and non-use values of the environmental good being considered and general applicability to a host of real world or potential future decision questions. However, these methods are relatively poor at generating data for the estimation of arrivals functions. This is because, by definition, survey respondents find it difficult to quantify essentially subjective and even subconscious factors which impinge upon their decision to visit a given site. For example, the influence of substitute availability may be a significant factor in determining a visit. However, expression of this factor within terms which can readily be incorporated in transferable arrivals models involves questions which are inherently difficult for survey respondents to answer. Such difficulties are directly addressed through the application of the revealed preference travel cost technique.

As typically applied, the travel cost method seeks to place a value on non-market recreational goods by using the costs of consuming the services of the recreational asset as a proxy for price. These costs include those associated with travel, entry fees and on-site expenditure (Freeman, 1993; Hanley and Spash, 1994). The basic premise of the travel cost model is that the number of visits from a locality to a recreational site will vary with the distance from that site. As travel distance to the site increases, and thus the costs of travelling to the site increases, the number of visits decreases. Hence it is assumed that as travel costs increase, the net benefits derived for a given potential visitor from a visit diminish. This is captured in the coefficient relating travel costs to visits; a coefficient which can then be used as a proxy for the value placed on a recreational visit (Fix et al. 2000). However, importantly, when structured appropriately the functional relationship between travel costs, other predictors and visits also permits estimation of the number of arrivals to a site.

There are a variety of permutations of travel cost analyses. (Hufschmidt et al., 1983; Freeman, 1979; 1993; Bockstael et al., 1991; Herriges and Kling, 1999). One such is the zonal travel cost technique which focuses upon the prediction of visit rates to a given site or set of sites from a set of outset zones. The use of averages or other descriptive statistics to characterise outset zones has made the method less popular in recent years as a route for estimating recreational values. Instead benefits assessment work has focussed upon individual travel cost methods (Freeman, 1993) and random utility models (Bockstael et al., 1991; Herriges and Kling, 1999) which are based on the revealed preferences of individuals and are therefore more tractable and theoretically consistent for benefits assessment work. However, this reliance upon individual level data makes such methods less amenable for the estimation of transferable arrivals functions, as such data is by definition not available for unsurveyed sites. Conversely the reliance upon area statistics inherent in the zonal method opens up the possibility of estimating arrivals functions which are based upon information available at both surveyed and unsurveyed sites. Data sources such as the UK census (from which measures of socio-economic characteristics may be obtained), road network information (yielding measures of accessibility, travel time and associated costs) and a variety of locational data (such as digital maps providing information on the location of potential substitutes) are available as national coverages. If functions drawing upon such data can be demonstrated to yield reliable estimates of arrivals at surveyed sites they, in principle, should be readily transferable to other, unsurveyed locations. Such promise has rekindled interest in the application of zonal travel cost models for function transfer purposes (Loomis et al., 1995).

The full range of information necessary for full and complete function transfer is provided within a transferable demand function such as that detailed in Equation {1}. This links the number of visits to a site to the time and distance cost of those visits (thereby raising the possibility of simultaneous transferral of benefits value estimates across sites), and other predictors including the type and quality of facilities at the recreational site, the availability and type of substitutes, socio-economic characteristics of the population and other explanatory variables.

Equation {1}:

Visits	= f	(Travel time,	SocEcon,	Quality,	Subs,	X)
↑		↑	↑	↑	↑	↑
Number of visits to the site under consideration. Expressed as a visit rate in zonal models		Disutility of a visit, e.g. travel time. May be reformulated as travel cost to permit the estimation of monetary recreational values	Socio-economic determinants of visits (e.g. family structure, social class, etc.)	Type an quality of facilities provided at the site under consideration	Availability of a set of substitute sites.	A matrix of other explanatory variables

Previous criticisms of travel cost analyses have included the difficulty in obtaining accurate information on several of the variables defined in Equation {1}. Measurement issues arise concerning the accurate assessment of visitor outset locations, the distances travelled and travel time taken for visitors to reach sites, and travel times to potential substitute amenities. The spatial complexity inherent in many of these issues has resulted in a number of studies adopting simplifying assumptions to obtain basic measures such as travel time (e.g., Rosenthal et al., 1986; Mendelsohn et al., 1992); assumptions which have been shown to yield substantial errors in recreational benefit estimates (Bateman et al., 1999a). However, recent advances in Geographical Information Systems (GIS) technology have provided a superior foundation for tackling some of these data requirements (Brainard *et al*, 1999, 2001). In particular, GIS offers the technology to help resolve some of the spatial and data-handling problems associated with travel cost and function transfer, whilst facilitating several methodological improvements.

The research described in this part of the report concerns the application of a GIS-based zonal function transfer model to the estimation of recreational visitor numbers at a specific set of open-access, informal recreation sites. The model is constructed and calibrated by the use of data from the visitor surveys, and then applied to unsurveyed sites where there is no direct survey information available. The overall objective of the project has been to develop a model to predict informal recreation use at target sites, taking account of competing recreation sites, the accessibility and quality of the resource, and the characteristics of local populations.

The empirical focus of this research concerns the estimation of recreational visitor numbers at a sample of Forestry Commission woodlands across Great Britain. A significant complication concerning this application arose from the fact that the available data is in the form of interviews gathered from surveys which were conducted at short periods during the year and with relatively sparse information concerning annual arrival numbers. Furthermore these surveys are mostly concentrated during summer months and, as a consequence of this, only imperfect information exists regarding the distribution of visits across the year. This makes the aggregation of findings up to an annual basis problematic. To reduce the promulgation of errors it was decided that aggregation to an annual basis should occur at the end rather than outset of the modelling exercise. Consequently the empirical focus rests initially upon the prediction of the numbers interviewed at any one site during a standardised survey period, which in the case was set to be a 24 hour day. This standardisation for time allowed for variations in survey effort between sites to be controlled for. Aggregation of visitor estimates up to an annual basis is considered once the modelling exercise is complete.

The background to this particular application is presented in the following sub-section while a detailed description of the development of the methodology is given in Section 2. Section 3 presents details of the various models estimated in the course of this research while Section 4 concludes this part of the report (discussion of our meta-analysis of woodland recreation values being given in the second, concluding part of the report).

1.2 Background to the empirical analysis: The Forestry Commission

The Forestry Commission of Great Britain is a government department, which through its agency Forest Enterprise is responsible for the protection and expansion of more than 860,000 hectares of Britain's forests and woodlands (Forestry Commission, 2001). The aims of the Forestry Commission are to produce environmental, economic and social benefits from the forests it manages. Achieving these objectives involves balancing timber production with the wider benefits from recreational and environmental programmes. A key objective of the Forestry Commission is the requirement to develop opportunities for woodland recreation and to increase public understanding and community participation in forestry.

Forest and woodland cover varies from country to country in Britain. England's woodland cover currently remains at 8% of its land area. Around three-quarters of this is privately owned, with the remaining quarter being managed by the Forestry Commission. Scotland's woodland cover currently stands at 1.2 million hectares, 17% of its land area, of which around 40% is managed by the Forestry Commission. Half of Britain's forest and woodland is in Scotland and over 17,000 hectares of trees are being planted in Scotland each year. In total, forests and woodlands make up more than 14% of the land area in Wales, comprising over 280,000 hectares, 40% of which is managed by the Forestry Commission (Forestry Commission, 2001). Although forests and woodlands are managed for their timber production, this activity has a lower strategic importance to the nation than has been the case in the past. Despite this, UK timber production is still increasing rapidly and is expected to do so over the next 15-20 years. In addition to their role in producing timber, the recreational uses of forests are becoming a much more significant aspect of woodland management. This is largely due to the value of recreation being more fully recognised than in the past.

Over 70% of all adults have visited British forests in the last few years, making around 350 million annual trips to woods (Forestry Commission, 2001). Woodlands provide a wide range of recreational pursuits including walking, cycling, horse riding, orienteering, camping, fishing and bird watching. Major public forests also provide a range of key facilities including car parks, forest drives, picnic sites, camping sites, holiday cabins, marked trails, cycle ways, horse riding routes, maps and information centres. In general, the Forestry Commission encourages public access to the forests and woodlands it manages and also encourages private woodland owners to manage their forests for public access by providing grants to help pay for activities.

2. METHODOLOGY

This section describes the methodology used to generate the range of variables required in function transfer analysis, as outlined in equation {1} above. The Arc/Info GIS package was used to calculate the spatial data required for use in the zonal travel cost models presented here. The initial work involved the Georeferencing of visitor survey postcodes using the Central Postcode Directory (CPD). A method was developed to allocate visitors to outset zones and the GIS was used to calculate travel time from each outset zone to the recreational sites. Various data sources including the satellite-produced Institute of Terrestrial Ecology UK Land Cover Database and Bartholomew's digital map database were employed to identify potential substitute resources that could be visited by the residents of each outset zone, and their accessibility was then estimated using trip modelling functions. These results were combined with demographic characteristics of populations, obtained from the 1991 National Census, so that the influence of social and economic factors such as levels of unemployment, social class, age and urbanisation on visitor recreation demand could be determined.

2.1 Georeferencing of woodland sites and visitor outset origins.

Responses from the 1996, 1997 and 1998 visitor surveys, consisting of data collected at 40 sites across Great Britain, were provided by the Forestry Commission and used in the models presented below. It is important to note that, although the surveys were undertaken at a wider range of sites, only those for which information on the provision of services (such as forest trails and a visitor centre) could be provided were included in the models developed here. Data for sites not meeting this criterion were discarded. The questionnaire used in this work is available from the Commission who also selected the survey sites. Each site had one interviewer, who interviewed on a continuous survey basis such that when one interview was completed the next individual passing was then interviewed. For groups of two or more people, one person was selected to be interviewed. Individuals or groups not interviewed were recorded by the interviewer.

Table 1 lists the sites for which survey information was provided. These sites are mapped in Figure 1.

Site No.	Site Name	Forest District and Country
3	Afan Argoed	Coed y Cymoed, Wales
6	Alice Holt	South East England
8	Back O Bennachie	Buchan, Scotland
9	Beechenhurst	Forest of Dean, England
14	Black Rocks	Sherwood and Lincolnshire, England
15	Blackwater	New Forest, England
17	Blidworth Woods	Sherwood and Lincolnshire, England
18	Bolderwood	New Forest, England
20	Bourne Wood	Northants, England
33	Chopwell	Kielder, England
34	Christchurch	Forest of Dean, England
40	Countesswells	Kincardine, Scotland
43	Cycle Centre	Forest of Dean, England
44	Dalby	North York Moors, England
46	Delamere	West Midlands, England
49	Dibden	New Forest, England
51	Donview	Buchan, Scotland
61	Garwnant	Coed y Cymoed, Wales
66	Glentrool	Newton Steward, Scotland
68	Grizedale	Lakes, England
72	Hamsterley	Kielder, England
80	Kielder	Kielder, England
83	Kings Wood	South East England
84	Kirkhill	Kincardine, Scotland
86	Kylerhea	Fort Augustus, Scotland
95	Mabie	Ae, Scotland
111	Queens View	Tay, Scotland
117	Salcey	Northants, England
119	Sherwood Pines	Sherwood and Lincolnshire, England
121	Simonside Hills	Kielder, England
126	Symonds Yat	Forest of Dean, England
128	Thetford High Lodge	East Anglia, England
129	Thieves Wood	Sherwood, England
130	Thrunton Woods	Kielder, England
134	Tyrebagger	Kincardine, Scotland
137	Waters Copse	New Forest, England
141	Wendover	South East England
143	Westonbirt Arboretum	Westonbirt Arboretum, England
147	Willingham Woods	Sherwood and Lincolnshire, England
153	Wyre	West Midlands, England

Table 1: Site Location Names, Forest District and Country

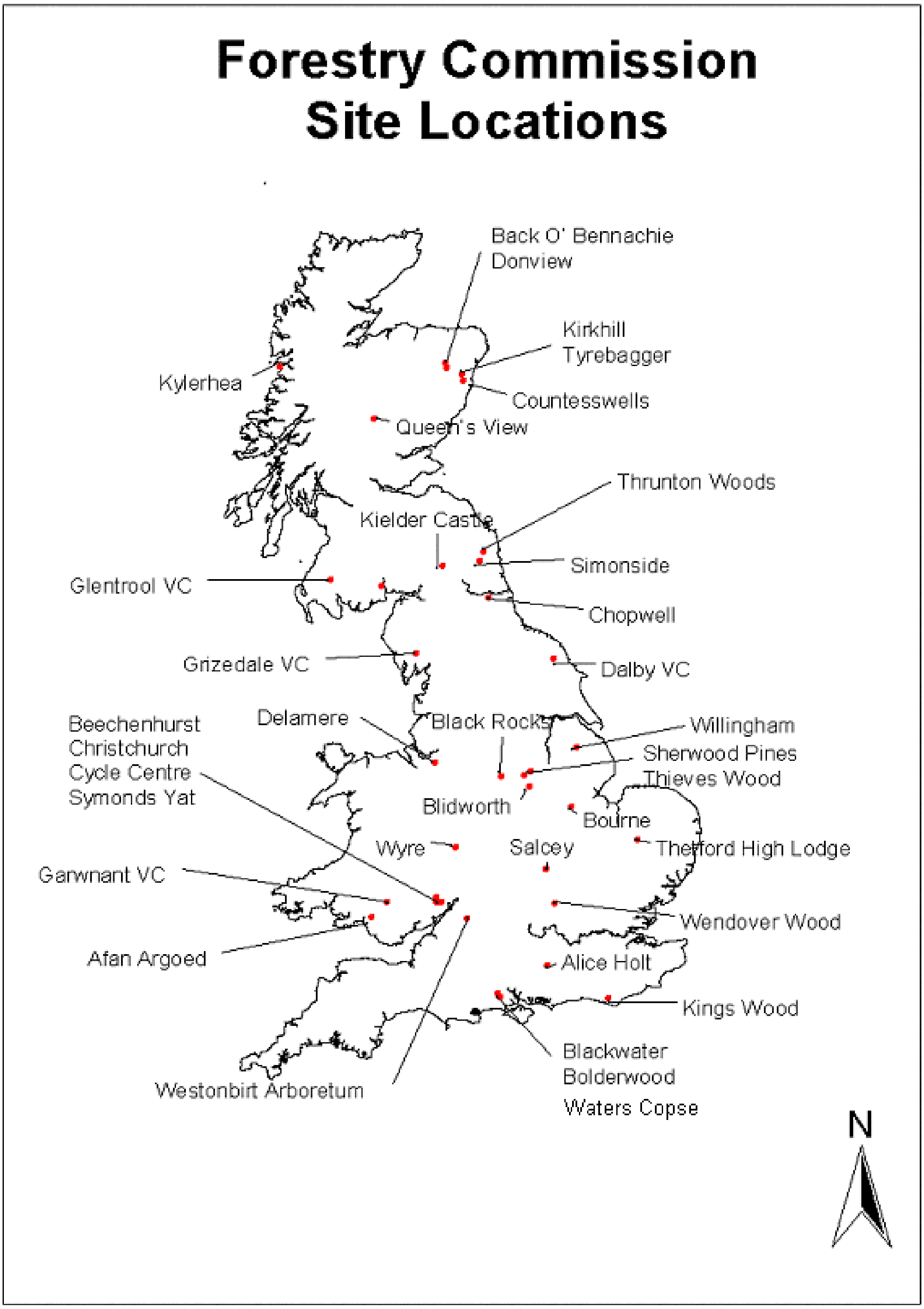


Figure 1: Survey sites used in this analysis

Table 2 shows the number of interviews conducted at each site. This number is subdivided into day-trippers and holidaymakers, based upon individuals' responses to Question 7 of the survey instrument (*"Have you travelled from home today?"*). In total 13,198 visitor records were analysed. Table 2 also details the survey effort expended at each site as measured by survey hours during which interviewing took place This is an important determinant of the number of interviews completed at each site and subsequent visitor predictions are adjusted for this factor.

Site No.	Site Name	Numbers of Visitor Surveyed			Survey Effort (Hours)
		Total Visitors	Day Visitors	Holiday Visitors	
3	Afan Argoed	458	381	76	157.0
6	Alice Holt	217	209	6	82.0
8	Back O Bennachie	100	92	8	69.0
9	Beechenhurst	128	84	43	34.0
14	Black Rocks	161	123	38	72.0
15	Blackwater	179	82	93	35.0
17	Blidworth Woods	216	211	2	106.0
18	Bolderwood	343	148	194	58.0
20	Bourne Wood	211	200	11	59.5
33	Chopwell	125	123	2	31.0
34	Christchurch	132	26	104	24.0
40	Countesswells	212	209	3	64.0
43	Cycle Centre	222	154	67	88.0
44	Dalby	305	157	148	72.0
46	Delamere	684	264	6	153.0
49	Dibden	215	206	9	89.0
51	Donview	144	126	18	66.0
61	Garwnant	358	274	83	80.5
66	Glentrool	321	114	205	100.0
68	Grizedale	265	68	197	51.0
72	Hamsterley	160	119	40	52.0
80	Kielder	104	38	64	26.5
83	Kings Wood	102	95	7	72.0
84	Kirkhill	207	197	10	107.0
86	Kylerhea	210	9	200	95.0
95	Mabie	686	355	315	108.0
111	Queens View	270	41	228	96.0
117	Salcey	196	185	9	54.0
119	Sherwood Pines	680	517	163	208.5
121	Simonside Hills	136	98	37	45.5
126	Symonds Yat	255	103	152	66.0
128	Thetford High Lodge	687	535	149	148.5
129	Thieves Wood	307	304	2	108.0
130	Thrunton Woods	142	89	52	48.0
134	Tyrebagger	149	139	9	71.0
137	Waters Copse	172	75	97	86.5
141	Wendover	117	112	5	42.0
143	Westonbirt Ab	440	349	86	44.5
147	Willingham Woods	176	163	12	124.0
153	Wyre	670	567	101	130.5

Table 2: Surveyed visitors, by type, at each site and site survey effort

The zonal basis of the function transfer models employed in this research meant that the key piece of information underpinning all analysis was accurate determinantion of the visitors' home outset location. This provides the reference point for identification of the explanatory variables specified in Equation {1} as relevant to a particular visitor. It is important to note that, for both day visitor and holiday maker models, measures are taken from the home location rather than any temporary address (e.g. a hotel or guesthouse). This is because, in the case of holidaymakers, no information was available on the location of their holiday residence. A consequence of this was that our analysis was unable to differentiate the factors that may determine choice of region within which to take a holiday from those which influence the locations visited each day during that vacation period. There is certainly potential for future work to be undertaken to determine the role of the explanatory variables analysed in this work in influencing these two sets of choices. However, this would require an extension of the survey questionnaire so as to obtain information on the temporary residence of holidaymakers. Although many of the day trippers in the sample will have set out from their homes, some may have also travelled to Forestry Commission sites from temporary addresses. Nevertheless, such locations would not provide an accurate description of the accessibility, substitute availability, socio-economic, and other variables which are pertinent to the visitor. If all sites attract a similar proportion of visitors from holiday as opposed to home outset locations then a model which does not distinguish between the two may prove suitable for general transfer purposes. However, such an assumption is unlikely to hold across a study area as large and diverse and Great Britain. Therefore our analysis makes allowance for the impact of such variation by estimating three types of model:

(i) undifferentiated all visitor models
(ii) models for only those visitors who set out from their home address (Day visitor models)
(iii) models for those setting out from temporary addresses (Holiday visitor models)

Visitors' home address locations were identified as follows. Interviewees were asked to state their home postcode. These were related to Ordnance Survey grid coordinates through interrogation of the 1995 release of the Central Postcode Directory (CPD) via the MIMAS system at the University of Manchester. A postcode covers a group of approximately 25 houses and the CPD provides a grid reference to a resolution of 100m of the location of the first delivery point in each postcode (Raper et al, 1992). It is important to recognise that rural postcodes may cover a large area and hence their grid references may not have such spatial precision as an urban postcode which, due to higher housing densities, will generally reference a much smaller area. However, postcoded grid references are widely used and, given the distances involved, are acceptable for this research. After primary interrogation with the CPD, those records which contained a postcode that could not be matched with a record in the database were manually checked. Where possible, typographical errors, such as the use of 'O' rather than '0' were corrected, and the records were re-interrogated. In total 2336 records did not contain any details of a postcode or contained an untraceable postcode, and hence were omitted from further analysis, as the presence of a valid postcode is a key requirement for travel cost methodology. This has left a valid dataset of 10,862 visitor records. The residential location of these visitors is mapped in Figure 2.

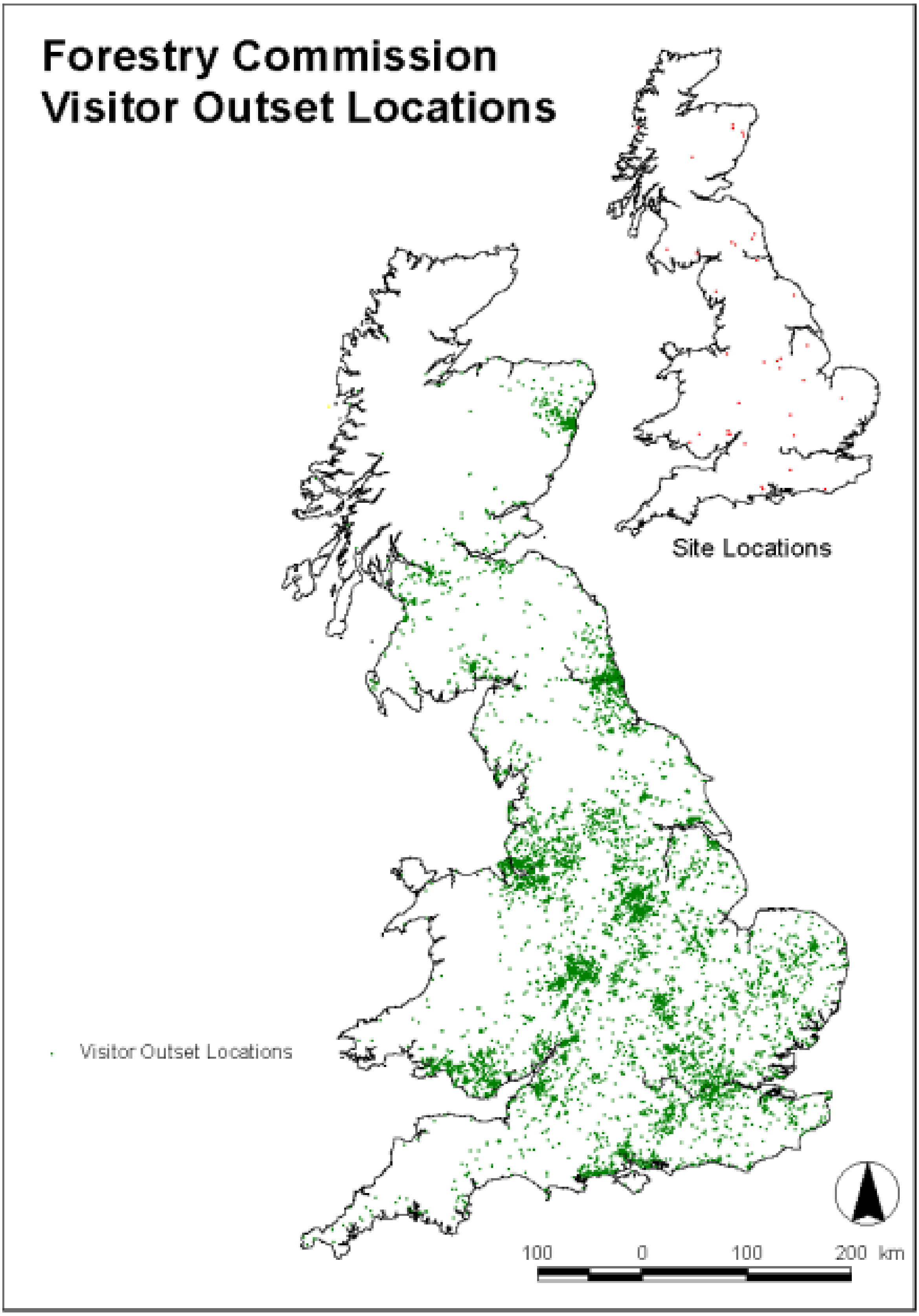


Figure 2: Visitor outset locations

Table 3 shows the ratio of the original survey responses provided to those with a valid postcode for each Forestry Commission site. All sites had some invalid postcode responses which were reasonably evenly, with the ratios of valid to original numbers being between 0.39 and 0.99 and the average ratio being 0.82. Six sites (Beechenhurst, Black Rocks, Hamsterley, Kings Wood, Kylerhea and Willingham Woods) had a ratio below 0.60, and hence had a substantial number of visitor surveys which could not be used in this research. The performance of these sites in the statistical models that were developed was examined closely, although we have no reason to believe that the missing visitor characteristics were not representative of those for which information was available.

Site ID No	Site Name	Original Totals	Valid Totals	Ratio Valid:Original
3	Afan Argoed	483	458	0.95
6	Alice Holt	254	217	0.85
8	Back O Bennachie	112	100	0.89
9	Beechenhurst	219	128	0.58
14	Black Rocks	276	161	0.58
15	Blackwater	257	179	0.70
17	Blidworth Woods	317	216	0.68
18	Bolderwood	403	343	0.85
20	Bourne Wood	225	211	0.94
33	Chopwell	145	125	0.86
34	Christchurch	156	132	0.85
40	Countesswells	236	212	0.90
43	Cycle Centre	336	222	0.66
44	Dalby	323	305	0.94
46	Delamere	718	684	0.95
49	Dibden	239	215	0.90
51	Donview	161	144	0.89
61	Garwnant	395	358	0.91
66	Glentrool	434	321	0.74
68	Grizedale	310	265	0.85
72	Hamsterley	336	160	0.48
80	Kielder	167	104	0.62
83	Kings Wood	173	102	0.59
84	Kirkhill	230	207	0.90
86	Kylerhea	325	210	0.65
95	Mabie	748	686	0.92
111	Queens View	348	270	0.78
117	Salcey	225	196	0.87
119	Sherwood Pines	722	680	0.94
121	Simonside Hills	138	136	0.99
126	Symonds Yat	354	255	0.72
128	Thetford High Lodge	731	687	0.94
129	Thieves Wood	348	307	0.88
130	Thrunton Woods	147	142	0.97
134	Tyrebagger	169	149	0.88
137	Waters Copse	217	172	0.79
141	Wendover	139	117	0.84
143	Westonbirt Arboretum	481	440	0.91
147	Willingham Woods	446	176	0.39
153	Wyre	755	670	0.89

Table 3: Ratio of Original to Valid Records for Forestry Commission Dataset

2.2 Site characteristics

It is important to consider information on the characteristics of each Forestry Commission site, as the facilities present may influence both the number and type of visitors attending. Whilst is this undoubtedly the general nature of a woodland environment that attracts visitors to forests, photographs in Figures 3 to 8 illustrate the diversity of features provided at woodland locations. For example, the small information board at Blidworth Woods in Nottinghamshire is shown in Figure 3 and indicates the presence of signposted walks. This is typical of the type of facility available at small sites. In comparison, a larger Forestry Commission site is likely to offer the range of facilities on offer at Sherwood Pines, also in Nottinghamshire, shown in Figure 4. This site offers cycle paths, a cycle hire centre, café, adventure park and car park, all clearly signposted. Figure 5 shows one of the cycle trails at Sherwood Pines, a feature of many woodland sites. Figures 6 to 8 contain photographs taken at Wendover Woods in Southeast England. They illustrate the types of facilities available at many of the woodland sites in this dataset. Figure 6 and 7 are typical of the type of children's play facilities available at many of the Forestry Commission sites introduced to encourage families to visit woodlands. Figure 8 illustrates a typical walking area which many visitors look for when visiting woods and forests.

Figure 3: Blidworth Woods Information Board

Figure 4: Sherwood Pines Information Board

Figure 5: Sherwood Pines Cycle Trials

Figure 6: Wendover Woods Children's Play Area

Figure 7: Wendover Woods Children's Play Area and Cycle Path

Figure 8: Wendover Woods Cycle Route and Walk

The variety of site characteristics illustrated in the Figures 3 to 8 may influence the number of visitor numbers at woodland sites. For example, the number of visits to a woodland site may be related to the car parking capacity at that site. To address this issue, a list of potential woodland facilities and quality attributes was developed in cooperation with the Forestry Commission who then supplied corresponding data for all sites considered in the analysis. Table 4 illustrates the list of characteristics supplied.

Table 4: Forestry Commission Site Characteristics Checklist

Forestry Commission Site Characteristics	
	Tick if Present
Car park	
Picnic site	
Forest walk	
Cycle trail	
Horse riding route	
Orienteering course	
Children’s play facilities	
Forest drive	
Viewpoint	
Hides	
Camping/caravan site	
Fishing allowed	
Water feature, e.g. lake, river	
Bothies	
Visitor centre	
Interpretation point	
Café	
Shop	
Cycle hire	
Forest classroom	
Toilets	
Disabled toilet	
Disabled access to shop/cafe	
Disabled walks	

2.3 Delineation of outset zones for calculation of visitation rates

Zonal travel cost models work by modelling visitor rates. Hence there is a requirement to delineate defined population catchments or outset zones, from which visitor rates may be calculated. One possibility would have been to calculate a visitor rate for each electoral ward in Great Britain. There are approximately 8,985 electoral wards in England and Wales and 1,158 in Scotland (Denham, 1993), each covering an average population of just over 5000. However, a comparison of the ward boundaries with the distribution of visitor origins indicated that this methodology would be problematic because the majority of wards would house no visitors to any given site. For example, Figure 9 shows the distribution of visits to the Forestry Commission site at Salcey. Here, 196 valid visitor surveys were completed with visits originating from just 79 different wards. Hence around 10,000 wards provided no visitors in this case. Adopting such a classification is likely to lead to estimation problems even for a count data approach such as the Poisson methodology discussed in the overview of this report. This is because a large number of zero visitor rates in the model could introduce problems associated with the presence of extra-Poisson variance. This occurs when the dependent variable shows a strong degree of right skew that is associated with an abnormally high number of low values in its distribution. As a solution, Local Authority Districts, of which there as 453 in Great Britain, were chosen as the units from which to delineate outset zones. Districts are large enough to each provide an adequate number of visitor arrivals at sites, and hence avoid the problems of extra-Poisson variance, whilst they are still small enough to preserve an acceptable amount of homogeneity of population characteristics within their boundaries.

Figure 9: Visitors to the Forestry Commission site at Salcey

2.4 Calculation of travel times to sites

As indicated in Equation {1} a key factor influencing the number of visitors to a site is some measure of the disutility of making visits to that site. In valuation studies this is typically taken as an inferred estimate of the travel cost incurred by the visitor. However, Randell (1994) notes, in a key criticism of the validity of such values, that visitors themselves do not explicitly observe this travel cost estimate and therefore are not necessarily reacting to it in the way implicitly assumed by travel cost estimates of consumer surplus. Partly in response to this criticism (and partly because of previous experience) it was decided to measure the disutility of visits via the travel time required to reach sites. Unlike inferred travel cost this variable is directly experienced by visitors and can therefore reasonably be assumed to impinge upon their journey decisions.

The calculation of travel times from the outset zone of residents to the site at which they were surveyed was undertaken using the GIS. This analysis used a digital version of the UK road network within which each road section was coded according to the estimated amount of time a vehicle travelling at a typical speed would take to traverse it. A simulation was developed in Arc-Info that predicted visitor routes, and estimated the associated travel distances and times.

Although the population is, of course, distributed in a non-continuous fashion throughout each district, it was necessary to estimate just a single travel time measure for each outset zone to each site. Generally such measures are taken from the centroid, or geographical centre point, of each zone. However, districts are quite large and the centroid is not necessarily representative to the location of most of the population. Hence, a population weighting methodology was developed[1]. Firstly ward centroids were chosen as the base point from which to measure travel times. For each electoral ward, population centroid data was downloaded from MIMAS. These centroids had been generated based on the work of Martin and Bracken (1991). This provided an easting and northing grid reference for each ward that corresponded to the average location of the population of that ward. The calculation of travel times from the centroids of all wards to all sites was undertaken through GIS using the methodology described below. Following the work of Brainard *et al.*, (1997) a simulation was developed in Arc/Info that predicted visitor routes, and estimated the associated travel times. This analysis used Bartholomew's digital version of the UK road network obtained from MIMAS, an extract from which is illustrated in Figure 10. The travel time value obtained for each ward was then grossed up to the district level by, rather then simply taking an average value, estimating a weighted average for each district, where the weighting factor was set to be the population of the ward. This had the effect of correcting for geographical variations in the location of populations within the district.

Each section of the road network was coded according to the estimated amount of time a car travelling at a typical speed would take to traverse it. Roads of all classes were included and typical speeds for each class of road were calculated using information from Department of Transport publications (DoT 1993) supplemented by research undertaken by Brainard *et al*. (1996) who empirically compared journey times along 31 known routes with those generated using the DoT data. Roads were classified according to both class and type of area they passed through since speeds in urban areas are generally slower than over the same class of road in rural areas. The final speeds used for this research are shown in Table 5.

[1] Bateman et al., (1999a) contrast travel cost measures of recreation value based upon both population weighted and geographical centroids

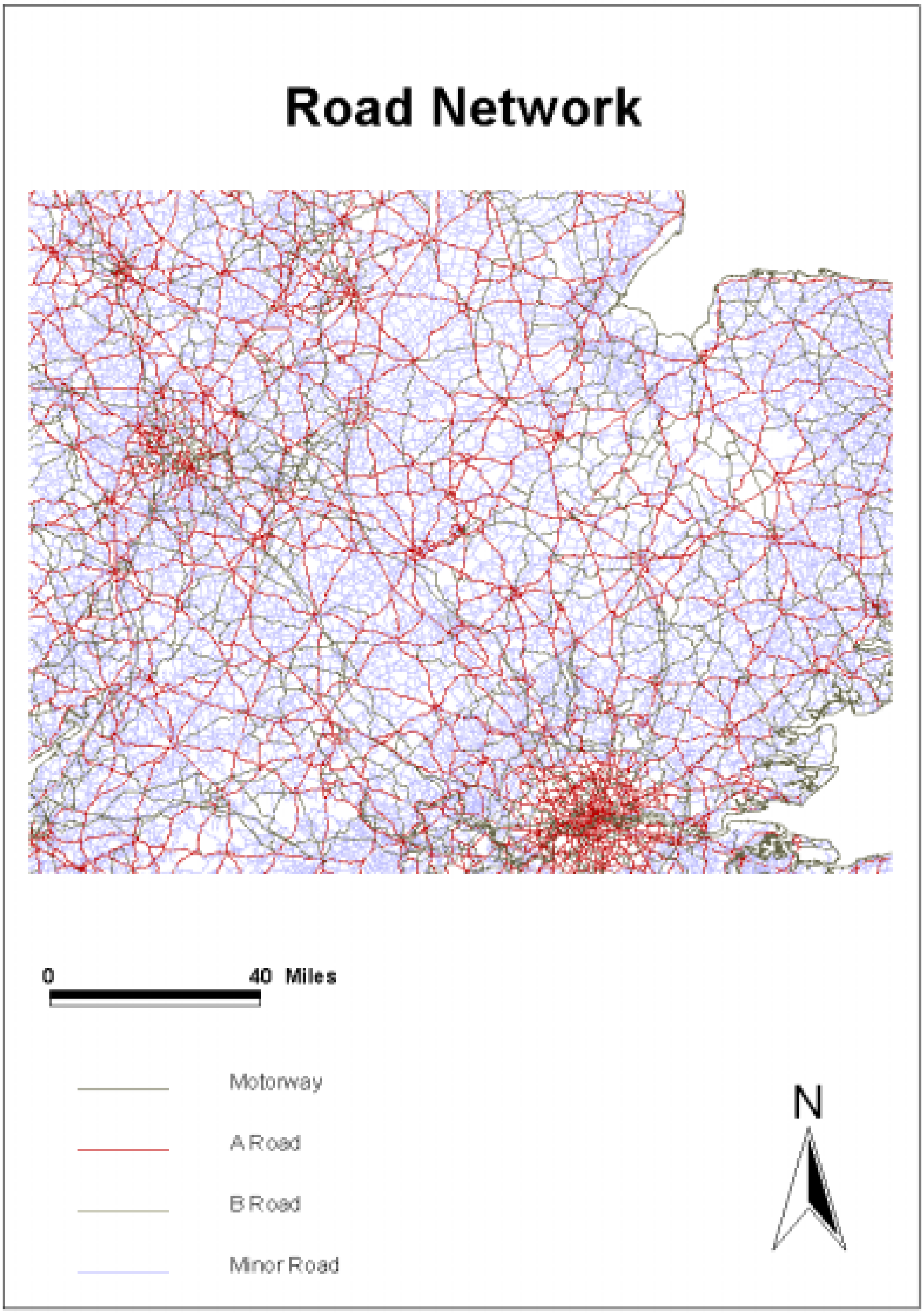


Figure 10: An Example of the Road Network Showing an Area of East Midlands and Southeast England

Road Type	Average Road Speed (mph)	
	Rural	**Urban**
Minor Road	14	11
B-Road Single Carriageway	24	12
B-Road Dual Carriageway	36	18
A-Road Single Carriageway	32	18
A-Road Single Carriageway Trunk Road	45	25
A-Road Dual Carriageway	50	25
A-Road Dual Carriageway Trunk Road	54	28
Motorway	63	35

Table 5: Road Classes and Speeds Adopted for this Analysis

A digitised road network only represents the centre lines of roads and, of course, most postcodes do not fall exactly on this road centre line. Hence a method to assign them to a point on the road network was required. The original road network was produced in vector format (point and line data) this needed to be converted to a raster (an image that has been sub-divided into regular tiles to represent a grid surface) to enable those postcodes that did not fall directly on a road to be analysed. Hence the road network with corresponding travel speeds was then converted into an 'impedance surface' within the GIS.

An impedance surface describes the travel effort (measured here in terms of time) associated with moving through any particular grid cell. As such it can be used to calculate the total time taken to move from any one location to another. The value of each cell was set to represent the time-per-unit distance of passing through the cell. The road network was converted to a raster surface of 500 * 500 m cells, the smallest size deemed to be manageable in terms of data storage for an area the size of the United Kingdom. This process left many empty areas between roads which were not assigned an impedance value. These areas were assigned values associated with increasing distance from the road network. Empty cells within 1500m of the known road network were filled with values of their nearest neighbours. Any cells still empty after this step were filled by incrementing the nearest cell values by 7 minutes. This was because it was assumed that visitors would walk to the nearest cell with an impedance value and 7 minutes is the assumed typical walking speed to traverse 500m (Brainard *et al.*, 1997). Hence every cell on the rasterised surface contained an impedance value related to the time taken to traverse that cell. An extract from the derived impedance surface is illustrated in Figure 11.

The impedance surface was then used to determine the accumulated travel time between ward centroids and Forestry Commission sites, by using the Arc/Info cost allocation command. This cost command calculates the travel time associated with moving from one cell to another from a starting point (in this case the centroid) to a finishing point (in this case a site).

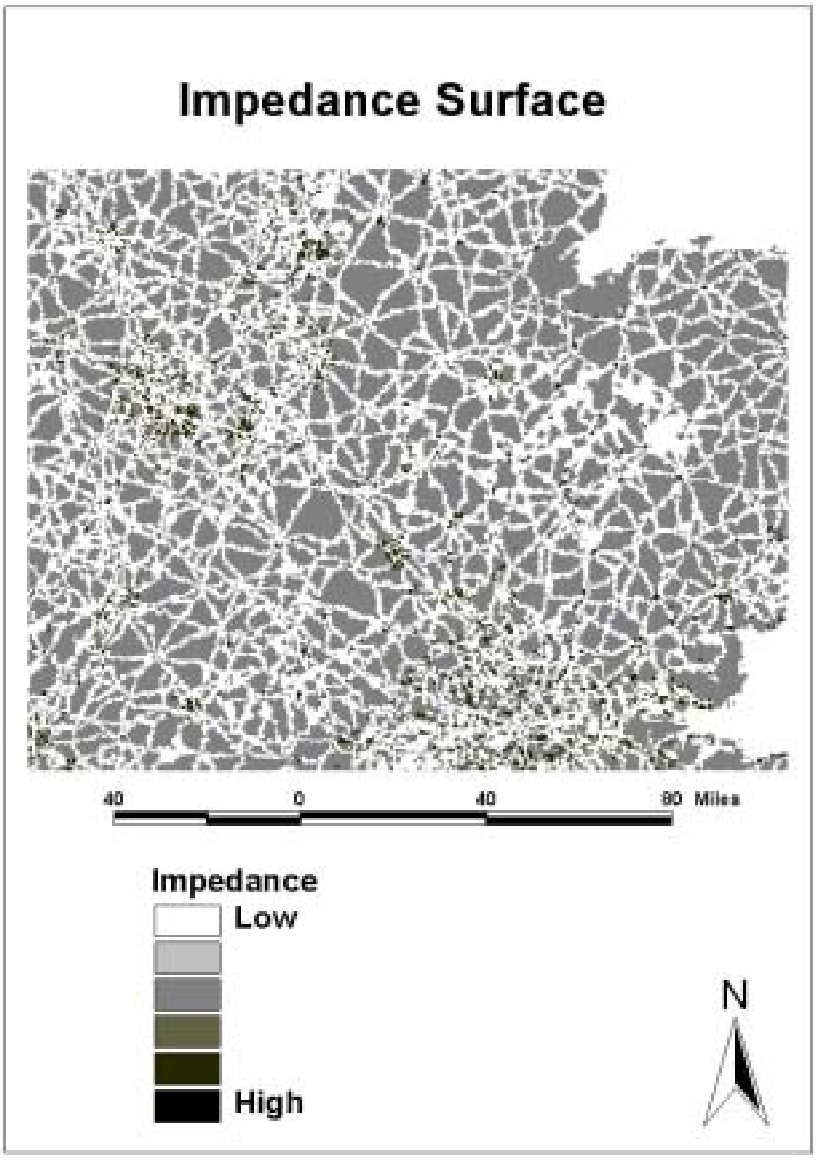


Figure 11: Section of the Impedance Surface in Eastern England

The output of this cost command, for the site at Salcey in Northamptonshire, is shown in Figure 12. A concentric pattern with prominent, irregular extensions coinciding with faster major roads is evident. Each cell on this image represents the estimated amount of time it would take to drive from that location to Salcey, assuming the fastest route was taken. It was these values that were initially assigned to ward centroids. This assignment was implemented by the development of a GIS routine that read off the travel time value from the cell that was spatially concurrent with each centroid, and assigned this value to each ward. The ward values were then grossed up to district level in the manner described previously.

