Evaluation of the recreational services from Danish Forests combining discrete choice models and GIS

Corresponding author: Dr Mette Termansen, Sustainability Research Institute, School of Earth and Environment, University of Leeds, Leeds LS2 9JT, UK, Email: m.termansen@leeds.ac.uk; tel.no. +44 113 343 6411; fax.no. +44 113 343 6716.

Dr Colin J. McClean, Environment Department, University of York, Heslington, York YO10 5DD, UK

Dr Riccardo Scarpa, Waikato Management School, The University of Waikato, Private Bag 3105, Hamilton New Zealand

Dr Frank Søndergaard Jensen, Danish Centre for Forest, Landscape and Planning, University of Copenhagen, Rolighedsvej 23, DK-1958 Frederiksberg C, Denmark

ABSTRACT

This paper combines recreational choice modelling and economic valuation with GIS based techniques to allow an assessment of the spatial diversity of forest recreation. The paper reports on a national study of the recreational use of Danish forest sites. A random utility framework is used and a comparison between the standard fixed coefficient model and a mixed logit model is made. The results show that the different specifications of the random utility models reveal similar preferences for the measured forest attributes in terms of sign and magnitude. The spatial predictions, however, reveal a considerable difference in the variation of economic benefits from recreation between the two models.

Key words: Mixed Logit, Forest Recreation, GIS.

1. Introduction

The use of discrete choice models for studying recreational activities and assessing the economic values of these activities is well established in environmental economics (Bockstael et al., 1987; Morey et al., 1993). Recent advances in modelling techniques include the development of simulated choice probabilities and in particular mixed logit models (Revelt and Train, 1998; Train, 1998). Using such approaches environmental economists have begun to investigate the effect of heterogeneous preferences on benefit estimations of outdoor recreation (Breffle and Morey, 2000; Chen and Cosslett, 1998; Provencher and Barenklau, 2002; Train, 1998). Furthermore, the development of mixed logit models has facilitated an implementation of more general patterns of substitution between sites (Herriges and Phaneuf, 2002), which have addressed one of the main criticisms of the conditional logit model, namely its restrictive substitution assumptions.

However, heterogeneity is not only present in people's preferences, but is also present in the spatial distribution of recreational opportunities. Most applications of standard conditional logit models aggregate the recreational sites to such an extent, that a large part of the spatial heterogeneity, represented by the variability of the attributes of the sites in the choice set, is lost from the data. The use of Geographical Information Systems (GIS) to process spatial data should allow a much larger proportion of the heterogeneity of spatial elements to be captured. In addition to site attributes, the spatial elements include population densities, road networks and broader landscape characteristics. Recently, GIS have been implemented in travel costs studies (Bateman et al., 1996; Bateman et al., 2002; Brainard et al., 1999), but despite the success of these studies environmental economists have not yet fully exploited the capabilities of GIS in capturing behavioural responses to spatial heterogeneity in recreation studies.

In this paper we aim to capture both of these sources of heterogeneity by using mixed logit models with a high resolution spatial dataset within a GIS environment. We apply these techniques to model forest recreation patterns in Denmark using a national data set collected by the Danish Forest and Landscape Research Institute. This gives information on 28,947 recreational trips with precise locations of trip origin and destination. Landscape indices derived from the GIS are used to characterise the individual destinations. The spatial distribution of the population, socio economic characteristics and the road network enables simulation of the spatial patterns of recreational values and the spatial distribution of values associated with alternative policy initiatives. The estimated recreational choice models incorporate taste variation and error components in the choice probabilities by using a mixed logit model. To evaluate the mixed logit model's ability to capture the spatial variability in the data it is bench-marked against the more conventional conditional logit approach, which does not allow for taste variation and has a restrictive substitution pattern.

The choice probability models specify the choice destinations, they do not however address total demand. To estimate total demand we use a household data set that gives the information to model demand separately by using a count data approach as previously done by others (Hutchinson et al., 2003; Romano et al., 2000) who followed a number of seminal papers (Feather et al., 1995; Hausman et al., 1995; Parsons and Kealy, 1995) add Bpckstael**. The Danish Forest and Landscape Research Institute also collected this dataset as part of the national study of outdoor recreational of the Danish public.

Using spatial demographic and socio-economic national data the predictive differences between the mixed logit (MXL) and the conditional logit (CL) models are evaluated for all woodlands in Denmark above 10 hectares. We assess the predictive values associated with three policy initiatives: I) increasing the area of all existing forest areas by 10 hectares, II) increasing the proportion of the broadleaved areas by 5% and III) altering land uses adjacent to forested areas increasing the boundaries with semi-natural vegetation by 5%. The expectation is that the mixed logit model will capture much more variation in the range of predictive values associated with the policy under evaluation than the fixed taste coefficient logit model can. Coupling the results of the model estimations and high resolution national datasets within GIS enables a more accurate evaluation of the spatial diversity in recreational values. We illustrate this by mapping the spatial pattern of economic values associated with individual forest sites and the pattern of welfare change resulting from the above mentioned policies by changing site attributes at individual sites. We also consider the spatial distribution of welfare changes accruing to the place of origin of individuals. This analysis generates spatial patterns of economic values highlighting the residential locations of individuals benefiting the most/least from implementation of individual policies.

The nature of the data is particularly important for the work presented in this paper. We therefore choose firstly to describe the data source in some detail before we introduce the modelling methodology. This is followed by the results of the analysis and discussion of the findings.

2. Data

Data from two separate surveys are used in this study. An on-site questionnaire survey is used to estimate the Random Utility Models while a household survey is used to estimate the total demand for car borne recreational trips. The evaluation of policies at the national level further requires data on the spatial distribution of the population and their socioeconomic characteristics.

2.1 The on-site survey data

The on-site questionnaire survey of car borne recreation covering the most important recreational areas in Denmark were used for this study. The sampled area accounts for approximately 75% of the annual recreational trips (Jensen, 2003). Data were collected simultaneously at all sites, 22 times during the year-long period of the study (December 1996 to November 1997) by the Danish Forest & Landscape Research Institute. The sampling is therefore choice-based, however contrary to most choice-based surveys, the survey has been designed so that sampling effort on each site is identical. Questionnaires were distributed to all cars visiting the forest areas during one ½ hour period. The identical sampling effort at all sites implies that the population proportions visiting individual sites can be assumed identical to the sample proportions. Questionnaires were distributed at 2,095 locations. A response rate of approximately 50% resulted in the return of around 40,000 questionnaires. Questionnaires where the recreational trips could not be precisely identified because of incomplete trip origin data were excluded and trips within, to or from the island of Bornholm, situated approx. 200 km from mainland Denmark, were also excluded. This left a sample of 28,947 questionnaires for the analysis. The origins of the trips were digitised through postal addresses using a software developed to assign postal addresses to the nearest node in the Danish road network (Carl Bro, 1997). The 2,095 survey locations were grouped into 581 continuous recreational areas identified by local forest managers to be separate recreational sites. The most centrally located questionnaire distribution point was chosen as the representative location of each recreational site. Travelling distances from the trip origin to the trip destination and to alternative sites were calculated using a 1:200,000 scale vector road map (VejnetDK) (Kort-og Matrikelstyrelsen/Vejdirektoratet, 1995). It is assumed that people have used the shortest route possible.

