Demand analysis projections for recreational visits to countryside woodlands in Great Britain

GARY W. HILL and PAUL R. COURTNEY

1 The Macaulay Institute, Craigiebuckler, Aberdeen AB15 8QH, Scotland
2 Countryside and Community Research Unit, University of Gloucestershire, Dunholme Villa, The Park, Cheltenham, Gloucestershire GL50 2RH, England
*Corresponding author. E-mail: pcourtney@glos.ac.uk

Summary
Forests and woodlands have long been associated with tourism and recreation and there is a growing demand for improved planning tools to measure and maximize amenity value from policy intervention. Building on earlier studies, this study develops a trip generation function (TGF) to predict visit rates to public and privately owned forest sites in Great Britain. Using a zonal approach, the models were based on visit counts to 100 countryside woodland sites. A survey of forest managers was made to collect data on facilities and characteristics, and geographical information systems analysis was used to define outset zones and compile data on the population living within the zones and the quality and accessibility of alternative woodlands. A series of stand-alone TGFs were developed. In addition to size of population living within a 2-h travel time and substitute-site accessibility, the results highlighted the importance of the number of site facilities in determining visit rates. The results of a cross-validation suggest that the TGFs are sufficiently robust to predict visitor numbers at an aggregate level. However, the current value of this type of modelling to forest management decision making, and further advancement in modelling to allow transferability of results to unsurveyed sites is currently constrained by a number of data limitations, including the quality of available visit data.

Introduction
As well as the conventional range of forest products, forests, woodlands and trees provide what can be termed amenity services. It is the existence of these amenity services that attract people to visit forest and countryside areas. While forests and woodlands have long been associated with tourism and recreation, recent years have seen a significant shift towards multi-purpose forestry, with increasing management effort focused on balancing timber production with management for recreational and environmental benefits (Mather, 2001). In line with changing priorities, there is a growing demand for improved planning tools to measure and maximize amenity value from policy intervention. Single-site valuation studies of amenity and recreational resources have limited relevance for wider decision making (Lovett et al., 1997; Bateman et al., 1999). As
such, economic tools to help evaluate past and prioritize future, forestry investments and to assist with resource allocation decisions are deemed increasingly useful. 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 (Jones et al., 2002). This understanding will assist forest managers in effective management to target resources at appropriate sites, for example, to maximize public good benefits and to ensure that visitor carrying capacities are not exceeded.

Consequently, the development of transferable recreational demand functions, where the findings of evaluation studies can be transferred to unsurveyed or planned forest sites, has become a focus for research (Garrod and Willis, 1994; Lovett et al., 1997). Such tools ‘potentially’ offer considerable financial savings for forest managers and are an attractive alternative to expensive and time-consuming visitor surveys.

Considerable research effort has focussed on the transfer of benefits between woodlands (e.g. Loomis, 1995). The benefits of recreational demand are a function of the unit value of demand and the size of that demand measured in terms of visitor numbers. It has been argued that whilst the unit value of demand may vary between woodlands, the greatest variation in recreational value between sites is due to differences in the number of recreational visits meriting further research this area (Lovett et al., 1997). However, to date only a few studies have focussed on the development of models for predicting visitor numbers (Bateman et al., 1999; Jones et al., 2002). A key reason for this bias in research effort is likely to be the large volume of literature on the valuation of the non-market benefits of woodlands relative to the limited availability of reliable visit count data at a representative range of forests and woodland sites across the country. Data availability is also a key factor limiting the transferability of existing models, most being developed from datasets of questionable reliability and limited to forests and woodlands located within a relatively small geographical area. Other limitations include the restricted ability of earlier analyses to capture the spatial complexity of demand functions (Bateman et al., 1999), although recent advances in computing technology and, in particular geographical information systems (GIS), have raised expectations in this respect.

The aim of this study is to test a series of models that explain recreational visits to countryside woodlands in Great Britain (GB). The analysis draws on the analytical capabilities of GIS to address spatial variability in factors influencing recreation demand, and on visitor data collected at 100 forest and woodland sites, the largest dataset of visit counts compiled to date for forests and woodlands in GB. The method follows that employed by Brainard et al. (2001) who were successful in developing simple, yet robust functions to explain recreation demand in English woodlands. As these authors state, the simplicity of the approach is one of its key merits, and unlike similar studies which have tended to concentrate on measuring the distance-decay relationship between travel costs and recreation sites (Willis, 1991; Willis and Garrod, 1992; Bateman et al., 1996), the approach is not dependent upon the costly and time-consuming collection of visitor data at each site of interest. The research extends previous studies by incorporating a greater number of observation sites in the analysis, by expanding the geographical coverage of the research to include England, Scotland and Wales, and by encompassing data from private as well as publicly owned woodlands. It also tests the efficacy of using the developed models to predict visitor numbers at unsurveyed sites and advances discussion on some of the related issues that would help improve this procedure.

Method

The methods employed here can be traced back to ‘gravity’ models, which were typically used to model commuting decisions by regional economists. Elements of these models were evident in Hotelling’s (1949) Travel Cost Models (TCMs) which used the number of visitors from an origin zone as a dependent variable and travel cost from the zone as an explanatory variable (Hanley et al., 2001). TCMs subsequently evolved into trip generation functions (TGFs) which predicted an individual’s demand for a recreational trip, incorporating visitor and, later, site characteristics.
DEMAND ANALYSIS PROJECTIONS FOR RECREATIONAL VISITS

as explanatory variables. In recent years TGFs have been developed to predict visit rates and associated values for unsurveyed sites incorporating the type of variables described in this paper (see, e.g. Willis and Benson, 1989; Bateman et al., 1999).

Alternatives to the TGF approach include Expenditure Partition methods and Contingent Visit Approaches (CVAs), which can also be used to generate estimations of visitation behaviour at recreational sites. The partition method suffers from the fact that it only provides an indication of the relative importance of attributes, and is not able to quantify how these attributes affect visitation behaviour. The main drawback of the CVA approach, which relies on perceived changes in visitation behaviour given hypothetical changes in site characteristics, can result in the role of forest attributes in visitation behaviour being exaggerated relative to other factors (Roberts et al., 2000).

The approach adopted in this study to derive the transferable TGF is the zonal technique, where visit rates to given sites from a set of outset zones are predicted (Willis and Benson, 1989). The zonation is based on travel time(s) from the outset zone to the forest. The information required to model visit arrivals is provided within the TGF, the basic functional form of which is presented in equation (1):

\[ \text{Visits} = f(\text{Pop}, \text{Char}, \text{Att}, \text{Sub}, x), \]  

where Visits equals the annual number of visits made to the forest site under consideration, Pop is a variable for the size of the population that lives within each outset zone, Char are variables to indicate the socio-economic characteristics of the population within each outset zone (e.g. age, unemployment), Att are variable(s) to reflect the type and quality of facilities (e.g. forest trails) and other attributes (e.g. species mix) of the forest site, Sub is a variable to represent the type, quality and accessibility of substitute forest sites (based on an index of accessibility to alternative woodlands of varying type) and \( x \) describes a vector of other explanatory variables.

