Using GIS for the social evaluation of rural natural environments

The Manual for the CLF-report (2013/26):

"Towards multifunctionality of rural natural environments? - A social evaluation of the extended buffer zones along Danish rivers, streams and lakes"

Angela Münch
Viktor Jozsef Racz

March 2014

Danish Centre for Rural Research and Development (CLF)
University of Southern Denmark

CLF Report 31/2014

The authors

Danish Centre for Rural Research
Department of Environmental and Business Economics
University of Southern Denmark
Niels Bohrs Vej 9-10
DK-6700 Esbjerg
Tel: 6550 4221
E-mail: clf@sam.sdu.dk

Contents

Figures

Tables

1. Introduction

In recent decades, there has been increased emphasis on the protection of the aquatic environment, and the Danish government has consistently implemented the EU regulations on water protection. One of the measures to improve the surface water quality is to dedicate buffer zones at rivers, streams and lakes, where farmers are not allowed to plant, grow or fertilize. In 2012, buffer zones of 10 meters were introduced around water bodies in Denmark in order to create a shield for Phosphor and Nitrate intake from agricultural zones.

There has been a significant debate about these buffer zones, and questions have been raised, particularly about the demonstrable environmental effects compared to the costs and production loss that farmers may experience. Questions have also been raised about other impacts such as social and recreational benefits for those who live nearby or use the, in principle, enlarged natural resources for recreational purposes. Up until now, the latter categories of questions have received more limited research attention, and there is not any solid knowledge about the impacts in either Denmark or internationally.

The purpose of this study is to initiate the research into the field of social benefits with Denmark as a case and example. More specifically, we are interested in researching whether: 1) The extended buffer zones add significant value in terms of open space for recreational use; and 2) The buffer zones are enhancing aesthetic values of nature/landscape for those who live nearby and hereby positively affect the property values.

The research results are already published in Münch et al. (2013). This paper is a report on how GIS (in particular ArcGIS) is used in this research. It should provide the reader with a step-by-step guideline on the usage of GIS for a hedonic price model.

To begin, a description of the implemented hedonic price model (chapter 2) as well as the GIS system in general (chapter 3) are discussed whereas in chapter 4, focus is turned towards the calculation of the several variables which are included in our analysis. In chapter 5, some features of the spatial statistics which are conducted in ArcGIS will be explained. The final chapter (chapter 6) briefly illustrates the way the data can be exported to be used in various other programs (e.g. R, STATA or Excel).

2. Environmental Valuation

2.1.　General

Assessing the economic value of the environmental goods and services is one of the major topics within the field of environmental economics. Due to the fact that most of these goods and services (e.g. biodiversity, scenic view, recreation facilities, or indirect processes e.g. water supply, ecosystem conservation etc.) have no observed market prices, several valuation techniques have been developed over the years in order to assess a price for environmental goods (Dixon, 2008). These different valuation techniques can be classified as either: (1) *stated preference methods,* which apply people´s responses to hypothetical questions related to environmental goods and services, or (2) *revealed preference methods,* which observe people´s actual behavior assuming utility maximization (Mäler and Vincent, 2005).

The stated preference approach encompasses such techniques as contingent valuation or choice experiment. These methods are widely applied although they also involve several weaknesses, e.g. answering on non-use values of an environmental good can be challenging for a respondent. The revealed preference approach (e.g. travel-cost or hedonic price methods), on the contrary, indirectly observes the market price of the good, which at the same time involves a bundle of characteristics that can be difficult to disentangle into the respective WTP per characteristic on its own versus the WTP for the collective bundle (Garrod and Willis, 1999). Hence, the feasibility of the revealed preference approach is determined by data availability and reliability. Moreover, the revealed preference method can only be applied to use values; hence it does not give a hint towards non-use or intrinsic values of nature protection measures in comparison to the stated preference method.

In spite of its weakness, we apply a revealed preference method, in particular the hedonic price method, in this study with the help of a Geographic Information System (GIS). Hereafter, we will introduce the hedonic price method (HPM) in more detail, since this technique will be used to study the benefit of the extended buffer zones for lakes and streams in Denmark, which were introduced in 2012. In other words, the HPM is applied in order to investigate if changes in natural amenities (natural improvements have been shown to occur owing to the existence/expansion of the buffer zones in Denmark) lead to changes in the willingness-to-pay (WTP) for real estates which are located close to these buffer zones. In our hypothesis, we assume a negative relationship between the prices of the properties and the distance of these properties from the protected streams and lakes. Thus, we predict that the farther away the real estate is from the buffer zone, the lower the sales price of the real estate (due to the decreased amenity values). As a result of this research, the social value of the improved environmental (amenity) characteristics –owing to the extended buffer zone – will be determined. Integrating spatial information with the help of GIS increases the accuracy of the calculation by incorporating not only house characteristics but also the surrounding of the house (Lake et al., 2000; Paterson and Boyle, 2002).

2.2.   The Hedonic Price Method

2.2.1. Literature Review

Using the hedonic price method (HPM), the values of environmental goods and services can be calculated by estimating the effect of changes in environmental amenities on e.g. real estate prices. In this case, the HPM – also called hedonic regression method – assumes that the price of a house is a function of the real estate´s underlying (e.g. structural, locational, neighborhood and environmental) characteristics. Therefore, the variations of the house transaction prices can be the result of the changes of the previously mentioned characteristics.

In an international context, the hedonic price approach has been widely applied and methodologically improved over the years in order to accommodate spatial effects and the non-linear function of house prices. In the last decades, intensive empirical research was conducted on the effect of close-by water streams or lakes on property values. Most of these studies differentiate between the aesthetic value, the recreational value and the value of improved water quality. In respect to the aesthetic value and recreational value, several studies have so far found that in general, view and proximity to lakes or rivers increases property values (e.g. Knetsch, 1964; d'Arge and Shogren, 1989; Kulshreshtha and Gillies, 1993). However, incorporating spatial information, Cho et al. (2011)could show that the effect of the aesthetic view on house prices only holds true if recreational and aesthetic value is generated by the water body at the same time; thus, the effect is dependent on the size and management of the water body (Netusil, 2005; Cho et al., 2011). Additionally, the direct access to the water body enhances the valuation of the water body (Lansford and Jones, 1995; Muller, 2009).

So far, most studies only consider the proximity or quality of water bodies and, in particular, lake or beach regions, but do not consider the value riparian buffer zones may create. One of the few studies which seeks to address this valuation created by riparian buffer zones is Mooney and Eisgruber (2001). Using market-assessed valuation data for single family residences and the proximity to riparian protection measures within the Mohawk watershed, western Oregon (USA), they estimated that although houses closer to water streams are on average 7 percent more highly valued, riparian buffer zones (on average ca. 9m wide) decrease the market value of the property by about 0.06 percent/foot. In other words, a buffer zone on the river which is 50ft. wide would reduce the market value of an average house by USD$ 4,650. They explain this result by the fact that these buffer zones are normally treed and therefore the visibility of the water stream is reduced. A second study which seeks to evaluate riparian buffer zones is Netusil (2006), who used sales prices for single-family residential properties close to the Fanno Creek Watershed in Portland, Oregon (USA). Differentiating between the kinds of wildlife habitat provided by the buffer zone and their riparian class, this study estimated a positive (decreasing) valuation for large forest patches, wetland areas, and large contiguous patches in uplands. The proximity to forest patches with low structure connector patches along streams and rivers, as well as semi-developed rivers accompanied by low structure vegetation and a forest canopy is, in

contrast, estimated to decrease the sales price of the property. Hence, the coverage of the riparian buffer zone in this study seems to determine the valuation of the buffer zone. Bin et al. (2009) compare the effect of the introduction of a mandatory riparian buffer zone with data for the Neuse River Basin in North Carolina (USA). By disentangling the valuation for the pre- and post-buffer zone riparian area, this study concluded that although riparian properties achieve a premium on the housing market, the mandatory buffer zone implemented in 1997 had no significant impact on the value of the property in the researched time period (1992-2002).

So far, all studies have been conducted for US cases, but no Danish application, which seeks to disentangle the Willingness-to-Pay for the proximity to a water body and the area surrounding the water (riparian buffers as well as buffer zones around lakes), is known to the authors. One of the few and the first applications of the hedonic price approach in Denmark according to Møller et al. (2000) is Hjorth-Andersen (1976) on the effect of noise emission on house prices. Regarding natural amenities, Præstholm et al. (2002) investigate the Danish house owners` willingness to pay for proximity to forests with the help of a hedonic price method. A more recent hedonic price study for the city of Aalborg and the effect of green space on house prices was conducted by Panduro and Veie (2013). By differentiating between types of green space, this study finds that green buffer areas as such are unattractive, while lake view and proximity to parks are increasing house prices in the area.

Another stream of the HPM literature follows the approach of Geoghegan et al. (1997), which focuses on the impact of the surrounding area (diversity, fragmentation) on the property sales price. Geoghegan et al. (1997) detect that individuals value the diversity and fragmentation of land use around their homes; open space in particular seems to be highly valued (Acharya and Bennett, 2001; Geoghegan, 2002).

The study is in the realm of research on riparian buffer zones, but also draws on the hedonic price approaches on the valuation of aesthetic and recreational values of water bodies, and at the same time considers the interfering effect of landscape fragmentation on the revealed values.

2.2.2. Theoretical Approach

The hedonic price approach draws on the theory of revealed preferences – revealed in the activity of buying e.g. a house. A house is in this sense a good with a bundle of characteristics (e.g. size, material quality, age, design). Besides these 'individual' characteristics of the house (*S*), additional surrounding factors may also support the decision of the individual to buy this specific house for this price. These additional surrounding factors may be (1) economic/social (*L*): e.g. positive: distance to place of employment, public transport facilities, and/or negative: closeness to street noise/emission, crime; (2) natural (*N*): e.g. positive: distance to beach and recreational areas, negative: closeness to waste site, windmills (not in my backyard); and (3) aesthetic value (*F*) which is created by the natural fragmentation of the surrounding area.

$$P = P(S, L, N, F) \tag{1}$$

As a house is a bundle of goods, the buyer always faces a trade-off between these different characteristics. The actual buying decision should reflect the preference ordering of the different characteristics within the price (i.e. WTP). The hedonic price method seeks to disentangle sales prices for houses on the market into WTPs for different characteristics of the houses as well as neighborhood characteristics. We use a simplistic version of the method by disentangling the prices of houses with the help of a linear regression model on the form (based on Gibbons et al., 2011):

$$P_{it} = \alpha + \beta_1 S_{it} + \beta_2 L_i + \beta_3 N_i + \beta_4 F_i + \beta_5 T + \varepsilon \qquad (2)$$

where the dependent variable (P_{it}) is the sales price of house i at the t time of sale, while S_{it} is the vector of the structural characteristics of house i at the time of sale (e.g. building size, number of rooms, number of bathrooms, age); L_i indicates the locational characteristics of house i, such as distance to economic variables – e.g. transport infrastructure (public transport, motorway etc.), distance to service provisions – e.g. hospitals and schools, and distance to a labor market, etc.; N_i denotes the vector of the neighborhood characteristics, in our case: the natural amenities (e.g. proximity to recreational facilities and buffer zones); while F_i captures the aesthetic value derived from the natural fragmentation around the house. T is a time dummy that captures exogenous, unobserved developments in the housing markets in the respective year of sale, and β_1, β_2, β_3 and β_4 are the coefficients for the structural, locational, natural and aesthetic attributes. Finally, the unobserved components are included in the error term ε.

Regarding spatial autocorrelation, which might bias the estimation results (Anselin, 2001), examining our dataset the test for global spatial correlation (in detail: Moran's I & Geary's C, see following chapters) displayed significant results. Hence, the sales price of house i is influenced by the sales price of the neighboring house j. Given these high z-scores of Moran's I, the likelihood that this clustered pattern could be the result of random chance is less than one percent. Thus, the above mentioned simplistic hedonic price approach needs to be extended towards a spatial regression model and should also incorporate the natural logarithm of the price function.

Following Elhorst (2010), further tests were undertaken to specify which spatial model would best reflect the detected spatial correlation in our sample data with the help of STATA 12 (not further described here). The spatial Durbin model (see below) was found to capture best the nature of the spatial autocorrelation. This model is computed as follows:

$$Y = \rho WY + \alpha \tau_N + X\beta + WX\theta + \varepsilon \qquad (3)$$

where Y denotes the dependent variable (i.e. natural log of sales price), WY is the spatial lag of the dependent variable, ρ is the spatial autoregressive coefficient, τ_N refers to the constant term,

$α,β$ are the associated estimated coefficient vectors, and X denotes the exogenous independent variable. For simplicity in display, X here is the combined vector of the above mentioned independent vectors S_{it}, L_i, N_i, F_i and T. WX are the spatially lagged independent variables, $θ$ the fixed but unknown parameters, and $ε$ the error term.

In order to estimate the effect of the introduction of the buffer zone on house prices, two additional dummy variables are implemented: (1) D^r shall capture the time-invariant effect of having a water stream on the property, and (2) $D^b_{>2008}$ shall capture the effect of the introduction of the buffer zone if the buffer zone is on the lot and the property was sold after 2008 (Bin et al., 2009). For the two variables it is necessary to indicate whether a water body is on the property.

The spatial Durbin model incorporates a spatial weighting matrix on the dependent and independent variables. An inverse distance matrix serves as spatial weight matrix (W) (i.e. $W = 1/d_{ij}$ with d as distance between house i and house j), which is row standardized (i.e. takes the interval $(w^2_{min}, 1)$). Thus, it is assumed that the price of house i has an impact on the price of house j, and the other way around. This impact is declining linearly with increasing distance between houses i and j.

In the following section, the so-called geographic information system (GIS) as used with the software ArcGIS 10.1 will be shortly introduced, since this tool will be used for the provision of the above described different spatial variables and the weighting matrix necessary for calculating the hedonic price function (equation 3).

3. Geographic Information System (GIS)

A Geographic Information System (GIS) is defined as "*a system for capturing, storing, checking, integrating, manipulating, analyzing, and displaying data that are spatially referenced to the earth*" (Department of the Environment 1987, according to Bateman et al., 2002). The crucial part of the definition is the phrase "spatially referenced to the earth" as it emphasizes that the location-dependent features are part of the analysis. Thus, spatial effects are included into the analysis. Spatial effects are commonly divided into spatial dependence and spatial heterogeneity. Spatial dependence refers to the relationship between spatially referenced data in the meaning that the structure of the correlation or covariance between random variables at different locations is derived from a specific ordering, determined by the relative position (distance, spatial arrangement) of the observations in geographic space (or, in general, in network space). Spatial heterogeneity occurs when there is a lack of spatial uniformity of the effects of spatial dependence and/or of the relationships between the variables under study. Spatial heterogeneity is often difficult to separate from spatial dependence. As cross-sectional data is applied, it is also often only possible to identify clusters and spatial patterns but not the processes that led to the patterns. Although, patterns can be theoretically created by structural change (apparent contagion) or follows from a true contagious process; an empirical differentiation between the two processes is so far not feasible (Fotheringham and Rogerson, 2009; Anselin, 2010; Anselin and Rey, 2010). Hence, sound theory and data testing is necessary in order to specify the model correctly.

Another way of theoretically conceptualizing spatial dependence and heterogeneity is to differentiate between a spillover model and a resource flow model. In the spillover model, it is assumed that an agent chooses the level of a decision variable by considering the values of the same variable chosen by other agents close by. In contrast, in the resource flow model, the agent's decision variable is not directly affected by the levels chosen by other agents, but only indirectly. The indirect effect follows from the presence of the value of a "resource" in the individual agent's objective function. Therefore, socio-economic decisions might be influenced by e.g. nearby amenities or the proximity of a pollution source (Anselin, 2002; Bateman et al., 2002).

