

# Using geographically weighted choice models to account for the spatial heterogeneity of preferences

W Budziński, University of Warsaw

D Campbell, University of Stirling

M Czajkowski, University of Warsaw

U Demšar, University of St Andrews

N Hanley, University of St Andrews

Accepted for publication in *Journal of Agricultural Economics* published by Wiley-Blackwell.

## Abstract

In this paper, we investigate the prospects of using geographically weighted choice models for modelling of spatially clustered preferences. We argue that this is a useful way of generating highly-detailed spatial maps of willingness to pay for environmental conservation, given the costs of collecting data. The data used in this study comes from a discrete choice experiment survey regarding public preferences for the implementation of a new country-wide forest management and protection program in Poland. We combine it with high-resolution spatial data related to local forest characteristics. Using locally estimated discrete choice models we obtain location-specific estimates of willingness to pay (WTP). Variation in these estimates is explained by characteristics of the forests in their place of residence. The results are compared with those obtained from a more typical, two stage procedure which uses Bayesian posterior means of the mixed logit model random parameters to calculate location-specific estimates of WTP. We find that there are indeed strong spatial patterns to the benefits of changes in management to national forests. People living in areas with more species-rich forests and those living nearer to higher areas of mixed forests have significantly different WTP values than those living in other locations. This kind of information enables a better distributional analysis of the gains and losses from changes to natural resource management, and better targeting of investments in forest quality.

1 **Keywords:** discrete choice experiment, contingent valuation, willingness to pay, spatial  
2 heterogeneity of preferences, forest management, passive protection, litter, tourist infrastructure,  
3 mixed logit, geographically weighted model, weighted maximum likelihood, local maximum  
4 likelihood

5 **JEL classification:** Q23, Q28, I38, Q51, Q57, Q58

6 **Highlights:**

- 7 - A discrete choice experiment related to forest management and protection in Poland is  
8 conducted
- 9 - The spatial heterogeneity of respondents' preferences and willingness to pay is investigated
- 10 - The possibility to use the geographically weighted multinomial choice model is explored and  
11 compared with the typical 2-step approach
- 12 - The analysis highlights advantages and disadvantages of the two approaches

13

14

## 1 Introduction

2 Preferences for environmental goods are likely to display spatial patterns such as clustering. One  
3 reason for this is that since there are differences in the spatial configuration of these goods.  
4 Peoples' preferences can be expected to adapt to their local environments (Nielsen, Olsen and  
5 Lundhede, 2007) and to the availability of substitutes (Munro and Hanley, 1999). So, for instance,  
6 living in an area with high levels of forest cover may induce people to value the conservation of  
7 forests more highly than those who live in less-forested landscapes. Another line of reasoning  
8 concerns residential sorting: people's preferences for environmental goods partly determine  
9 where they choose to live, so that measures of preferences tend to be correlated with measures  
10 of environmental quality or with distance to environmental amenities (Timmins and Murdock,  
11 2007; Timmins and Schlenker, 2009; Baerenklau *et al.*, 2010). These two views on the reasons why  
12 spatial patterns are present in values and preferences for the environment can be thought of as  
13 different perspectives in the *directional drivers* which link values with environmental features.

14 From the policy and management perspective, enhancing our ability to produce detailed spatial  
15 maps of willingness to pay is important. For instance, national forest planners might want to target  
16 forest regeneration or investments in forest recreational resources in areas where the benefits of  
17 such actions would be highest. National water quality managers might, similarly, be interested in  
18 targeting costly actions to reduce pollution in a way which reflects the variation in values across a  
19 population. The application of benefit-cost thinking to spatial planning indeed requires such highly  
20 disaggregated benefits information. However, given the costs of undertaking a large number of  
21 original Willingness to Pay surveys at multiple sites, there is an interest in new methods which  
22 allows spatial maps of values to be generated in a data-efficient manner as a form of benefits  
23 transfer.

24 Accounting for spatial dependencies is insightful for various reasons, including in the aggregation  
25 of benefits, as illustrated by Bateman *et al.* (2006) in terms of identifying what "relevant  
26 population" of beneficiaries from a resource quality change should be considered. One of the most  
27 commonly encountered examples is the distance decay relationship (e.g., Jørgensen *et al.*, 2013).  
28 However, distance is not the only spatial factor which should be considered. For example,

1 availability of site substitutes can vary across space, and it is likely to influence willingness to pay  
2 considerably. However, because in most applied cases researchers are not able to account for all  
3 possible spatial factors, there is a need for a method, which would allow to take into account  
4 different spatial dimensions of preferences, with possible nonlinear relationships and clusters.

5 Recent developments in Geographical Information Science (GIScience) allow the researcher to  
6 obtain rich datasets containing detailed information about the spatial configuration of  
7 environmental goods and indeed in the socio-economic characteristics of households, and then  
8 use these to investigate spatial patterns in stated and revealed preferences for such environmental  
9 goods. In this paper, we employ a method which, while widely used in other areas of social science,  
10 remains relatively unknown in agricultural and resource economics: Geographically Weighted  
11 Regression (GWR) as proposed by Fotheringham, Charlton and Brunsdon (1998). Specifically, we  
12 apply geographical weighting to choice models to investigate the spatial relationship of willingness  
13 to pay for landscape characteristics, using the example of national forest management in Poland.  
14 The rationale of this statistical approach is that, if spatial clusters of preferences do exist, a locally-  
15 weighted maximum likelihood method can be used to derive location-specific estimates.  
16 Estimation of such models can provide us with insight regarding multiple possible spatial patterns  
17 of preferences and welfare measures. This is a semi-parametric approach in that no *a priori*  
18 assumptions about the spatial distribution of preferences are made.

19 Economists have used a number of ways for spatial value interpolation and mapping (Johnston and  
20 Ramachandran, 2014). This includes the use of micro-simulation methods such as combinatorial  
21 optimization, as used by Hynes, Hanley and O'Donoghue (2010) for spatial aggregation on  
22 contingent valuation survey data and a "two stage" analysis, which involves estimating a Mixed  
23 Logit (MXL) model to derive location-specific WTP values, and using these WTP estimates as  
24 dependent variables in a GIS-based spatial regression (see for example Campbell, Hutchinson and  
25 Scarpa (2009) and Czajkowski *et al.* (forthcoming-a)). The relevant question is thus whether the  
26 geographical weighting approach offers advantages relative to existing methods. We therefore  
27 compare the results obtained using geographically weighted multinomial logit (GWMNL) model  
28 with the latter, "two stage" approach. This comparison reveals that although there are similarities

1 in the spatial distributions of preferences identified using the two methods, the results differ in  
2 several important ways.

3

#### 4 **Geographically weighted models in the literature**

5 Geographically weighted models belong to the general class of “locally estimated” models. These  
6 recognize that the relationships between analyzed variables may be highly nonlinear which are  
7 therefore difficult to represent parametrically. Early examples of such models consist of use of  
8 spline functions (Wahba, 1990), LOWESS regression (Cleveland, 1979) and kernel regressions (Fan  
9 and Gijbels, 1996). The geographically weighted approach differs, because it recognizes nonlinear  
10 relationships with respect to spatial dimensions.

11 Early applications of geographically weighted models were based solely on linear local models.  
12 They were used for analysis of morbidity (Fotheringham, Charlton and Brunson, 1998), house  
13 price data (Brunson, Fotheringham and Charlton, 1999), economic growth (LeSage, 1999), school  
14 performance (Fotheringham, Charlton and Brunson, 2001) and urban temperatures (Páez, Uchida  
15 and Miyamoto, 2002). In the context of non-market valuation, this approach has been used with  
16 hedonic price models of house prices in Cho, Poudyal and Roberts (2008) or Saphores and Li (2012).

17 Local likelihood models were introduced by Fan, Heckman and Wand (1995) and Fan, Farmen and  
18 Gijbels (1998). These use weighted maximum likelihood estimators for inference. Applications of  
19 these techniques in discrete choice models are very limited, and were undertaken mostly in  
20 context of transportation. Locally estimated models are used either to recover WTP distribution  
21 non parametrically such as in Fosgerau (2007), Börjesson, Fosgerau and Algers (2012) and Koster  
22 and Koster (2015) or to analyze behavioral tendencies such as the implications of prospect theory  
23 (Hjorth and Fosgerau, 2012) and preference dynamics (Dekker, Koster and Brouwer, 2014).  
24 However, none of these approaches used local discrete choice models to analyze spatial  
25 heterogeneity – the issue considered in the present paper. Geographically weighted models for  
26 discrete response variables have been employed in the past, but in rather different contexts, not

1 connected with valuation of public goods, such as modelling urban growth (Luo and Kanala , 2008)  
 2 or predicting land use changes (Wang, Kockelman and Wang, 2011).

### 3 Methodology

4 We begin this section with a description of the geographically weighted multinomial logit. We  
 5 follow this with an explanation of the two-stage approach using location-specific WTP estimates  
 6 retrieved from the MXL model.

7

#### 8 *The Geographically weighted multinomial logit model*

9 GWML model is defined as follows. A respondent  $n$ 's utility from choosing alternative  $i$  in the  $j$ -th  
 10 choice task is given by:

$$11 \quad U_{ijn} = V_{ijn} + \varepsilon_{ijn} = \beta_l' \mathbf{X}_{ijn} + \varepsilon_{ijn} , \quad (1)$$

12 where the error term  $\varepsilon_{ijn}$  is assumed to be i.i.d with a Gumbel distribution.  $\beta_l$  is a set of parameters  
 13 for location  $l$ . The assumption which allows for the estimation of such a model is that individuals  
 14 located close to each other are assumed to have more similar preference parameters than  
 15 individuals located far away from each other, which is consistent with either of the directional  
 16 drivers in the introduction. As a result, the parameters become spatially correlated. For  
 17 convenience and ease of comparison between this approach and the approach used in Campbell,  
 18 Hutchinson and Scarpa (2009) and Czajkowski *et al.* (forthcoming-a) (described in detail below) we  
 19 estimated the GWML in WTP-space (Train and Weeks, 2005). This means that equation (1) was  
 20 reformulated as:

$$21 \quad \begin{aligned} U_{ijn} &= \beta_l^{\text{non-cost}} \mathbf{X}_{ijn}^{\text{non-cost}} - \beta_l^{\text{cost}} X_{ijn}^{\text{cost}} + \varepsilon_{ijn} = \\ &= \beta_l^{\text{cost}} \left( \frac{\beta_l^{\text{non-cost}}}{\beta_l^{\text{cost}}} \mathbf{X}_{ijn}^{\text{non-cost}} - X_{ijn}^{\text{cost}} \right) + \varepsilon_{ijn} = \beta_l^{\text{cost}} \left( \mathbf{a}_l' \mathbf{X}_{ijn}^{\text{non-cost}} - X_{ijn}^{\text{cost}} \right) + \varepsilon_{ijn} \end{aligned} \quad (2)$$

22 where now  $\mathbf{a}_l \beta_l^{\text{cost}}$  are parameters to be estimated.