The use of GIS in this manner enabled the calculation of travel time based on assumed routes from each ward. However, this procedure does have some limitations. Firstly it was assumed that vehicles would travel at average speeds over each road type, but in reality this will be affected by factors such as rush hour traffic, the time of day and length of journey. Furthermore this process cannot account for 'meanderers', those people who enjoy travelling and include this as part of their enjoyment of a visit. These individuals may not take the most direct or fastest route to the site being visited. However, although individual travel time may vary, Bateman et al., (1996) show a very strong correspondence between GIS generated travel times and those reported by woodland visitors interviewed in an on-site survey and we are confident, as borne out by the results of the models described below, that this indicator provides a highly accurate measure of accessibility for the vast majority of visitors.

.

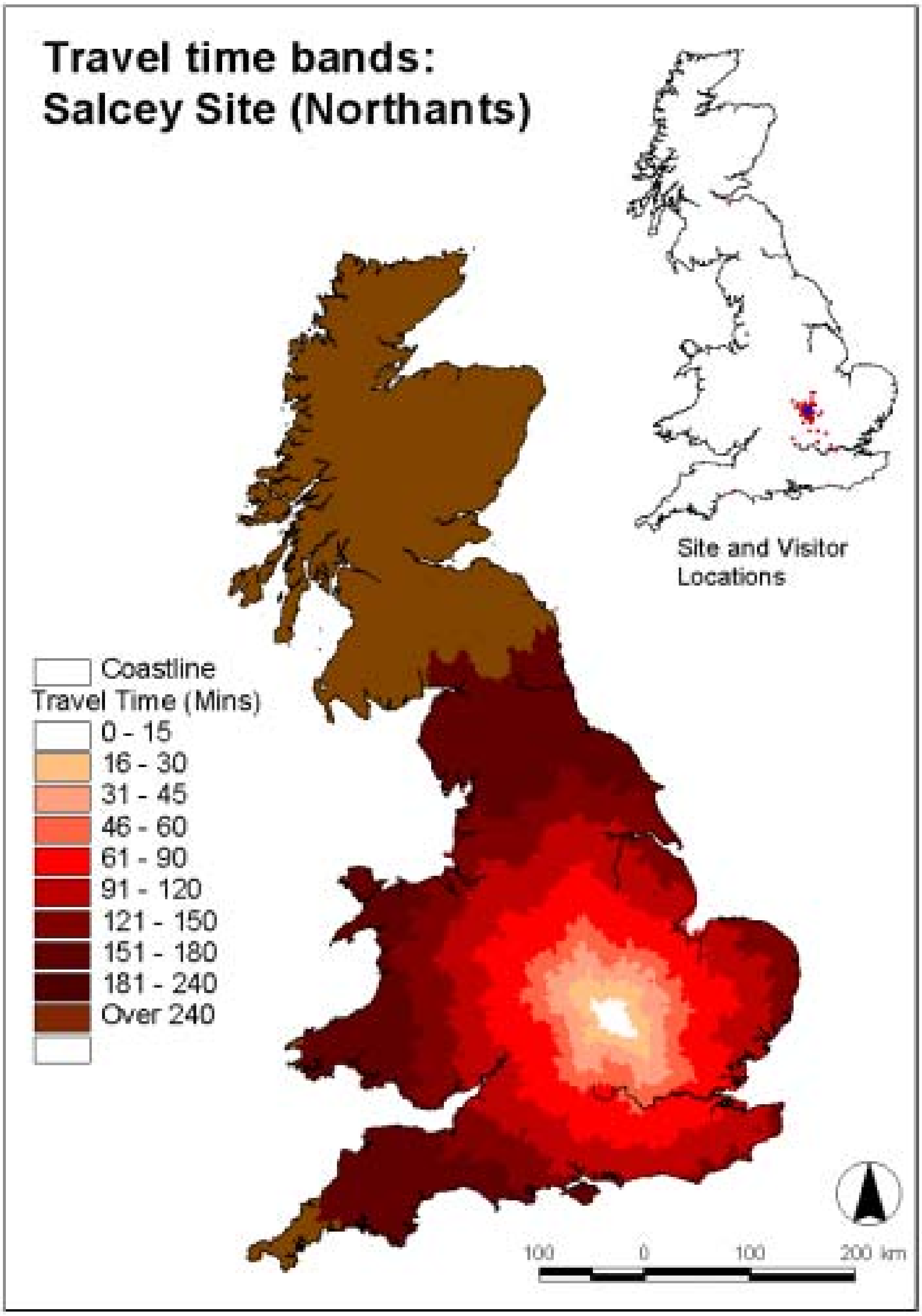


Figure 12: Estimated Travel Time Bands (in Minutes) for Salcey Site (Insert Shows Location of Site and Visitor Outset Locations)

2.5 Calculation of Travel Times to Substitute Attractions

2.5.1 Substitute Choice

In addition to the consideration of travel times to each site from outset zones, the existence of substitute attractions may influence the number of visitors to woodland locations. The availability of substitute attractions has often been ignored in travel cost studies (Willis and Benson, 1988). However, their presence may be an important factor in determining visitor rates.

The nature of the substitute relationship may be complex. Generally analyses consider substitutes at a micro level where they may act to draw visitors away from a given site to alternative recreation opportunities. However, at a more macro level the presence of a concentration of substitutes may draw people to an area (especially holidaymakers who by definition have greater discretion about their choices than do day-trippers). Hence, our analysis considers both a wide range of substitutes and a number of spatial scales in the preparation of substitution accessibility indicators for each outset zone.

Substitute attractions cover a wide and complex array of activities and it would not be possible to consider all of them in a single study. Consequently some key attractions, such as, rivers, beaches and National Parks were selected that were felt to offer similar or complementary recreational experiences to woodlands. A further series of amenities were also selected as being representative of developed attractions, such as, wildlife parks, theme parks and National Trust properties. Towns and cities contain numerous different types of attraction, hence large towns and cities were included as substitutes so as to provide a surrogate indicator for some of the attractions found within them but were not directly measured by us, such as cinemas, shopping centres and sports centres. Substitutes for which accessibility indicators were produced are outlined in Table 6

Countryside/Natural Attractions	Developed Attractions
Main Rivers	Large Towns and Cities
Woodlands	Zoos and Wildlife Parks
Forest Parks	Theme Parks
Heathland	National Trust Properties
Sandy Beaches	Historic Houses
Inland Waterways and Canals	
Coastal Areas	
Scenic Areas	
National Parks	

Table 6: Substitute Types for which Measures of Accessibility were Calculated

A wide range of data sources were employed to identify potential substitutes. These are discussed in some detail in sections 2.5.1.1 to 2.5.1.5 below.

2.5.1.1 The Institute of Terrestrial Ecology

The Institute of Terrestrial Ecology 1990 Land Cover Map of Great Britain contains 26 land cover types at an output resolution of 25 m resolution cells. This dataset has been derived from the classification of images provided by the Landsat-5 Thematic Mapper satellite. Classes in the land cover types identified in this map include such land covers as woodland, heathland, agricultural land and inland water. This data source was interrogated in Arc/Info and all cells containing a given land cover code were identified and used to create separate new GIS layers that could be employed in subsequent accessibility analysis.

2.5.1.2 British Waterways

A digital representation of locations of all British Waterway features (including recognised recreation access points) was provided by the organisation in vector format. This was used to estimate the accessibility of British Waterways canalside facilities to outset zones.

2.5.1.3 Bartholomew's Digital Database

The Bartholomew's 1:250,000 Digital Database for Great Britain was used to identify the location of main rivers, woodlands, forest parks, heathland, sandy beaches, scenic areas, National Parks, coastline, urban areas, historic houses, theme parks, zoos, wildlife parks and National Trust properties. This data source is a vector dataset digitised from 1:250,000 paper maps. However, when checking for completeness of feature representation some facilities, such as certain theme parks, zoos and wildlife parks, were found to be missing. Hence the *www.daysoutuk.com* internet site was used to obtain the postcode of these missing features. Grid references of these features were subsequently obtained from the Central Postcode Directory, and these were used to update the relevant GIS layers with the missing substitutes. The postcode locations of a number of wildlife parks and zoos not included in the Bartholomew's coverages were not found on the *daysout* site. These were hence obtained from other published sources such as International Zoo Yearbook (Olney and Fisken, 1998) and the Good Zoo Guide (Ironmonger, 1992).

2.5.1.4 Other Published Sources

The National Parks coverage was based on designated parks but there are officially no National Parks in Scotland. However, there are areas of Scotland with the same characteristics as National Parks and thereby these constitute similar substitute attractions. The advice of the Scottish Council for National Parks and the former Countryside Commission for Scotland was taken which led to the identification of six areas for National Park equivalent designation. These were: The Cullins of Skye, Assynt-Colgach, Wester Ross, Cairngorms, Loch Lomond and The Trossachs (Sharpley, 1993). The boundaries of these areas were manually digitised from paper maps and were added to other designated National Parks throughout England and Wales to yield a complete GIS layer for Great Britain.

2.5.2 Calculation of Area Weighted Travel Time Values for Substitutes

For those substitutes comprising areas rather than linear features, such as woodland, lakes and scenic areas, both point and area weighted accessibility measures were derived. While the former treat the substitute as a single location, the latter allow for the fact that the larger the area of the substitute is, the more accessible it may be. For example, in the calculation of point based measures described above, travel times to the closest feature that were independent of feature size were assigned to outset locations. Hence travel times to patches of wood just $1km^2$ in size were assigned the same precedence as those to a large afforested area such as Thetford Forest. These variations in feature size may be important influences on the likely number of visitors attending sites. Hence, for the purposes of comparison with the other substitute accessibility measures, weighted estimates were calculated as an alternative index of travel time accessibility.

The procedure developed for defining area weighted accessibility measures is illustrated using the woodland coverage as an example. The original woodland GIS layer (from the 1990 Land Cover Map) contained all woodland of all sizes. This was used to select woodland based on area size which was categorised according to tersiles (three equal parts) of area. Four woodland layers were produced to represent small, medium, large and all woodlands. Illustrations for East Anglia are shown in Figure 13.

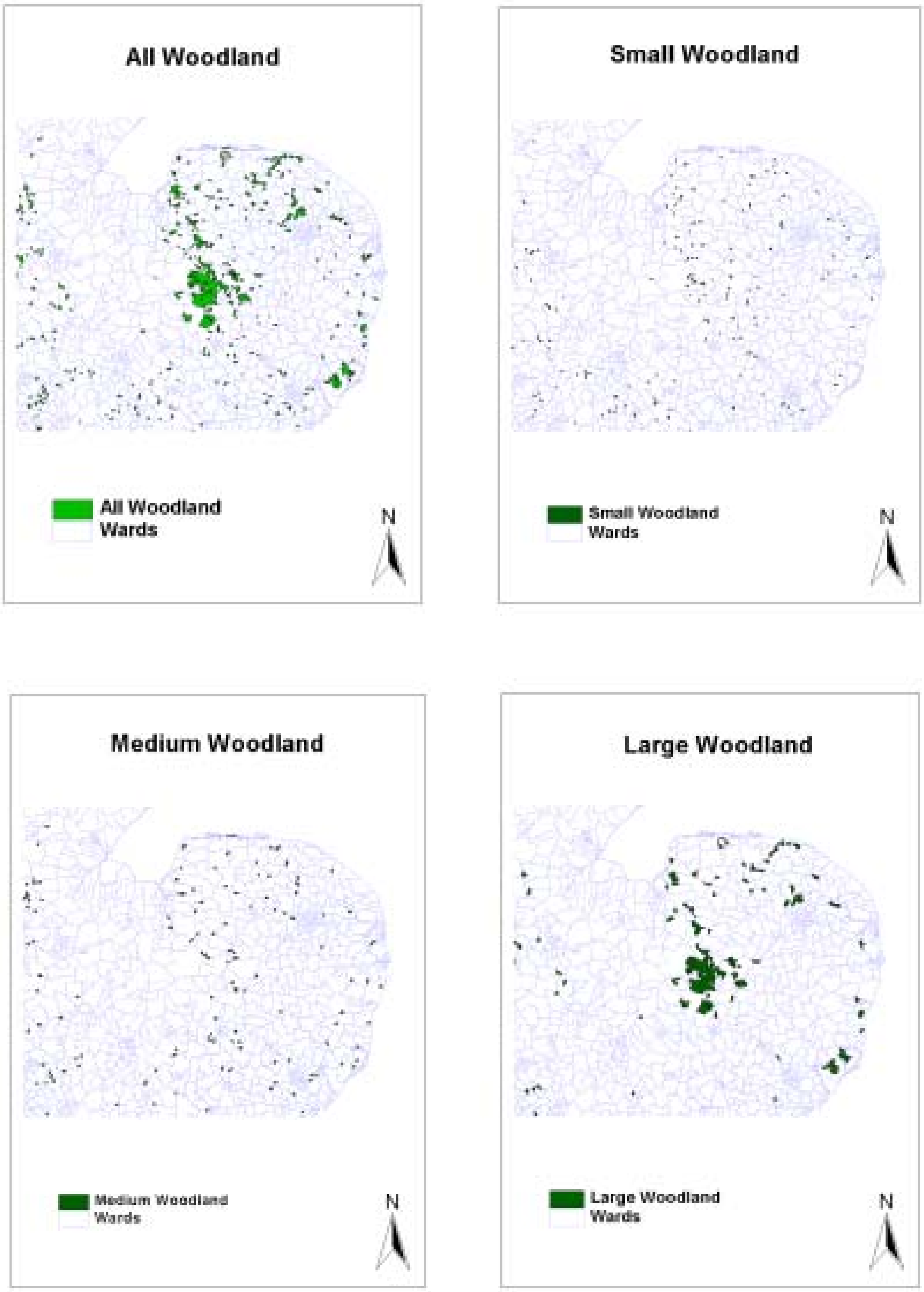


Figure 13: Woodland – All, Small, Medium and Large Woodland Areas for an Example Area in East Anglia

A surface of estimated vehicle travel times was created for each small, medium and large woodland coverage, using the same methodology as described above. This procedure produced the raster surfaces depicted in Figure 14 which show the travel time estimates produced for small, medium and large woodlands respectively.

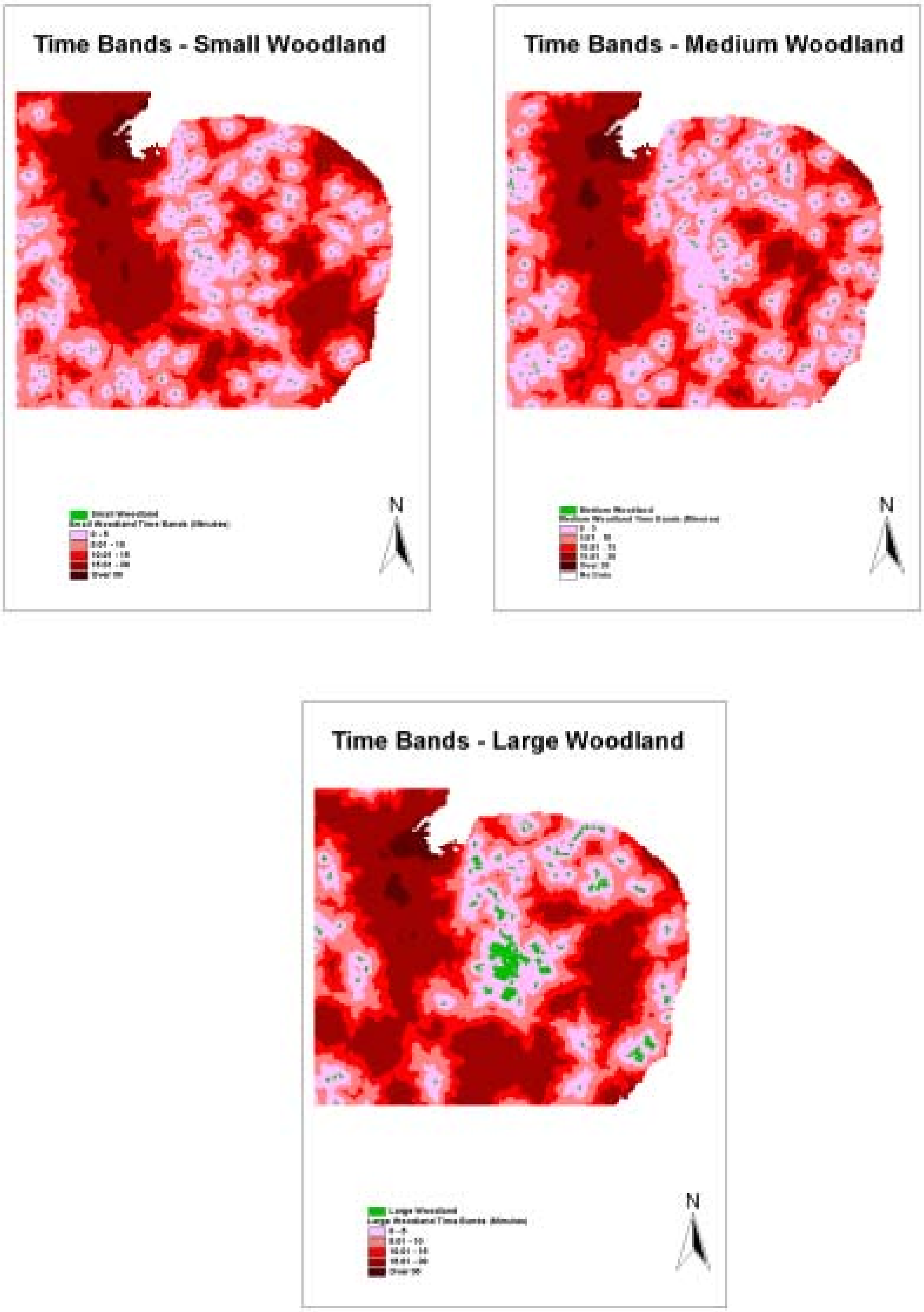


Figure 14: Time Bands – Small, Medium and Large Woodland for an Example Area in East Anglia

A gravity model was then developed to weight travel time to the different sized areas of woodland. The mean size of small, medium and large woodland polygons was extracted and the ratio of each mean to total area calculated as follows:

Small woodland mean:	314961.22m^2
Medium woodland mean:	733097.60 m^2
Large woodland mean:	5606223.70 m^2
Total woodland area:	6654282.50 m^2
Ratio of small woodland mean/total woodland:	0.05
Ratio of medium woodland/total woodland:	0.11
Ratio of large woodland/total woodland:	0.84

The calculated ratios were used as weighting values to describe the presence of each category of woodland within possible visitors perceptions of substitute availability. Area weighed substitute accessibility coverages were created for each of the small, medium and large woodland categories by multiplying each travel time accessibility coverage by its weighting value. This produced a set of 'score values' which are higher for larger as opposed to smaller woodlands thus allowing for a greater impact of the former upon substitute availability perceptions.

These values were further modified to place proportionally greater weight upon substitute features which were proportionally more proximal to visitor outset zone locations. This modification was achieved by dividing the weighted accessibility scores by the square of travel time from each outset origin (previous research by Bateman et al., (2002) having suggested that a squared power produced a good fit to observed visitor patterns). Figure 15 illustrates an example of the output from this procedure as applied to the large woodland category for East Anglia. The final area and proximity weighted substitutes accessibility surfaces were interrogated using an Arc/Info macro and the index values assigned to every ward level outset zone. Whilst the values on Figure 15 are in minutes, the weighting procedure means that they are best viewed as an index of accessibility as opposed to actual estimated travel times. These estimates were assigned to outset zones using the same ward based population weighting methodology applied to the generation of travel to site measures in Section 2.4 above.

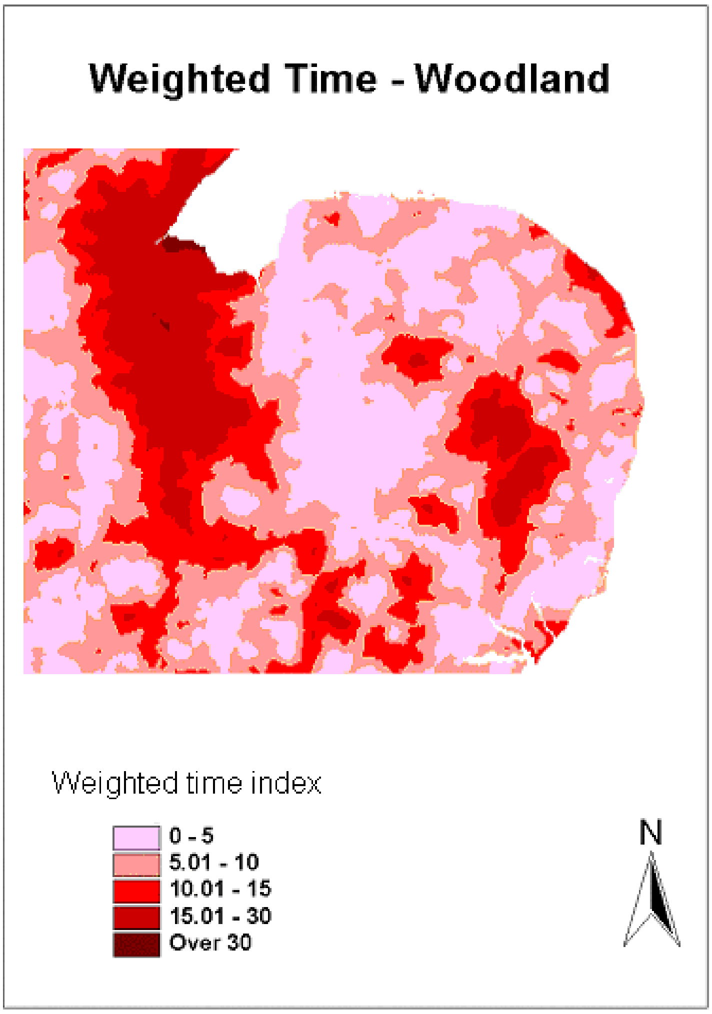


Figure 15: Area and proximity weighted substitutes accessibility surface for large woodlands in East Anglia

2.5.2 Travel Times To Other Substitutes

The methodologies described above were also used to generate substitute accessibility measures from each outset zone to all of the substitutes specified in Table 6. For those substitutes that were area based, both weighted and un-weighted indicators were calculated. For point based substitutes, where no weighting factor was available, unweighted indicators were calculated. The results of this procedure are exemplified by Figures 16 and 17 which depict area and proximity weighted substitute accessibility surfaces for Wildlife Parks/Zoos and Historic Houses/Castles respectively. The inset maps show the locations of individual substitute attractions. As before, these estimates were assigned to outset zones using the same ward based population weighting methodology that is described in Section 2.4.

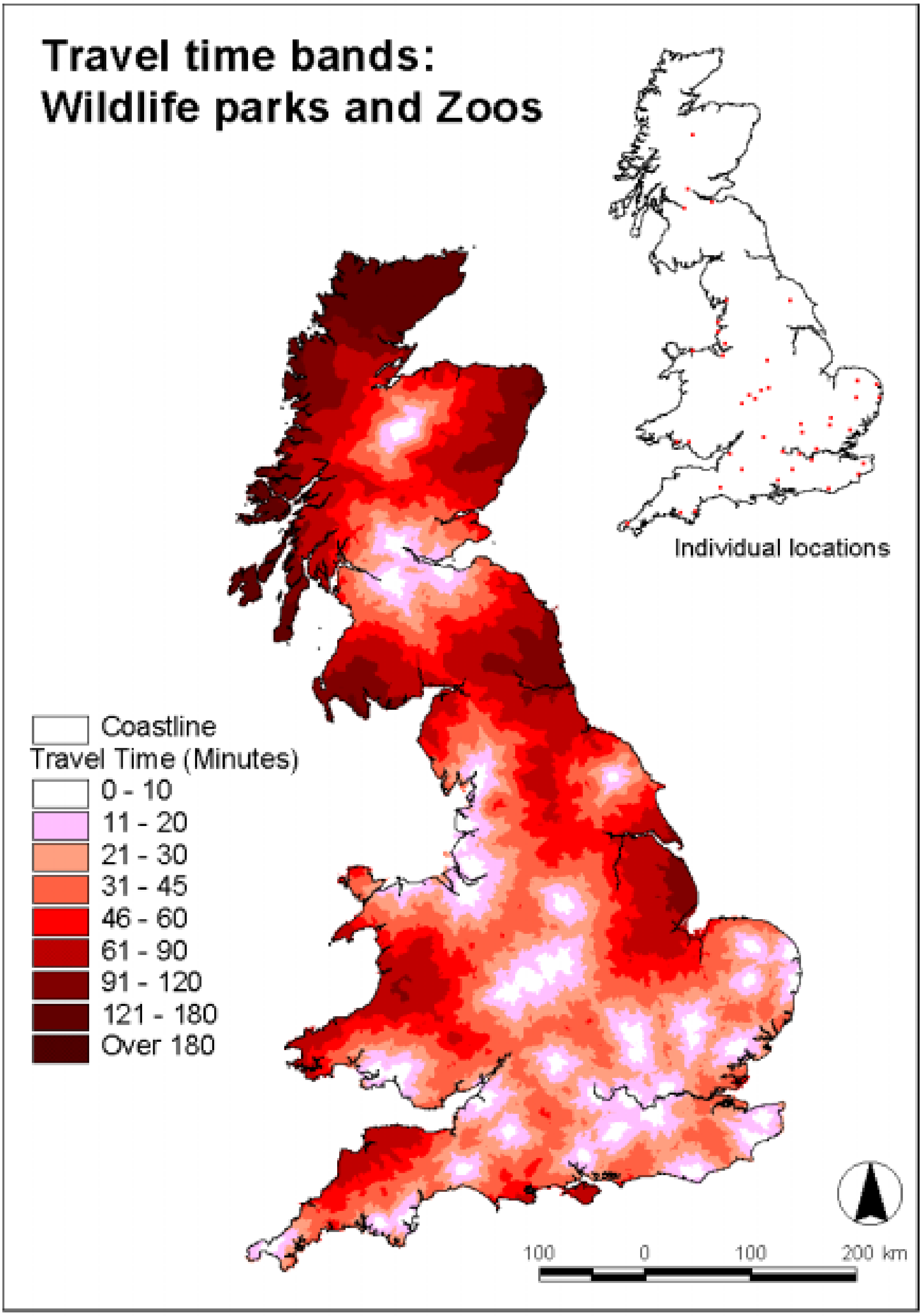


Figure 16: Area and proximity weighted substitutes accessibility surface for Wildlife Parks and Zoos

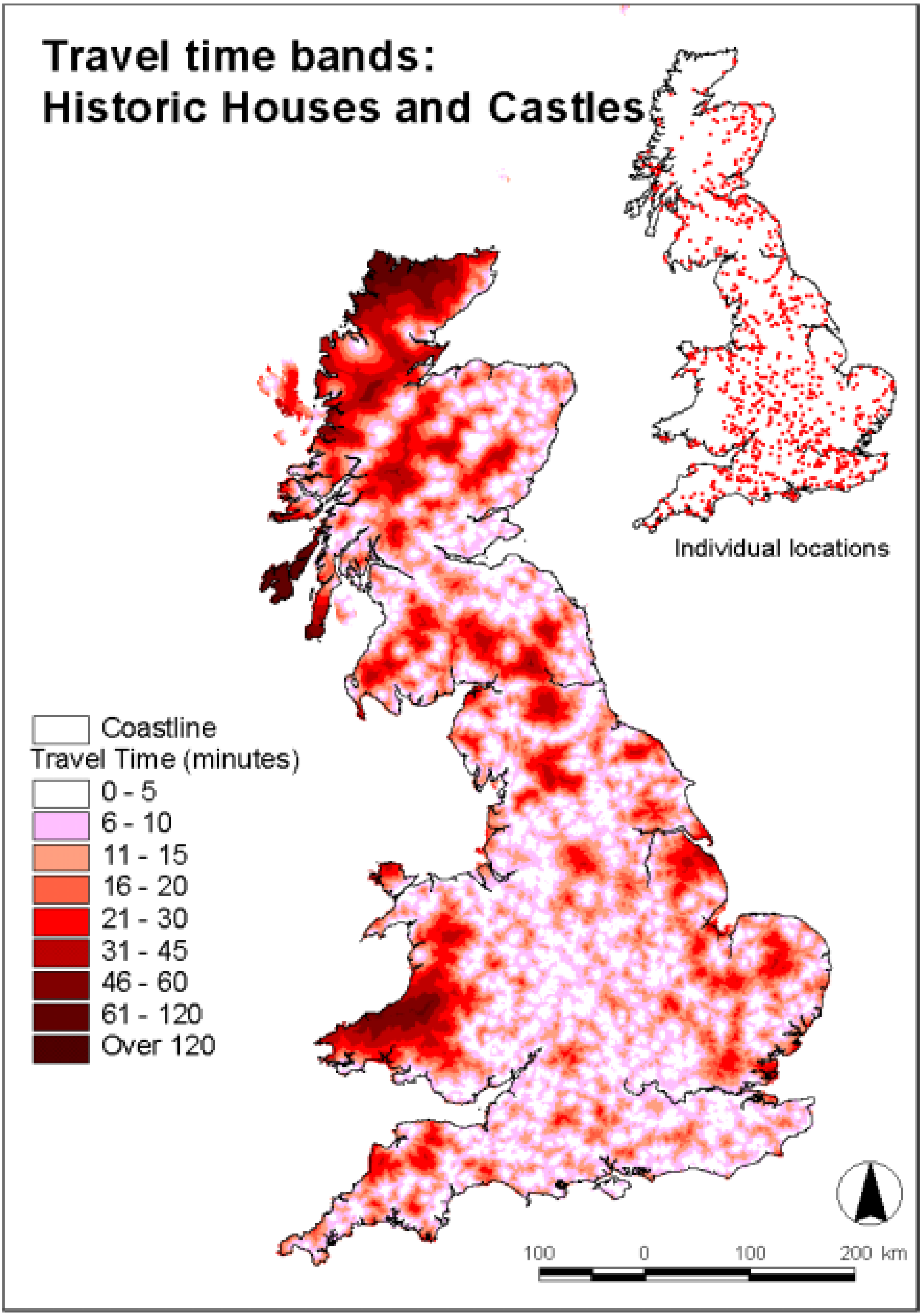


Figure 17: Area and proximity weighted substitutes accessibility surface for Historic Houses and Castles

2.5.3 Area of Substitute Provision in Outset Region

While, at a micro level, substitutes compete for visitors and thereby may be negatively associated with visits from a given population, the size of that visiting population may be positively related to the overall density of substitutes in an area. For example, an area of outstanding natural beauty encompassing a number of attractions may draw holiday visitors from some distance. Such areas may also contribute to high numbers of day trip visits from home if individuals relocate to live in such areas (Gibson, 1978; Parsons, 1991). To allow for this possibility, further measures were calculated representing the percentage of each outset zone and its surroundings covered by each substitute considered (e.g. the percentage of the surrounding area covered by woodland).

Although it would have been possible to simply calculate the percentage area of each district covered by the substitutes under consideration (an example for woodland is given in Figure 18), this measure would have been somewhat simplistic as, of course, some residents may live very close to district boundaries. Hence it is preferable to additionally consider access to substitute facilities in neighbouring districts. Therefore a procedure was developed whereby each district outset zone was amalgamated with its contiguous neighbours (contiguous districts are those which share a common boundary) using Arc/Info. For each of these attraction zones, the area of each substitute was calculated and assigned to the principal district. This was repeated for all districts within England, Wales, and Scotland. This process was undertaken for those substitutes with defined areas, such as woodland, inland water, urban areas and scenic areas. Linear features, such as rivers and canals are represented by lines in Arc/Info and have no width. Consequently it is not possible to calculate an area for these features. Hence a buffering technique was used to move the boundaries outwards to give a mean 20-metre width for all linear features (the assumption being that this gave a reasonable approximation of the real world width of these features). This provided a measure of area which enabled them to also be assessed using this method.

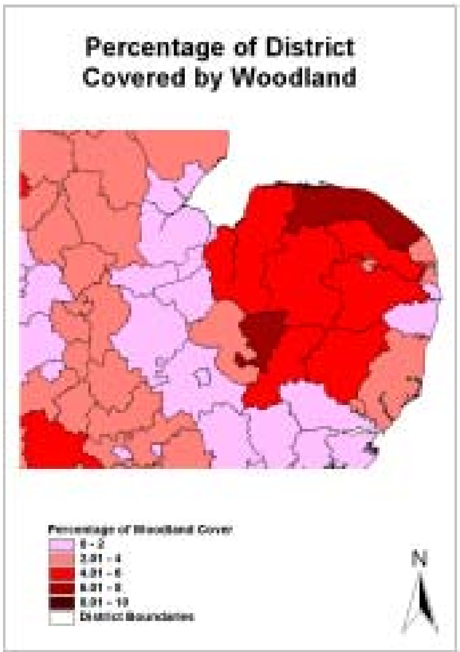


Figure 18: Percentage of District Covered by Woodland for an Example Area of Eastern England

2.6 Estimation of Population Demographic Characteristics

Variations in the demographic structure of populations of outset zones may be a further influence upon the number of visitors to Forestry Commission sites. For example, there may be variation in the propensity to seek this type of recreational experience associated with factors such as wealth, ethnicity, or household structure. Furthermore, due to the specific recreational experience provided, some sites may be inherently more or less attractive to certain population groups; children's play facilities may attract families with younger children whilst those with cycle trails may attract young adults. Hence measures were computed of the geographical distribution of population characteristics. For these, key demographic indicators that were felt may particularly influence visitor numbers were chosen.

Demographic data from the 1991 Census, at the level of local authority districts, was downloaded from the Census archive at MIMAS. In total 27 indicators (covering indicators of transport availability, affluence, deprivation, education, ethnicity, age and family size) were obtained from the 1991 UK Census of Population. Table 7 shows the complete list of demographic variables used for this research. The categories were selected in order to identify characteristics which were felt to be indicative of the propensity of individuals to visit the recreational resources being considered. For each outset zone, population density was also calculated using population and area figures downloaded from MIMAS.

Demographic Indicators
Transport Availability Indicator
Percentage of population with no car
Affluence indicators
Percentage of population of social class 1 and 2
Percentage of households owned or buying their home
Deprivation indicators
Percentage of households with over 1 occupant per room
Percentage of population of social class 4 and 5
Percentage of population with long term illness
Percentage of female population with long term illness
Percentage of lone parent households
Percentage of economically active male population unemployed
Percentage of economically active female population unemployed
Percentage of adult population in temporary accommodation
Percentage of adult population in Local Authority/Housing Authority accommodation
Higher education indicator
Percentage of population with a higher degree
Ethnicity indicators
Percentage of population black and over 16
Percentage of population Irish and over 16
Percentage of population ethnic (Black,Indian,Pakistan,Bangladesh,Chinese), over 16
Percentage of population black,indian,pakistan,Chinese, over 16
Population age indicators
Percentage of population retired
Percentage of households with head retired
Percentage of population over 16
Percentage of population over 16 male
Percentage of population under 5 years
Percentage of population under 9 years
Percentage of households with 1 dependent child
Percentage of households with 2 dependent children
Percentage of households with 3 or more dependent children
Percentage of households with no dependent children

Table 7: Demographic indicators calculated for each outset zone

2.7 Data checking and processing

The range of variables generated by the processes set out in preceding sections provides the necessary basis for estimation of the transferable demand function described in Equation {1}. The number of visits to a site can now be linked to the travel time of those visits and other predictors including the type and quality of facilities, the availability and type of substitutes and the demographic characteristics of the population of outset zone. The remainder of Section 2 describes the modelling methodology developed and applied in this study.

Once the calculation of variables was complete, all data was sorted, checked and imported into a common statistical package (SPSS) for initial analysis. Natural logarithms were calculated for variables where these provided a better fit for modelling, and the visitor survey data was split. Alongside data concerning all visitors, two additional sub-databases were prepared. One was for those visitors self-defined as holidaymakers and was for those who were on day trips from their home addresses.

Before any analysis could be undertaken, it was necessary to define the structure that the response variable would take in subsequent regression models. The general requirement was that it should take the form of an indicator of the number of visitors from each outset zone that were surveyed at each site. However, there were two issues associated with the specific definition of the variable. The first was a design concern whereby the amount of effort, measured in terms of time, that was expended to collect the original survey data varied between sites. This fact would be expected to influence the number of interviews undertaken, whereby a higher sample of responses may be anticipated at sites that were surveyed for a longer period. The second issue was that variations in the population size of each outset zone may be expected to affect the number of visitors each zone generated. Here it is anticipated that more highly populated zones would generate more visitors.

To account for variations in survey effort, the number of interviews recorded at each site from each outset zone was divided by the amount of survey effort expended at that site. Survey effort was initially measured in hours, but this figure was divided by 24 so that the variable became a measure of effort in 24 hour periods. In order to account for variations in population size, the result of the above calculation was then further divided by the outset zone population number. Hence the response variable became an indicator, over a 24 hour period, of the average number of interviews undertaken at each site with the inhabitants of each outset zone that was standardised for the size of the outset zone population. As there were 451 outset zones and 40 sites included in the models, the data matrix modelled consisted of a total of 18040 (451 multiplied by 40) observations. This was because every outset zone was represented by 40 response variables, each measuring visitors to each site.

Of course the response variable analysed here has the limitation that it reflects the number of *interviews* undertaken over the survey period rather than, necessarily, the actual number of *visitor arrivals* at each site. Furthermore, as interviews were often undertaken with a individual who was actually part of a group of visitors, the models hence predict the number of *parties* interviewed rather than *individuals*. In travel cost modelling these limitations are unavoidable as a key requirement for the inclusion of observations is that an outset location can be identified for each, and information on outset locations can only be elicited by surveying individuals. Nevertheless the reliance on interviews rather than measures of actual visitor numbers may be a problem for some sites if the number of interviews are not representative of the number of visitors attending. We investigated the impact that this issue may have in the methodology we developed to gross up our results to annual predicted visitor numbers, and we describe this work later in the report.

Initial regression analysis in SPSS revealed that the response variable, the total number of group interviews undertaken at each site divided by the survey effort (measured in hours and then divided by 24 to convert to 24 hour periods) and then further divided by the outset zone population, was not normally distributed. In fact it showed a strong right-skewed distribution. This non-linear distribution is due to the low number of visitor counts and high number of zones containing zero visitors, as illustrated for one site, Salcey, in Figure 19. This results in a distribution with a mean which is low relative to the overall range of observed visits. This generally conforms to a Poisson distribution. A form of regression analysis predicated upon this underlying distribution and capable of predicting visitor counts was therefore adopted for the modelling work.