In this study, methods developed in spatial econometrics are applied, i.e. spatial dependence is controlled for but not further explored as in spatial planning literature (Maguire et al., 2005). In the latter stream of literature, GIS is not only a tool but also a mean and is defined software centered as "*an integrated software system for the collection, storage, query, analysis, and presentation of geographic information.*" (Maguire et al., 2005).

The key parts of GIS as a software system (Figure 1) are:

1. The *information model,* which defines the basic design pattern applied to describe the real world and its abstraction within the computer system (e.g. network, raster/grid, vector topology).

2. In the *data manager* information is organized mainly within a database management system (DBMS) table. Thus, objects are collected and connected with properties/class templates.

3. The *process functions* are the tools within GIS to process the information. Hence, these functions are at the core of the definition of GIS as given before by Bateman et al. (2002). The system around the process function is not further considered in spatial econometrics but is a special interest for spatial planning/modelling.

4. Interfaces are differentiated into *user interface* and *application programming interface*. While the former one enables the end users to interact with the system via graphical and command line interfaces, the latter one targets the interaction of the developer with the underlying component building blocks.
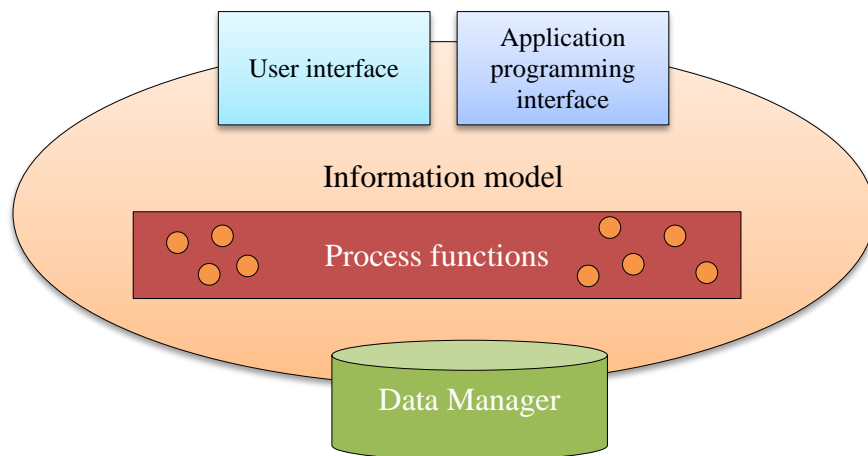
Figure 1: Key parts of GIS as software system
(Source: Maguire, 2005, p. 21)

Due to increased electronic data availability and enhanced GIS software, methods of spatial econometrics enjoy rising popularity within in the scientific community (Anselin and Rey, 2010). In particular, the use of spatial regression for the hedonic price method is nothing new and already advertised by e.g. Lake et al. (2000), Paterson and Boyle (2002) or Osland (2010). However, applications are still not very common although standard regression methods fail to predict correct standard errors if spatial autocorrelation and/or spatial heterogeneity exist in the data set (Anselin, 1990, 1999; Corrado and Fingleton, 2012).

The manual at hand seeks to show how GIS, in particular the software ArcGIS 10.1, can be used to support spatial econometrics in respect to the hedonic price method. Hence, both approaches of spatial information shall be here combined and technicalities explained in more detail. While the data manager and the interfaces are given by software ArcGIS 10.1, the information model and the process function are up to the user and are at the core of this manual. Regarding the interfaces, the manual will be restricted to the user interface to ensure readability. Please note that the terms ArcMap and ArcGIS are used interchangeably here. Commands as stated below, however, refer to the path within the ArcGIS catalogue instead of the ArcMap toolbar.

4. Applying ArcGIS as a Tool for Data Attainment

4.1. Data Preparation

As written in section 2.2.2, in order to perform a hedonic price regression, housing data and data on the surrounding area have to be incorporated into ArcGIS 10.1. Data on the structural characteristics of the houses (S) are provided by KMD A/S. The dataset contains data on house characteristics but also on sales of the houses in the last years. We limited the data with the help of Microsoft Excel to detached single-family houses sold between 1996-2013, hereby excluding holiday and farm houses as well as multi-family dwellings. All our houses (sample size: 4002) are located in the postal code areas: Skjern (Postal code 6880), Tarm (Postal code 6900), and Ribe (Postal code 6760). The advantage of the database is that it also provides the XY - coordinates of the respective house (i.e. longitude and latitude); hence, the geographic location is identified and a connection with the plot belonging to the house is possible.

4.1.1. Add Data

As the first step, this structural housing data needs to be transferred from Microsoft Excel to ArcMap 10.1. After using the command `Add data´ in ArcMap 10.1 and choosing the Excel-file, the XY-coordinates of the houses need to be specified further to display the data within ArcGIS. With a right click on the layer which contains the housing data, one can select the command `Display XY data`. The window 'Display XY Data' will be shown, in which the fields (columns in the excel file) for the X and Y coordinate (also for Z, if available) can be specified. Here, the projection can be set as well (Figure 2). In the X Field, choose the longitude field from the Excel-file and the in the Y Field choose the latitude field. The Z-field would be the altitude, if available. In this analysis, we use the ETRS (1989) – UTM Zone 32N as a projected coordinate system, which was also specified here for the XY-data with the help of the command 'Edit…'.

In order to enable calculation with the data, the newly created data layer needs to be transformed into a shapefile (see section 6.1).
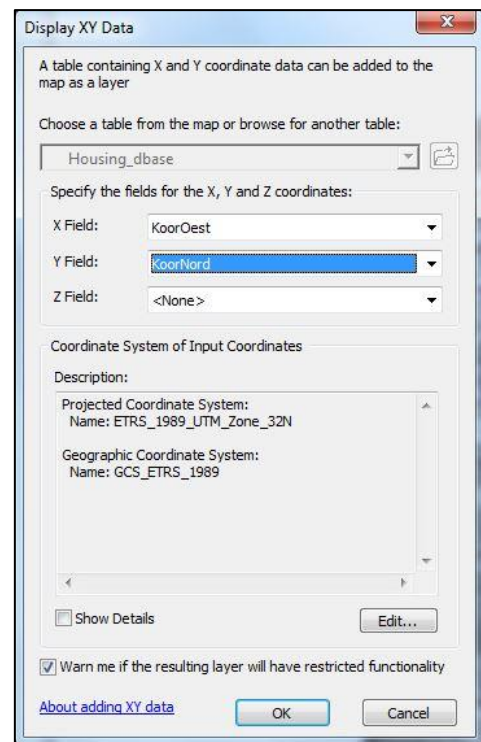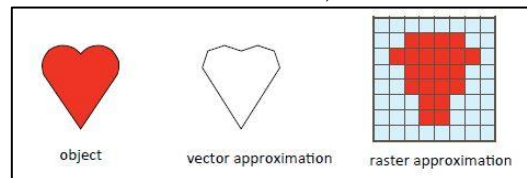

Figure 2: Display XY-Data

Additional shapefiles are provided by Geodatastyrelsen, NaturErhvervstyrelsen and Miljøpor-talen, which contain the geographic location of the incorporated variables referring to N and L (see also section 2.2.2):

L_i	N_i
City center *(Source: Geodatastyrelsen)*	Water streams & lakes *(Source: Geodatastyrelsen)*
Streets *(Source: Geodatastyrelsen)*	10m buffer zone *(Source: NaturErhvervstyrelsen)*
Parking *(Source: Geodatastyrelsen)*	Beach *(Source: Geodatastyrelsen)*
Property plots *(Source: Geodatastyrelsen)*	Forests*(Source: Geodatastyrelsen)*
	Farmland *(Source: Miljøportalen)*
	Windmills *(Source: Miljøportalen)*

Table 1: Shapefiles in the Analysis

This data can just simply be added into an ArcGIS map document without any issues if the projected coordinate systems match. Should the data be displayed in another coordinate system, one can easily transform these under: Toolboxes\System Toolboxes\Data Management Tools.tbx\Projections and Transformations. However, it is easier to define the coordinate system of the map document before adding the layers. By adding the data layers, ArcGIS already standardly transforms the data to match the pre-defined coordinate system or formerly imported data.

Please note: ArcGIS uses two basic data models, namely vector and raster. Vector maps symbolize geographic phenomena with points (e.g. tree, house), line (e.g. road, river) and polygons (e.g. building plot, lake). The location of a point is given by its X and Y coordinates, while lines and polygons are their sets of coordinate pairs (in the case of a polygon, these coordinates describe the boundary of the enclosed area). In contrast, raster is a matrix of cells in which the geographic phenomena are projected. The location of the cells is determined by the row and column locations of the cells within the matrix. An example of the difference between the vector and

Figure 3: Difference between vector and raster models
(Source: Mamoulis, 2011, p. 3.)

raster model projections is shown in Figure 3. More details about the vector and raster models can be found on the ArcGIS Desktop help site.

In the analysis described here, only vector data is used. Although this kind of projection allows for a more accurate calculation, it is also data and calculation capacity intense. Moreover, it does not resolve any kind of issues related to the sample strategy or in respect to the Modifiable Areal Unit Problem (MAUP) (Anselin and Getis, 2010; Anselin and Rey, 2010).

4.1.2. Data Restriction (Select by Location)

All these shapefiles contain data for the geographic scope of all of Denmark (see e.g Figure 4 in which the geographic location of the buffer zones and houses in the sample are displayed). We are only interested in the aforementioned postal code areas. Using country wide maps in the calculations slows down the calculation processes, especially when the scope of the project is as narrow as it is in our project case. Thus, restricting the information on the study area will make the work more efficient.
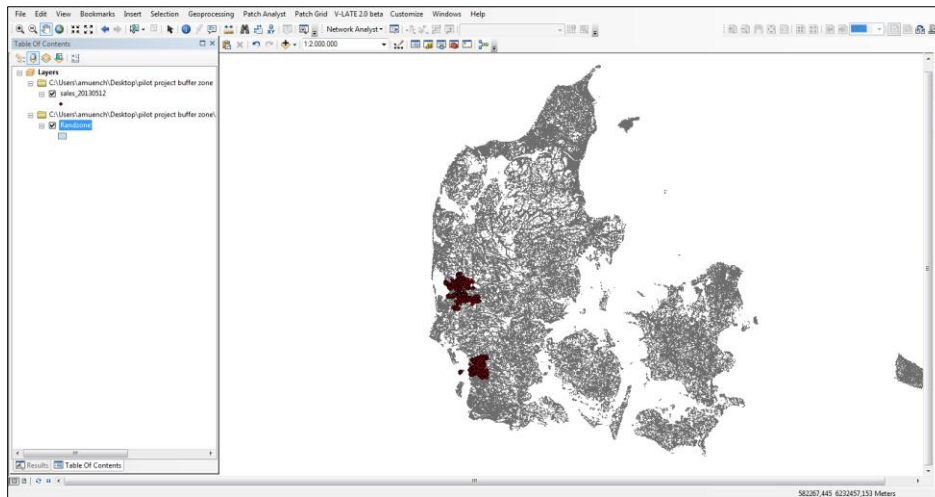
Figure 4: Added Shapefiles (e.g. buffer zone map and housing map)

To select a geographic region or a part of a map, one should follow the next steps:

1. Right click on the name of the layer which is to be reduced followed by the command 'Selection' and the command 'Make This the Only Selectable Layer'
2. Select the command 'Selection' in the toolbar, choose one of the options (e.g. rectangle, see Figure 5) and highlight the area which should be selected (i.e. the study area) by clicking (left mouse button) in the map.

Figure 5: Selection options

3. The study area should now be highlighted or marked differently. With a right click of the mouse on the layer name in the Table of Contents, again, and choosing the command 'Selection', one should now be able to choose the command 'Create Layer From Selected Features` which creates a layer from the chosen area and restricts the geographic extension to the information needed for the analysis. An example for selecting the sample area from extended buffer zone map can be seen in Figure 6.
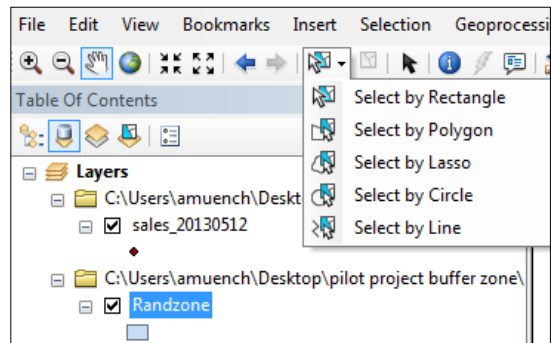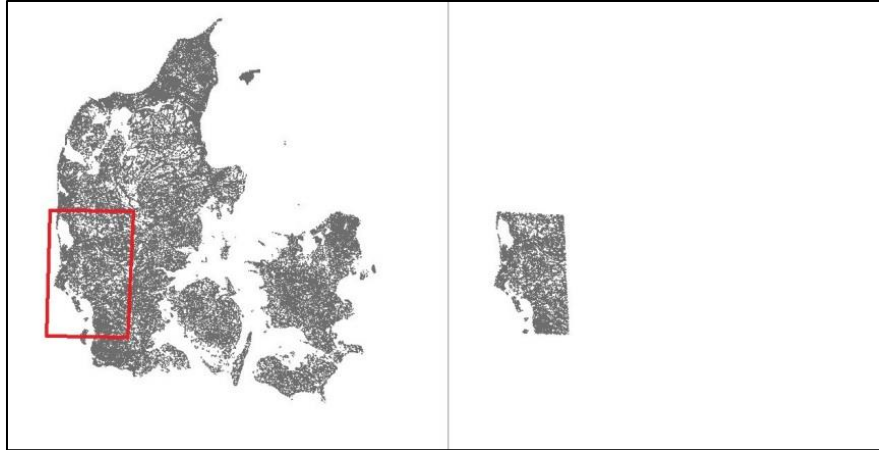
Figure 6: Example of Select by location (e.g. buffer zones in Ringkøbing)

4.1.3. Data Filtering (Select by Attribute)

Besides restricting the sample area, one might also be interested in the effect of the special feature of one layer. So, e.g. the shapefile 'streets' contains all kinds of streets (e.g. foot paths as well as highways), but one can assume that the effect of a public foot path next to a lot will be different than the adjacency of a highway. The same holds true for the shapefile 'rivers' which e.g., also contains covered water streams, i.e., water bodies which are not used for recreational activities. Moreover, it can be assumed that a big river has different impact on house prices in the neighborhood than a small creek. Therefore, the data within the layer should be filtered or selected by characteristic (compared to above: selected by location). One way of doing this is:
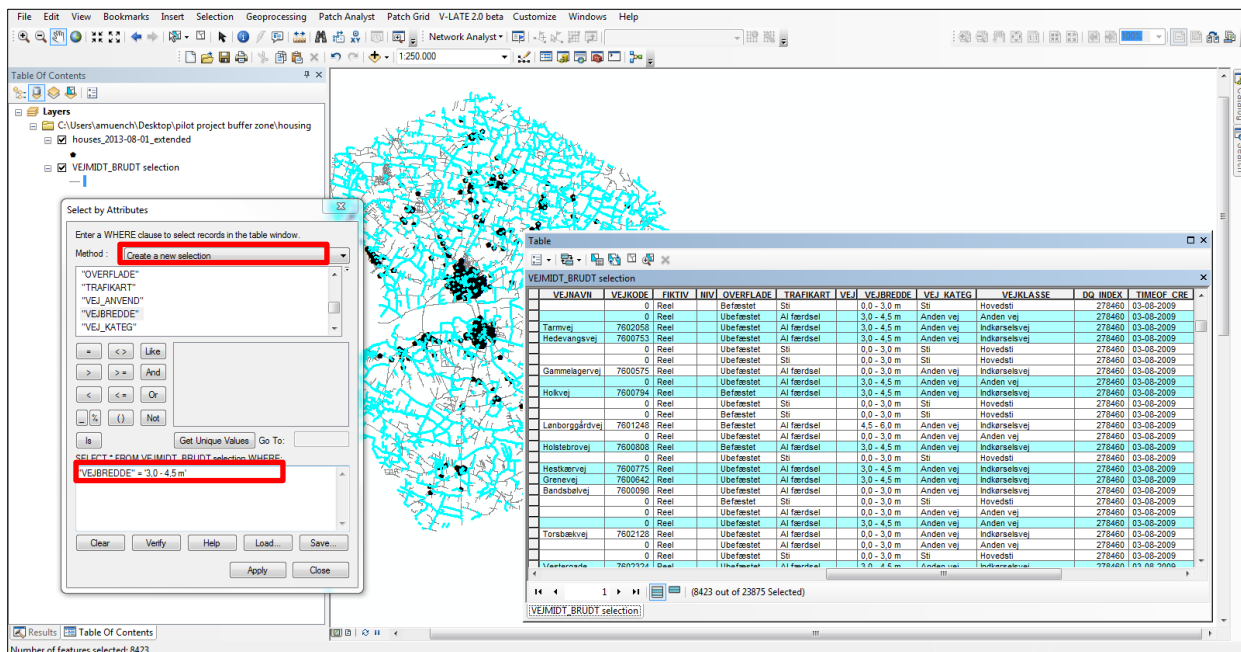

Figure 7: Select by attribute (e.g. street width in Ringkøbing)

16

1. Open the attribute table: right mouse click on the layer where data should be filtered and choose the command 'Open Attribute Table'
2. Within the toolbar of the attribute table choose 'Select by attribute'. In the window 'Select by Attribute' choose the attribute which shall characterize the selection (e.g. street width) and define the selection criteria (e.g. "VEJBREDDE" = '3,0 - 4,5 m'). After, press 'Apply' and the streets which fulfill the criteria will be highlighted (see Figure 7).
3. As above, a layer of the selection is created by using a right click of the mouse on the layer name in the Table of Contents, and choosing the command 'Selection' and then the command 'Create Layer From Selected Features`.