1 Estimation of the GWMNL model is conducted by estimating  $L$  ‘local’ models, where  $L$  is a number  
 2 of distinct locations. In the case of our study, there were 253 distinct locations of respondents  
 3 (unique postal codes) and therefore this is the number of the local models. Each local model is  
 4 estimated via the weighted maximum likelihood method. The likelihood of individual  $n$  making a  
 5 choice in a  $j$ -th choice task in the  $l$ -th local model is given by a standard multinomial logit formula:

$$6 \quad L_{j,n}^l = \prod_{i=1}^l \left( \frac{\exp(\boldsymbol{\beta}_l' \mathbf{X}_{ijn})}{\sum_k \exp(\boldsymbol{\beta}_l' \mathbf{X}_{kjn})} \right)^{y_{ijn}}. \quad (3)$$

7 The weighted log-likelihood for  $l$ -th model is defined as follows:

$$8 \quad WL^l = \sum_{n=1}^N \sum_{j=1}^J \lambda(Lat_n, Long_n, b, l) \log(L_{j,n}^l), \quad (4)$$

9 where  $\lambda(Lat_n, Long_n, b, l)$  is a geographical weight (kernel), which depends on latitude and  
 10 longitude of individual  $n$ 's location,  $b$  which is called the ‘bandwidth parameter’ and the location  $l$   
 11 for which the local model is estimated. In order to take the panel nature of the data into account,  
 12 we have calculated robust standard errors that are clustered at the individual level. Note that  
 13 geographically weighted models normally use projected data, with the location given as metric  
 14 coordinates  $X$  and  $Y$  (easting and northing), to avoid having to use the complex and computationally  
 15 time-consuming 3D calculation of geographic distance with the two angular coordinates (latitude  
 16 and longitude) and indeed this was the same in our case. However, for clarity and to avoid potential  
 17 confusion with independent variables  $\mathbf{X}_{ijn}$  we refer to the two projected coordinates ( $X$  and  $Y$ )  
 18 here as longitude and latitude respectively. There are a few functional forms of  $\lambda(\cdot)$  proposed in  
 19 the literature. In what follows, we use the Gaussian kernel defined as:

$$20 \quad \lambda(Lat_n, Long_n, b, l) = \exp \left( -0.5 \frac{(Lat_n - Lat_l)^2 + (Long_n - Long_l)^2}{b^2} \right) \quad (5)$$

21 This is simply an exponential function of minus half of squared Euclidean distance of individual  $n$ 's  
 22 location from location  $l$  divided by the square of the bandwidth parameter. If a respondent lives

1 exactly in location  $l$  – this weight is equal to 1.<sup>1</sup> The use of this weight implies the clustering of  
2 similar values because observations near to location  $l$  have a larger bearing on the local model’s  
3 log-likelihood compared to observations that are further away. The bandwidth parameter  
4 therefore determines what “further away” means. If the bandwidth is low, then practically only  
5 the observations in very close proximity of given location influence the local model. Specifically,  
6 when  $b \rightarrow 0$  each local model is estimated using observations only from the given location.  
7 Analogously, when bandwidth is high, all local models will have similar parameter estimates, with  
8  $b \rightarrow \infty$  leading to a simple MNL model for the whole sample.

9 It is worth noting, that the choice of bandwidth may have a greater impact on the results than the  
10 choice of a specific weighting scheme (Fosgerau, 2007). There are several methods for choosing  
11 the bandwidth parameter available in the literature, with no apparent dominant approach. We  
12 tested three approaches, namely: using the corrected Akaike Information Criterion (AIC, Dekker,  
13 Koster and Brouwer, 2014), taking the lowest bandwidth for which all local models converge  
14 (Dekker, Koster and Brouwer, 2014), and using leave-one-individual-out cross-validation criterion  
15 (Fotheringham, Brunson and Charlton, 2003). To evaluate these, we used simulated data which  
16 utilized the designs utilized in our study. The results indicated that the available methods led to  
17 either under or over-smoothing and were unsatisfactory – a conclusion also voiced by Koster and  
18 Koster (2015). We therefore decided to use the ‘eye-balling’ approach they propose. In this  
19 approach, a researcher chooses the lowest bandwidth for which the model estimates satisfy a set  
20 of *a priori* specified conditions (e.g., achieving identification of all the models or avoiding extreme  
21 estimates). Pagan and Ullah (1999) recommend using this approach when the number of  
22 bandwidth parameters is not greater than 2, which is the case in our analysis. We decided that all

---

<sup>1</sup> For robustness check, we also tried different weighting functions, such as the spatially varying kernel (Fotheringham, Brunson and Charlton, 2003):  $\lambda(Lat_n, Long_n, b, l) = \exp\left(-\frac{R_{n,l}}{b}\right)$ , where  $R_{n,l}$  is the rank of the  $n$ -th location from  $l$ -th location in terms of the distance  $n$  is from  $l$ . The results were not much different from the Gaussian kernel.



1 WTP estimates should lie in interval [-100, 100] EUR, for results to be reliable. Bandwidths that  
2 result in WTP estimates outside of this range can be considered as inappropriate.

3

#### 4 ***Sample size***

5 There is a concern regarding the size of sample needed to calculate a local model with reliable  
6 parameter estimates. Generally, the literature provides little guidance in this regard. Sample sizes  
7 and the number of local models vary greatly depending on the application. In the cases where  
8 secondary survey data are used, such as in the case of house prices (Cho, Poudyal and Roberts,  
9 2008; Saphores and Li, 2012), school performance (Fotheringham, Charlton and Brunson, 2001)  
10 or land use (Wang, Kockelman and Wang (2011), the number of observations is typically high,  
11 ranging from around 3,700 to 50,000. Although, sometimes, for each observation there is  
12 estimated a separate local model. Applications using stated preference methods usually make use  
13 of much smaller datasets. For example, Fosgerau (2007) used data from 2000 respondents with 8  
14 choice tasks per person and estimate 441 local models. Börjesson, Fosgerau and Algers (2012)  
15 estimated the local models using responses from 1317 individuals. Koster and Koster (2015) used  
16 a dataset of 487 individuals, and reported a local model for each. In this respect, our sample does  
17 not seem to be “too small”, especially when considering the number of individuals per location.  
18 Nevertheless, as noted by one of our reviewers, for the GWMNL model the distribution of the  
19 individuals across the space maybe a bigger issue than the sample size. When there are locations  
20 with a low number of individuals, which are far away from any other individuals, the bandwidth  
21 needs to increase to provide any meaningful estimates (the bias-variance trade-off). Also, the  
22 locations with very high number of individuals may influence estimates of other local models  
23 unproportionally. Unfortunately, for now, the effect of the sampling on spatial distribution of WTP  
24 is not well researched, and it is not clear how it may affect the results.

25

#### 26 ***The Location specific mixed logit model***

27 The baseline for the comparison of the performance of GWMNL approach is provided by the  
28 location-specific conditional distributions retrieved from the mixed logit model ([Czajkowski et al.](#),

1 [forthcoming](#)) which can be estimated in WTP-space. In this model, respondent  $n$ 's utility from  
 2 choosing alternative  $i$  in the  $j$ -th choice task is given by:

$$3 \quad U_{ijn} = \beta_l^{\text{cost}} (\mathbf{a}_l' \mathbf{X}_{ijn}^{\text{non-cost}} - X_{ijn}^{\text{cost}}) + \varepsilon_{ijn} . \quad (6)$$

4 We assume that each location  $l$  has a separate, independent set of parameters and therefore, we  
 5 assume that all individuals within given location have homogeneous preferences. We prefer this  
 6 specification over usual, aspatial individual-specific one, because it is more comparable with the  
 7 GWMNL approach. The usual, individual-specific MXL takes into account different sources of  
 8 heterogeneity, while the GWMNL accounts for the spatial heterogeneity only.

9 Location-specific parameters are not directly observed, but it is possible to estimate their values  
 10 implied by each respondents' choices conditional on the population-level estimates of parameter  
 11 distributions (Bayesian posterior means) using the Bayes theorem:

$$12 \quad E(\mathbf{a}_l | \mathbf{y}_l, \mathbf{X}_l, \theta) = \int \mathbf{a}_l \frac{p(\mathbf{y}_l | \mathbf{X}_l, \theta, \mathbf{a}_l, \beta_l^{\text{cost}}) f(\mathbf{a}_l, \beta_l^{\text{cost}} | \theta)}{p(\mathbf{y}_l | \mathbf{X}_l, \theta)} d(\mathbf{a}_l, \beta_l^{\text{cost}}), \quad (7)$$

13 where  $p(\mathbf{y}_l | \mathbf{X}_l, \theta, \mathbf{a}_l, \beta_l^{\text{cost}})$  is the likelihood of all individuals from location  $l$  making the observed  
 14 choices conditional on the values of random parameters,  $p(\mathbf{y}_l | \mathbf{X}_l, \theta)$  is the same likelihood but  
 15 unconditional (so it is likelihood function for MXL) and  $f(\mathbf{a}_l, \beta_l^{\text{cost}} | \theta)$  is the assumed pdf function  
 16 of random parameters (normal distribution for all attributes except for the cost, which was  
 17 assumed to be log-normally distributed). For more details about this approach see [Czajkowski et](#)  
 18 [al. \(forthcoming\)](#).

19 Note, that contrary to the MXL model, in the GWMNL there is no need to specify a distribution  
 20 from which the parameters are drawn. It is also important to note that in this specification the MXL  
 21 model parameters associated with different locations are independent. Spatial dependence is,  
 22 therefore, accommodated indirectly as it arises from the calculation of conditional expected values  
 23 of random parameters.

24

## 1 Data

2 The original survey was conducted in 2010 on a representative sample<sup>2</sup> of 1001 Polish adults. The  
3 main objective of the survey was estimate public preferences over management options for the  
4 national forest area (rather than specific local forests). The attributes used to describe these  
5 management options were (1) passive protection of the most ecologically valuable forests,<sup>3</sup> (2)  
6 reducing the amount of litter (garbage, rubbish) in forests through tougher law enforcement and  
7 by increasing forest cleaning services and (3) increasing the level of recreational infrastructure,  
8 such as improved signposting of forest trails. The dataset used in this study was also exploited in  
9 in Czajkowski et al. (2014) and Czajkowski, Giergiczny and Greene (2014), where the attributes,  
10 experimental design and sampling strategy are described in detail.

11 Information on forest characteristics used in this study was obtained from two different sources.  
12 Firstly, the CORINE Land Cover (CLC) dataset was used. This project is coordinated by the European  
13 Environment Agency with the objective of collecting high resolution data for the whole continent.<sup>4</sup>  
14 CLC databases contain area data for objects with a minimum area of 5 ha and a width of more than  
15 100 meters. The second source of information used was the Polish Information System of State  
16 Forests. This contains very precise data about the characteristics of forests in Poland. The data  
17 from these sources was aggregated to 10x10 km squares.<sup>5</sup> In total, 3,307 such squares cover the

---

<sup>2</sup> We hired a professional polling agency that collected the questionnaires using high-quality, face-to-face computer-assisted surveying techniques. A multi-stage sampling strategy was employed, in which communities were randomly selected to represent different community types, and then within each of the selected communities. A starting point address was randomly selected and then a set of addresses was chosen using the random route method. Finally, a random selection of an adult household member was employed.

<sup>3</sup> By passive (as opposed to active) protection of the forest, we mean leaving the forest ecosystem without any human intervention, even if this results in (natural) changes in ecosystems. It was highlighted that passive protection does not preclude recreational use.

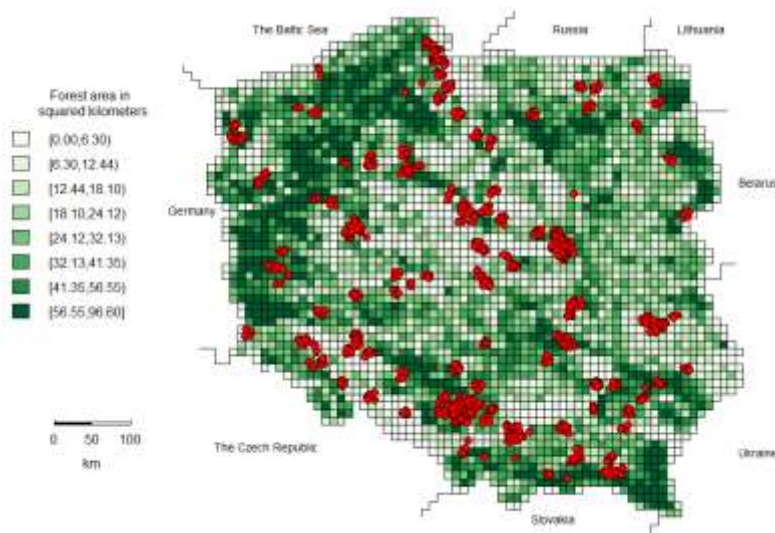
<sup>4</sup> See <http://www.eea.europa.eu/publications/COR0-landcover> for further information on the CORINE program.

<sup>5</sup> We also tested aggregating the 50x50 km resolution which provided equivalent results, although model fits were

1 area of Poland. Figure 2 presents a map with a distribution of DCE study respondents. The GIS data  
2 were associated with particular respondents using their ZIP-codes identifying their place of  
3 residence. For every respondent, the explanatory variables were calculated as weighted averages  
4 of forest characteristics in the 10x10 km area common with respondents' ZIP area code. The GIS  
5 variables used in this study are described in Table 1.

6

7 Figure 2. Respondents and forest area spatial distribution (red dots denote respondents' place of  
8 living)<sup>6</sup>



9

10

11

---

inferior.

<sup>6</sup> As some respondents reported the same ZIP-codes we jittered all of them with uniform random variable on  $[-5,5] \times [-5,5]$  km square. This also allow us to compute spatial weights matrix.