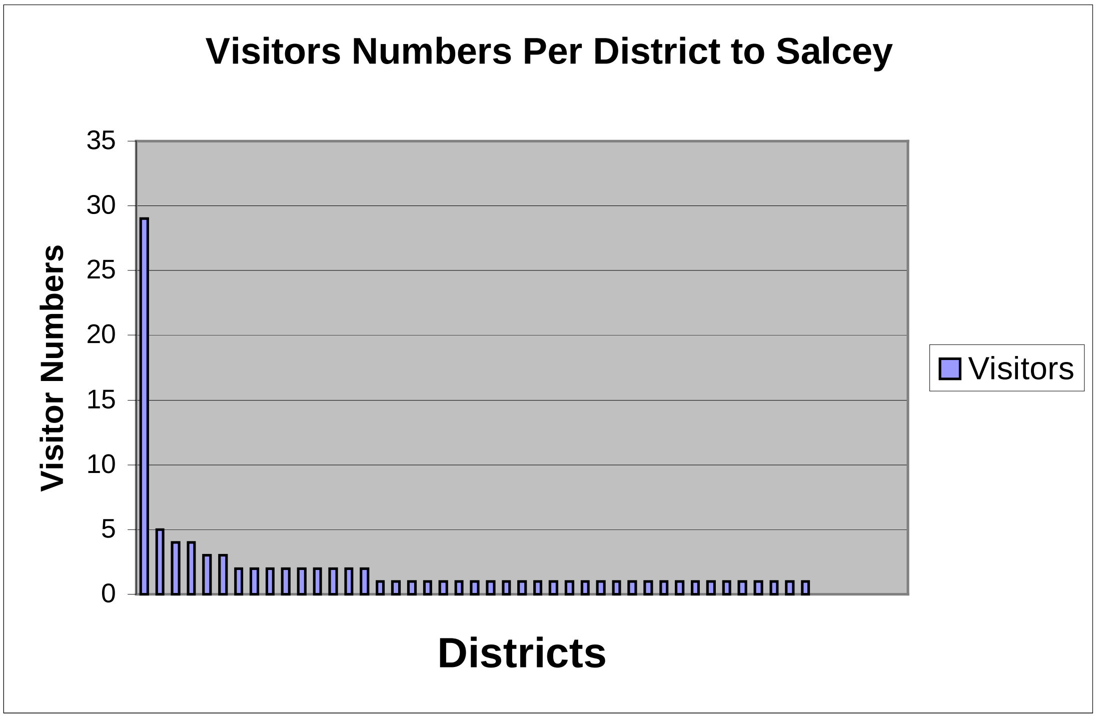


Figure 19: Frequency Distribution of Visitors to Salcey

2.7.1 The regression modelling methodologies adopted

To study variations in visitor rates between recreational sites, a statistical model which will predict visitors for an individual site based upon a range of explanatory variables is generally required. This is why the most frequently used technique is regression analysis. The simplest case is linear regression. Linear regression estimates the coefficients of a linear equation, involving one or more independent variables, that best predict the value of the dependent variable. For example, it is possible to predict a site visitor rate (number of interviews divided by survey effort and outset zone population size) from an outset zone from independent variables such as travel time from site visited, site characteristics, social class, etc.

The ordinary regression relationship for a single site visitor rate may be expressed as:

$$y_i = a + bx_i + e_i \qquad \{2\}$$

where subscript i takes values from 1 to the number of outset zones in the dataset. In Equation {2} y_i and x_i are respectively the visitor rate and travel time to site visited for the i-th outset zone, a is the intercept where the regression line meets the y axis and b is the slope coefficient. In contrast e_i is the departure of the i-th outset zone actual visitor rate to each site from the predicted visitor rate, and is known as the error term or residual. In other words the expression $a+bx_i$ forms a fixed model and e_i is that part of y_i which is not predicted by that part of the equation. For each site, the value y_i may be summed for all outset zones to produce a prediction for total visitor numbers.

As has been noted, our analysis revealed that the majority of the 453 districts provided no visitors to any given site. The resultant skewed distribution can lead to problems with the development of statistical models to predict visitor rates. Here, Ordinary Least Squares (OLS) regression models, which assume that the frequency distribution is normally distributed and the variance between districts is constant, are inappropriate. Instead, Poisson regression was used.

Poisson regression presents a convenient way of modelling a sparse visitor rate. Briefly, the Poisson model differs from OLS because it describes the probability that an event occurs t times in a fixed period, given that each occurrence is independent and has a constant probability. The assumption that the frequency of events is normally distributed with constant variance is therefore not required. The Poisson variant of the traditional OLS regression is presented in Equation {3}, where $\hat{\lambda}$ is the maximum likelihood estimate of the mean of the Poisson distributed response variable y_i (Lovett and Flowerdew, 1989) and X_0 is the outset zone population. In Poisson models the numerator and denominator elements of the response variable are separated. Hence for the models described in this report, the response variable for each outset zone was set to be the average number of surveys undertaken at each site with visitors from that zone over a 24 hour period of survey effort. The natural logarithm of the outset zone population was modelled as an offset to this. The natural logarithm of this estimate is hence equal to a linear combination of the corresponding values of the independent, x, variables. These independent variables represent the explanatory variables (such as travel times) used in this analysis to quantify and predict variations in visitor numbers between sites.

$$\hat{\lambda}_i = \exp((\beta_o X_{oi} + \beta_1 X_{1i} + (e_i X_{oi})) \qquad \{3\}$$

The popularity of regression models stems from their ease of interpretation, widespread acceptability, and the provision of suitable estimation routines in most popular statistical packages. However there can be problems with their application in certain areas of visitor prediction. One of the most fundamental of these arises when the factors influencing the probability of visitors attending any individual site are seen to be operating at a variety of scales. For example, irrespective of their measured characteristics, some sites may be more attractive to visitors than others, and hence may generate more or less interviews than would be predicted from the values of the predictor variables used to describe them. Similarly, some outset zones may also generate fewer or greater visitors than may be predicted from the model. If this is the case the assumption of independence in the residuals from the regression model is violated, and the parameters estimated may be consequently unreliable.

As described earlier, the dataset being analysed here comprises a series of response variables for each outset zone, with each corresponding to a derived measure of the number of respondents that zone generated to each site. Because both outset zone and site related characteristics may affect visitor numbers, the dataset can be conceptualised as corresponding to a hierarchical structure of outset zones nested within sites. The possible existence of such hierarchical structures within geographical data is commonly ignored. However, disregarding hierarchies where they are present can lead to the production of models giving unreliable estimates, incorrect standard errors, confidence limits, and tests (Skinner et al. 1989). Furthermore, the resultant model will present an over-simplistic picture of a complex reality, and hence may poorly predict visitor numbers (Goldstein, 1995).

In recent years a new form of statistical modelling has been developed that allows hierarchical data structures to be easily specified and their influence to be eloquently and efficiently estimated (Duncan et al., 1998). Several terms have been used to describe this new development: multilevel models (Goldstein, 1995), random coefficient models (Longford, 1993), and hierarchical linear models (Bryk and Raudenbush, 1992). Hereinafter, only the term “multilevel models” is used. Multilevel analysis has been applied in a number of fields, including education (Goldstein et al., 1993), medicine (Goldstein et al., 1994) and population geography (Jones and Duncan, 1996). It is becoming particularly popular in the field of geography since geographical analysis often involves the grouping of elementary units of interest, for example, households and individuals, into higher spatial amalgamations, such as, neighbourhoods and communities. In such a context it is important to recognise and preserve the intrinsic differences across these (Bhat, 2000).

Work we have also undertaken with an additional dataset supplied by British Waterways (parts of this analysis are discussed below) showed that the factors that are statistically significant predictors of visitor numbers are subject to considerable heterogeneity between sites. This was illustrated by the fact that, when separate regression models were fitted for each site, many of the predictor variables (with the exception of travel time) were found to be quite site-specific and did not appear in a high percentage of models. Given the similarity of these recreational resources, we have no reason to believe that this observation would be any different for the Forestry Commission data. This has obvious implications for any methodology developed to predict visitor numbers in the absence of actual survey data because it is possible that a generic model developed for all sites will be limited by the site heterogeneity. We have investigated this issue using multilevel models.

The Forestry Commission visitor dataset may be viewed as actually corresponding to a two level hierarchy of visitor rates (level 1) nested within sites (level 2) or of visitor rates (level 1) nested within districts (level 2). Hierarchical data structures cannot be easily accommodated within the traditional generalised linear estimation framework. Here, the values of site related variables must be collapsed to the level of the individual outset zone and simply replicated across all zones providing a response variable to each site. Fitting regression models which include dummy variables to indicate this may circumvent this limitation. However, it is readily apparent that any model estimated using dummies will quickly become extremely large and complex if the dataset contains many observations at each level of the hierarchy, such is the case here, as hundreds of dummy variables will be required. This complexity of model would be of little use in any attempt at function transfer. Furthermore an alternative solution of fitting separate regression models for each site would also not be viable for function transfer since there is the requirement of a single model to be used across a number of sites (we subsequently call this a meta model).

Aside from methodological considerations, a further limitation of traditional generalised linear estimation methodologies stems from the fact that non-multilevel models are likely to contain poorly estimated parameters and standard errors (Skinner et al, 1989). Problems with standard error estimation arise due to the presence of intra-unit correlation; the fact that observed visit rates to sites from a single district, or to a single site from a number of districts, may be expected to be more similar than those drawn from a random sample. If this is the case, the assumption that the residual, or error terms, produced by the model will be uncorrelated will not be met. If this intra-unit correlation is small, then reasonably good estimates of standard errors may be expected (Goldstein, 1995). However, where intra-unit correlation is large then traditionally employed estimation strategies such as weighted least squares (WLS) will tend to under-estimate the standard error, meaning that confidence intervals will be too narrow and significance tests will too often reject the null hypothesis.

The fitting of true multilevel models overcomes many of the limitations outlined above. We have also adopted this strategy for the production of meta models from our work examining British Waterways visitor survey data (Jones, et al., 2000). For algebraic simplicity, a two level hierarchy of *i* visitor rates (at level 1) within *j* sites (at level 2) is considered here to illustrate the procedure. In other words, for each outset zone we have not one visitor rate but *n* rates, where *n* corresponds to the number of sites in the model (because visitors from any outset zone can attend any site). Each of those *n* rates will correspond to a separate site, and there will be a rate for that site associated with every local authority district. Because the multilevel meta models encompassed all visitors to all sites, the problem of sparse visitor rates associated with the individual site models was not so apparent.

In the multilevel case, the visitor rates included in the model are regarded as a random sample from a population, and hence a regression relation (although not a separate regression model) is assumed for each. Considering a situation with just one explanatory variable, the proportion of visitors in each outset zone in Social Class 4&5 (*sclass*), the model may be written as:

$$y_{ij} = \beta_{0j} + \beta_1 sclass_{ij} + \in_{ij} \qquad \{4\}$$

In Equation {4} the subscript *i* takes the value from 1 to the number of visitor rates calculated for all outset zones (in the case of this dataset the value corresponds to the number of districts multiplied by the number of sites as each district has a visitor rate to associated with it for every site), and the subscript *j* takes the value from 1 to the number of sites in the sample. Using this notation, items with two subscripts *ij* vary between outset zones, where an item that has a *j* subscript only varies across sites but is constant for all outset zones potentially producing visitors to that site (and hence having a response variable). If an item has neither subscript then it is constant across both levels.

As the sites included in the analysis are treated as a random sample from a population, Equation {4} may be re-expressed as ;

$$\beta_{0j} = \beta_0 + \mu_j \qquad \{5\}$$

$$\hat{y}_{ij} = \beta_0 + \beta_1 sclass_{ij} + \mu_j$$

Where β_0 is a constant and μ_j is the departure of the *j*-th site's intercept from the overall value, which means that it is a site level (level 2) residual. Therefore this term describes, after holding constant the effect of the explanatory variables within the model, the residual influence of the site (over and above that of the predictor variables in the model) in determining the number of visitors to it. It therefore allows the identification to be undertaken of sites that for which visitor numbers are more poorly predicted than would be expected from the values of the explanatory variables.

The notations expressed in Equation {5} can be combined. Introducing an explanatory variable *cons*, which takes the value 1 for all visitor rates (and hence forms a constant or intercept term), and associating every term with an explanatory variable, we obtain Equation {6};

$$\beta_{0ij} = \beta_0 + \mu_{0j} + \in_{0ij}$$
$$y_{ij} = \beta_0 cons + \beta_1 sclass_{ij} + \mu_{0j} cons + \in_{0ij} cons \qquad \{6\}$$

In Equation {6} both μ_j (the level 2 or site level residuals) and $\in_{ij}$ (the level 1 or outset zone level residuals) are random quantities whose means are estimated to be equal to zero. It is assumed that, being at different levels, these variables are uncorrelated. Traditionally the residual error term of a non-hierarchical model, $\in$, is seen as an annoyance and the aim of the modelling process is to minimise its size. With multilevel models the error term is of pivotal importance in model estimation. Rather than a single error term being estimated it is stratified into a range of terms, each representing the residual variance present at each level of the hierarchy. Viewed in this sense, μ_j represents site level effects, whilst $\in_{ij}$ represents those that are unexplained by between site differences in model performance.

If, after holding constant the influence of the x_{ij} explanatory variables in the model, $\mu_j > \in_{ij}$, then this would suggest that some factors associated with the site itself are of greatest importance in determining the number of visitors. If $\mu_j < \in_{ij}$ then other, outset zone related, characteristics may be the most important. A common scenario is that, whilst both μ_j and $\in_{ij}$ are large in a model containing few x_{ij} explanatory variables, μ_j will decrease as explanatory variables associated with the site are added, and $\in_{ij}$ will decrease with the addition of information on individual visitors.

The model presented in Equation {6} is known as a variance components model, where the only random parameters are the intercept variances at each level (Lin, 1997). There are various methods available for parameter estimation in multilevel models (Gilks et al. 1996, Hoijtink, 1998). The least computationally intensive approach involves the use of Iterative Generalised Least Squares (IGLS). IGLS is adequate for situations where there is a large sample of responses at each level of the hierarchy, for example many visitor rates nested within many sites (as would be the case if the sample of sites was expanded and visitor rates from electoral wards were modelled instead of those from districts). IGLS was the estimation method used in the following section.

Goldstein (1995) describes the theory of IGLS in detail. Briefly, initial estimates of the fixed terms are derived by Ordinary Least Squares ignoring the higher-level random terms. The squared residuals from this initial fit are then regressed on a set of variables defining the structure of the random part to provide initial estimates of the variances/covariances. These estimates are then used to provide revised estimates of the fixed part, which is in turn used to revise the estimates of the random part, and so on until convergence. Crucially, a difficult estimation problem is decomposed into a sequence of linear regressions that can be solved efficiently and effectively, providing maximum-likelihood estimates. However, a limitation of IGLS for models with a binomial or Poisson distributed response variable is that it uses a method based on either marginal or penalised quasilikelihood. This requires assumption of normally distributed variance above level one of the hierarchy. Hence, it is assumed that the variation in model residuals between sites and districts is normally distributed in the dataset (although the assumption that the response variable is Poisson distributed remains unchanged).

It is important to note that the slopes and intercepts that are estimated for units within level 2 and above of the hierarchy will not be the same as those that would be obtained from an OLS solution; they are in-fact shrunken residuals. They have, to a greater or lesser extent, been shrunken towards the overall mean relationship. Taking a 2-level model, if $\sigma_{e0}^2 = \mathrm{var}(\in_{0ij})$ and $\sigma_{u0}^2 = \mathrm{var}(\mu_{0j})$ then each site level residual is estimated using Equation {7};

$$\hat{u}_j = \frac{n_j \sigma_{u0}^2}{n_j \sigma_{u0}^2 + \sigma_{e0}^2} \tilde{y}_j \qquad \{7\}$$

In this equation $\tilde{y}_j$ is the raw OLS residual. From Equation {7} it can be seen that if n_j is large and there are many outset zones potentially providing visitors to the site the site, then the predicted level-2 residuals will be closer in value to the raw OLS residual than when n_j is small. If n_j is small, then the residual will be shrunken towards the mean. Similarly if σ_{e0}^2 is large and there is a lot of variability in of the predicted visitor numbers to the site between outset zones, then the predicated residual will also be shrunken. In this sense, multilevel estimates can be seen as conservative estimates of variability at different levels of the hierarchy, where units based on a small sample or a very variable outcome are considered to provide little information, and are shrunken towards a mean. In the case of the models developed here the first consideration (different numbers of outset zones potentially providing visitors) is not important as the dataset is fully rectangular. However, second consideration (variability in performance in predictions between outset zones) is important.

The methodological sophistication afforded through the application of multilevel models to high quality, GIS generated variables provides a superior basis for the development of robust, transferable models of visitor arrivals. In the following section we describe the results obtained through application of these techniques.

3. RESULTS: MODELS PREDICTING THE NUMBER OF VISITORS INTERVIEWED AT SITES

In this section the results of the regression models developed for this research are given. Recollect that the aim of this exercise is to produce a series of models that may be used to predict the number of visitors interviewed at Forestry Commission sites based on a matrix of explanatory variables. These explanatory variables include indices of the accessibility of every site and potential substitute recreational resources from each outset zone, measures of the provision of facilities at each site, and indicators of the socio demographic characteristics of outset zone populations. In these models the response variable was set to be the average number of interviews undertaken in a 24 hour period (adjusted for the size of outset zone populations) and therefore we are modelling the number of visitors interviewed rather than the number of arrivals. Hence the regression coefficients should be interpreted as the effect each parameter has on determining the number of interviews with parties. The full suite of explanatory variables considered in these models, along with their definitions and value ranges, is given in Table 8 below.

Table 8: Explanatory variables used in the analysis

Variable Description	Name	Min	Max
Accessibility indicators			
Travel time to site visited	ETIMESITE	6.39	712.80
Percentage of population with no car	ENOCAR	10.97	65.92
Affluence indicators			
Percentage of households with head in Social Class 1 or 2	ESC12	9.66	45.03
Percentage of households owned or buying	EOWNBUY	16.05	88.55
Deprivation indicators			
Percentage of households with over 1 occupant per room	EHHOVER1	0.65	11.11
Percentage of households with head in Social Class 4 or 5	ESC45	3.64	39.90
Percentage of population with long term illness	ETOTILL	6.32	26.73
Percentage of female population with long term illness	EMILL	3.22	12.77
Percentage of lone parent households	ELPARENT	1.42	9.05
Percentage of economically active male population unemployed	EMUNEMP	2.74	25.33
Higher education indicator			
Percentage of population with a higher degree	EDEGREE	13.00	33.21
Ethnic indicators			
Percentage of population black and over 16	EBLACK16	0.02	20.43
Percentage of population Irish and over 16	EIRE16	0.20	11.09
Percentage of population ethnic (Black,Indian,Pakistan,Bangladesh,Chinese)	EETHNIC	0.05	40.26
Population age indicators			
Percentage of population retired	ERETIRE	10.31	37.64
Percentage of households with head retired	ERETHH	4.38	42.12
Percentage of population over 16	ETOT16	62.29	91.20
Percentage of population over 16 and male	EMALE16	39.40	61.33
Percentage of population under 5 years	EPOPU5	2.18	9.65
Percentage of population under 9 years	EPOPU9	4.21	18.50
Percentage of households with children	ECHILD	8.61	39.16

Percentage of households with no dependent children	E0CH	60.84	91.39
Coastal indicators			
Variable Description (continued)	**Name**	**Min**	**Max**
Travel time to nearest coastal region	ECOAST	0.42	75.84
Travel time to nearest sandy beach	ESAND	1.15	99.82
Water feature indicators			
Percentage of district and adjoining districts - main rivers	EPERDRIV	0.00	0.17
Percentage of district and adjoining districts - BW canals	EPERDBW	0.00	0.26
Percentage of district and adjoining districts - inland water	EPERDINW	0.00	4.37
Percentage of district and adjoining districts - all water	EPERWAT	0.00	4.43
Travel time to nearest inland water (weighted by size)	EALLINW	1.06	39.17
Travel time to nearest primary/secondary river	ERIVS	0.60	70.65
Average Travel time to BW canal	EBWCAN	0.55	144.81
Woodland indicators			
Percentage of district and adjoining districts - woodland	EPERDWD	0.00	25.22
Percentage of district and adjoining districts - forest park	EPERDFOR	0.00	30.22
Travel time to nearest woodland (weighted by size)	EALLWD	0.51	19.79
Travel time to nearest forest park	ETMFORPA	3.17	183.99
Scenic area indicators			
Percentage of district and adjoining districts - scenic areas	EPERDSCN	0.00	79.56
Percentage of district and adjoining districts - National Park	EPERDNP	0.00	66.54
Percentage of district and adjoining districts - National Trust	EPERDNT	0.00	10.68
Travel time to nearest National Park	ENATP	0.08	168.02
Travel time to nearest large heathland area	EHTHBG	0.00	79.63
Travel time to nearest scenic area (weighted by size)	EALLSC	4.25	69.30
Travel time to nearest National Trust site	ENT	2.54	134.45
Population distribution indicators			
Population density of outset zone	EPOPDEN	0.01	9.92
Percentage of district and adjoining districts - all urban	EPERDALU	0.16	99.90
Percentage of district and adjoining districts - large urban	EPERDBGU	0.00	99.90
Travel time to nearest urban area	EALLURB	0.00	70.93
Travel time to nearest large urban area (over 3500000sq m)	EBGURB	0.00	146.80
Other recreational indicators			
Travel time to nearest theme park	ETHEME	2.93	313.77
Travel time to nearest Wildlife Park/Zoo	EWILD	2.84	183.15
Travel time to nearest historic house or castle	EHSE	1.17	44.08
Site Characteristics			
Presence of a car park at site	CARPARK	0	1
Presence of a picnic area at site	PICNIC	0	1
Presence of marked walking trails at site	WALKING	0	1
Presence of marked cycle trails at site	CYCLE	0	1
Presence of marked bridleways at site	BRIDLE	0	1
Presence of orienteering course at site	ORIENT	0	1
Presence of children’s play facilities at site	CHPLAY	0	1
Presence of forest drives at site	DRIVES	0	1
Presence of viewpoints at site	VIEWS	0	1
Presence of birdwatching hides at site	BIRDS	0	1
Presence of camping / caravan facilities at site	CAMPING	0	1
Fishing allowed at site	FISHING	0	1

Presence of water body (such as river or lake) at site	WATER	0	1
Presence of bothies at site	BOTHIES	0	1
Variable Description (continued)	**Name**	**Min**	**Max**
Presence of visitor centre at site	VISITOR	0	1
Presence of interpretation point at site	INTERPET	0	1
Presence of café at site	CAFÉ	0	1
Presence of shop at site	SHOP	0	1
Presence of cycle hire facilities at site	CYCLE	0	1
Presence of classroom at site	CLASSRM	0	1
Presence of toilets at site	TOILETS	0	1
Presence of disabled toilet at site	DISTOIL	0	1
Disabled access to shop / cafe	DISHOP	0	1
Disabled walks	DISWALK	0	1

3.1. Initial models

3.1.1 Non-multilevel analysis

All of the regression modelling work presented in this report utilised multilevel modelling approaches. However, to provide a comparison with the output of those models, we also fitted an initial non-multilevel Poisson model to determine the predictors of interview numbers for all visitor types (daytrippers and holidaymakers combined) at Forestry Commission sites. The results of this exercise are shown in Table 9 below.

Table 9 shows that there are a large number of variables that exhibit a statistically significant relationship with party interview numbers. In terms of the variables included, this model is rather similar to the multilevel ones presented later in this report, although the values of the parameter estimates do differ. A full consideration of the interpretation and meaning of these indicators is therefore left until the multilevel analyses are discussed from Section 3.2 onwards. However, it is worth at this point noting from Table 9 that the measure that exhibited by far the strongest association with the response variable is the estimate of travel time to each site ($T=109.7$, $p<0.001$). The coefficient of this variable is negative, showing that visitation rates show an inverse association with estimated travel times. This observation conforms to economic theory and suggests that the principal influence on geographical variations in visitors to each site will be its accessibility (measured here in terms of travel time) from outset zone locations. Although this travel time measure dominated the model, a rather wide range of other variables are also significant. These include a number of measures of substitute accessibility (to woodlands, the coast, canals, inland water, heathlands, and National Trust properties, urban areas), and outset zone socio-economic characteristics (high social class, ethnicity, and children and retired populations). The only indicator of facility provision at sites that proved to show a statistically significant relationship with interview numbers was the presence of a visitor centre. The coefficient for this parameter was positive suggesting that sites with such a facility may attract more visitors than those without.

Table 9: Initial non-multilevel model of interview numbers (all visitor types) to a sample of Forestry Commission woodland sites across Britain.

Variable	Coefficient	SE	T value	P
Constant	-11.853	1.676	-7.07	***
Travel time to site	-2.194	0.020	109.7	***
Travel time to nearest inland water	0.358	0.042	8.53	***
Travel time to nearest heathland	0.091	0.019	4.80	***
Travel time to nearest coast	0.043	0.020	2.15	*
Travel time to nearest National Trust site	0.192	0.036	5.34	***
Percentage of outset zone district classified as Social Class 1 or 2	0.578	0.078	7.41	***
Percentage of outset zone district classified as ethnic	-0.231	0.029	-7.97	***
Percentage of outset district and surrounding districts classified as woodland	-0.043	0.012	3.58	***
Percentage of outset district and surrounding districts classified as British Waterways canals	-0.014	0.002	7.00	***
Early site visitors (7am to 10am)	-0.089	0.005	-17.80	***
Travel time to nearest large urban area	0.045	0.014	3.21	**
Presence of visitor centre at site	0.383	0.039	9.82	***
Percentage of population aged under 5 years	1.059	0.280	3.78	***
Percentage of households with retired head	0.515	0.221	2.33	*
* = p<0.05, **=p<0.01, ***=p<0.001				

As the model presented in Table 9 was based upon a Poisson distributed response variable it was not possible to calculate an R^2 value in order to assess goodness of fit. Whilst some pseudo-R^2 measures have been developed to approximate goodness-of-fit for Poisson models our experience is that they are very unreliable. Instead we assessed the goodness of fit by comparing the scaled deviance of a null model (i.e. that with no explanatory variables included) with that of the final model described above. The scaled deviance is an indicator of the variance in the response variable that is unexplained by the explanatory variables. Because of this, its value reduces as more variables are added to the model, and the magnitude of each reduction is associated with the explanatory power of each new addition. The reduction in scaled deviance observed by moving from null to final models follows the chi-squared distribution, and hence can be employed to calculate a probability based indicator of model goodness of fit. Here the null model scaled deviance was -9021 whilst that of the final model presented above was –8605. The difference was 416, which was statistically significant at well below the 0.001 value. Hence the above model gave a very good fit to the data, which is unsurprising given the strength of the observed relationship with travel times.

Unfortunately it was not possible to produce similar goodness of fit measures for the multilevel models outlined below. This is because there is still considerable uncertainty regarding the formulation of deviance statistics for non-linear (i.e. Poisson) multilevel models. This is still an area of active research (see Rasbash et al. 2000) and hence no facility for deviance calculations is provided by the MLWin software we used to fit the multilevel models. Nevertheless, our observations confirm that these models generally fit the data well, and hence it may be assumed that their goodness of fit is also highly statistically significant.

3.1.2 Comparison with models produced for British Waterways datasets

In addition to the development of a non-multilevel model to provide a comparison with later analyses, initial analyses were also guided by our experience of estimating visitor arrivals models for visitors to British Waterways (BW) recreational sites (Jones et al., 2000). This research utilised a survey dataset of 5058 valid interviews, conducted between 1997 and 2000, with visitors at 53 BW sites throughout England, Scotland and Wales. Unlike the Forestry Commission data, that supplied by British Waterways was collected during an identical eight-day period each year. While this allows comparison between years it does not provide information concerning seasonal fluctuations in visitation patterns. Furthermore, no record of people or groups of people not interviewed was recorded.

The best fitting model of visitors to BW sites combined a variety of access (travel time), substitute availability, socio-economic and site characteristic variables. As an initial investigation of the Forestry Commission dataset, an identical set of variables were applied to provide a comparative multi-site or meta-model of all visitors to woodlands (i.e. not distinguishing between day-trippers, holidaymakers, etc.). Results for this model are presented in Table 10 .

Table 10: Initial meta-model of interview numbers (all visitor types) to a sample of Forestry Commission woodland sites across Britain.

Variable	**Coefficient**	**SE**	**T value**	**P**
Constant	2.353	0.455	5.177	***
Travel time to site	-2.496	0.024	-105.94	***
Travel time to nearest historic house or castle	0.038	0.044	0.875	
Travel time to nearest wildlife park or zoo	0.231	0.031	7.477	***
Percentage of outset zone and surrounding districts classified as urban	-0.032	0.010	-3.126	***
Percentage of outset zone population aged under 5 years	-1.145	0.175	-6.558	***
Percentage of outset zone population reporting limiting long term illness	-1.373	0.100	-13.774	***
Presence of fishing facilities at site	0.966	0.461	2.094	*
Scottish site	1.081	0.367	2.945	***
σ^2_{u0}	0.918	0.208	4.413	***
* = p<0.05, **=p<0.01, ***=p<0.001				

The model described in Table 10 conforms well to prior expectations. All of the explanatory variables used were statistically significant in explaining visits to BW sites and most remain so when applied to our data on woodland visits. The offset in the Poisson regression models has been constructed so that they control for the effects of zonal population in boosting visitor numbers, thus all results have already controlled for this factor. Given this, by far and away the strongest predictor is that of travel time from the outset origin to the site. As expected, this yields a negative coefficient reflecting the fact that increasing ‘cost’ (here measured as the untransformed travel time variable rather than some inferred travel cost valuation) is associated with lower numbers of visits.

The best fit BW model contained two substitute accessibility variables, the first of which, travel time to nearest historic house or castle, proves to be the only insignificant variable when the model is applied to the Forestry Commission data. However, the other substitute variable, travel time to nearest wildlife park or zoo, proves to be highly statistically significant. The positive coefficient on this variable is as expected indicating that as distance to such substitutes increases so the number of arrivals at woodland sites also rises.

Three socio-economic variables were included in the optimal BW model and these also prove significant in predicting woodland visits. All three can be loosely described as deprivation indicators and all yield negative associations with visitation.

Two site characteristic variables prove statistically significant in both the BW and Forestry Commission models, both yield positive coefficients showing visits are higher than would otherwise be expected for sites with fishing facilities and for Scottish sites. While the former seems self explanatory the second appears to reflect higher than expected visits (especially amongst holidaymakers) at otherwise relatively remote, low visitation sites.

One of the key interests of fitting a multilevel model here was to determine if, after controlling for the explanatory variables in the model, there remained statistically significant variations in unexplained residual variance in interview numbers (as described by the u_j parameter discussed in Section 2.7.1) undertaken between sites. These random effects are summarised by the σ^2_{u0} parameter at the bottom of Table 10. Interpretation of this part of the model is relatively simple. Although, as outlined above, the multilevel methodology we used involves estimating a separate u_j value for each site, the variance present in u_j values between sites can be neatly summarised by σ^2_{u0}. This is the same parameter used in the calculation of the shrinkage factor illustrated in Equation 7. It is known as a variance parameter, as it measures the variance in u_j values estimated for each site. In other words, it shows the variability in the model residuals that may be attributed to unexplained differences between sites. For a large sample such as that used here, the statistical significance of σ^2_{u0} may be assessed by using a Wald test (Korn & Graubard, 1990). For a smaller sample (for example if there had only been far fewer sites included in the analysis), the assumption of normality between higher levels of the hierarchy may not hold true, and higher level variances are better modelled using simulation methods (Browne and Draper, 2000). In this case, from an examination of the T value (4.413) and associated probability ($p<0.001$) of σ^2_{u0}, it is clear that statistically significant residual variance in interviews is present between sites. Hence there appears to be significant variability in the performance of the model between sites, suggesting that there may be factors that make certain sites more or less attractive to visitors than may be expected based upon predictions made from the explanatory variables we have considered in the meta-model. To some extent this observation was unexpected as our other work using British Waterways data found, when separate regression models were fitted for each site, there was considerable variability observed in the predictors of visitor numbers. The nature and implications of this site level variance is discussed in more detail in the sections below.

3.2 An initial transfer exercise

As noted, the model described in Table 10 is a meta-model drawing upon information from all the sites for which we have survey data (as opposed to a single site model). As such it can, in principle, be used for transfer purposes to predict arrivals at other sites. However, given that we do not have information on any other sites we need to invoke other strategies with which to test the efficacy of this model for transfer purposes. Therefore, a series of 'omit' models are estimated. Here, for each site in turn, data from that site is dropped from the analysis and the model re-estimated, drawing upon data from all other sites (i.e. treating these as 'survey' sites and the omitted site as the 'target'). The coefficients from this exercise are then used in conjunction with information on the values of the explanatory variables at the omitted site to predict arrivals at that target site. In this manner we obtain a transferred estimate of arrivals.

Table 11 details resulting transferred estimates of arrivals at each of the 40 sites in the Forestry Commission dataset. As can be seen, they provide an excellent prediction of actual arrivals. In the table shaded cells denote Scottish sites which, as indicated above, may be statistically distinct from other sites. It is important to note that the inclusion of the multilevel residual in this set of models means that the high degree of accuracy with which the models predict the actual number of interviews undertaken needs to be interpreted with caution, as the model has a in-built correction for disparities in goodness of fit at each site. This is because these predictions were made with the site level multilevel parameter u_j included in the model. The inclusion of u_j during model development and testing of models is normal practice, as the parameter corrects for unexplained variability between sites when predictions are being made. However, as u_j is, in reality, forming part of the residuals from the model, its inclusion will lead to better fitting predictions than those that would be obtained if a coefficient for the parameter had not been estimated. This observation is important because, if the model was being used to predict the number of visitors that may attend sites for which no survey information was available, it would not be possible to estimate a u_j value. Hence in the more rigorous testing of the models, as detailed in later in this section, we did not include values for u_j in the predictions being made.

Table 11: Results from the initial transfer exercise: Predicting numbers surveyed (all visitor types). Note observed visitors surveyed are those for which valid outset locations could be determined

Site No	Site Name	Observed Visitors Surveyed	Survey Effort (24hr Days)	Predicted Vis Numbers Surveyed	Ratio Obs:Predicted
3	Afan Argoed	458	6.54	456.08	1.00
6	Alice Holt	217	3.42	223.48	1.03
8	Back O Bennachie	100	2.88	99.91	1.00
9	Beechenhurst	128	1.42	127.56	1.00
14	Black Rocks	161	3.00	163.41	1.01
15	Blackwater	179	1.46	178.37	1.00
17	Blidworth Woods	216	4.42	223.87	1.04
18	Bolderwood	343	2.42	342.30	1.00
20	Bourne Wood	211	2.48	210.38	1.00
33	Chopwell	125	1.29	125.35	1.00
34	Christchurch	132	1.00	131.29	0.99
40	Countesswells	212	2.67	215.53	1.02
43	Cycle Centre	222	3.67	222.38	1.00
44	Dalby	305	3.00	303.79	1.00
46	Delamere	684	6.38	685.00	1.00
49	Dibden	215	3.71	215.97	1.00
51	Donview	144	2.75	141.99	0.99
61	Garwnant	358	3.35	358.92	1.00
66	Glentrool	321	4.17	319.05	0.99
68	Grizedale	265	2.13	261.04	0.99
72	Hamsterley	160	2.17	157.82	0.99
80	Kielder	104	1.10	102.57	0.99
83	Kings Wood	102	3.00	103.76	1.02
84	Kirkhill	207	4.46	216.25	1.04
86	Kylerhea	210	3.96	202.49	0.96
95	Mabie	686	4.50	680.18	0.99
111	Queens View	270	4.00	265.48	0.98
117	Salcey	196	2.25	198.02	1.01
119	Sherwood Pines	680	8.69	686.33	1.01
121	Simonside Hills	136	1.90	133.38	0.98
126	Symonds Yat	255	2.75	254.46	1.00
128	Thetford High Lodge	687	6.19	682.43	0.99
129	Thieves Wood	307	4.50	312.67	1.02
130	Thrunton Woods	142	2.00	139.97	0.99
134	Tyrebagger	149	2.96	155.37	1.04
137	Waters Copse	172	3.60	173.85	1.01
141	Wendover	117	1.75	118.89	1.02
143	Westonbirt Ab	440	1.85	437.82	1.00
147	Willingham Woods	176	5.17	178.25	1.01
153	Wyre	670	5.44	671.74	1.00

3.3 Refining the meta-model: A best-fit model for woodland visits

In providing a richer picture of the source of variation within hierarchical data, the multilevel residual (u_j) estimated in models such as that described in Table 10 can be used to investigate omitted variables which can in turn be defined as predictors of arrivals within the conventional (or 'fixed') part of the meta-model. This process is graphically illustrated in Figure 20 which shows values for u_j for each site ordered from the most negative to the most positive value. Sites towards the left of this plot (with negative u_j values) are associated with fewer interviewed parties than predicted based on the values of the explanatory variables within the model, whilst sites at the right of the plot (with positive u_j values) had more interviews than predicted. For each site, comparative 95% percent confidence intervals are provided. If these confidence intervals do *not* overlap when two sites are compared, then this signifies that the performance of those sites in the model is significantly different (i.e. they appear, from a statistical sense, not to be drawn from the same populations). Conversely the performance of sites with overlapping confidence intervals in the model is statistically indiscernible. The dashed line in the centre of Figure 20 corresponds to a u_j value of zero. As a u_j value of zero for a site would be observed for one that sat at the mean of the population (i.e. the u_j residual was not significantly high or low), the line is useful. This is because the location of sites along the ranking can be examined with respect to it; sites which have confidence intervals that do not overlap with the line show statistically significant over prediction (if they are below the line) or under prediction (if they are above the line) of party interviews. It is this examination of the ranking of sites that can elucidate information on potential new explanatory variables that were not included in the original model used to generated the u_j values. This exercise revealed that Scottish sites yielded the four highest positive multilevel residuals, the most extreme of which (Kylerhea and Glentrool) are identified in the figure. Conversely the English site of Blidworth Woods is identified as having the highest negative residual. Overall analysis of trends in the multilevel residual support the argument that Scottish sites are somewhat distinct from English or Welsh woodlands. This observation gives further weight to the use of a dummy variable differentiating Scottish sites from those located elsewhere, and this variable was hence included in the models described below.

Figure 20: Rank ordered values of the multilevel residual (u_j) estimated for the initial meta-model

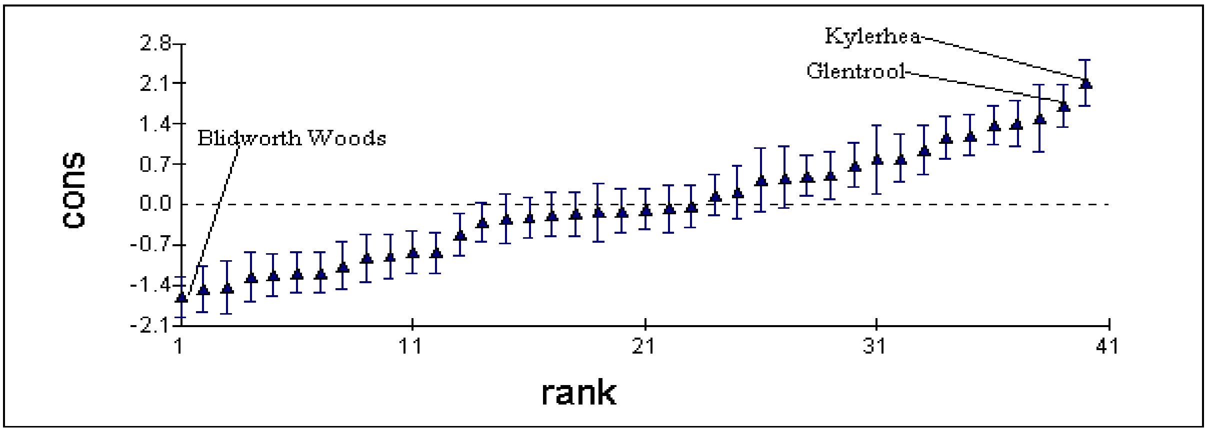


Table 12: Best-fit meta-model of interview numbers (all visitor types) to a sample of Forestry Commission woodland sites across Britain.

Two Level Site Model (All Visitors)				
Variable	**Coefficient**	**SE**	**t value**	**p**
Constant	-11.730	1.775	-6.608	***
Travel time to site	-2.563	0.026	-98.615	***
Travel time to nearest inland water	0.226	0.044	5.199	***
Travel time to nearest heathland	0.170	0.023	7.274	***
Travel time to nearest coast	0.153	0.022	6.888	***
Travel time to nearest National Trust site	0.105	0.040	2.642	**
Travel time to nearest large urban area	0.044	0.014	3.095	**
Percentage of outset district and surrounding districts classified as woodland	-0.048	0.012	-4.105	***
Percentage of outset district and surrounding districts classified as BW canals	-0.018	0.002	-9.588	***
Percentage of outset district classified as households with children	1.157	0.293	3.952	***
Percentage of outset district classified with household head retired	0.668	0.234	2.854	**
Percentage of outset district classified as Social Class 1 or 2	0.703	0.086	8.173	***
Percentage of outset district classified as ethnic	-0.109	0.029	-3.710	***
Early site visitors (7am to 10am)	-0.093	0.030	-3.082	**
Presence of Information Centre at site visited	0.640	0.273	2.341	*
Scottish site indicator	1.485	0.299	4.967	***
$\sigma^2_{u}0$	0.581	0.133	4.368	***
* 0.05 probability				
** 0.01 probability				
*** 0.001 probability				

Extensive analyses were undertaken to identify a best fitting model for visits to our sample of British woodlands. These analyses ranged across a full set of the potential explanatory variables identified in Equation {1}. Table 12 details our best-fit model.