The basic analytical method involves fitting various linear regression models with the forest visitor counts from surveyed sites as the response variable. The resulting TGF predicts the mean values of the response variable for a linear combination of the explanatory variables, and shows, all things being equal, impacts of each explanatory variable on visit numbers. The objective of the modelling is to test regression models that help project the demand for recreational trips to woodland destinations. Inclusion of the described variables in the models is conducted on the basis of \( R^2 \) performance in univariate regressions and stepwise methods. Model performance is evaluated on the basis of the out-of-sample predictive power. To inform the multivariate regression models, a series of univariate regressions were first fitted between the explanatory variable and visit number data to assess independent influences over the variation in visit data. Following Brainard et al. (2001) the response variable used was the natural logarithm of the visit count.

Data

Visitor counts

For this study, visit number data were available for a total of 100 countryside woodland sites across GB, 42 in England, 41 in Scotland and 17 in Wales. ‘Countryside woodlands’ refers to woodlands in rural and semi-rural areas including woodlands on the urban fringe. Of these, 68 sites were managed by Forest Enterprise (FE), 17 by the Royal Society for the Protection of Birds (RSPB) and 15 sites by the Woodland Trust in Scotland (WTS) (Figure 1). Site names along with the visit numbers and methods used to collect the data are presented in Table 1.

The quality of the visit data (the response variable) is a key determinant of the ability of regression models to explain the variability in the data and therefore the ability of the TGF to predict future visits, while the extent to which the data is representative of all woodlands and forests in GB is a key determinant of the transferability of the TGF model. Data quality has been given surprisingly little consideration in earlier related studies. From a consideration of the geographical spread of sites, it is clear that very few of the woodlands in the sample are around the main centres of population. However, reliable visitor data is scarce, being collected on a systematic basis by only a small number of landowners, such as FE and a few not-for-profit organizations. What little data exists is subject to considerable error.
According to the National Inventory of Woodland and Trees (Smith and Gilbert, 2001), there are around 2.7 million ha of forest and woodland in GB. Around 52 per cent is owned by private/commercial interests. The nature of these sites and the extent to which they are publicly accessible is unknown. FE manages about 70 per cent of the remaining 1.2 million ha, the majority of which is open to the public. Charitable organizations collectively own about 90 000 hectares, about 3.5 per cent of all woodland, of which the RSPB and WTS own around a third. Data is
Table 1: Woodland sites with visitor count data

<table>
<thead>
<tr>
<th>Forest name</th>
<th>Count type</th>
<th>Visit numbers</th>
<th>Owner</th>
<th>Forest name</th>
<th>Count type</th>
<th>Visit numbers</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>England</td>
<td></td>
<td></td>
<td></td>
<td>Scotland continued</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Beechenhurst</td>
<td>V</td>
<td>182298</td>
<td>FC</td>
<td>Brighty Wood</td>
<td>M</td>
<td>1000</td>
<td>WTS</td>
</tr>
<tr>
<td>Bickley Gate</td>
<td>V</td>
<td>229258</td>
<td>FC</td>
<td>Cardrona</td>
<td>V</td>
<td>22353</td>
<td>FC</td>
</tr>
<tr>
<td>Birches Valley</td>
<td>VC</td>
<td>89877</td>
<td>FC</td>
<td>Ciste Dubh</td>
<td>V</td>
<td>12485</td>
<td>FC</td>
</tr>
<tr>
<td>Blean Woods</td>
<td>*</td>
<td>8915</td>
<td>RSPB</td>
<td>Countesswells</td>
<td>V</td>
<td>100404</td>
<td>FC</td>
</tr>
<tr>
<td>Blidworth Bottoms</td>
<td>V</td>
<td>233597</td>
<td>FC</td>
<td>Crinan</td>
<td>P</td>
<td>4000</td>
<td>WTS</td>
</tr>
<tr>
<td>Bull Crag</td>
<td>V</td>
<td>9626</td>
<td>FC</td>
<td>Den Wood</td>
<td>M</td>
<td>2000</td>
<td>WTS</td>
</tr>
<tr>
<td>Cannock</td>
<td>V</td>
<td>143425</td>
<td>FC</td>
<td>Dog Falls</td>
<td>V</td>
<td>21460</td>
<td>FC</td>
</tr>
<tr>
<td>Chopwell</td>
<td>V</td>
<td>99903</td>
<td>FC</td>
<td>Formonthills</td>
<td>M</td>
<td>11000</td>
<td>WTS</td>
</tr>
<tr>
<td>Clay Bank</td>
<td>V</td>
<td>94140</td>
<td>FC</td>
<td>Glen Doll</td>
<td>P&amp;D</td>
<td>56603</td>
<td>FC</td>
</tr>
<tr>
<td>Coombs &amp; Churnet Valley</td>
<td>*</td>
<td>8636</td>
<td>RSPB</td>
<td>Glen Fingus</td>
<td>P</td>
<td>13000</td>
<td>WTS</td>
</tr>
<tr>
<td>Cannop</td>
<td>P&amp;D</td>
<td>26603</td>
<td>FC</td>
<td>Glenmore</td>
<td>C, ME</td>
<td>68380</td>
<td>FC</td>
</tr>
<tr>
<td>Dalby Forest</td>
<td>V</td>
<td>411338</td>
<td>FC</td>
<td>Glenfress</td>
<td>V</td>
<td>55372</td>
<td>FC</td>
</tr>
<tr>
<td>Derbyshire Bridge</td>
<td>V</td>
<td>148711</td>
<td>FC</td>
<td>Inch Marshes</td>
<td>*</td>
<td>8900</td>
<td>RSPB</td>
</tr>
<tr>
<td>Grizedale</td>
<td>VC</td>
<td>44406</td>
<td>FC</td>
<td>Inchree</td>
<td>P</td>
<td>26695</td>
<td>FC</td>
</tr>
<tr>
<td>Guisborough</td>
<td>V</td>
<td>125355</td>
<td>FC</td>
<td>Ken-Dee Marshes</td>
<td>*</td>
<td>3815</td>
<td>RSPB</td>
</tr>
<tr>
<td>Hamsterley</td>
<td>V</td>
<td>172425</td>
<td>FC</td>
<td>Kirkhill Main</td>
<td>V</td>
<td>39780</td>
<td>FC</td>
</tr>
<tr>
<td>Hepburn</td>
<td>V</td>
<td>13125</td>
<td>FC</td>
<td>Ledmore and Migdale</td>
<td>M</td>
<td>1000</td>
<td>WTS</td>
</tr>
<tr>
<td>High lodge</td>
<td>FDT</td>
<td>127068</td>
<td>FC</td>
<td>Moncreif Hill</td>
<td>P</td>
<td>16000</td>
<td>WTS</td>
</tr>
<tr>
<td>Kielder Castle</td>
<td>V</td>
<td>23464</td>
<td>FC</td>
<td>Otter Hide</td>
<td>P</td>
<td>29936</td>
<td>FC</td>
</tr>
<tr>
<td>Leighten Moss</td>
<td>Veh</td>
<td>85546</td>
<td>RSPB</td>
<td>Portmoak Moss</td>
<td>P</td>
<td>11000</td>
<td>WTS</td>
</tr>
<tr>
<td>Longdale Lane</td>
<td>Veh</td>
<td>101126</td>
<td>FC</td>
<td>Pressmennan Wood</td>
<td>P</td>
<td>9000</td>
<td>WTS</td>
</tr>
<tr>
<td>May Beck</td>
<td>V</td>
<td>33343</td>
<td>FC</td>
<td>Queen Elizabeth FP</td>
<td>V</td>
<td>50383</td>
<td>FC</td>
</tr>
<tr>
<td>Minsmere</td>
<td>*</td>
<td>69371</td>
<td>RSPB</td>
<td>R.Affric</td>
<td>V</td>
<td>31717</td>
<td>FC</td>
</tr>
<tr>
<td>Nagshead</td>
<td>*</td>
<td>13860</td>
<td>RSPB</td>
<td>Rogie Falls</td>
<td>P</td>
<td>56052</td>
<td>FC</td>
</tr>
<tr>
<td>Newtondale</td>
<td>V</td>
<td>63863</td>
<td>FC</td>
<td>Scolty</td>
<td>V</td>
<td>57641</td>
<td>FC</td>
</tr>
<tr>
<td>Nobel Knott</td>
<td>V</td>
<td>10723</td>
<td>FC</td>
<td>Shooting Greens</td>
<td>V</td>
<td>18353</td>
<td>FC</td>
</tr>
<tr>
<td>Normans Hill</td>
<td>V</td>
<td>248044</td>
<td>FC</td>
<td>St. Ronans</td>
<td>M</td>
<td>8000</td>
<td>WTS</td>
</tr>
<tr>
<td>Northward Hill</td>
<td>*</td>
<td>983</td>
<td>RSPB</td>
<td>Strathyre Car Park</td>
<td>V</td>
<td>4356</td>
<td>FC</td>
</tr>
<tr>
<td>Pexton</td>
<td>V</td>
<td>515705</td>
<td>FC</td>
<td>Suttershills Grove</td>
<td>V</td>
<td>25018</td>
<td>FC</td>
</tr>
<tr>
<td>Pinchinthorpe</td>
<td>VC</td>
<td>16111</td>
<td>FC</td>
<td>Tyrebagger West</td>
<td>V</td>
<td>23492</td>
<td>FC</td>
</tr>
<tr>
<td>Rigg Lane Tower</td>
<td>V</td>
<td>166273</td>
<td>FC</td>
<td>Vane Farm</td>
<td>*</td>
<td>39729</td>
<td>RSPB</td>
</tr>
<tr>
<td>Sherwood Forest Drive</td>
<td>V</td>
<td>82953</td>
<td>FC</td>
<td>Warout Wood</td>
<td>M</td>
<td>100000</td>
<td>WTS</td>
</tr>
<tr>
<td>Sherwood Pines Forest Park*</td>
<td>V</td>
<td>227795</td>
<td>FC</td>
<td>Wood of Cree</td>
<td>*</td>
<td>5816</td>
<td>RSPB</td>
</tr>
</tbody>
</table>
### Table 1: Continued