1 Table 1. GIS variables used to characterize the locations in which respondents' lived

Variable name	Description	Source	Mean	St. Dev.
Area of coniferous forests	Sum of areas of all coniferous forests [km <sup>2</sup> ]	Corine Land Cover	11.3202	13.3060
Area of deciduous forests	Sum of areas of all deciduous forests [km <sup>2</sup> ]	Corine Land Cover	4.2290	3.9805
Area of mixed forests	Sum of areas of all mixed forests [km <sup>2</sup> ]	Corine Land Cover	6.5767	6.1084
Average Euclidean distance to forest	It is average distance from any point in 10x10 km square to the nearest forest	Corine Land Cover	1.3075	0.8921
Area of forests with age > 120	Sum of areas of all forests older than 120 years [km <sup>2</sup> ]	Information System of State Forests	0.9586	1.3336
Area of forests with the number of species > 6	Sum of areas of all forests with the number of tree species greater than 6 [km <sup>2</sup> ]	Information System of State Forests	5.9285	7.1911
Built-up area	Built-up area [km <sup>2</sup> ]	Corine Land Cover	19.5532	19.3520

2  
 3 The input for our geographically weighted models was the spatial data set of the respondents,  
 4 where the location was given as coordinates of ZIP-codes, and the locations were linked to  
 5 responses and environmental variables. Prior to running the GWMNL model, the original WGS1984  
 6 coordinates were projected using the ETRS\_1989\_Poland\_CS92 coordinate system and the  
 7 projected coordinates were normalized.

8 The sample used in our study, 1001 individuals with 26 choice tasks per respondent, is sufficient  
 9 for estimation of 253 local models. Note that number of individuals vary between locations, with  
 10 some locations having only one individual, and some having more than 10 individuals. This result  
 11 in unbalanced panel, which may make estimates for locations with small number of individuals  
 12 imprecise. Specifically, in the case of MXL, when number of individuals in the given location is small,  
 13 the Bayesian posterior will be dominated by the chosen mixing distribution. In the case of GWMNL,  
 14 the estimation of local models involves a tradeoff of bias and variance of estimates ([Fotheringham,  
 15 Brunsdon and Charlton, 2003](#)). Nevertheless, if parameters values change continuously  
 16 throughout space the properly chosen bandwidth should provide reliable estimates for local  
 17 models.

## 1 Results

2 Following the approach outlined earlier, we estimated the GWMNL models for each of the 253  
 3 distinct locations in which our 1,001 respondents were located.<sup>7</sup> Table 2 presents the summary  
 4 statistics of the estimated parameters for this model, which are compared with results from the  
 5 MNL and the location-specific MXL models.<sup>8</sup> For the GWMNL model, we present means and  
 6 standard deviations of parameter estimates across 253 local models, which allows for  
 7 straightforward comparison with parameters from MXL model.

8  
 9 Table 2. Results of the MNL, location specific MXL and GWMNL models (standard errors in brackets,  
 10 coefficients in WTP-space, in EUR per year<sup>9</sup>).

Variable	MNL model	Location specific MXL model		GWMNL	
		Mean	Std. Dev.	Mean	Std. Dev.
	coef. (st. err.)	coef. (st. err.)	coef. (st. err.)	coef. (st. err.)	coef. (st. err.)
<b>NAT<sub>1</sub></b> (passive protection of most valuable forests – partial improvement)	14.8307*** (0.5673)	11.7961*** (0.4288)	7.6114*** (0.4559)	15.7107*** (0.1258)	6.8786*** (0.1701)
<b>NAT<sub>2</sub></b> (passive protection of most valuable forests – substantial improvement)	21.8207*** (0.7248)	16.8630*** (0.6129)	12.1756*** (0.6531)	23.0767*** (0.1826)	10.0173*** (0.2575)
<b>TRA<sub>1</sub></b> (the amount of litter in forests – partial improvement)	26.6697*** (0.8298)	17.4477*** (0.6066)	8.3347*** (0.4433)	28.3018*** (0.2001)	11.0293*** (0.2316)
<b>TRA<sub>2</sub></b> (the amount of litter in forests – substantial improvement)	35.6782*** (1.0664)	25.2329*** (0.9080)	14.1371*** (0.7508)	37.8590*** (0.2721)	14.7594*** (0.3624)

<sup>7</sup> In the estimation, we used the bandwidth parameter of 0.475 which was the lowest value to satisfy our a priori (albeit arbitrarily) specified condition that all models converge and in no location the estimated WTP is larger than 100 EUR. See section 3.1 for discussion and Appendix A for the robustness analysis of this assumption.

<sup>8</sup> Standard errors for the GWMNL model estimates were Monte-Carlo simulated using 10,000 repetitions, in which parameters of every locally estimated model were assumed to follow multivariate normal distribution.

<sup>9</sup> At 1 PLN  $\approx$  0.23 EUR  $\approx$  0.25 USD.

<i>INF</i> <sub>1</sub>	12.1400***	8.2641***	4.5826***	12.7121***	5.1219***
(tourist infrastructure – partial improvement)	(0.5269)	(0.3997)	(0.3251)	(0.0915)	(0.1003)
<i>INF</i> <sub>2</sub>	19.5598***	12.1145***	6.5138***	20.6085***	8.2931***
(tourist infrastructure – substantial improvement)	(0.6553)	(0.5044)	(0.3723)	(0.1388)	(0.1543)
<i>SQ</i>	37.2443***	-3.2461***	43.2764***	39.3765***	26.2943***
(alternative specific constant for the no-choice alternative)	(1.4032)	(0.7457)	(2.1355)	(0.3866)	(0.4566)
<i>COST</i>	0.0538***	-2.2472***	0.7066***	0.0575***	0.0240***
(annual cost – tax increase)	(0.0014)	(0.0400)	(0.0245)	(0.0004)	(0.0005)

---

#### Model characteristics

Log-likelihood (constant only)	-36,045.3765	-36,045.3765	-36,045.3765
Log-likelihood	-29,708.2771	-22,632.3017	-28,555.9663 <sup>10</sup>
Ben-Akiva Lerman's pseudo-R <sup>2</sup>	0.3282	0.4626	0.3550 <sup>14</sup>
McFadden's pseudo-R <sup>2</sup>	0.3308	0.3721	
AIC/ <i>n</i>	2.2836	1.7426	2.1950 <sup>14</sup>
<i>n</i> (observations)	26,026	26,026	26,026
<i>k</i> (parameters)	8	44	

---

\*\*\* p-value < 1%, \*\* p-value in [1%,5%), \* p-value in [5%, 10%)

1

2

3 As our model was estimated in WTP-space, parameters for all attributes can be interpreted directly

4 as willingness to pay. Qualitative results mimic those found in [Czajkowski et al. \(forthcoming\)](#),

5 namely that on average the individuals are willing to pay the most for reducing of amount of litter

6 and the least for improvements in infrastructure.

7 The comparison of WTP characteristics between the models reveals that means of the GWMNL

8 estimates are very close to MNL estimates. In contrast, for the location-specific MXL the mean WTP

9 values are significantly lower. The biggest difference is observed for the mean estimates of

10 alternative specific constant parameter (*SQ*) which has a reversed sign. Finally, we note that the

11 standard deviations are of similar magnitude in both approaches (except for *SQ* which has a higher

12 standard deviation in the location-specific MXL model).

---

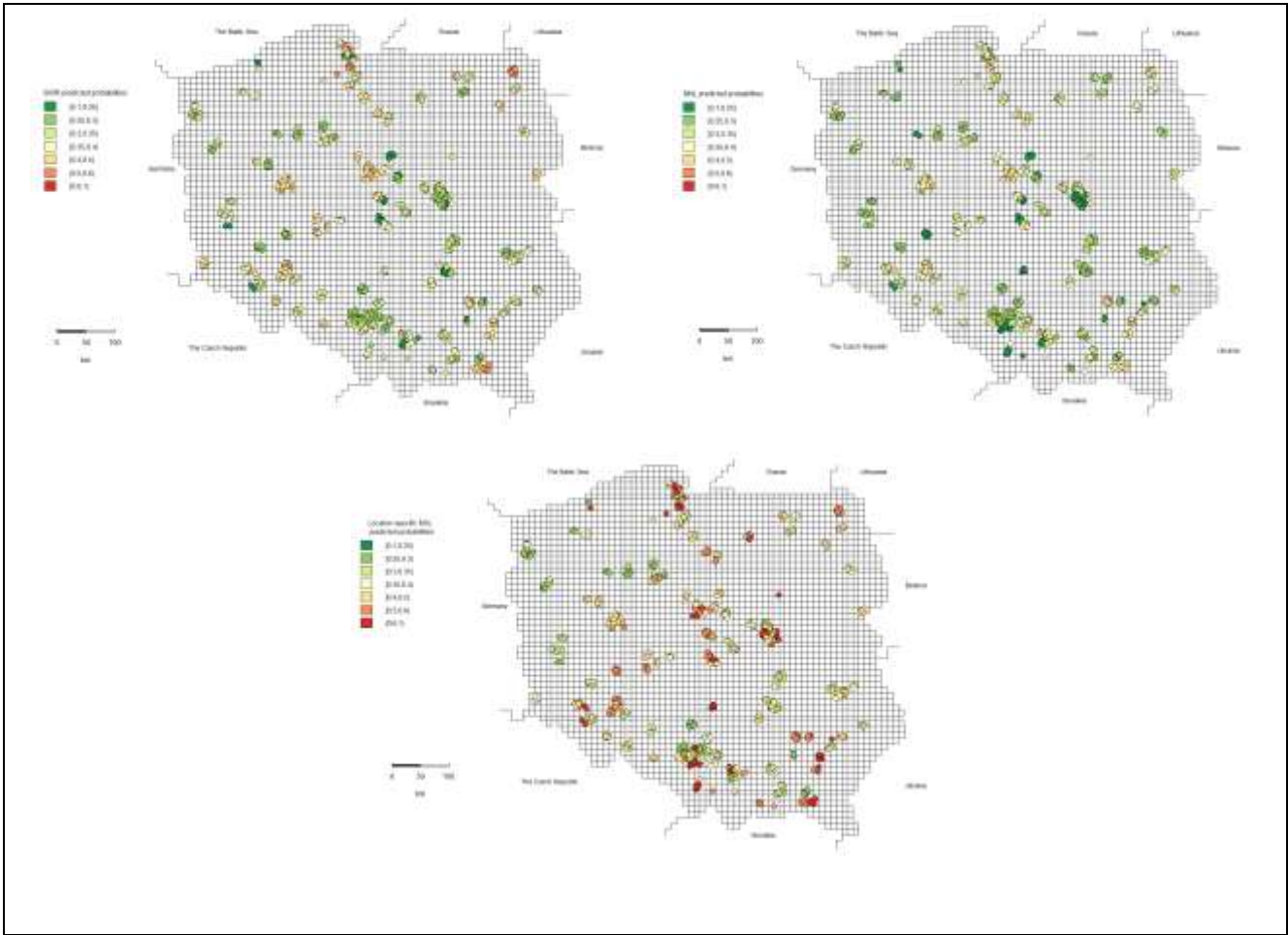
<sup>10</sup> For GWMNL these are mean values across 253 local models

1 There are, at least, three possible reasons for why we observe significant changes in the mean WTP  
2 values. Firstly, one can expect that not allowing for spatial correlation in the specification of the  
3 MXL model may lead to biased estimates. Secondly, the assumption of the MNL model form of  
4 local models in GWMNL may not be justified, because the error terms could in reality be correlated  
5 across alternatives. Lastly, and crucially, it may be driven by the distributional assumptions of the  
6 MXL model. While the MXL model assumes that the cost\**scale* parameter is log-normally  
7 distributed and that the marginal WTP distributions are all normally distributed, the GWMNL  
8 model is a semi-parametric approach and thus makes no such assumptions.

