Title: Identifying and assessing the residency effect in Pocatello, Idaho, using combined census and parcel data

Author

Tanh T. N. Nguyen, Ph.D.

- 1) College of Forestry and Agriculture, Stephen F. Austin State University, 419 E. College St., Nacogdoches, TX 75962-6109 (Current), Email: ntntanh@gmail.com
- 2) Idaho State University/Geosciences Department (Sponsored)
921 S 8th Ave, Stop 8072, Pocatello, ID 83209-8072
- 3) An Giang University/Department of Environment and Sustainable Development
18 Ung Van Khiem St., Long Xuyen, An Giang, Vietnam (Affiliated)

Abstract

Human residency is the spatial effect source on ecosystem and thus it should be studied and assessed. Supporting residency effect research, this study developed and applied procedures and a model to combine census and parcel data for the assessment. The case study is in Pocatello, Idaho, where revealing land service associated with flood control and locating/evaluating resident effect are needed. Methods include (1) data mining, (2) land service valuation, (3) data screening, (4) integration of census and parcel data, (5) data screening, and (5) analysis and modeling with R programming language and ArcMap. Results are, for land service assessment, land value per area unit in residence areas (LAND) along the concrete channel (for flood control) was less than that along the Portneuf River. Spatial responses under LAND to a source effect (either the concrete channel or the river) are the same. The applied methods helped locate and assess a variety of residency effects spatio-temporally. Results informed the human preferences under LAND and the effect distribution to support decision-making. Technically, using the parcels as a baseline provided comprehensive results with a fine resolution for the effect study, particularly as combined with the census data. This study suggests using a data screening and validation procedure besides the mining approach to minimize outcome uncertainty.

Keywords: Census and parcel; data mining; residency; impact assessment; modeling; R programming.

Identifying and assessing the residency effect in Pocatello, Idaho, using combined census and parcel data

Abstract

Human residency is the spatial effect source on ecosystem and thus it should be studied and assessed. Supporting residency effect research, this study developed and applied procedures and a model to combine census and parcel data for the assessment. The case study is in Pocatello, Idaho, where revealing land service associated with flood control and locating/evaluating resident effect are needed. Methods include (1) data mining, (2) land service valuation, (3) data screening, (4) integration of census and parcel data, (5) data screening, and (5) analysis and modeling with R programing language and ArcMap. Results are, for land service assessment, land value per area unit in residence areas (LAND) along the concrete channel (for flood control) was less than that along the Portneuf River. Spatial responses under LAND to a source effect (either the concrete channel or the river) are the same. The applied methods helped locate and assess a variety of residency effects spatio-temporally. Results informed the human preferences under LAND and the effect distribution to support decision-making. Technically, using the parcels as a baseline provided comprehensive results with a fine resolution for the effect study, particularly as combined with the census data. This study suggests using a data screening and validation procedure besides the mining approach to minimize outcome uncertainty.

Keywords: Census and parcel; data mining; residency; impact assessment; modeling; R programing.

Introduction

Residency is the core activity of human and thus the foremost source of human effect on ecosystem. When residing, human affects his surrounding environment and this forms a spatio-temporal residency effect. Assessing the effect is necessary to support decision-making on ecosystem and land use management, which locating and valuing the effect spatio-temporally are core tasks. However, the assessment is difficult due to data constraints and thus needs a comprehensive solution.

In US, parcel data is the public data containing residency information that are collected and managed by local governments, e.g. counties. The parcel data are the available spatial individual plots indicating where residents live and act (Manson *et al.*, 2009), which contain spatio-temporal residency information for tax collection (Jarosz, 2008, Swanson and Tayman, 2012). Fundamentally, the parcels are representative for individual home ranges with the attributes of geographic identity, land use type, land-associated business type, land/house value, and history. Thus, they reflect the realistic residency boundaries with a fine resolution. Because of their fine resolution, the parcel data have been used to address issues spatial-explicitly in details, e.g. on land fragmentation (Shrestha *et al.*, 2012), socio-ecological metabolism (Hall, 2011), land-cover fragmentation and configuration (Robinson, 2012), carbon storage potential (Visscher *et al.*, 2014), and segmenting private landowner motivations, properties, and management (Nielsen-Pincus *et al.*, 2015). Nevertheless, due to lacks of human information, the application of parcel data for assessing the effects of human residency has been limited.

In a different approach, census data are provided as the public data about human population supplied by the US Census Bureau with resolution levels of state, county, county subdivision, place, census tract, block group, and block (Seidelmann and Merry, 2007). The

primary use of census data is for issues of population. As combined with ecological data, census data can be used for investigating the interactions between human and ecosystem. For example, census data may be linked to ecological records to assess human effects on ecosystems using agent modeling (Smajgl *et al.*, 2011, Rounsevell *et al.*, 2012), human impacts on land resource (Erb, 2012), and urban vulnerability to extreme heat (Wilhelmi and Hayden, 2010). The applicability of census data for human impact assessment depends on their information resolution levels. For instance, the census tract data can be used for neighboring socio-economic analysis (Pickett *et al.*, 2002, Schuurman *et al.*, 2008) as outcomes are expected with a broadly spatial spectrum. For a research targeted with a fine information resolution, the census block group is often used. A census block group level may help provide spatial results with more details than the census tract level does. The application of census block group data is diverse, e.g. assessing microclimate and vulnerability to heat stress (Harlan *et al.*, 2006), socio-ecological dynamics of urban residential landscapes (Roy Chowdhury *et al.*, 2011), community evaluation (Aronson *et al.*, 2007), conservation valuation of residential yards (Lerman and Warren, 2011), decadal progress of urban ecological systems (Pickett *et al.*, 2011), and local level vulnerability (Heaton *et al.*, 2014). One challenge is that census data are collected for the purpose of summarizing population attributes and thus they have a limitation to reflect human decisions, e.g. human perceptions (Wilhelmi and Hayden, 2010). Moreover, census information resolution depends much on census cell sizes, which their large cell sizes may not represent the realistic interaction between human and ecosystem. These problems may cause difficulties for residency studies.