<table>
<thead>
<tr>
<th>Forest name</th>
<th>Count type</th>
<th>Visit numbers</th>
<th>Owner</th>
<th>Forest name</th>
<th>Count type</th>
<th>Visit numbers</th>
<th>Owner</th>
</tr>
</thead>
<tbody>
<tr>
<td>Simonside V</td>
<td>V</td>
<td>23598</td>
<td>FC</td>
<td>Wales</td>
<td>V</td>
<td>20430</td>
<td>FC</td>
</tr>
<tr>
<td>Symonds Yat Rock</td>
<td>V</td>
<td>176195</td>
<td>FC</td>
<td>Black Covert</td>
<td>P&amp;D</td>
<td>5635</td>
<td>FC</td>
</tr>
<tr>
<td>The Lodge</td>
<td>*</td>
<td>39603</td>
<td>RSPB</td>
<td>Coed Y Brenin VC</td>
<td>*</td>
<td>7190</td>
<td>RSPB</td>
</tr>
<tr>
<td>The Street</td>
<td>V</td>
<td>255327</td>
<td>FC</td>
<td>Coed Aber</td>
<td>VC</td>
<td>34129</td>
<td>FC</td>
</tr>
<tr>
<td>Thieves Wood</td>
<td>V</td>
<td>225710</td>
<td>FC</td>
<td>Cwn Clydach</td>
<td>*</td>
<td>32253</td>
<td>RSPB</td>
</tr>
<tr>
<td>Westonbirt Arboretum</td>
<td>P</td>
<td>156773</td>
<td>FC</td>
<td>Llyn Parc Mawr</td>
<td>P&amp;D</td>
<td>8690</td>
<td>FC</td>
</tr>
<tr>
<td>Whinlatter Main</td>
<td>V</td>
<td>75140</td>
<td>FC</td>
<td>Lake Vyrnwy</td>
<td>*</td>
<td>32253</td>
<td>RSPB</td>
</tr>
<tr>
<td>Wolves and Ramsey Woods</td>
<td>*</td>
<td>3650</td>
<td>RSPB</td>
<td>Llyn Crafnant</td>
<td>P&amp;D</td>
<td>8988</td>
<td>FC</td>
</tr>
<tr>
<td>Wyre</td>
<td>P&amp;D</td>
<td>71550</td>
<td>FC</td>
<td>Abernethy Forest</td>
<td>*</td>
<td>31477</td>
<td>RSPB</td>
</tr>
<tr>
<td>Abernethy Forest</td>
<td>M</td>
<td>1000</td>
<td>WTS</td>
<td>Lyn Geirionydd</td>
<td>M</td>
<td>1000</td>
<td>WTS</td>
</tr>
<tr>
<td>Abriachan Wood</td>
<td>V</td>
<td>11065</td>
<td>FC</td>
<td>Moel Famau</td>
<td>P&amp;D</td>
<td>32695</td>
<td>FC</td>
</tr>
<tr>
<td>Alt Na Calliche</td>
<td>M</td>
<td>38000</td>
<td>WTS</td>
<td>Nant yr Arian</td>
<td>V</td>
<td>87833</td>
<td>FC</td>
</tr>
<tr>
<td>Beinn a'Mheadhoin</td>
<td>V</td>
<td>8494</td>
<td>FC</td>
<td>Newborough Beach</td>
<td>V</td>
<td>147279</td>
<td>FC</td>
</tr>
<tr>
<td>Bellsquary</td>
<td>M</td>
<td>15000</td>
<td>WTS</td>
<td>The Arch</td>
<td>V</td>
<td>23651</td>
<td>FC</td>
</tr>
<tr>
<td>Ben Venue Dukes Pass</td>
<td>P</td>
<td>15872</td>
<td>FC</td>
<td>Tynybedw</td>
<td>V</td>
<td>4995</td>
<td>FC</td>
</tr>
<tr>
<td>Blaeberry Community Wood</td>
<td>M</td>
<td>15000</td>
<td>WTS</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Veh = vehicle counter (pneumatic); P&D = pay and display ticket; Cyc = cycle; VC = visitor centre; M = aggregated manual counts; V = vehicle counter; P = pedestrian counter; C = counter; ME = magic eye; FDT = forest drive tickets.