The model described in Table 12 has expected signs on a large number of significant predictors. As before, controlling for population in each outset zone, the dominant factor determining visits is the negative influence of increasing travel time. This is modified by a number of substitute availability variables all of which indicate that arrivals at any given woodland are positively related to increases in travel time (i.e. lower accessibility) from outset locations to substitutes. Interestingly, our GIS based methodology has allowed a far larger set of substitutes to exert a significant effect upon arrivals, including a range of outdoor activity attractions (inland water, heathland, coast and National Trust sites) but also large urban centres. This suggests that many potential woodland visitors do consider both similar, natural environment, outdoor sites and manmade attractions as substitutes for woodland recreation. Further substitute relationships are expressed in the variables detailing the percentage of outset and surrounding districts which is either woodland or canalside. Here the expected relationship is again observed as increases in these figures (i.e. increases in local substitute availability) are associated with reductions in the number of visitors interviewed at any given woodland.

Continuing down Table 12 we see that, as before, a number of socio-economic and demographic variables prove to be significant predictors of the numbers interviewed. Areas which have higher levels of young children, retired or higher social classes are all associated with elevated numbers of visitor interviews. A number of potential explanations can be put forward for these results all of which seem plausible. Families with young children may well be more disposed to outdoor activities while the retired have less time constraints than others. Similarly higher social groups typically have higher incomes and so have greater mobility and ability to afford travel costs. Conversely areas with more ethnic populations yield less visitors, a result which may reflect tastes or associated lower income levels or may be a proxy for the lower accessibility of woodlands to the primarily urban ethnic community.

The 'Early site visitors' variable controls for sites where a relatively high proportion of interviews were conducted early on in the day (defined as between 07.00 and 10.00). The lower number of visitors during this time period is reflected in lower numbers of interviews being completed than would otherwise be expected. Interestingly only one of the numerous site facility and quality variables gathered proved to exert a significant impact upon the numbers interviewed (and even then the effect is relatively weak if in accord with expectations). Visitor interview numbers were higher at sites with interpretation points (labelled information centre in the above model). Taken at face value this may seem a strong assertion and it may well be that such a variable is either a proxy for other facilities or site characteristics. Alternatively this may be an endogenous effect if the decision to install a notice board may well depend on there being sufficient visitors to warrant its erection. Indeed the analysis we present later where we control for the mix of holidaymakers and day visitors at each site suggests that this latter explanation may well be true. If so its inclusion is dubious, but given the relatively weak nature of this effect it would not pose a major endogeneity problem and so is retained.

Finally, as per our findings for British Waterways canals, Scottish sites were found to be significantly and positively related to visitor interview numbers. Such sites appear to generate more recreational demand than would be expected given their other characteristics (note that this does not mean that they yield more visits in total, just that there are more than expected). One possibility may be the influence of holiday visitors boosting the potential visitor pool above that associated with the relatively sparsely distributed local population.

Overall the best-fit model seems highly satisfactory from a theoretical perspective being considerably richer than that provided by most previous research and being consistently in accordance with prior expectations derived from theory and previously observed empirical regularities.

Figure 21 details the ranked multilevel residuals (u_j values) from the best-fit model. Comparison of the scale on this graph with that for our initial model as shown in Figure 20, indicates a substantial reduction in the size of u_j across sites, as would be anticipated from the introduction of new explanatory variables.

Figure 21: Rank ordered values of the multilevel residual (u_j) estimated for the best-fit meta-model

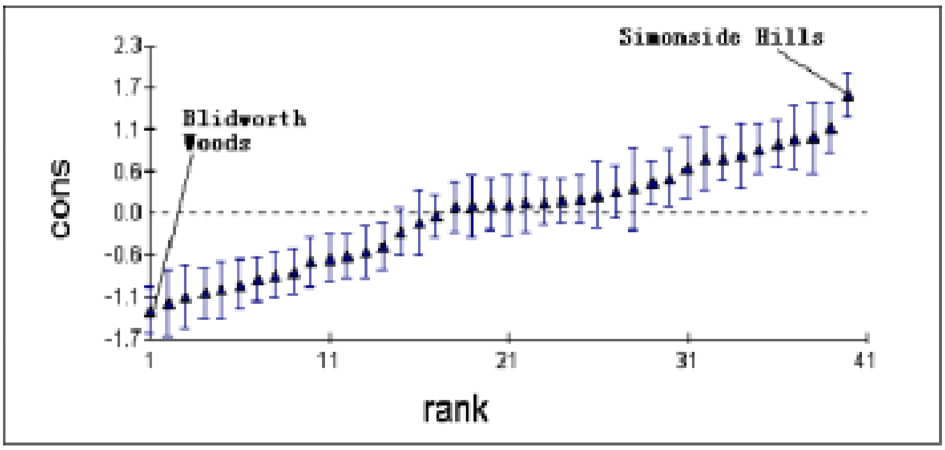


Table 13 provides further information regarding the nature of the multilevel residual estimated from our best-fit meta-model. Here sites are ordered according to residuals with names of Scottish sites shown in shaded cells. No particular trend could be discerned within the distribution of these values and together with the reduced size of residuals these findings provide further support for the best-fit nature of this model.

Table 13: Sites ranked by size of multilevel residual (u_j)estimated from best-fit meta-model

Rank	Site No	Site Name	Observed Vis Numbers Surveyed
1	17	Blidworth Woods	216
2	84	Kirkhill	207
3	134	Tyrebagger	149
4	119	Sherwood Pines	680
5	6	Alice Holt	217
6	117	Salcey	196
7	129	Thieves Wood	307
8	153	Wyre	670
9	3	Afan Argoed	458
10	18	Bolderwood	343
11	14	Black Rocks	161
12	141	Wendover	117
13	15	Blackwater	179
14	46	Delamere	684
15	83	Kings Wood	102
16	111	Queens View	270
17	33	Chopwell	125
18	147	Willingham Woods	176
19	51	Donview	144
20	9	Beechenhurst	128
21	95	Mabie	686
22	128	Thetford High Lodge	687
23	72	Hamsterley	160
24	43	Cycle Centre	222
25	61	Garwnant	358
26	8	Back O Bennachie	100
27	44	Dalby	305
28	40	Countesswells	212
29	20	Bourne Wood	211
30	137	Waters Copse	172
31	66	Glentrool	321
32	143	Westonbirt Ab	440
33	126	Symonds Yat	255
34	34	Christchurch	132
35	68	Grizedale	265
36	130	Thrunton Woods	142
37	86	Kylerhea	210
38	49	Dibden	215
39	80	Kielder	104
40	121	Simonside Hills	136

These positive attributes suggest that the best-fit model does provide a plausible basis for function transfer; to which we now turn.

3.4 Transferring the best-fit meta-model for woodland visits

In order to demonstrate the impact upon arrival estimates of including or excluding the multilevel residual within our transfer exercise, we adopt both approaches in the following analyses. Table 14 details predicted arrivals when the target site is omitted but the multilevel residuals for that site are retained. As can be seen predictions of the number of visitors interviewed are extremely close to the observed number of interviews. By contrast Table 15 repeats this exercise but excludes both the target site and its multilevel residual thus providing a more realistic assessment of the likely performance of the model in a real world policy situation from which no survey based information is known about a given target site.

Table 14: Transferred predictions of interview numbers for all visitor types from the best-fit meta-model including multilevel residuals for all sites (including target site)

Site No	Site Name	Observed Vis Numbers Surveyed	Survey Effort (24hr Days)	Predicted Vis Numbers Surveyed	Ratio Obs:Predicted
3	Afan Argoed	458	6.54	466.97	1.02
6	Alice Holt	217	3.42	223.15	1.03
8	Back O Bennachie	100	2.88	98.71	0.99
9	Beechenhurst	128	1.42	127.71	1.00
14	Black Rocks	161	3.00	164.18	1.02
15	Blackwater	179	1.46	180.33	1.01
17	Blidworth Woods	216	4.42	226.10	1.05
18	Bolderwood	343	2.42	345.77	1.01
20	Bourne Wood	211	2.48	209.19	0.99
33	Chopwell	125	1.29	125.07	1.00
34	Christchurch	132	1.00	130.64	0.99
40	Countesswells	212	2.67	210.44	0.99
43	Cycle Centre	222	3.67	220.87	0.99
44	Dalby	305	3.00	303.42	0.99
46	Delamere	684	6.38	689.02	1.01
49	Dibden	215	3.71	208.33	0.97
51	Donview	144	2.75	143.52	1.00
61	Garwnant	358	3.35	356.79	1.00
66	Glentrool	321	4.17	316.46	0.99
68	Grizedale	265	2.13	261.75	0.99
72	Hamsterley	160	2.17	159.42	1.00
80	Kielder	104	1.10	101.75	0.98
83	Kings Wood	102	3.00	103.31	1.01
84	Kirkhill	207	4.46	216.44	1.05
86	Kylerhea	210	3.96	202.94	0.97
95	Mabie	686	4.50	685.06	1.00
111	Queens View	270	4.00	270.83	1.00
117	Salcey	196	2.25	199.77	1.02
119	Sherwood Pines	680	8.69	696.35	1.02
121	Simonside Hills	136	1.90	130.66	0.96
126	Symonds Yat	255	2.75	251.43	0.99
128	Thetford High Lodge	687	6.19	685.52	1.00
129	Thieves Wood	307	4.50	314.10	1.02
130	Thrunton Woods	142	2.00	138.73	0.98
134	Tyrebagger	149	2.96	154.83	1.04
137	Waters Copse	172	3.60	169.00	0.98
141	Wendover	117	1.75	118.72	1.01
143	Westonbirt Ab	440	1.85	437.60	0.99
147	Willingham Woods	176	5.17	175.13	1.00
153	Wyre	670	5.44	677.96	1.01

Table 15: Transferred predictions of interview numbers for all visitor types from the best-fit meta-model excluding observations and multilevel residual for the target site

Predictions for each site using best fit 'all visitor' model with each site removed from the model					
Site No	**Site Name**	**Observed**	**Predicted**	**Difference**	**Ratio**
		Vis Numbers	**Vis Numbers**	**Obs-Predicted**	**Obs-Predicted**
		Surveyed	**Surveyed**		
3	Afan Argoed	458	1086.61	-628.61	2.37
6	Alice Holt	217	701.55	-484.55	3.23
8	Back O Bennachie	100	71.95	28.05	0.72
9	Beechenhurst	128	111.33	16.67	0.87
14	Black Rocks	161	303.21	-142.21	1.88
15	Blackwater	179	305.06	-126.06	1.70
17	Blidworth Woods	216	906.70	-690.70	4.20
18	Bolderwood	343	668.10	-325.10	1.95
20	Bourne Wood	211	132.35	78.65	0.63
33	Chopwell	125	94.50	30.50	0.76
34	Christchurch	132	50.38	81.62	0.38
40	Countesswells	212	121.69	90.31	0.57
43	Cycle Centre	222	180.60	41.40	0.81
44	Dalby	305	225.60	79.40	0.74
46	Delamere	684	1086.18	-402.18	1.59
49	Dibden	215	57.71	157.29	0.27
51	Donview	144	128.88	15.12	0.90
61	Garwnant	358	286.10	71.90	0.80
66	Glentrool	321	167.45	153.55	0.52
68	Grizedale	265	107.09	157.91	0.40
72	Hamsterley	160	134.23	25.77	0.84
80	Kielder	104	28.08	75.92	0.27
83	Kings Wood	102	134.84	-32.84	1.32
84	Kirkhill	207	935.21	-728.21	4.52
86	Kylerhea	210	59.65	150.35	0.28
95	Mabie	686	618.66	67.34	0.90
111	Queens View	270	307.90	-37.90	1.14
117	Salcey	196	550.64	-354.64	2.81
119	Sherwood Pines	680	2317.93	-1637.93	3.41
121	Simonside Hills	136	25.42	110.58	0.19
126	Symonds Yat	255	112.73	142.27	0.44
128	Thetford High Lodge	687	650.38	36.62	0.95
129	Thieves Wood	307	801.27	-494.27	2.61
130	Thrunton Woods	142	49.94	92.06	0.35
134	Tyrebagger	149	612.53	-463.53	4.11
137	Waters Copse	172	97.09	74.91	0.56
141	Wendover	117	211.98	-94.98	1.81
143	Westonbirt Ab	440	174.80	265.20	0.40
147	Willingham Woods	176	158.90	17.10	0.90
153	Wyre	670	1749.85	-1079.85	2.61
	Scottish Site				

Inspection of Table 15 shows that while the overall trend of results is encouraging, nevertheless there is substantial error at certain sites. While some are over predicted, others are under predicted.

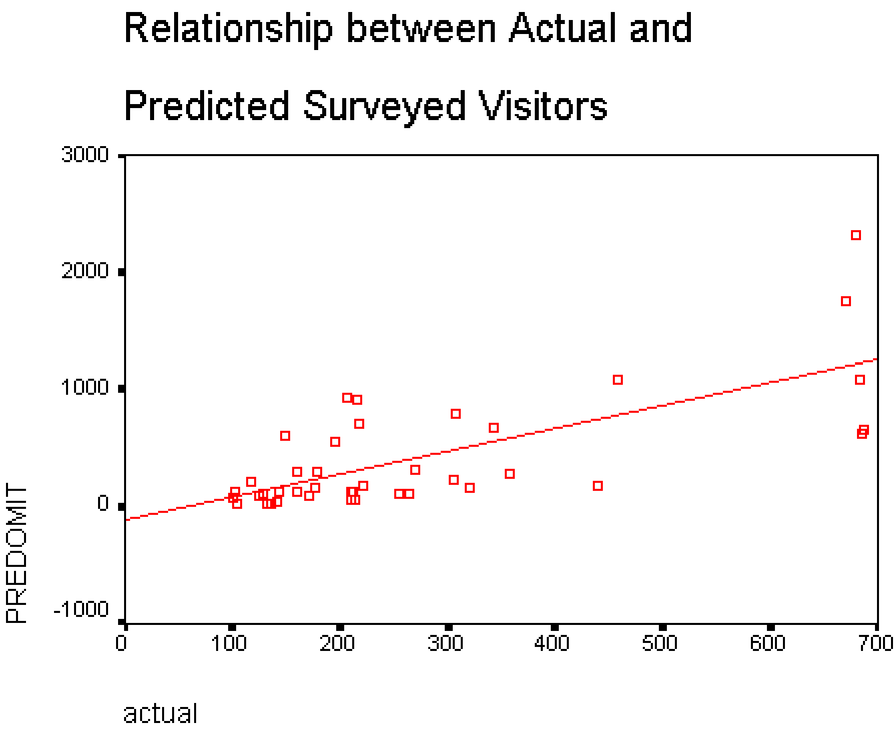


Figure 22: Plot of observed and predicted visitor interview numbers for all sample sites.

Can such a model provide useful input to the real world planning and decision making process? In order to address this issue the relationship between observed and predicted interview numbers was inspected further. Figure 22 details this relationship (with predicted interviews for each site on the vertical *Y* axis, and actual interviews on the horizontal *X* axis) and includes a regression line indicating the expected positive relation between observed and predicted values.

Inspection of Figure 22 suggests that the observed number of interviews may not entirely be an unbiased reflection of the characteristics of the site, as there seems to be some truncation of these values at roughly 700 interviews. Factors underlying this are not clear but it may be that interviewers were instructed to finish interviewing, or decided to do so of their own accord, once this level was reached. Accepting that this will militate against a clean test of our model it is nevertheless clear that our best-fit meta-model differentiates well between sites with high and low visitor interview numbers. A simple correlation test gives a values of 0.709 ($p<0.001$) suggesting that the model performs well overall. A further Chi-square test of this assertion examines the ability of the model to predict sites which are either above or below the mean number of surveys was undertaken. Results showed that the model readily satisfies such a test ($X^2 = 11.00$; $p<0.001$).

3.5 Subsample models for day-visitors and holidaymakers

As discussed previously, our sample of interviewed visitors actually consists of at least two distinct sub-samples; that of day visitors on trips from their home address; and that of holidaymakers setting out from temporary addresses. Recollect however that, for holidaymakers, no information was available on their place of stay during their holiday. Thus, for both sets of visitors, measures of travel time to each site and to substitute resources were computed based on residential locations rather than those which necessarily formed the outset location on the day in which each interview was undertaken. Table 16 details the total number of visitors interviewed and the number and proportion of these which were day trippers (from which holidaymaker numbers can be inferred).

Table 16: Total number of visitors interviewed and the number and proportion of these which were day trippers.

Rank	Site No	Site Name	Observed Total Visitors Surveyed	Observed Day Visitors Surveyed	Percentage of Total Visitors which are day
1	17	Blidworth Woods	216	211	97.69
2	119	Sherwood Pines	680	517	76.03
3	18	Bolderwood	343	148	43.15
4	15	Blackwater	179	82	45.81
5	134	Tyrebagger	149	139	93.29
6	129	Thieves Wood	307	304	99.02
7	117	Salcey	196	185	94.39
8	84	Kirkhill	207	197	95.17
9	3	Afan Argoed	458	381	83.19
10	46	Delamere	684	264	38.60
11	153	Wyre	670	567	84.63
12	141	Wendover	117	112	95.73
13	33	Chopwell	125	123	98.40
14	43	Cycle Centre	222	154	69.37
15	6	Alice Holt	217	209	96.31
16	126	Symonds Yat	255	103	40.39
17	14	Black Rocks	161	123	76.40
18	34	Christchurch	132	26	19.70
19	86	Kylerhea	210	9	4.29
20	137	Waters Copse	172	75	43.60
21	128	Thetford High Lodge	687	535	77.87
22	9	Beechenhurst	128	84	65.63
23	61	Garwnant	358	274	76.54
24	72	Hamsterley	160	119	74.38
25	143	Westonbirt Ab	440	349	79.32
26	111	Queens View	270	41	15.19
27	147	Willingham Woods	176	163	92.61
28	83	Kings Wood	102	95	93.14
29	130	Thrunton Woods	142	89	62.68
30	20	Bourne Wood	211	200	94.79
31	68	Grizedale	265	68	25.66
32	8	Back O Bennachie	100	92	92.00
33	40	Countesswells	212	209	98.58
34	80	Kielder	104	38	36.54
35	51	Donview	144	126	87.50
36	49	Dibden	215	206	95.81
37	44	Dalby	305	157	51.48
38	95	Mabie	686	355	51.75
39	66	Glentrool	321	114	35.51
40	121	Simonside Hills	136	98	72.06

Two Level Site Model Using All Visitor Model (Day Visitors)				
Variable	**Coefficient**	**SE**	**t value**	**p**
Constant	-10.060	2.541	-3.959	***
Travel time to site	-3.292	0.039	-84.065	***
Travel time to nearest inland water	0.346	0.062	5.601	***
Travel time to nearest heathland	0.152	0.037	4.087	***
Travel time to nearest coast	0.089	0.033	2.713	**
Travel time to nearest National Trust site	0.232	0.057	4.087	***
Travel time to nearest large urban area	0.093	0.024	3.906	***
Percentage of outset district and surrounding districts classified as woodland	-0.082	0.019	-4.320	***
Percentage of outset district and surrounding districts classified as BW canals	-0.022	0.003	-7.604	***
Percentage of outset district classified as households with children	1.289	0.424	3.044	***
Percentage of outset district classified with household head retired	0.656	0.338	1.938	
Percentage of outset district classified as Social Class 1 or 2	0.655	0.123	5.343	***
Percentage of outset district classified as ethnic	-0.122	0.045	-2.713	**
Early site visitors (7am to 10am)	-0.108	0.036	-3.018	***
Presence of Information Centre at site visited	0.300	0.323	0.929	
Scottish site indicator	1.160	0.365	3.182	***
σ^2_u0	0.800	0.185	4.324	***
* 0.05 probability				
** 0.01 probability				
*** 0.001 probability				

Table 17: Applying the best-fit meta-model for all visitor types to the sub-sample of day trippers (other visitors excluded)

As an initial comparison the specification of our best-fit meta-model (Table 12) was applied in turn to both the day-tripper and then holidaymaker subsamples. Table 17 details the resultant model for daytrippers.

Comparison of our models for all visitors (Table 12) with the same model applied to day trippers (Table 17) shows that they are similar in most but not all respects. A minor difference concerns the variables for information centres and retired populations, both of which are now insignificant although their signs remain unchanged. More importantly the coefficient on the key 'travel time to site' variable is substantially more negative in the model applied to day trippers (-3.292) than when applied to all interviewed visitors (-2.563). This suggests that consumer surplus values for a single day-trip visit are likely to be lower than those for the overall sample of visitors and hence even lower than those for the holidaymakers who constitute the remainder of that overall sample. This needs to be offset against the fact that day-trippers may well make more trips per annum than holidaymakers and therefore have higher annual values for the woodlands concerned. Again this accords with expectations and reflects comparisons of travel cost consumer surplus and contingent valuation willingness to pay values for woodland recreation when assessed for both day trippers and holidaymakers (Bateman et al., forthcoming).

As noted, the model given in Table 17 relies exclusively upon those variables included in our best-fit model for all visitors (Table 12) and is useful for comparative purposes. However, we also estimate a best-fit model for day-trip visitors alone the details of which are presented in Table 18.

Table 18: Best-fit meta-model for predicting the number of day trip visitors interviewed (other visitors excluded).

Two Level Site Model (Day Visitors)				
Variable	**Coefficient**	**SE**	**t value**	**p**
Constant	-7.829	1.252	-6.253	***
Travel time to site	-3.246	0.038	-85.421	***
Travel time to nearest inland water	0.340	0.061	5.574	***
Travel time to nearest National Trust site	0.175	0.058	3.017	***
Travel time to nearest Wildlife Park or Zoo	0.341	0.055	6.200	***
Travel time to nearest large urban area	0.076	0.023	3.304	***
Travel time to nearest coast	0.099	0.033	3.016	***
Percentage of outset district and surrounding districts classified as National Park	-0.013	0.002	-6.031	***
Percentage of outset district and surrounding districts classified as BW canals	-0.015	0.003	-5.084	***
Percentage of outset district and surrounding districts classified as woodland	-0.062	0.021	-2.925	***
Percentage of outset district population classified as Social Class 1 or 2	0.781	0.122	6.417	***
Percentage of outset district population classified as ethnic	-0.228	0.048	-4.757	***
Percentage of outset district population classified as under 9 years	1.184	0.353	3.354	***
Presence of Toilet at Site Visited	0.691	0.334	2.069	*
Early site visitors (7am to 11am)	-0.068	0.024	-2.833	**
$\sigma^2_u 0$	0.751	0.173	4.341	***
* 0.05 probability				
** 0.01 probability				
*** 0.001 probability				

Consideration of Table 18 shows it to be very similar to that given in Table 17. The one major difference is that the Scottish site indicator has now proved insignificant. This suggests that the variable is important in explaining holidaymaker visits but not those of day trippers (i.e. it is not reflecting a cultural difference between those who live in Scotland and elsewhere regarding their attitudes to woodland recreation).

Turning to consider the sub-sample of holidaymakers, estimated models identified a considerably shorter list of significant predictors than applied for either the all visitor or day tripper samples, a finding which may well reflect the considerably lower number of holidaymakers observations collected. Table 19 below details the best fitting model for holidaymakers. As expected, the travel time coefficient is substantially less negative than for the all sample model, the difference being even greater when compared to the day-tripper model. The reduced absolute magnitude of this coefficient is associated with higher per visit consumer surplus values, but as discussed, the lower number of visits made by holidaymakers means that their annual consumer surplus value for all visits made is likely to be smaller than that for day trippers. Coefficients on all other explanatory variables have consistent signs with previous models.

Table 19: Best-fit meta-model for predicting the number of holidaymaker visitors interviewed (other visitors excluded).

Variable	**Coefficient**	**SE**	**T value**	**P**
Constant	-13.604	0.656	-20.74	***
Travel time to site	-1.065	0.052	-20.48	***
Travel time to nearest inland water	0.205	0.069	2.97	**
Travel time to nearest large urban area	0.054	0.011	4.91	***
Percentage of outset district and surrounding districts classified as British Waterways canals	-0.009	0.003	-3.00	***
Travel time to nearest woodland	0.189	0.068	2.78	**
Percentage of outset zone district classified as Social Class 1 or 2	0.786	0.111	7.08	***
Percentage of outset zone district classified as Forest Park	0.191	0.058	3.30	***
Presence of visitor information centre at site	1.155	0.442	2.61	**
Early site visitors (7am to 1pm)	-0.134	0.048	-2.79	**
σ^2_{u0}	1.299	0.320	4.05	***
* = p<0.05, **=p<0.01, ***=p<0.00				

Figure 23 illustrates residuals from the above model while Table 20 ranks sites by residual value (ranked from most negative to most positive). Inspection of Table 20 revealed an interesting trend; it appears that those sites that have the largest number of holidaymakers visiting them also tend to have the most positive residuals. As the presence of a positive residual for a site signifies that more interviews have been undertaken than would be predicted based on the values of the regression coefficients, this observation suggests that the model is tending to under-predict visitor numbers at sites that appear, based on the observed data, to particularly appeal to holidaymakers. To examine this issue further, a new explanatory variable was created. This was set to be the percentage of total numbers of interviews at each site that were undertaken with holidaymakers, as opposed to day trippers. The model was re-fitted with this new explanatory variable included. This gave the results detailed in Table 21.

Figure 23: Rank ordered values of the multilevel residual (u_j) estimated for the best-fit meta-model for holidaymakers

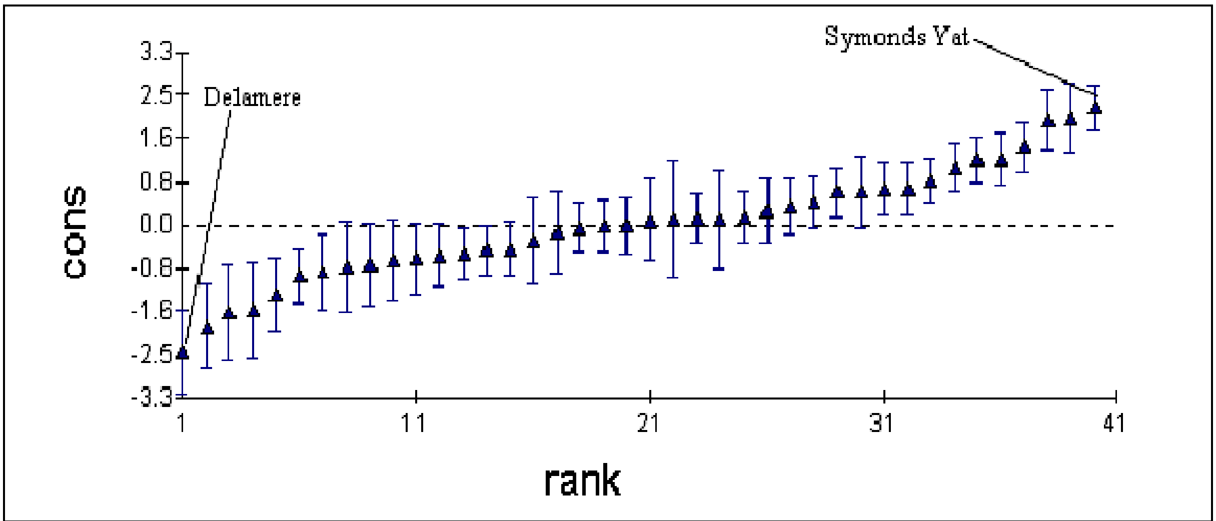


Table 20: Sites ordered from most negative to most positive values of the multilevel residual (u_j) estimated for the best-fit meta-model for holidaymakers

Rank	Site No	Site Name	Observed Total Visitors Surveyed	Observed Hol Visitors Surveyed	Percentage of Total Visitors which are hol
1	46	Delamere	684	6	0.88
2	6	Alice Holt	217	6	2.76
3	129	Thieves Wood	307	2	0.65
4	17	Blidworth Woods	216	2	0.93
5	117	Salcey	196	9	4.59
6	3	Afan Argoed	458	76	16.59
7	141	Wendover	117	5	4.27
8	33	Chopwell	125	2	1.60
9	83	Kings Wood	102	8	7.84
10	147	Willingham Woods	176	12	6.82
11	20	Bourne Wood	211	11	5.21
12	51	Donview	144	18	12.50
13	153	Wyre	670	101	15.07
14	72	Hamsterley	160	40	25.00
15	119	Sherwood Pines	680	163	23.97
16	84	Kirkhill	207	10	4.83
17	8	Back O Bennachie	100	8	8.00
18	15	Blackwater	179	93	51.96
19	44	Dalby	305	148	48.52
20	14	Black Rocks	161	38	23.60
21	134	Tyrebagger	149	9	6.04
22	40	Countesswells	212	3	1.42
23	18	Bolderwood	343	194	56.56
24	49	Dibden	215	9	4.19
25	61	Garwnant	358	83	23.18
26	128	Thetford High Lodge	687	149	21.69
27	9	Beechenhurst	128	44	34.38
28	80	Kielder	104	64	61.54
29	68	Grizedale	265	197	74.34
30	143	Westonbirt Ab	440	86	19.55
31	111	Queens View	270	228	84.44
32	43	Cycle Centre	222	67	30.18
33	95	Mabie	686	315	45.92
34	86	Kylerhea	210	201	95.71
35	66	Glentrool	321	205	63.86
36	130	Thrunton Woods	142	52	36.62
37	121	Simonside Hills	136	37	27.21
38	137	Waters Copse	172	97	56.40
39	34	Christchurch	132	104	78.79
40	126	Symonds Yat	255	152	59.61

Table 21: Best-fit meta-model for predicting the number of holidaymaker visitors interviewed (other visitors excluded) including measure of percentage of interviews undertaken at each site that are with holidaymakers.

Variable	**Coefficient**	**SE**	**T value**	**P**
Constant	-14.707	0.525	-28.07	***
Travel time to site	-1.067	0.052	-20.52	***
Travel time to nearest inland water	0.207	0.069	3.00	***
Travel time to nearest large urban area	0.054	0.011	4.91	***
Percentage of outset district and surrounding districts classified as British Waterways canals	-0.009	0.003	-3.00	***
Travel time to nearest woodland	0.189	0.068	2.78	**
Percentage of outset zone district classified as Social Class 1 or 2	0.783	0.111	7.05	***
Percentage of outset zone district classified as Forest Park	0.190	0.058	3.27	***
Presence of visitor information centre at site	0.195	0.228	0.85	
Early site visitors (7am to 1pm)	-0.073	0.024	-3.05	***
Percentage of visitors who are holidaymakers	0.042	0.004	10.50	***
σ_{u0}^2	0.255	0.077	3.31	***
* = p<0.05, **=p<0.01, ***=p<0.00				

A comparison between Tables 19 (model without percentage of holidaymakers included) and 21 (the model with this new variable added) reveals a number of interesting differences. Firstly, it is apparent from Table 21 that, as anticipated from the examination of rankings in Table 20, the indicator of holidaymakers is highly statistically significant, with a positive coefficient of 0.042 and a T value of 10.50 (p<0.001). The positive sign on this new indicator confirms that, holding constant the effect of the other variables in the model, sites which have a higher proportion of holidaymakers visiting them tend to be associated with a higher number of interviews. This suggests that they may be particularly attractive to visitors.

A second observation made from this new model is that, whilst the presence of a visitor information centre at a site was found to be a statistically significant predictor of the number of interviews undertaken in Table 19, this is now no longer the case (in the above table p=0.85). This suggests that those sites that have this facility may also tend to be those that attract holidaymakers. Indeed a comparison between sites with and without a visitor information centre of the percentage of interviews that were with holidaymakers shows this to be the case; at sites with a visitor information centre an average of 42.9% of interviews were with holidaymakers, whilst at sites without the facility only 15.8% of interviews were with holidaymakers. Hence in the earlier model it may not be the presence of an information centre that is attracting visitors, but rather that such features tend to be provided at sites which already attract large numbers of holidaymakers.

A third observation when the models in Tables 19 and 21 are compared is that, although still statistically significant, the size of the variance of the site-level multilevel residuals has reduced from 1.299 to 0.255 after the introduction of the new variable. Given that the indicator of holidaymaker numbers is itself highly statistically significant this observation is unsurprising, as it may be expected that at least some of the previously observed variations in model performance between sites may be due to the differing number of holidaymakers visiting each.

The finding that suggests that sites that attract more holiday visitors than the previous model predicts also tend to attract more visitors in general is an interesting one. Of course, the new variable measuring the proportion of interviews at each site that are undertaken with holidaymakers does not in itself explain the reasons why these sites appear to be particularly attractive to this group of visitors, and the precise reasons for this require further research. It is likely that some, as yet unmeasured, characteristics of the sites are pertinent. For example it may be that they tend to be located in areas that are particularly popular holiday destinations and hence have a propensity to attract more visitors because there are many people staying in the vicinity. Alternatively it may be that these sites have some facility, unmeasured by us, then tends to attract holidaymakers to them. Whilst these possibilities suggest that further work should be undertaken, this line of enquiry is not followed further in the work presented here. This is because, from a benefit transfer point of view, information would not be available on the percentage of holiday makers visiting unsurveyed sites, and hence this indicator cannot be included in transfer models.

Table 22 below details transferred predictions of the number of holidaymakers interviewed from each site obtained from applying the best-fit meta-model as detailed in Table 19. This is calculated as per Table 18 (for day-trippers) by excluding observations for target sites but retaining this information for the purposes of estimating the multilevel residual. Comparisons between these results and those presented previously shows that, as we may expect, the holidaymaker transfer model performs relatively poorly for sites with very few (say, less than 10) holiday maker interviews.

Table 22: Transferred predictions of holidaymaker interview numbers from the best-fit meta-model including multilevel residuals for all sites (including target site)

Site No	Site Name	Observed No. Hol Visitors Surveyed	Survey Effort (24hr Days)	Predicted Vis Numbers Surveyed	Ratio Obs:Predict
3	Afan Argoed	76	6.54	81.14	1.07
6	Alice Holt	6	3.42	11.34	1.89
8	Back O Bennachie	8	2.88	8.34	1.04
9	Beechenhurst	43	1.42	42.61	0.99
14	Black Rocks	38	3.00	38.03	1.00
15	Blackwater	93	1.46	93.05	1.00
17	Blidworth Woods	2	4.42	7.77	3.89
18	Bolderwood	194	2.42	193.78	1.00
20	Bourne Wood	11	2.48	12.30	1.12
33	Chopwell	2	1.29	2.84	1.42
34	Christchurch	104	1.00	102.37	0.98
40	Countesswells	3	2.67	2.76	0.92
43	Cycle Centre	67	3.67	65.00	0.97
44	Dalby	148	3.00	148.09	1.00
46	Delamere	6	6.38	18.37	3.06
49	Dibden	9	3.71	8.66	0.96
51	Donview	18	2.75	19.31	1.07
61	Garwnant	83	3.35	82.65	1.00
66	Glentrool	205	4.17	200.87	0.98
68	Grizedale	197	2.13	195.94	0.99
72	Hamsterley	40	2.17	40.82	1.02
80	Kielder	64	1.10	63.63	0.99
83	Kings Wood	7	3.00	8.79	1.26
84	Kirkhill	10	4.46	11.09	1.11
86	Kylerhea	200	3.96	196.50	0.98
95	Mabie	315	4.50	311.96	0.99
111	Queens View	228	4.00	225.83	0.99
117	Salcey	9	2.25	11.40	1.27
119	Sherwood Pines	163	8.69	166.19	1.02
121	Simonside Hills	37	1.90	34.74	0.94
126	Symonds Yat	152	2.75	147.09	0.97
128	Thetford High Lodge	149	6.19	147.71	0.99
129	Thieves Wood	2	4.50	7.98	3.99
130	Thrunton Woods	52	2.00	49.97	0.96
134	Tyrebagger	9	2.96	8.80	0.98
137	Waters Copse	97	3.60	91.21	0.94
141	Wendover	5	1.75	6.27	1.25
143	Westonbirt Ab	86	1.85	85.07	0.99
147	Willingham Woods	12	5.17	14.78	1.23
153	Wyre	101	5.44	103.47	1.02

4. CONVERTING FROM PREDICTING THE NUMBER OF VISITORS INTERVIEWED AT SITES TO PREDICTING ANNUAL VISITS

The models detailed above focus upon the prediction of the number of visitors interviewed at each site (adjusted to a standard unit of survey time). However, for decision making purposes this needs to be converted to some estimate of total visitation numbers within some specified period which, for the purposes of this research, is taken to be one year.

Such conversion is in some respects more difficult than the modelling exercise undertaken above, not because of its intrinsic complexity, but because of a lack of accurate information concerning the relationship between the numbers interviewed and annual arrivals. Indeed the accuracy of estimates of the annual number of visitors to sites is itself an issue of some concern. Furthermore, as we have shown elsewhere (Bateman et al., 2002), errors within the aggregation process can induce far greater variability in estimates of total demand than do errors within the modelling exercise outline above.

In order to facilitate testing of any predictions of annual arrivals, the Forestry Commission supplied their own estimates for a subset of five sites, included within our preceding analysis, for which they held annual arrivals data. This data is summarised in Table 23 however it should be emphasised that, while these are best estimates, the Forestry Commission recognises that they may be subject to significant error due to problems with the means of data collection available (traffic counters, etc.). Note that these estimates, like those modelled in the preceding section, refer to party visits rather than necessarily those made by individuals. However, conversion from the former to the latter is a relatively trivial task providing that information on party size is held at a site level.

Table 23: Forestry Commission estimates of annual party visits at five sites.

Site number	Site name	Total number of party visits per annum
9	Beechenhurst	72,845
17	Blidworth Woods	63,849
33	Chopwell	33,708
95	Mabie	51,704
126	Symonds Yat	77,525
	Total:-	299,631

Remember that the figures given in Table 23 are in terms of party visits. This accords with the units used in our modelling exercise and therefore no mismatch occurs here (although readers should remember that these values do not necessarily equate with those of individual visitors).

The simplest method of converting the number of predicted interviews to the number of annual visits is to simply multiply the predictions obtained from the main meta model described in Table 12 by a constant value. It is important to note that the models described in the previous sections of this report predict the number of interviews undertaken at each site during a period which has been set at 24 hours. However, this period does not correspond to a calendar day as visits are not uniformly spread across a 24 hour period. The Forestry Commission have estimated that the average number passing during one interview hour is likely to be around 1/8th of the daily total (*pers comm.*, Simon Gillam 16 April 2002).

Therefore, in order to scale the model predictions to annual numbers, the predicted number of interviews in each 24 hour period was divided by 24 (to give an hourly prediction) and then multiplied by 8 (to scale up to a daily value). The result was then further multiplied by 365.25 in order to give an annual prediction.

A further complication arises whereby the number of interviews included in the model (and hence predicted) at each site does not necessarily equate to the original number of interviews undertaken. This is because it was necessary to exclude from the modelling process interviews where no outset location could be determined. For example, at Beechenhurst, only 58% of interviews were associated with a valid postcode. Hence an additional correction was undertaken whereby the predicted number of visits at each site was scaled upwards in order to account for these non-modelled visitors.

An further issue concerned the relationship between interview effort, the number of completed interviews and the number of visitors during the interview period, as an interviewer can only interview one party at a time. This consideration is important as the above models are based on the number of interviews undertaken at sites rather than, necessarily, the actual number of visitors. Therefore allowance had to be made for those who entered the site while others were being interviewed and therefore could not be interviewed themselves. Information regarding the necessary adjustment to allow for this factor was provided by the Forestry Commission This consisted of an estimate of the number of groups who had visited each site, but were not interviewed, in the period during which surveyors were present. This value used to calculate a further scaling factor for each site, which was set to be the ratio of unsurveyed to surveyed groups, and predictions made for each site were then multiplied by this ratio (The ratio of visting parties not interviewed to valid interviews was 0.69 for Beechenhurst, 1.62 for Blidworth Woods, 0.73 for Chopwell, 0.16 for Mabie, and 1.43 for Symonds Yat). This scaled up predictions of interview numbers to provide estimates of the corresponding number of arrivals during that period. The results of this entire exercise are given in Table 24 below.

Table 24: Comparison of Forestry Commission and model estimates of annual visits to five woodlands: Multiplication by 365.25.

Site name	Forestry Commission estimate of the number of party visits per annum.	Predicted number of party visits per annum based on upwards scaling (see text)
Beechenhurst	72845	22024
Blidworth Woods	63849	78532
Chopwell	33708	17854
Mabie	51704	21204
Symonds Yat	77525	16267
Total:-	299631	155881

A comparison of the estimated number of party visits provided by the Forestry Commission with those produced from the model show that the predictions of party numbers for all sites agree within one order of magnitude. However, there are relatively large discrepancies between predicted values obtained from this method and those values provided by the Forestry Commission. Across all five sites, the Forestry Commission estimates suggest there may be 299631 annual party visits whilst the results from our analysis scaled in the manner described above give a lower figure of 155881 visits. This provides a ratio of our own to Forestry

Commission predictions of 0.52. Given that there may be considerable uncertainty associated with the Forestry Commission estimates themselves this degree of agreement is not entirely unexpected. However, by simply scaling a 24 hour based model in the manner described above, a number of factors that may be important were ignored. In particular, using this method it was not possible to account the fact that the surveys at each site were undertaken at different times of the year, and hence it may be expected that those undertaken during the summer would provide more interviews due to anticipated seasonality in visit rates. This issue could explain the general tendency of our model to under-predict compared to Forestry Commission estimates, and are addressed in the subsequent analyses described below. This is a potentially important issue as we know that the distribution of total visits is not uniform across the year. Therefore adjustment had to be made to allow for the effect of surveys being undertaken in more or less popular times of the year.