In an integrated direction, several studies tried combining parcel and census with an assistance of environmental and ecological data. There are several applications for this

integration, e.g. for land use studies of the conversion of agricultural to residential use (Polimeni, 2005), the effects of land changes on residence tree cover (Landry and Pu, 2010), land use change (Zhou and Kockelman, 2008), agricultural conversion to residential use (Polimeni, 2005), and effects of land development regulation on tree cover in residence areas (Landry and Pu, 2010). Other applications include the optimization for urban water use sectors (Friedman *et al.*, 2014), interrelationships between physical and social environment of discarded needles (de Montigny *et al.*, 2011), characterization of urban vulnerability to heat stress (Heaton *et al.*, 2014), and interaction between public and private for community gardens and services (Knigge, 2009). The residency effect information can be found in such research but under a defragmented form in terms of its spatial distribution and effect types.

The census and parcel data themselves are the public data and thus exploring their applicability to residency impact assessment can be very supportive to management and decision-making. Recent studies have indicated a lack of evidences of using the combined census and parcel data for understanding spatial residency human impacts. Thus, as a contribution to residency research, this study developed and use procedures and a simulation model (using R programing language) to combine the census and parcel data to investigate issues of the human residency. Pocatello, Idaho, was selected as the case study to demonstrate the applicability of the developed methods and at the mean time to (1) assess land service associated with flood control and (2) locate/assess the spatial residency effect, which are crucial research demands.

Methods

Study site and situations

The research site is in Pocatello, Idaho (Figure 1), that includes Pocatello city and Chubbuck city with the population of 54,292 and 14,229 in 2014, respectively (US Census

Bureau, 2015). Pocatello belongs to the southeast of Idaho (Jacox and Bauer, 2015), which its transition is from a transportation hub. The transition of Pocatello results in an urbanization with a spatio-temporal residence expansion.

(Figure 1)

Similar to urbanization in other cities, the development of Pocatello is associated with residency needs. Currently, it has two crucial residency issues including decision-makings for a concrete channel and a residency effect. The concrete channel issue is related to the Portneuf River, which flows through Pocatello (Marcarelli *et al.*, 2010). Historically, for flood protection, the channel was completely constructed in 1968 to replace partially the river part crossing out Pocatello to protect residence areas along the river (Eppley, 2015). As a result, this impact forms two groups of residence areas neighboring to each other in the same geographical area; one is along the concrete channel while the other is along the river (Figure 2). Decisions for the concrete channel are difficult because stakeholders have had different points of view. The discussions among the stakeholders have led to three proposed solutions consisting of enforcement, removal and reconstruction with greening, and complete removal for the concrete channel (Stoutenborough, 2015). To come up with a comprehensive solution for this issue, besides economic and policy issues, assessing residency asset values associated with the concrete channel has been requested.

The next issue is the demand of identifying the spatio-temporal residence human effect on ecosystem for the entire Pocatello to support land use management regarding ecosystem protection. This issue is complex since it is related to the development of Pocatello with human settlement (Puffett and McNary, 1977, Torregrosa and Devoe, 2008). Dating back to history, Pocatello has been known having a transition to urbanization, which recently has residency

issues, e.g. the river water polluted of nutrients (Hopkins *et al.*, 2011). Due to residency problems, the residency effect assessment has been demanded.

Data mining

A data mining procedure was established and applied to obtain proper data from the parcel and census data for analysis. This process includes an analysis and assessment of data structure and pattern (organization, classification, and sampling) to extract the desired data supporting the study objectives. The mining procedure has the following steps

- Step 1: determine attributes of the census and parcel data that are potential to support research objectives. This step was related to the study purpose and then any information associated with residency was identified and gathered
- Step 2: evaluate the applicability of the data gathered from the step 1 in which data attributes were narrowed to match the study scopes
- Step 3: select the desired data with preferred attributes from the data sets obtained from the step 2.

Economic valuation for land residency service

Economic valuation was applied to assess the potential effect of flood control on residence land using the parcel data solely (case study 1). The residence parcels include land and house values. Because house properties are different and can be modified by residents for their needs, house values are not stable by time. In contrast, land structure is stable by time. Thus, within the same area with the same land use purpose for residence, land value is applicable for use to assess the residency effect under the indicator of land value per area unit (LAND). Two

residence areas next to each other within the same geographic elevation were selected for the valuation using the following experimental design (Figure 2)

- Experiment A: comparing LAND of the parcels within the following distances from the concrete channel: within 100m, between greater than 100m and 200m, and between greater than 200m and 300m (Figure 3). The purpose was to assess how LAND spatially distributed regarding distances from the parcels to the concrete channel. Because the width of the residence area adjacent to the concrete channel is about 300 m, the boundary of 300m was selected to reduce result uncertainty due to the effect of mountainous conditions on LAND because beyond this boundary has several parcels adjacent to a mountain (Figures 2 and 3).