*Unknown.
collected at only a small number of sites by each organization. Due to lack of data, the extent to which these sites are representative of the holdings of these organizations cannot be determined, nor can we determine the extent to which the sites included in the analysis are representative of publicly accessible woodlands in GB.

The scarcity of data meant that visit counts for different 12-month periods had to be combined for the three ownership groups to form a single dataset. FE visit data were for the calendar year 1999, RSPB visit data were for the year April 1999–March 2000 and only visit data for the 12 months from March 2000 were available for WTS sites (Woodland Trust Scotland, 2001). This approach was considered acceptable based on a comparison of means for 1999 and 2000 FE visit data, which showed no significant difference in the visit data between the 2 years ($t = 0.40, P = 0.968$). However, the rounded nature of the WTS data raised further concerns over its reliability.

As with previous studies, data were collected using a variety of different counting mechanisms, some of which may be more reliable than others, including electronic vehicle counters, pay and display counters, pedestrian counters, pneumatic counters, ‘magic eyes’, forest drive tickets, cycle counters and manual ‘hand’ counts. Depending on the mechanism, it can be necessary to translate counts into individual visits using assumptions such as the average number of individuals per car parked at a pay and display car park. Where data were available for less than 12 months, data for missing months were estimated by the managing organization. Where possible, manual data collected for certain months were assumed to be representative of visitor numbers for missing months allowing an annual estimate to take place.

There was also some inconsistency in the definitions of a ‘site’ between the three organizations introducing further uncertainty into the data and a degree of ambiguity in terms of what was being modelled. The RSPB and WTS sites are nature reserves or areas of estate characterized by a significant presence of woodland. Visits are generally counted at all access points to each site and represent the visits to the whole area, not just to the woodland. Consequently, woodland visits may be overstated. In contrast, the FE data were collected at single access points to a forest area (not specifically a block or compartment). These forest areas may have had more than one access point; thus, the visit numbers for FE woodland may be understated.

Site attributes data

Previous studies have shown site attributes to be significant factors influencing visit rates to woodland recreation sites. Based on a review of the literature – assuming a reasonable degree of correspondence between the attributes that influence willingness to pay (WTP) and those that influence visit rates – (Englin and Mendelsohn, 1991; Hanley and Ruffel, 1993; Bostedt and Mattsson, 1995; Scarpa et al., 2000; and Brainard et al., 2001) a range of factors were selected for inclusion in the analysis, as determined by the forest manager. Data on site facilities and characteristics were collected for the 100 sites through a survey of forest managers (Table 2).

Site characteristics included forest age, size and species mix. Following Brainard et al. (2001) data were collected on both the main woodland stand (BROAD) and on the woodland experienced by the majority of the visitors at the forest entrance (SPENT). Data on 27 site facilities were recorded including presence of a car park, picnic site, and visitor centre and data on the number and length of trails. Three additional ‘facility index’ variables were calculated to indicate the number of facilities present at each site. The interpretation of index variables in the models is less straightforward than the individual variables. However, they are included as the likelihood of multicollinearity between attributes and the influence over degrees of freedom can present problems in the analysis. An unweighted facility index was calculated for all facilities included in the survey (UNWT27). All attributes were assigned a value of 0.1. Thus, if a site had five attributes it would have an index of 0.5. This index, which was in the interval 0,2.7, included eight sub-attributes (wildlife activities, cafe, shop, bike hire, toilets, disabled walks and access to shop and fishing), which could not be easily recognized as distinct site attributes in their own right (i.e. they were usually provided jointly with other attributes). Consequently, a second unweighted index was calculated for the 19 main attributes in the interval 0,1.9 (UNWT19). To improve on the method of Brainard et al. (2001), a third index was calculated which weighted these 19 attributes...
based on their popularity with visitors to forest sites. The weighting was derived from rankings of facilities from a random sample survey of 1900 respondents at 44 of the 100 forest sites undertaken in the summer of 2002. Respondents were asked to identify which attributes had had a positive influence on their decision to visit the site and to rank these selected attributes in terms of the influence on their decision to visit. A weighted index was then created which took into account the frequency that an attribute was cited as being important in the decision to visit, and the mean importance of the attribute in the decision to visit. For convenience the index was taken in the interval 0,1.

Population and socio-economic variables

Previous studies have found that the distance-decay function between a site and its potential users is the dominant population-related factor in explaining visitor numbers (Brainard et al., 2001). However, the socio-economic characteristics of the population are also likely to be key determinants of recreation demand. In order to represent both the proximity and nature of the local population, demographic data drawn from the 1991

<table>
<thead>
<tr>
<th>Table 2: Sample of site attributes collected in forest manager survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest attributes</td>
</tr>
<tr>
<td>---------------------------------------------------------------</td>
</tr>
<tr>
<td><strong>Forest characteristics</strong></td>
</tr>
<tr>
<td>Forest size</td>
</tr>
<tr>
<td>Average age of whole forest stand</td>
</tr>
<tr>
<td>Species mix for main stand*</td>
</tr>
<tr>
<td>Species mix around site entrance*</td>
</tr>
<tr>
<td><strong>Facilities</strong></td>
</tr>
<tr>
<td>Car park</td>
</tr>
<tr>
<td>Car park capacity</td>
</tr>
<tr>
<td>Picnic site</td>
</tr>
<tr>
<td>Forest walk (number and length)</td>
</tr>
<tr>
<td>Cycle trail</td>
</tr>
<tr>
<td>Horse riding route</td>
</tr>
<tr>
<td>Number of trails (all types)</td>
</tr>
<tr>
<td>Average length of trails (all types)</td>
</tr>
<tr>
<td>Orienteering course</td>
</tr>
<tr>
<td>Play equipment</td>
</tr>
<tr>
<td>Forest drive</td>
</tr>
<tr>
<td>Viewpoint</td>
</tr>
<tr>
<td>Hides</td>
</tr>
<tr>
<td>Camping/caravan site</td>
</tr>
<tr>
<td>Visitor centre</td>
</tr>
<tr>
<td>Cafe</td>
</tr>
<tr>
<td>Shop</td>
</tr>
<tr>
<td>Interpretation centre</td>
</tr>
<tr>
<td>Forest classroom</td>
</tr>
<tr>
<td>Toilets</td>
</tr>
<tr>
<td>Disabled facilities (toilets, walks etc.)</td>
</tr>
<tr>
<td>Water feature/fishing</td>
</tr>
<tr>
<td>Unweighted index of 19 attributes</td>
</tr>
<tr>
<td>Unweighted index of 27 attributes</td>
</tr>
<tr>
<td>Weighted index of 19 attributes</td>
</tr>
</tbody>
</table>