9 In order to compare the relative fit to the data provided by each of the three models (GWMNL,  
10 MNL and location -specific MXL models) we propose to use the Ben-Akiva-Lerman's pseudo-R<sup>2</sup>  
11 ([Ben-Akiva and Lerman, 1985](#)). This is a measure of predicted probabilities of choosing the  
12 alternatives which were actually chosen by respondents – an intuitive way of illustrating how well  
13 a model predicts the observed choices. We adapt this measure to the panel character of our data  
14 – because each respondent or each location was associated with *n* choice tasks, the joint  
15 probability of the observed series of *n* choices is normalized by taking its *n*-th root. Mean  
16 probabilities are presented in Table 2 (Ben-Akiva-Lerman's pseudo-R<sup>2</sup>), while their spatial  
17 distribution is illustrated in Figure 3. The pattern that emerges is clear – although the GWMNL  
18 approach provides a better fit than the MNL model, it is worse than the location-specific MXL  
19 model. Apparently the ability to generically account for the unobserved preference heterogeneity  
20 offers more of an improvement in fit than explicitly accounting for spatial correlations in the MNL  
21 model. However, we note that the predicted probabilities are highly correlated – the regions in  
22 which respondents' choices are relatively better or worse predicted are unchanged across the four  
23 models. This observation is further illustrated with the results provided in Table 3 – the correlation  
24 coefficients between location specific choice probabilities predicted by different models. Indeed,  
25 the predictions from GWMNL are more correlated with MNL, than with MXL.

26 Figure 3. Model-specific predicted probabilities of the observed choices (Ben-Akiva-Lerman's  
27 pseudo-R<sup>2</sup>)





1  
 2 Table 3. Correlation coefficients of model-specific predicted probabilities of the observed choices  
 3 (Ben-Akiva-Lerman's pseudo-R<sup>2</sup>)

	GWMNL ( $b=0.475$ )	MNL	Location specific MXL model
GWMNL ( $b=0.475$ )	1.0000	0.7782	0.4110
MNL	0.7782	1.0000	0.2313
Location specific MXL model	0.4110	0.2313	1.0000

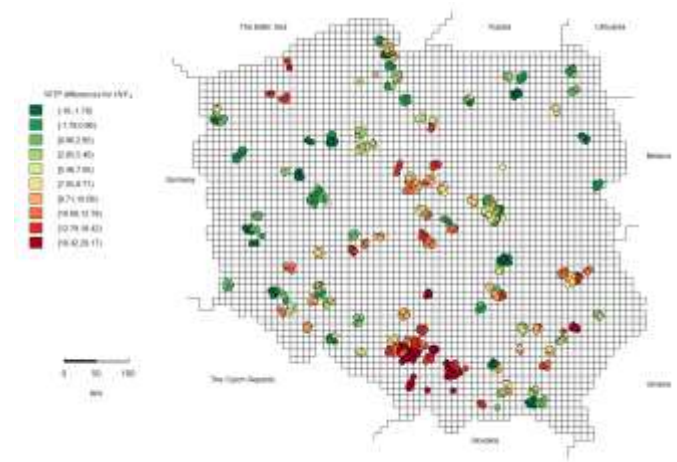
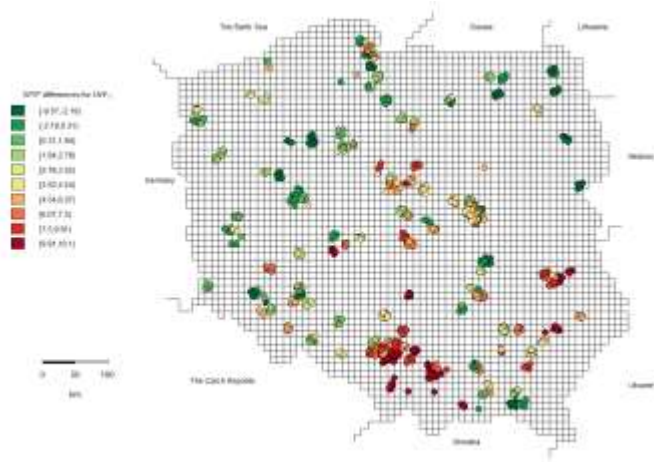
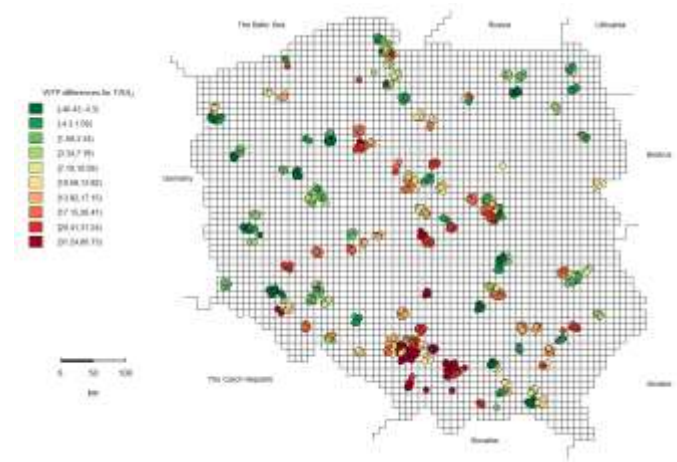
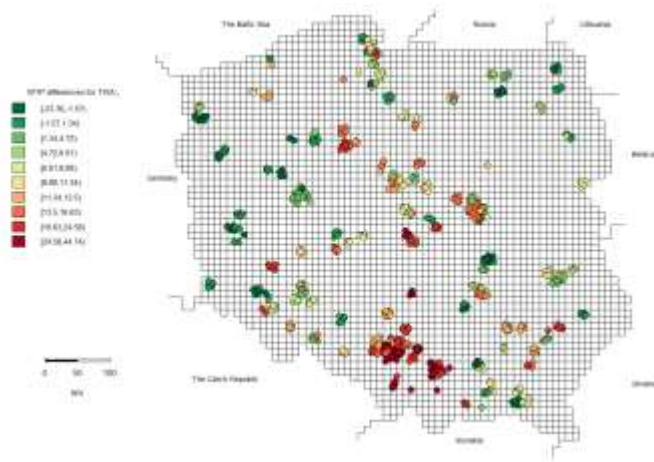
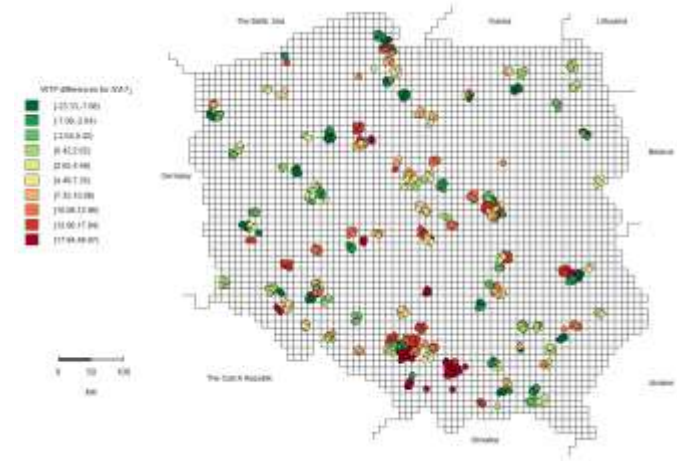
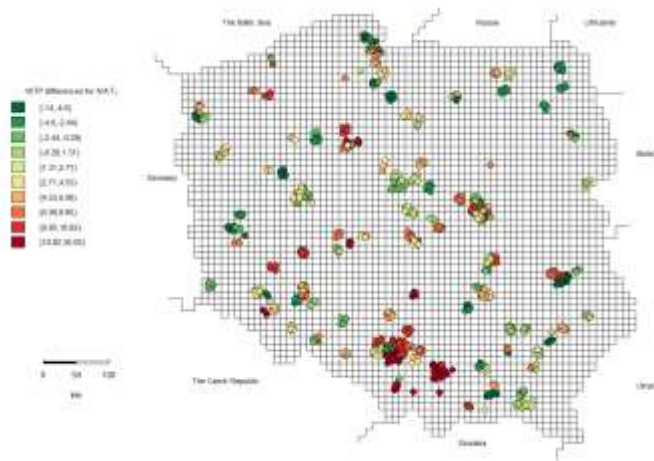
4  
 5 In order to analyze the discrepancies between the geographically weighted and the traditional two-  
 6 step approach further we can compare the differences between WTP estimates for every location  
 7 presented on map of Poland. This is done in the 7 panels of Figure 4. In these panels positive values

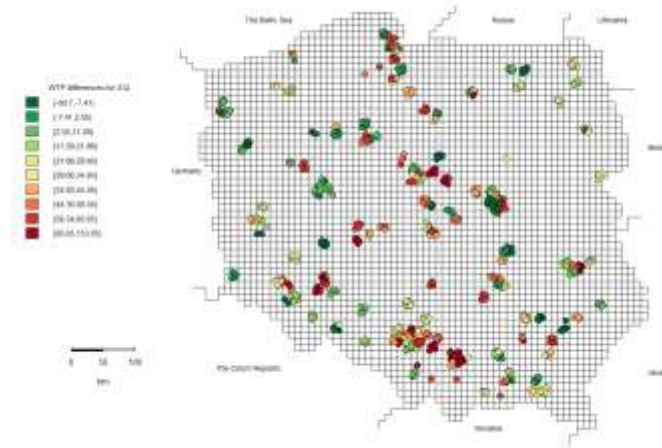
1 depict locations where GWMNL ( $b=0.475$ ) estimates are higher (red) and negative values depict  
2 locations where conditional expected values from location-specific MXL are higher (green).  
3 Distributions of these differences are not symmetric with respect to 0 (as every interval consists of  
4 10% of the sample). For all attributes, more than 80% of observations have positive values (higher  
5 values of WTP from GWMNL).

6 Graphical analysis reveals several spatial patterns of between-estimate differences which are  
7 consistent across all attributes. First of all, the largest differences can be observed in the central-  
8 south part of Poland near Cracow and Katowice cities, where GWMNL approach leads to much  
9 higher estimates of WTP. Secondly, in the west and north-east parts of Poland the differences seem  
10 to be much lower, sometimes negative. In the other parts of the country there is no clear spatial  
11 pattern, although it seems that also in the central part the differences are rather low (but positive).  
12 The fact, that some spatial patterns can be observed is an indication that these the two approaches  
13 recover different spatial dependencies of preferences. As GWMNL is designed to recover such  
14 spatial dependencies we can expect it to work better in this regard.

15 In order to investigate any systematic dependencies in these differences we estimated simple  
16 linear regressions in which their absolute values are explained by GIS variables and the number of  
17 observations per location. The full results are available in Appendix B. In short, we found that for  
18 locations with higher number of observations differences in WTP are significantly lower. This may  
19 indicate, that with more homogeneous sampling (with multiple observations per location) the two  
20 methods become more similar. What is more, we found that the differences are lower in areas  
21 with high coverage of very old forests, but are higher in built-up areas and areas with high coverage  
22 of forests with more than 6 tree species.

23  
24 Figure 4. Spatial distribution of differences between WTP estimates from GWMNL and conditional  
25 expected values from location-specific MXL.





1

2 Lastly, it is possible to perform a decomposition of the estimated WTP using GIS variables, similar

3 to [Czajkowski et al. \(forthcoming\)](#). To this end, seven regressions were estimated in which WTP

4 were explained by the same GIS variables as in [Czajkowski et al. \(forthcoming\)](#). The results from

5 linear regression model on the GWMNL ( $b=0.475$ ) results are given in Table 4, whereas the spatial

6 lag models based on the conditional expected values of random parameters from location-specific

7 MXL are presented in Table 5.<sup>11</sup> In all cases we decided to use only GIS variables and omit socio-

8 demographic variables, since these were insignificant in most cases for the GWMNL and location-

9 specific MXL approaches. Testing their joint significance we found that at 5% significance level

10 socio-demographic variables are significant only in  $NAT_1$  and  $NAT_2$  equations, and not significant in

11 any equation at 1% significance level. We think that these results may be caused partly by the fact

12 that we needed to average socio-demographic variables over individuals in the same location, as

13 in [Czajkowski et al. \(forthcoming\)](#), where this relationship was analyzed on the individual level,

14 socio-demographic variables were more significant.

---

<sup>11</sup> Note that in current analysis we used Ordinary Least Square instead of spatial lag model. It was not possible to estimate spatial lag model on WTP estimated from GWMNL as from definition these are highly spatially autocorrelated and therefore coefficient  $\rho$  (which is coefficient for spatial lag term) was almost equal to 1 and no other variable would then occur significant.