(Figure 3)

- Experiment B: comparing LAND of the parcels within distances from the river: within 100m, between greater than 100m and 200m, between greater than 200m and 300m, and between greater than 300m and 500m (Figure 4). This experimental design was to understand how LAND spatially distributes within distances from the river. This study selected the boundary of 500m to avoid the effect of mountainous conditions on LAND because several parcels are adjacent to a mountain exceeding the 500m limit (Figures 2 and 4).

(Figure 4)

- Experiment C: comparing LAND of the parcels in an area within 100 m from the river to that from the concrete channel (Figure 5). These areas were selected because they were the most adjacent to the effect source of either the river or the concrete channel, which are the most sensitive to the impacts of the source. This case was

aimed at revealing the difference of the residency land service of the parcels between
along the river and the concrete channel.

(Figure 5)

Integration of census and parcel data

The census and parcel data were integrated to model the spatio-temporal distribution of
human residency effect (case study 2). This study developed and applied a model using the R
programming language to assess the human residency effect. The integration concept is
considering the parcel cells as a baseline for the integration because they represent realistic
residency home ranges. This integration works on data information, time stamp, type, structure,
and cell size.

For the data information, a data evaluation was applied to find residency information that
both data contained. The residence parcels were selected as

$$ParcelData = Residence\ Building \cap YearBuilt \cap LandValue \cap GISArea \quad (Eq.1)$$

where ParcelData is the desired residence parcels, Residence Building is the residence type
among parcel building types including residence, agriculture, and commercial (residence
complexes and public residence buildings were excluded), YearBuilt is the year built of
buildings, LandValue is the land values, and GISArea is the area.

For the block group data, this study selected the census cells containing the attributes of total
house units for various occupancy statuses representative for the residency effect. They included
the state of being occupied (indicating the current settlement of residents or a current residency
effect) and vacancy (showing the potential residency effects) that included five house states of

“House-for-rent”, “House-for-sale”, “House-sold-but-not-occupied”, “House-rented-but-not-occupied”, and “House-for-seasonal- recreational-or-occasional”.

For the time stamp, data was prepared using a grouping method. The US census block group data from five-year estimates that can provide yearly data from 2010 to 2013 (<https://www.census.gov/geo/maps-data/data/tiger-data.html>) were selected. Because the parcel data contain timestamps, they can be grouped timely as desired, e.g. yearly, five-year, or ten-year. This study grouped the parcels yearly to match with the time stamp of the census data.

For data type and structure, the census and parcel data were managed differently. While the parcel data shapefile contained the residency attribute, the shapefile of census data for 2013 prediction but other periods had house occupancy attributes created in a separate table. Thus, the developed model included a joining step to add the occupancy attribute for 2013 to the shapefile.

For the cell size, both data were integrated by the model using a computation process. It is noted that this integration process uses a raster calculation instead of overlaying the census over the parcel data because the overlaying may result in unreasonable spatial cells due to the difference between census and parcel data. Using the raster approach, the developed model from this study allows a selection of raster resolution as desired. The integration procedure is

Step 1: create census raster maps

- Create census block group raster maps by the desired attributes of residency status for periods of 2010, 2011, 2012, and 2013
- Reclassify the census raster maps, where all cell values greater than “0” are replaced by “1” for each map. This step is to create the raster maps with two information levels, “1” and “0” are for house availability and unavailability, respectively

Step 2: create parcel raster maps

- Select the residence parcels with the same time periods as those of the census data to create new parcels maps and then, for each map, adding a new field in which its values are assigned with “1” for all cells to locate residence areas and residence boundary (a source of residence human effect and its boundary). The equation for selecting the parcel time periods is

$$ParcelTime \leq CensusTime \quad (Eq. 2)$$

where ParcelTime is the parcel time periods to be matched with CensusTime as the census time periods. For example, for the census occupancy status of houses in 2010, the match time of parcel data is the year built of residence houses in 2010 or earlier to locate all housing parcels in 2010.

- Convert the residence parcel maps to parcel raster maps

Step 3: multiply the census raster maps by the parcel raster maps to obtain new raster maps of the distribution of residence effect. The multiplication created new raster cells with two values of “0” and “1” representative for residency statuses. This process is to reduce the uncertainty of residency boundaries in census data and add the house occupancy information to the parcel data to create new maps with residency information with a fine resolution.

Data screening

A data screening was applied for parcel and census data. The metadata of both data were checked to ensure they agreed in coordinates, reference system, and units. For the case study 1, the parcels were selected in the same geographic areas (belonging to the same river/concrete

channel side and having the same geographical elevation). The information of public residence land and private residence land for building complexes were excluded using an outlier removal method. The selected parcels are also not adjacent to parks or schools or mountains. The goal of this screening step is minimizing result uncertainty due to the influences of parks, schools, geographic conditions on land value and data biases. For the case study 2, census and parcel data are descriptively analyzed to identify outliers and remove them as needed.

Analysis and modeling tools

For both objectives, ArcMap (version 10) and R programming language (R Core Team, 2013) with R packages of "maptools" (Lewin-Koh *et al.*, 2011), "rgeos" (Bivand and Rundel, 2013), "raster" (Hijmans and van Etten, 2012), "rgdal" (Bivand *et al.*, 2013), "ggmap" (Kahle and Wickham, 2013), and "foreign" (R Core Team, 2014) were used. Data normality was tested using a Shapiro-Wilk test and used median comparisons between data samples using Wilcoxon tests for data without normality.

Datasets

This study used the available public census data for 2010, 2011, 2012, and 2013 from five-year estimates for 2006-2010, 2007-2011, 2008-2012, and 2009-2013 by American Community Survey, respectively (United State Census Bureau-TIGER/Line, 2015, 2015a, 2015b, 2015c). The parcel data in 2014 was obtained from Bannock County (Bannock County, 2014), which contain timely information of land and house. The parcel data with a ten-year period were used solely for the valuation of land service and the identification of residency effect distribution while those with an annual period were combined with the census to assess residency effects.