Potentially important site attributes have been derived from several data sources and handled within the GIS. The area information system, AIS (Miljø & Energiministeriet and Danmarks miljøundersøgelse, 2000), contains information on the spatial distribution of forests, semi natural vegetation such as heather moorland, meadows, rough pasture, lakes, bogs and rivers and different categories of built up areas. This area information system has been used to derive site attributes such as the size of the contiguous forested area associated with each site, distance to water features, adjacency of the forested area to other areas of semi natural vegetation. The forest area was further classified into broadleaved forest, coniferous forest and scrub vegetation, based on LandSat TM satellite images (Miljø og Energiministeriet and Danmarks miljøundersøgelse, 2000). This data source enabled the computation of the fraction of each forest type for each site. Location of marked nature trails and viewpoints were obtained from “the interactive map of Denmark” (Kort-og Matrikelstyrelsen, 2001). This enabled derivation of the distance from location of the site to the origin of marked nature trails and viewpoints. Classification of the terrain was based on a 50 x 50 metre digital elevation model. The mean slope calculated for a 1 km grid was used as a measure of the terrain (Skov-Petersen, 2002). Information on parking facilities was available from the survey. The site attributes used in the final model specification, how they have been measured and their sources are given in Table 1.

2.2 The household survey data

A household data set available for the study was collected in 1993/1994 by the Danish forest and Landscape Research Institute (Jensen and Koch, 1997). 2,895 people were randomly selected from the Danish population in the age group between 15 and 76 years, from the national register. The questionnaires were posted at 12 individual times during one year, 243 questionnaires at each selected time. This resulted in 2,424 completed questionnaires, a response rate of 83.7% (Jensen and Koch, 1997). However, the usable sample was reduced as we only included respondents we were able to geo-reference, not all questionnaires were complete and individuals from the island of Bornholm were excluded, leaving 1,494 questionnaires for the analysis (Table 2).

2.3 Demographic and socio-economic data

The demographic data originates from a digital parish map of 2,116 parishes with information on male and female population divided into 6 age classes. Finer resolution data, allocating population segments to nodes in the road network, were available from the Danish Forest and Landscape Research Institute. The spatial disaggregation of the population was generated using an urban land use map (100x100m resolution) (Skov-Petersen, 2002). Information on average income was available from Danish Statistic at the parish level. The data are explained in Table 3.

3. Methodology

The trip allocation models are specified as Random Utility Models, estimated using different assumptions about heterogeneity of taste and patterns of substitution. We use a MXL specification to allow for continuous taste heterogeneity and to introduce correlation among utilities from substitute sites. The standard fixed coefficients CL specification is reported as a benchmark. The total number of trips per individual is specified as a count process, to account for the integer nature of the dependent variable (Hellerstein and Mendelsohn, 1993).

3.1 The Random Utility Models

We choose a discrete choice approach (Ben-Akiva and Lerman, 1985; Haab and McConnell, 2002; Train, 2003) to model choices between recreational sites for a single choice occasion. This assumes that an individual chooses the site *j* from the set of available sites *J* with the highest expected utility from the visit. A general representative utility function of visit to site *j* of individual *n*, v_{nj}, is specified. The utility function includes taste variability in site characteristics x_{nj}, and an unobserved component, u_{nj}, which account for flexible substitution patterns. The utility that individual *n* gains from visiting site *j* is therefore specified as

$$U_{nj} = v_{nj} + u_{nj} \quad [1]$$

where $v_{nj} = \boldsymbol{\beta}'\mathbf{x}_{nj}$ and $\boldsymbol{\beta}$ is a taste parameter vector specified according to a taste distribution function with parameter vector $\boldsymbol{\theta}$, $f(\boldsymbol{\beta}|\boldsymbol{\theta})$. A variety of distributions have been used in the literature (Hensher and Greene, 2003). Usually, $\boldsymbol{\theta}$ includes a vector of means and elements of a variance-covariance matrix.

The generalisation of the substitution patterns is achieved in an equivalent way by introducing error components (Herriges and Phaneuf, 2002; Train, 2003). The error components, z_{nj}, defines clusters of sites which are believed to have a higher degree of substitutability, z_{nj} equals 1 if site *j* belongs to a cluster and z_{nj} equals 0 if it does not. The unobserved part of utility, u_{nj}, is therefore specified as,

$$u_{nj} = \boldsymbol{\eta}'\mathbf{z}_{nj} + \varepsilon_{nj} \quad [2]$$

where $\boldsymbol{\eta}$ is a vector of random normal terms with zero mean and a vector of standard deviations $\boldsymbol{\sigma}_{ec}$, and ε_{nj} is iid extreme value (Train, 2003).

The probability of individual *n* choosing site *i* can be derived given the parameter vector $\boldsymbol{\theta}$ and the parameter vector $\boldsymbol{\sigma}_{ec}$ by integrating over the domain of the parameter distribution for $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ as follows:

$$\Pr_{ni} = \int_{\beta}\int_{\eta}\left(\frac{\exp(\boldsymbol{\beta}'\mathbf{x}_{ni}+\boldsymbol{\eta}'\mathbf{z}_{ni})}{\sum_{j\in J}\exp(\boldsymbol{\beta}'\mathbf{x}_{nj}+\boldsymbol{\eta}'\mathbf{z}_{nj})}\right)\cdot g(\boldsymbol{\eta}\,|\,\boldsymbol{\sigma}_{ec})\,f(\boldsymbol{\beta}\,|\,\boldsymbol{\theta})d\boldsymbol{\eta}d\boldsymbol{\beta} \qquad [3]$$

This makes the MXL probability a weighted average of the logit formula evaluated at different values of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$ with the weights given by the density $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ and $g(\boldsymbol{\eta}|\boldsymbol{\sigma}_{ec})$.

Restricting the model to a fixed parameter model, the probability specification reduces to:

$$\Pr_{ni} = \frac{\exp(\bar{\boldsymbol{\beta}}'\mathbf{x}_{ni})}{\sum_{j\in J}\exp(\bar{\boldsymbol{\beta}}'\mathbf{x}_{nj})} \qquad [4]$$

where $\bar{\boldsymbol{\beta}}$ is a fixed parameter vector.