1 Looking firstly at the results based on the GWMNL results, we find that most of GIS variables are  
2 highly significant. We can compare these estimates with results in Table 5 in order to investigate  
3 structural differences in spatial heterogeneity in WTP estimates between the GWMNL and  
4 Bayesian posterior means from location-specific MXL. Greyed out cells of Table 4 indicate  
5 coefficients which have a different sign than equivalent coefficients in Table 5. This issue is the  
6 most prominent in regression for the *SQ*, where 3 variables have different signs, although they are  
7 all insignificant in this case. This problem also occurs, for almost every attribute, with “*Area of*  
8 *forests with age > 120*” variable. In case of location-specific MXL, this variable has a positive  
9 (although insignificant) coefficient for most attributes. What is also interesting is that in the current  
10 analysis variables “*Built-up area*” and “*Area of forests with no. of species > 6*” were significant in  
11 almost all cases. This differs to what we discovered when using the conditional expected values of  
12 random parameters from the MXL model where these variables were insignificant in all cases.

13

14 Table 4. Results of regressions in which WTP estimates from GWMNL are explained by GIS  
15 variables

	<b>SQ</b> (alternative specific constant for the no-choice alternative)	<b>NAT<sub>1</sub></b> (passive protection of most valuable forests – partial improvement)	<b>NAT<sub>2</sub></b> (passive protection of most valuable forests – substantial improvement)	<b>TRA<sub>1</sub></b> (the amount of litter in forests – partial improvement)	<b>TRA<sub>2</sub></b> (the amount of litter in forests – substantial improvement)	<b>INF<sub>1</sub></b> (tourist infrastructure – partial improvement)	<b>INF<sub>2</sub></b> (tourist infrastructure – substantial improvement)
Constant	34.7688*** (5.7861)	17.6132*** (1.5842)	27.0083*** (2.4027)	30.8844*** (2.4394)	44.8497*** (3.4203)	11.8696*** (1.0303)	22.0946*** (1.8038)
Area of coniferous forests	-0.1544 (0.1362)	-0.0727* (0.0373)	-0.1088* (0.0566)	-0.0732 (0.0574)	-0.1519* (0.0805)	0.0239 (0.0243)	-0.0449 (0.0425)
Area of deciduous forests	1.0730** (0.4567)	-0.1207 (0.1251)	-0.3464* (0.1897)	-0.0479 (0.1926)	-0.3411 (0.2700)	0.2376*** (0.0813)	0.2164 (0.1424)
Area of mixed forests	-0.1680 (0.3356)	-0.2419*** (0.0919)	-0.3022** (0.1393)	-0.2930** (0.1415)	-0.4331** (0.1984)	-0.0364 (0.0598)	-0.2270** (0.1046)
Area of forests with age >120	-7.0385*** (1.4597)	-0.4937 (0.3996)	-0.5343 (0.6061)	-1.3759** (0.6154)	-1.0958 (0.8628)	-1.4039*** (0.2599)	-2.0912*** (0.4550)
Average Euclidean distance to a forest	-1.7794 (2.3522)	-1.1623* (0.6440)	-1.7118* (0.9768)	-2.3111** (0.9917)	-3.8154*** (1.3904)	0.0347 (0.4189)	-1.1040 (0.7333)
Built-up area	0.2859*** (0.0793)	0.0813*** (0.0217)	0.1123*** (0.0329)	0.1454*** (0.0334)	0.1703*** (0.0469)	0.0129 (0.0141)	0.0405 (0.0247)
Area of forests with no. of species > 6	1.0447***	0.2774***	0.3147**	0.3567**	0.4133**	0.1161*	0.3829***

	(0.3331)	(0.0912)	(0.1383)	(0.1405)	(0.1969)	(0.0593)	(0.1039)
<b>Model characteristics</b>							
R <sup>2</sup>	0.2195	0.1215	0.0980	0.1293	0.0925	0.1556	0.1648
n (observations)	253	253	253	253	253	253	253
k (parameters)	8	8	8	8	8	8	8

1 \*\*\* p-value < 1%, \*\* p-value in [1%,5%), \* p-value in [5%, 10%)

2

3

4 Table 5. Results of spatial lag models in which Bayesian posterior means from MXL model are  
5 explained by GIS variables.

	<b>SQ</b> (alternative specific constant for the no-choice alternative)	<b>NAT<sub>1</sub></b> (passive protection of most valuable forests – partial improvement)	<b>NAT<sub>2</sub></b> (passive protection of most valuable forests – substantial improvement)	<b>TRA<sub>1</sub></b> (the amount of litter in forests – partial improvement)	<b>TRA<sub>2</sub></b> (the amount of litter in forests – substantial improvement)	<b>INF<sub>1</sub></b> (tourist infrastructure – partial improvement)	<b>INF<sub>2</sub></b> (tourist infrastructure – substantial improvement)
<b>Location-specific MXL</b>							
Constant	-35.0845*** [9.0012]	15.2120*** [1.9167]	23.2420*** [2.9933]	16.2849*** [2.1964]	26.6299*** [3.5287]	5.9459*** [1.0342]	8.9798*** [1.4828]
Area of coniferous forests	0.4807** [0.2108]	-0.0828** [0.0371]	-0.1442** [0.0595]	-0.0674* [0.0402]	-0.1537** [0.0668]	0.0054 [0.0202]	-0.0129 [0.0284]
Area of deciduous forests	2.5479*** [0.6502]	-0.3785*** [0.1139]	-0.6469*** [0.1829]	-0.1931 [0.1226]	-0.5320*** [0.2043]	0.0471 [0.0616]	0.0096 [0.0867]
Area of mixed forests	0.7688* [0.4489]	-0.1880** [0.0791]	-0.3133** [0.1270]	-0.2060** [0.0859]	-0.3404** [0.1425]	-0.0634 [0.0430]	-0.1022* [0.0607]
Area of forests with age >120	-0.6381 [2.2125]	0.6013 [0.3907]	0.9906 [0.6268]	0.4653 [0.4225]	0.9354 [0.7025]	-0.0776 [0.2117]	0.0971 [0.2983]
Average Euclidean distance to a forest	10.8442*** [3.3104]	-1.6825*** [0.5811]	-2.8926*** [0.9328]	-1.6013** [0.6309]	-3.2686*** [1.0499]	-0.2967 [0.3153]	-0.6871 [0.4454]
$\rho$	0.2555*** [0.0723]	0.1810** [0.0760]	0.1824** [0.0758]	0.3397*** [0.0681]	0.3325*** [0.0682]	0.3700*** [0.0664]	0.4090*** [0.0637]
<b>Model characteristics</b>							
Log Likelihood	-1255.8570	-814.7967	-934.2825	-838.5166	-966.3454	-665.4564	-753.4436
AIC/n	10.0246	6.5194	7.4643	6.7622	7.7718	5.4062	6.1279
n (observations)	253	253	253	253	253	253	253
k (parameters)	8	8	8	8	8	8	8

6 \*\*\* p-value < 1%, \*\* p-value in [1%,5%), \* p-value in [5%, 10%)

7

8 Overall, the results indicate that there are significant discrepancies with regard to spatial patterns  
9 recovered with the two methods. Differences in the signs and significance of coefficients for GIS  
10 variables demonstrate that the WTP distributions differ in structure between these two  
11 approaches. As the GWMNL model explicitly deals with spatial heterogeneity (rather than trying  
12 to recover it indirectly post estimation) it could be considered more reliable. It is also important to

1 note that in all models using the GWMNL estimates the  $R^2$  in all models is very low, which suggest  
2 that the GIS variables we used explain only a small fraction of the observed variance. Therefore,  
3 assuming that the estimated values obtained from the GWMNL models are the true values, most  
4 of the heterogeneity in WTP is caused by some other factors, which we do not account for. This  
5 may partly be due to the spatial distribution of forest characteristics in Poland.

6  
7 In some cases the differences between forests which lie next to each other are large, and therefore  
8 significant variance in their values may occur even on a local level. Because of that, our model may  
9 not be able to recover relationship between these factors and preference heterogeneity correctly,  
10 as we do not have sufficiently detailed data to model this local variation. It is possible that the  
11 GWMNL approach would perform better in the case of preferences for environmental goods which  
12 change more gradually.

13

#### 14 **Discussion and conclusions**

15 In this paper we investigate two alternative methods for addressing spatial patterns in willingness  
16 to pay for changes to an environmental good. We argued that knowledge of how willingness to pay  
17 for a specific environmental change varies across space is useful from a policy and management  
18 perspective. Knowing the spatial pattern of values can help resource managers target investments  
19 in site quality, or in new forests, as investments can be directed at locations where they are most  
20 valued. It also enables a higher-resolution identification of the gainers and losers from changes in  
21 resource management, since now the location of individuals who gain by a given amount from a  
22 policy can be mapped. An example of how important this might be is provided by Hynes et al (2010)  
23 using contingent valuation survey data for Ireland. They found that allowing for spatial differences  
24 in the characteristics of individuals (in their case, farmers) via a micro-simulation approach resulted  
25 in significantly different estimates of willingness to pay for biodiversity conservation at the regional  
26 level, and thus significantly different measures of aggregate benefits, compared to an aggregation  
27 method which did not take into account spatial variation in the characteristics of beneficiaries.



1 Given the costs of original survey work, the ability to produce such maps cost-effectively is highly  
2 desirable so long as results are sufficiently robust. The case study used to generate the data with  
3 which the two alternative methods are tested concerns the management of forests in Poland. We  
4 introduced a novel method of geographically weighted discrete choice models to account for  
5 spatial heterogeneity of model parameters, and compared this to a standard “two-step” approach  
6 using the MXL model and posterior Bayesian means of random parameters ([Czajkowski et al.,  
7 forthcoming](#)). The comparison focused on the consequences for estimates of willingness to pay, as  
8 this is typically the focus in environmental economics. Our analysis revealed significant differences  
9 in the estimates of WTP between the two methods. Specifically, for all forest attributes, mean WTP  
10 was much higher when obtained using GWMNL. We also found some important structural  
11 differences – several land cover variables appeared to have a reversed effect on WTP when the  
12 models are compared.

13 We note that both methods have shortcomings. MXL assumes that random parameters are drawn  
14 from spatially independent distributions and relies heavily on distributional assumptions. Any  
15 spatial correlations that are observed are obtained from posterior Bayesian means, and they are,  
16 obviously, conditional on these assumptions. On the other hand, GWMNL ignores other (non-  
17 spatial) sources of preference heterogeneity, such as variations in income, which other research  
18 has shown to matter. This shortcoming of the GWMNL model could be addressed by using a more  
19 complicated weighting function which accounts for socio-demographic characteristics. However,  
20 this may require the use of multiple bandwidth parameters. We tried to incorporate heterogeneity  
21 with respect to socio-demographic characteristics by using a second bandwidth parameter, but the  
22 results were not satisfactory, particularly when using the eye-balling technique to determine the  
23 optimal bandwidth value. First of all, in such a setting there was no unique ‘lower’ bandwidth for  
24 which our criterion was fulfilled, e.g. in case of having criterion of all WTP being below 100 EUR,  
25 there could exist several pairs of bandwidths for which all WTP values are below this chosen level,  
26 such that an increase in any of bandwidths would lead to exceeding this level. Researchers would  
27 therefore need some additional criterion to choose the preferred model. Second, the number of  
28 models that one needs to estimate increases very quickly with the number of bandwidth  
29 parameters, e.g. if for a single bandwidth researcher wants to evaluate 20 different models, then



1 for two bandwidth parameters about  $20 \times 20 = 400$  models need to be evaluated. Another possibility  
2 is to include unobserved heterogeneity in local models via estimation of latent class models instead  
3 of simple multinomial logits. Such approach was used by [Koster and Koster \(2015\)](#) (although not in  
4 the spatial context) and may be a preferable approach to GWMNL. This also introduces additional  
5 computational burdens, however.

6 In light of the above, it is difficult to conclude that either of the two methods presented here for  
7 the spatial modelling of willingness to pay is superior. Additional analysis of the reliability of  
8 Bayesian posterior means and their vulnerability to MXL assumptions is needed. To some extent,  
9 [Hess \(2010\)](#) approach this issue, but with no focus on welfare measures. Moreover, methods for  
10 choosing an appropriate bandwidth parameter in geographically weighted discrete choice models  
11 are currently under-developed. Many of the methods proposed are unsatisfactory and lead to poor  
12 results. This is especially important for more advanced kernels and with multiple bandwidth  
13 parameters, which would allow for spatial sources of heterogeneity. Lastly, more research on  
14 sampling design for such spatial models is needed, in terms of the implications for estimates of  
15 spatial relationships. We could expect that more homogenous geographic sampling would provide  
16 better results, although then the sample will no longer be representative.