Results

Case study 1—Assess residency land values associated with flood control

Study results showed various effects of the concrete channel under LAND indicator. The outcome from this study indicated that the distance from home ranges to the river or concrete channel does not likely influence differently on land values per area unit spatially as tested at a confidence level of 95% (Experiment A and B). For the case of the concrete channel, LAND was 29.06 USD/m² spatially (Table 1). In contrast, the spatial LAND for residence areas along the river is 32.29 USD/m² (Table 2). These results informed that given the same geographic elevation, land use purpose, and effect source (river or concrete channel), the spatial response of the residence land service under LAND is likely the same. It is noted that LAND is the human preference regarding a free market (since public land was excluded) and thus reflects the perception of residents. As such regards, the study outcome showed a trend of the preference of the residents that they placed the same values on LAND spatially responding to the effect source, either the river or the concrete channel.

(Tables 1 and 2)

In another assessment, study results reported the different values of LAND in the area adjacent to the concrete channel from those for the river case. LAND in the area proximate to the concrete channel is likely less than that to the river at a confidence level of 95% (Table 3) (Experiment C). As indicated in the experiment A and B, these two areas had the same spatial LAND regarding either the river or the concrete channel. In addition, they are adjacent to each other and have the same geographic elevation and land use purposes. The potential effects such as school, park, and public land were also excluded. The assumption in this study was

environmental conditions are the same in these two areas, which is reasonable because they are located in the same environment, neighboring to each other, and there has not found any report stating the difference of environmental conditions between them, e.g. water and air quality (Figure 2). Thus, results indicated a trend that LAND in the concrete channel is less than that in the river case. These results also suggest that decision-making for the fate of the concrete channel should be concerned to the effect source (the concrete channel or the river).

(Table 3)

The outcome from this study is similar to other cases that often happened for flood protection projects that replaced natural water systems by the man-made levees with a result of costs. For example, Nguyen and Tran (2009) reported that the residence areas with flood protection dikes had lower net present values (combining socio-economic and environmental valuation) than those with more-natural systems.

Linked to historical records, this study also reported a transition of human perception on residency related to flood control. For example, in Florida, due to the demand of flood control to protect residence areas, levees were developed, which altered ecosystems and landscapes (Haman and Svendsen, 2006). Later on, the new public demand of hydrologic restoration has raised and the restoration, with the removal of levees, has implemented (Light and Dineen, 1994, Haman and Svendsen, 2006). This study found a similarity of such transition of the public demands in Pocatello. The flood control with a replacement of the river part by a concrete channel was requested for the flood control. The situation now has changed that there are raising public opinions on the demands of the removal of the concrete channel as described by Stoutenborough (2015). These showed that the residency land service is a historical issue associated with changes of human demands and perceptions.

Technically, results also showed that the procedure using the parcel data with home range boundaries is applicable to address residency land service (an effect of residence on land resource). The parcel land use values, the data records on land ownership (Brownson *et al.*, 2009), showed their competency to reflect the preferences of the public in market. This study helped reveal the preferences of residents, which was often unknown (Stoutenborough, 2015), as an useful information to support decision-making. Results from this study not only revealed the difference of land service values regarding the impact sources of the concrete channel and river but also assisted in informing residents' decisions. It is noted that the land values were provided by counties, e.g. Bannock County in this study. Thus, any decision-making regarding LAND should conduct additional surveys and interviews, e.g. interviewing local people directly, to understand their preferences more.

Case study 2-Assessing the spatio-temporal residency effect

Residency effect distribution

Results from using the parcel data in this study provided a picture of the transition of the residency effect on ecosystem spatio-temporally. From combing historical and spatial analyses, results indicated that the spatio-temporal residency transition in Pocatello was due to house demands. This transition was similarly reported by Jacox and Bauer (2015). In 1939 and earlier, Pocatello was as a transportation hub that needed workers. When settling in Pocatello, the workers raised their demands of residence houses. Thus, houses were gradually developed. After the World War I, house shortages occurred in Pocatello and later in 1945s; this made Pocatello become more urbanization. The other factors for the house demand's increase are the development of Idaho State University and other businesses. Of the factors, land use planning and management was a crucial driver for the transition.

Study results from spatial analyses also showed that the residency transition varied spatio-temporally. The residency distribution was in the North-West direction (with a wide expansion for the West-East direction) and Southeast direction (with a narrow expansion for the West-East direction) (Figures 6, 7, and 8). It was found in this study that the narrow West-East distribution southward of Pocatello was due to the geographical elevation restriction (hills on either side of residence areas). Our results are similar to the report of McMahan et al. (2002). In other words, the geographical condition, e.g. elevation, is a crucial factor affecting the distribution of the human residency transition. Results also indicated that the residency source distributed initially at the center of Pocatello city and Chubbuck city and later in cluster.

(Figures 6, 7, and 8)

Residency effect

This study noted that the results in Figures 5, 6, and 7 were from the parcel data solely that contained current and potential states of residency because the parcels did not have the information of occupancy status of houses in residence areas. In more details of residency, by combining the parcel data and the house occupancy information in the census data, results provided various residency types that reflect a variety of the residency effect on ecosystem. Results revealed the various residency effect types of spatio-temporal house occupancy status with a fine resolution (Figures 10-15). For instance, “House-occupied” status indicated the current presence of residents, which mean a current effect on ecosystem. “House-for-sale” and “House-for-rent” are uncertain states of the current residency. “House-sold-but-not-occupied” and “House-rented-not-occupied” obviously have no present residency and thus have no current effect on ecosystem.