The MXL specification increases the degrees of freedom in the simulated maximization of the sample likelihood, and hence nearly always significantly increases the statistical fit of the model in comparison to the CL, which is a special case of the MXL when the spread is constrained to zero. For further information on the MXL readers are referred to (Train, 2003).

3.2 Specification of the Random Utility Models

We estimate a CL model and a MXL model using the same attribute specification for the indirect utility function. The total choice set is specified as the sites within a travel distance of 250 km, as only 0.1 % of the sample is observed to travel beyond this distance. It has been proved that using a subset of the alternatives in model estimation will provide consistent model parameters in a fixed coefficient model (McFadden, 1978). This approach was shown to generate very similar relative parameter estimates to those obtained when the estimation choice set is the same as the total choice set in a lake recreation study (Parsons and Kealy, 1992). For the present forest recreation dataset a similar result was obtained (Termansen et al., 2004). However, their analysis using conditional logit models showed that in this data set a much larger choice set was needed to achieve stability in parameter estimates. A similar theoretical proof is not available for mixed logit models, however Nerella and Bhrat (2004) provide a guide for the number of alternatives needed in the sample to generate good numerical performance. They suggest using a minimum of ¼ of the total choice set and recommend using ½ when it is not possible to estimate the models using the full choice set (Nerella and Bhat, 2004). To limit computing time in estimation, we define an estimation choice set as a random subset containing 120 alternatives from the total choice set. In particular, for each individual the estimation choice set contains the observed individual choice and 119 randomly selected sites from the sites in the total choice set. Our choice of 120 alternatives is consistent with the available recommendation. We used 100 Halton draws in the simulation, as Halton draws have better equidispersion properties than random draws (Train, 2004). The presence of taste heterogeneity of the different site attributes is tested in turn by restricting the relative spread parameter to be zero. Although statistical evidence of heterogeneity is found at this large sample size for most taste parameters, we focus, in the final specification, on those attributes with the largest taste variability; percent of broadleaved woodland and travel cost (marginal utility of income). Substitution patterns are investigated with two error component specifications. The first based on an increased substitutability between sites, which are close to the coast, the second clustering the sites that are managed to a larger degree for recreation as opposed to timber production. In the final specification only the former is reported as evidence for a common correlation in the latter is not found.

The travel cost parameter is expected to have a negative sign, and it is therefore assumed to have a constrained distribution. We choose a log-normal distribution for the negative of travel cost, so that the taste intensities for travel cost, (β_c), are distributed log-normally, $\ln(\beta_c) \sim N(\mu_c, \sigma_c^2)$.

The distribution of taste intensity for broadleaves is not necessarily constrained to a given sign. Visitors may like or dislike the fraction of broadleaved trees present in a given woodland.

Therefore, we assume that the distribution of this variable is normal, $\beta_b \sim N(\mu_b, \sigma_b^2)$. All estimations were conducted with the GAUSS routine available from Prof. K. Train's webpage.

3.3 Specification of the model for total demand of car borne recreational trips

The total annual number of recreational trips to woodlands by the Danish population is specified using a zero inflated count model to account for the two stages in the decision process (Cameron and Trivedi, 1998). The first stage is the decision on participation in car borne recreational activities or using other modes of transport or not participating in forest recreation. 91% of the Danish population participated in forest recreation in the year preceding the survey (Jensen and Koch, 1997), however only 49% used the car as a mode of transport. We model the participation stage using a normal distribution. The probability of individual *n* not using the car as a mode of transport, q_n, is modelled using a normal cumulative distribution function, $\Phi(w_n)$, where w_n is a linear function, $w_n = \boldsymbol{\gamma}_1' \mathbf{s}_n$, of socio-economic characteristics of individual *n*, $\mathbf{s}_n$, and $\boldsymbol{\gamma}_1$ is a parameter vector. The second stage is the decision on the number of annual car borne trips given that the individual *n* travels by car. If the choice is to participate in the recreational activity the number of trips *t* is bigger than zero, $t>0$, otherwise $t = 0$. This stage of the process is specified as a negative binomial, as we find evidence of over-dispersion. The probabilities of an individual undertaking a given number of car borne recreational trips to woodlands are therefore specified as;

$$
\begin{aligned}
Pr[t_n = 0] &= q_n + (1-q_n) R_n(0) \\
Pr[t_n = r] &= (1-q_n) R_n(r), \qquad r \in \mathbf{Z}^+
\end{aligned}
\qquad [5]
$$

where $R_n(t)$ is the negative binomial probability of observing visits equal to *t*.

$$
R_n(t) = \Gamma(\theta+t) \big/ [t!\Gamma(\theta)] u_n^{\theta} [1-u_n]^t, \qquad t \in \mathbf{Z} \qquad [6]
$$

$\theta = 1/\alpha$, where α is the over dispersion parameter, $u_n = \theta \big/ [\theta + \lambda_n]$, and λ_n is specified as a linear exponential, $\exp(\boldsymbol{\gamma}_2 ' \mathbf{s}_n)$, in the socio-economic characteristics $\mathbf{s}_n$ given the parameter vector $\boldsymbol{\gamma}_2$.

3.4 Predictions of the spatial distribution of the annual number of visits

The predictions of the spatial distribution of the annual number of car borne visits are made by combining the model for total demand and the random utility models. The total demand for car born recreational trips is predicted for each node in the road network using the estimates from the zero inflated negative binomial model and the national data on socioeconomic variables. The probabilistic allocation of the total number of trips from each node in the road network to the individual sites in the choice set is predicted using the estimated random utility models. Choice sets vary between origins as they are defined as the accessible sites within a radius of 250 km through the road network. Repetition of this process over all origins results in an allocation of the predicted total number of trips to all sites.

We compare two predictions of the spatial distributions of the annual number of trips; the first based on the CL model with fixed parameters, the second based on the MXL model. For the MXL model the probability distribution is evaluated using draws from the density function $f(\boldsymbol{\beta}|\boldsymbol{\theta})$ and $g(\boldsymbol{\eta} \mid \boldsymbol{\sigma}_e)$ given the the estimates of $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}_e$. The number of draws is in proportion to the number of people resident at each origin, to reduce computational time each draw is taken to represent 100 residents.