17

## 1 References

- 2 Baerenklau, K. A., González-Cabán, A., Paez, C., and Chavez, E., 2010. Spatial allocation of forest recreation  
3 value. *Journal of Forest Economics*, 16(2):113-126.
- 4 Bateman, I. J., Day, B. H., Georgiou, S., and Lake, I., 2006. The aggregation of environmental benefit values:  
5 Welfare measures, distance decay and total WTP. *Ecological Economics*, 60(2):450-460.
- 6 Ben-Akiva, M., and Lerman, S. R., 1985. Discrete Choice Analysis: Theory and Application to Travel  
7 Demand. MIT Press, Cambridge, MA.
- 8 Bliemer, M. C. J., Rose, J. M., and Hess, S., 2008. Approximation of Bayesian Efficiency in Experimental  
9 Choice Designs. *Journal of Choice Modelling*, 1(1):98-127.
- 10 Börjesson, M., Fosgerau, M., and Algers, S., 2012. Catching the tail: Empirical identification of the  
11 distribution of the value of travel time. *Transportation Research Part A: Policy and Practice*,  
12 46(2):378-391.
- 13 Brunsdon, C., Fotheringham, A. S., and Charlton, M., 1999. Some notes on parametric significance tests for  
14 geographically weighted regression. *Journal of Regional Science*, 39(3):497-524.
- 15 Campbell, D., Hutchinson, W. G., and Scarpa, R., 2009. Using Choice Experiments to Explore the Spatial  
16 Distribution of Willingness to Pay for Rural Landscape Improvements. *Environment and planning*  
17 *A*, 41(1):97-111.
- 18 Cho, S.-H., Poudyal, N. C., and Roberts, R. K., 2008. Spatial analysis of the amenity value of green open  
19 space. *Ecological Economics*, 66(2):403-416.
- 20 Cleveland, W. S., 1979. Robust locally weighted regression and smoothing scatterplots. *Journal of the*  
21 *American Statistical Association*, 74(368):829-836.
- 22 Czajkowski, M., Bartczak, A., Giergiczy, M., Navrud, S., and Żylicz, T., 2014. Providing Preference-Based  
23 Support for Forest Ecosystem Service Management. *Forest Policy and Economics*, 39:1-12.
- 24 Czajkowski, M., Budziński, W., Campbell, D., Giergiczy, M., and Hanley, N., forthcoming. Spatial  
25 heterogeneity of willingness to pay for forest management. *Environmental and Resource*  
26 *Economics*.
- 27 Czajkowski, M., Giergiczy, M., and Greene, W. H., 2014. Learning and fatigue effects revisited.  
28 Investigating the effects of accounting for unobservable preference and scale heterogeneity. *Land*  
29 *Economics*, 90(2):323-350.
- 30 Dekker, T., Koster, P., and Brouwer, R., 2014. Changing with the Tide: Semiparametric Estimation of  
31 Preference Dynamics. *Land Economics*, 90(4):717-745.
- 32 Fan, J., Farmen, M., and Gijbels, I., 1998. Local maximum likelihood estimation and inference. *Journal of*  
33 *the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):591-608.
- 34 Fan, J., and Gijbels, I., 1996. Local polynomial modelling and its applications: monographs on statistics and  
35 applied probability 66. CRC Press.
- 36 Fan, J., Heckman, N. E., and Wand, M. P., 1995. Local polynomial kernel regression for generalized linear  
37 models and quasi-likelihood functions. *Journal of the American Statistical Association*,  
38 90(429):141-150.
- 39 Ferrini, S., and Scarpa, R., 2007. Designs with a priori information for nonmarket valuation with choice  
40 experiments: A Monte Carlo study. *Journal of Environmental Economics and Management*,  
41 53(3):342-363.
- 42 Fosgerau, M., 2007. Using nonparametrics to specify a model to measure the value of travel time.  
43 *Transportation Research Part A: Policy and Practice*, 41(9):842-856.
- 44 Fotheringham, A. S., Brunsdon, C., and Charlton, M., 2003. Geographically weighted regression: the  
45 analysis of spatially varying relationships. John Wiley & Sons.

- 1 Fotheringham, A. S., Charlton, M. E., and Brunson, C., 2001. Spatial variations in school performance: a  
2 local analysis using geographically weighted regression. *Geographical and Environmental*  
3 *Modelling*, 5(1):43-66.
- 4 Fotheringham, S., Charlton, M., and Brunson, C., 1998. Geographically weighted regression: a natural  
5 evolution of the expansion method for spatial data analysis. *Environment and planning A*,  
6 30(11):1905-1927.
- 7 Hess, S., 2010. Conditional parameter estimates from Mixed Logit models: distributional assumptions and  
8 a free software tool. *Journal of Choice Modelling*, 3(2):134-152.
- 9 Hjorth, K., and Fosgerau, M., 2012. Using prospect theory to investigate the low marginal value of travel  
10 time for small time changes. *Transportation Research Part B: Methodological*, 46(8):917-932.
- 11 Jørgensen, S. L., Olsen, S. B., Ladenburg, J., Martinsen, L., Svenningsen, S. R., and Hasler, B., 2013. Spatially  
12 induced disparities in users' and non-users' WTP for water quality improvements—Testing the  
13 effect of multiple substitutes and distance decay. *Ecological Economics*, 92(0):58-66.
- 14 Koster, P. R., and Koster, H. R. A., 2015. Commuters' preferences for fast and reliable travel: A semi-  
15 parametric estimation approach. *Transportation Research Part B: Methodological*, 81, Part 1:289-  
16 301.
- 17 LeSage, J. P., 1999. A spatial econometric examination of China's economic growth. *Geographic*  
18 *Information Sciences*, 5(2):143-153.
- 19 Luo, J., and Kanala, N. K., Year. Modeling urban growth with geographically weighted multinomial logistic  
20 regression. Geoinformatics 2008 and Joint conference on GIS and Built Environment: The Built  
21 Environment and its Dynamics, International Society for Optics and Photonics, 71440M-71440M-  
22 71411.
- 23 Munro, A., and Hanley, N., 1999. Information, Uncertainty and Contingent Valuation. In: *Contingent*  
24 *Valuation of Environmental Preferences: Assessing Theory and Practice in the USA, Europe, and*  
25 *Developing Countries*, I. J. Bateman and K. G. Willis, eds., Oxford University Press.
- 26 Nielsen, A. B., Olsen, S. B., and Lundhede, T., 2007. An Economic Valuation of the Recreational Benefits  
27 Associated with Nature-Based Forest Management Practices. *Landscape and Urban Planning*,  
28 80:63-71.
- 29 Páez, A., Uchida, T., and Miyamoto, K., 2002. A general framework for estimation and inference of  
30 geographically weighted regression models: 1. Location-specific kernel bandwidths and a test for  
31 locational heterogeneity. *Environment and planning A*, 34(4):733-754.
- 32 Pagan, A., and Ullah, A., 1999. Nonparametric econometrics. Cambridge university press.
- 33 Sándor, Z., and Wedel, M., 2001. Designing conjoint choice experiments using managers' prior beliefs.  
34 *Journal of Marketing Research*, 38(4):430-444.
- 35 Saphores, J.-D., and Li, W., 2012. Estimating the value of urban green areas: A hedonic pricing analysis of  
36 the single family housing market in Los Angeles, CA. *Landscape and Urban Planning*, 104(3):373-  
37 387.
- 38 Scarpa, R., and Rose, J. M., 2008. Design Efficiency for Non-Market Valuation with Choice Modelling: How  
39 to Measure it, What to Report and Why. *Australian Journal of Agricultural and Resource*  
40 *Economics*, 52(3):253-282.
- 41 Timmins, C., and Murdock, J., 2007. A revealed preference approach to the measurement of congestion in  
42 travel cost models. *Journal of Environmental Economics and Management*, 53(2):230-249.
- 43 Timmins, C., and Schlenker, W., 2009. Reduced-Form Versus Structural Modeling in Environmental and  
44 Resource Economics. *Annual Review of Resource Economics*, 1(1):351-380.
- 45 Train, K., and Weeks, M., 2005. Discrete Choice Models in Preference Space and Willingness-to-Pay Space.  
46 In: *Applications of Simulation Methods in Environmental and Resource Economics*, R. Scarpa and  
47 A. Alberini, eds., Springer Netherlands, 1-16.
- 48 Wahba, G., 1990. Spline models for observational data. Siam.

1 Wang, Y., Kockelman, K., and Wang, X., 2011. Anticipation of land use change through use of  
2 geographically weighted regression models for discrete response. *Transportation Research*  
3 *Record: Journal of the Transportation Research Board*, (2245):111-123.

4

5

## 1 **Appendix A**

2 In this section we check robustness of our results by analyzing differences in estimates for 3  
3 different values of bandwidth parameter: 0.275, 0.475 and 0.6. 0.475 is a value used throughout  
4 the paper, 0.275 is the lowest bandwidth value for which all WTP except SQ are below 100 EUR  
5 and 0.6 is the highest value we considered (as it is generally recommended to use lowest  
6 bandwidth value which fulfill chosen criteria). In Table A1. we provide characteristics of WTP for  
7 these values of bandwidth analogously as it was done in Table 2. We note that mean WTP are quite  
8 robust with respect to choice of a bandwidth. Although mean estimates for bandwidth values of  
9 0.275 and 0.6 lie more than two standard errors from estimates for bandwidth = 0.475, they are  
10 fairly close to each other. Dispersion of WTP change significantly between different bandwidth  
11 values. To investigate how change in dispersion influence our results we conducted a graphical  
12 analysis presented in Figure A1. In every panel we plotted parameters estimates for chosen  
13 bandwidth values: 0.275 (blue lines), 0.475 (red lines) and 0.6 (yellow lines) with their 95%  
14 confidence intervals (dotted lines) against longitude. We see that, indeed increased dispersion for  
15 bandwidth of 0.275 leads to several 'peaks', with very wide confidence intervals, for every  
16 parameter. Because of that for lower bandwidth value many WTP estimates in local models  
17 occurred insignificant. Estimates for 0.6 bandwidth usually lies within confidence interval for  
18 estimates for bandwidth of 0.475. We therefore conclude that choosing bandwidth of 0.475 is  
19 justified as it allows to avoid implausible 'peaks' in WTP (especially for SQ) and insignificant  
20 estimates. We also do not see any reason to choose bandwidth above 0.475 as the plots does not  
21 reveal any particular under-smoothing when comparing lines for bandwidth of 0.475 with lines for  
22 bandwidth of 0.6.

23

- 1 Table A1. Characteristics of distributions of estimated WTP from GWMNL model for different
- 2 bandwidth values (standard errors in [] brackets, coefficients in WTP-space, in EUR per year)

	Bandwidth = 0.275		Bandwidth = 0.475		Bandwidth = 0.6	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
<b>NAT<sub>1</sub></b> (passive protection of most valuable forests – substantial improvement)	16.4541*** [0.2225]	10.9360*** [0.5033]	15.7107*** [0.1258]	6.8786*** [0.1701]	15.5389*** [0.0968]	5.7223*** [0.1138]
<b>NAT<sub>2</sub></b> (passive protection of most valuable forests – partial improvement)	24.0868*** [0.3136]	15.4706*** [0.5758]	23.0767*** [0.1826]	10.0173*** [0.2575]	22.7602*** [0.1390]	8.1253*** [0.1682]
<b>TRA<sub>1</sub></b> (the amount of litter in forests – partial improvement)	29.6563*** [0.3318]	16.7661*** [0.4383]	28.3018*** [0.2001]	11.0293*** [0.2316]	27.8463*** [0.1530]	9.0786*** [0.1704]
<b>TRA<sub>2</sub></b> (the amount of litter in forests – substantial improvement)	39.8062*** [0.4591]	22.7725*** [0.6970]	37.8590*** [0.2721]	14.7594*** [0.3624]	37.1348*** [0.2041]	11.8575*** [0.2446]
<b>INF<sub>1</sub></b> (tourist infrastructure – partial improvement)	12.9375*** [0.1469]	7.5310*** [0.1862]	12.7121*** [0.0915]	5.1219*** [0.1003]	12.6410*** [0.0736]	4.3811*** [0.0793]
<b>INF<sub>2</sub></b> (tourist infrastructure – substantial improvement)	21.1835*** [0.2223]	11.5530*** [0.2611]	20.6085*** [0.1388]	8.2931*** [0.1543]	20.4056*** [0.1064]	7.1172*** [0.1127]
<b>SQ</b> (alternative specific constant for the no-choice alternative)	38.7161*** [0.6308]	37.5901*** [1.1447]	39.3765*** [0.3866]	26.2943*** [0.4566]	39.5860*** [0.3046]	23.3231*** [0.3690]
<b>COST</b> (annual cost – tax increase)	0.0626*** [0.0005]	0.0403*** [0.0008]	0.0575*** [0.0004]	0.0240*** [0.0005]	0.0561*** [0.0003]	0.0182*** [0.0003]