Results showed the variation of residency states by time. The “House-for-sale” state decreased temporally, which indicated a less trade-off for sale by time. In other words, the availability for a house move-in under the sale status reduced and then the residency effect for this category declined. The similar situation of the decreasing was found for the state of “House-sold-but-not-occupied” (Table 4). In a meantime, there was a significant increase of total houses in 2013 compared to the previous years, which indicated the increasing effect for other categories. Results also showed a trend of housing, which potentially assists in quantifying the carrying capacity of residence under house status to assess the transition of the residency impact. Therefore, it is applicable to estimate where new move-in persons (internally from inside Pocatello or externally from other cities) may locate and, for extent, assess the level of residency impacts.

(Table 4)

Study outcome reported that the types of house vacancy may help determine levels of human impacts on ecosystem for new move-ins. As new people move in houses, they may have various impacts on ecosystem depending on their vacancy status. Since the status of house occupancy represents human existence, their occupancy levels may result in short-term, long-term, or uncertainty impacts. Results showed the potential impact levels for new move-ins (Table 5). For “House-rented-not-occupied” and “House-sold-not-occupied”, there are fewer chances for new people to move in while more chances may occur in the areas with “House-for-rent” and “House-for-sale” statuses.

(Table 5)

Benefits of combined census and parcel data for residency effect assessment

Results from this study demonstrated the benefits of integrating the census data to the parcel data to assess the residency effect, which the parcel home ranges are used as a baseline. Each data had their residency information and as they are combined, they complement to each other. For instance, the parcel data only contain the spatial residency boundaries under home ranges without occupancy statuses. As a modification for the parcel data, the census data have the information of occupancy statuses under different categories as discussed above but their cell sizes are large. Thus, the integration helped create a result including both residency information carried from them, which included both fine spatial residency boundaries and occupancy statuses (residency effect distribution).

The outcome from this study provided the realistic residency effect boundaries under individual home ranges in various states of the residency effect regarding occupancy statuses. The developed procedures from this study significantly assisted in solving problems previous studies indicated. For instance, Cho (2014) indicated that mismatching of time-periods may be a problem for the integration. As a solution, this study took the advantage of the timestamp information of the parcel data and then was able to select the periods as desired. Particularly, results emphasized that the parcel data can help to assess the residency effects yearly because they contain yearly time stamp data. Using this integration approach with a flexible time arrangement helped the parcel data agreeing with the census data about time stamps. In another problem, Abraham et al. (2015) reported that the parcel and census block groups often do not match on boundaries and may create errors for data allocation. Indeed, in the case study, the census block group cell sizes were larger than the parcel cell size that caused difficulties for the integration; which this problem was solved by using the parcel cells as a spatial baseline using

the developed model in this study. The next problem is that the census cells may contain unrealistic data information at a sub-cell level, e.g. census sub-cells may not contain values that census cells carry. For example, the northeast census cell contains the data of availability of “House-for-rent” in 2013 but, at a census sub-cell look, some areas inside that census cell do not contain residence houses as proved from the parcel data (Figure 9). Similarly, the census cells contain the state of “House-for-rent” that as matched with elevation data it appears that this state occurs on hills, which is not reasonable (Figure 9). The developed procedure used the residence information of parcels to validate the census cells to remove the spatial outliers to provide the comprehensive results. As assistance, the integration approach developed in this study significantly helped reduce uncertainty of the information of locating the spatial residency effect (Figure 9).

This study also recognized several uncertainties for assessing the residency effect. First, the census data are the five-year estimates and thus result uncertainties depend on the uncertainty of census sampling and estimation. This uncertainty may similarly happen for the parcel data. Second, parcel and census data are collected and managed by different governmental offices and thus the dissymmetry may occur. This research set the parcels as the baseline but they may contain uncertainties of spatial boundaries due to data sampling and zoning. These uncertainties should be considered.

Conclusions

The combination of census and parcel data helped reveal the spatial residency effect distribution, in terms of location, boundary, and types. Using the parcels as a baseline helped reveal monetary values of residency land reflecting preferences of local community. The point-source of either the concrete channel or the river has the same influence on LAND spatially.

LAND in the residence area affected by the concrete channel was found less than that for the river case. These results are useful for decision-making that this study suggests a consideration of preferences of local community on LAND and perhaps additional interviews or surveys about LAND should be additionally conducted to assist in decision-making for the concrete channel. For extent, because parcels are often consistent and updated as an accurate and detailed of resident's house/land (Swanson and Tayman, 2012), they are applicable for residency impact assessment as well as residence and land use planning/management.

The integration procedure from this research successfully and comprehensively combined census and parcel data to assess the residency effect. Results provided various types residency effects with a comprehensive and fine information resolution. The results also reveal the states of resident effect, which can be used for land use and ecosystem management. Although census and parcel data are very different, particularly data structure and cell units (Leyk *et al.*, 2012), this study demonstrated the applicability and benefits of combining them for residency assessment.

The data mining approach developed in this paper helped assess the residency effect regarding flood control, determine its distribution, and detect its impact levels. Since parcel data may contain different types of property building, e.g. residence, commercial and agricultural, as combined with census data, they can potentially help to determine and assess various effects on ecosystem beyond the residency. Particularly, in this study, the data mining approach showed its competency to determine and model the effect with a fine resolution. The study approach is applicable to assist further research for tracking point-sources of pollutants from residence (or other types of human activity depending on availability of the attributes of census and parcel data). The application may be extended to environmental data collection and experimental designs to reduce uncertainty using spatial maps with large cell sizes.