Welfare calculations

The welfare calculations used in random utility modelling are based on the indirect utility function. The researcher's estimation of the individuals' indirect utility function given that the recreationists choose the site which maximise utility is given by;

$$
\boldsymbol{E}\{max(\boldsymbol{u})\} = \ln\Big(\sum_{j \in J} exp(\boldsymbol{v}_j)\Big) \qquad [7]
$$

where $v_{nj} = v(y\text{-}c_{nj}, \mathbf{q}_j)$, y is income, c_{nj} is the cost for individual n visiting site j and $\mathbf{q}$ is a vector of site attributes. WTP_n is calculated as the willingness to pay to achieve conditions $\boldsymbol{c}^*$, $\mathbf{q}^*$ when the current conditions are $\boldsymbol{c}$, $\mathbf{q}$ (Hanemann, 1999; McConnell, 1985).

$$WTP_n = \int_{\beta}\int_{\eta} \frac{1}{\beta_C}\left[\ln\left(\sum_{j\in J} exp(\boldsymbol{v}^*)\right) - \ln\left(\sum_{j\in J} exp(\boldsymbol{v})\right)\right] g(\boldsymbol{\eta}) f(\boldsymbol{\beta}) d\boldsymbol{\eta} d\boldsymbol{\beta} \qquad [8]$$

In the prediction of WTP_n at the population level utilities were evaluated by randomly drawing from the estimated distribution of $\boldsymbol{\beta}$ and $\boldsymbol{\eta}$.

3.5 Prediction of the spatial distribution of the value of access

We estimate the loss in welfare from loss of access to sites by simulating the loss in expected value of the maximum utility function $E\{\max(u)\}$ by letting the costs go to infinity for the site which is being assessed.

The welfare loss for site j, WTP_j, is the sum of the welfare losses over all origins, O, for which site j is included in the choice set, $O(j)$. The aggregation over origins takes into account whether a site j is accessible from an origin. $J(o)$ is the set of sites which are within 250 km travel distance through the road network from the origin o. $J^*(o)$ is the set of sites which are within 250 km travel distance through the road network from the origin o excluding the site which is being assessed.

$$WTP_j = \sum_{o\in O(j)} \left\{ -\frac{1}{\beta_c}\left[\quad - \quad \right] \right\} \qquad [9]$$

where β_c is the travel cost coefficient and IV and IV^* are the inclusive values.

$$= \ln\left\{ \sum_{j\in J(o)} exp(\boldsymbol{v}_j) \right\}; \qquad = \ln\left\{ \sum_{j\in J^*(o)} exp(\boldsymbol{v}_j) \right\} \qquad [10]$$

3.6 Prediction of the spatial distribution of the value of quality changes

The estimated welfare change associated with changes in the attributes of the sites is evaluated from two different perspectives. Firstly, it is evaluated separately for each individual site. The general formula [8] is used for this evaluation, but in this case the inclusive value (IV^*) is the expected value of the maximum utility given a change in one of the site attributes, q_j, at site j. Each site is evaluated sequentially assuming that site characteristics at other sites remain unchanged. Secondly, welfare changes are evaluated from the perspective of individual origins. Here we assess the spatial pattern of economic value associated with a policy initiative implemented at all sites. We measure the welfare change accruing to individual origins represented by the nodes in the road network. The policy relevant changes in terms of site attributes that we consider are: I) An increase in the size of the individual forested area by 10 hectares; II) an increase in the fraction of the area being broadleaved at individual sites by 5%; III) an increase in the fraction of the adjacent land classified as semi-natural land cover by 5%.

4. Results

4.1 Estimates for the random utility models and the count regression

The parameter estimates and standard errors for the CL and the MXL model for the final model specification are given in Table 4. The estimates of the two models are similar, both with respect to sign and magnitude. An important site attribute is parking facility, which increases the likelihood of the site being selected. Larger forest areas have a higher probability of selection, however this attribute has a declining marginal effect. Sites close to the coast are found to be more popular than inland sites as the coefficient on the distance from coasts is negative. The error component for coastal sites is significant demonstrating a common substitutability between sites within this category and a difference in the substitutability with other forests. Distance to marked nature trails were found to decrease the selection probability, however the analysis did not find that areas with marked nature trails had a higher degree of substitutability between them, as this error component was not significant. Recreationists were found to prefer sites, which were adjacent to other semi natural areas and undulating topography was also preferred to topographically flat areas. On average broadleaved areas are preferred to coniferous woodland, however a large variation of intensity exists for this taste parameter. The results suggest that 40 % of the population dislike broadleaves.

The parameter estimates for the ZIP model are given in Table 5. These show that the proportion of people choosing to travel by car decreases as the distances to the nearest forest increases and, on average, people from higher income parishes are more likely to travel on foot. The amount of car borne trips taken in a year decreases with distance to the nearest forest and increases with income, in accordance with expectations.

4.2 Value of site access

The average values for site access for the two model specifications are similar (Table 6, 1st row). The minimum and maximum values to individual sites, however, vary considerably (Table 6, 2nd and 3rd row) and disclose how the various sources of heterogeneity (spatial and individual) captured by the MXL model implies a much larger variation in values of access to forest sites than the CL model.

Maps of the spatial variation in values in the Eastern part of Demark are shown in Figure 1a (MXL) and Figure 1b (CL). Comparison of the two figures reveals the larger diversity in values resulting from the predictions based on the mixed logit model. Predicting with a fixed coefficients CL model, nearly all sites fall into the intermediate category of *WTP* which is between 8000 and 800,000 €/site/year. When the random effects MXL model is applied more sites have lower values (between 0 and 8000 €/site/year) and a few sites generate very high *WTP* values (between 800,000 – 2,000,000 €/site/year).

4.3 Values for improved site quality

Deriving the welfare changes associated with changes in the attributes assesses the importance of individual site attributes. The welfare implications of the management policies we explored are given in Table 7. Again, the values derived from the MXL specification display a much larger variability between individual sites. The mean values are, however, quite similar.

We illustrate the welfare implication of policy initiative I for the area north of Copenhagen in Figure 2. The spatial variation of predicted changes of values per recreational site, using the MXL specification, due to an increase of the area by 10 hectares is shown in Figure 2a. The variability of predicted values associated with the increase in the size of the forest area is partly due to the log specification of this variable. The analysis indicates that there seems to be a marginally decreasing value of increasing forest area, as a natural logarithm transformation was found to be statistically superior to a linear specification during the model specification search. This effect partly explains the spatial pattern of values. Increasing a small woodland by a given area is more valuable to recreationists than when the same increase is made in a larger woodland.

The spatial variation in the values per individual (at the place of residence) from the implementation of policy I uniformly at all sites is illustrated in Figure 2b. There are two effects that generate this spatial pattern. One effect it that origins in areas where there are either few forest or mainly small ones benefit more from the implementation of this policy, because of the marginally decreasing value of increasing forest area. The other effect is that origins from which individuals are predicted to undertake recreational trips more frequently benefit more from improvements in forest attributes. As the number of recreational trips is related to income, the spatial pattern of *WTP* per individual for improvements in recreational facilities is also determined by the spatial distribution of income. The type of detailed results illustrated in Figure 1 and 2 would not be attained without the use of GIS.