In our example, there should now be a specific layer for streets with a width between 3.0 and 4.5m. However, there are streets which are wider (4.5 – 6.0m). By choosing in the window 'Select by Attribute' the method 'Add to current selection' one can easily extend the selection of the attributes, e.g. accommodate in the selection wider streets also (see Figure 8).
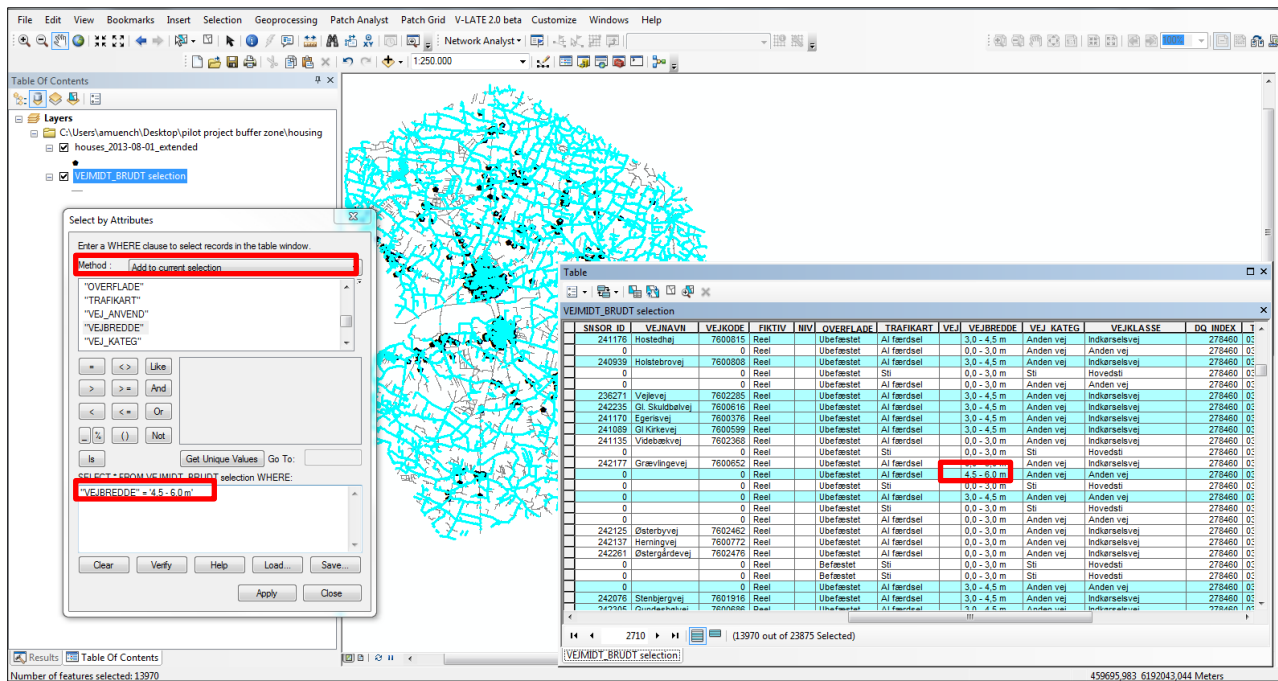

Figure 8: Select by attribute (e.g. extended street width in Ringkøbing)

4.2.   Distance Calculations

In our theoretical model (section 2.2.2), we consider distances from the house to the buffer zone as well as to several other variables (L_i, N_i), which can be significant for the buyer's decision, and then reflect the WTP for the proximity to this feature. To measure the distance, we used in this project two different approaches: (1) the Euclidean distance and (2) the travel distance (e.g. via road, walking path).

17

4.2.1. Calculate Euclidean Distances

Most of our variables describing L_i and N_i are polyline or polygon data which surround the point data (house). Hence, we are interested in the direct distance between our houses in the sample and the natural or economic features, irrespective of the access to this feature. Any barriers to the points of interest will be therefore ignored. Moreover, we only calculate the distance to the boundary of a polygon feature, not to the center or centroid of the polygon.

In order to calculate the distance to the nearest feature of the variable L_i / N_i for each observation point (house), the following steps are necessary:

1. After adding the relevant shapefiles and the housing data, chose within the 'System Toolbox' of ArcGis under 'Analysis Tools' the category 'Proximity'. There, one can find the command 'Generate Near Table' (see Figure 9). Using the command allows one to define more than only the closest feature of the layer, but also the second or third closest. Additionally, results are saved in an extra layer, which can easily be exported and/or joined to the original database.
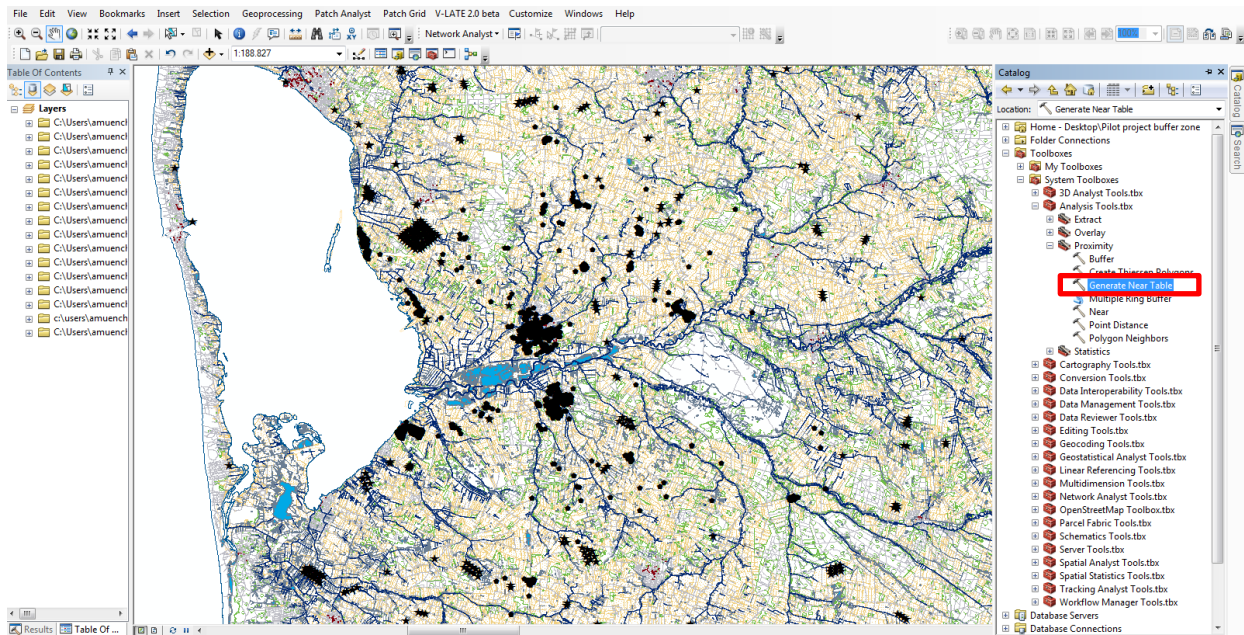

Figure 9: Command 'Generate Near Table'

2. In the window 'Generate Near Table' (see Figure 10) the following features need to be defined:
    a. Input feature: layer of houses.
    b. Near feature: layer of L_i or N_i. Please note each variable should be considered separately as otherwise only the distance to the nearest of all feature will be calculated.
    c. Output table: defines the name and (file) location of the generated table.

18

d. Search radius: can be used if the distance to the feature should be restricted to a defined radius. Hence, one can assume that the proximity to a forest does not create any positive externality after a specific distance; i.e. the distance after the threshold is not important and therefore can be neglected in the calculation.

e. Location: if this box is ticked, not only the distance towards the feature is written into the output table, but also the coordinates (X and Y) of the feature which was calculated to be the closest to the house.

f. Angle: if this box is ticked, a near angle measures from the x-axis (horizontal axis) to the direction of the line connecting the house to the nearest feature within the range of 0 to 180 or 0 to -180 decimal degrees - 0 to the east, 90 to the north, 180 (-180°) to the west, and -90 to the south- will be written in the output table.

g. Closest feature vs. number of closest match: if the box 'find only closest feature' is not ticked one can also specify here the number of closest features to be found (e.g. the two nearest forests). The latter can be interesting if one is interested to know how many of the same feature is in a given radius of the house.
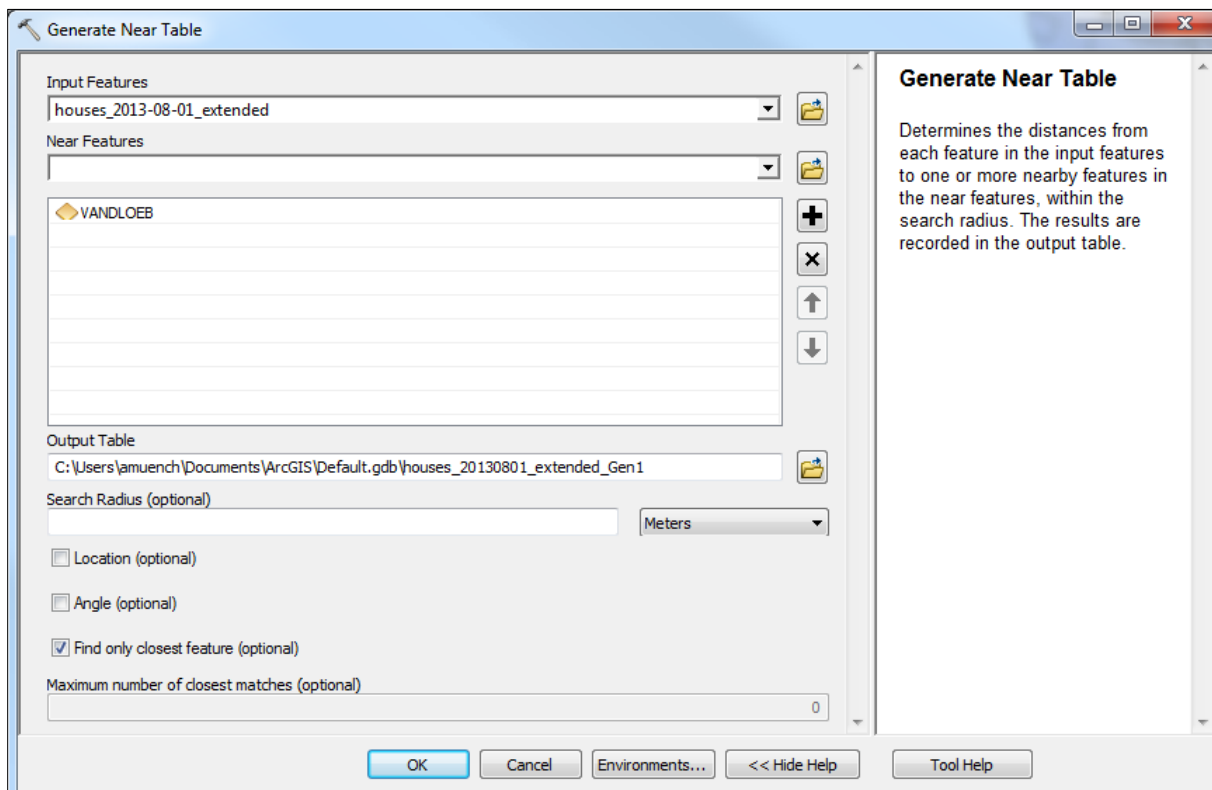

Figure 10: Window 'Generate Near Table'

Using the command 'Generate Near Table' instead of the command 'Near' allows one to define not only the closest feature of the layer, but also the second or third closest. Additionally, results are saved in an extra table layer that can easily be exported (see section 6.2) and/or joined to the original database.