This study suggests the mining should be conducted properly to maximize the benefits of parcel data, in terms of relating available parcel attributes to residency issues. In addition, data screening should be always implemented to reduce data uncertainty. The application should consider uncertainties of census and parcel data due to their sampling, management, and estimation. While parcel data are updated yearly, the census data may be delayed for public releases, which may cause a time-lag for research.

Acknowledgement

This study was given supports from the MILES project (Managing Idaho's Landscapes for Ecosystem Services) at Idaho State University (the NSF Idaho EPSCoR Program and the National Science Foundation under award number IIA-1301792) for funding; Dr. David W. Rodgers, Dr. Shannon Kobs-Nawotniak, and Dr. Donna Delparte for advices; University of Florida for a permission to access online library resources; and An Giang University for comprehensive supports.

References

- Abraham, J.E., Weidner, T.J., Gliebe, J., Willison, C. & Hunt, J.D., 2015. *Synthesizing building space data for integrated land use transport models* [online].
<http://128.40.111.250/cupum/searchpapers/papers/paper368.pdf> [Accessed Access Date 2015].
- Aronson, R.E., Wallis, A.B., O'campo, P.J., Whitehead, T.L. & Schafer, P., 2007. Ethnographically informed community evaluation: A framework and approach for evaluating community-based initiatives. *Maternal and Child Health Journal*, 11, 97-109.
- Bannock County, 2014. Bannock County Parcel Data. In I. Bannock County (ed.) Bannock County, ID: MILES project (Managing Idaho's Landscapes for Ecosystem Services).
- Bivand, R., Keitt, T., Rowlingson, B., Pebesma, M. & Summer, M., 2013. rgdal: Bindings for the geospatial data abstraction library. *R package version*, 1.
- Bivand, R. & Rundel, C., 2013. rgeos: interface to geometry engine-open source (GEOS). *R package version 0.2-13*.
- Brownson, R.C., Hoehner, C.M., Day, K., Forsyth, A. & Sallis, J.F., 2009. Measuring the Built Environment for Physical Activity: State of the Science. *American Journal of Preventive Medicine*, 36, S99-S123.e12.

- Cho, S.H., Kim, S.G., Roberts, R.K., Lambert, D.M. & Kim, T., 2014. Effects of Land-Related Policies on Land Development during a Real Estate Boom and a Recession. *Growth and Change*.
- De Montigny, L., Vernez Moudon, A., Leigh, B.C. & Kim, S.-Y., 2011. A spatial analysis of the physical and social environmental correlates of discarded needles. *Health & Place*, 17, 757-766.
- Eppley, J., 2015. *Exploration of Pocatello's flooding history* [online]. Idaho State University. Available from: <http://miles.isu.edu/Timeline/Overview/> [Accessed Access Date 2015].
- Erb, K.-H., 2012. How a socio-ecological metabolism approach can help to advance our understanding of changes in land-use intensity. *Ecological Economics*, 76, 8-14.
- Friedman, K., Heaney, J.P., Morales, M. & Palenchar, J., 2014. Analytical optimization of demand management strategies across all urban water use sectors. *Water Resources Research*, 50, 5475-5491.
- Hall, M.H.P., 2011. A preliminary assessment of socio-ecological metabolism for three neighborhoods within a rust belt urban ecosystem. *Ecological Modelling*, 223, 20-31.
- Haman, D.Z. & Svendsen, M., 2006. Managing the Florida Everglades: Changing values, changing policies. *Irrigation and Drainage Systems*, 20, 283-302.
- Harlan, S.L., Brazel, A.J., Prashad, L., Stefanov, W.L. & Larsen, L., 2006. Neighborhood microclimates and vulnerability to heat stress. *Social Science & Medicine*, 63, 2847-2863.
- Heaton, M.J., Sain, S.R., Greasby, T.A., Uejio, C.K., Hayden, M.H., Monaghan, A.J., Boehnert, J., Sampson, K., Banerjee, D., Nepal, V. & Wilhelmi, O.V., 2014. Characterizing urban vulnerability to heat stress using a spatially varying coefficient model. *Spatial and Spatio-temporal Epidemiology*, 8, 23-33.
- Hijmans, R.J. & Van Etten, J., 2012. raster: Geographic analysis and modeling with raster data. R package version 2.0-12.
- Hopkins, J.M., Marcarelli, A.M. & Bechtold, H.A., 2011. Ecosystem structure and function are complementary measures of water quality in a polluted, spring-influenced river. *Water, Air, & Soil Pollution*, 214, 409-421.
- Jacox, E. & Bauer, B.P., 2015. *Old Pocatello* [online]. TAG Historical Research & Consulting. Available from: <http://sspa.boisestate.edu/idaho/exhibitions-tours/idaho-issues-online/cities/old-pocatello/> [Accessed Access Date 2015].
- Jarosz, B., 2008. Using assessor parcel data to maintain housing unit counts for small area population estimates. *Applied demography in the 21st century*. Springer, 89-101.
- Kahle, D. & Wickham, H., 2013. ggmap: A package for spatial visualization with Google Maps and OpenStreetMap. R package version 2.3. See <http://CRAN.Rproject.org/package=ggmap>.
- Knigge, L., 2009. Intersections between public and private: community gardens, community service and geographies of care in the US City of Buffalo, NY. *Geographica Helvetica*, 64, 45-52.
- Landry, S. & Pu, R., 2010. The impact of land development regulation on residential tree cover: An empirical evaluation using high-resolution IKONOS imagery. *Landscape and Urban Planning*, 94, 94-104.
- Lerman, S.B. & Warren, P.S., 2011. The conservation value of residential yards: linking birds and people. *Ecological applications*, 21, 1327-1339.