N = a continuous numeric variable; P/A = presence/absence (0/1).
Log refers to the transformation of the data by natural logarithm.
*Species mix coded: 1 = broadleaved, 0 = coniferous/mixed.
were calculated for the population of six travel time zones around each forest site. Following the techniques described in Lovett et al. (1997) and Jones et al. (2002), zones were defined using cost-distance modelling based on travel time within 30, 45, 60, 90 and 120 min of the forest site. Based on findings in earlier studies, data for 11 key demographic characteristics were assembled for each zone. Demographic data was extracted from the 1991 census based on enumeration district polygons in England and Wales and census output area polygons in Scotland. These spatial units represented the possible outset zones for visitors. GIS techniques were subsequently employed to create a unique identifier for each spatial unit which was then used to extract the relevant demographic data from the Small Area Statistics data tables. In addition to population size, data were collected on indicators to represent affluence, deprivation, age, ethnicity, access to transport and higher education (Table 3). Individual demographic variables were summed for each time band to obtain the total population for that variable. These variables were then respecified to represent proportional indicators (taken in the interval 0,1) of the population or household in the time band as appropriate.

The Census data only provide information on the resident population. However, tourists can account for a significant proportion of visits to many forest sites. In this study the definition of a tourism visit is based on the distinction between leisure and tourism day trips in the GBDVS, (2003), which defines as tourists those visitors on holiday staying away from home or on a day visit from home of duration longer than 3 h. As part of this study, a survey of 1900 forest visitors found that the average proportion of visits made by tourists was 0.52 (SD = 0.31, N = 44). This proportion tended to be higher in more rural areas and popular holiday destinations. This suggests that the size of the tourist population around forest sites may be a significant factor in influencing visit numbers. However, data on tourist populations is generally unavailable at the equivalent spatially explicit level as residence data, and consequently was not included in the population variables.

**Substitute woodland destination variables**

The proximity of alternative recreation sites to visitor outset zones is also likely to be a key influence on the number of visitors to any given woodland location (Bockstael, 1995). For any individual visitor there may be a great many substitute recreation sites, including attractions in towns, cities, beaches and other countryside destinations as well as other woodlands. Calculating a substitution function for each substitute site presents a considerable challenge, as the substitution function is likely to differ between types of substitute sites as well as between sites within a single type. Lovett et al. (1997) suggest that the data available for these calculations, such as spatially explicit information of average driving speeds, are not yet adequate to inform their calculation, whilst Brainard et al. (2001) argue for simple models that exclude measures of access to substitute sites altogether due to the complexity of their

**Table 3: Socio-economic variables considered in correlation analysis**

<table>
<thead>
<tr>
<th>Demographic variable for each time zone</th>
<th>Short name</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total population</td>
<td>POP</td>
</tr>
<tr>
<td>Number of households classified as social class 1 or 2*</td>
<td>SOC</td>
</tr>
<tr>
<td>Number of owner-occupied households</td>
<td>OWN</td>
</tr>
<tr>
<td>Number of economically active male population unemployed</td>
<td>UNPER</td>
</tr>
<tr>
<td>Number of adult population in local authority/housing association accommodation</td>
<td>PERLA</td>
</tr>
<tr>
<td>Number of population aged &lt;9 years</td>
<td>PERUN9</td>
</tr>
<tr>
<td>Number of households with dependent children</td>
<td>DEPCH</td>
</tr>
<tr>
<td>Number of households with retired head</td>
<td>RETIR</td>
</tr>
<tr>
<td>Number of population classified as ethnic (Black, Indian, Pakistan, Bangladesh, Chinese)</td>
<td>PERETH</td>
</tr>
<tr>
<td>Number of households with no car</td>
<td>CAR</td>
</tr>
<tr>
<td>Number of population over 18 with higher degrees</td>
<td>PERHI</td>
</tr>
</tbody>
</table>

*Occupational (social) group: 1 = professional and technical occupations; 2 = managerial occupations.
measurement. However, advancements in GIS technology make the calculation of these variables increasingly straightforward, even if the data on which they are based are limited. Here, only indicators of proximity to substitute woodland sites were estimated. Size of woodland was used as a proxy for woodland quality. Based on the method outlined by Jones et al. (2002), substitution indices were created for four woodland types in each time zone: small (<50 ha), medium (50–100 ha), large (>100 ha) and 'all woodlands' (i.e. small, medium and large). This allowed the availability of other woodlands in the proximity of the visitor outset locations to be quantified, using data assigned to enumeration districts or census output area centroids. First, Bartholomew’s digital database was used to compute a travel time surface for each woodland type using GIS. A ratio for each woodland type was then calculated based on the ratio between the area of woodland type to total woodland area. Because it is assumed that larger forests are more likely to draw visitors than smaller forests, the calculated ratios were then used as a weighting mechanism by multiplying the ratio for each woodland type by the corresponding travel time surface. These resulting accessibility surfaces constituted indices of substitute accessibility for the four woodland types. To create a final set of substitute accessibility scores for inclusion in the analysis, the three accessibility indices were assigned to each enumeration district or census output area centroid cell and an average for the six time bands was calculated in GIS for each of the three substitute accessibility scores.

Results

Univariate regression analysis

Of the forest characteristics, forest size and the dominant species mix were found to be significant explanatory variables with an $R^2$ of 0.157 and 0.258, respectively ($P < 0.1$). Ownership explained the greatest amount of variability in the data ($R^2 = 0.277$) of all attribute variables. Of the individual site facility variables, the presence of a visitor centre ($R^2 = 0.115$), picnic sites ($R^2 = 0.181$), toilets ($R^2 = 0.147$) and cycle trails ($R^2 = 0.111$) were all significant explanatory variables. However, as with forest size and species mix, picnic sites, toilets and cycle trails were shown to be highly correlated with ownership. When RSPB and WTS woodlands are removed from the dataset on visit rates, the explanatory power of these variables is significantly reduced. In contrast, the explanatory power of the presence of a visitor centre increased when only FE sites were considered. Car park capacity was another facility attribute found to be an important explanatory factor. However, as in the study of Brainard et al. (2001), the correlative, rather than causative, nature of this link with visit rates also limits its value for use in predictive models. While the same could be true for other variables (e.g. toilets), it is likely to be more prominent for car parks because, on the one hand, ease of parking is likely to be a particularly important factor in the decision to visit. However, on the other hand, one cannot assume that large car parks will attract more visitors than smaller ones. Both the unweighted (UNWHT19) and weighted attribute index (ATTINDEX) were statistically significant ($P < 0.01$), although the unweighted index had a considerably higher $R^2$ of 0.138. This result is contrary to expectation and may be due to the quality of the visit number data, the survey data used to derive the weightings or to the weighting method itself. Nevertheless, the unweighted attribute index overcomes the problem of accounting for a large number of attributes within a relatively small sample size, as well as issues of multi-collinearity within the attribute set.