4.2.2. Calculate Road Distances

Besides the direct distance, one major feature of the buffer zone regulation is, in general, that people are allowed to access this area if there is a public path into the buffer zone. So, it can be that, e.g. two houses (A;B) are equidistant to the buffer zone regarding the linear distance (as measured above). However, taking the road for the owner of house A will take more time to access the buffer zone than the owner of house B due to the lack of direct road access to the buffer zone via public streets. This example is illustrated in Figure 11. While house owner A needs to take the detour north which eventually leads to the road with access to the buffer zone, house owner B is located directly on a road with access to the buffer zone.
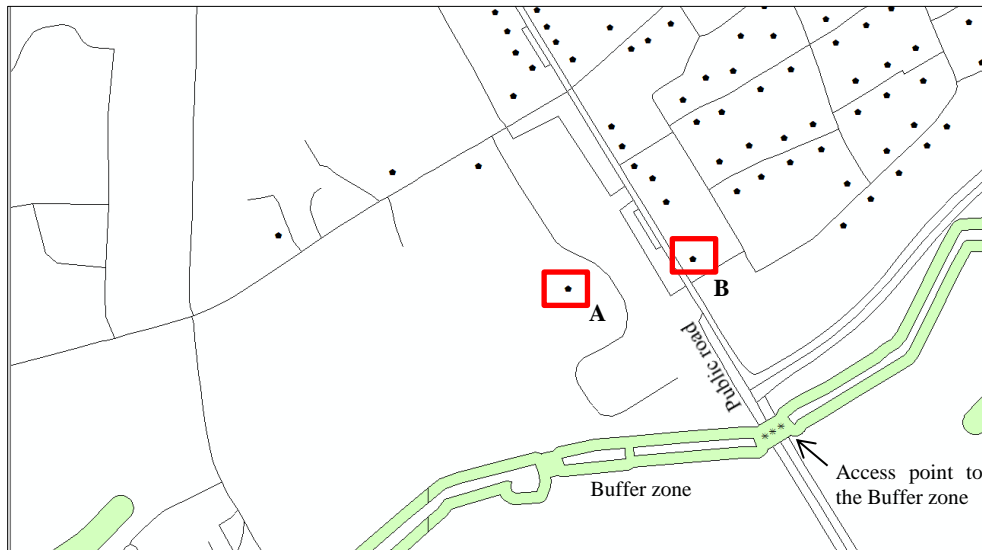
Figure 11: Example road access buffer zone Skjern
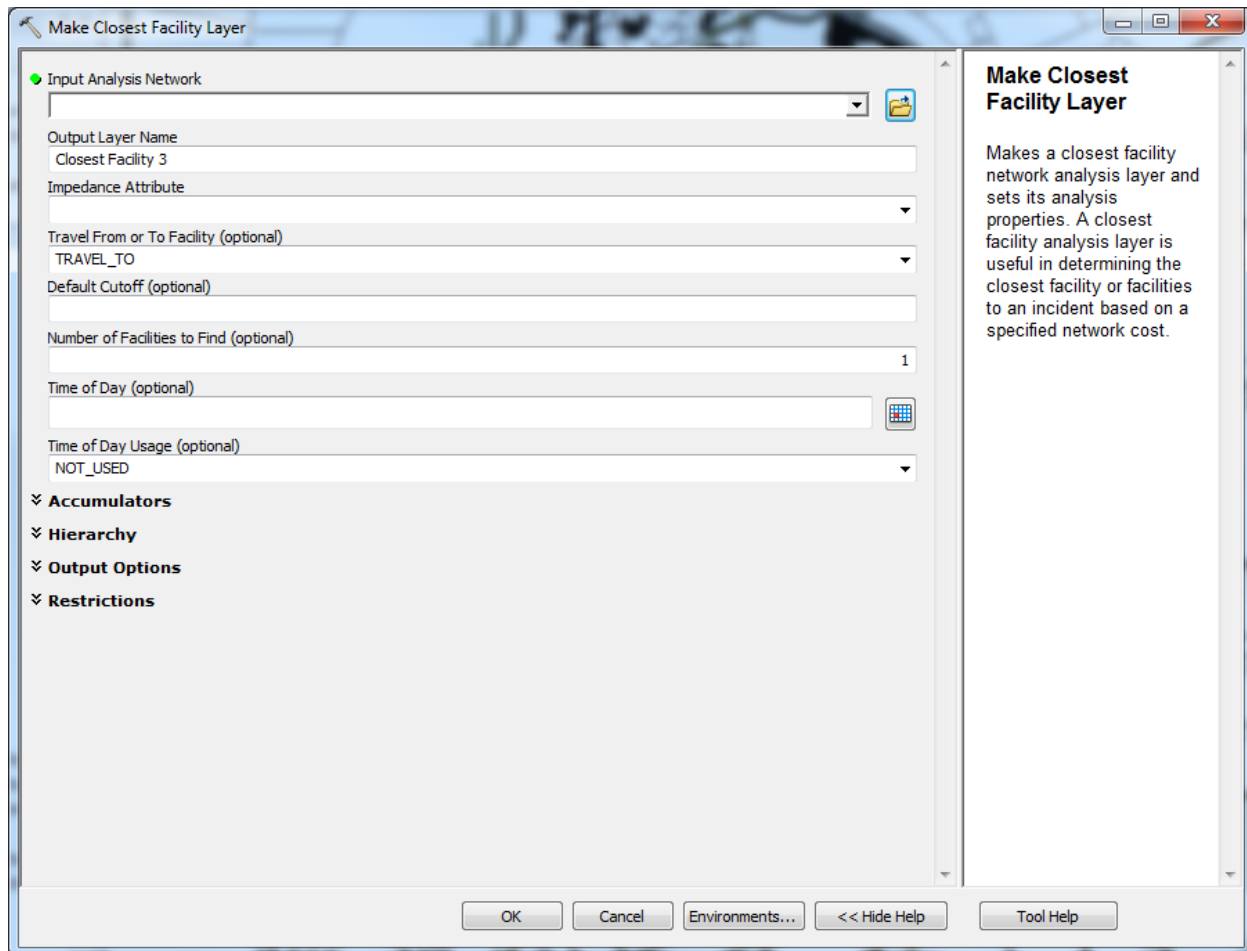
Create a Network Dataset

In order to integrate this difference due to access, a network calculation for the road distance is calculated. For this calculation a *Network Dataset* is necessary. To create such a network dataset, a polyline shapefile is necessary. We use the above mentioned shapefile on the road network (including foot path) provided by Geodatastyrelsen. With a right mouse click in the GIS catalogue in the folder connections on the shapefile 'roads', one can select the command 'New Network Dataset …'. After entering a name for the new network dataset, in the next steps one can define whether turns are allowed, or special requirements regarding the connectivity between the polylines. Additionally, the variable which defines elevation within the road network and any other requirements/specific attributes one may want to add as a cost barrier into the network layer are defined in this step as well.

In our analysis, we defined that no turns are allowed, but also did not give any driving directions. Neither elevation nor any other cost changing features are defined. Hence, the length of the path is the only cost attribute in the calculation.

20

Create a Closest Facility Layer

ArcGIS offers different options within the network analyst tool. Besides, for example, the route solver (to identify the quickest or shortest route), one can also can find the service areas around any location in a network (i.e. zone of attraction) or find the closest facility of interest. The latter one is used in this analysis in order to obtain the distance between the houses and the access points to the buffer zones. Within the 'System Toolboxes' in the category 'Network Analyst Tools' in the folder 'Analysis' the command 'Make Closest Facility Layer' can be found. In the window that opens, called 'Make Closest Facility Layer' (see Figure 12), several options should be further specified:

Figure 12: Window 'Make Closest Facility Layer'

1. Input Analysis Network: this field requires the path of the above created network dataset.
2. Output Layer Name: name of the newly created layer.
3. Impedance Attribute: length (as predefined in the network dataset).
4. Travel from or to Facility: specifies the direction of the travel path.
5. Default Cutoff: should be defined if a specific radius in search of the closest facility should not be exceeded.

21

6. Number of Facilities to Find: we specified here '1' as we are only interested in the closest access point to the buffer zone. However, one could also implement higher numbers (i.e. second closest and third closest facility).
7. Time of Day: can be specified – only makes sense if the network dataset contains information on route utilization at specific times.
8. Time of Day usage: here the same applies as in the field 'Time of Day'. If no time information is included in the network dataset, this field can be left as 'not_used'.
9. Restrictions: here one can again define if turns are allowed.

After obtaining the closest facility layer from the network dataset, we also need to define the points of origin and the points of destination. The *point of origin* will be defined by our point layer set of the houses (as created in section 4.1.1). For the *point of destination,* we need to determine the access points to the buffer zone. A simplistic approach is to intersect the buffer zone layer with the road map layer.

Intersect Data Layers

Within the 'System Toolboxes' under the category 'Analysis Tools' in the Folder 'Overlay', one can find the command 'Intersect'. By selecting this command, a window is opened (see Figure 13). In our case the 'Input Features' are the shapefiles 'roads' and 'buffer zone'. The field 'Output Feature Class' is the name of the created intersection layer. In order to reduce the size of the newly created shapefile and with this the calculation time, one can also choose to restrict the 'Join Attributes' to 'only FID'. As we use the command 'intersect' to create a point layer, which should serve as point of destiny layer later on, we need to define the 'Output Type' as 'Point'.
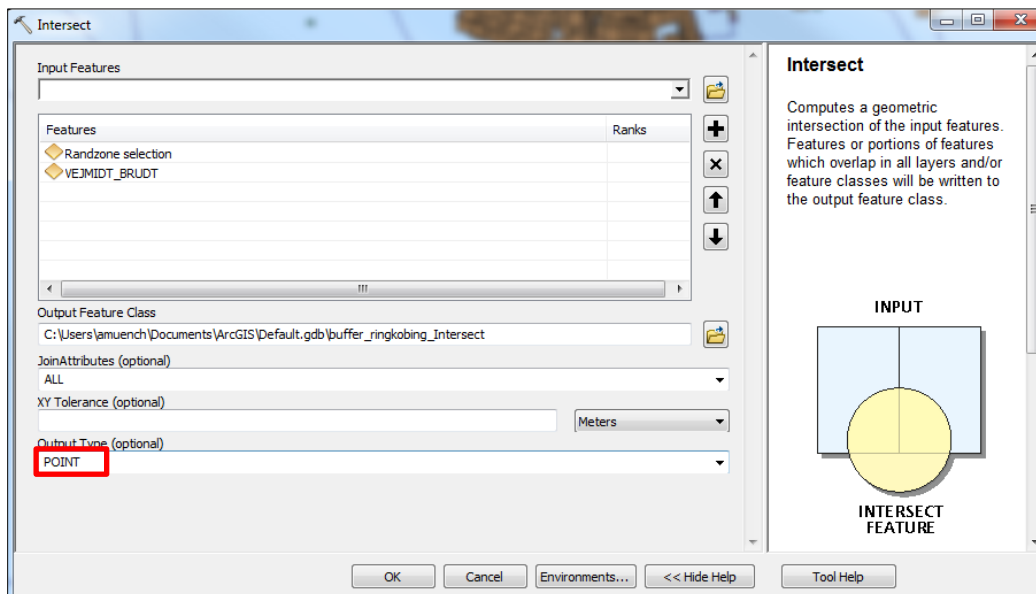
Figure 13: Window 'Intersect'

As the features in the layer 'street' are in the format of polylines and the features of the layer 'buffer zone' are in the format of polygons, a multipoint layer will be created instead of a point

feature class. In order to be used within the network analyst tool, this multipoint format needs to be redefined into point data. Therefore, we use the command 'Feature To Point' within the 'System Toolboxes' in the folder 'Data Management Tools' in the category 'Features'. In the opening window, the only items to define are the 'Input Features', which is the above created layer, and the 'Output Feature Class', which is the name of the transformed layer which shall be created (see Figure 14).
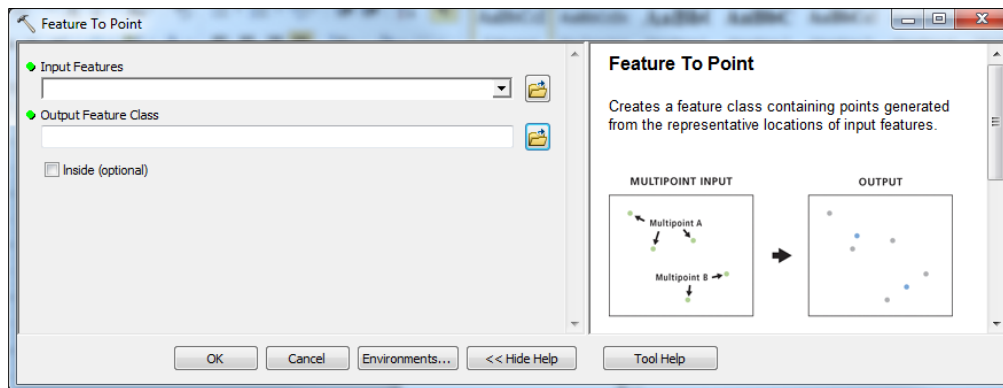

Figure 14: Transform Multipoint into Point Features

Add location

Upon opening the attribute table (described above) of the 'Closest Facility' layer, it becomes obvious that the attribution tables are empty due to the fact that we still need to specify our point of origin and our point of destination. In order to proceed, we need to 'Add Locations' (can be found under System Toolboxes\Network Analyst Tools\Analysis\Add Locations). In the window 'Add Locations' (see Figure 15) we need to specify:

1. Input network analysis layer: in our case, the closest facility layer created before.
2. Sub layer: facilities are the points of destination, i.e. the access points to the buffer zone; incidents are the points of origin, i.e. the housing data.
3. Search tolerance: specifies the tolerance between the road and the facilities/incidents. Therefore, e.g. houses are seldom directly at the road, but rather back some meters on the lot. The here defined tolerance considers this deviation. Please note: the higher the tolerance is defined here, the less accurate the results are for the route length.
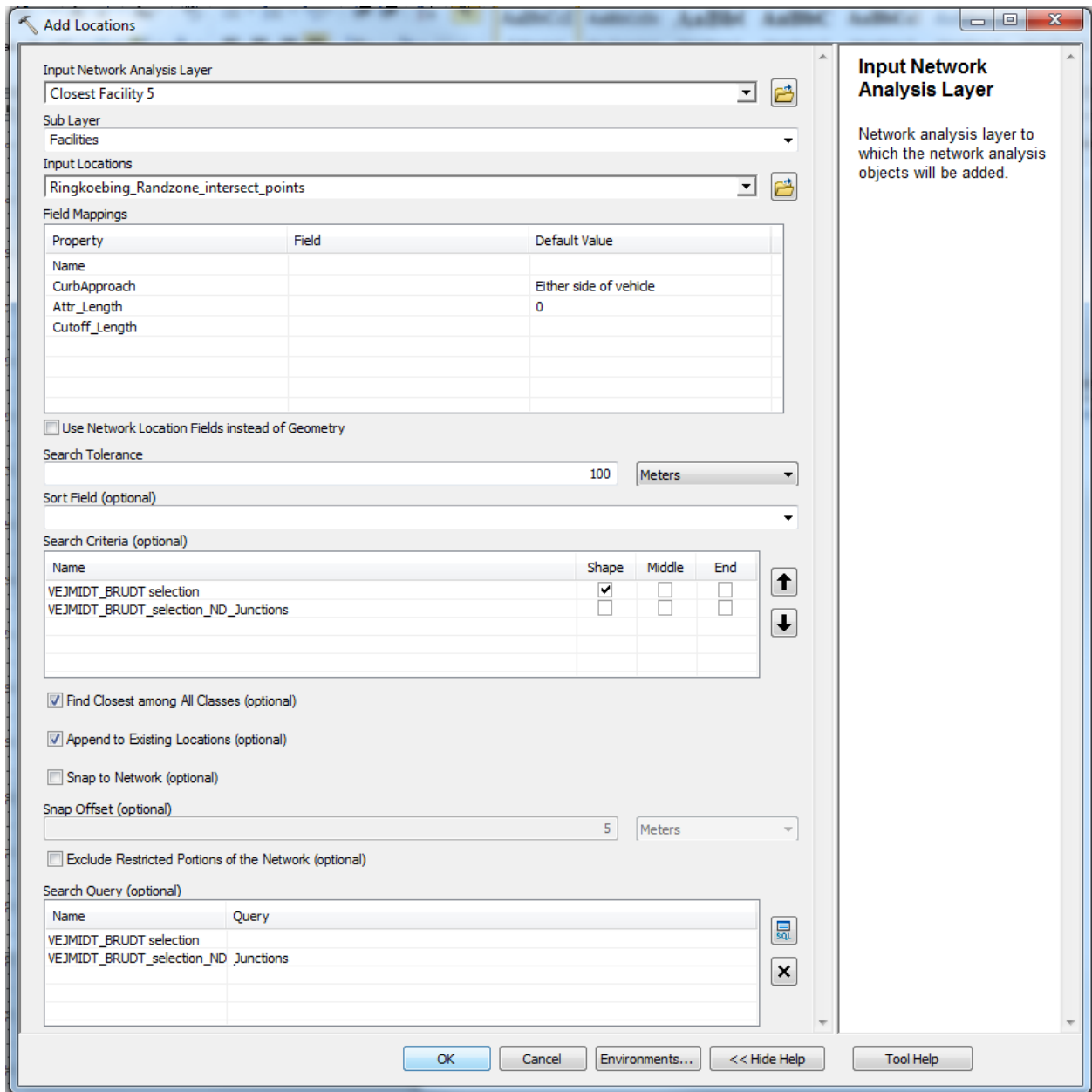
Figure 15: Window 'Add Locations'

Calculate the Road Distance

After creating the layer and adding the necessary locations (facility and incidents) the final step is to 'Solve' this network problem. Therefore, one can either use the ⊞ -button in the taskbar or the command 'Solve' in the 'System Toolboxes' under the category 'Network Analyst Tools' and folder 'Analysis'.