- Lewin-Koh, N.J., Bivand, R., Pebesma, E., Archer, E., Baddeley, A., Bibiko, H., Dray, S., Forrest, D., Friendly, M. & Giraudoux, P., 2011. maptools: Tools for reading and handling spatial objects. *R package version 0.8-10*, URL <http://CRAN.R-project.org/package=maptools>.
- Leyk, S., Buitenfield, B.P., Nagle, N.N. & Stum, A.K., 2012. Landcover Data to Derive Related Ancillary Variables for Dasymetric Modeling.
- Light, S.S. & Dineen, J.W., 1994. Water control in the Everglades: a historical perspective. *Everglades: The ecosystem and its restoration*, 47-84.
- Manson, S.M., Sander, H.A., Ghosh, D., Oakes, J.M., Orfield Jr, M.W., Craig, W.J., Luce Jr, T.F., Myott, E. & Sun, S., 2009. Parcel data for research and policy. *Geography Compass*, 3, 698-726.
- Marcarelli, A.M., Kirk, R.W.V. & Baxter, C.V., 2010. Predicting effects of hydrologic alteration and climate change on ecosystem metabolism in a western US river. *Ecological Applications*, 20, 2081-2088.
- Mcmahan, J.B., Weber, K.T. & Sauder, J., 2002. Using Remotely Sensed Data in Urban Sprawl and Green Space Analyses. *Intermountain Journal of Sciences*, 30-37.
- Nguyen, T.T.N. & Tran, N.T.H., 2009. Socio-economic and environmental impact assessment in agricultural cultivation, case studies in rice cultivation and shrimp farming in the Mekong River Delta, Vietnam. *Environmental Impact Assessment, Korean Society of Environmental Impact Assessment*, 18.
- Nielsen-Pincus, M., Ribe, R.G. & Johnson, B.R., 2015. Spatially and socially segmenting private landowner motivations, properties, and management: A typology for the wildland urban interface. *Landscape and Urban Planning*, 137, 1-12.
- Pickett, K.E., Ahern, J.E., Selvin, S. & Abrams, B., 2002. Neighborhood socioeconomic status, maternal race and preterm delivery: a case-control study. *Annals of epidemiology*, 12, 410-418.
- Pickett, S., Cadenasso, M., Grove, J.M., Boone, C.G., Groffman, P.M., Irwin, E., Kaushal, S.S., Marshall, V., Mcgrath, B.P. & Nilon, C.H., 2011. Urban ecological systems: Scientific foundations and a decade of progress. *Journal of Environmental Management*, 92, 331-362.
- Polimeni, J.M., 2005. Simulating agricultural conversion to residential use in the Hudson River Valley: scenario analyses and case studies. *Agriculture and Human Values*, 22, 377-393.
- Puffett, W.P. & McNary, S.W., 1977. *Status of Mineral Resource Information for the Fort Hall Indian Reservation, Idaho*: US Bureau of Indian Affairs.
- R Core Team, 2013. R: A language and environment for statistical computing. R Foundation for Statistical Computing.
- R Core Team, 2014. Foreign: read data stored by minitab, S, SAS, SPSS, Stata, Systat, Weka, dBase,... R package version 0.8-59.
- Robinson, D., 2012. Land-cover fragmentation and configuration of ownership parcels in an exurban landscape. *Urban Ecosystems*, 15, 53-69.
- Rounsevell, M., Robinson, D. & Murray-Rust, D., 2012. From actors to agents in socio-ecological systems models. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 367, 259-269.
- Roy Chowdhury, R., Larson, K., Grove, M., Polsky, C., Cook, E., Onsted, J. & Ogden, L., 2011. A multi-scalar approach to theorizing socio-ecological dynamics of urban residential landscapes. *Cities and the Environment (CATE)*, 4, 6.