Two approaches were developed to allow for the distribution of visits across the year within estimates of annual arrivals. The first approach drew upon information derived from the 1998 UK Day Visit Survey (UKDVS), (National Centre for Social Research, 1999) details of which are provided in Table 25. It is important to note that one limitation of the UKDVS that it only covers visits from home. This partly explains the low figure for August when a higher proportion of visits are made from holiday bases.

Table 25: Distribution of day visits to woodland by month

Proportion of Visitors to Woodland According to Day Visitors Survey	
Month	**Percent**
Jan	10
Feb	5
Mar	9
Apr	9
May	7
Jun	15
Jul	11
Aug	7
Sep	7
Oct	7
Nov	5
Dec	6
Total	98
Missing	2
Total	100

Source: UK Day Visit Survey

The information given in Table 25 shows that visits peak during the months of June and July but are fairly uniform during other months with a further peak during January. We were somewhat surprised by this information as it did not accord with that used by Bateman (1996) in an analysis of weekly arrivals over a two year period at a site in Thetford Forest. In the latter study a much less bimodal distribution was observed, devoid of the January peak seen in Table 25. Furthermore, while the distribution given in the UK Day Visits Survey is relatively flat across the months of February, March, April, May, August, September, October, November and December. In contrast the study by Bateman (1996) shows a clear upward trend as the seasons pass from Winter to Summer and a decline from Summer to Winter (with significant impacts from unusually adverse weather within each season and bank holiday effects). Given these contrasting results, we describe the work and results of Bateman (1996) in Appendix A of this report.

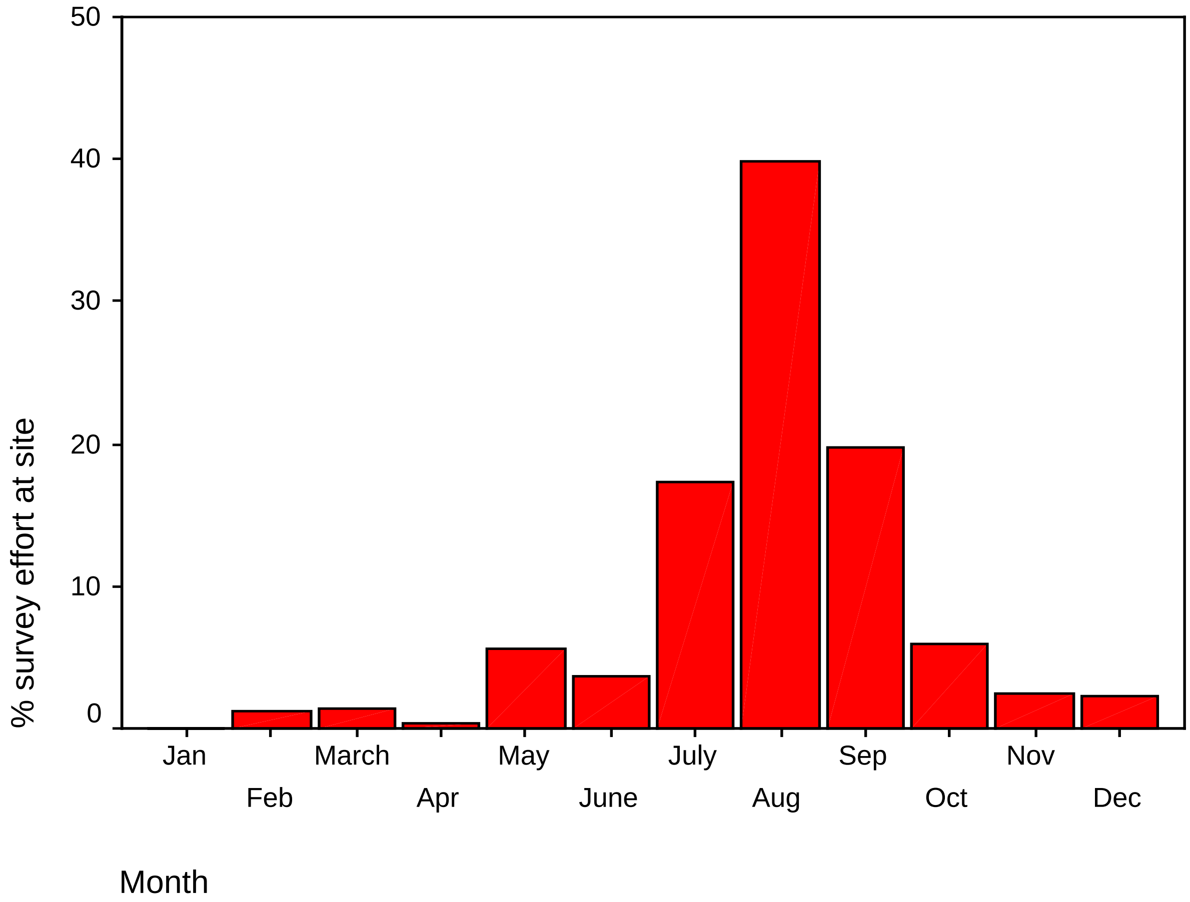
Given the above observation, which may be associated with the fact that the survey only measures visits from home, we have some concerns about using the UK Day Visit Survey data as the basis for any conversion from per day to per annum predictions of arrivals at woodlands. Nevertheless, this approach was tested as follows. For each site, note was taken of the proportion of total survey effort which occurred in each month. A methodology was developed whereby for the months surveyed the estimate of the number of visitors in the survey period was converted to a monthly value. Based on these monthly value estimates the number of visitors at other times of the year when no survey had taken place was calculated. Scaling was undertaken according to the extent to which these estimates were above or below average in terms of visitor numbers according to the UK Leisure Day Visit survey. Applying this scaling factor to the prediction of arrivals during a standard day period takes account of the uplift exerted by surveying in high-visitation months (or the drag exerted by surveying in low-visitation months). A simple example illustrates the procedure. First consider the percentage of visitors which would arrive at forest locations each month if there was no monthly variation in visitation levels. This value would be 100/12 = 8.3%. In reality, according to the UK Day Visit Survey, 11% of annual visits to forests occur in July. Now assuming that all interviews at a given site were conducted in July and that, based upon the model predictions, some 200 visitor parties arrive at this site per day, this estimate would be elevated relative to the annual average number of visits because the survey was conducted at the most popular time of the year. Multiplying the estimate of 200 visitor parties by the number of days in July (31) gives a monthly prediction for July of 6200 which represents 11% of the annual visitor according to the Day Visit Survey. Hence multiplying 6200 by 9.091 (100/11) gives an annual estimate of 56,363 visitors. This estimate is adjusted for the period in which the survey took place.

Table 26: Comparison of Forestry Commission and model estimates of annual visits to five woodlands: Aggregation using UK Day Visit Survey data.

Site name	Forestry Commission estimate of the number of party visits per annum.	Predicted number of party visits per annum based on UK Day Visit Survey data regarding annual distribution of visits
Beechenhurst	72845	20603
Blidworth Woods	63849	82644
Chopwell	33708	20950
Mabie	51704	21403
Symonds Yat	77525	15177
Total:-	299631	160777

Consideration of Table 26 shows that aggregation on the basis of UK Day Visit Survey information results in model estimates rather lower to those provided by the Forestry Commission, the ratio between these two sets of estimates being 0.54. In general this confirms that there is fair agreement between the Forestry Commission estimates of arrivals and our own predictions once they are adjusted for trends in day visits across the year, but that the upscaling methodology based on the UKDVS provides little improvement over the more simple scaling illustrated in Table 24. An examination of the rankings of sites in Table 26 shows that there is still some substantial disagreement between our predictions and the estimates of the Forestry Commission. For example, we predict that the lowest numbers of visitors will be at Symonds Yat whereas this location has the highest number in the Forestry Commission estimates. Therefore, a third approach as a basis for aggregation was also investigated.

Figure 24: Distribution of survey effort by month



The third approach devised to address the aggregation issue involved incorporation of the temporal distribution of survey effort directly within the model. Figure 24 describes, for the entire sample of observations, the distribution of survey effort across the year.

As can be seen from Figure 24, survey effort varies highly significantly across the year with the summer months being by far the most prevalent. In order to incorporate this within our models variables were created detailing, for each site, the proportion of total survey effort undertaken in each month of the year. Statistical investigation indicated that there was too few surveys (and too little corresponding variation) within the months from October to April to justify their separate inclusion and so only separate variables for each of the months May to September were included within revised models of visitor interview numbers (i.e. coefficients on these variables, measured as proportional survey effort in each of the latter months, reflect departures from the base case of interviews outside this period). These models were estimated and provided the coefficient values illustrated in Figure 25. These parameters estimate the marginal impact on interviews that a single unit of survey effort has in each month. Hence, compared with the Oct-Apr average, 18% more interviews per hour are achieved in May, 28% more in June and so forth. These seasonal differences are smaller than might be expected, although the model is measuring seasonality in interviews achieved, not in numbers passing; there is a constraint on the number that can physically be interviewed per hour, and this is not accounted for here as the models do not include information on unsurveyed visitors. Furthermore, surveys in winter months may have been at busier times to avoid interviewers standing idle for long periods. These surveys were also less likely to have been at sites that have significant seasonality, because managers might have been less interested in hosting winter surveys where there are relatively few winter visitors. For all these reasons, we would expect the estimated coefficient to be smaller than average seasonality in visitor numbers.

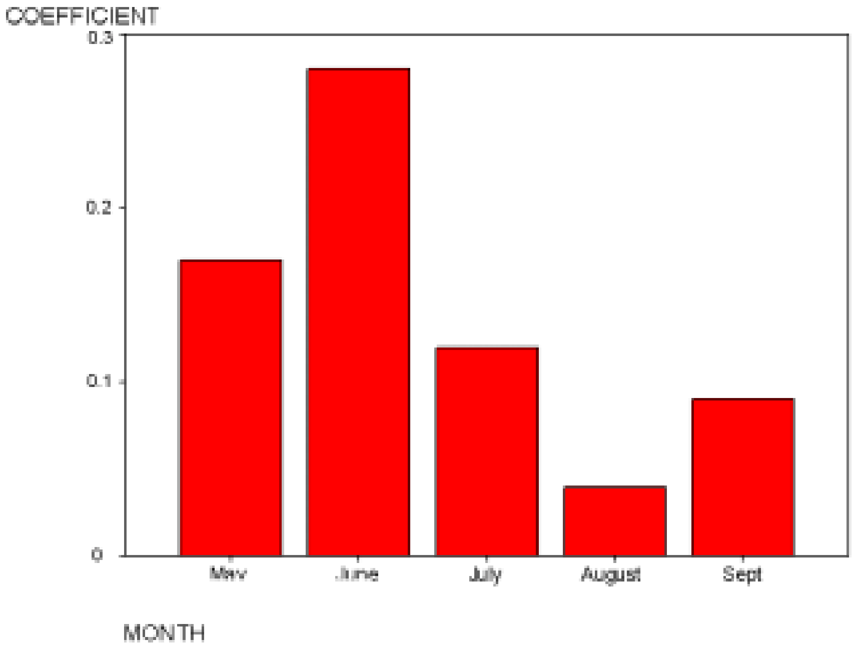


Figure 25: Model parameter coefficients for variables measuring the proportion of survey effort expended in each of the months shown.

Examining Figure 25 we can see that, as expected, all coefficient values are positive, reflecting the higher number of visits taking place in these months as opposed to other periods in the year. The pattern of these values also seems reasonable. Although none of these variables were significant at the 5% level, the fact that their values conform to expectations and control for seasonality effects in survey timing meant that their inclusion within the multilevel model predicting interview numbers was justified on the grounds that it is likely to provide a superior basis for aggregation to annual visitor predications. Estimates from a model including these parameters were adjusted as set out at the start of this section (e.g. allowing for visitors during the survey period who had not been interviewed) but the aggregation to annual visits was achieved by simply multiplying the derived 24 hour predictions of arrivals (which had been adjusted for the seasonality of the survey period) by 365.25. Resultant estimates of arrivals are presented in Table 27.

Table 27: Comparison of Forestry Commission and model estimates of annual visits to five woodlands: Aggregation using models with predictors defining the proportion of survey effort falling in each of the months May to September.

Site name	Forestry Commission estimate of the number of party visits per annum.	Predicted number of party visits per annum based on UK Day Visit Survey data regarding annual distribution of visits
Beechenhurst	72845	45755
Blidworth Woods	63849	57211
Chopwell	33708	27688
Mabie	51704	20521
Symonds Yat	77525	50142
Total:-	299631	201317

The estimates detailed in Table 27 are encouraging and the degree of agreement is similar to that obtained from the work with the Day Visit Survey data (as detailed in Table 26). The overall correspondence between Forestry Commission and model derived estimates of arrivals has a ratio of 0.67. The value of this ratio is improved compared to that for our previous analysis, and there is a substantial improvement in the consistency of predictions in terms of the ranking of sites. The top three sites and lower two sites are consistently ranked across the two sets of estimates suggesting that this approach allows the ability of the underlying models to distinguish between high and low visitation models (as noted with respect to Figure 22) to be preserved. Given the acknowledged uncertainty with regard to the Forestry Commission estimates we feel that this is a satisfactory finding.

4.1 Disaggregation of predicted annual visits by visitor outset zone.

The annual visitor estimates detailed above are derived from our interview prediction model which takes as its base unit the number of party visits generated from each zone. This model can therefore be used to provide per annum estimates for each zone. Figure 26 illustrates results for area around the site at Salcey while Figure 27 repeats this analysis for the entire area of Great Britain. As expected, the highly significant travel time variable is reflected in the relative concentration of visitors around the site. This provides a useful insight into the demand characteristics of particular sites.

Figure 26: Annual party visits to Salcey disaggregated by outset zone: Area around the site

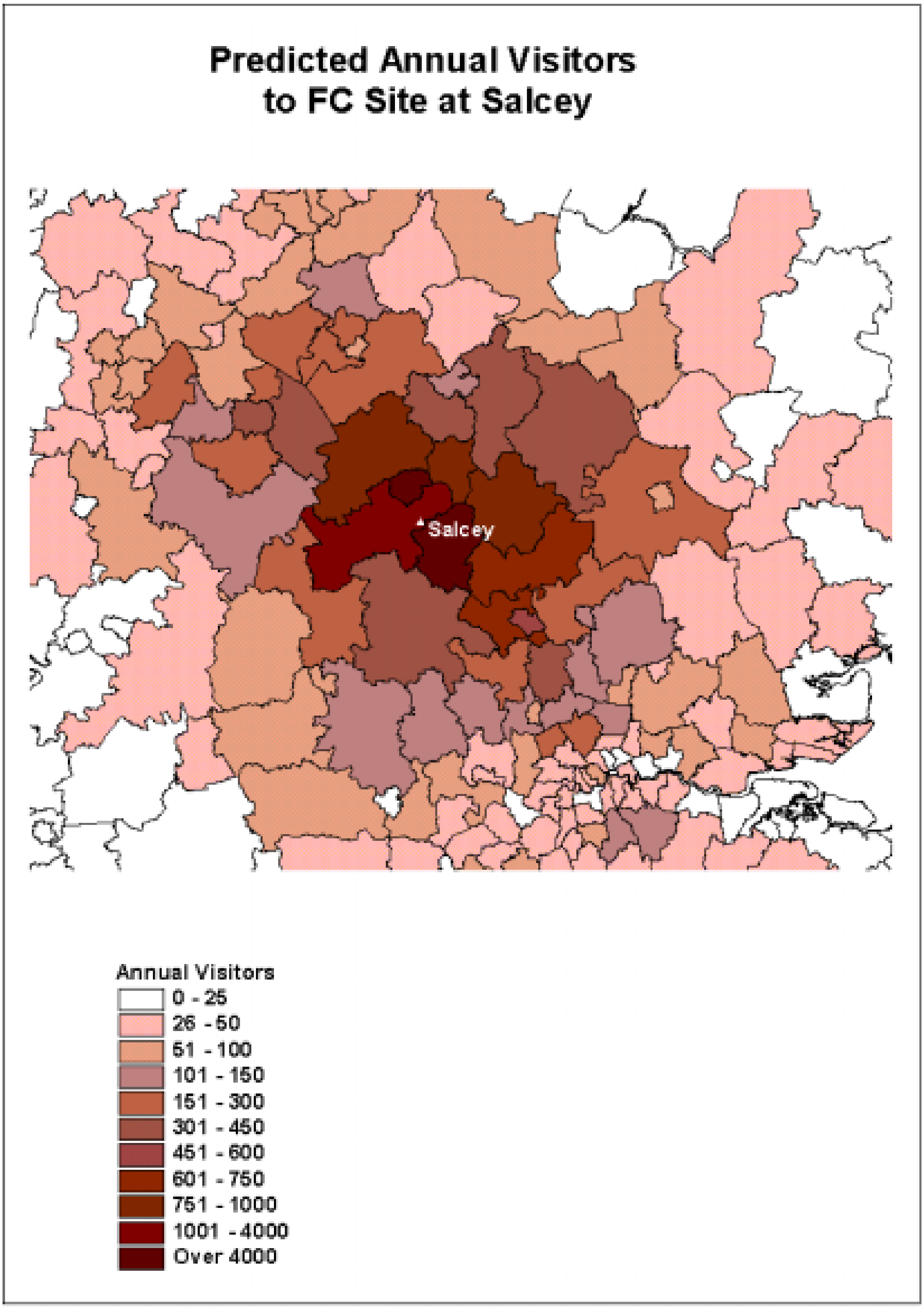
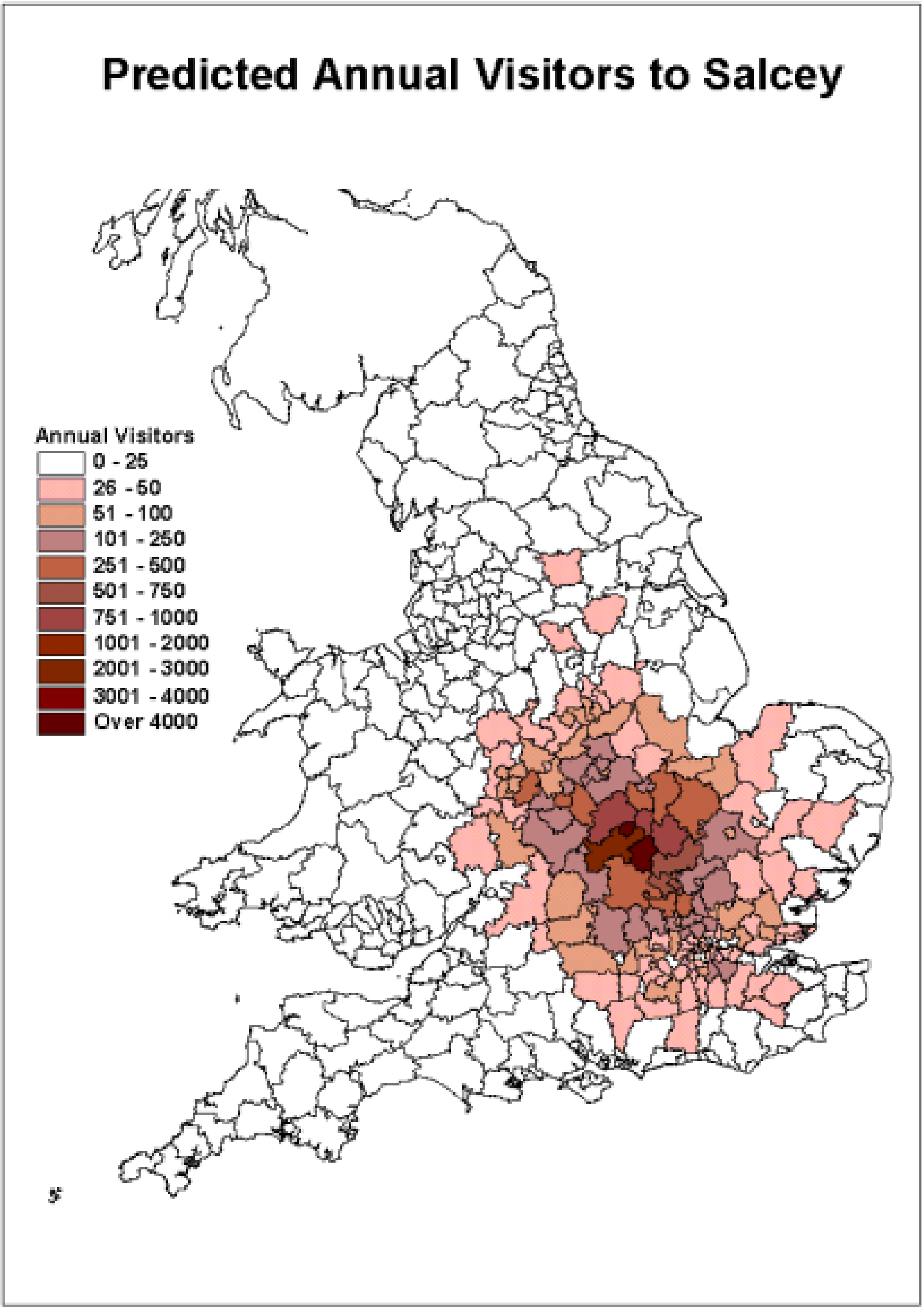


Figure 27: Annual party visits to Salcey disaggregated by outset zone: All areas across Great Britain.



5. PREDICTING THE IMPACT OF ADDING FACILITIES AT FORESTRY COMMISSION WOODLANDS.

The estimated models are also useful for investigating changes to the stock of Forestry Commission woodlands such as the addition of new sites or facilities. To test the latter five sites were chosen for analysis, site selection being determined so as to cover the geographic extent of Great Britain. Locations of these sites are indicated in Figure 28.

Figure 28: Location of five Forestry Commission woodlands used in facility impact testing.

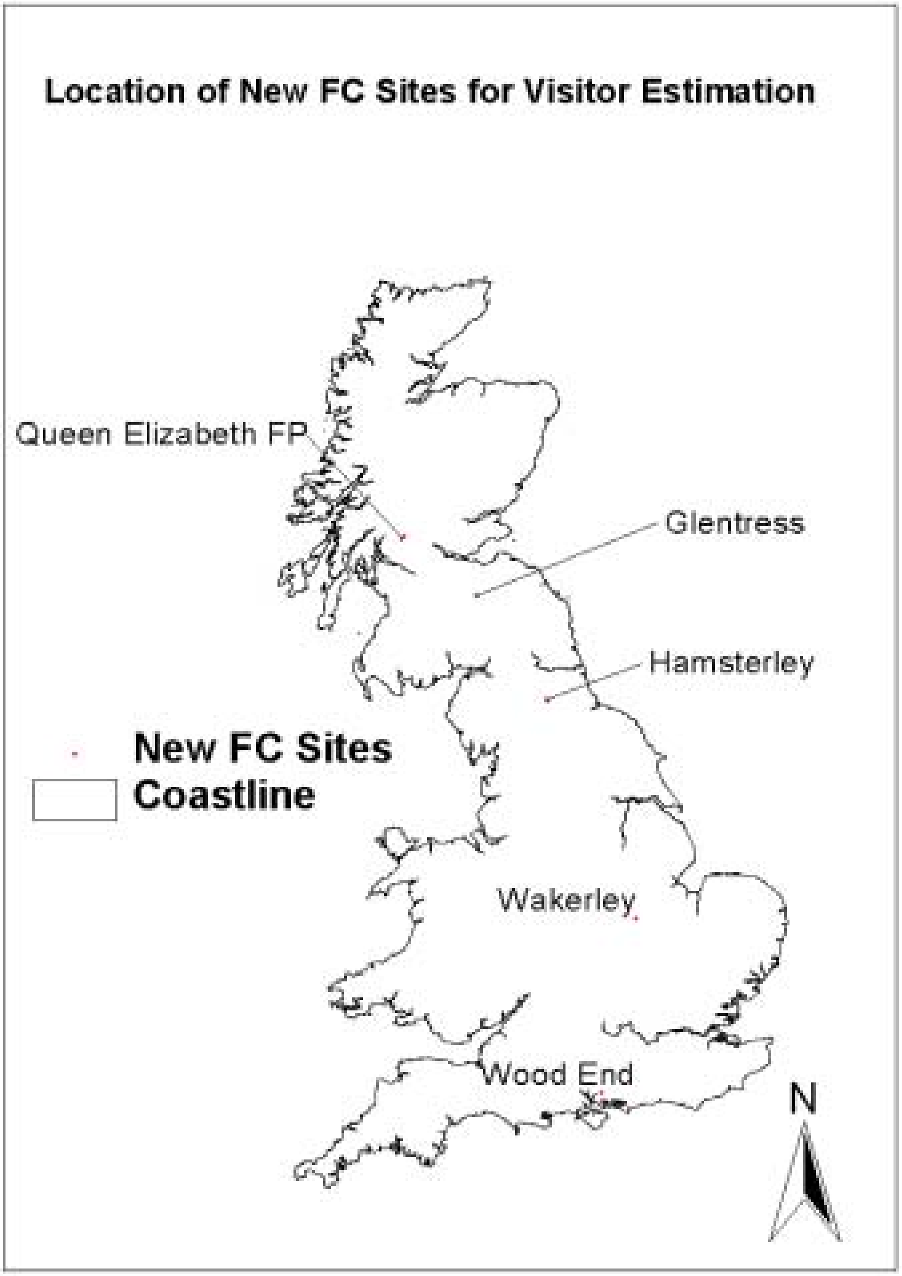


Table 28: The impact upon predicted visitor numbers of adding a visitor centre at five sites.

Site No.	Site Name	No allowance made for seasonality of survey effort		Allowance made for seasonality of survey effort	
		Predicted number of daily party visits without visitor centre	Predicted number of daily party visits with visitor centre	Predicted number of daily party visits without visitor centre	Predicted number of daily party visits with visitor centre
64	Glentress	163	309	147	286
72	Hamsterley	47	90	103	194
110	Queen Elizabeth	107	202	168	318
136	Wakerley	153	291	181	343
151	Woodend	199	378	142	268

A number of models were estimated for the above sites, looking at the presence or absence of a visitor centre. Results from these analyses are presented in Table 28

The findings detailed in Table 28 suggest that the presence of a visitor centre will have a substantial positive impact upon the number of visits (although these values have not be corrected to account for missed visitors or invalid postcodes). However, these findings should be treated with caution as they infer that the general relationship between a facility, like a visitor centre, derived from the overall sample, will apply to the addition of a new centre in a wood which previously had none. This may be erroneous because the presence of a visitor centre in the existing set of sites may proxy other characteristics of those sites which may not be endowed by simply building a new visitor centre in a given woodland. For example, if sites which presently have visitor centres are, say, bigger, more diverse, have other facilities or be located near to major population densities, then the mere construction of a visitor centre at a new site may well not endow that site with such characteristics. In such cases the above findings may well represent over-estimates of the impact of such facilities.

Despite the fact that the findings from this part analysis should be treated with caution, they do illustrate how the models we have presented here could be used to test what impact the introduction of new visitor facilities, or the removal of existing ones, may have at existing sites. They also illustrate how the models could be used to predict visitor numbers at 'new' sites for which no survey information is available. To use the models for new sites it would, of course be necessary to estimate travel time values to each site from outset zones. Hence, if the parameter coefficients presented in the models here were used in this prediction process, it would be necessary to ensure that any GIS methodology adopted to provide these new estimates produced comparable travel time estimates to our own. This is particularly important given that the estimated travel time to each site has been shown to be by far the most dominant predictor of surveyed visitor numbers. Aside from this issue it would also be important that the original data matrix used in the development of these models was available. If this information was at hand, it would be possible to make predictions of visitor numbers by multiplying the parameter coefficients presented in the tables above by the corresponding values of the variables in the data matrix, and then summing the product of this multiplication for each site. An alternative to this process would involve the re-estimation of models. In the future, this further model development may be desirable as information from the 2001 UK Census of Population becomes available, the nature and form of Forestry Commission visitor surveys changes, and new methodologies for parameter estimation (particularly associated with the development of simulation methods in the field of multilevel modelling) become available. Further model development will also allow some of this issues identified in this report, such as the relationship between model performance and visitor type mix, to be investigated more fully.

6. TRAVEL COST VALUES

Part Two of this report discusses consumer surplus and willingness to pay measures of the recreational value of trips to woodlands. However, a further aspect of such values concerns the travel cost values incurred by visitors for such trips. The models estimated in this research are amenable to these calculations as the travel time variable can readily be replaced by travel cost estimates (although the critique provided by Randell (1994) and discussed above, is relevant here).

In calculating travel cost allowance was made for both travel expenditure and travel time values. The procedure used was as follows:

(i) To convert travel time to the cost of travel time: Travel time (as calculated in the GIS) from any outset location was multiplied by the regional hourly wage rate (taken from CSO, 1998) appropriate to that outset location. This value was then multiplied by one-third following the work of Cesario (1976) as applied to woodlands by Benson and Willis (1992)

(ii) To calculate travel expenditure: Travel distance was calculated as the product of travel time and an assumed average speed of 40mph. Travel expenditure was then calculated by multiplying travel distance by average running costs per mile obtained from the Automobile Association (figures for 1998). Note that these values will tend to slightly over-state marginal costs for rural trips when fuel consumption is better than in the urban cycle.

(iii) To calculate travel cost per group from each outset location: The travel cost value obtained at (i) was added to the travel expenditure value derived at (ii).

(iv) To calculate travel cost per outset location: The value derived at (iii) was multiplied by the total number of parties arising from each outset location

(v) To obtain the total value of travel costs per site: The values estimated at (iv) were summed on a per site basis.

(vi) To calculate the average cost of a group visit: the value estimated at (v) was divided by the number of visitors to the site.

The above calculation was carried out separately for all visitors, day-trippers and holidaymakers to yield the values detailed in Table 29. The values in Table 29 correspond to one way travel only, and hence the subsequent valuation exercise they were doubled to encompass the entire cost of trips. It should be noted that, for holidaymakers, only a proportion of the full travel costs are relevant to the woodland visit. The final column in this table details all visitor values estimated from a benefit transfer analysis in which the site in question is omitted and values estimated from the remaining dataset. As per the comparable analysis for visitor numbers given in Table 15, this final column conforms reasonably well to the full-information result. As expected travel costs are higher for holidaymaker than day-trip visitors, the former reflecting travel from home to the woodland area.

Table 29: One way travel cost values for all visitors, day-tripper and holidaymakers

Average Travel Cost Per Group Visit Per Site				
Siteno	**All visitors**	**Day Visitors**	**Hol Visitors**	**Omit All Visitors**
3	8.62	4.02	31.76	7.85
6	4.30	4.09	11.10	8.48
8	12.24	6.61	77.01	21.70
9	14.00	8.79	24.26	12.46
14	9.32	5.17	22.78	9.61
15	19.62	8.16	29.85	10.99
17	2.95	2.74	26.46	4.84
18	18.71	6.58	27.97	9.52
20	5.73	4.88	21.10	13.10
33	4.02	4.00	5.25	5.25
34	18.15	14.72	18.79	12.03
40	3.63	2.79	62.24	5.41
43	13.13	7.93	25.21	12.58
44	20.54	12.06	29.54	18.90
46	4.53	4.44	6.62	8.65
49	3.85	3.48	12.25	12.16
51	15.48	8.13	66.92	24.92
61	11.12	5.22	30.72	10.46
66	35.77	13.84	47.78	36.58
68	25.59	10.38	30.83	23.75
72	16.49	11.14	32.65	16.47
80	35.79	19.50	45.00	29.42
83	17.38	15.62	41.27	11.54
84	5.11	3.26	41.53	4.55
86	85.83	58.30	87.13	60.60
95	20.28	6.43	36.47	25.06
111	57.25	27.19	62.84	30.48
117	3.78	3.14	17.22	7.46
119	8.95	3.85	25.13	7.06
121	17.66	8.74	41.57	18.45
126	19.76	11.76	25.19	11.20
128	11.64	7.74	25.74	12.58
129	2.58	2.52	12.78	5.56
130	19.34	6.01	42.45	14.63
134	4.56	3.39	22.93	4.04
137	18.98	8.24	27.28	11.43
141	4.96	4.61	12.78	8.32
143	11.75	8.59	24.91	12.66
147	8.28	7.18	22.92	14.29
153	6.81	4.65	18.95	5.94

The values detailed in Table 29 are travel costs rather than consumer surplus values. However they may be used with the predicted annual visitor numbers we produced from the 5 sites for which Forestry Commission estimates were available in order to estimate the *travel cost* value placed by visitors on each site. Recall that we used 3 different methods to gross up 24 hour model predictions to annual values. The first (Method 1) involved simply multiplying the predictions by the annual values and correcting for invalid interviews. The second methodology (Method 2) fitted the predicted values from the model to the monthly distribution of visitors to forest sites based on data in the UK day visits survey. The third methodology (Method 3) involved modelling the period of survey effort at each site by adding a new suite of explanatory variables to the model. For the purposes of comparison, all three methodologies were used to produce the travel cost values detailed in Table 30 below.

Site name	Travel cost estimated from Method 1	Travel cost estimated from Method 2	Travel cost estimated from Method 3
Beechenhurst	£548,838	£513,426	£1,140,214
Blidworth Woods	£760,189	£799,994	£553,802
Chopwell	£187,467	£219,976	£290,724
Mabie	£1,062,744	£1,072,718	£1,028,512
Symonds Yat	£364,380	£339,947	£1,123,180
Total:-	£2,923,619	£2,947,606	£4,136,432

Table 30: Estimated travel costs to five woodlands using three methods of converting model predictions to annual visitor numbers.

As would be expected from the differences in the predictions of annual group visits made from the three models, there is considerable variability between methods in the travel cost estimates made for each site. The lowest overall costs are produced from Method 1. Similarly, the finding that Method 2, including seasonality variables in the modelling process, gave the highest total estimated costs. The variability in costs within sites across the three methods is a function of how many parties at each site each method predicted. For example, Method 2 predicted 21,204 party visits for Mabie, whilst method 1 predicted that 21,403 parties would visit the site annually. This discrepancy is reflected in the rather different travel cost (£1,072,718 and £1,062,744 respectively) that the two methods placed on each site. Observed variations in travel costs between sites that have been estimated using the same method are a function of both the number of predicted visits to each site, and the travel cost for the site taken from Table 29. For example, Mabie has a large number of holidaymaker visitors coming some distance to reach the site, and hence has a mean cost of £50.12 per visit. On the other hand, the mean cost for Blidworth Woods is just £8.96.

It should be stressed once again that these value are travel costs rather than measures of the consumer surplus that visitors will place on each woodland. For discussion of the latter we move to Part Two of this report.

Part Two:
Estimating the value of informal recreation at British woodlands:
A multilevel meta-analysis.

by

Ian J. Bateman and Andrew P. Jones

Overview

This second part of the report presents a variety of meta analysis models of woodland recreation benefit estimates, contrasting conventionally estimated models with those provided by novel, multi-level modelling (MLM) techniques (Goldstein, 1995). Our conventional models suggest that studies carried out by certain authors are associated with unusually large residuals within our meta-analysis. However, the MLM approach explicitly incorporates the hierarchical nature of meta-analysis data, with estimates nested within study sites and authors. Allowing for this reveals that these residuals are not a significant determinant upon values, suggesting that, at least in this aspect, estimates may be more robust than indicated by less sophisticated models. However, previously noted differences in benefit estimates between alternate valuation methods persist across our various analyses and remain a cause for concern.

1. Introduction

The past two decades have witnessed an increasing reliance upon cost-benefit analysis (CBA) as a tool for project appraisal and to inform decision making. In the UK, a typical example of this trend is provided by the 1995 Environment Act which brought into being the Environment Agency (EA) and imposed 'general duties' upon the Agency to take account of the costs and benefits arising from its policies (H.M. Government, 1995). For many agencies, particularly those which have an explicitly environmental or public goods remit, the assessment of benefits necessitated by adopting CBA approaches has led to a growing interest in tools for the monetary valuation of preferences for environmental goods and services. Consequently, expressed preference methods such as contingent valuation (CV) and conjoint analysis (CA) together with revealed preference techniques such as hedonic pricing (HP) and individual and zonal travel cost (TC) have enjoyed an unprecedented increase in application. However, such applications themselves incur both direct and time related costs. Consequently the proliferation of these studies has coincided with increased interest in the potential for benefit transfer.

Rosenberger and Loomis (2000) define benefit transfer as 'the application of values and other information from a "study" site with data to a "policy" site with little or no data' (p1097). A number of approaches to undertaking transfers are available[2] including simple transfer of unadjusted point estimates, transfer of benefit demand functions and meta-analysis. As the simplest approaches cannot incorporate the characteristics of a given site within the transfer exercise, considerable attention is being given to the development of methods for transferring benefit demand functions (Loomis, 1992; Bergland *et al*., 1995; Loomis et al., 1995; Downing and Ozuna, 1996; Kirchhoff et al., 1997; Brouwer and Spaninks, 1999; Brouwer and

[2] For reviews of the issues raised by benefit transfer applications see Brookshire and Neill (1992), OECD (1994), Pearce and Moran (1994), Bergland *et al*., (1995) and Desvousges et al., (1998).

Bateman, 2000). However, results are mixed with some studies reporting considerable success while others indicate abject failure. Given this and the empirical difficulties of such studies, a substantial literature has developed regarding the applications of meta-analysis techniques as a basis for benefit transfer.

Meta-analysis is the statistical analysis of the summary findings of prior empirical studies for the purpose of their integration (Glass, 1976; Wolf, 1986). Developed over the last thirty years, it has most commonly been applied in the fields of experimental medical treatment, psychotherapy, and education. Typically, these experiments took place in well-controlled circumstances with standard designs. Deviation from such specifications increases the problems with any cross-analysis (Glass *et al.*, 1981)[3].

Despite problems, meta-analysis offers a transparent structure with which to understand underlying patterns of assumptions, relations and causalities, so permitting the derivation of useful generalizations without violating more useful contingent or interactive conclusions (Hunter *et al*., 1982). It permits the extraction of general trend information from large datasets gleaned from numerous studies which would otherwise be difficult to summarize. In comparison with other benefit transfer techniques, Rosenberger and Loomis (2000) identify three advantages of adopting a meta-analysis approach: (i) it typically collates information from a greater number of studies, (ii) it is relatively straightforward to control for methodological differences between valuation source studies, (iii) benefit transfer is readily effected by setting explanatory variable values to those at the desired target site be it a previously surveyed, unsurveyed or just proposed (i.e. currently non-existent) site.

Table 1: Meta-analysis studies in environmental and resource economics.

Subject area	Study authors
Recreation benefits	Bateman et al., (1999b, 2000), Markowski, et al., (2001), Rosenberger and Loomis (2000), Shrestha and Loomis (2001), Smith and Kaoru (1990a), Sturtevant et al. (1995), Van Houtven et al., (2001), Walsh et al. (1990, 1992)
Price elasticity in TC studies	Smith and Kaoru (1990b)
CV versus revealed preference	Carson et al. (1996)
Multiplier effects of tourism	Baaijens, et al. (1998), Van den Bergh et al., (1997, Ch9)
Wetland functions	Brouwer et al., (1999), Woodward and Wui (2001)
Groundwater quality	Boyle et al., (1994), Poe et al., (2001)
Price elasticity for water	Espey et al., (1997)
Urban pollution valuation	Smith (1989), Smith and Huang (1993), Smith and Huang (1995), Schwartz (1994), Van den Bergh et al., (1997, Ch10)
Noise nuisance	Button (1995), Nelson (1980), Van den Bergh et al., (1997, Ch4)
Congestion and transport	Button and Kerr (1996), Van den Bergh et al., (1997, Ch13 and 14), Waters (1993)
Visibility and air quality	Desvousges et al., (1998), Smith and Osborne (1996)
Endangered species	Loomis and White (1996)
Valuation of life estimates	Van den Bergh et al., (1997, Ch11)

[3] Meta-analyses also face the problem that studies published in the available literature may over represent that subset of all studies which produce ‘positive’ or significant results if studies yielding ‘negative’ or non-significant findings tend not to be published.

Table 1 extends reviews by Van den Bergh et al., (1997) and Smith and Pattanayak (forthcoming) to provide a brief summary of studies in this area. The empirical applicability of meta analysis to any given context is determined by the number, quality and comparability of studies available to the researcher (Desvousges, et al., 1998). Here there is a difficult trade-off between the desire to extend the remit of analysis so as to enhance the applicability of results to different goods, provision changes, locations and contexts, and the consequent increase in data demands which such extensions entail. For example, Rosenberger and Loomis (2000) consider a wide range of outdoor recreation activities (10 separate categories ranging from fishing to rock climbing to snowmobiling) across a very extensive locational remit, the US and Canada. This analysis requires a large valuation dataset and their study utilizes 682 value estimates from 131 separate projects. By contrast the meta-analysis presented in this paper considers just one type of activity, recreation in open-access woodlands, and just one geographical area, Great Britain, a land area just over 1% the size of that considered by Rosenberger and Loomis. Our analysis is initially restricted just to measures obtained by application of the CV method yielding a dataset of 44 value estimates from 11 studies. A second analysis supplements these data with results obtained from 6 TC studies, bringing the total number of value estimates to 77. While this is less than the size of the Rosenberger and Loomis dataset (reflecting the fewer number of studies conducted in Great Britain) the much smaller geographical remit of our study, and its focus upon just one activity, mean that data are placed under considerably less stress, enhancing the reliability of resultant benefit transfer estimates. The disadvantage of this focus is that our results are not readily applicable to other activities or to areas outside Great Britain.