The total population within a 2-h drive of the forest sites explained the greatest variability in the data of the population variables ($R^2 = 0.144$). It was noted that when RSPB and WTS woodlands were removed from the dataset, the explanatory power of this variable significantly increased ($R^2 = 0.304$). This effect was consistent across all time zones, possibly raising questions over the reliability of the RSPB and WTS data. Interestingly, when only FE data were considered, the population within the 90-min time zone explained the greatest variability in the data of the population variables ($R^2 = 0.304$). Of the socio-economic characteristics of the population, the proportion of the population which is <9 years (PERUN9120), the proportion of households where the head is retired (RETIR120), the proportion of owner-occupied households (OWN120) and the proportion of the
population in local authority housing (PERLA120) were all shown to be statistically significant. Surprisingly, car ownership (CAR120) and social class (SOC120) showed no statistically significant explanatory power. The substitute index for all woodlands within a 2-h travel time from outset zones (SALL120) was found to be significant ($R^2 = 0.111$). Again, when only FE sites were considered, the explanatory power of the variable increased ($R^2 = 0.153$), although accessibility to large substitute woodland sites had the greatest explanatory power ($R^2 = 0.168$). Table 4 summarizes the results of the univariate regressions for those variables that were significant and which were selected to form the set of explanatory variables entered into the stepwise multivariate regression models. It was necessary to preselect variables in such a way due to restrictions on degrees of freedom imposed by the sample size, which in this case was limited to 100 observations, and problems of multi-collinearity of variables which effectively acted as surrogates for others. Using the results of the univariate regression, the most significant variables were selected from four predictor groups where potential for multi-collinearity was high. These groups were attribute variables, population variables, substitute woodland variables and socio-economic variables. Within the attribute group, the unweighted index proved a significant predictor of recreation demand (more significant than the weighted index) and overcame the problem of accounting for a large number of attributes in a relatively small sample size. Its use also overcame the problem of multi-collinearity which was inherent in the presence/absence attribute set. For example, many sites with a cycle trail also have a forest walk and those with a visitor centre also have toilets etc. Some attributes, such as cycle trails and picnic sites, were also found to act as surrogates for forest ownership, as was the species dummy which was dropped for the same reason.

### Multivariate regression models

Backward stepwise multiple linear regression was used to generate a number of TGF models to forecast visitor arrivals at GB woodlands. Tolerance statistics were examined in all models to check for possible multi-collinearity among predictor variables, of which no instances were found. As the univariate regression analyses suggested that the FE data may be more reliable than the RSPB and WTS data, models were run for all-site data and FE data only. To avoid problems of multicollinearity with the size of population and substitute woodland variables, only one was considered at any time in the multiple regression analysis. In other words, in cases where the population variable was entered, the substitute woodland variable was excluded and vice versa.

In model I (Table 5), four variables explain considerably more than half of the variation in

### Table 4: Results of the simple regressions for variables included in the multiple regression

<table>
<thead>
<tr>
<th>Variable name</th>
<th>Variable description</th>
<th>t ($P &lt; 0.01$)</th>
<th>$R^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>OWNDUM</td>
<td>Forest Ownership (1 FE, 0 WT and RSPB)</td>
<td>6.239</td>
<td>0.277</td>
</tr>
<tr>
<td>COUNTRY*</td>
<td>Country (1 England, 0 Scot and Wales)</td>
<td>5.320</td>
<td>0.216</td>
</tr>
<tr>
<td>(log)CARCAP</td>
<td>Car park capacity (no. of spaces)</td>
<td>5.321</td>
<td>0.216</td>
</tr>
<tr>
<td>(log)UNWHT19</td>
<td>Unweighted attribute index</td>
<td>4.099</td>
<td>0.138</td>
</tr>
<tr>
<td>(log)FORSIZE</td>
<td>Forest size in ha</td>
<td>4.413</td>
<td>0.157</td>
</tr>
<tr>
<td>(log)POP120</td>
<td>Population within 120 min travel time</td>
<td>4.208</td>
<td>0.144</td>
</tr>
<tr>
<td>(log)POP90</td>
<td>Population within 90 min travel time</td>
<td>4.075</td>
<td>0.136</td>
</tr>
<tr>
<td>SALL120</td>
<td>Substitute index for all woodlands within 120 min travel time of outset zone</td>
<td>3.655</td>
<td>0.111</td>
</tr>
<tr>
<td>SLRG120</td>
<td>Substitute index for large woodlands within 120 min travel time of outset zone</td>
<td>3.482</td>
<td>0.101</td>
</tr>
<tr>
<td>RETIR120</td>
<td>Proportion of households with retired head</td>
<td>2.839</td>
<td>0.067</td>
</tr>
<tr>
<td>PERUN9120</td>
<td>Proportion of population &lt;9 years</td>
<td>3.427</td>
<td>0.098</td>
</tr>
</tbody>
</table>

*1 = England; 0 = Scotland and Wales (Scotland and Wales combined due to degrees of freedom).
the visitor counts \((R^2 = 0.552)\) at all forest sites: forest ownership, the population within a 2-h drive time, the number of forest attributes and car park capacity.

Confidence intervals are also included to indicate the degree of certainty that can be applied to the coefficients. The range of values between the lower and upper bound limits indicates how certain we can be about the values of \(\beta\) computed for each variable. These relatively wide intervals are likely to be indicative of the quality of visit data available for the study. Acquisition of more, and certainly better quality, visit data would be advisable before applying these parameter estimates to predict visitor demand at unsurveyed sites. However, the variation in visit rates explained by the predictors, and overall statistical significance of the model implies that some useful observation can be made.

As discussed previously, the value of forest ownership and car park capacity as explanatory variables is questionable. Excluding these from the all-site analysis, along with forest size which correlated with ownership, generates model II (Table 6). This has less explanatory power \((R^2 = 0.408)\), but is made up of more easily explained variables: the number of facilities present at the site, the population within a 2-h drive time of the site and two socio-economic characteristics of the local population – the proportion of retired households and households with young families. It is not unreasonable to expect both these groups to visit forests for recreation more than other socio-economic groups. All variables are highly significant and have a positive influence on visit numbers. Thus on balance, model II presents a fairly robust TGF for all sites.