5. Discussion

In this paper we have combined MXL models of destination site selection and count models of total trip demand with detailed spatial information from GIS in order to assess the implications of allowing for taste heterogeneity and flexible substitution patterns. For the purpose of illustrating the potential of such an approach we focused on the prediction of the spatial distribution of forest values consequent to a set of management policies at the national level. This has given us the opportunity to model recreational behaviour and derive policy relevant predictions at a geographical scale that is appropriate for forest recreation behaviour of the general public and therefore for evaluating forest policy schemes aimed at increasing the welfare derived from forest recreation.

We find that, on average, the Danish population value access to forested sites at approximately 3.3 €/visit. This is 5 times higher than the previous Danish study based on a zonal travel cost model, but lower than most other European studies (Zandersen and Toll, 2003). We speculate that this might be due to the fact that significant population clusters are situated in regions with a high percentage of forest cover. The predicted recreational values per site vary between about 2 thousand and 1.8 million Euros per year. On a per hectare basis for forest sites utilised for timber production (larger than 100 hectare), the predicted recreational values vary between less than 20 to more than 5,000 Euros per hectare per year. According to statistics from the Danish Forest and Nature Agency, the average annual timber values are 45 Euros per hectare (Miljø og Energiministeriet, 1999). This illustrates that for some sites the recreational values contributes significantly to the overall generation of economic benefits.

The results of our analysis show that forest recreation in Denmark is dominated by frequent short distance trips. We predict that an adult member of the Danish population undertakes an average of 13 car borne recreational trips per year, and a large proportion of these is shorter than 10 kilometres. Modelling such an activity in a satisfactory way is demanding in terms of spatial accuracy. The precise origin and destination locations used in this survey have therefore been essential for this research. Furthermore, a cruder method of spatial aggregation of alternative destinations would clearly be unsatisfactory when modelling recreational activities characterised by short distance trips. The need for spatial accuracy has made the on-site survey essential for modelling the recreational choices. As an on-site sample will include frequent visitors in a larger proportion than a sample of the general public, we have estimated total demand based on a household survey. Our predictions implicitly assume that the choices people make between sites when they have decided to use the car as a mode of transport is independent of their total demand. If this is not the case the estimates based on the on-site sample will give biased results. The median travel distance for car users in the household sample is 10 km, and the average is 14 km. The median travel distance in the on-site sample is equally 10 km and the average is 17 km. This gives us some confidence that the bias due to on-site sampling does not invalidate our results and the use of the on-site sample is justified in order to achieve increased spatial accuracy.

The high resolution spatial data sets have also enabled us to identify many policy-relevant site attributes, one of which is the size of the forested area. Both linear and natural log transformation are tested. The analysis shows that a log specification is superior. This implies that there is a marginally declining improvement in visitor attraction with size. This effect is also found in an Irish study on willingness to pay for creation of nature reserves in Ireland (Scarpa et al., 2000). This result has implications for the allocation of resources to reafforestation initiatives as it gives an indication of the trade-offs between size of existing forests, size of the reafforestation project and proximity to populated areas.

The use of GIS derived data also enhances our results by allowing us to account for the spatial composition of the landscape. For example, it allows us to highlight that the presence of other semi-natural areas adjacent to the forested area affects recreational choices. We find that the fraction of the forest border adjacent to other semi-natural areas is positively associated with visit probability. Recreationists, therefore, seem to prefer visiting a site with a mix of forest and other semi-natural environments. Furthermore, proximity to the coast is identified as being preferred by visitors, as has been identified in other studies (Skov-Petersen, 2002). Our analysis also shows that coastal woodlands are perceived differently from a recreational viewpoint than inland woodlands, leading to distinct substitution patterns.

This study also identifies site attributes, which have direct implications for forest management. The species composition of forests seems to have an impact on recreational choices, as broadleaved forest seems to be preferred to conifer forest. This is to some extend also supported by empirical evidence from other Danish (Jensen and Koch, 1997; Skov-Petersen, 2002), Irish (Scarpa et al., 2000) and Canadian (Boxall et al., 1996) studies. Marked nature trails and parking facilities are other management-dependent site attributes. This analysis suggests that both seem to have a positive effect on probability of visit. It should be noted that the derived values associated with information leaflets are likely to be overestimated as these are more likely to have been produced for popular areas than for areas with few visitors.

An additional aim of the paper was to assess the implications of allowing for taste heterogeneity. The analysis shows very clearly that accounting for taste heterogeneity captures additional diversity in forest values, thereby impacting on the predicted values of forest policy initiatives. The diversity in the travel cost parameter is particularly important in valuation studies. Using a stochastic specification assumes that the researcher cannot observe an association between socio-economic variables and the variation in the marginal utility of income. It may have been expected that the variation in the travel cost parameter could be associated with income as argued by others (Herriges and Kling, 1999; Shonkwiler and Shaw, 2003). In this study we tested for such association but did not find support for a link between socio-economic status and the travel cost coefficient. This may be due to the use of aggregated data on income, but given the data, the treatment of variation in willingness to travel as unobserved heterogeneity seems a suitable approach. There is clear evidence that fixed parameter linear travel cost models do not represent the data used in this study well, as a non-linear transformation of distance markedly improves the statistical fit (Termansen et al., 2004). Including taste heterogeneity in willingness to travel allows us to account for the observed non-linearity in the data but maintains the assumption that each individual has a constant marginal utility of income. This assumption seems reasonable given that the expenditure on recreational activities will only constitute a minor proportion of the household budget (Haab and McConnell, 2002). Furthermore, the linearity assumption greatly eases the calculation of welfare effects, which for our application was an important consideration given the computational costs.

Our research therefore does not show an income effect on destination choice at the individual level. Individuals from higher income parishes do not seem to be willing to travel further than individuals from low income parishes. We do, however, find income effects on the total demand for recreational trips. Our results show that for higher income parishes the probability of travelling on foot is higher than for lower income parishes. Furthermore, the number of annual recreational trips is higher in parishes with high income.

An additional policy relevant conclusion, which was obtained from adopting the MXL approach, is that even though on average the results show a positive preference for broadleaved species, 40% of the individuals seem not to like broadleaves. Afforestation of agricultural land with broadleaved species is approximately 2-3 times more costly than afforestation with coniferous species (Miljø og Energiministeriet, 2003). This therefore suggests that some caution in the adoption of management policies to increase the area of broadleaved purely for recreational benefits is warranted.