Full details of the dataset assembled for this research are given in Appendix B to this report.

The first meta-analysis presented here is conducted using conventional generalized linear model (GLM) regression techniques (McCullagh and Nelder, 1989). However, the second and third analyses apply multilevel modelling (MLM) techniques (Goldstein, 1995) to the full dataset of observations. For further information on MLM readers should refer to the Goldstein book or refer the multilevel.ioe.ac.uk website. Because the MLM approach allows the researcher to explicitly incorporate potential nested structures within the data, it is possible to examine a number of key issues and criticisms of both meta-analysis and valuation studies. For example, it is possible to control for the fact that the number of estimates provided is not constant across authors[4] (as in the case presented here) such that a conventional GLM analysis may give results which are heavily weighted towards the most prolific authors. Another important issue concerns whether different authors consistently implement valuation techniques in a manner which is liable to lead to upward or downward pressure upon resultant valuation estimates. The detection of evidence of these problems would constitute a substantial criticism of methods for valuing preferences for non-market goods.

2. The recreational value of forests: Background and data sources

In terms of land use, British forestry has always been the poor cousin of agriculture. A history of deforestation meant that, by 1900, only 4% of England and Wales and 2% of Scotland and Ireland was forested, by far the lowest level in Europe (Rackham, 1976). The establishment of the FC in 1919 has done much to reverse this trend and over 10%[5] of the land area of Great Britain is under woodland today. The FC woodland, the largest single source of open-access

[4] One could identify a number of possible hierarchies for example value estimates could be nested within studies and then within authors. As discussed subsequently, they can also be 'cross classified' when different authors conduct studies at the same (as well as differing) forest sites.

[5] This decomposes into 14.7% of Scotland, 12.0% of Wales and 7.4% of England. However, this is still well below an EU average of about 25% of land area under forestry (FICGB, 1992).

land, generates approximately 24-32 million recreational visits per annum (NAO, 1986; Benson and Willis, 1990; 1992), and produces a national aggregate consumer surplus value estimated at between £40 million (Bateman, 1996) and well over £50 million (Benson and Willis, 1992) at current prices. From an economic perspective, the recreational value of forestry is therefore one of its most important benefit streams.

The initial stage of any meta-analysis involves a survey of the relevant literature to identify potential base data studies. Table 2 presents summary details from some 30 studies of UK woodland recreation value yielding over 100 benefit estimates.

Table 2: Studies of open-access woodland recreation value in Great Britain.

Value type	Recreation value unit	Valuation method	No. of studies	Date conducted[1]	No. of value estimates	Value range (£, 1990) (m = million)
Use	Per person per visit.	CV	8[a]	1987 – 1993	28	£ 0.28 - £ 1.55
Use + option	Per person per visit.	CV	3[b]	1988 – 1992	16	£ 0.51 - £ 1.46
Use	Per person per visit.	ZTC	3[c]	1976 – 1988	17	£ 1.30 - £ 3.91
Use	Per person per visit.	ITC	3[d]	1988 – 1993	16	£ 0.07 - £ 2.74
Use	Per person per year	CV	3[e]	1989 – 1992	7	£ 5.14 - £ 29.59
Use	Per household capital[2]	CV	3[f]	1990	3	£ 3.27[3] - £ 12.89
Use	FC forests/ conservancy[4]	TC	1[g]	1970	13	£0.1m - £1.1m
Use	Total UK value	TC	6[h]	1970 – 1998	6	£6.5m - £62.5m
-	All studies	-	30	1970 - 1998	106	-

Notes:

1 = Dates refer to the year of study survey rather than publication date.
2 = These studies use a once-and-for-all willingness to pay per household question.
3 = We have recalculated this figure by including those who refused to pay as zero bids.
4 = The FC at the time divided the area of Great Britain into a number of Forest Conservancies and large forests to which these estimates relate.

Study references:

a = Whiteman and Sinclair (1994); Hanley and Ruffell (1991); Bishop (1992); Willis and Benson (1989); Hanley (1989); Willis et al (1988); Bateman and Langford (1997); Bateman (1996).
b = Bishop (1992); Willis and Benson (1989); Willis et al (1988)
c = Benson and Willis (1992); Hanley (1989); Everett (1979)
d = Willis and Garrod (1991); Bateman (1996); Bateman et al., (1996)
e = Whiteman and Sinclair (1994); Bishop (1992); Bateman (1996)
f = Hanley and Munro (1991); Hanley and Ecotec (1991); Hanley and Craig (1991).
g = H.M. Treasury (1972)

h = H.M. Treasury (1972); Grayson et al (1975); NAO (1986); Willis and Garrod (1991); Benson and Willis (1992); Bateman (1996).

An initial analysis focused solely upon those estimates of per person per visit recreation value obtained from applications of the CV method. Here survey respondents, typically interviewed on-site, were asked to state their willingness to pay (WTP) for the recreational value of the forests concerned[6]. Table 2 indicates that there are 8 studies yielding 28 estimates of the direct 'use value' of the recreational services provided by forests. Three studies also asked respondents about their WTP for both the present and possible future use (or 'option value'; Weisbrod, 1964; Pearce and Turner, 1990) of forests providing a further 16 estimates of this wider recreational value. In total therefore, these studies yield 44 value estimates[7].

Table 3: Per person per visit woodland recreation value estimates disaggregated by study author and valuation/estimation method

Method	Whiteman & Sinclair	Hanley et al.	Bishop	Willis et al.	Bateman et al.	Everett	All
CV	3 0.78 (0.66 - 0.93) [0.14]	6 1.30 (0.85 - 1.55) [0.27]	4 0.89 (0.46 - 1.46) [0.46]	28 0.71 (0.28 - 1.29) [0.27]	3 1.08 (0.47 - 1.55) [0.55]	0 -- -- --	44 0.84 (0.28 - 1.55) [0.36]
ITCols	0 -- -- --	0 -- -- --	0 -- -- --	6 1.46 (0.47 - 2.74) [0.84]	3 1.35 (1.07 - 1.58) [0.26]	0 -- -- --	9 1.42 (0.47 - 2.74) [0.68]
ITCml	0 -- -- --	0 -- -- --	0 -- -- --	6 0.57 (0.07 - 1.13) [0.47]	1 1.20 (1.20 - 1.20) [--]	0 -- -- --	7 0.66 (0.07 - 1.20) [0.49]
ZTC	0 -- -- --	1 2.14 (2.14 - 2.14) [--]	0 -- -- --	15 2.53 (1.58 - 3.91) [0.66]	0 -- -- --	1 1.30 (1.30 - 1.30) [--]	17 2.43 (1.30 - 3.91) [0.71]
All	3 0.78 (0.66 - 0.93) [0.14]	7 1.41 (0.85 - 2.14) [0.40]	4 0.89 (0.46 - 1.46) [0.46]	55 1.27 (0.07 - 3.91) [0.95]	7 1.21 (0.47 - 1.58) [0.38]	1 1.30 (1.30 - 1.30) [--]	77 1.24 (0.07 - 3.91) [0.83]

Cell contents are as follows:

Number of estimates
Mean value (£/person/visit)
(Range: minimum to maximum value)

[6] Note that CV studies can be adapted to ask either WTP or willingness to accept compensation questions in respect of either gains or losses of the resource concerned (Mitchell and Carson, 1989), although only the WTP format was used in the studies concerned.

[7] Further details of these studies are provided in Bateman (1996) and Bateman et al., (2000).

[StDev of values]

A second analysis was conducted by expanding the dataset to include a further 23 per person per visit value estimates obtained from TC studies. These estimates can be further subdivided. There are 16 individual TC estimates of which 9 use ordinary least squares (OLS) estimators (hereafter identified as *ITCols*; these are liable to lead to over-estimates of benefits due to an inability to reflect the truncation of non-visitors within an on-site TC survey sample). A further 7 use maximum likelihood (ML) estimators[8] (identified as *ITCml* studies; these explicitly model the truncation of non-visitors and are not upwardly biased in this respect). There are also 17 zonal TC (*ZTC*) estimates (all of which use OLS estimators).

All TC estimates refer solely to recreational use values alone and the addition of these to our CV values allow us to examine the influence of using different valuation methods upon the estimates obtained. Table 3 reports summary descriptive statistics for the various per person per visit values which constitute our wider dataset. Here estimates are disaggregated by both study author and the valuation method employed

Table 3 highlights two important features of the dataset that are the subject of subsequent investigation. First, the data is dominated by estimates derived from studies conducted by Willis et al., reflecting their leading role in this field. Second, while the number of estimates is too small to permit calculation of confidence intervals, values produced using the ZTC method appear to be substantially higher than those from other approaches. This may be attributed to a number of causes including the use of OLS predictors in such studies and the systematic upward bias in most zonal estimates of travel time and distance (and hence consumer surplus) recently identified by Bateman et al., (1999a). We return to this issue subsequently.

3. Conventional GLM based meta-analyses

3.1. GLM meta-analysis of the CV per person per day values

Our initial meta-analysis focused upon the 44 per person per visit value estimates collected using the CV method. Examination of these studies produced a number of variables which might influence estimated values. These variables are:

Option: 1 = use value plus option value requested in WTP question, 0 = use value alone;

Elicit (WTP elicitation method): 1 = open ended (OE), 2 = iterative bidding (IB), 3 = payment card (PC), 4 = high range payment card (PCH)[9], 5 = dichotomous choice (DC);

OE (recoding of the Elicit variable): 1 = open ended elicitation method used, 0 = other;

Forest: 20 categories identifying each of the forests included in at least one of the studies;

Year: Continuous variable; the number of years before (negative) or after (positive) the base year (1990);

[8] For a discussion of ML estimators see Maddala (1983).

[9] Bateman et al., (forthcoming) contrast a standard with a high range payment card.

Author variable*: 1 = Whiteman and Sinclair, 2 = Hanley et al., 3 = Bishop, 4 = Willis et al., 5 = Bateman et al.*

Economic theory and empirical regularities observed in earlier meta-analyses provide a number of expectations regarding the potential impact of these variables upon stated WTP values. Clearly *Option* would be expected to be positively related to higher levels of WTP as it defines cases where respondents were asked to consider a wider set of values than just present recreational forest use. Economic theory also has clear expectations regarding the *Elicit* variable. A simple open ended (OE) WTP question, such as 'What are you willing to pay?' is liable to free-riding behaviour, leading to understatement of WTP relative to the incentive compatible estimates from dichotomous choice (DC) approaches (Hoehn and Randall, 1987; Carson et al., 1999)[10]. This in turn will lead to the *OE* variable being associated with lower levels of stated WTP.

The nominal *Forest* variable is included to identify any influences that variations in the nature of individual sites (e.g. facilities) may have upon stated WTP. *Year* is open to interpretation as it might reflect perceived changes in the availability or desirability of open-access recreational goods, or changes in the types of data collected or estimators and methods employed over the period spanned by the studies. This is also addressed by *Author* which is included to allow for possible differences in study design and application across researchers.

The analysis was conducted using the generalized linear modelling (GLM) approach set out by McCullagh and Nelder (1989). This permits direct incorporation of nominal variables such as *Elicit* and *Author* in a way which yields estimates which are comparable with the use of separate binary variables for each category, but which is more efficient than adopting this explicit approach. Coefficient estimates for each category are interpreted normally.

Collinearity between the *Author* and *Forest* variables was too high to permit their simultaneous inclusion within a single model (e.g. all studies by Hanley et al., were conducted in Aberfoyle forest). Inspection of the relationship between the separate effects of these variables upon stated WTP suggested that *Author* provided the more interesting insight into the process of stated value formation. Hence, this variable was modelled, with analysis of *Forest* being reported subsequently. Analysis also showed that the variable *Year* gave a small, positive but statistically insignificant ($p>0.2$) and the variable was omitted from the present analysis. Tests indicated that a linear model performed better than other functional forms and the final model is given as Table 4[11].

Table 4: GLM meta-analysis of CV estimates of per person per visit recreation values (£, 1990) for open-access woodland in Great Britain.

Variable		Coefficient	95% CI	p
Constant		1.2822	(1.082 - 1.482)	0.000
Option		0.2094	(0.135 - 0.284)	0.000
Elicit[1]	*1 (OE)*	-0.3313	(-0.579 - -0.084)	0.013
	2 (IB)	-0.2980	(-0.681 – 0.085)	0.136
	3 (PC)	0.0753	(-0.172 – 0.323)	0.554
	4 (PCH)	0.5820	(0.146 – 1.018)	0.013

[10] While the DC method is incentive compatible, whether or not it is in practice also demand revealing (i.e. produces unbiased estimates of true WTP) is an ongoing source of debate (Green et al., 1998; Carson et al., 1999).

[11] Bateman et al., (1999a) use a reduced form of the model reported in Table 4 in their GIS based benefit transfer analysis of woodland recreation values.

Author[2]	*1 (Whiteman & Sinclair)*	0.0385	(-0.186 – 0.263)	0.739
	2 (Hanley et al.)	0.3652	(0.147 – 0.383)	0.002
	3 (Bishop)	-0.0584	(-0.265 – 0.149)	0.584
	4 (Willis et al.)	-0.2405	(-0.382 - -0.099)	0.002

Notes on above table:
1. Base case (category 5) = *DC*
2. Base case (category 5) = *Bateman et al.*
R^2 = 0.716; n = 44

Table 4 shows the *Option* variable provides the strongest influence upon stated WTP. Respondents facing a 'use plus option value' question stated very significantly higher WTP sums than those facing 'use value alone' questions; a result which conforms well with prior expectations. The *Elicit* variable shows that two elicitation methods produce estimates which differ significantly from others in the dataset[12]; OE values being substantially lower than the base case (the incentive compatible DC approach) with high range payment cards being, unsurprisingly, higher than all other approaches (although this latter result relies upon just a single value estimate). Again these results conform well with prior expectations. Two of the *Author* categories are also significant: category 2 (*Hanley et al.)*, which yields recreation value estimates that are higher than average, whilst category 4 (*Willis et al.*) yields lower than average estimates. This provides support for the contention that reported valuation estimates are partly dependent upon the researcher carrying out the study.

3.2. GLM meta-analysis of the CV and TC per person per day values

The analysis was subsequently expanded by the addition of the 23 estimates of per person per visit woodland recreation values obtained using TC methods. In addition to increasing the total observations to 77, this also adds a new categorical explanatory variable, *Method*, which defines the four method/estimation combinations used (*CV*, *ITCols*, *ITCml* and *ZTC*, of which the CV studies are held as the base case in subsequent analyses)[13]. Both economic theory and empirical studies suggest that the categories of *Method* may be associated with differing recreation value estimates. This is because, while the CV approach yields direct Hicksian welfare measures of WTP, TC methods provide Marshallian consumer surplus estimates. The relationship of these measures depends upon the relative shape of the underlying compensated and uncompensated demand curves for the goods and provision changes concerned (Just et al., 1982; Boadway and Bruce, 1984). Carson et al., (1996) review 83 studies from which 616 comparisons of CV to revealed preference (RP; including TC) estimates are drawn, yielding a whole sample mean CV:RP ratio of 0.89 (95% CI = 0.81 to 0.96), i.e. CV estimates were found to be significantly lower than TC values.

The *Elicit* and *OE* variables were omitted from this analysis as they do not apply to the TC studies. Models were estimated using GLM techniques, allowing for variation across categorical variables[14]. Table 5 details results for a number of model specifications. In each case, tests of functional form indicate that the linear specification performs roughly as well as other standard forms and is retained for comparability and ease of interpretation.

[12] Elicitation type 6 (DC) is used as the base category here.
[13] In addition we have one further *Author* category (*Everett*) and one extra forest study site.
[14] Equation (A1) in Bateman et al., (2000) details such a model showing effects for individual forests.

Table 5: GLM meta-analyses of CV and TC estimates of per person per visit recreation values (£, 1990) for open-access woodland in Great Britain.

	Models						
	A	B	C	D	E	F	G
Intercept	1.1980 (0.1057) [11.34] {0.000}	1.2781 (0.1177) [10.86] {0.000}	0.8368 (0.0764) [10.95] {0.000}	0.7523 (0.0852) [8.83] {0.000}	0.6687 (0.0862) [7.75] {0.000}	0.6796 (0.0886) [7.67] {0.000}	0.7697 (0.0910) [8.46] {0.000}
Option		-0.3489 (0.2338) [-1.49] {0.140}		0.1902 (0.1521) [1.25] {0.215}	0.2717 (0.1436) [1.89] {0.063}	0.2626 (0.1469) [1.79] {0.078}	0.3414 (0.1434) [2.38] {0.020}
Forest: **Cheshire**	-0.3780 (0.3839) [-0.98] {0.328}	-0.3883 (0.3808) [-1.02] {0.311}			-0.4029 (0.2163) [-1.86] {0.067}	-0.4153 (0.2203) [-1.88] {0.064}	-0.3962 (0.2109) [-1.88] {0.065}
Loch Awe	0.5653 (0.4881) [1.16] {0.251}	0.6015 (0.4847) [1.24] {0.219}			0.4379 (0.2760) [1.59] {0.117}	0.4212 (0.2812) [1.50] {0.139}	0.4154 (0.2690) [1.54] {0.127}
Aberfoyle	0.4445 (0.3104) [1.43] {0.156}	0.3644 (0.3124) [1.17] {0.247}			0.5491 (0.1799) [3.05] {0.003}		
Method[1]: **ITCols**			0.5876 (0.1854) [3.17] {0.002}	0.6722 (0.1884) [3.57] {0.001}	0.8005 (0.1767) [4.53] {0.000}	0.7910 (0.1805) [4.38] {0.000}	0.7994 (0.1727) [4.63] {0.000}
ITCml			-0.1811 (0.2062) [-0.88] {0.383}				
ZTC			1.5973 (0.1447) [11.04] {0.000}	1.6818 (0.1490) [11.29] {0.000}	1.6988 (0.1378) [12.33] {0.000}	1.7253 (0.1418) [12.17] {0.000}	1.8461 (0.1427) [12.94] {0.000}
Author: **Hanley**						0.4926 (0.1955) [2.52] {0.014}	0.4390 (0.1881) [2.33] {0.023}
Year							0.0755 (0.0276) [2.74] {0.008}
R^2 (adj. R^2) n	0.059 (0.020) 77	0.087 (0.036) 77	0.545 (0.531) 77	0.549 (0.534) 77	0.714 (0.690) 77	0.703 (0.678) 77	0.732 (0.705) 77

Cell contents are: Estimated coefficient
(StDev)
[t-value]
{p-value}

Table 5 (cont.)

where:

Dependent variable = recreational value per person per visit;

Option	= 1 where the value estimate relates to the sum of use plus option value and 0 where the value estimated is use value alone[15];
Cheshire	= 1 for studies conducted at Cheshire forest and 0 otherwise;
Loch Awe	= 1 for studies conducted at Loch Awe forest and 0 otherwise;
Aberfoyle	= 1 for studies conducted at Aberfoyle forest and 0 otherwise;
ITCols	= 1 if study uses the individual travel cost method with an OLS estimator and 0 otherwise;
ITCml	= 1 if study uses the individual travel cost method with a ML estimator and 0 otherwise;
ZTC	= 1 if study uses the zonal travel cost method and 0 otherwise;
Hanley	= 1 if study conducted by Hanley et al., and 0 otherwise.
Year	= Continuous variable; the number of years before (negative) or after (positive) the base year (1990);

Note: The CV method is held as the base case for the various categories of the *Method* variable.

In Table 5, Model A only uses the three *Forest* categories which were shown to be the most significant site related predictors in preliminary ANOVA investigations. All three prove statistically insignificant in the absence of other predictors. The addition of *Option* to yield Model B does little to improve overall explanatory power. However, when all these variables are removed in favour of *Method* to yield Model C, explanatory power increases dramatically, although there is good prior reason to believe that the ZTC results are upwardly biased. Indeed *ZTC* has a large, positive and highly significant coefficient. Hence caution is required regarding direct interpretation of the fit statistics reported for these models. Nevertheless, the direction of effect for this and the *ITCols* variable is in accordance with both theoretical and empirical expectations, particularly as all the estimates are obtained from OLS regression techniques which omit to model the truncation of non-visitors and are liable to overestimation of consumer surplus values (the positive and significant coefficients on *ZTC* and *ITCols* being in line with the findings of Carson et al., 1996). However, interestingly, *ITCml* is statistically insignificant, suggesting that estimates produced by this method are similar to those from the base case CV method. Accordingly *ITCml* is omitted subsequently.

Model D adds *Option* to the *ITCols* and *ZTC* valuation/estimation method variables. While taking the expected positive sign, the coefficient on the *Option* variable fails to be significant even at the 10% level until the three forest site variables are added to produce Model E. Model E provides a substantial improvement in overall fit compared to the preceding two models. The degree of fit does not change substantially in remaining models, the first of which (Model F) replaces the site variable *Aberfoyle* with the author variable *Hanley,* with which it is correlated (all of the Hanley et al., studies were conducted at Aberfoyle although other authors also provide estimates for this forest).

The *Hanley* variable has a significant and positive coefficient, as per Table 4[16]. However, unlike the model reported in Table 4, Model F does not control for the OE variable within its

[15] Note that all TC studies relate to use value alone.

CV based estimates (this is because such a variable cannot apply to TC estimates) and we know from inspection (Bateman, 1996) that a majority of the CV studies conducted by Hanley et al., did not use the downwardly biased OE approach. However, the fact that *Hanley* is significant and positive in both Model F *and* Table 4 (which does control for elicitation method) appears to support the argument that valuation estimates may be subject to authorship effects. An alternative explanation is that the Hanley et al., estimates are elevated because of some characteristics of the Aberfoyle site for which they were estimated. Yet a further explanation might be that this result is in some way a product of the GLM modelling approach adopted in this meta-analysis. All of these possibilities are explored subsequently.

Model G adds the final variable *Year* into the analysis. As per our analyses of CV estimates this gave a small, positive coefficient which in this expanded dataset proved statistically significant. This is an interesting finding, which seems likely to reflect a relative increase in the perceived value of woodland recreation over this longer time period. The result is not particularly robust, becoming insignificant ($p = 0.181$) when the oldest estimate (that provided by Everett (1976)) is omitted, yet even then the sign and size of the coefficient remain similar ($\beta = 0.0526$). This suggests that, given a longer data period, a positive trend in valuations might become more clearly established. While emphasizing statistical uncertainties regarding this result, its general message seems plausible, suggesting an increasing relative interest in outdoor, environmentally based recreation over the last three decades and echoing the seminal work of Krutilla and Fisher (1975).

The other relationships detailed in Model G also conform well to expectations. Values are positively related to the *Option* variable which is now significant at the 5% level. Similarly the *Method* variables *ITCols* and *ZTC* both have significant and positive coefficients reflecting their expected relationship with the CV values which form the majority of the base case of this analysis. Neither of the remaining *Forest* variables are significant at $p<0.05$ although *Cheshire* is significant at $p<0.10$. The negative coefficient on this variable may reflect the high visitor congestion noted in studies of this forest (Willis and Benson, 1989)[17]. Model G also provides the best fit to our data and, given the desirable characteristics noted above, provides a typical example of a meta-analysis estimated using conventional statistical modelling approaches. We now consider an alternative to this approach and examine the extent to which this may provide superior insight into the nature and robustness of these postulated relationships.

4. An MLM approach to meta-analysis

The various models reported in Tables 4 and 5 all assume independence between estimates. However, in a recent meta analysis of CV studies of wetlands, Brouwer et al., (1999) use Multilevel Modelling (MLM) techniques (Goldstein, 1995) to relax this assumption and consider the possibility that valuation estimates are clustered within authors, not by the use of *Author* variables such as that used above, but instead by modelling the residual variance of estimates in two parts; that due to the effect of authors on study estimates, and that remaining due to true unexplained error. In effect, this approach allows for the possibility that variation within value estimates may differ between authors. In order to test this possibility we now undertake an MLM exercise considering both the variation of estimates between authors and that between forests.

---

[16] Note that controlling for the Method variables makes the Willis et al. values no longer significantly different from other estimates.

[17] By contrast the *Loch Awe* coefficient is positive (although not statistically significant) possibly reflecting its somewhat remote and secluded location attracting a more 'dedicated' woodland user (a model excluding the *Loch Awe* variable is presented in Bateman et al., 2000).

A potential limitation of the application of GLM techniques in meta-analysis occurs if the observations being modelled possess an inherent hierarchy. For traditional GLM estimation strategies, some of the variables used to predict recreation may be specific to each individual study (examples being the study design and elicitation method used). However, others, such as the forest in which the study took place or the characteristics of the author, will be constant across many of the published value estimates. These can be conceptualised as higher level variables, and in this sense the data may be viewed as possessing a hierarchical structure. The data structure from the above examples can be seen as actually corresponding to a range of hierarchical levels; of value estimates (level 1) within studies (level 2), of value estimates (level 1) within forests (level 2), or alternatively of value estimates (level 1) within authors (level 2)[18]. Given sufficient data, this hierarchy could be extended with further levels representing, for example regions or even nations.

Hierarchical data structures cannot be easily accommodated within the traditional GLM framework. Here, the values of author or study location related variables must be collapsed to the level of the individual value estimate and simply replicated across all observations sharing those characteristics. This procedure is problematic in that it provides no information on, for example, the probability of estimates made in the same forests, or by the same authors producing similar value estimates. This limitation may be circumvented, as employed in the examples above, by the use of dummy variables to indicate forest location or authorship. However, this solution can present difficulties. With the present data, there are only a limited number of authors and forest sites, and hence the number of dummy variables that need to be added to the models are manageable. However, it is readily apparent that any model estimated using dummies will quickly become extremely large and complex if the dataset contains numerous observations at each level of the hierarchy.[19]

An alternative to the use of dummy variables to model hierarchical data structures is to fit a series of separate regression models. For example, separate models could be fitted for each forest or author. However, this approach defeats the objective of meta-analyses when the variables found to be significant may differ between models. Furthermore, unreliable results may be produced due to small sample sizes when there are relatively few estimates for each forest, as in the present case.

Aside from methodological considerations, a further limitation of traditional GLM meta-analyses stems from the fact that they may contain poorly estimated parameters and standard errors (Skinner et al, 1989). Problems with standard error estimation arise due to the presence of intra-unit correlation; the fact that recreation value estimates from studies within the same forest, or by the same author, may be expected to be more similar than those drawn from a random sample. If intra-unit correlation is small, then reasonably good estimates of standard errors may be expected (Goldstein, 1995). However, where intra-unit correlation is significant then traditionally employed GLM strategies will tend to under-estimate standard errors, meaning that confidence intervals will be too short and significance tests will too often reject the null hypothesis.

[18] If no two authors undertake a study in the same forest, then this may be extended to a three level hierarchy of WTP estimates (level 1) within forests (level 2) within authors (level 3). If multiple authors do study the same forests, then a more complex structure (known as cross classified) exists wherein estimates (level 1) are nested within a cross-classified level (2) of forests and authors. Such a case is not considered here (although it is the subject of ongoing research by the authors), but the theory of cross classified hierarchies is discussed in detail by Goldstein (1995).

[19] An example might be an international dataset of value estimates nested within hundreds of study locations.

For simplicity, a two level hierarchy of *i* value estimates (at level 1) within *j* authors (at level 2) is considered in the examples below. As with a traditional generalized linear model, the observed responses y_{ij} are the published mean per person per visit recreation value estimates in 1990 pounds sterling. Considering a situation with just one explanatory variable, *OPTION* (defined as before) being tested, a simple model may be written as:

$$y_{ij} = \beta_{0j} + \beta_1 OPTION_{ij} + \in_{ij} \tag{1}$$

The subscript *i* takes the value from 1 to the number of value estimates in the model, and the subscript *j* takes the value from 1 to the number of authors in the sample. Using this notation, items with two subscripts *ij* vary from estimate to estimate. However, an item that has a *j* subscript only varies across authors but is constant for all the estimates made by each author. If an item has neither subscript it is constant across all studies and authors.

As the authors included in the analysis are treated as a random sample from a population, Equation (1) may be re-expressed as:

$$\beta_{0j} = \beta_0 + \mu_j$$

$$\hat{y}_{ij} = \beta_0 + \beta_1 OPTION_{ij} + \mu_j \tag{2}$$

Where β_0 is a constant and μ_j is the departure of the *j*-th author's intercept from the overall value. This means that it is an author level (level 2) residual that is the same for all estimates nested within an author. In other words this term describes, after holding constant the effect of the explanatory variables within the model, the residual influence of the author in determining the outcome for each individual mean WTP estimate they published.

The notations expressed in Equation (2) can be combined. Introducing an explanatory variable *cons*, which takes the value 1 for all estimates (and hence forms a constant or intercept term), and associating every term with an explanatory variable, the model becomes as shown in Equation (3):

$$y_{ij} = \beta_0 cons + \beta_1 OPTION_{ij} + \mu_{0j} cons + \in_{0ij} \tag{3}$$

Finally the coefficients can be collected together and written as:

$$\beta_{0ij} = \beta_0 + \mu_{0j} + \in_{0ij}$$

$$y_{ij} = \beta_{0ij} cons + \beta_1 OPTION_{ij} \tag{4}$$

In Equation (4), both μ_j (the level 2 or author level residuals) and $\in_{ij}$ (the level 1 or estimate level residuals) are random quantities whose means are estimated to be equal to zero. A comparison between the multilevel model expressed in Equations (3) and (4) and the original non-hierarchical structure depicted in Table 4 illustrates the tenet of multilevel models. Traditionally the residual error term of a model, $\in$, is seen as an annoyance and the aim of the modelling process is to minimize its size. With multilevel models the error term is of pivotal importance in model estimation. Rather than a single error term being estimated, it is stratified into a range of terms, each representing the residual variance present at each level of the hierarchy. Viewed in this sense, μ_j represents author level effects, whilst $\in_{ij}$ represents those operating at the level of the value estimate.

If, after holding constant the influence of the x_{ij} explanatory variables in the model, $\mu_j > \in_{ij}$, then this would suggest that some factors associated with the authors themselves are of greatest importance in explaining the residual variation in WTP estimates. If instead $\mu_j < \in_{ij}$ then some un-modelled factor associated with the elicitation of each estimate (which, for example, could be associated with the characteristics of each specific study, or might simply be random variation in each elicited WTP value) is more important. A common scenario is that, whilst both μ_j and $\in_{ij}$ are large in a model containing few x_{ij} explanatory variables, both will decrease as further explanatory variables are added and the residual variance in the model is explained.

The structure presented in Equation (4) is known as a variance components model (Lin, 1997). For ease of interpretation the estimated parameters may be classified as either being of a *fixed* or *random* nature. The fixed parameters are those for which just a single coefficient is estimated, and hence correspond to those that would be found in a traditional GLM. In this example both *CONS* and *OPTION* are fixed. In contrast, the random parameters are those where individual estimates are made for every unit at each level of the hierarchy. Here both μ_j and $\in_{ij}$ are random, as a value of $\in_{ij}$ is estimated for each value estimate (at level 1 of the model) and a value of μ_j is estimated for each author (at level 2 of the model). Hence, in terms of model interpretation, it is the stratification of the error term to form these random parameters that differentiates a multilevel model from more traditional regression analysis techniques. Remembering that $OPTION_{ij}$ is a dummy variable that represents whether the elicited WTP requested use plus option value (OPTION = 1) or use value alone (OPTION = 0), the variance components model depicts the relationship between OPTION and the value estimate as being constant, but (provided $\mu_j > 0$) recreation values are modelled as being higher for some authors than others.

Whilst there are various methods available for parameter estimation in multilevel models, an approach known as Iterative Generalized Least Squares (IGLS) was adopted in our subsequent analysis. The statistical theory underpinning IGLS is described in detail by Goldstein (1995). Briefly, initial estimates of the fixed parameters are derived by traditional GLM methodologies ignoring the higher-level random terms. The squared residuals from this initial fit are then regressed on a set of variables defining the stricture of the random part to provide initial estimates of the variances/covariances. These estimates are then used to provide revised estimates of the fixed part, which is in turn employed to revise the estimates of the random part, and so on until convergence. Crucially, a difficult estimation problem is decomposed into a sequence of linear regressions that can be solved efficiently and effectively, providing maximum-likelihood estimates[20].

It is important to note that the slopes and intercepts that are estimated for units within level 2 and above of the hierarchy will not be the same as those that would be obtained from a traditional generalized linear solution. They are in-fact residuals which have, to a greater or lesser extent, been shrunken towards the average regression line giving the predicted relationship between mean WTP and the explanatory variables across all authors. Taking our example of a 2-level model, at the author level, if $\sigma^2_{e0} = \mathrm{var}(\in_{0ij})$ and $\sigma^2_{u0} = \mathrm{var}(u_{0j})$ then each author level residual is estimated using Equation (5):

$$\hat{u}_j = \frac{n_j \sigma^2_{u0}}{n_j \sigma^2_{u0} + \sigma^2_{e0}} \tilde{y}_j \tag{5}$$

Here, n_j is the total number of estimates produced by author j, $\tilde{y}_j$ is the raw residual associated with the author (the mean estimate level residual for all estimates made by author j) and $\hat{u}_j$ is the shrunken residual. From this, it can be seen that if n_j is large and there are many value estimates made by an author, then the predicted level-2 residuals will be closer to the raw residual than when n_j is small. If n_j is small, then the residual will be shrunken towards the mean. Similarly if σ^2_{e0} is large and there is a lot of variability in the recreation value estimates produced by an author, then the predicted residual will also be shrunken. In this sense, the MLM approach provides conservative estimates of variability at different levels of the hierarchy where units based on a small sample or a very variable outcome are considered to provide little information. This is particularly pertinent here because, as has already been considered, the statistically significant positive coefficient observed for *Hanley* in Table 5 was based on studies that were all conducted at a single forest (Aberfoyle).

A multilevel re-analysis of the meta-analysis data was undertaken using the MLwiN package (Rasbash et al., 2000) developed by the Multilevel Models Project at the Institute of Education, London. Two sets of model were produced; one with a hierarchy of WTP estimates nested within authors, and one of estimates within study locations. The results of the model of estimates nested within authors are given in Table 6

Although technically different, the fixed parameters in the model in Table 6 can be interpreted in the same way as an ordinary regression. They confirm the findings of Tables 4 and 5 that the highest estimates of recreation value are derived from *ZTC* models while the lowest come from OE CV formats and that studies for which the *Option* dummy applies yield higher recreation value estimates.

One of the objectives of fitting a multilevel model was to determine if, after controlling for the variables in the fixed part of the model, there was still statistically significant variation in WTP estimates between authors. These random effects are shown in the lower part of Table 6. This part of the model is relatively simple. Although the multilevel methodology involves estimating a separate intercept value for each author (μ_j) and a separate residual for each value

[20] A limitation of IGLS for models with a binomial or Poisson distributed response variable (neither of which were used in the present application) is that is uses a method based on either marginal or penalized quasi-likelihood. This requires assumption of normally distributed variance above level one of the hierarchy.

estimate ($\in_{ij}$), the variance between the two levels of the model may be neatly summarized by the two parameters σ^2_{u0} and σ^2_{e0}. These are the same parameters used in the calculation of the shrinkage factor illustrated in Equation (5) and are known as variance parameters, as they indicate the variance in the μ_j and $\in_{ij}$ terms respectively. Hence a comparison of the values of σ^2_{u0} and σ^2_{e0} shows the relative importance of author (level 2) and estimate (level 1) effects in determining the variability of WTP values that is not explained by the fixed parameters in the model.

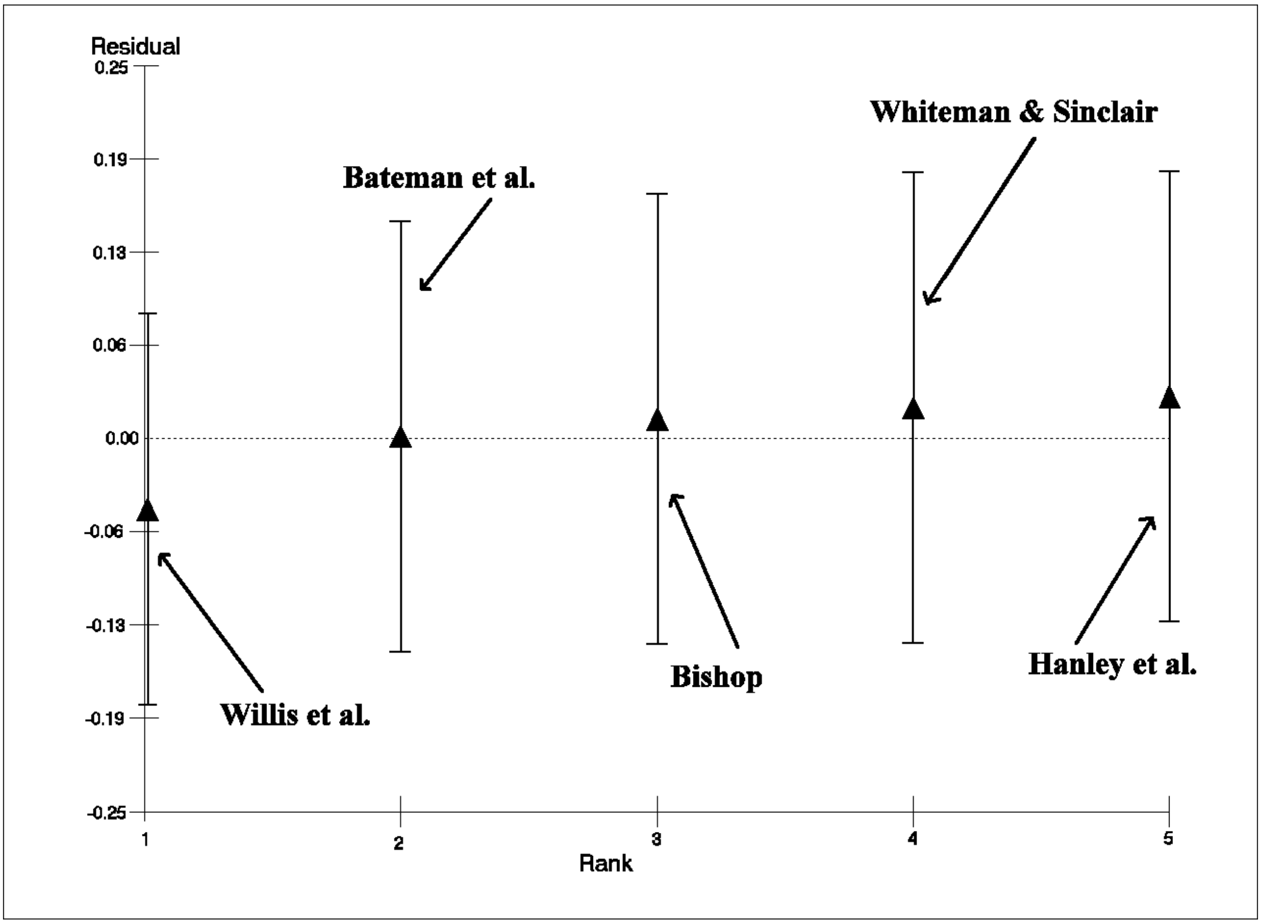
Table 6: MLM model estimates

Variable	**Coefficient**	**95% CI**	**P**
FIXED EFFECTS			
Cons	0.616	0.364 - 0.879	<0.001
Option	0.339	0.028 - 0.654	0.04
CVnonOE	0.706	0.243 - 1.176	<0.001
ITCols	0.813	0.442 - 1.198	<0.001
ITCml	0.052	-0.367 - 0.427	0.8
ZTC	1.837	1.520 - 2.153	<0.001
RANDOM (HIERARCHICAL) EFFECTS			
	Variance	**95% CI**	**P**
Level 1 (Value estimate) *Variance* σ^2_{e0}	0.229	0.156 - 0.313	<0.001
Level 2 (Author) *Variance* σ^2_{u0}	0.034	-0.154 - 0.021	0.7

The parameter estimates for both σ^2_{u0} and σ^2_{e0} are greater than zero, suggesting that variability between estimates and between authors remains after controlling for the explanatory variables that were included in the fixed part of the model. Taking the ratio of these estimates suggests that approximately 12% of unexplained variation in elicited recreation value is associated with author effects. However, the calculation of t-statistics for each coefficient shows that whilst statistically significant residual variation remains between estimates at level 1 ($t = 5.72$, $p<0.001$) the effect of authorship does not reach statistical significance ($t = 0.38$, $p>0.05$). In other words, the multilevel analysis suggests that an author effect is present but is not statistically significant.