Table 5: Model I

<table>
<thead>
<tr>
<th>Model I (all sites, all selected variables)</th>
<th>Coefficient ((\beta))</th>
<th>(t) Statistics</th>
<th>95% Confidence interval for (\beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>((\log))VIS_COUNT</td>
<td>3.980</td>
<td>4.108***</td>
<td>Lower 2.056 Upper 5.904</td>
</tr>
<tr>
<td>OWNDUM</td>
<td>+1.326</td>
<td>6.248***</td>
<td>Lower 0.905 Upper 1.748</td>
</tr>
<tr>
<td>((\log)) POP120</td>
<td>+0.340</td>
<td>4.883***</td>
<td>Lower 0.202 Upper 0.478</td>
</tr>
<tr>
<td>((\log)) UNWHT19</td>
<td>+0.537</td>
<td>3.315***</td>
<td>Lower 0.215 Upper 0.859</td>
</tr>
<tr>
<td>((\log)) CARCAP</td>
<td>+0.191</td>
<td>1.751*</td>
<td>Lower -0.026 Upper 0.408</td>
</tr>
<tr>
<td>(R^2 = 0.552)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Significance levels of \(t\): *\(P < 0.1\), **\(P < 0.01\).

Given concerns over the quality of the visit data provided by the RSPB and WTS, regressions are run on FE data only to generate model III (Table 7). Interestingly, the model with the greatest explanatory power retains the unweighted index of site attributes. Other explanatory variables in model III include a measure of population, this time the population within a 90-min drive time of the site, and the index of accessibility to alternative large woodlands. The explanatory power of the model is also increased, with over half the variation \((R^2 = 0.508)\) in the data explained by only three variables. Confidence intervals also imply a greater degree of certainty about the relative influence of included parameters on visitor demand. Again, all variables are highly significant and have a positive influence on visit numbers. The explanatory variables in model III reinforce expectations of the key factors influencing recreation demand.

The increased ability of the model to explain greater variation in the FE data is a further indication that the FE data may be more reliable than the WTS and RSPB data.

Model transfer and validation

In principle, these models can be used to predict visitor numbers to other forest sites and the efficacy of transferring the models can be tested. Following Jones et al. (2002), the models were cross-validated by refitting a series of ‘omit’ models to predict visitor numbers for forest sites systematically excluded from the sample. (This is preferable to excluding a subset of observations for validation at the outset, as it does not have adverse effects on degrees of freedom in the main model.) A series of regressions are run, each omitting one site. The revised coefficients are used...
to estimate the predicted annual number of visits to the omitted site. This is done for all sites and provides a form of cross-validation in that the same dataset is not being used to assess the quality of predictions as is used to develop the model. Thus, prediction errors will not have an over-optimistic bias. An observed-to-predicted ratio is then calculated to assess validity of the model. Table 8 shows the proportion of sites where the predicted number of visits fell within ±10, ±50, ±75 and ±100 per cent of the observed values respectively.

Model II predicted within ±10 per cent (observed-to-predicted ratio between 0.90 and 1.11) of the observed number of visits for only 4 per cent of all sites. Predictions of visits for 36 per cent of the sites were within ±50 per cent of the observed number of visits (observed-to-predicted ratio between 0.66 and 2), and predictions for 65 per cent were within ±75 per cent (ratio >0.57 and <4). A small number of sites had adversely high/low observed-to-predicted ratios. The results for model III, the model for FE sites only, show a significant improvement in the accuracy of the predictions over the ‘all sites’ model. Model III predicted within ±10 per cent of the observed number of visits for 11 per cent of the sites, within ±50 per cent on 44 per cent of the sites and within ±75 per cent of visits on 80 per cent of the sites. For this model, the difference between observed and predicted was greater than 100 per cent on only one site.

At the site level, estimates from both models II and III were relatively evenly split between underpredictions and overpredictions. Across all sites, total visit numbers were predicted within 33 and

**Table 6: Model II**

<table>
<thead>
<tr>
<th>Model II (all sites, excluding ownership, size and car park capacity)</th>
<th>Coefficient (β)</th>
<th>t Statistics</th>
<th>95% Confidence interval for β</th>
</tr>
</thead>
<tbody>
<tr>
<td>(log)VIS_COUNT</td>
<td>-14.061</td>
<td>-2.604**</td>
<td>(-24.790, -3.332)</td>
</tr>
<tr>
<td>RETIR120</td>
<td>+1.653</td>
<td>2.120**</td>
<td>(0.103, 3.202)</td>
</tr>
<tr>
<td>PERUN9120</td>
<td>+1.415</td>
<td>3.665***</td>
<td>(0.648, 2.182)</td>
</tr>
<tr>
<td>(log)UNWHT19</td>
<td>+0.733</td>
<td>4.241***</td>
<td>(0.390, 1.076)</td>
</tr>
<tr>
<td>(log)POP120</td>
<td>+0.160</td>
<td>1.815***</td>
<td>(-0.015, 0.334)</td>
</tr>
</tbody>
</table>

$R^2 = 0.408$

Significance levels of $t$: **$P < 0.05$, ***$P < 0.01$.

**Table 7: Model III**

<table>
<thead>
<tr>
<th>Model III (FE sites only)</th>
<th>Coefficient (β)</th>
<th>t Statistics</th>
<th>95% Confidence interval for β</th>
</tr>
</thead>
<tbody>
<tr>
<td>(log)VIS_COUNT</td>
<td>5.044</td>
<td>5.637***</td>
<td>(3.254, 6.834)</td>
</tr>
<tr>
<td>(log)SLRG120</td>
<td>+1.539</td>
<td>3.546***</td>
<td>(0.671, 2.408)</td>
</tr>
<tr>
<td>(log)UNWHT19</td>
<td>+0.735</td>
<td>4.083***</td>
<td>(0.375, 1.095)</td>
</tr>
<tr>
<td>(log)POP90</td>
<td>+0.247</td>
<td>3.813***</td>
<td>(0.118, 0.377)</td>
</tr>
</tbody>
</table>

$R^2 = 0.508$

Significance levels of $t$: ***$P < 0.01$.

**Table 8: Proportion of sites with observed-to-predicted ratios falling within specified range**

<table>
<thead>
<tr>
<th>Range of observation to prediction</th>
<th>Model II, all sites (%)</th>
<th>Model III, FE sites only (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within ±10%</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>Within ±50%</td>
<td>36</td>
<td>44</td>
</tr>
<tr>
<td>Within ±75%</td>
<td>65</td>
<td>80</td>
</tr>
<tr>
<td>Within ±100%</td>
<td>78</td>
<td>99</td>
</tr>
</tbody>
</table>
23 per cent of the observed number, respectively, indicating that at an aggregate level the models are moderately effective at predicting visitor arrivals to individual sites that have not been used to help derive the coefficients.

Overall, the results of the validation suggest that model III is a reasonably reliable predictor of visits to FE sites outside the sample on which it was based. Thus, the application of this model to predict visits at other FE sites could be justified if the sites used in the analysis were representative of other FE holdings across GB. However, this was not statistically established due to limited data availability.