A clear result of the comparison of the CL and MXL specification is that the spatial implementation of the models highlights a larger spatial diversity in forest values associated with different forest policy scenarios. Our interpretation of this is that the MXL model allows sites with desirable attributes to attract visitors from further away. At each node in the road network a fraction of the recreationists, given by the log normal distribution of the travel cost parameter, will be assigned a small marginal utility of income, and will therefore be willing to travel further. However, this effect is not homogeneous across space. Firstly, the effect will be magnified by population clusters, which are not evenly distributed. Secondly, the effect will be determined by the relative location of population clusters and recreational sites of varying quality. This is particularly important for area with high population density and relatively few attractive sites close by. Allowing for taste heterogeneity in the travel cost parameter shifts a proportion of the visits to more attractive sites located further away. Nodes located close to attractive sites, however, will not shift demand towards less attractive sites located further away. There is therefore a spatial asymmetry in the effects of allowing for heterogeneity. This effect will be expected for areas around highly populated areas with few recreational opportunities close by. As an example, our application to the area around Copenhagen illustrates this difference between the predictions from the CL and MXL models.

6. Conclusion

The work reported here is the first to demonstrate the potential in combining GIS and MXL approaches for modelling the spatial variation in recreational behaviour and associated economic values. Combining variability of taste, spatial features and substitution patterns proves effective in capturing individuals' responses to detailed landscape scale attributes in their recreational choices. We join previous researchers whose work has addressed the need to combine GIS and valuation techniques in the belief that this approach is valuable for environmental policy assessment because it produces more detailed, realistic and spatially explicit predictions.

Acknowledgements
We would like to thank Dr Hans Skov-Petersen for giving us access to useful forest attribute data and Dr Stephane Hess for useful suggestions on estimation of mixed logit models. We are also grateful to Prof. Kenneth Train for making his GAUSS code available on the internet. Finally, we acknowledge the financial support from the Danish Socio Economic Research Council.

Reference List

Bateman,I.J., Garrod,G.J., Brainard,J.S., and Lovett,A.A., 1996. Measurement Issues in the Travel Cost Method: a Geographical Information Systems Approach. Journal of Agricultural Economics, 47:191-205 pp.

Bateman,I.J., Jones,A.P., Lovett,A.A., Lake,I.R., and Day,B.H., 2002. Applying Geographical Information Systems (GIS) to Environmental and Resource Economics. Environmental and Resource Economics, 22:219-269 pp.

Ben-Akiva,M. and Lerman,S.R., 1985. Discrete Choice Analysis: Theory and Applications to Travel Demand. The MIT Press, Cambridge.

Bockstael,N.E., Hanneman,W.M., and Kling,C.L., 1987. Estimating the value of water quality improvements ina recreational demand framework. Water Resources Research, 23:951-960 pp.

Boxall,P.C., Watson,D.O., and Englin,J., 1996. Backcountry recreationists' valuation of forest and park management features in wilderness parks of the western Canadian Shield. Canadian Journal of Forest Research, 26:982-990 pp.

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 Science, 13:227-246 pp.

Breffle,W. and Morey,E.R., 2000. Investigating preference heterogeneity in a repeated discrete choice recreation demand model of Atlantic salmon fishing. Marine Resource Economics, 15:1-20 pp.

Cameron,A.C. and Trivedi,P.K., 1998. Regression Analysis of Count Data. Cambridge University Press.

Carl Bro, 1997. Befordringsbidrag. Carl Bro Informatik, Copenhagen, 1-8 pp.

Chen,H.Z. and Cosslett,S.R., 1998. Environmental quality preference and benefit estimation in multinomial probit models: a simulation approach. American Journal of Agricultural Economics, 80:512-520 pp.

Feather,P., Hellerstein,D., and Tomasi,T., 1995. A discrete-count model of recreation demand. Journal of Environmental Economics and Management, 29:214-227 pp.

Haab,T.C. and McConnell,K.E., 2002. Valuing Environmental and Natural Resources. The econometrics of non-market valuation. Edward Elgar., Cheltenham.

Hanemann,W.M., 1999. Welfare analysis with discrete choice models. In: J.A.Herriges and C.L.Kling (Editors), Valuing recreation and the environment: Revealed preference methods in theory and practise. Edward Elgar, Cheltenham, UK and Northhamton, MA.

Hausman,J.A., Leonard,G.K., and McFadden,D., 1995. A Utility-consistent, combined discrete choice and count data model Assessing recreational use losses due to natural resource damage. Journal of Public Economics, 56:1-30 pp.

Hellerstein,D. and Mendelsohn,M., 1993. A Theoretical Foundation for Count Data Models, with an Application to a Travel Cost Model. American Journal of Agricultural Economics, 75:604-611 pp.

Hensher,D.A. and Greene,W.H., 2003. The Mixed Logit model: The state of practice. Transportation, 30:133-176 pp.

Herriges,J.A. and Kling,C.L., 1999. Nonlinear income effects in random utility models. Review of Economics and Statistics, 81:62-72 pp.

Herriges,J.A. and Phaneuf,D.J., 2002. Inducing patterns of correlation and substitution in repeated logit models of recreation demand. American Journal of Agricultural Economics, 84:1076-1090 pp.

Hutchinson,W.G., Scarpa,R., Chilton,S.M., and McCallion,T., 2003. Spatial Distribution Versus Efficiency Effects Of Forest Recreation Policies: Using a Regional Travel Cost Model. In: N.Hanley, D.Shaw, and R.Wright (Editors), The New Economics of Outdoor Recreation. Edward Elgar, pp. 139-164.

Jensen,F.S., 2003. Friluftsliv i 592 skove og andre natur omraader [Outdoor recreation in 592 forests and other natural areas]. Forskningscenteret for skov og landskab, Horsholm, Denmark.

Jensen,F.S. and Koch,N.E., 1997. Friluftsliv i skovene 1976/77 - 1993/94. The Forest and Landscape Research Institute, Denmark.

Kort-og Matrikelstyrelsen, 2001. Det levende Danmarkskort [The interactive map of Denmark]. Kort-og Matrikelstyrelsen, Copenhagen.

Kort-og Matrikelstyrelsen/Vejdirektoratet. VejnetDK. 1995. Denmark, Kort- og Matrikelstyrelsen.

McConnell,K.E., 1985. The economics of outdoor recreation. In: A.V.Kneese and J.L.Sweeney (Editors), Handbook of Natural Resource and Energy Economics. North Holland, Amsterdam, pp. 677-722.

McFadden,D., 1978. Modeling the choice of residential location. Transportation research records, 673:72-77 pp.

Miljø & Energiministeriet and Danmarks miljøundersøgelse, 2000. Areal Informations Systemet [Land cover map]. Miljø- & Energiministeriet/Danmarks miljøundersøgelse, Copenhagen.

Miljø og Energiministeriet, 1999. Virksomhedregnskab [Annual Business Report]. Skov og Naturstyrelsen [Forest and Nature Agency], Miljø og Energiministeriet [Ministry of Environment and Energy], 1-157 pp.

Miljø og Energiministeriet. Skovkulturen i et kulturnaert perspectiv. 1-41. 2003. The Danish Forest and Nature Agency, Ministry of Environment and Energy.