Although in conflict with the earlier findings from the GLM analysis, such a result accords with theoretical expectations that recreation values should not vary significantly according to study authorship. This provides a substantial (if on its own insufficient) support for the practice of placing monetary values upon preferences for non-market environmental goods. The author specific results are illustrated in Figure 1 where the value of the intercept value u_j estimated for each individual author is presented in rank order along with corresponding 95% confidence intervals. The figure shows that, in the multilevel analysis, studies by *Hanley et al.* are still predicted to give the highest recreation values and those by *Willis et al.* the lowest. However, the confidence intervals now overlap. This represents a reduction in variance from the situation observed in Table 4 where both estimates provided by *Hanley et al.* and *Willis et al*. were found to significantly differ from that of other authors. The reduction of variance is due to the effects of the conservative estimation strategy implemented in Equation (5) where residuals are shrunken towards the mean value.

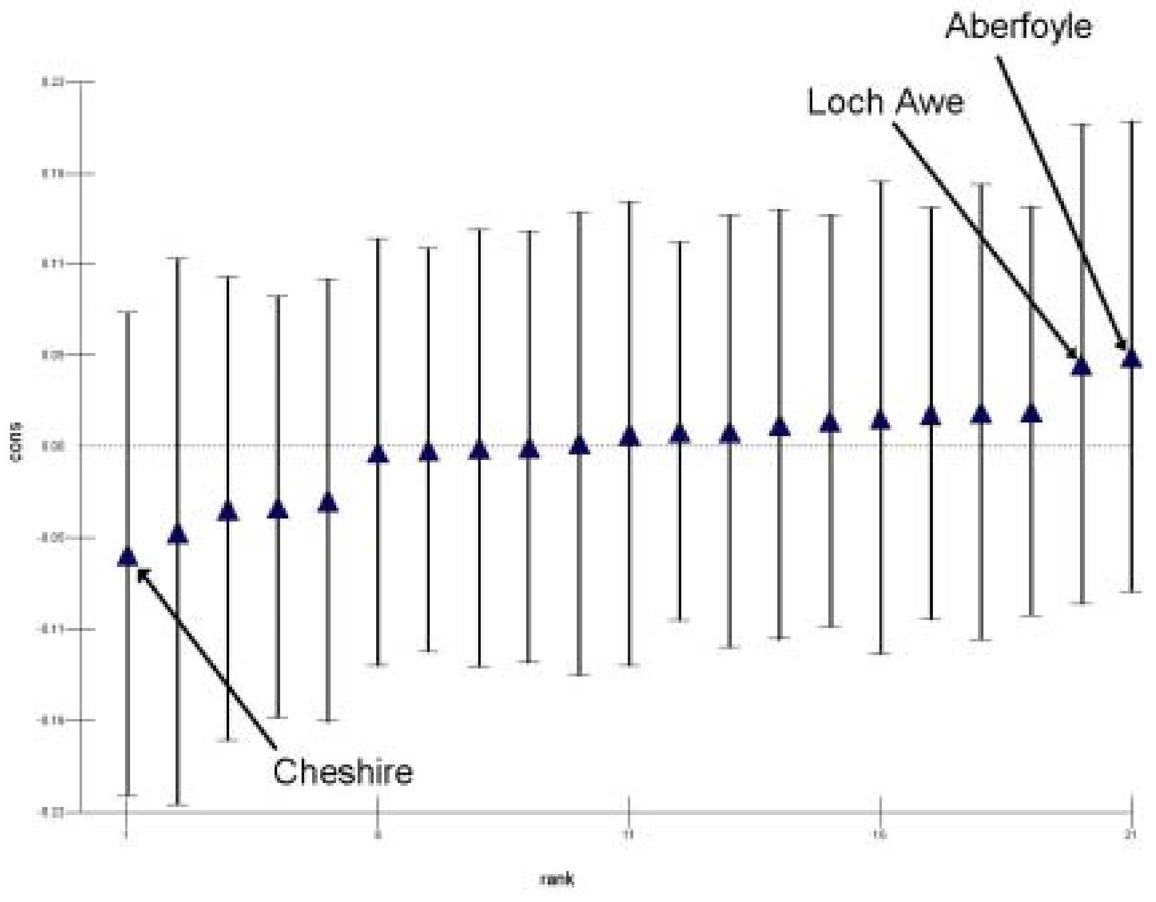
Figure 1: MLM author level residuals



The shrinkage illustrated by Figure 1 has interesting implications for the comparison of results between multilevel and non-multilevel models. The message from the multilevel model is that variation is present between authors but, because of the magnitude of the variance and the size of the sample, it cannot be said to be statistically significant. Hence we are making a statement about the importance of context (in this case authorship) and composition (the remaining unexplained variation in between WTP estimates). The traditional GLM approach used previously did the opposite; it told us little about the overall roles of context and composition, but it did highlight two authors with rather different patterns of responses from the rest of the sample. From this comparison, it is clear that, whilst the conclusions reached may be different from those of a traditional GLM analysis, the multilevel approach is prudent if the intention of the analysis is to quantify whether there are overall contextual influences (in this case associated with different authors) on the measured outcome (recreation value).

The earlier GLM analyses also found evidence of a *Forest* (site) effect where recreation values for *Cheshire* were somewhat lower than the rest of the sample, and those for *Loch Awe* and *Aberfoyle* were relatively higher (see Table 5). To test if any evidence of between-site heterogeneity remained after a multilevel approach was taken, the model presented in Table 6 was refitted, but this time authorship at level 2 was replaced by *Forest* identifiers. The fixed effect coefficient values and levels of significance were not found to differ greatly from the previous example and are hence not replicated here. However in this case the values of σ^2_{u0} (now for forests) and σ^2_{e0} (for value estimates) were estimated at 0.011 (t = 0.73, p>0.05), and 0.225 (t = 5.63, p<0.01) respectively. In similar fashion to the model for authors, these results show strong variation between estimates, but only a limited forest site effect (accounting for under 5% of the total residual variance). Figure 2 shows the forest level residuals ranked with 95% confidence intervals. In order to maintain legibility only those forests mentioned previously are identified. As with the original non-multilevel analysis, *Cheshire* shows the greatest negative residual (and hence correspondingly lower than predicted WTP values), whilst *Loch Awe* and *Aberfoyle* yield the highest positive residual values.

Figure 2: Forest level residuals



5. Discussion and Conclusions

Previously we have suggested a number of ways in which benefit transfer research may be taken forward (Bateman et al., 2000). These include improvements in the conduct and reporting of new studies, the specific incorporation of benefit transfer requirements within their design, and the reanalysis of past work. The present paper goes some way towards highlighting a novel way in which this latter aim might be best realized. We have compared the application of traditional GLM and novel MLM methodologies to meta-analyses of British woodland recreation values. Both sets of results generally conform well to expectations derived from either theoretical considerations or empirical regularities. However, one of the consistent messages derived from these results is the substantial impact of design choices (such as the elicitation technique, or the choice of valuation and/or estimation method) upon derived value estimates. This provides a cautionary note to the wider interpretability of any single study.

Our GLM findings suggest that certain authors and forests are associated with large recreation value residuals. However, the more sophisticated and conservative MLM approach shows that these residuals are not large enough (or are not based on a large enough sample size) to be differentiated from variation that might be expected by chance. Here we have fitted only simple two level models. More complex structures have not been implemented here for a number of reasons. No significant variation was observed between authors or survey site locations, and it is highly unlikely that a more detailed model hierarchy would have contradicted these findings. A second limitation to the use of more complex hierarchies concerns sample size; as models become more complicated there is an associated loss of degrees of freedom. In particular, the conservative estimation strategy used means that the presence of a small amount of level 2 variation in a simple two-level model may be shrunken to zero if either a more complex structure is attempted. Whilst the dataset we have studied is comprehensive, it is a based on a sample of just 77 observations, and hence has somewhat limited power. The increased number of observations that will result from more studies being undertaken will allow a greater complexity of models to be fitted.

Although the essential ideas of multilevel models were developed over 20 years ago, it is only recently that improvements in computing power and advances in our understanding of effective model implementation have meant that their execution has become a practical proposition (Bull et al., 1998). We are currently on a wave of innovation as use spreads from the original developers to the wider research community. Having said that, the multilevel approach retains some of the limitations of more traditional quantitative techniques, as well as introducing new ones.

In the MLM models presented here, influences on recreation values are modelled more powerfully than traditional techniques allow, yet the random parameters can ultimately offer only limited insight into the reasons behind between-author and between-forest variations in outcome. Preferences for complex, non-market environmental goods such as open-access recreation involve a detailed interplay between a wide range of factors that are difficult to quantify and may be subject to random variation. This unpredictability will undoubtedly introduce uncertainty into any model, multilevel or not, developed to identify and predict the important influences on such preferences. However, whilst multilevel models cannot remove this uncertainty, they can allow it to be more richly quantified and accounted for, and hence allow for systematic factors to be assessed.

Finally, considering the case at hand, our MLM estimated meta-analysis has some clear messages for policy makers within the UK Forestry Commission. While our results suggest that different authors provide consistent estimates of woodland recreation values, those values themselves do not appear particularly responsive to choice of study site. This finding is in line with other research showing that, while visitor arrivals at UK woodlands are highly responsive to a variety of locational factors, they are somewhat less responsive to the facilities on offer at these sites (Brainard et al., 1999; 2001)[21]. Given this, the onus upon woodland policy makers within the UK context, appears to be upon using scarce resources to optimise site location rather than to extend the diversity of facilities within woodlands.

References

Baaijens, S. R., Nijkamp, P. and Montfort, K.V. (1998) Explanatory meta-analysis for the comparison and transfer of regional tourism income multipliers, *Regional Studies*, 32 (9): 839-849.

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., Brainard, J.S., Garrod, G.D. and Lovett, A.A. (1999a) The impact of journey origin specification and other measurement assumptions upon individual travel cost estimates of consumer surplus: a geographical information systems analysis, *Regional Environmental Change*, 1(1): 24-30.

Bateman, I.J., Carson, R.T., Day, B., Hanemann, W.M., Hanley, N., Hett, T., Jones-Lee, M., Loomes, G., Mourato, S., Ozdemirogla, E., Pearce, D.W., Sugden, R. and Swanson, J. (2002) *Guidelines for the Use of Expressed Preference Methods for the Valuation of Preferences for Non-market Goods*, Edward Elgar Publishing (publication date: July 2002).

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

Bateman, I.J., Jones, A.P., Nishikawa, N. and Brouwer, R. (2000) Benefits transfer in theory and practice: A review, *CSERGE Global Environmental Change Working Paper GEC2000-25,* Centre for Social and Economic Research on the Global Environment, University of East Anglia and University College London.

Bateman, I.J. and Langford, I.H. (1997) Budget constraint, temporal and ordering effects in contingent valuation studies, *Environment and Planning A*, 29(7): 1215-1228.

Bateman, I.J. and Lovett, A.A., (2000) Valuing and mapping woodland access potential, *Quarterly Journal of Forestry*, 94(3), 215-222.

Bateman, I. J., Lovett, A.A. and Brainard, J.S. (forthcoming) *Applied Environmental Economics: A GIS Approach to Cost Benefit Analysis*, Cambridge University Press, Cambridge.

Bateman, I.J., Lovett, A.A. and Brainard, J.S. (1999b) Developing a methodology for benefit transfers using geographical information systems: modelling demand for woodland recreation, *Regional Studies*, 33(3): 191-205.

[21] This is not to suggest that site facilities are irrelevant in attracting visitors. However, as shown by Brainard *et al.*, locational factors provide much stronger predictors of demand. In part this may be because virtually all UK sites provide the basic walking and recreational amenities which characterize woodlands visits and are thus relatively little differentiated in terms of further cogent attributes.

Bateman, I.J., Willis K.G. (eds.) (1999) *Valuing Environmental Preferences: Theory and Practice of the Contingent Valuation method in the US, EU and Developing Countries*, Oxford University Press.

Benson, J.F. and Willis, K.G. (1990) The aggregate value of the non-priced recreation benefits of the Forestry Commission estate. *Report to the Forestry Commission*, Department of Town and County Planning, University of Newcastle upon Tyne.

Benson, J.F. and Willis, K.G. (1992) Valuing informal recreation on the Forestry Commission estate, *Bulletin 104*, Forestry Commission, Edinburgh.

Bergland, O., Magnussen, K. and Navrud, S. (1995) Benefit transfer: testing for accuracy and reliability, *Discussion Paper #D-03/1995*, Agricultural University of Norway.

Bhat C.R., 2000. A multi-level cross-classified model for discrete response variables. *Transportation Research Part B* 34 567-582.

Bishop, K.D. (1992) Assessing the benefits of community forests: an evaluation of the Recreational use benefits of two urban fringe woodlands, *Journal of Environmental Planning and Management*, 35(1):63-76.

Boadway, R. and Bruce, N. (1984) *Welfare Economics*, Basil Blackwell, Oxford.

Bockstael, N.E., McConnell, K.E. and Strand, I.E. (1991) Recreation, in Braden, J.B. and Kolstad, C.D. (eds.) *Measuring the Demand for Environmental Quality*, North-Holland, Elsevier Science Publishers, Amsterdam.

Boxall et al, 1996).

Boyle, K. J. (n.d.) *The practical implications of accomplishing an ideal benefits transfer with limited data availability*, Department of Resource Economics and Policy, University of Maine, Orono.

Boyle, K. J. & Bergstrom, J. C. (1992) Benefit transfer studies: myths, pragmatism, and idealism, *Water Resources Research*, 28(3):657-663

Boyle, K.J., Poe, G.L. and Bergstrom, J. (1994) What do we know about groundwater values?: Preliminary implications from a meta-analysis of contingent valuation studies, *American Journal of Agricultural Economics*, 76(5), 1055-1061.

Brainard, J.S., Bateman, I.J. and Lovett, A.A., (2001) Modelling demand for recreation in English woodlands, *Forestry*, 74(5): 423-438.

Brainard, J.S., Lovett, A.A. and Bateman, I.J. (1999) Integrating geographical information systems into travel cost analysis and benefit transfer, *International Journal of Geographical Information Systems*, 13(3): 227-246.

Brookshire, D. S. and Neill, H. R. (1992) Benefit transfers: conceptual and empirical issues, *Water Resources Research*, 28(3):651-655.

Brouwer, R., Langford, I.H., Bateman, I.J., and Turner, R.K., (1999) A meta-analysis of wetland contingent valuation studies, *Regional Environmental Change*, 1(1): 47-57.

Brouwer, R. and Spaninks, F.A. (1999). The validity of environmental benefits transfer: further empirical testing. *Environmental and Resource Economics*, 14(1): 95-117.

Bryk, A. S., Raudenbush, S. W., 1992. *Hierarchical Linear Models*. Sage, Newbury Park.

Bull, J. M., Riley, G. D., Rasbash, J. and Goldstein, H., (1998) *Parallel implementation of a multilevel modelling package*, University of Manchester, Manchester.

Button, K.J. (1995) Evaluation of transport externalities: what can we learn using meta-analysis?, *Regional Studies* **29**, 507-517.

Button, K.J. and Kerr, J. (1996) Effectiveness of traffic restraint policies: a simple meta-regression analysis, *International Journal of Transport Economics* **23**, 213-225.

Carson, et al., Flores, N. E., Martin, K. M. and Wright, J. L. (1996) Contingent valuation and revealed preference methodologies: comparing the estimates for quasi-public goods, *Land Economics*, 72:80-99.

Carson, R.T., Groves, T. and Machina, M.J., (1999) Incentive and informational properties of preference questions, *Plenary Address, Ninth Annual Conference of the European Association of Environmental and Resource Economists (EAERE),* Oslo, Norway, June 1999.

Cesario, F.J. (1976) Value of time in recreation benefit studies, *Land Economics*, 55:32-41.

CSO (Central Statistical Office) (1998) Regional Trends, HMSO, London.

Desvousges, W. H., Johnson, F.R. and Banzaf, H.S. (1998) *Environmental Policy Analysis With Limited Information: Principles and Applications of the Transfer Method*, Edward Elgar, Northampton, Mass.

Desvousges, W. H., Naughton, M. C. & Parsons, G. R. (1992) Benefit transfer: conceptual problems in estimating water quality benefits using existing studies, *Water Resources Research*, 28(3):675-683

Downing, M. and Ozuna, T. (1996) Testing the reliability of the benefit function transfer approach, *Journal of Environmental Economics and Management*, 30: 316-322.

Duncan, C., Jones, K., Moon, G., 1998. Context, composition, and heterogeneity: using multilevel models in health research. *Social Science & Medicine* 46 (1), 97-117.

Dunteman and George, 1989

Espey, M., Espey, J. and Shaw, W.D. (1997) Price elasticity of residential demand for water: A meta-analysis, *Water Resources Research*, 33(6): 1369-1374.

Everett, R.D. (1979) The monetary value of the recreational benefits of wildlife, *Journal of Environmental Management*, 8:203-213.

Forestry Commission, 2001 Website:- http://www.forestry.gov.uk

Freeman, A. M. III (1979) *The benefits of environmental improvement: theory and practice*, Resources for the Future. John Hopkins University Press, Baltimore and London.

Freeman, A. M. III. (1993) *The Measurement of Environmental and Resource Values: Theory and Methods*. Resources for the Future: Washington, D. C.

FICGB (Forestry Industry Committee of Great Britain), (1992) *The Forestry Industry Year-Book 1991-92*, FICGB, London.

Fix P, Loomis J, Eichhorn R (2000) Endogenously chosen travel costs and the travel cost model: an application to mountain biking at Moab, Utah. *Applied Economics*, 32(10): 1227-1231.

Garrod, G.D. and Willis, K.G. (forthcoming). Using Contingent Ranking to Estimate the Loss of Amenity Value for Inland Waterways from Public Utility Service Structures. *Environmental and Resource Economics*, in press.

Gibson, J.G. (1978) Recreation land use, in Pearce, D.W. (ed.). *The Valuation of Social Cost*, Allen and Unwin, London.

Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall, London.

Glass, G.V. (1976) Primary, secondary and meta-analysis of research, *Educational Researcher*, 5(10): 3-8.

Glass, G.V., McGaw, B. and Smith, M.L. (1981) *Meta-analysis in social research*, Sage Publications, Beverley Hills, CA.

Goldstein, H., 1995. Multilevel Statistical Models (2nd edition). Edward Arnold, London.

Goldstein, H., Rasbash, J., Yang, M., Woodhouse.G, Pan. H., Nuttall, D., Thomas, S., 1993. A multilevel analysis of school examination results. *Oxford Review of Education* 19, 425-433.

Grayson, A.J., Sidaway, R.M. and Thompson, F.P. (1975) Some aspects of recreation planning in the Forestry Commission in Searle, G.A.C. (ed.). *Recreational Economics and Analysis: Papers presented at the Symposium on Recreational Economics and Analysis, London Graduate School of Business Studies, January 1972*, Longman, Essex.

Green, D., Jacowitz, K.E., Kahneman, D. and McFadden, D. (1998) Referendum contingent valuation, anchoring, and willingness to pay for public goods, *Resource and Energy Economics*, 20 (2): 85-116.

Hanley, N.D. (1989) Valuing rural recreation benefits: an empirical comparison of two approaches, *Journal of Agricultural Economics*, 40(3):361-374.

Hanley, N.D. and Craig, S. (1991) Wilderness development decisions and the Krutilla-Fisher model: the case of Scotland's Flow Country. *Ecological Economics*, 4:145-164.

Hanley, N.D. and Ecotec Ltd (1991) *The Valuation of Environmental Effects: Stage Two Final Report*. The Scottish Office Industry Department and Scottish Enterprise, Edinburgh.

Hanley, N.D. and Munro, A. (1991) Design bias in contingent valuation studies: the impact of information, *Discussion Paper in Economics 91/13*, Department of Economics, University of Stirling.

Hanley, N.D. and Ruffell, R.J. (1991) Recreational use values of woodland features. *Report to the Forestry Commission*, University of Stirling.

Hanley and Spash (1994) Cost-Benefit Analysis and the Environment, Edward Elgar.

Herriges, J.A. and Kling C.L. (eds.) (1999) *Valuing Recreation And The Environment: Revealed Preference Methods in Theory and Practice*, Edward Elgar Publishing, Cheltenham.

H.M. Government (1995) *Environment Act 1995*, HMSO, London.

H.M. Treasury (1972) *Forestry in Great Britain: An Interdepartmental Cost/Benefit Study*, HMSO, London.

Hoehn, J. P. and Randall, A. (1987) A satisfactory benefit cost indicator from contingent valuation, *Journal of Environmental Economics and Management*, 14(3):226-247.

Hoijtink, H., 1998. *Bayesian and Bootstrap Approaches to Testing Homogeneity and Normality in a Simple Multilevel Model*. Utrecht University, Utrecht.

Hufschmidt, M.M., James, D.E., Meister, A.D., Bower, B.T. and Dixon, J.A. (1983) *Environment, Natural Systems and Development: An Economic Valuation Guide*. John Hopkins University Press, Baltimore and London.

Hunter, J. E., Schmidt, F. L. and Jackson, G. (1982) *Advanced Meta-analysis: Quantitative Methods for Cumulating Research Findings A Cross Studies*, Sage, Beverly Hills.

Jones, A.P., Bateman, I.J. and Wright, J. (2000) *Predicting and Valuing Informal Recreation Use of Inland Waterways: Report to British Waterways*, CSERGE, University of East Anglia.

Just, R.E., Hueth, D.L. and Schmitz, A. (1982) *Applied Welfare Economics and Public Policy*, Prentice Hall, Englewood Cliffs, N.J.

Kirchhoff, S., Colby, B. G. and LaFrance, J. T. (1997) Evaluating the performance of benefit transfer: an empirical inquiry, *Journal of Environmental Economics and Management*, 33:75-93.

Korn, E. L., Graubard, B. I., 1999. *Analysis of Health Surveys*. John Wiley, New York.

Krupnick, A. J. (1993) Benefit transfers and valuation of environmental improvements, *Resources: Resources for the Future*, 110:1-7

Krutilla, J.V. and Fisher, A.C. (1975) *The Economics of Natural Environments: Studies in the Valuation of Commodity and Amenity Resources*, Johns Hopkins University Press (for Resources for the Future), Baltimore, MD.

Lin, X., 1997. Variance component testing in generalised linear models with random effects. *Biometrika* 84, 309-25.

Liston-Heyes and Heyes, (1999). Recreational benefits from the Dartmoor National Park. Journal *of Environmental Management* 55, 69-80.

Loomis, J. B. (1992) The evolution of a more rigorous approach to benefit transfer: benefit function transfer, *Water Resources Research*, 28(3):701-705.

Loomis, J.B., Roach, B., Ward, F. and Ready, R. (1995) Testing transferability of recreation demand models across regions: A study of Corps of Engineers reservoirs, *Water Resources Research*, 31: 721-730.

Loomis, J.B. and White, D.S. (1996) Economic benefits of rare and endangered species: Summary and meta-analysis, *Ecological Economics*, 18 (3): 197-206.

Longford, N. T., 1993. *Random Coefficient Models*. Clarendon Press, Oxford.

Lovett, A. A., (1984) Poisson Regression using the GLIM package, *Computer Package Guide No. 5*, Department of Geography, University of Lancaster.

Lovett, A.A., Brainard, J.S. and Bateman, I.J. (1997) Improving benefit transfer demand functions: a GIS approach, *Journal of Environmental Management*, 51: 373-389.

Lovett, A. A. and Flowerdew. R., (1989) Analysis of Count Data Using Poisson Regression, *Professional Geographer*, 41 (2), 1989, pp. 190-198.

Maddala, G.S., (1983) *Limited-Dependent and Qualitative Variables in Econometrics*, Cambridge University Press, Cambridge.

Markowski, M. A., Boyle, K.J., Bishop, R.C., Larson, D.M. and Paterson, R.W. (2001) A cautionary note on interpreting meta analyses, unpublished paper, Industrial Economics Inc.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models, 2nd Edition*, Monographs on Statistics and Applied Probability 37, Chapman and Hall, London.

Menard, S., (1995). Applied Logistic Regression Analysis, *Quantitative Applications in the Social Sciences*, Vol 106. Sage Publications, New York.

Mendelsohn, R., Hof, J., Peterson, G. and Johnson, R. (1992) Measuring recreation values with multiple destination trips, *American Journal of Agricultural Economics*, 24(4): 926-933.

Mitchell, R.C. and Carson, R.T. (1989) *Using Surveys To Value Public Goods: The Contingent Valuation Method*, Washington D.C., Resources for the future.

NAO (National Audit Office), (1986) *Review of Forestry Commission Objectives and Achievements*. Report by the Comptroller and Auditor General, National Audit Office, HMSO, London.

Nelson, J.P. (1980) Airports and property values: a survey of recent evidence, *Journal of Transport Economics and Policy* **19**, 37-52.

OECD (Organisation for Economic Co-operation and Development), (1994) *Project and policy appraisal: integrating economics and environment*, OECD, Paris.

Parsons, G.R. (1991) A note on choice of residential location in travel cost demand models, *Land Economics*, 67:360-364.

Parsons, G.R. and Kealy, M.J. (1994) Benefits transfer in a random utility model of recreation, *Water Resources Research*, 30(8):2477-2484.

Pearce, D.W. and Moran, D. (1994) *The Economic Value of Biodiversity*, Earthscan, London.

Pearce, D.W. and Turner, R.K. (1990) *Economics of Natural Resources and the Environment*, Harvester Wheatsheaf, Hemel Hempstead.

Poe, G.L., Boyle, K.J. and Bergstrom, J.C. (2001) A preliminary meta analysis of contingent values for ground water revisited, in Bergstrom, J.C., Boyle, K.J. and Poe, G.L. (eds.) *The Economic Value of Water Quality*, Edward Elgar, Northampton, MA.

Rackham, O. (1976) *Trees and Woodland in the British Landscape*, Dent, London.

Randall, A. (1994) A difficulty with the travel cost method, *Land Economics*, 70(1): 88-96.

Rasbash. J., Browne, W., Goldstein, H., Yang, M., Plewis, I., Healy, M., Woodhouse, G., Draper, D., Langford, I., Lewis, T., (2000). *A Users Guide to MlwinN: Version 2.1*, Institute of Education, London.

Rosenberger, R.S. and Loomis, J.B. (2000) Using meta-analysis for benefit transfer: In-sample convergent validity tests of an outdoor recreation database, *Water Resources Research*, 36(4): 1097-1107.

Rosenthal, D.H., Donnelly, D.M., Schiffhauer, M.B. and Brink, G.E. (1986) User's guide to RMTCM: software for travel cost analysis, *General Technical Report RM-132*, United States Department of Agriculture: Forest Service, Rocky Mountain Forest and Range Experiment Station, Fort Collins, Colorado.

Schwartz, J. (1994) Air pollution and daily mortality: a review and a meta analysis, *Environmental Economics* **64**, 36-52.

Shrestha, R. K. and Loomis, J.B. (2001) Testing a meta-analysis model for benefit transfer in international outdoor recreation, *Ecological Economics*, 39: 67-83.

Skinner, C.J., Holt, D., Smith, T.M.F. (1989) *Analysis of Complex Surveys*. Wiley, Chichester, UK.

Smith, V.K. (1989) Can we measure the economic value of environmental amenities?, *Southern Economic Journal* **56**, 865-878.

Smith, V.K. and Huang, J.C. (1993) Hedonic models and air pollution: Twenty five years and counting, *Environmental and Resource Economics* **3**, 381-394.

Smith, V. K. and Huang, J. C. (1995) Can markets value air quality? A meta-analysis of hedonic property values models, *Journal of Political Economy*, 103:209-227.

Smith, V. K. and Kaoru, Y. (1990a) Signals or noise? Explaining the variation in recreation benefit estimates, *American Journal of Agricultural Economics*, May 1990: 419-433.

Smith, V.K. and Kaoru, Y. (1990b) What have we learned since Hotelling's letter? A meta-analysis, *Economics Letters*, 32: 267-272.

Smith, V.K. and Osborne, L. (1996) Do contingent valuation estimates pass a "Scope" test? A meta-analysis, *Journal of Environmental Economics and Management* **31**, 287-301.

Smith, V.K. and Pattanayak, S.K. (forthcoming) Is meta-analysis a Noah's Ark for non-market valuation?, *Environmental and Resource Economics*, in press.

Sturtevant, L.A., Johnson, F.R. and Desvousges, W.H. (1995). A meta-analysis of recreational fishing. *Triangle Economic Research*, Durham, North Carolina.

Van den Bergh, J. C. J. M., Button, K. J., Nijkamp, P. and Pepping, G. C. (1997) , *Meta-Analysis in Environmental Economics*, Kluwer Academic Publishers, AH Dordrecht, The Netherlands.

Van Houtven, G.L., Pattanayak, S.K., Pringle, C. and Yang, J-C. (2001) Review and meta-analysis of water quality valuation studies, Draft Report, prepared for U.S. Environmental Protection Agency, Office of Water and Office of Policy, Economics, and Innovation, Washington, D.C.

Walsh, R. G., Johnson, D. M. and McKean, J. R. (1992) Benefit transfer of outdoor recreation demand studies, 1968-1988, *Water Resources Research*, 28(3):707-713.

Walsh, R. G., Johnson, D. M. and McKean, J. R. (1990), Nonmarket values from two decades of research on recreation demand, in Link, A. and Smith, V.K. (eds.) *Advances in Applied Micro-Economics*, Vol. 5: 167-193, JAI Press, Greenwich, CT.

Waters, W.G. (1993) Variations in the value of travel time savings: empirical studies and the values for road project evaluation, *mimeo*, cited in Van den Bergh, J. C. J. M., Button, K. J., Nijkamp, P. and Pepping, G. C. (1997), *Meta-Analysis in Environmental Economics*, Kluwer Academic Publishers, AH Dordrecht, The Netherlands.

Weisbrod, B.A. (1964) Collective - consumption services of individual consumption goods, *Quarterly Journal of Economics*, 78:471-477.

Whiteman, A. and Sinclair, J. (1994) *The Costs and Benefits of Planting Three Community Forests: Forest of Mercia, Thames Chase and Great North Forest*, Policy Studies Division, Forestry Commission, Edinburgh.

Woodward, R. T. and Wui, Y-S (2001) The economic value of wetland services: A meta-analysis, *Ecological Economics*, 37: 257-270.

World Health Organisation (WHO) (1997) *Health and environment in sustainable development: Five years after the earth summit*, WHO, Geneva.

Willis, K.G. and Benson, J.F. (1989) Values of user benefits of forest recreation: some further site surveys. *Report to the Forestry Commission*, Department of Town and County Planning. University of Newcastle upon Tyne.

Willis K.G., Benson, J.F and Whitby, M.C. (1988) Values of user benefits of forest recreation and wildlife. *Report to the Forestry Commission*, Department of Town and County Planning, University of Newcastle upon Tyne.

Willis, K.G. and Garrod, G.D. (1991) An individual travel cost method of evaluating forest recreation, *Journal of Agricultural Economics*, 42(1):33-41.

Willis, K.G. and Garrod, G.D. (1992). On-Site Recreation Surveys and Selection Effects: Valuing Open Access Recreation on Inland Waterways, in J. Van der Stratton and H. Briassoulis (eds.) *Tourism and the Environment: Regional, Economic and Policy Issues*, Kluwer Academic Publishers.

Willis, K. & Garrod, G. (1994) Transferability of benefit estimates, *in* Willis, K. G. & Corkindale, J. T. (eds.) *Environmental Valuation: New Perspective,* CAB International

Wolf, F. (1986) Meta-Analysis: Quantitative Methods for Research Synthesis, Sage, Beverly Hills.

APPENDIX A: MODELLING ANNUAL VISITS (FROM BATEMAN, 1996)

OVERVIEW:

The study by Bateman (1996) of visitors to Lynford Stag, a recreational site within Thetford Forest, East Anglia, included a previously unpublished appendix concerning the modelling of annual visits from a visitor survey occurring at one particular period in the year. Given the relevance of this work to the present study (and its unpublished nature) this analysis is reported in full within this appendix.

A.0 INTRODUCTION

Our estimated arrivals function only relates to those visitors who were interviewed during those days which were sampled during the survey period. If we wish to extrapolate our arrivals function to estimate annual arrivals we need to take account of the following:

i. Visits which occur while interviewers were occupied with other visitors or which occur outside interview hours;
ii. Visits which occur on non-sampled days during the survey period;
iii. Visits which occur outside the survey period.

In the following sections we make all of the adjustments outlined above and in so doing develop a model of annual visitation pattern at the Thetford site. This allows us to extrapolate our arrivals function onto an annual basis. This *adjusted annual arrivals function* is subsequently used to predict per annum visit totals for five sites in Wales for which information on actual arrivals is available, thus permitting an actual versus predicted validation test of the applicability of our adjusted annual arrivals function to other sites.

A.1: ALLOWING FOR NON-INTERVIEWED VISITORS DURING SURVEY DAYS

The 1993 Thetford Forest survey interviewed 351 parties over 17 survey days (one of which was curtailed due to poor weather) spread across the period 26.3.93 to 25.4.93. From 1.4.93 an electronic induction loop car counter operated at the site giving accurate information regarding the number of party visits per week. Table A1 details visit and survey data for the overlapping period of 12 days[22].

Table A1: Overlap of interview days and electronic counter operation days

Overlap period	Counter days (1)	Interview days (2)	Interviews (No. of parties) (3)	Cars (4)	(4)* [(2)/(1)] =(5)	Interview rate [(3)/(5)] (%)
1.4.93-4.4.93	4	2	45	658	329	13.6778
5.4.93-11.4.93	7	6	103	844	723	14.2378
12.4.93-19.4.93	7	2	59	1436	410	14.3802
19.4.93-26.4.93	7	2	50	1099	314	15.9236
Totals	25	12	257	4037	1776	14.4707

Table A1 shows that we achieved just over a 14% interview rate on the 12 days for which data is available. Assuming that this rate also applies for the full 16 effective days which were sampled then multiplying our total sample size by a factor of 1/0.144707 gives our best estimate of the total number of visitors during those days of the

[22]A potential problem would arise at sites with a large number of pedestrian visitors not arriving by car. As confirmed by our survey this is not a significant problem at Lynford Stag which can only be reached on foot by a lengthy walk.

survey on which interviewing took place. Therefore the estimated number of arrivals during those 16 days is as follows:

$$\text{Estimated arrivals (parties)} = 351 * \frac{1}{0.144707}$$

$$= 351 * 6.9105$$

$$= 2426$$

However, the sample period was not evenly distributed throughout the days of the week. In order to increase sample size, 7 of the sample days were on weekends (3 Saturdays and 4 Sundays). We therefore need to examine whether arrival rates are significantly larger on weekend days than weekdays as, if they are, then our arrivals function will overestimate visitors.

Table A2 shows on which days the survey was conducted and the number of interviews on each day. Average interview numbers are calculated in the final column. The average number of interviews/day over all days was 20.16 ($\sigma = 5.68$). Weekend days did record a higher interview rate of 22.55 interviews/day ($\sigma = 2.75$) compared to a mean for weekdays of 19.20 interviews/day ($\sigma = 6.24$). However, upon testing, this difference was found to be highly insignificant ($t = 0.62$).

Table A2: Interview rates across survey days

Day	Interviews (No. of parties)	Cumulative interview count	% of total sample	Cumulative % of total sample	Survey (No. of days)	Average interviews/day
Mon	29	29	8.26	8.26	1	29.0
Tues	32	61	9.12	17.38	2	16.0
Wed	20	81	5.70	23.08	1	20.0
Thurs	10	91	2.85	25.93	1	10.0
Fri	105	196	29.91	55.84	5	21.0
Sat	76	272	21.65	77.49	3	25.3
Sun	79	351	22.51	100.00	4	19.8
N = 351						mean=20.16 $\sigma = 5.68$

One possible complicating factor was very adverse weather conditions on one of the weekend sample days. Removing this from the dataset raised the weekend day mean to 25.65 ($s = 0.35$). However, whilst this increased the overall apparent contrast between weekend and weekday distributions this difference remained statistically insignificant ($t = 1.26$) and, as Britain is no stranger to adverse weather we feel that our initial findings are more defensible.

In summary we can conclude that our inflation factor of 6.9105 is unbiased in relating survey day interviews to the total number of parties visiting per survey day.

A.2: ALLOWING FOR NON-SURVEYED DAYS DURING THE SURVEY PERIOD

Ultimately we need to relate arrivals during our sample period to annual arrivals. As arrivals data is recorded upon a weekly rather than daily basis it will be convenient to convert our 16 day estimate to one which relates to the entire encompassing five week (35 day) period. Given the above investigation, a justifiable and simple conversion is to multiply by a factor of 35/16. Therefore our party arrivals estimate for a five week period from late March to the end of April 1993 is as follows:

$$= 351 * 6.9105 * 2.1875$$

= 351 * 15.1167

= 5306

A.3: RELATING SAMPLE PERIOD TO ANNUAL VISITS

We now need to consider evaluation of a factor to relate estimates for the sample period to annual arrivals. To do this we first require an accurate estimate of annual visits.

A.3.1: Adjusting for systematic errors in pneumatic visit counters

Table A3 details weekly visitor data collected via pneumatic counter from March 1990 up to the installation of an electronic loop counter on 1 April 1993. Table A4 details weekly visitor data from the latter electronic counter from its installation to the end of July 1993.

A major problem facing UK forest recreation research has been the acknowledged deficiencies of pneumatic visit counters[23]. Pneumatic counters tend to suffer from systematic errors, that is they record the overall pattern of visits reasonably well but tend to be systematically inaccurate in recording absolute numbers. For example, a particular pneumatic counter may, on average, fail to register one car in ten whilst another pneumatic counter may double count on average one car in fifteen. Each pneumatic counter seems to have its own idiosyncrasies. This means that we have to calculate adjustment factors for any individual pneumatic counter whose data we wish to use.

This situation has been considerably improved by the recent introduction of electronic loop counters which are considered to be far more accurate. However, such counters have only been installed at a few sites and since early 1993. One of the major reasons determining our choice of survey site was the installation of an electronic loop counter at Lynford Stag. Because errors in the pneumatic counters tend to be systematic, comparison of the data obtained from a particular pneumatic counter with that derived for the same period from electronic loop counters allows estimation of a pneumatic/electronic loop conversion factor. Such a factor can then be applied to the pneumatically derived annual visitor estimates to adjust these for error in such counters. While we did not have a period over which both pneumatic and electronic loop counters were in operation, we can compare counts made by the electronic loop device with those made by the pneumatic counter for the same period in the previous year. While this is perhaps less than ideal such a comparison can be improved by ensuring that factors such as the number of bank holidays and wet weather days is the same during compared periods. Such checks were made and appropriate comparable periods defined. Table A5 compares data from the electronic loop counter in 1993 with data from the relevant pneumatic counter[24] for identical periods in 1992 and 1991. The weighted mean adjustment factor implied from table A5 was 0.7427, i.e. the annual visitor totals recorded by the pneumatic counter should be adjusted by this factor.

---

[23] Pers. comm., Roger Oakes, Forestry Commission Statistics Branch, Edinburgh, August 1993.

[24] It is very important to note that each individual pneumatic counter is liable to exhibit its own idiosyncratic systematic error. The counter in table A3.17 systematically overestimated arrivals whereas analysis of an earlier counter used in 1990 showed that it underestimated arrivals (electronic loop/pneumatic = 1.3483). It is therefore important individual adjustment factors are calculated for each pneumatic counter.

Table A3: Lynford Stag traffic count 1990-93: pneumatic counter

Note: some pneumatic counters are designed to record the number of vehicle axles while others record the number of tyres. Weekly readings have to be divided by either 2 or 4 accordingly. Counter types are recorded in the remarks column.

Date of Reading			Traffic Counter Reading					Difference from previous reading				Remarks (e.g. about counter type, weather, public holidays, closures, special events)	Cars: (party visits)
Date	Month	Year											
			4	1	8	2	0					DIVIDE BY 2 UNTIL 31.5.90	
1	3	90	4	2	8	5	0	1	0	3	0	Wet & windy	515
8	3	90	4	3	5	6	9		7	1	9	Wet & windy-generally cloudy	359
15	3	90	4	4	5	7	6	1	0	0	7	Windy at weekend-fine & warm week	503
22	3	90	4	5	9	6	3	1	3	8	7	Warm & windy	639
29	3	90	4	6	9	0	5		9	4	2	Cold & windy weekend	471
5	4	90	4	8	4	1	3	1	5	0	8	Warm weekend otherwise variable	754
12	4	90	4	9	7	4	3	1	3	3	0	Sunny but cold wind	665
19	4	90	5	1	7	7	3	2	0	3	0	Windy, cold & showery	1015
26	4	90	5	3	1	7	2	1	3	9	9	Improvement from weekend	699
3	5	90	5	4	9	5	4	1	7	8	2	Warm & dry becoming hot	891
10	5	90	5	7	9	2	8	2	9	7	4	Hot & dry w/e, cooler BH Monday	1487
17	5	90										COUNTER DEFECTIVE - REMOVED	
13	8	90			9	1	4					NEW COUNTER INSTALLED, DIVIDE BY 4	
16	8	90		3	5	1	7	2	5	9	3	Hot & Sunny becoming cooler	648
23	8	90	1	2	6	2	9	10	0	3	6	Rain at w/e becoming hot	2509
30	8	90	2	9	0	5	2	16	4	2	3	Hot Bank Holiday	4106
6	9	90	3	9	6	8	1	10	6	2	9	Becoming changeable & cooler	2657
13	9	90	3	8	0	0	8	8	3	2	7	Warm	2082
20	9	90	5	5	3	3	9	7	3	2	1	Pleasant w/e	1830
27	9	90	6	1	2	2	4	5	9	3	5	Showery & cool	1484
4	10	90	6	3	5	5	0	2	2	6	6	" " "	566
11	10	90	6	4	8	8	9	1	3	2	9	Warmer	332
18	10	90		6	4	4	1	1	5	6	2	Sunny w/e & generally warm	390

Table A3 (cont.)