**Discussion**

The models presented here show how a relatively high degree of variation in visit data can be explained by only a few variables. The results of the analysis reinforce many of the findings from previous studies. As in Brainard et al. (2001), car park capacity was found to be a significant factor influencing visit rates, although interpretation of this remains ambiguous. Here, it was a measure of the number of site attributes as measured by the index, rather than individual site facilities, that were found to be significant. Alongside these relatively easily collected and reliably measured variables, the models show the key importance of the size and nature of the local population, as well as measures of accessibility to alternative forest recreation sites in understanding visit rates to forests. Again, as with Brainard et al. (2001), of the population variables it was the number of people living within a 2-h drive of the forest site that had the strongest statistical relationship with the number of visits.

Population and substitute-site variables are less easy to calculate than attribute variables, relying on sophisticated GIS analysis to generate the data. Brainard et al. (2001) suggest using simpler models that require less complex data. While this may be appropriate where a lower level of explanatory power is all that is required, this is unlikely to be adequate to reliably inform site-level management decisions. Furthermore, the highly significant relationship between visits and the socio-economic characteristics of the population in model II and a measure of access to substitute woodlands in model III would suggest that they are too important to ignore.

Essentially, the research described here represents a good example of what could be achieved if a number of data limitations were addressed. These limitations, which relate to both the measurement of forest recreation visits and the acquisition of reliable data to predict such visits, are discussed below.

Although developments in GIS technology make the generation of spatially explicit input data increasingly straightforward, there are some high-quality datasets available, such as the socio-economic data on population in the national census data, in which significant improvement is unlikely. Nonetheless, as noted by Lovett et al. (1997), many of the generic and specific issues related to the use of GIS in the analysis of recreation demand have yet to be adequately addressed, particularly in relation to the estimation of reliable explanatory variables such as accessibility indices for substitute recreation sites. In addition, data availability in some areas remains a barrier to progress, for example, the lack of spatially explicit data on tourism populations. This may be one reason why many earlier studies have failed to consider the tourism population in their analysis. The results of the survey of forest visitors associated with this study found that tourists accounted for a significant proportion of visitors across all sites and that the proportion of tourists to residents varies considerably between sites. Further consideration reveals a relatively distinct bimodal pattern in the visit data, with visits to the majority of sites being dominated by either tourists or local residents. Thus, the absence of tourism population data undoubtedly accounts for some of the unexplained variation in the visit data. Perhaps more importantly, if there are grounds for suspecting that visit decisions made by tourists are influenced by a significantly different set of factors to those that influence visit decisions of local residents, then this pattern in visits may suggest the need for future TGF analysis to distinguish between woodlands on the basis of their primary user group, i.e. those servicing local residents, and those with a higher profile primarily servicing tourists.

Ultimately, from this analysis it is not possible to know whether the unexplained variation in visit data is due to ‘missing’ or inaccurate explanatory
variables or due to inaccuracies in the original visitor counts. Essentially, where the method, timing and timeframe of visit count data collection vary, potential errors in the data are introduced. The likelihood of error is even greater where visitor numbers to forests are based on extrapolated estimates rather than counts, whilst the lack of a common definition of what constitutes a recreation site introduces further ambiguity into the modelling process.

Currently, reliable visit data do not exist for a sufficiently large and representative range of publicly accessible woodland to develop models that will be universally transferable. Systematic collection of data across a representative range of sites would require considerable investment and co-ordination between woodland site owners/managers. In theory, overcoming the limitations described here and developing a reliable tool with which to predict visits to unsurveyed sites could be of great benefit to forest management decision making, especially if the role of woodlands as providers of amenity services continues to increase in importance. However, any commitment of future resources by those involved in forest management to rectify the identified shortfalls should take into account the decision-making benefits resulting from improved TGFs in relation to the inherent ‘noise’ that is always likely to affect recreation demand modelling, where many facets of visitor behaviour and local contextual factors will inevitably remain unaccounted for, however reliable the data inputs.

Conclusions

Increasing demand for countryside recreation is a primary driver of the trend towards multi-purpose forestry in the UK. In the next few decades the landscape of the UK is set to undergo significant change, with large areas of forest due for harvesting, whilst shifts in land use policy, such as the reforms of the Common Agricultural Policy, provide opportunities to redirect resources away from forests managed for timber towards forests managed for a wider range of benefits including recreation and amenity.

Together, these provide a strong argument for the development of tools to assist in the planning of forests for recreation in the countryside. However, robust and statistically defensible models for widespread application have yet to be developed. Previous studies have generally been constrained by limited visitor data and inadequate treatment of spatially explicit explanatory factors. This study has built on previous studies in this important area by including visit data from 100 sites around GB. The developed models illustrate how recreational trips to woodland destinations can be explained using variables describing the recreational catchments around the destination, its population and the attributes of the destination.

The study represents good illustration of what could be achieved if data limitations, in particular those relating to the quality of visit data itself, the definition of what constitutes a forest site and the lack of spatially explicit data on tourism populations, could be solved. In the same way, whilst transfer of TGFs to specific forest sites to inform site-level micro-management decisions cannot yet be statistically defended, the results of this study suggest that the more reliable models presented here are sufficiently robust to be used for the purpose of aggregate-level analysis, for example, to improve understanding of the important role that forest recreation plays in the rural economy. This is subject to the sites used in the model being representative of other sites, a strong assumption in itself. As with other recent studies in this area of research (Lovett et al., 1997; Jones et al., 2002) use of GIS analytical techniques has facilitated continual improvement in the calculation of spatially explicit explanatory factors. This can only be encouraging in the search to develop tools to forecast user demand for woodlands specific in both type and facilities. However, it is likely to be the quality of visit data that is the ultimate constraint in this respect. In order to achieve models of sufficient reliability for site-level decisions, there is a clear need for private land owners to work alongside public organizations like FE to improve the quality of data relating to the number of visits to publicly accessible woodlands in GB.

Acknowledgements

The work was funded by the Economics and Statistics Unit, Policy and Practice Division of the Forestry
Commission, and supported by the Scottish Executive Environment and Rural Affairs Department (SEERAD). The authors would like to acknowledge and thank John DeGroote and Peter Shannon of the Macaulay Institute for their assistance with the GIS analysis, Jackie Potts from BioSS for statistical support and Nick Hanley of the University of Glasgow for his advice and comments on the modelling. The authors would also like to thank the staff of the Forestry Commission, RSPB and the WTS for providing visit data.

References
Jones, A., Bateman, I. and Wright, J. 2002 Estimating arrival numbers and values for informal recreational use of British woodlands. Final Report to the Forestry Commission, CERGSC School of Environmental Sciences, University of East Anglia.
Loomis, J.B. 1995 Four models for determining environmental quality effects on recreational demand and regional economics. Ecol. Econ. 12, 55–65.
Office for National Statistics. 1991b Census: small area statistics and local base. ESRC/JISC Census Programme, Census Dissemination Unit, University of Manchester.

Received 27 September 2004