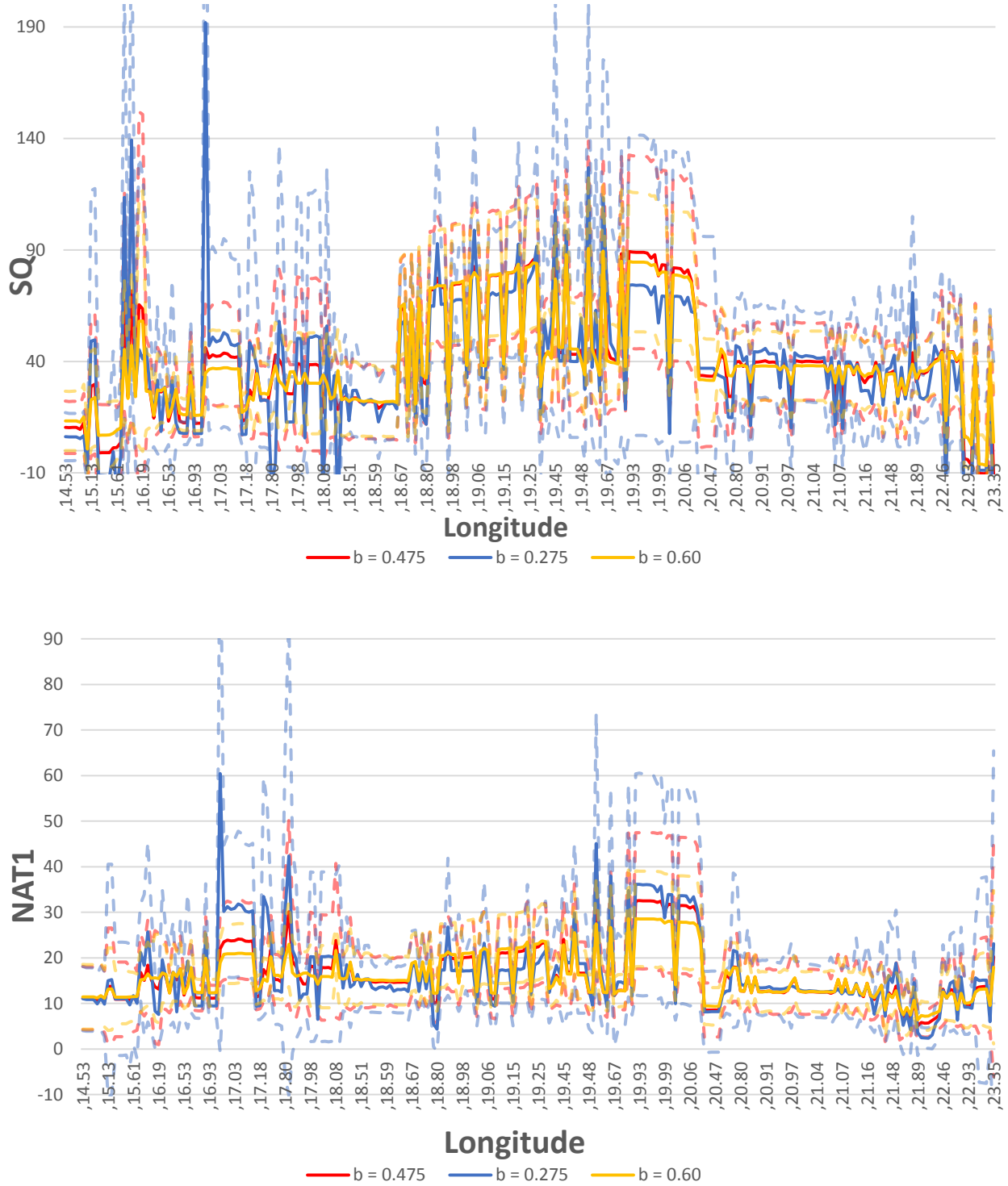
3 \*\*\* p-value < 1%, \*\* p-value in [1%,5%), \* p-value in [5%, 10%)

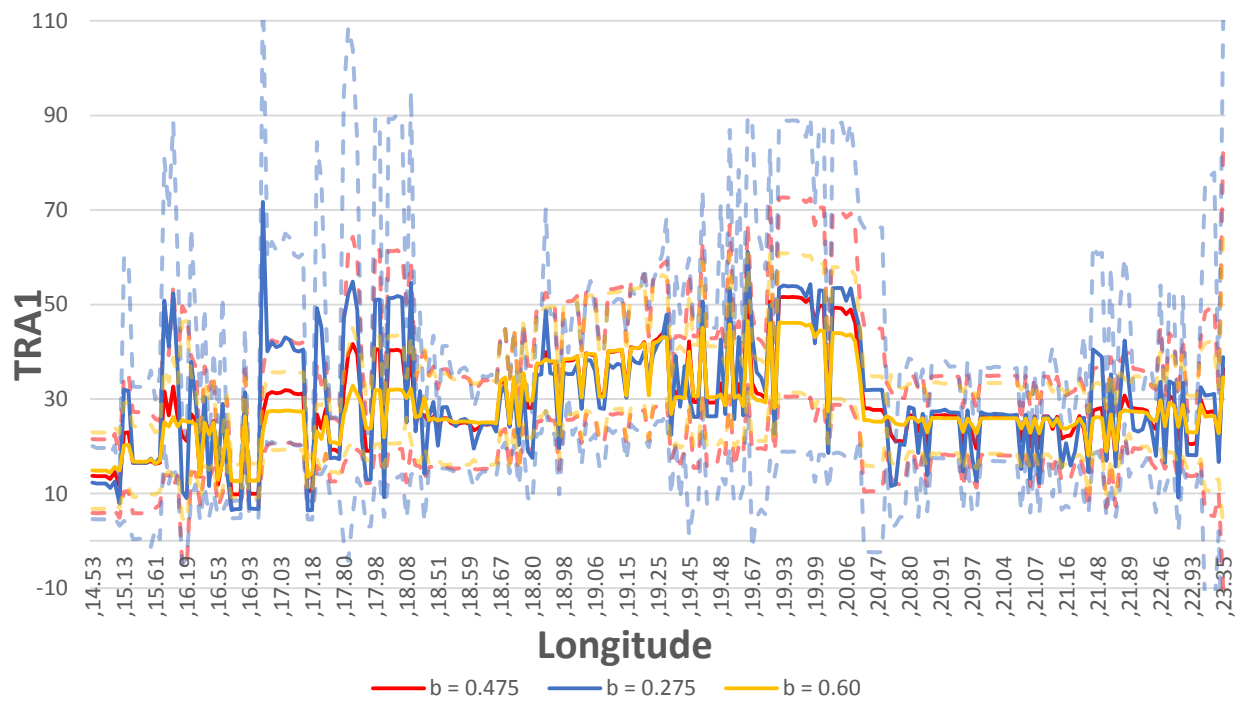
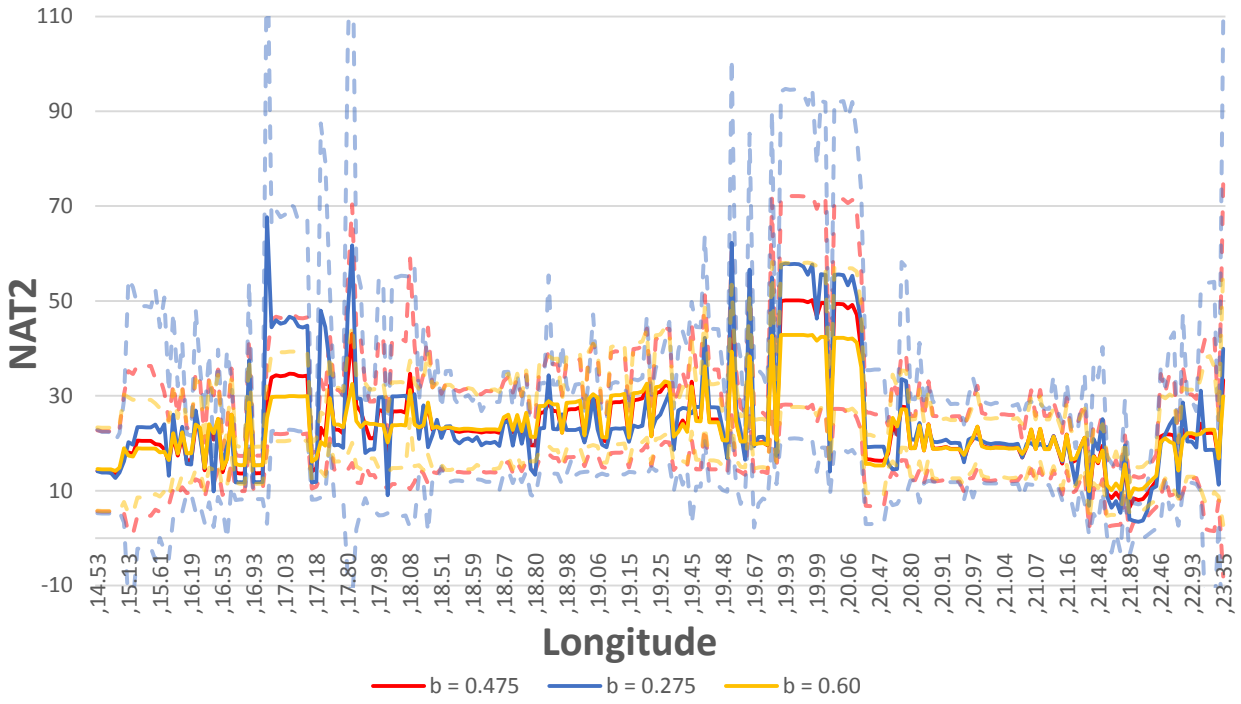
4