Date of Reading			Traffic Counter Reading					Difference from previous reading				Remarks (e.g. about weather, public holidays, closures, special events)	Cars: (party visits)
Date	Month	Year											
			6	6	4	2	1						
25	10	90	6	7	8	5	3	1	4	3	2	Variable sunshine	358
1	11	90	6	9	2	7	3	1	4	2	0	" "	355
8	11	90	7	0	1	1	9		8	4	6	" " + rain	211
15	11	90	7	1	1	2	2	1	0	0	3	" " " "	251
22	11	90	7	2	0	1	8		8	9	6	Dull wet & cool	224
29	12	90	7	2	7	8	5		7	6	7	" " " "	192
6	12	90	7	3	4	1	5		6	3	0	Variable	157
10	12	90			5	3	0					DEFECTIVE COUNTER-NEW ONE INSTALLED Variable, cold	-
13	12	90		1	1	3	1		6	0	1	& snow	150
20	12	90		3	0	0	1	1	8	7	0	" " " "	467
27	12	90										Not read: 1/2 of following week total	660
3	1	91		8	2	8	2	5	2	8	1	Variable but windy	660
10	1	91	1	0	7	0	6	2	4	2	4	" " "	606
17	1	91	1	3	1	3	3	2	4	2	7	Variable	607
24	1	91	1	5	3	7	3	2	2	4	0	"	560
31	1	91	1	7	5	0	9	2	1	3	6	Cloudy & cold	534
7	2	91	1	9	6	0	9	2	1	0	0	" " cold	525
14	2	91	1	9	8	5	1		2	4	2	Snow & very cold	60
21	2	91	2	1	9	8	8	2	1	3	7	Milder	534
28	2	91	2	4	7	7	6	2	8	8	8	"	722
7	3	91	2	8	2	6	1	3	4	8	5	"	871
14	3	91	3	1	4	1	6	3	1	5	5	Cool	789
21	3	91	3	5	5	3	9	4	1	2	3	Variable	1031
28	3	91	3	9	9	1	4	4	3	7	5		1094
4	4	91	5	0	8	1	9	10	9	9	5	Bank holiday weekend BH	2749
11	4	91	5	6	0	9	4	5	2	7	5		1319
18	4	91	6	1	6	0	2	5	5	0	8	Cool	1377
25	4	91	6	5	2	9	9	3	6	9	7	Cool, showery	924

Table A3 (cont.)

Date of Reading			Traffic Counter Reading					Difference from previous reading				Remarks (e.g. about weather, public holidays, closures, special events)	Adjusted cars:
Date	Month	Year											
			6	5	2	9	9						
2	5	91	7	1	7	1	3	6	4	1	4	Cool	1603
9	5	91	8	0	2	3	4	8	5	2	1	BH Monday - cool	2130
16	5	91	8	7	3	0	0	7	0	6	6	Cool, sunny intervals	1766
23	5	91	9	3	9	8	8	6	6	8	8	" " "	1672
30	5	91		5	6	6	4	12	6	7	6	" " " BH	3169
6	6	91	1	3	6	4	2	6	9	7	8	" " "	1744
13	6	91	1	9	4	4	9	5	8	0	7	Cool, showery & windy	1452
20	6	91	2	5	4	3	1	5	9	8	2	" " " "	1495
27	6	91	3	2	0	1	7	6	6	8	6	" " " "	1671
4	7	91	3	9	6	1	5	7	5	9	8	" " " "	1899
11	7	91	4	8	4	1	4	8	7	9	9	Hot & sunny	2200
18	7	91	5	6	7	9	1	8	3	7	7	" " "	2094
25	7	91	6	6	2	0	2	9	4	1	1	" " "	2353
2	8	91	7	8	0	5	3	11	8	5	1	Mainly hot & sunny, some showers	2963
8	8	91	9	0	1	3	0	12	0	7	7	" " " " " "	3019
15	8	91	0	1	5	7	6	11	4	4	6	" " " " cool on Sunday	2861
22	8	91	1	3	6	7	6	12	1	0	0	Warm & dry	3025
29	8	91	3	1	4	0	3	17	7	2	7	Warm & dry except Friday BH	4432
5	9	91	4	3	4	7	4	12	0	7	1	" " "	3018
12	9	91	5	1	9	7	6	8	5	0	2	" " "	2125
19	9	91	5	8	5	9	4	6	6	1	8	Variable but mainly sunny	1654
26	9	91	6	4	3	7	3	5	7	7	9	Showery	1445
3	10	91	6	8	8	5	0	4	4	7	7	"	1119
10	10	91	7	4	2	2	7	5	3	7	7	Dry & sunny	1344
17	10	91	7	9	3	5	7	5	1	3	0	Windy, cool & showery	1282
24	10	91	8	3	6	8	0	4	3	3	7	Cloudy & cool	1084
31	10	91	8	8	0	0	1	4	3	2	1	" " "	1080
7	11	91	9	1	5	7	0	3	4	6	9		867

Table A3 (cont.)

Date of Reading			Traffic Counter Reading					Difference from previous reading				Remarks (e.g. about weather, public holidays, closures, special events)	Adjusted cars:
Date	Month	Year											
			9	1	5	7	0						
14	11	91	9	4	8	5	0	3	2	8	0	Cloudy & cool	820
21	11	91	9	7	4	8	6	2	6	3	6	" " "	659
28	11	91	0	0	3	3	0	2	8	4	4	" but milder	711
5	12	91	0	2	9	4	0	2	5	9	0	" " "	647
12	12	91	0	4	6	3	0	1	6	9	0	Cold & frosty	422
19	12	91	0	6	6	9	0	2	0	6	0	Milder	515
26	12	91	0	8	7	7	3	2	0	8	3	Windy & mild	521
2	1	92	1	2	4	3	2	3	6	5	9	Mild	915
9	1	92	1	5	2	6	8	2	8	2	6	"	706
16	1	92	1	7	3	1	7	2	0	4	9	" & cloudy	512
23	1	92	1	9	8	4	5	2	5	2	8	Sunny spells - cold	632
30	1	92	2	2	0	3	4	2	1	8	9	Variable but mainly cloudy	547
6	2	92	2	4	0	9	0	2	0	6	4	Cold variable cloud	516
13	2	92	2	6	7	0	3	2	6	1	3	Changeable	653
20	2	92	2	9	3	7	2	2	6	6	9	"	667
27	2	92	3	3	4	3	0	4	0	5	8	Half term - variable	1014
5	3	92	3	7	3	1	9	3	8	8	9	Variable	972
12	3	92	4	0	8	9	4	3	5	7	5	"	894
19	3	92	4	3	8	4	3	2	9	4	9	" & wet	737
26	3	92	4	7	0	7	2	3	2	2	9	" " "	807
2	4	92	4	9	9	4	9	2	8	7	7	" " "	719
9	4	92	5	3	9	4	1	3	9	9	2	" " "	998
16	4	92	5	8	9	0	7	5	9	6	6	" " "	1491
23	4	92	7	0	0	3	3	11	1	2	6	Variable Easter BH	2781
30	4	92	7	6	3	2	3	6	2	9	0	" & cool	1572
7	5	92	8	6	4	1	6	10	0	9	3	May BH Mainly cold became warmer	2523
14	5	92	9	1	3	3	0	4	9	2	4	Variable - wet w/e	1231
21	5	92	9	9	4	0	3	8	0	7	3	Mainly hot & dry	2018

Table A3 (cont.)

Date of Reading			Traffic Counter Reading					Difference from previous reading				Remarks (e.g. about weather, public holidays, closures, special events)	Adjusted cars:
Date	Month	Year											
			9	9	4	0	3						
28	5	92	1	1	8	3	2	12	4	2	9	Mainly hot & dry. BH	3107
4	6	92	1	7	8	9	7	6	0	6	5	Variable	1516
												COUNTER REMOVED	-
												REPAIRED-COUNTER REPLACED AT STAG	-
6	11	92	2	2	6	6	4					WITH THIS READING. DIVIDE BY 2.	-
12	11	92	2	4	4	7	7	1	8	1	3	Cold & dry	906
19	11	92	2	5	8	8	0	1	4	0	3	Frosty a.m. overcast drizzle	701
26	11	92	2	7	1	1	7	1	2	3	7	Clear & dry	618
3	12	92	2	8	5	7	0	1	4	5	3	Wet & windy	726
10	12	92	2	9	4	5	9		8	8	9	Cold & damp	444
17	12	92	3	0	3	2	5		8	8	6	Dry & frosty	443
24	12	92	3	1	1	2	8		8	0	0	Cold & frosty	400
31	12	92	3	1	8	5	8		7	3	0	" " "	365
7	1	93	3	2	2	7	2		4	1	4	Wet & mild	207
14	1	93	3	3	5	0	4	1	2	3	2	Cold/very windy	616
21	1	93	3	4	3	7	4		8	7	0	" " "	435
28	1	93	3	5	3	5	5		9	8	1	Wet & mild	490
3	2	93	3	6	3	5	2		9	9	7	Dry & overcast	498
10	2	93	3	7	3	3	0		9	7	8	" " "	489
18	2	93	3	8	5	0	2	1	1	7	2	" " " (mild for Feb)	586
25	2	93	3	9	9	6	9	1	4	6	7	Dry & mild	733
4	3	93	4	1	2	5	1	1	2	8	2	Snow/cold/overcast	641
11	3	93	4	2	6	3	1	1	3	8	0	Mild/frosty/dry	690
18	3	93	4	4	6	3	4	2	0	0	3	Warm/sunny	1001
25	3	93	4	6	2	3	0	1	5	9	6	Cool/frosty start	798

Table A4: Lynford Stag traffic count 1993: electronic loop counter

Week no.	Week commencing	Start of week reading	End of week reading	Change	Cars (party visits)	Remarks
1	1.4.93	46230	47546	1316	658	Dry/cool
2	5.4.93	47546	49234	1688	844	Overcast/drizzle BH
3	12.4.93	49234	52106	2872	1436	Warm/sunny BH
4	19.4.93	52106	54304	2198	1099	Cool/sunny
5	26.4.93	54304	56471	2167	1083	Cold/windy
6	3.5.93	56471	59700	3229	1615	Warm/sunny BH
7	10.5.93	59700	61816	2116	1058	Hot/dry
8	17.5.93	61816	63735	1919	959	Hot/dry
9	24.5.93	63735	66514	2779	1389	Hot/dry
10	31.5.93	66514	70384	3870	1935	Warm/sunny BH
11	7.6.93	70384	73863	3479	1739	Warm/sunny
12	14.6.93	73863	76436	2573	1287	Warm/overcast
13	21.6.93	76436	79072	2636	1318	Dull/drizzle
14	28.6.93	79072	82460	3388	1694	Warm/sunny
15	5.7.93	82460	85624	3164	1582	Cool/damp
16	12.7.93	85624	88329	2705	1352	Warm/damp
17	19.7.93	88329	91193	2864	1432	Wet/warm
18	26.7.93	91193	94845	3652	1826	Cool/windy

Notes: The 'Change' column refers to the difference between the start and end of week readings. This tells us the number of cars both entering and leaving the car park. Dividing this number by 2 gives the number of cars visiting the site during the week (shown in the 'Cars' column). Total cars for the entire period shown = 24307.
BH = Bank holiday

Table A5: Comparison of visitor counts by pneumatic and electronic loop counters Lynford Stag (1991-93)

Period	Counter	Cars	Electronic Loop / Pneumatic
1.4.93-10.6.93	Electronic loop	13,071	0.7233
1.4.92-10.6.92[1]	Pneumatic	18,071	
4.4.93-1.8.93	Electronic loop	23,850	0.7543
4.4.91-1.8.91	Pneumatic	31,617	

Note: 1. Counter removed due to failure 11.6.92
2. Ratios are 0.7233136 and 0.754341 respectively. Weighted mean = 0.7427264.

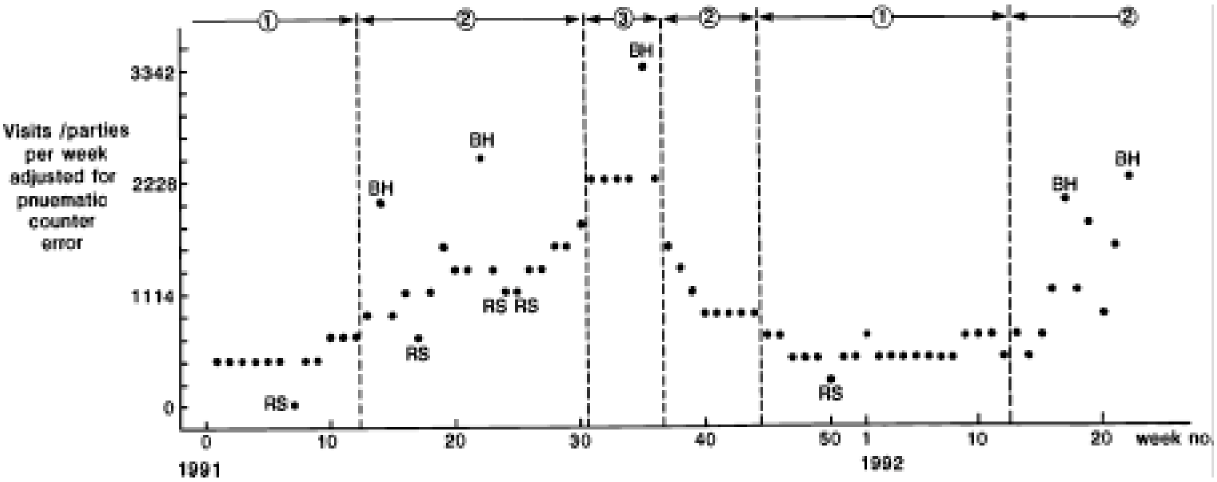
A.3.2: Modelling Annual Visit Trends

In Bateman (1996) we estimate an arrivals function which can predict the number of visitors which will arrive during the survey period[25]. We now need to convert this to an estimate of annual arrivals. This is achieved by examining the relationship between arrivals in our sample study period and annual arrivals. But for a derived factor to be reliable this relationship needs to be stable.

Data from the same pneumatic counter analysed in table A5 was held for the period w/c 13.12.90 to w/c 4.6.92. An initial analysis investigated two 12-month periods within this dataset: 3.1.91 to 1.1.92 and 6.6.91 to 4.6.92 (both consist of 364 days). Adjusting for pneumatic counter error, annual arrivals were 56316 and 56843 parties respectively, a difference of less than 1% between the two annual sums. Figure A1 shows the frequency of visits per week for virtually the entire operation of this single pneumatic counter. Note that the seasonal periods are defined to reflect the interaction of both seasonal and holiday period dates. These two factors are highly collinear and may not be entered separately into the model.

[25]Which we have previously adjusted for those who were not interviewed during the survey period.

Figure A1: Visits to Lynford Stag (parties per week): w/c 3.1.91 to w/c 4.6.92



Key: BH = Bank holiday (excluding Christmas)
RS = Raining or snowing/frosty
1 = Winter period
2 = Spring/Autumn period
3 = Summer period

Considering our two 12 month periods (3.1.91 to 1.7.92) and (6.6.91 to 4.6.92), there is clearly a considerable area of overlap. However, the non-overlapping weeks of the first period show a striking similarity to corresponding weeks in the second. In both cases relevant variables determining visits appear to be seasonal factors, bank holidays (BH) and rain or snow (RS).

In order to test the stability of the relationship between our survey period and annual visits, a simple statistical model of the latter was constructed. Using this, a hypothesis as to the stability of the sample/annual relationship could be tested. All visitor data was adjusted for pneumatic counter error prior to modelling.

As figure A1 indicates, there is clearly a strong seasonal pattern to arrivals[26]. This reflects a mixture of annual weather patterns heightened by the distribution of holidays. Visits are roughly constant at a low level for approximately the first 12 weeks of the year after which visit frequency grows at a fairly steady rate until a plateau is reached at about week 31. Visit frequency falls relatively sharply from week 37 to return to winter levels by about week 45. Oneway analysis of variance tests showed that the pre and post New Year winter visitation rates were insignificantly different as were the spring and autumn periods. Figure A2 details this analysis. Three highly distinct seasonal periods could then be defined: Winter; Spring/Autumn; and Summer. Figure A3 details statistical analysis of such a three level seasonality variable[27].

Further investigation revealed two further important explanatory variables affecting weekly visit totals. Firstly, weeks which contained a bank holiday[28]: recorded significantly higher visit numbers than comparable weeks within the same season but without bank holidays. Secondly, weeks characterised by unusually high levels of rain for the season or which experienced snow[29] recorded significantly lower visit rates than comparable weeks in the same season without such adverse weather conditions.

[26]Both monthly and weekly variables were examined with the latter, as expected, providing a better fit to the data.

[27]Initial analyses (such as those detailed in figures A2 and A3) examined the dataset for 3.1.91 to 1.1.92. The full regression model, detailed subsequently, examines data for the full period from 13.12.90 to 4.6.92.

[28]Tests were run to examine the effect of the double bank holidays of the Easter period. These proved to be insignificantly different from single day bank holidays. We conclude that a bank holiday significantly raises the probability that a household will visit but a double bank holiday does not lead to two visits being made.

[29]Data from Forestry Commission records. Various permutations of weather were investigated with rain/snow being the only factor not collinear with the seasonality variable ie. rain/snow depresses visitor rate irrespective of the season.

Figure A2: Oneway analysis of variance of weekly visits (parties) on 5 seasonal periods during the year 3.1.91 to 1.1.92. Visits adjusted for pneumatic counter error.

```
Individual 95% CI's for mean visitor
        numbers in each period               Period
      (based on pooled st. dev.)
--------+---------+---------+---------
  (--*-)                              1. Post New Year
               (-*-)                  2. Spring
                            (---*---) 3. Summer
         (---*--)                     4. Autumn
  (--*--)                             5. Pre New Year
-------+---------+---------+---------
      700      1400      2100
   Mean visitor numbers per period
```

Analysis of variance on weekly visits (pooled st. dev = 322.5)

Source	df	SS	MS	F	p
Period Error Total	4 47 51	19082608 4887401 23970008	4770652 103987	45.88	0.000

Period	n	mean	st.dev
1. Post new year	12	464.1	175.4
2. Spring	18	1349.7	418.9
3. Summer	6	2391.3	443.5
4. Autumn	8	1033.6	264.8
5. Pre new year	8	479.2	114.5

Figure A3: Oneway analysis of variance of weekly visits (parties) on 3 seasonal periods for the year 3.1.91 to 1.1.92. Visits adjusted for pneumatic counter error.

```
Individual 95% CI's for mean visitor
        numbers in each period               Period
      (based on pooled st. dev.)
--------+---------+---------+---------+
  (-*-)                                   1. Winter
              (-*-)                        2. Spring/Autumn
                            (---*---)     3. Summer
--------+---------+---------+---------+
       700      1400      2100      2800
    Mean visitor numbers per period
```

Analysis of variance on weekly visits (pooled st. dev = 333.3)

SOURCE	DF	SS	MS	F	p
Period(3) ERROR TOTAL	2 49 51	18527962 5442047 23970008	9263981 111062	83.41	0.000

LEVEL	N	MEAN	STDEV
1.Winter	20	470.2	150.7
2.Spring/Autumn	26	1252.5	401.3
3.Summer	6	2391.3	443.5

With our three explanatory variables defined, regression models of weekly visitor rate could be estimated. A linear specification provided a best fit to the data, the model being reported as table A6[30].

Table A6: Generalised linear model of weekly visitation pattern at Lynford Stag, Thetford Forest.
Dependent variable = visits per week (adjusted for pneumatic counter error)
Data period: 13.12.90 to 4.6.92

Term	Coeff	Stdev	t-value	p
Constant	1623.58	56.01	28.99	0.000
Season:				
Winter	-803.40	40.48	-19.85	0.000
Spring/Autumn	-120.85	40.84	-2.96	0.004
Rain/Snow	-131.93	30.92	-4.27	0.000
Bank Holiday	432.65	44.97	9.62	0.000

Analysis of variance

Source	df	Seq SS	Adj SS	Adj MS	F	p
Season	2	23452372	17710412	8855206	197.09	0.000
RainSnow	1	1980331	817955	817955	18.21	0.000
BankHol	1	4159146	4159146	4159146	92.57	0.000
Error	73	3279836	3279836	44929		
Total	77	32871684				

where:

VISITS = Dependent variable: Number of parties (cars) visiting Lynford Stag per week
SEASON = 1 for a winter week; 2 for a spring/autumn week; 3 for a summer week
RAINSNOW = 1 if significant rain or snow during the week; 0 otherwise
BANKHOL = 1 if week contains a bank holiday; 0 otherwise

The model given in table A6 describes the data well (R^2 = 90.0%) with expected relationships on all explanatory variables, the latter all being significant at the 1% level.

The model was then used to examine those periods which are relevant to the relationship between our survey observations and annual visits. Our survey spanned the five week period from 26.3.93 to 25.4.93 and so we wish to see how robust the relationship between such a period and annual arrivals might be. The model given in table A6 uses data for similar periods both in 1991 and 1992 and period/annual relationships can be calculated for both. However, the model gives us reason to believe that this relationship will not be completely stable for these two years. Whilst the 5 week period 28.3.91 to 1.5.91 contains just one rainy week, the 5 week period 26.3.92 to 29.4.92 contains three (bank holidays being constant between the periods). The model therefore predicts that the latter period will have less visitors than the former but that the annual totals for 1991 and 1992 are likely to be similar(from our previous observations and because annual weather patterns are similar). Examining actual arrivals for these periods we find that the predictions of our model are borne out. Table A7 details visits for the survey periods in previous years (1991, 1992), along with respective annual totals and resultant ratios.

[30]The model is fitted using a generalised linear modelling (GLM) package where the upper level of each categorical variable is taken as the reference point from which category coefficients for that variable are calculated. Thus the upper level of the season variable (3; summer) is not explicitly shown as it is the default when season is neither winter or spring/autumn. Standard output from the GLM package would refer to the absence of rain/snow or of bank holidays, however as this is counter to the conventional approach in specifying dummy variables, the variables in table A3.18 have been respecified to give standard outputs ie visit rate is lowered by rain/snow and raised by bank holidays.

Table A7: Survey period/annual arrivals conversion factors

Year	Party visits during 'survey period'	Annual visits	Ratio
1991	5543	56316	10.1598
1992	5048	56843	11.2605

Although the two periods appear dissimilar a oneway analysis of variance rejected such a hypothesis ($p = 0.796$). In effect the relationship between visits in the sample period and annual visits varies logically with the explanatory variables in table A7 and seems to exhibit reasonable temporal stability. Given that the 1991 period appears to be one of relatively good weather, and that for 1992 seems relatively poor, a reasonable assumption would be to adopt a midway point between the two resulting values, this being 10.7102. Applying such a factor to our survey sample (after allowing for those not surveyed on survey days and for those days not sampled during the survey period) gives a predicted arrival total for the 1993 survey period of 5306 party visits, an estimate which accords well with actual visits in 1993 and lies midway between actual visits in 1991 and 1992 further justifying our choice of a mean period/annual conversion ratio.

A.3.3: Summary: relating survey period to annual visits

A number of conversion factors have now been calculated such that we can now relate our observed sample of 351 party visits to an estimated annual visit total as follows:

i. Allowing for those who were not interviewed on survey days:

= 351 * 6.9105

= 2426 parties

ii. Allowing for days not surveyed during the survey period:

= 351 * 6.9105 * 2.1875

= 351 * 15.1167

= 5306 parties

iii. Relating the survey period to annual totals (after first allowing for pneumatic counter error in the latter):

= 351 * 6.9105 * 2.1875 * 10.7102

= 351 * 161.9031

= 56828 parties

Our estimated total party visits based upon our survey observations accords well with both the 1991 and 1992 totals detailed in table A7 being within 1% of the former and almost identical to the latter.

The arrivals function detailed in chapter 5 operates in terms of parties for which the above conversion factors are appropriate. However, we could further convert this to estimate the number of person visits per annum. Table A8 provides descriptive statistics regarding party and household size and composition gathered from our survey at Lynford Stag.

Table A8: Party and household size and composition: Thetford survey

Variable	N[1]	mean	median	tr. mean[2]	st. dev	s.e. mean	min	max	Q1	Q3
Party16+	350	2.374	2	2.067	2.174	0.116	0	25	2	2
Party<16	350	1.480	1	1.229	2.230	0.119	0	28	0	2
House16+	351	2.234	2	2.152	0.924	1.049	1	11	2	2
House<16	351	1.137	1	1.044	1.204	0.064	0	6	0	2

Notes: 1. 1 missing observation with regard to party age
2. 5% trimmed mean

Table A8 shows that the average party consisted of 2.37 adults and 1.48 children. These means are somewhat inflated by a very few large parties and it may therefore be more valid to consider the median party which consists of 2 adults and 1 child[31]. Using such an estimate implies that nearly 170,000 person visits are made to Lynford Stag every year[32].

[31]Interestingly this coincides with the Forestry Commission's own working estimate of party size being 3 persons (Anna Chylak, Forestry Commission, Thetford Forest, pers comm, August 1993).

[32]Precise estimate is 169,739 person visits, the majority of which are repeat visits. On average each visitor visits 14.65 times per year, implying that some 11,586 individual people visit Lynford Stag p.a.

APPENDIX B: DATA COMPILED FOR META-ANALYSES OF WOODLAND RECREATION VALUE

Table B1: Assembled database of studies providing per person per visit valuation estimates for UK woodland recreation benefits.

Est. No.	*Method*	Authors of study	*Author* (No.)	*Study* (No.)	Study forest	*Forest* (No.)	Value type/ elicit. method	*Option* (value type)	*Elicit*	*OE*	(Study year value £)	*Year* (of study)	*Value* (£ 1990)
1	1	Whiteman and Sinclair (1994)	1	1	Mercia	1	Use/OE	0	1	1	1.00	1992	0.93
2	1	ibid.	1	1	Thames Chase	2	Use/OE	0	1	1	0.71	1992	0.66
3	1	ibid.	1	1	Gt. Northern Forest	3	Use/OE	0	1	1	0.81	1992	0.75
4	1	Hanley and Ruffell (1991)	2	2	Aberfoyle	4	Use/OE	0	1	1	0.90	1991	0.85
5	1	ibid.	2	2	Aberfoyle	4	Use/IB	0	2	0	1.21	1991	1.14
6	1	ibid.	2	2	Aberfoyle	4	Use/PC	0	3	0	1.39	1991	1.31
7	1	ibid.	2	2	Aberfoyle	4	Use/DC	0	6	0	1.49	1991	1.41
8	1	Bishop (1992)	3	3	Derwent Walk	5	Use/OE	0	1	1	0.42	1989	0.46
9	1	ibid.	3	3	Derwent Walk	5	use+ option/OE	1	1	1	0.97	1989	1.06
10	1	ibid.	3	3	Whippendell Wood	6	Use/OE	0	1	1	0.54	1989	0.59
11	1	ibid.	3	3	Whippendell Wood	6	use+option/OE	1	1	1	1.34	1989	1.46
12	1	Willis and Benson (1989)	4	4	New Forest	7	Use/OE	0	1	1	0.43	1988	0.47
13	1	ibid.	4	4	Cheshire	8	Use/OE	0	1	1	0.47	1988	0.51
14	1	ibid.	4	4	Loch Awe	9	Use/OE	0	1	1	0.50	1988	0.55
15	1	ibid.	4	4	Brecon	10	Use/OE	0	1	1	0.46	1988	0.50
16	1	ibid.	4	4	Buchan	11	Use/OE	0	1	1	0.57	1988	0.62

Est. No.	*Method*	Authors of study	*Author* (No.)	*Study* (No.)	Study forest	*Forest* (No.)	Value type/ elicit. method	*Option* (value type)	*Elicit*	*OE*	(Study year value £)	*Year* (of study)	*Value* (£ 1990)
17	1	ibid.	4	4	Newton Stewart	12	Use/OE	0	1	1	0.73	1988	0.80
18	1	ibid.	4	4	Lorne	13	Use/OE	0	1	1	0.72	1988	0.79
19	1	ibid.	4	4	Ruthin	14	Use/OE	0	1	1	0.44	1988	0.48
20	1	ibid.	4	4	New Forest	7	use+option/OE	1	1	1	0.88	1988	0.96
21	1	ibid.	4	4	Cheshire	8	use+option/OE	1	1	1	0.72	1988	0.79
22	1	ibid.	4	4	Loch Awe	9	use+option/OE	1	1	1	0.76	1988	0.83
23	1	ibid.	4	4	Brecon	10	use+option/OE	1	1	1	0.66	1988	0.72
24	1	ibid.	4	4	Buchan	11	use+option/OE	1	1	1	0.79	1988	0.86
25	1	ibid.	4	4	Newton Stewart	12	use+option/OE	1	1	1	1.18	1988	1.29
26	1	ibid.	4	4	Lorne	13	use+option/OE	1	1	1	1.02	1988	1.12
27	1	ibid.	4	4	Ruthin	14	use+option/OE	1	1	1	0.63	1988	0.69
28	1	Hanley (1989)	2	5	Aberfoyle	4	Use/OE	0	1	1	1.24	1987	1.53
29	1	ibid.	2	5	Aberfoyle	4	Use/PC	0	3	0	1.25	1987	1.55
30	1	Willis et al (1988)	4	6	Castle Douglas	15	Use/OE	0	1	1	0.37	1987	0.46
31	1	ibid.	4	6	South Lakes	16	Use/OE	0	1	1	0.39	1987	0.48
32	1	ibid.	4	6	North York Moors	17	Use/OE	0	1	1	0.53	1987	0.66
33	1	ibid.	4	6	Durham	18	Use/OE	0	1	1	0.31	1987	0.38
34	1	ibid.	4	6	Thetford	19	Use/OE	0	1	1	0.23	1987	0.28
35	1	ibid.	4	6	Dean	20	Use/OE	0	1	1	0.28	1987	0.35
36	1	ibid.	4	6	Castle Douglas	15	use+option/OE	1	1	1	0.80	1987	0.99
37	1	ibid.	4	6	South Lakes	16	use+option/OE	1	1	1	0.86	1987	1.06

Est. No.	*Method*	Authors of study	*Author* (No.)	*Study* (No.)	Study forest	*Forest* (No.)	Value type/ elicit. method	*Option* (value type)	*Elicit*	*OE*	(Study year value £)	*Year* (of study)	*Value* (£ 1990)
38	1	ibid.	4	6	North York Moors	17	use+option/OE	1	1	1	1.03	1987	1.27
39	1	ibid.	4	6	Durham	18	use+option/OE	1	1	1	0.56	1987	0.69
40	1	ibid.	4	6	Thetford	19	use+option/OE	1	1	1	0.41	1987	0.51
41	1	ibid.	4	6	Dean	20	use+option/OE	1	1	1	0.63	1987	0.78
42	1	Bateman and Langford (1997a)[2]	5	7	Thetford 2 (f2NB)[3]	19	Use/OE	0	1	1	0.52	1993	0.47
43	1	Bateman (1996)	5	8	Thetford 1	19	Use/PCL	0	4	0	1.21	1990	1.21
44	1	Ibid.	5	8	Thetford 1	19	Use/PCH	0	5	0	1.55	1990	1.55
45	2	Willis and Garrod (1991)	4	9	Brecon	10	Use	0	-9	-9	1.40	1988	1.65
46	2	Ibid.	4	9	Buchan	11	Use	0	-9	-9	0.50	1988	0.59
47	2	Ibid.	4	9	Cheshire	8	Use	0	-9	-9	0.40	1988	0.47
48	2	Ibid.	4	9	Lorne	13	Use	0	-9	-9	1.53	1988	1.80
49	2	Ibid.	4	9	New Forest	7	Use	0	-9	-9	2.32	1988	2.74
50	2	Ibid.	4	9	Ruthin	14	Use	0	-9	-9	1.29	1988	1.52
51	3	Ibid.	4	9	Brecon	10	Use	0	-9	-9	0.66	1988	0.78
52	3	Ibid.	4	9	Buchan	11	Use	0	-9	-9	0.20	1988	0.24
53	3	Ibid.	4	9	Cheshire	8	Use	0	-9	-9	0.06	1988	0.07
54	3	Ibid.	4	9	Lorne	13	Use	0	-9	-9	0.96	1988	1.13
55	3	Ibid.	4	9	New Forest	7	Use	0	-9	-9	0.12	1988	0.14
56	3	ibid.	4	9	Ruthin	14	Use	0	-9	-9	0.88	1988	1.03
57	3	Bateman (1996)	5	10	Thetford 2[2]	19	Use	0	-9	-9	1.32	1993	1.20

Est. No.	*Method*	Authors of study	*Author* (No.)	*Study* (No.)	Study forest	*Forest* (No.)	Value type/ elicit. method	*Option* (value type)	*Elicit*	*OE*	(Study year value £)	*Year* (of study)	*Value* (£ 1990)
58	2	Bateman (1996)	5	11	Thetford 1[3,4]	19	Use	0	-9	-9	1.07	1990	1.07
59	2	Ibid.	5	11	Thetford 1[3,5]	19	Use	0	-9	-9	1.19	1990	1.40
60	2	Ibid.	5	11	Thetford 1[3,6]	19	Use	0	-9	-9	1.34	1990	1.58
61	4	Benson and Willis (1992)	4	12	New Forest	7	Use	0	-9	-9	1.43	1988	1.69
62	4	ibid.	4	12	Cheshire	8	Use	0	-9	-9	1.91	1988	2.26
63	4	ibid.	4	12	Loch Awe	9	Use	0	-9	-9	3.31	1988	3.91
64	4	ibid.	4	12	Brecon	10	Use	0	-9	-9	2.60	1988	3.07
65	4	ibid.	4	12	Buchan	11	Use	0	-9	-9	2.26	1988	2.67
66	4	ibid.	4	12	Durham	18	Use	0	-9	-9	1.64	1988	1.94
67	4	ibid.	4	12	N York Moors	17	Use	0	-9	-9	1.93	1988	2.28
68	4	ibid.	4	12	Aberfoyle	4	Use	0	-9	-9	2.72	1988	3.21
69	4	ibid.	4	12	South Lakes	16	Use	0	-9	-9	1.34	1988	1.58
70	4	ibid.	4	12	Newton Stewart	12	Use	0	-9	-9	1.61	1988	1.90
71	4	ibid.	4	12	Lorne	13	Use	0	-9	-9	1.44	1988	1.70
72	4	ibid.	4	12	Castle Douglas	15	Use	0	-9	-9	2.41	1988	2.85
73	4	ibid.	4	12	Ruthin	14	Use	0	-9	-9	2.52	1988	2.98
74	4	ibid.	4	12	Dean	20	Use	0	-9	-9	2.34	1988	2.76
75	4	ibid.	4	12	Thetford	19	Use	0	-9	-9	2.66	1988	3.14
76	4	Hanley (1989)	2	13	Aberfoyle	4	Use	0	-9	-9	1.70	1987	2.14
77	4	Everett (1979)	6	14	Dalby	21	Use	0	-9	-9	0.41	1976	1.30

Variable coding for Table B1 is as follows (all variable values were re-coded to individual dummies for the regression analysis):

Method (valuation method):

1 = CV
2 = individual TC (OLS estimator used)
3 = individual TC (ML estimator used)
4 = zonal TC

Author

1= Whiteman and Sinclair
2 = Hanley et al.
3 = Bishop
4 = Willis et al.
5 = Bateman et al.
6 = Everett

Option

1 = use value plus option value explicitly requested in WTP question
0 = use value

Elicit

WTP elicitation methods are as follows:
1= open ended (OE)
2 = iterative bidding (IB)
3 = payment card (PC), only one range used in study
4 = low range payment card (PCL)
5 = high range payment card (PCH)
6 = dichotomous choice (DC)

OE

1= open ended elicitation method used
0 = other

Investigating the size of forest effects in the conventional meta-analysis model

The GLM estimated model detailed below includes estimated coefficients for each of the forests studied in the dataset. Note that in this model the author variable *Hanley* becomes insignificant. This is due to very high colinearity with the Aberfoyle site (Forest number 4) with 7 of the 8 studies conducted at this site being undertaken by Nick Hanley.

Analysis of Variance for Value, using Adjusted SS for Tests

Source	DF	Seq SS	Adj SS	Adj MS	F	P
ZTC	1	31.0209	30.6757	30.6757	127.94	0.000
ITCols	1	2.8698	3.6617	3.6617	15.27	0.000
Year1990	1	1.4840	0.0895	0.0895	0.37	0.544
Option	1	0.8683	0.9549	0.9549	3.98	0.051
Hanley	1	1.1742	0.0681	0.0681	0.28	0.596
Forest	20	3.1926	3.1926	0.1596	0.67	0.840
Error	51	12.2281	12.2281	0.2398		
Total	76	52.8379				

$R^2 = 0.768$

Term	Coef	StDev	T	P
Constant	2.0064	0.3181	6.31	0.000
ZTC	0.90282	0.07982	11.31	0.000
ITCols	0.38088	0.09746	3.91	0.000
Year1990	0.03625	0.05933	0.61	0.544
Option	0.16016	0.08026	2.00	0.051
Hanley	-0.1452	0.2724	-0.53	0.596
Forest				
1	0.1498	0.5273	0.28	0.777
2	-0.1202	0.5273	-0.23	0.821
3	-0.0302	0.5273	-0.06	0.955
4	0.7692	0.4834	1.59	0.118
5	-0.0716	0.3466	-0.21	0.837
6	0.1934	0.3466	0.56	0.579
7	-0.0127	0.2219	-0.06	0.955
8	-0.3927	0.2219	-1.77	0.083
9	0.4195	0.2795	1.50	0.140
10	0.1313	0.2219	0.59	0.557
11	-0.2167	0.2219	-0.98	0.333
12	-0.0138	0.2795	-0.05	0.961
13	0.0953	0.2219	0.43	0.669
14	0.1273	0.2219	0.57	0.569
15	0.1137	0.2819	0.40	0.688
16	-0.2797	0.2819	-0.99	0.326
17	0.0837	0.2819	0.30	0.768
18	-0.3163	0.2819	-1.12	0.267
19	0.0995	0.1988	0.50	0.619
20	-0.0230	0.2819	-0.08	0.935

Forest numbers are as per Table B1

Stability of the best fit conventional meta-analysis model

Stepwise forward entry regression models were drawn from 13 potential predictors of woodland recreation value. Stability of coefficient estimates across steps (and standard diagnostics) suggests that multicollinearity is not a significant problem here. The final step (7) gives us the Model G from Table 5 of Part Two of this report.

Step	1	2	3	4	5	6	7
Constant	0.9038	0.8120	0.9162	0.9390	0.8671	0.7860	0.7697
ZTC	1.53	1.62	1.72	1.71	1.82	1.86	1.85
T-Value	10.33	11.45	12.08	12.31	12.40	12.91	12.94
ITCols		0.61	0.60	0.63	0.72	0.79	0.80
T-Value		3.35	3.42	3.61	4.07	4.51	4.63
Year1990			0.072	0.069	0.082	0.076	0.075
T-Value			2.49	2.45	2.89	2.72	2.74
Cheshire				-0.45	-0.45	-0.41	-0.40
T-Value				-2.02	-2.05	-1.94	-1.88
Option					0.29	0.35	0.34
T-Value					1.98	2.44	2.38
Hanley						0.42	0.44
T-Value						2.22	2.33
Loch_Awe							0.42
T-Value							1.54
S	0.539	0.506	0.489	0.479	0.470	0.457	0.453
R-Sq	58.71	64.14	66.95	68.72	70.36	72.31	73.24

We can re-estimate Model G excluding the Loch Awe variable to yield the following model:

Predictor	Coef	StDev	T	P
Constant	0.78597	0.09125	8.61	0.000
Option	0.3529	0.1446	2.44	0.017
Cheshire	-0.4135	0.2127	-1.94	0.056
Year1990	0.07580	0.02785	2.72	0.008
Hanley	0.4214	0.1896	2.22	0.029
ZTC	1.8572	0.1439	12.91	0.000
ITCols	0.7855	0.1741	4.51	0.000

R-Sq = 72.3%   R-Sq(adj.) = 69.9%

As can be seen, omission of the Loch Awe variable makes no substantive difference to this model.