After resolving the 'closest facility layer', one can find the length of the road distance from the house in the attribute table of the layer routes (see Figure 16 and Figure 17).
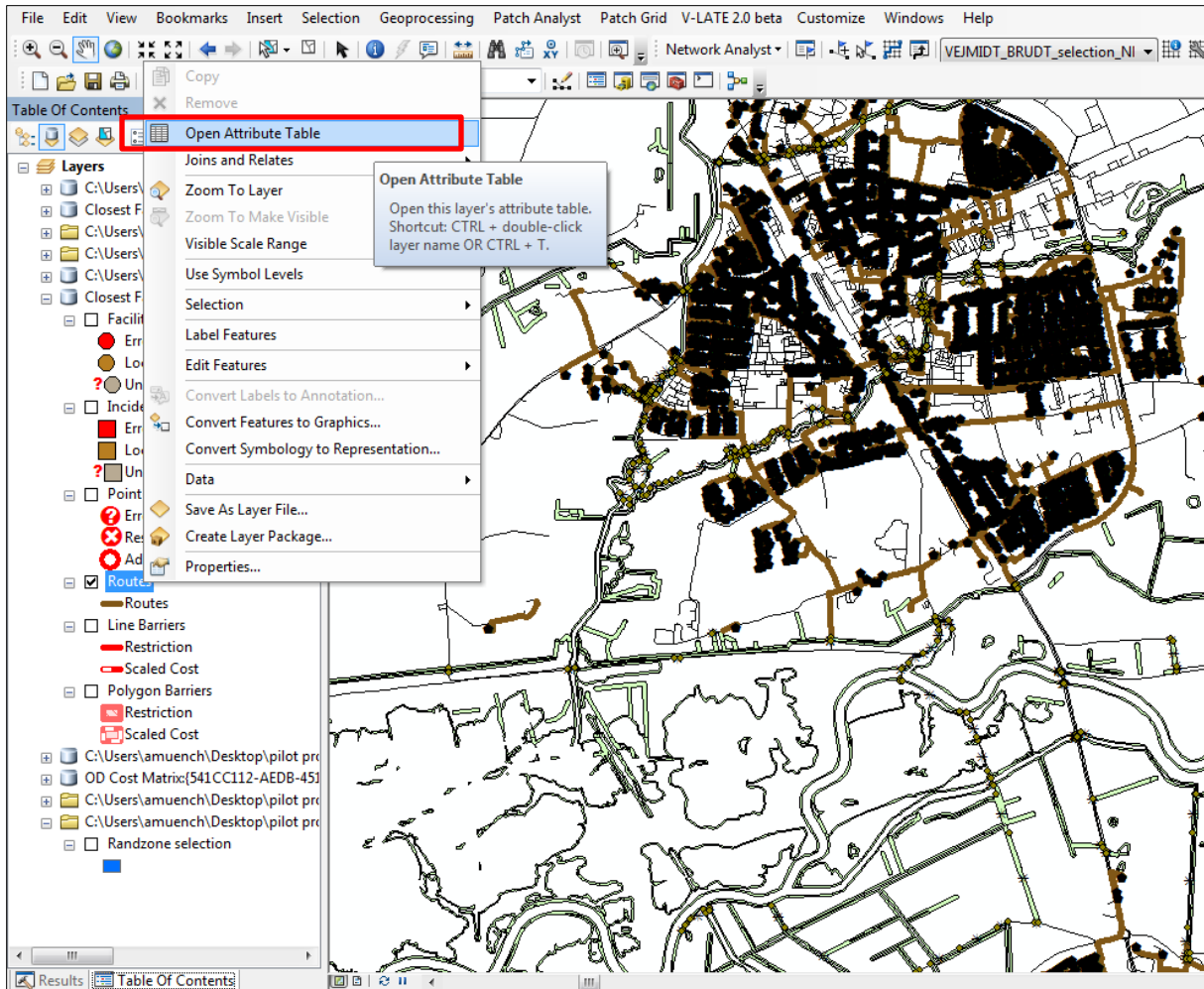
Figure 16: Solved 'Closest Facility' layer

Figure 17: Attribute table of 'Routes' of the 'Closest Facility' layer

A description of how to export the data stored in the attribute table is given in section 6.3.

4.3. Fragmentation Calculation

In section 2.2.2, it is noted that the fragmentation of the surrounding environment of the houses (Fi) gives a utility to the house owner (see also section 2.2.1). In this analysis to assess the utility created by open space, the mean patch fractal dimension (MPFD) is included and serves as a fragmentation index (Raines, 2002). This index measures the shape complexity, which equals 1 for shapes with simple perimeters and approaches 2 when shapes are more complex; thus, more space is taken up by the natural open space elements. Calculations will be executed with the help of the extension 'Patch Analyst' available for ArcGIS (Rempel et al., 2012). This extension allows the use of shapefiles with polygon features as a basis for analyzing the fragmentation of an area. The standard analyzed unit in this tool is a hexagon structure, which can be directly created by the extension as well. However, it was opted to use a classical buffer around the house as a unit of analysis instead of the hexagon structure. The hexagon structure would encompass the entire area in the calculation, i.e. the areas of the postal codes Skjern, Tarm and Ribe. Of interest for the analysis is the open space directly surrounding each house and not the open space available in the area in general. Using the buffer structure instead of a general hexagon structure, however, leads to the issue of the optimal range of the buffer. According to the shapefile 'Jordstykke' provided by Geodatastyrelsen, the average lot around the house is 140 meters long. It was therefore decided to calculate the fragmentation of an area of 150 meters around the house. Landscape elements like water streams, lakes and agricultural land served as open space elements.

4.3.1. Create Buffers

In the first step, the buffer around the house is drawn into the house layer. Selecting the command 'Buffer' in the folder 'Proximity' in the category 'Analysis Tools' in the 'System Toolboxes', the window 'Buffer' opens (see Figure 18). The layer of houses as created above is selected as 'Input Features'. The field 'Output Feature Class' contains the name and location of the new created layer. In the field 'Distance' the linear unit of 150 meters is entered as this should
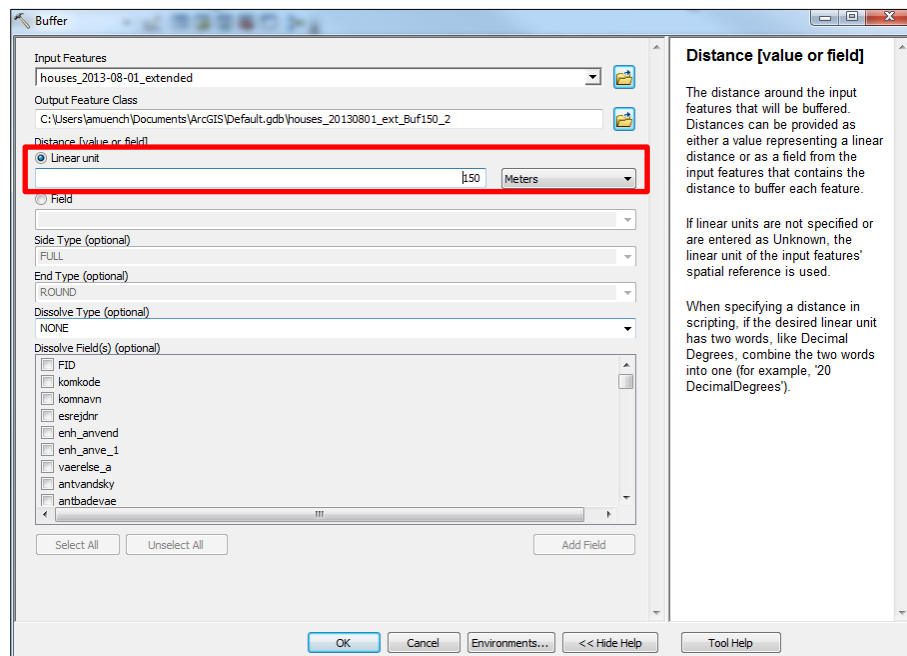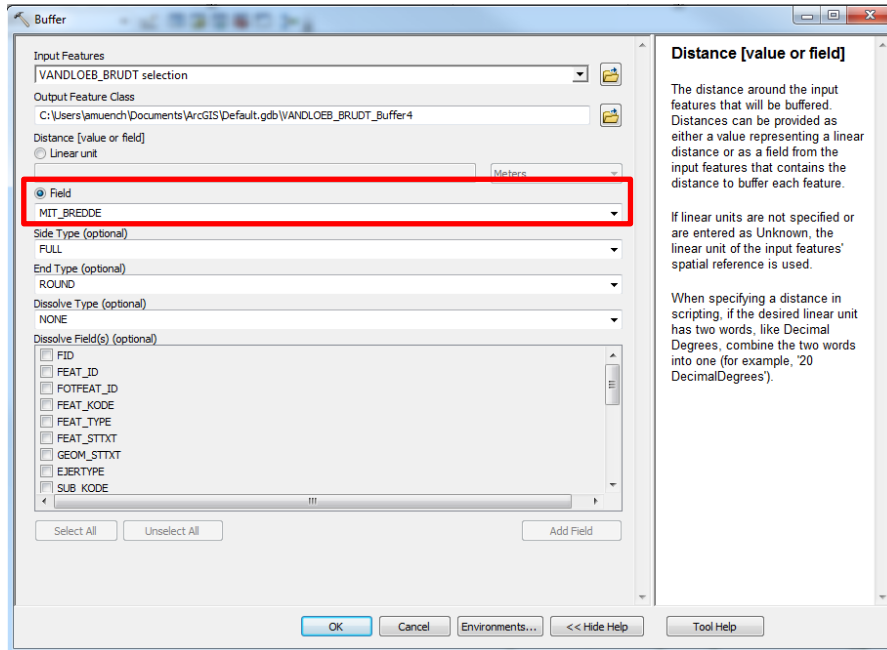

Figure 18: Window 'Buffer with uniform width'

be the radius of the buffer around the house to be considered in the fragmentation analysis.

As preparation for the next step (i.e. the Union of Layers), we need to transform the layer

displaying the water streams from a polyline to a polygon feature based layer as all layers need to be polygon features. A simple way to transform polyline into polygon features is to use the command 'Buffer'. However, as rivers tend to vary in width, it is refrained from using a uniform buffer width; instead the width of the water stream, as stored in the attribute table, should determine the width of the buffer around the stream. Therefore, in the window 'Buffer' as 'Input Features' the layer describing the water streams is chosen. The 'Output Feature Class' is the name and (file) location of the newly created layer. For 'Distance' of the buffer, the row in the attribute table 'stream width' is selected (see Figure 19).

4.3.2. Calculate the Geometric Union

After buffers around the house samples are generated, a geometric union of the open space features (lakes, rivers and agricultural land) is computed in order to have one layer of all relevant features as the basis for the 'Patch Analyst'. This union layer is created by selecting the command 'Union' in the 'System Toolboxes' in the
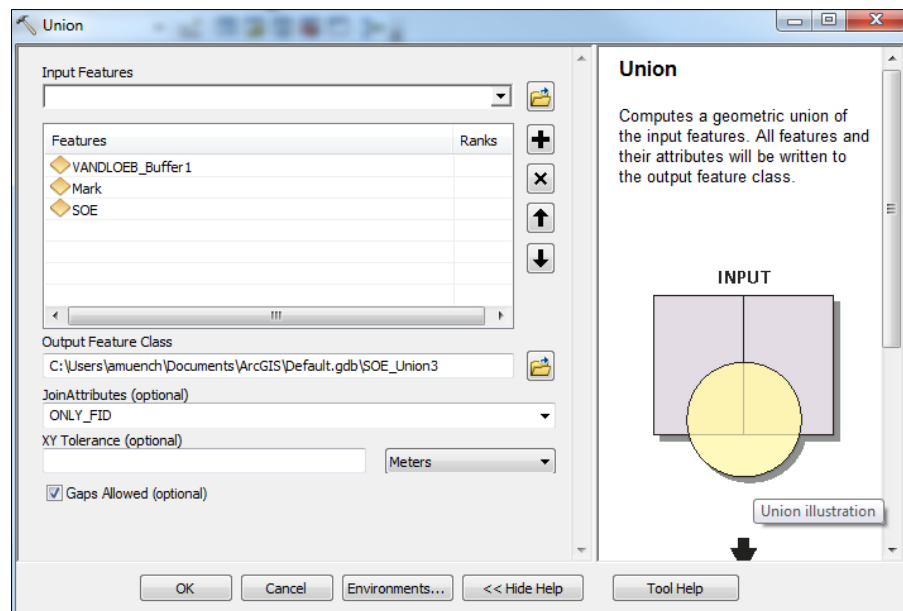
Figure 20: Window 'Union'

27

category 'Analysis Tools' in the folder 'Overlay'. In the window 'Union' (see Figure 20), we specify as 'Input Features' all layers which display the open space features as polygon (buffer of water streams, lakes and agricultural land). The field 'Output Feature Class' contains the name and the location of the new layer. In the field 'Join Attributes', the option 'Only_FID' is selected in order to keep the size of the united layer small as well as the fact that the specific information of the landscape element for the following analysis is not relevant. The box 'Gaps Allowed' is marked in order to acknowledge the area taken up by other landscape elements (e.g. neighboring houses, trees). The 'XY Tolerance' is set to 'Null' in order to keep accuracy as much as possible.

4.3.3. Clip the Open Space with the House Buffer

After the different open space layers are bundled with the help of the command 'Union' into one layer, this common layer is reduced to the information necessary for the patch analyst. The relevant information is the open space in the buffer zone around the house. Therefore, the command 'Clip' is employed which can be found in the 'System Toolboxes' in the category 'Analysis Tools' in the folder 'Extract'. In the 'Clip' window (see Figure 21), we denote the layer of the open space elements as 'Input Features'. The layer with the buffers around
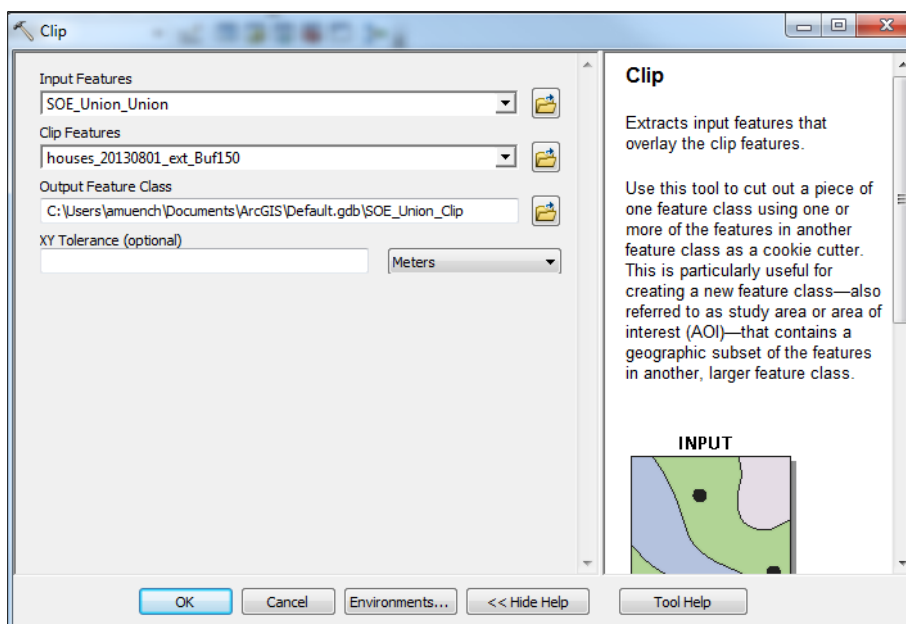
Figure 21: 'Window 'Clip'

the houses serves as 'Clip Features'. Again, the field 'Output Feature Class' contains the name and the location of the new layer, while the 'XY Tolerance' is set to 'Null' in order to preserve accuracy.