Miljø og Energiministeriet and Danmarks miljøundersøgelse, 2000. Areal Informations Systemet [Land cover map]. Miljø- og Energiministeriet/Danmarks miljøundersøgelse, Copenhagen.

Morey,E.R., Rowe,R.D., and Watson,D.O., 1993. A repeated nested logit model of atlantic salmon fisheries. American Journal of Agricultural Economics, 75:578-592 pp.

Nerella,S. and Bhat,C.R., 2004. Numerical Analysis of Effect of Sampling of Alternatives in Discrete Choice Models. Transportation Research Record, 1894:11-19 pp.

Parsons,G.R. and Kealy,M.J., 1995. A demand theory for number of trips in a random utility model of recreation. Journal of Environmental Economics and Management, 29:357-367 pp.

Parsons,G.R. and Kealy,M.J., 1992. Randomly Drawn Opportunity Sets in a Random Utility Model of Lake Recreation. Land Economics, 68:93-106 pp.

Provencher,B. and Barenklau,K.B.R.C., 2002. A finite mixture model of recreational angling with serially correlated random utility. American Journal of Agricultural Economics,1066-1075 pp.

Revelt,D. and Train,K., 1998. Mixed Logit with Repeated Choices: Households' Choices of Appliance Efficiency Level. Review of Economics and Statistics, LXXX:647-657 pp.

Romano,D., Scarpa,R., Spalatro,F., and Viganò,L., 2000. Modeling determinants of participation, number of trips and site choice for outdoor recreation in protected areas. Journal of Agricultural Economics, 51:224-238 pp.

Scarpa,R., Chilton,S.M., Hutchinson,W.G., and Boungiorno,J., 2000. Valuing the recreational benefits from creation of nature reserves in Irish forests. Ecological Economics, 33:237-250 pp.

Shonkwiler,J.S. and Shaw,W.D., 2003. A finite mixture approach to analyzing income effects in random utility models: reservoir recreation along the Columbia river. In: N.Hanley, W.D.Shaw, and R.E.Wright (Editors), The new economics of outdoor recreation. Edward Elgar, Cheltenham, pp. 268-279.

Skov-Petersen,H., 2002. GIS, Accessibility, and Physical Planning. Exemplified by Models of Recreational Activities. University of Copenhagen.

Termansen,M., McClean,C.J., and Skov-Petersen,H., 2004. Recreational site choice modeling using high spatial resolution data. Environment and Planning A, 36:1085-1099 pp.

Train,K., 1998. Recreation demand models with taste differences over people. Land Economics, 74:230-239 pp.

Train,K., 2003. Discrete choice methods with simulation. Cambridge University Press.

Train, K. Halton sequences for mixed logit. Working Paper No. E00-278. 2004. University of California, Berkeley, Department of Economics.

Zandersen, M and Toll, R. 2003. A meta analysis of forest recreation in Europe. Working paper. Research unit sustainability and global change, Hamburg University. Germany.

TABLE 1
Data used for specification of the random utility models.

Variable	Measurement	Data Source
Travel costs	Round trip costs using the shortest distance through road network from the origin of the trip given by the respondents to the sites. The travelled distance is measured to the visited site. The distance to the alternative sites are measured to the aggregated sites. Travel costs calculated using variable driving costs only.	On-site questionnaire Road network Variable travel costs
Forest area	Size of the nearest contiguous forest area.	AIS
Fraction of broadleaved area	Fraction of Forest area classified as broadleaved	AIS
Coast	Euclidian distance from aggregate site to nearest coast line. Assignment of Error Component: If Coast is less than 2 km, Coast_EC = 1, otherwise Coast_EC = 0.	AIS
Natural area edge fraction	Length of external and internal forest edge neighbouring area classified as semi-natural or wet area divided by total forest edge.	AIS
Parking	Dummy variable. If parking lots are available at one of the sites, the aggregated site is classified as Parking = 1, otherwise Parking = 0.	On-site questionnaire data
Info	Euclidian distance from aggregate site to nearest marked nature trail described in an information leaflet.	Interactive map of Denmark
View point	Dummy variable. If euclidian distance from aggregate site to nearest view point is less than 5 km, View point = 1, otherwise View point = 0.	Interactive map of Denmark

TABLE 2
Data used for estimating annual demand of recreational visits

Variable	Measurement	Data Source
Distance	Shortest Euclidian distance from each individuals home address to a forest patch larger than 10 hectares	AIS Road network
Income	Average parish council income	Danish Statistics
Trips	Total number of car borne trips per year	Household questionnaire

TABLE 3
Data used for predicting the spatial distribution of the total number of visits

Variable	Measurement	Data Source
Population	Total adult population allocated to nodes in the road network.	Danish Statistics
Distance	Shortest Euclidian distance from each node in the road network to a forest patch larger than 10 hectares	AIS
Income	Average parish council income	Danish Statistics

TABLE 4

Parameter estimates for fixed parameter and mixed logit specification.
Choice set size = 120, N = 28,947.

Variables	Fixed parameter model		Mixed logit model	
	Estimates	Asymptotic-z	Estimates	Asymptotic-z
Travel costs (mean)	-0.044	366.6	-2.607	258.4
Travel costs (standard div.)	n.a.	n.a.	1.070	87.1
Fraction broadleaved (mean)	0.451	14.0	0.314	6.8
Fraction broadleaved (standard div.)	n.a.	n.a.	1.218	9.8
Parking	0.407	20.1	0.548	23.6
Ln(Info)	-0.300	34.6	-0.440	39.6
View point	0.238	13.9	0.314	16.3
Ln(Coast)	-0.504	60.3	-0.628	49.0
Ln(Forest area)	0.328	76.2	0.388	58.1
Fraction natural area edge	2.515	60.3	2.892	56.4
Coast_ Error component	n.a.	n.a.	0.722	6.2
Mean log-likelihood	-2.44		-2.18	

TABLE 5
Parameter estimates for zero inflated count regression model of trip frequency. N = 1494.
Log-likelihood (Poisson) = -35482.5; (Zero Inflated Negative Binomial) = -4133.6

	Variable	Estimates	Asymptotic-z
Negative binomial	Constant	1.720	5.7
	Income	7.44×10^{-3}	5.8
	Distance	-1.08×10^{-4}	4.7
Dispersion parameter	Alpha	3.105	33.1
Zero inflation model	Constant	-0.296	2.1
	Income	3.88×10^{-3}	17.4
	Distance	-9.44×10^{-4}	5.5

TABLE 6
Value of site access

Economic values	Fixed parameter model	Mixed logit model
Mean value per trip [€/trip]	3.1	3.3
Maximum value for a site [€/site]	0.63×10^6	1.8×10^6
Minimum value for a site [€/site]	3,128	2,315

TABLE 7
Changes in economic values associated with changes in management on all forest sites. Values are in Euros/trip.