5

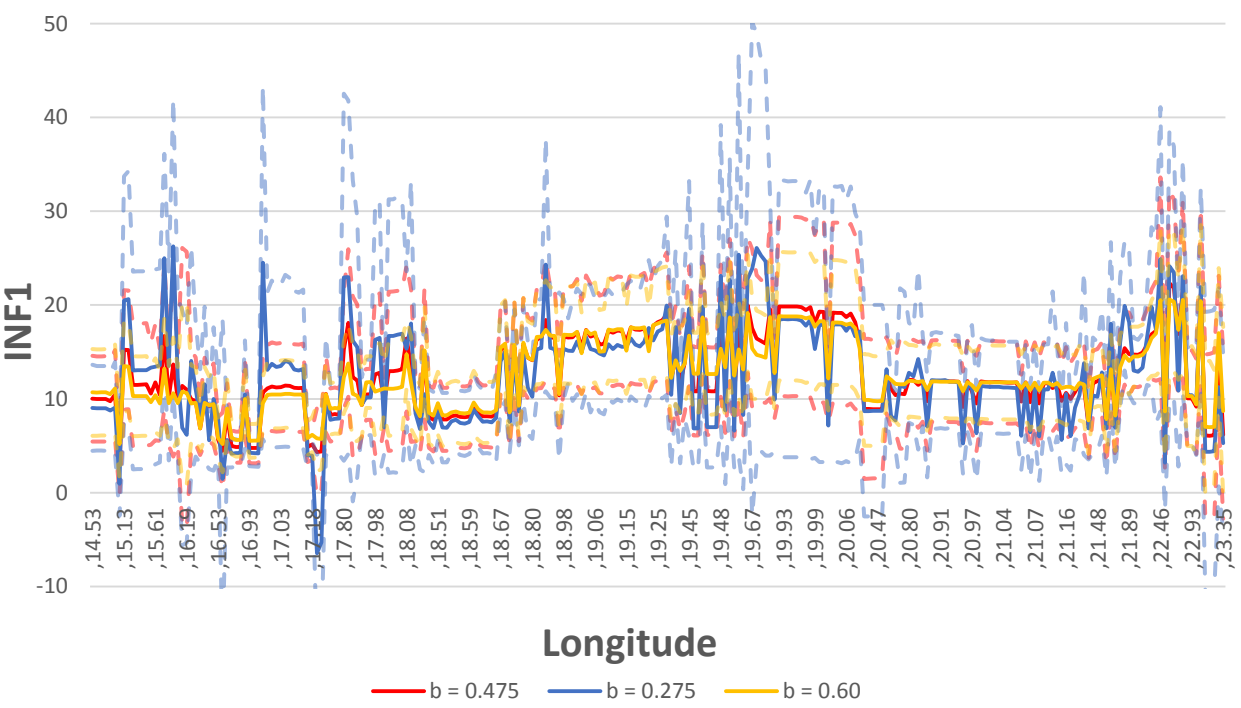
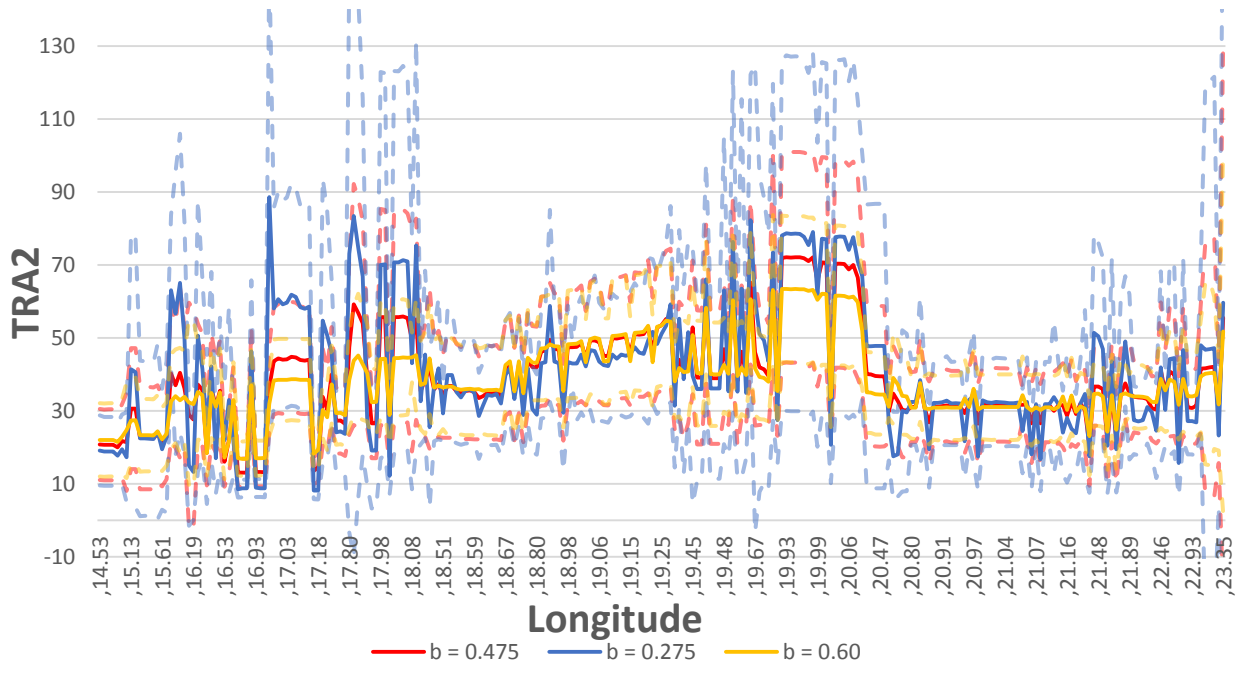
6

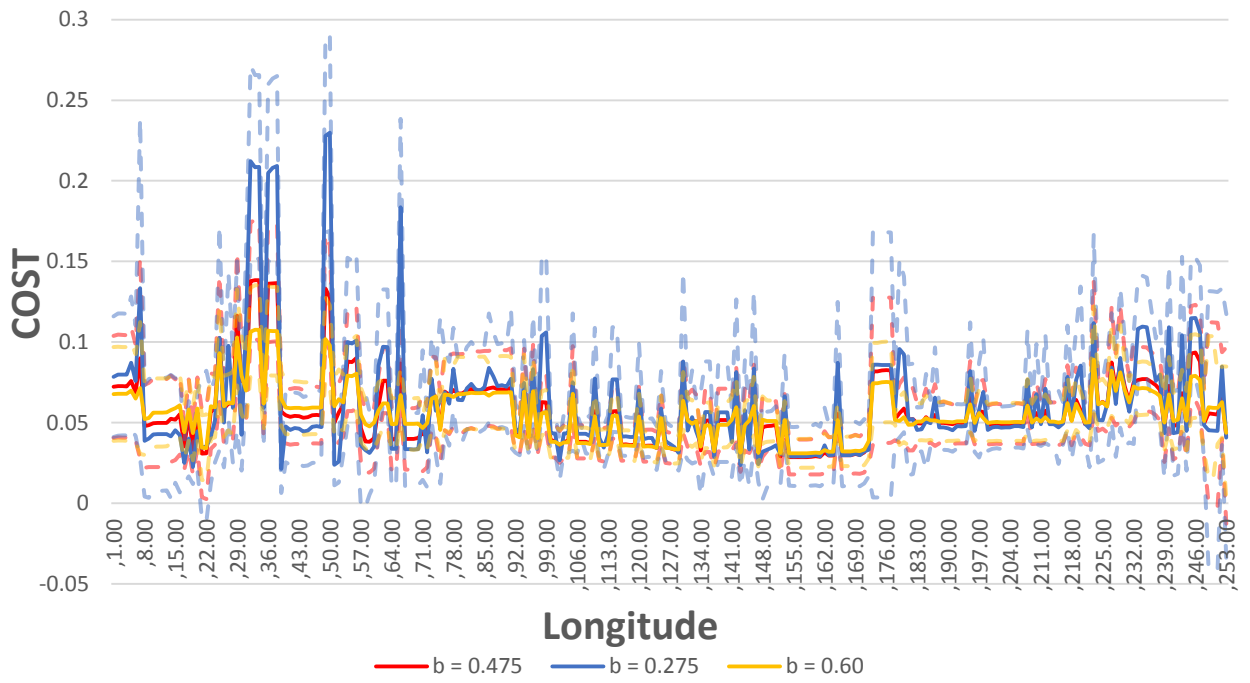
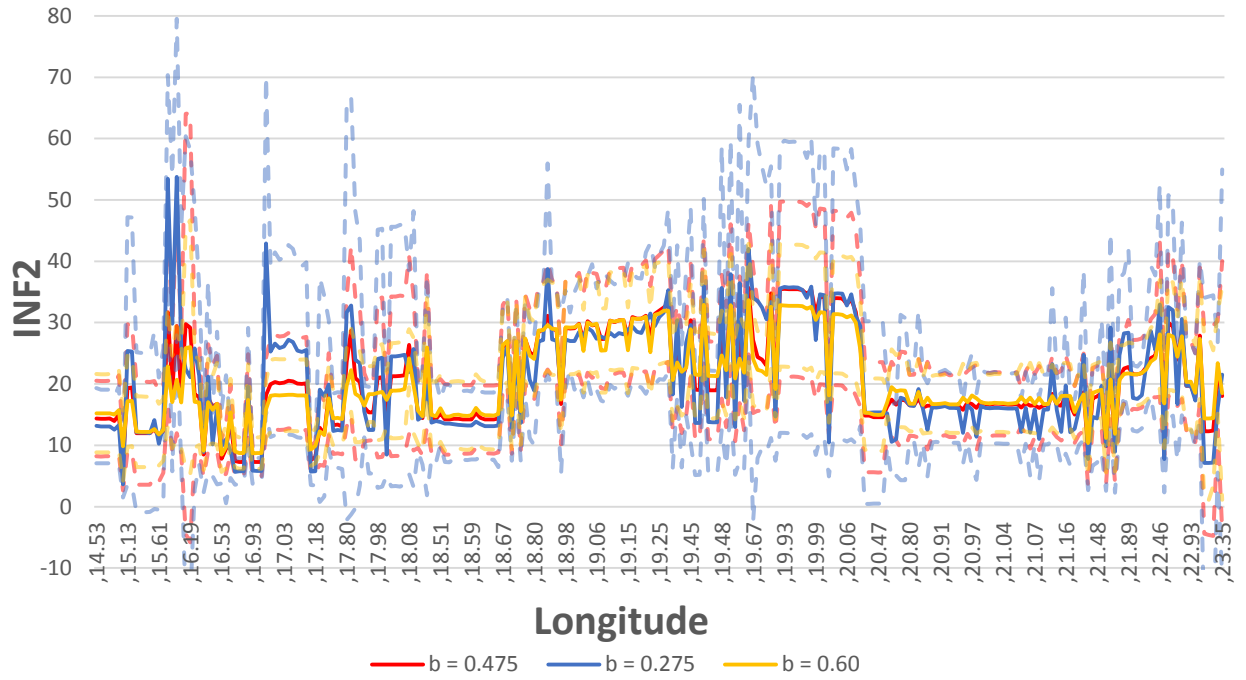
Figure A1. Parameters values with different bandwidths values plotted against longitude











Appendix B. Results of regressions where the dependent variables are the absolute values of differences between WTP from GWMNL and location-specific MXL.

	<b>SQ</b> (alternative specific constant for the no-choice alternative)	<b>NAT<sub>1</sub></b> (passive protection of most valuable forests – partial improvement)	<b>NAT<sub>2</sub></b> (passive protection of most valuable forests – substantial improvement)	<b>TRA<sub>1</sub></b> (the amount of litter in forests – partial improvement)	<b>TRA<sub>2</sub></b> (the amount of litter in forests – substantial improvement)	<b>INF<sub>1</sub></b> (tourist infrastructure – partial improvement)	<b>INF<sub>2</sub></b> (tourist infrastructure – substantial improvement)
Constant	82.0358*** (9.2459)	5.7755*** (1.5418)	9.6175*** (2.4648)	9.8597*** (2.6067)	12.7036*** (3.6717)	4.1844*** (1.0185)	8.8477*** (1.8850)
Area of coniferous forests	-0.4873** (0.1989)	0.0137 (0.0332)	0.0114 (0.0530)	0.0813 (0.0561)	0.1330* (0.0790)	0.0176 (0.0219)	0.0009 (0.0405)
Area of deciduous forests	-1.2194* (0.6669)	0.3164*** (0.1112)	0.4116** (0.1778)	0.2939 (0.1880)	0.3938 (0.2648)	0.1304* (0.0735)	0.2372* (0.1360)
Area of mixed forests	-0.6284 (0.4917)	-0.0237 (0.0820)	-0.0621 (0.1311)	-0.0474 (0.1386)	-0.0675 (0.1952)	0.0244 (0.0542)	-0.0575 (0.1002)
Area of forests with age >120	-4.2394** (2.1335)	-0.9866*** (0.3558)	-1.3444** (0.5688)	-2.4452*** (0.6015)	-3.1542*** (0.8473)	-1.1758*** (0.2350)	-2.1505*** (0.4350)
Average Euclidean distance to a forest	-11.0065*** (3.4348)	0.4652 (0.5728)	1.0371 (0.9157)	-0.1832 (0.9684)	0.3181 (1.3640)	0.1845 (0.3784)	-0.1190 (0.7003)
Built-up area	0.3118*** (0.1208)	0.0367* (0.0202)	0.0435 (0.0322)	0.1361*** (0.0341)	0.1378*** (0.0480)	0.0221* (0.0133)	0.0455* (0.0246)
Area of forests with no. of species > 6	0.1275 (0.4912)	0.1858** (0.0819)	0.2741** (0.1310)	0.4277*** (0.1385)	0.6317*** (0.1951)	0.1482*** (0.0541)	0.3786*** (0.1002)
No. of observations per location	-2.9135*** (0.7404)	-0.4663*** (0.1235)	-0.8032*** (0.1974)	-0.5535*** (0.2088)	-0.8810*** (0.2940)	-0.1038 (0.0816)	-0.3383** (0.1510)
<b>Model characteristics</b>							
R <sup>2</sup>	18.19%	18.04%	16.58%	20.07%	17.08%	14.49%	18.11%
n (observations)	253	253	253	253	253	253	253
k (parameters)	9	9	9	9	9	9	9

\*\*\* p-value < 1%, \*\* p-value in [1%,5%), \* p-value in [5%, 10%)