After the data displayed in the open space layer is reduced to the buffer zone area around the house, the housing data is related to the reduced open space layer. This step is taken in order to identify the houses in the attribute table of the reduced open space layer as well as possessing an ID-variable for the patch (i.e. house ID) which can be used in the 'Patch Analyst'. In order to relate both of these layers, the 'Union' command is chosen again. However, instead of selecting the option 'Only_FID' as 'Join Attributes' as above, 'All Join Attributes' will be kept in the new layer.

After these steps, the layer which is further used for the 'Patch Analyst' should look similar to Figure 22. Within the buffer zones around the house, the different characteristics are pointed out. Moreover, considering the 'Attribute Table' (Figure 23), the different open space elements are still differentiated for each house (observation). The elements in the table are increased tremendously to the fact that each element in the 150-meter buffer zone is listed separately per house. Hence, the same house should be in the table more than once. See, for example, ESREJD-Nr. 312340 (i.e. Identification number for the houses).
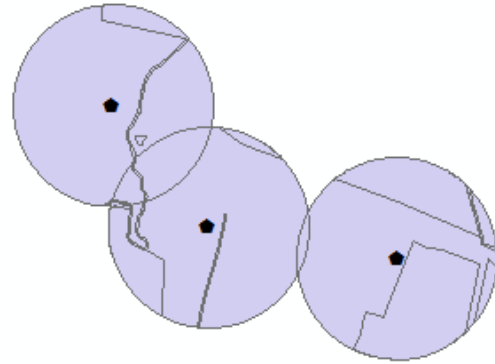

Figure 22: Example display of the 'Open Space' layer

Apparently, within a 150-meter radius around this house, one can find at least four different fields. In contrast, ESREJD-Nr. 312311 possesses at least one water stream within the 150-meter radius. How much of the area is taken up by these different open space elements can be seen in the column 'Shape Area'.

OBJECTID *	Shape *	FID houses	esrejdnr	FID SOE Uni	FID SOE	FID Mark	FID VANDLO	Shape Length	Shape Area
1439	Polygon ZM	12	312311	-1	0	0	0	153,788618	432,537119
1440	Polygon ZM	11	312309	-1	0	0	0	284,732887	1432,64448
1441	Polygon ZM	12	312311	-1	0	0	0	284,732887	1432,64448
1442	Polygon ZM	12	312311	-1	0	0	0	237,736114	400,156198
1443	Polygon ZM	1245	319390	-1	0	0	0	237,736114	400,156198
1444	Polygon ZM	12	312311	1023	0	0	408200	49,992718	62,927066
1445	Polygon ZM	19	312326	-1	0	0	0	48,879432	123,206152
1446	Polygon ZM	331	314388	-1	0	0	0	48,879432	123,206152
1447	Polygon ZM	19	312326	-1	0	0	0	134,58638	612,437088
1448	Polygon ZM	446	314946	-1	0	0	0	134,58638	612,437088
1449	Polygon ZM	21	312340	-1	0	0	0	1057,798015	20536,771497
1450	Polygon ZM	22	312341	-1	0	0	0	1057,798015	20536,771497
1451	Polygon ZM	21	312340	-1	0	0	0	157,960817	1215,255455
1452	Polygon ZM	206	313416	-1	0	0	0	157,960817	1215,255455
1453	Polygon ZM	21	312340	2124	-1	339023	-1	388,113071	8937,441445
1454	Polygon ZM	21	312340	2130	-1	402434	-1	513,500617	14737,685522
1455	Polygon ZM	21	312340	2146	-1	581907	-1	217,107087	2445,898681
1456	Polygon ZM	21	312340	2147	-1	581911	-1	164,985481	988,802086
1457	Polygon ZM	22	312341	-1	0	0	0	71,198979	167,975138
1458	Polygon ZM	206	313416	-1	0	0	0	71,198979	167,975138
1459	Polygon ZM	22	312341	2096	-1	56511	-1	376,200308	7351,003844
1460	Polygon ZM	22	312341	2110	-1	294876	-1	434,526972	11078,492993
1461	Polygon ZM	22	312341	2111	-1	294877	-1	401,130198	3361,535362
1462	Polygon ZM	22	312341	2134	-1	402442	-1	88,494373	275,696755
1463	Polygon ZM	22	312341	2147	-1	581911	-1	40,7473	72,463269
1464	Polygon ZM	23	312342	-1	0	0	0	158.055949	850.267303

(0 out of 3572982 Selected)

houses_20130801_ext_Buf150_U2

Figure 23: Attribute Table of the Open Space Layer

4.3.4. The Patch Analyst

After the data layer is prepared, one can start the 'Patch Analyst', which should give a more detailed picture on the fragmentation of the area surrounding the houses. This extension for ArcGis can be downloaded from the Centre for Northern Forest Ecosystem Research (Rempel et

al., 2012). This extension will be added to the ArcMap toolbar instead of to the catalogue folders. Selecting the command 'Spatial Statistics' in the drop down menu of the 'Patch Analyst' will open a window (see Figure 24). The above created layer, which contains information on houses and open space elements, is selected as 'Layers' to be analyzed. The variable which identifies the 'Class' is the house-ID (i.e. ESREJDNR). As we are interested in the fragmentation around the house, 'Analyze by Class' is selected. The option 'Analyze by landscape' refers to general landscape fragmentation analysis, which is not feasible in this case as the data does not contain all of the necessary information anymore (due to the former clipping).
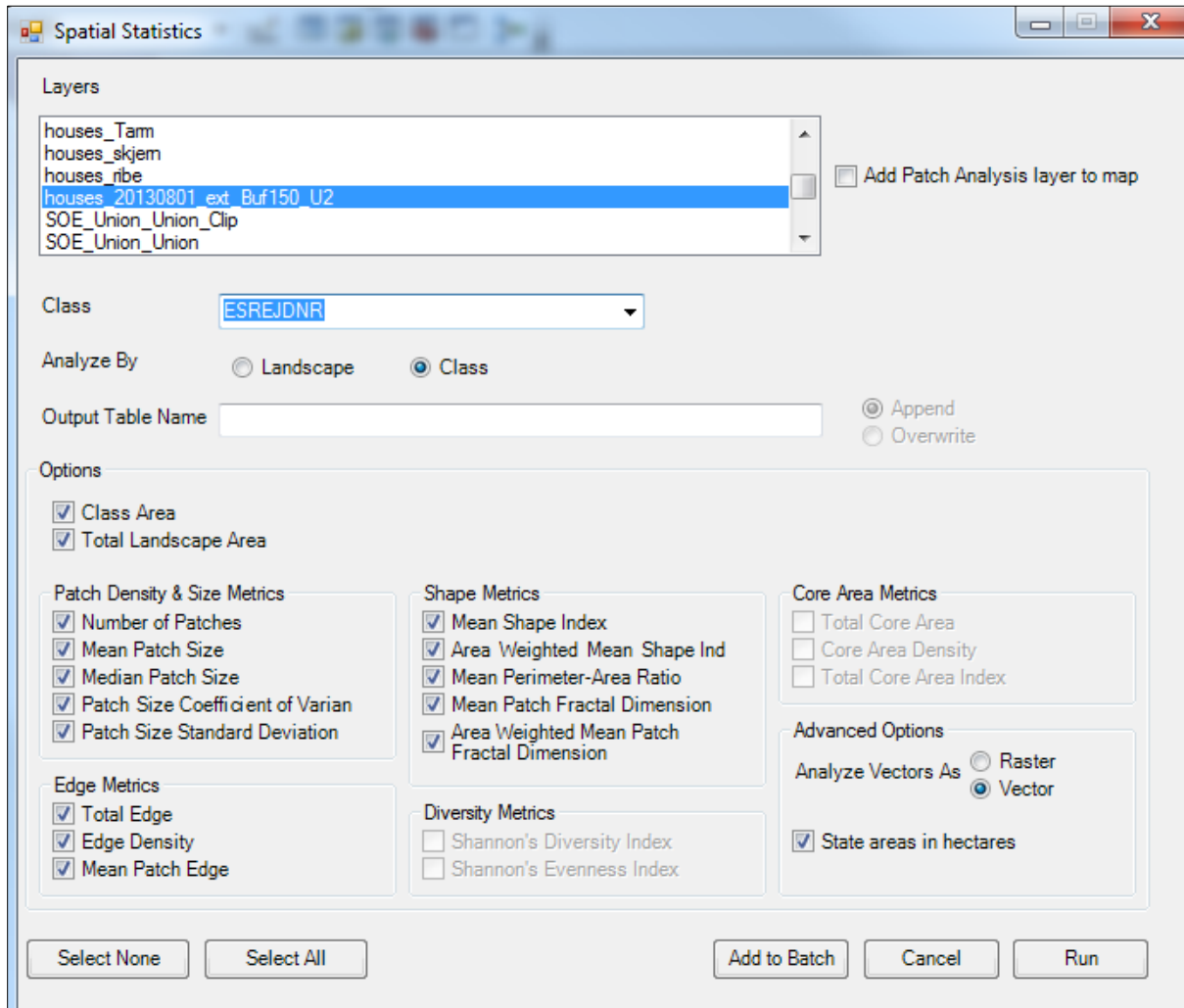

Figure 24: Spatial Statistics of the Patch Analyst

After running the 'Patch Analyst' on our prepared data layer, we obtain a table layer which contains all the information on the patch we selected before. The variable of interest is the 'Mean Patch Fractal Dimension' (MPFD), which can be found in the rubric 'Shape Metrics'. The standard measures like the Shannon Diversity or Evenness Index are only selectable if a landscape is examined. As described above, this is not feasible in our case. Therefore, the MPFD is chosen although it is not highly taken up and discussed in the literature.

4.4. Creation of a Dummy-Variable

In the theoretical approach in section 2.2.2, it was discussed that in order to measure the effect of the buffer zone regulation on house price, a dummy variable should be integrated which takes the value '1' if the newly created buffer zone is on the plot of the sold house, and '0' if otherwise. Additionally, a dummy variable is included to capture the effect of having a water stream on the plot (see Figure 25). Both variables will be calculated in the same way as the only difference is the data involved. In the following, the way the second dummy variable 'riparian property' is calculated is described. Riparian properties are defined as plots belonging to houses which are crossed by a water stream. Thus, the water stream needs to be either on the property or adjacent in such a way that the outer border of the plot is overlapping with the outer border of the water stream. Hence, a property is not considered to be a riparian property if a water stream is further than one meter away from the property. The tolerance of one meter is considered due to variation in water levels of the rivers.
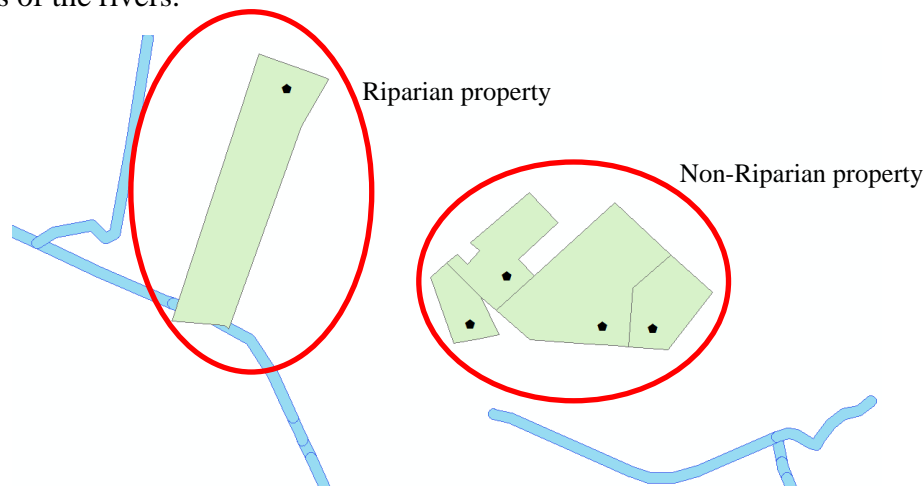

Riparian property

Non-Riparian property

Figure 25: Definition of Riparian Property

4.4.1. Spatial Join the sample of the houses with its plots

In the first step, we need to assign the plots to the houses in our database. Geodatastyrelsen provides the shapefile 'Jordstykke' which will be used here as a basis. Within the 'System Toolboxes' under the category 'Analysis Tools' in the folder 'Overlay', one can select the command 'Spatial Join'. Within the 'Spatial Join' window (see Figure 26), the following specifications were taken:

1. Target Features: the layer displaying the plots.
2. Join Features: the layer of our housing sample as created above.
3. Output Feature Class: the name and (file) location of the newly created layer.
4. Join Operation: as each house only has one plot, the option 'Join_one_to_one' is selected.
5. To keep all information on the houses as well as have some control options by adding the information on the plot, the box 'Keep All Target Features' is ticked.

6. Match option: 'Contains' is selected in order to identify which plot contains a house in our database.
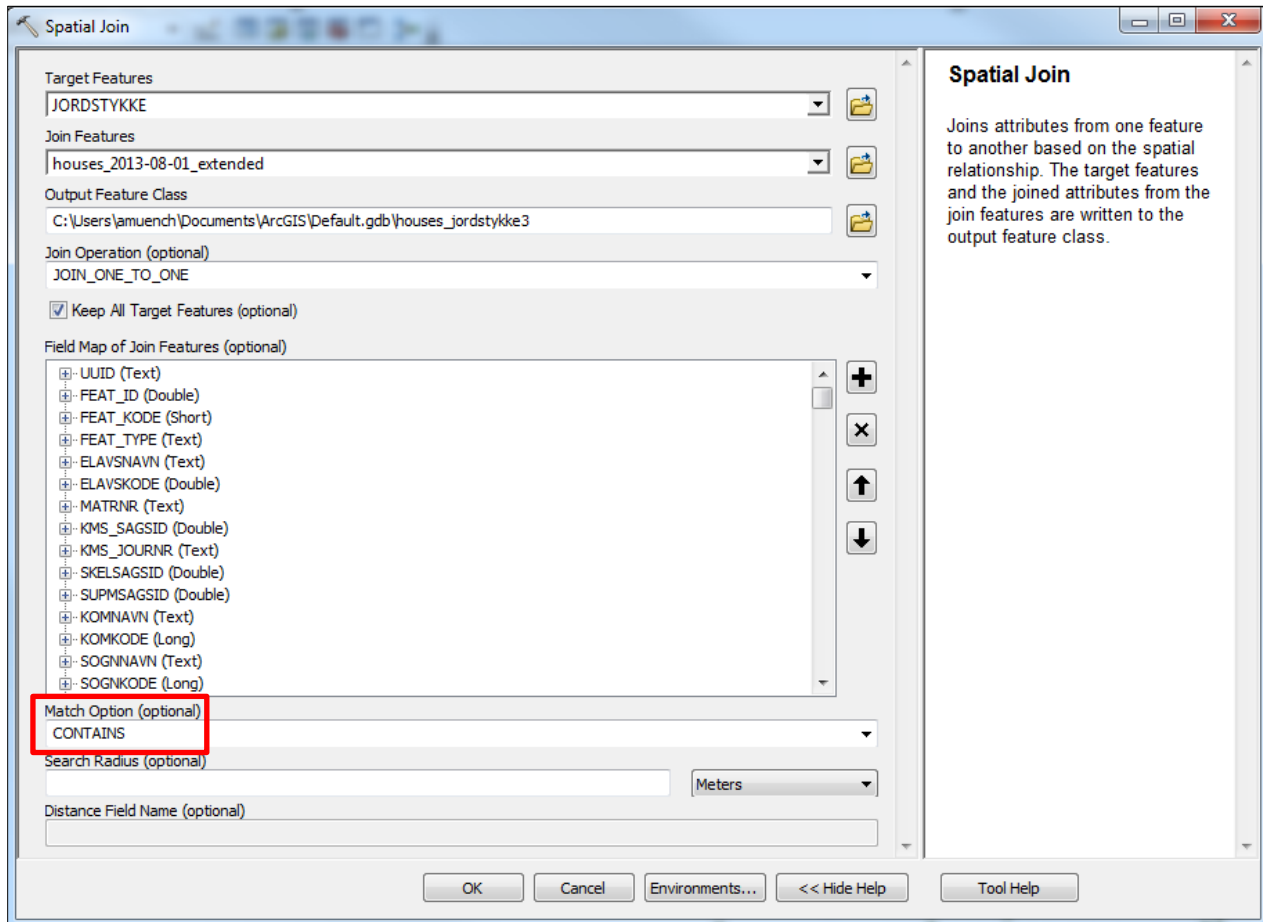

Figure 26: Window 'Spatial Join' relating houses to the plots

Please note: as we are interested in the characteristics of the plots, the target feature is the plot. In this step, we simply relate the database of the houses to the property plots. In this step, the tolerance of calculation ('Search Radius') is zero.