		Fixed			Mixed logit model		
Policy	Change in attribute	Mean	Min	Max	Mean	Min	Max
I: Forest area	Increase 10 ha	0.31	0.0001	0.62	0.42	0.001	2.05
II: Broadleaved	Increase 5 %	0.07	0.07	0.07	0.06	-0.09	0.11
III: Natural area edge	Increase 5 %	0.36	0.34	0.36	0.53	0.06	1.34

FIGURES

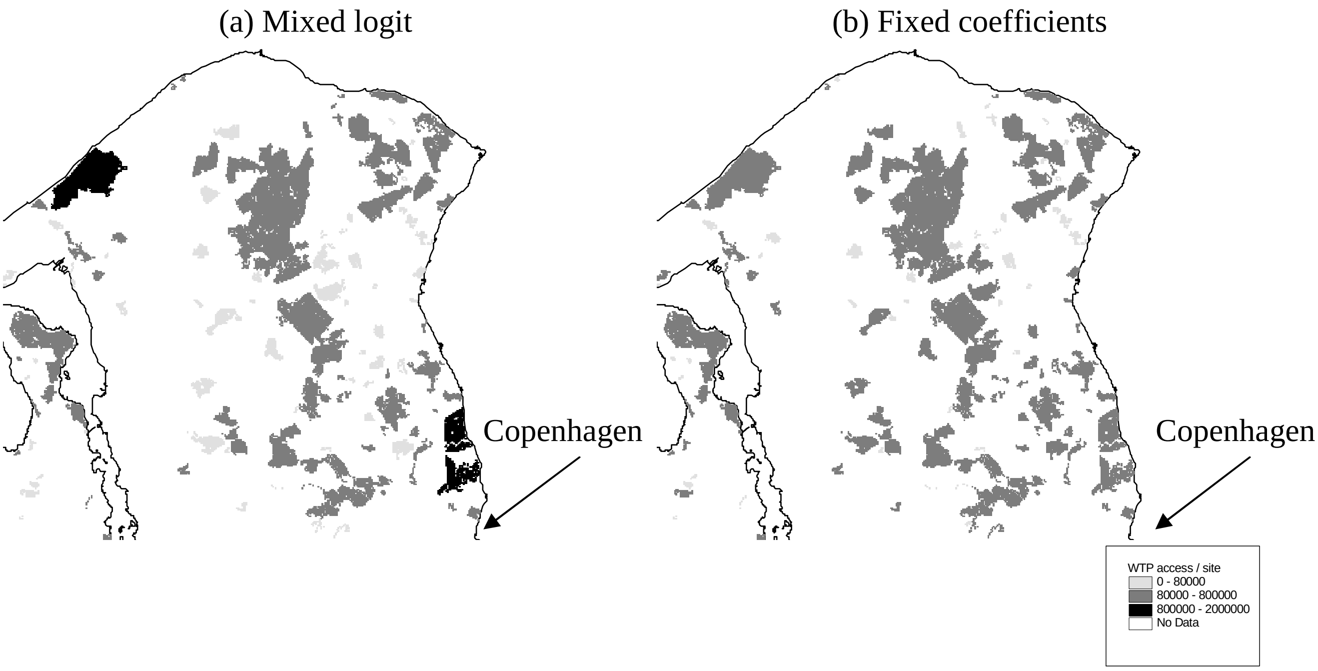

Figure 1: WTP for access to individual recreational sites [Euros/site/year]. (a) Mixed logit model; (b) Fixed coefficient model

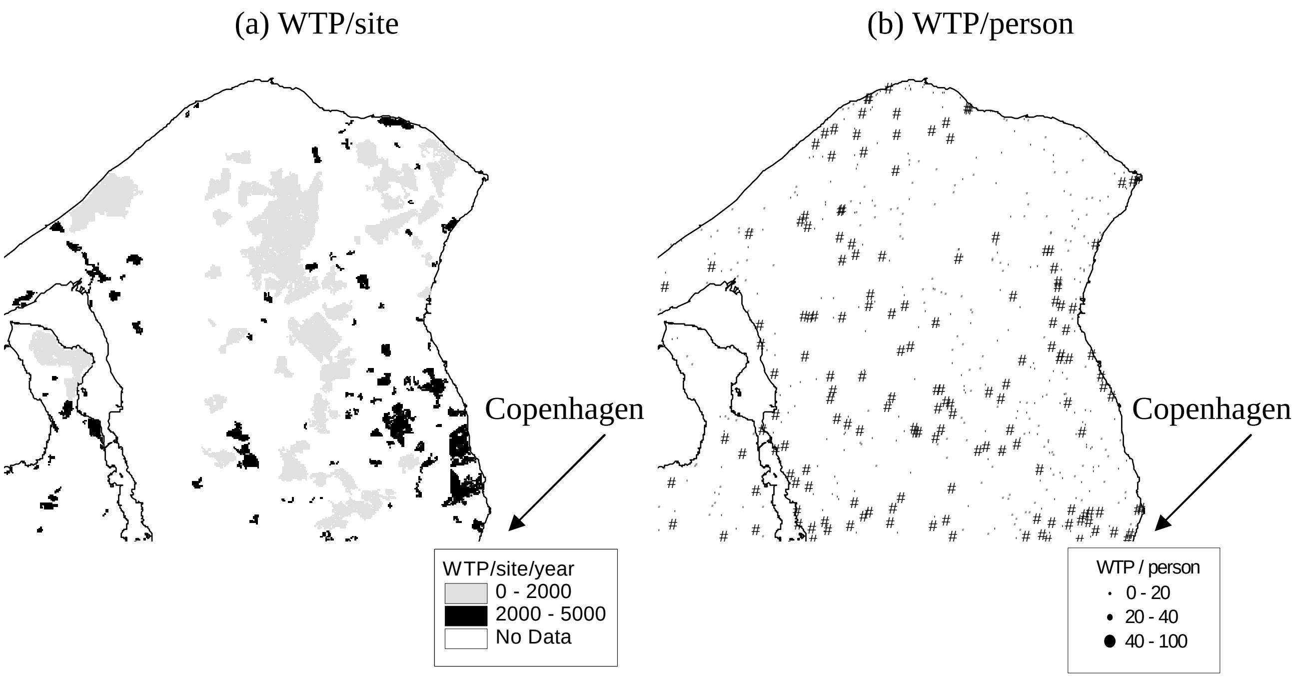

Figure 2: WTP for increase in forest area: (a) WTP for increase in area at individual sites [Euros/site/year]; (b) WTP per individual for increase in area at all sites [Euros/person/year]