4.4.2. Identify the Riparian Properties in the Sample

In order to relate the plots we are interested in to the water stream layer, both layers need to be the same format (i.e. polygon feature). Therefore, instead of using the original shapefile display-ing the water streams as polylines, the polygon layer of water streams created in 4.3.1 with the help of buffers is here employed further.

In the next step, we intersect the spatial join layer with the water stream buffer layer. Using the 'Spatial Join' command again, this time the 'Target Features' is the layer describing the plots/houses. The 'Join Features' is the created layer of the buffers around water streams. As we are only interested in the houses in our sample and not all possible plots in the area, and because it is enough for us to know if there is at least one water stream on the property (i.e. the number of

water streams on the property is not important), the options 'Join_one_to_one' and 'Intersect' are chosen. In order to allow some tolerance, the 'Search Radius' is defined as one meter (see Figure 27).
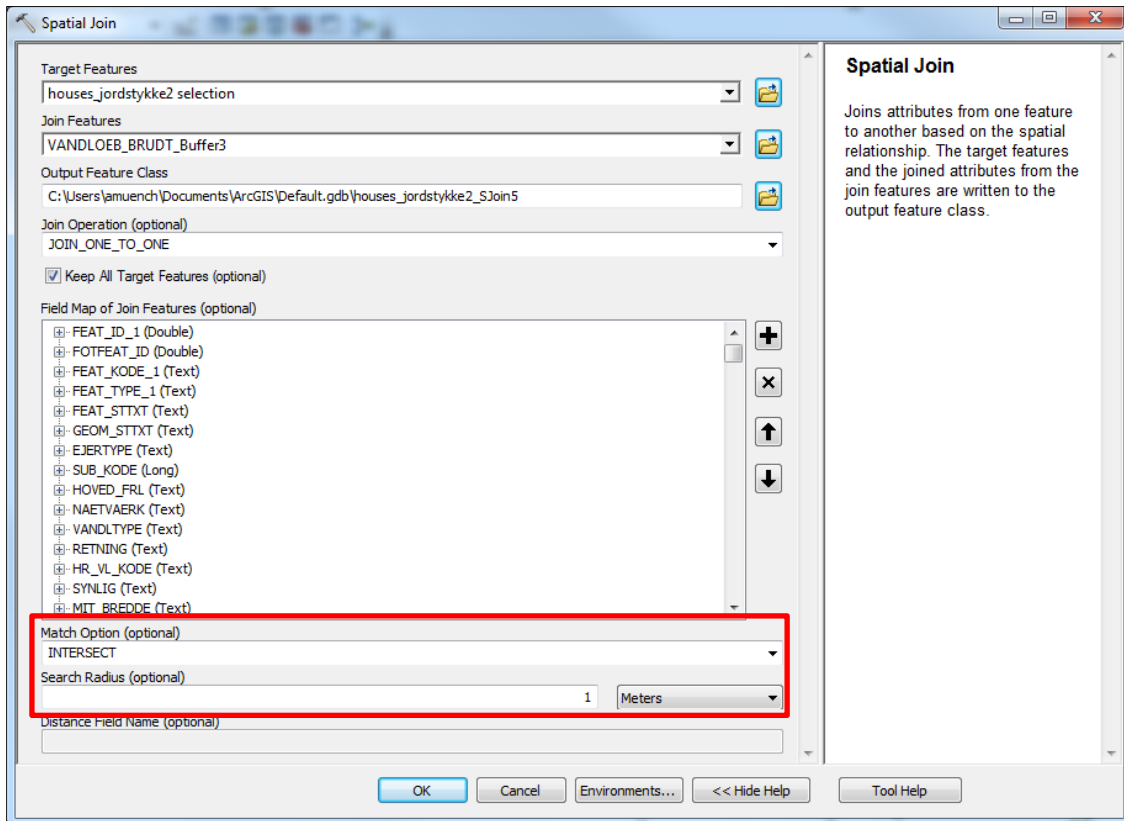

Figure 27: Window 'Spatial Join' relating plots to the water streams

In the newly created layer, one can find in the attribute table the column 'Shape Length' and 'Shape Area' (one of the last columns – see Figure 28). The rows containing the value 'Null' signal that this property is a non-riparian property. In these rows, if there is a value it captures the length of the connection property – water stream ('Shape Length') and how much of the area of the property is in fact a water body ('Shape Area').

Figure 28: Attribute Table Spatial Join relating plots to the water streams

A description of how to export the data stored in the attribute table is given in section 6.3.

5. Spatial Statistics with ArcGIS

In section 3, spatial dependence is explained based on the work of Fotheringham and Rogerson (2009) as well as Anselin and Getis (2010). Differentiating between spatial autocorrelation and spatial heterogeneity, the two works claim that spatial processes or pattern needs to be integrated if calculation should not be biased. ArcGIS offers different tools within its Spatial Statistics Package in order to allow such analysis of the spatial patterns.

5.1. Generate Spatial Weights Matrix

To identify and incorporate spatial autocorrelation into a quantitative analysis, a spatial weight matrix is necessary. This spatial weight matrix denotes the spatial relationship of observation i to all other observations j. Dependent of the dataset or research question, a different weighting matrix can be calculated with the help of ArcGIS. The spatial weights matrix is a $N \times N$ matrix with the diagonal '0'. Classical weighting matrixes are the neighborhood matrix or the inverse distance matrix, although more special matrixes, like a space-time-window or shared boundaries, can be generated as well. The neighborhood matrix takes the value '1' if observation i is the neighbor of observation j while in the inverse distance matrix, a higher weight will be devoted to observation j as it is closer to observation i.

In the 'System Toolboxes' in the category 'Spatial Statistics Tools' in the folder 'Modeling Spatial Relationships', the command 'Generate Spatial Weights Matrix' is found, which opens the following window (see Figure 29):
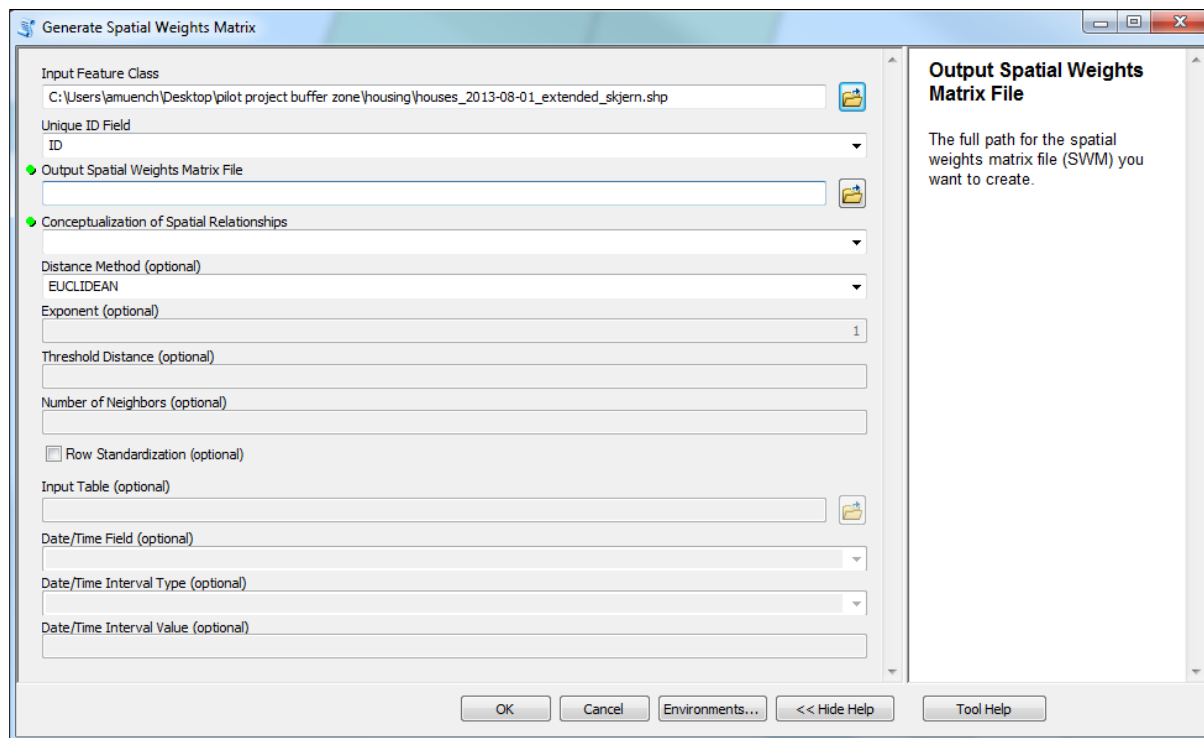

Figure 29: Window 'Generate Spatial Weights Matrix'

The 'Input Feature Class' is, in our case, the layer of houses in the sample as we would like to account for the violation of the independence of the observation axiom created by proximity between houses (e.g. houses close to each other have a similar sales price as they are located in the same area) later in the analysis. For 'ID', the field that is unique for each observation needs to be selected. Please note: It might be necessary to create an extra unique identifier for the dataset as ArcGIS does not use its own FID for this analysis. If, as above, ESREJDNR shall serve as 'ID', a transformation of the format of this column is necessary, as the ESREJDNR is stored in the 'double' format but the 'ID' field needs to be an integer. After opening the 'Attribute Table', one can select 'Add Field' in the menu of the attribute table. The new field is then defined as type 'long integer' and will be added as an empty space in the column in the attribute table. In order to fill this column with the respective ESREJDNR, one opens the field calculator with a right mouse click on the column name and defines 'ID = ESREJDNR' whereby a new ID-field is calculated which can then be used for the calculation of the geographic weights.

The field 'Output Spatial Weights Matrix File' requires the name and location of the newly created spatial weight matrix file (.swm). In the field 'Conceptualization of Spatial Relationships', one can select from various forms (e.g. neighborhood matrix, cost-time matrix or shared boundary matrix). In our case, we chose the 'inverse distance matrix' as the conceptualization method. Depending on the conceptualization method, one can further refine the exponent for the inverse distance, the threshold distance or number of neighbors it should consider (i.e. restricting the extent of the matrix).

Distance can be measured as Euclidean ('as the crow flies') or as Manhattan (summing the difference between x- and y-coordinates). It is decided to use the 'Euclidean' 'Distance Method' here.

If the field 'Row Standardization' is ticked, the spatial weights are standardized by dividing each weight by its row sum. In our case, the spatial weight matrix is further used in STATA; our matrix is not standardized at this point but will be later by transferring the matrix within STATA.

5.2. Spatial Autocorrelation & Spatial Statistics Tools

ArcGIS 10.1 has already implemented several tools to analyze spatial autocorrelation and to perform spatial statistics. In the study of Münch et al. (2013), STATA was used instead of ArcGIS. Therefore, these tools won't be described in more detail here. The test for spatial autocorrelation can still be found in the 'System Toolboxes' in the category 'Spatial Statistics Tools':

- *Local Moran's I*: command 'Cluster and Outlier Analysis (Anselin Local Morans I)' in the folder 'Mapping Clusters'.
- *Local Gettis Ord G*: command 'Hot Spot Analysis (Getis-Ord Gi*)' in the folder 'Mapping Clusters'.
- *Global Moran's I*: command 'Spatial Autocorrelation (Morans I)' in the folder 'Analyzing Patterns'.

- *Global Gettis Ord G*: command 'High/Low Clustering (Getis-Ord General G)' in the folder 'Analyzing Patterns'.
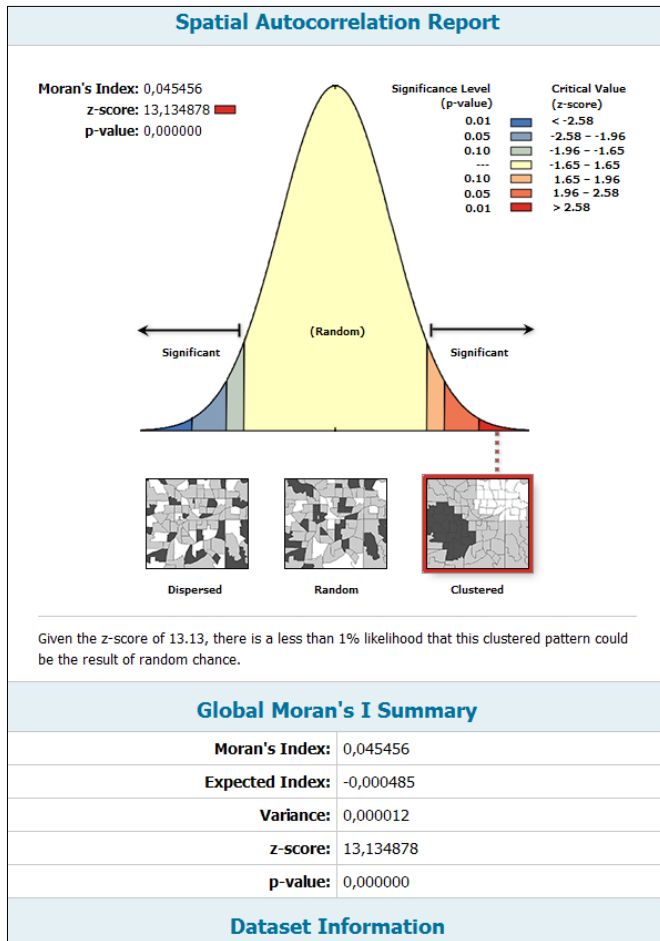

Figure 30: Spatial Autocorrelation Report

ArcGIS further provides the user with a report which offers information and an interpretation of the found spatial correlation among the observations of the variable under consideration. Figure 31 shows an example of the report created for our dependent variable of sales prices of houses in Ribe. The high Z-score and the p-value, respectively, do not allow the rejection of the Null-hypothesis of 'no spatial autocorrelation in the distribution of the variable'. Hence, an estimation model needs to be found which considers these underlying structures and will therefore not be biased when estimating the standard errors (and with this, the significance of the results).

Should further test results and estimation model comparisons hint to Spatial Heterogeneity (i.e. observations of the dependent variable react differently towards the change of an independent variable due to geographic location), ArcGIS 10.1 offers to calculate a 'Geographically Weighted Regression' (in System Toolboxes\Spatial Statistics Tools\Modeling Spatial Relationships). For a detailed description of this estimation method, see Fotheringham and Rogerson (2009).

6. Data Export

6.1.   XY-layer to Shapefile

In order to enable calculation with the data, the layer displaying XY-data needs to be transformed into a shapefile. After a right mouse click on the layer, one can choose the commands 'Data'/'Export Data' (see Figure 31).
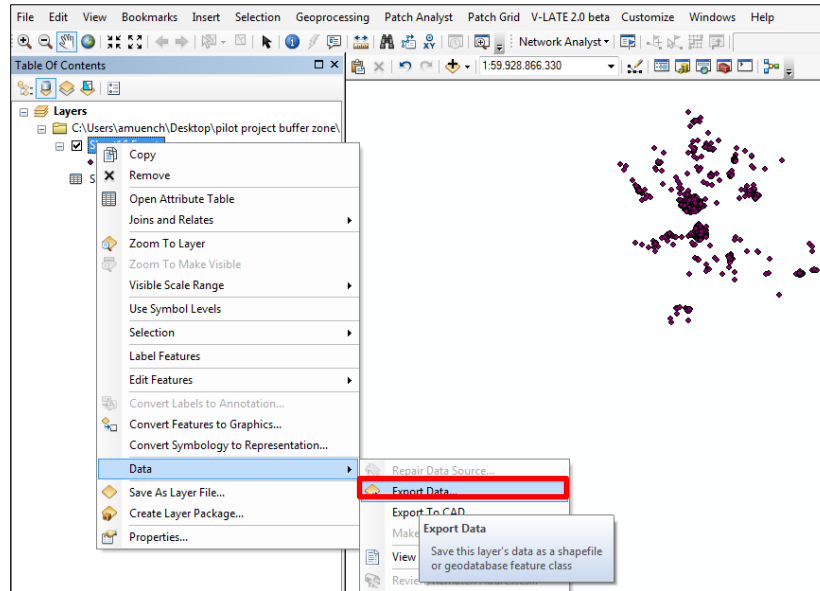

Figure 31: Creating a shapefile from the XY-data

After exporting all features of the layer by using the coordinate system specified before for the layer, one has created a shapefile of point data from an XY-Data set which can now be added into a ArcGIS map document and processed further. In the attribute table of the shapefile, one finds all the variables which were stored in the former Window Excel-file.

6.2.   Layer Table to .csv or .txt File

Additionally, results are saved in an extra table layer, which can easily be exported and/or joined to the original database. Exporting the data basically requires the same step as described above in section 6.1. With a right mouse click on layer one, select 'Data' and then 'Export …'. The window as displayed in Figure 32 should open. Define export as 'All records', and write the location where the output file should be stored. While choosing the storage location (selecting), one can at the same time specify the file type e.g. save the output table as a .csv or a .txt file in order to process it further in

Figure 32: Window 'Export Data'

programs such as Microsoft Excel.

6.3. Attribute table to .csv or .txt File

In order to use the data within, e.g. Microsoft Excel, the attribute table can be exported into a .csv or .txt file After opening the attribute table (right mouse click on the layer and choose the command 'Open Attribute Table'), in the drop-down menu the command 'Export…' is chosen (see Figure 33). The same window as in section 6.2 is displayed, i.e. same steps apply.
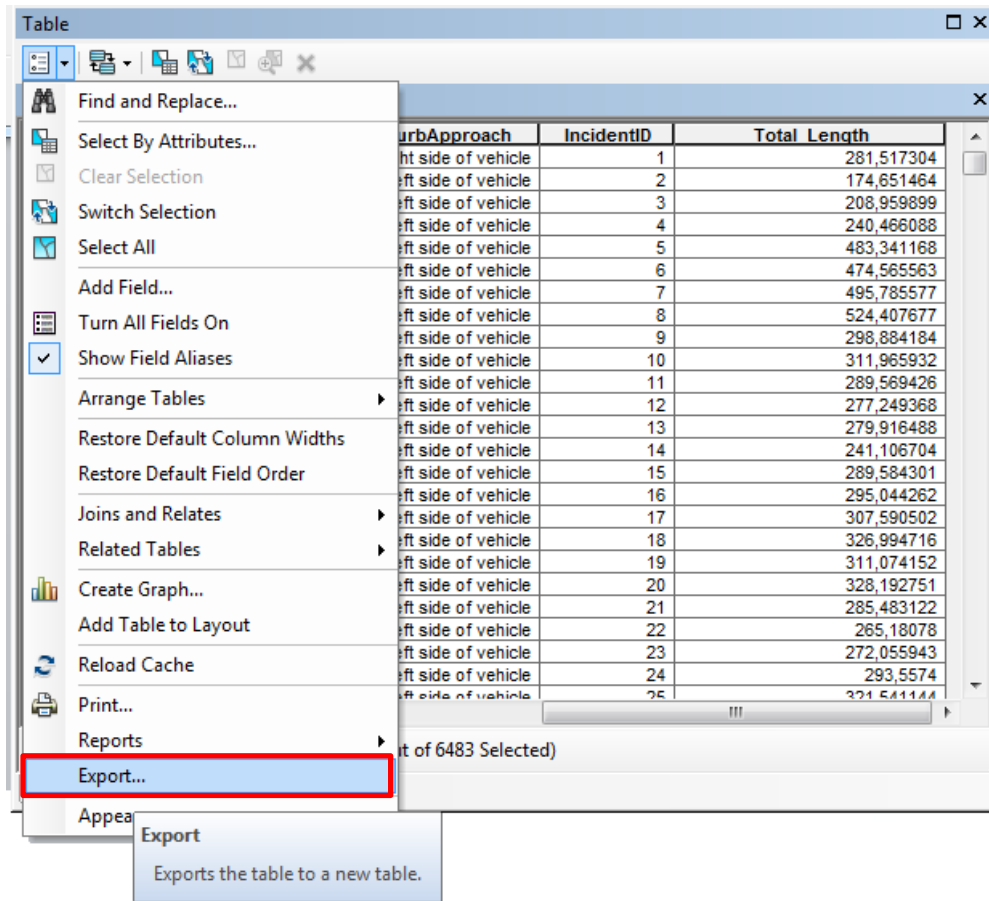
Figure 33: Export Attribute Table

6.4. Weighting Matrix to Layer Table

In section 5.1, a spatial weights matrix was created that captures the distance between observations. In order to use this matrix in other programs, the file needs to be converted. Within the 'System Toolboxes' in the category 'Spatial Statistics Tools' in the folder 'Utilities, one finds the command 'Convert Spatial Weights Matrix to Table'. In the following window (see Figure 34), information is required regarding the name and (file) location of the created spatial weights matrix in the field 'Input Spatial Weights Matrix File' as well as the name and file location of the table to be created (i.e. 'Output table').

After using this command, a layer table will be created in the current ArcGIS map document, which can be easily exported as described in section 6.2.
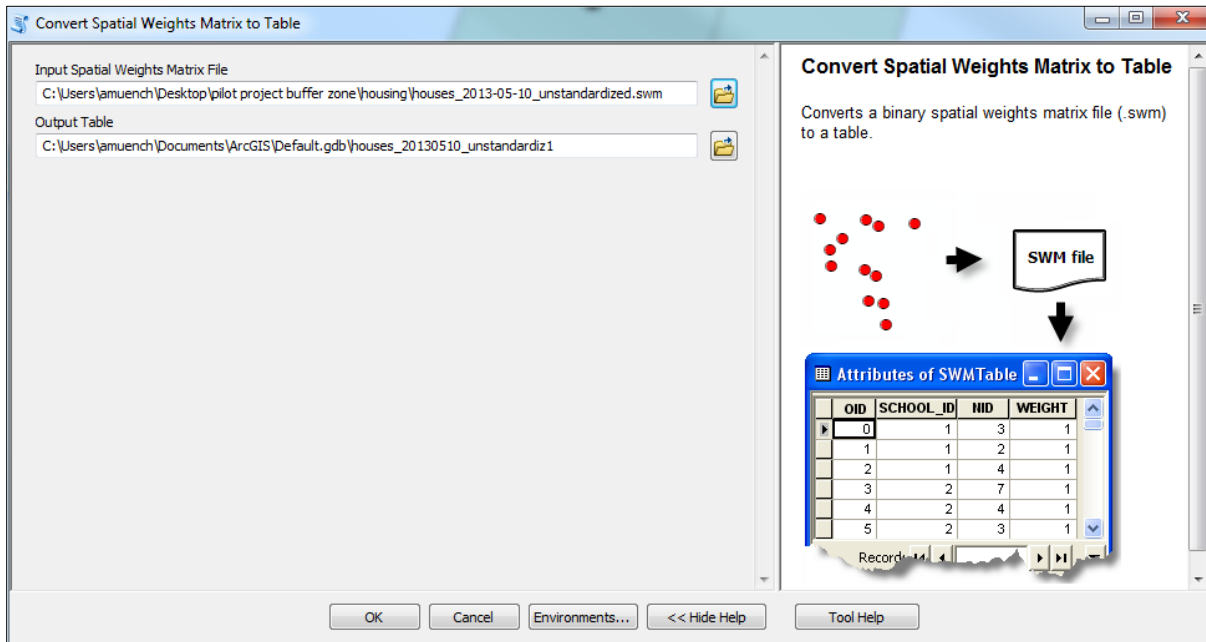

Figure 34: Window 'Convert Spatial Weights Matrix to Table'

References

Acharya, G., Bennett, L.L., 2001. Valuing open space and land-use patterns in urban watersheds. *The Journal of Real Estate Finance and Economics* 22, 221-237.

Anselin, L., 1990. Spatial Dependence and Spatial Structural Instability in Applied Regression Analysis. *Journal of Regional Science* 30, 185-207.

Anselin, L., 1999. *Spatial Econometrics*, Tutorial. University of Texas at Dallas.

Anselin, L., 2001. Spatial effects in econometric practice in environmental and resource economics. *American Journal of Agricultural Economics* 83, 705-710.

Anselin, L., 2002. Under the hood issues in the specification and interpretation of spatial regression models. *Agricultural Economics* 27, 247-267.

Anselin, L., 2010. Thirty years of spatial econometrics. *Papers in Regional Science* 89, 3-25.

Anselin, L., Getis, A., 2010. Spatial statistical analysis and geographic information systems. In: Anselin, L., Rey, S.J. (ed.) *Perspectives on Spatial Data Analysis*. Springer, pp. 35-47.

Anselin, L., Rey, S.J., 2010. *Perspectives on spatial data analysis*. Springer.

Bateman, I.J., Jones, A.P., Lovett, A.A., Lake, I., Day, B., 2002. Applying geographical information systems (GIS) to environmental and resource economics. *Environmental and Resource Economics* 22, 219-269.

Bin, O., Landry, C.E., Meyer, G.F., 2009. Riparian buffers and hedonic prices: a quasi-experimental analysis of residential property values in the Neuse River basin. *American Journal of Agricultural Economics* 91, 1067-1079.

Cho, S.-H., Kim, S.G., Roberts, R.K., 2011. Values of environmental landscape amenities during the 2000–2006 real estate boom and subsequent 2008 recession. *Journal of Environmental Planning and Management* 54, 71-91.

Corrado, L., Fingleton, B., 2012. Where is the Economics in Spatial Econometrics?. *Journal of Regional Science* 52, 210-239.

d'Arge, R.C., Shogren, J.F., 1989. Okoboji experiment: Comparing non-market valuation techniques in an unusually well-defined market for water quality. *Ecological Economics* 1, 251-259.

Dixon, J., 2008. Environmental valuation: Challenges and Practices a personal view. *Economics and Conservation in the Tropics: A strategic dialogue*.

Elhorst, J.P., 2010. Applied Spatial Econometrics: Raising the Bar. Spatial Economic Analysis 5, 9-28.

Fotheringham, A.S., Rogerson, P., 2009. *The Sage handbook of spatial analysis*. Sage Publ., Los Angeles, Calif. [u.a.].

Garrod, G., Willis, K.G., 1999. *Economic valuation of the environment: methods and case studies*. Edward Elgar Cheltenham.

Geoghegan, J., 2002. The value of open spaces in residential land use. *Land Use Policy* 19, 91-98.

Geoghegan, J., Wainger, L.A., Bockstael, N.E., 1997. Spatial landscape indices in a hedonic framework: an ecological economics analysis using GIS. *Ecological Economics* 23, 251-264.

Gibbons, S., Mourato, S., Resende, G., 2011. The amenity value of English nature: a hedonic price approach. *Environmental and Resource Economics* 57, 175-196

Hjorth-Andersen, C., 1976. *Måling af vejstøjens påvirkning af parcelhuspriserne*. Cykelafdelingen, Økonomisk Institut, Københavns Universitet, Kbh.

Knetsch, J.L., 1964. The influence of reservoir projects on land values. *Journal of Farm Economics* 46, 231-243.

Kulshreshtha, S.N., Gillies, J.A., 1993. Economic Evaluation of Aesthetic Amenities: A Case Study of River View. *JAWRA Journal of the American Water Resources Association* 29, 257-266.

Lake, I.R., Lovett, A.A., Bateman, I.J., Day, B., 2000. Using GIS and large-scale digital data to implement hedonic pricing studies. *International Journal of Geographical Information Science* 14, 521-541.

Lansford, N.H., Jones, L.L., 1995. Marginal price of lake recreation and aesthetics: a hedonic approach. *Journal of Agricultural and Applied Economics* 27, 212-212.

Maguire, D.J., Batty, M., Goodchild, M.F., 2005. *GIS, spatial analysis and modelling*. ESRI, Redlands, CA.

Mäler, K.-G., Vincent, J.R., 2005. *Handbook of Environmental Economics: Valuing environmental changes*. Access Online via Elsevier.

Mamoulis, N., 2011. Spatial data management. *Synthesis Lectures on Data Management* 3, 1-149.

Møller, F., Andersen, S.P., Grau, P., Huusom, H., Madsen, T., Nielsen, J., Strandmark, L., 2000. *Samfundsøkonomisk vurdering af miljøprojekter*. Danmarks Miljøundersøgelser, Miljøstyrelsen og Skov- og Naturstyrelsen.

Mooney, S., Eisgruber, L.M., 2001. The influence of riparian protection measures on residential property values: the case of the Oregon plan for salmon and watersheds. *The Journal of Real Estate Finance and Economics* 22, 273-286.

Muller, N.Z., 2009. Using hedonic property models to value public water bodies: An analysis of specification issues. *Water Resources Research* 45.

Münch, A., Porner Nielsen, S.P., Racz, V.J. & A. Hjalager (2013): *Towards multifunctionality of rural natural environments? – A social evaluation of the extended buffer zones along Danish rivers, streams and lakes*, CLF Report 26

Netusil, N.R., 2005. The effect of environmental zoning and amenities on property values: Portland, Oregon. *Land Economics* 81, 227-246.

Netusil, N.R., 2006. Economic valuation of riparian corridors and upland wildlife habitat in an urban watershed. *Journal of Contemporary Water Research & Education* 134, 39-45.

Osland, L., 2010. An application of spatial econometrics in relation to hedonic house price modeling. *Journal of Real Estate Research* 32, 289-320.

Panduro, T.E., Veie, K.L., 2013. Classification and valuation of urban green spaces—A hedonic house price valuation. *Landscape and Urban Planning* 120, 119-128.

Paterson, R.W., Boyle, K.J., 2002. Out of sight, out of mind? Using GIS to incorporate visibility in hedonic property value models. *Land economics* 78, 417-425.

Præstholm, S., Jensen, F.S., Hasler, B., Damgaard, C., Erichsen, E., 2002. Forests improve qualities and values of local areas in Denmark. *Urban Forestry & Urban Greening* 1, 97-106.

Raines, G.L., 2002. Description and comparison of geologic maps with FRAGSTATS—a spatial statistics program. *Computers & Geosciences* 28, 169-177.

Rempel, R., Kaukinen, D., Carr, A., 2012. *Patch Analyst and Patch Grid*. Ontario Ministry of Natural Resources. Centre for Northern Forest Ecosystem Research, Thunder Bay, Ontario, Canada.