- Schuurman, N., Hameed, S.M., Fiedler, R., Bell, N. & Simons, R.K., 2008. The spatial epidemiology of trauma: the potential of geographic information science to organize data and reveal patterns of injury and services. *Canadian journal of surgery*, 51, 389.
- Seidelmann, M.E. & Merry, C.J., Year. Use of Landsat imagery for evaluation of land cover/land use changes for a 30-year period for the Lake Erie watershed. *Proceedings of the ASPRS 2007 Specialty Conference, Ottawa, Ontario*, 1-11.
- Shrestha, M.K., York, A.M., Boone, C.G. & Zhang, S., 2012. Land fragmentation due to rapid urbanization in the Phoenix Metropolitan Area: analyzing the spatiotemporal patterns and drivers. *Applied Geography*, 32, 522-531.
- Smajgl, A., Brown, D.G., Valbuena, D. & Huigen, M.G., 2011. Empirical characterisation of agent behaviours in socio-ecological systems. *Environmental Modelling & Software*, 26, 837-844.
- Stoutenborough, J.W., 2015. *Determinants of Stakeholders' Preferred Policy Solution: Comparing Competing Policy Proposals to Address Degraded Levees* [online]. <http://wpsa.research.pdx.edu/papers/docs/Stoutenborough%20-%20WPSA%20-%20Determinants%20of%20Stakeholders%20Preferred%20Policy%20Solution.pdf> [Accessed Access Date 2015].
- Swanson, D.A. & Tayman, J., 2012. *Subnational population estimates*: Springer Science & Business Media.
- Torregrosa, A. & Devoe, N., 2008. Urbanization and changing land use in the Great Basin. *Collaborative Management and Research in the Great Basin*—, 9.
- United State Census Bureau-Tiger/Line, 2015. American Community Survey 5-Year Estimates-Geodatabase Format/2007-2011 Detailed Tables-Block Group. *In* U.S.C. Bureau (ed.) U.S.
- United States Census Bureau-Tiger/Line, 2015a. American Community Survey 5-Year Estimates-Geodatabase Format/2006-2010 Detailed Tables-Block Group. *In* U.S.C. Bureau (ed.) U.S.
- United States Census Bureau-Tiger/Line, 2015b. American Community Survey 5-Year Estimates-Geodatabase Format/2008-2012 Detailed Tables-Block Group. U.S: United States Census Bureau.
- United States Census Bureau-Tiger/Line, 2015c. American Community Survey 5-Year Estimates-Geodatabase Format/2009-2013 Detailed Tables-Block Group. *In* U.S.C. Bureau (ed.) U.S.
- Us Census Bureau, 2015. *State & County QuickFacts* [online]. US Census Bureau. Available from: <http://quickfacts.census.gov/qfd/states/16/1664090.html> [Accessed Access Date 2015].
- Visser, R.S., Nassauer, J.I., Brown, D.G., Currie, W.S. & Parker, D.C., 2014. Exurban residential household behaviors and values: Influence of parcel size and neighbors on carbon storage potential. *Landscape and Urban Planning*, 132, 37-46.
- Wilhelmi, O.V. & Hayden, M.H., 2010. Connecting people and place: a new framework for reducing urban vulnerability to extreme heat. *Environmental Research Letters*, 5, 014021.
- Zhou, B. & Kockelman, K.M., 2008. Neighborhood impacts on land use change: a multinomial logit model of spatial relationships. *The Annals of Regional Science*, 42, 321-340.

List of figures

Figure 1: Pocatello research site within Idaho adapted from United States Census Bureau-TIGER/Line (2015c)

Figure 2: Research areas for residency land valuation with the parcels, the Portneuf river part, and the concrete channel.

Figure 3: Experiment A-Residence parcels within distances from the concrete channel.

Figure 4: Experiment B-Residence parcels within distances from the Portneuf River.

Figure 5: Experiment C-Residence parcels within 100 m from the Portneuf River and the concrete channel.

Notes: the residence areas along to the river and the channel have some parcels adjacent to a public park in between that were excluded to reduce result uncertainty.

Figure 6: Spatio-temporal distribution of overall residency effect on ecosystem in periods of 1939-earlier, 1940-1949, 1950-1959, and 1960-1969.

Figure 7: Spatio-temporal distribution of overall residency effect on ecosystem in periods of 1970-1979, 1980-1989, 1990-1999, and 2000-2009.

Figure 8: Spatio-temporal distribution of overall residency effect on ecosystem in the 2010-or-later period.

Figure 9: A demonstration of integrating census to parcel data using the case “House-for-rent” in 2013.

Notes: the census map on the top left shows the availability of “House-for-rent” in 2013, the parcel map on the top right shows where houses exist in 2013, and the integration map indicates the validated map with a fine resolution for the availability of “House-for-rent” in 2013.

Figure 10: Spatio-temporal distribution of standing residency.

Notes: “1” indicates the availability and results are from using the “houses occupied” census attribute.

Figure 11: Spatio-temporal distribution of residency effect for “House-for-rent” status.

Notes: “1” is available and “0” is unavailable.

Figure 12: Spatio-temporal distribution of residency effect for “House-for-sale” status.

Notes: “1” is available and “0” is unavailable.

Figure 13: Spatio-temporal distribution of residency effect for “House-for-seasonal-recreational-or occasional” status.

Notes: “1” is available and “0” is unavailable.

Figure 14: Spatio-temporal distribution of residency effect for “House-rented-but-not-occupied” status.

Notes: “1” is available and “0” is unavailable.

Figure 15: Spatio-temporal distribution of residency effect for “House-sold-but-not-occupied” status.

Notes: “1” is available and “0” is unavailable.

List of tables

Table 1: Spatial land values per area unit of the residence parcels within distances from the channel (unit: USD/m²).

Statistics	Land values per area unit			Wilcoxon test
	$0 < D_1 \leq 100\text{m}$	$100 < D_1 \leq 200\text{m}$	$200\text{m} < D_1 \leq 300\text{m}$	p-value
Sample size	122	134	86	
Median	29.06	29.06	29.06	> 0.05

Notes: D_1 is the distance from the residence parcels the concrete channel and p-values were estimated from Wilcoxon tests at a confidence level of 95%.

Table 2: Spatial land values per area unit of the residence parcels within distances from the Portneuf River (unit: USD/m²).

Statistics	Land values per area unit				p-values
	$0 < D_2 \leq 100\text{m}$	$100 < D_2 \leq 200\text{m}$	$200\text{m} < D_2 \leq 300\text{m}$	$300\text{m} < D_2 \leq 500\text{m}$	
Sample sizes	37	58	59	77	
Median	32.29	32.29	32.29	32.29	> 0.05

Notes: D_2 is the distance and p-values were from Wilcoxon tests at a confidence level of 95%.

Table 3: Comparison of land values per area unit of the residence parcels within 100 m from the river to that from the concrete channel (unit: USD/m²).

Statistics	Land values per area unit		Wilcoxon test p-value
	Control (100m from the river)	Test (100m from the channel)	
Sample sizes	37	122	
Median	32.29	29.06	< 0.05

Notes: p-values were estimated from Wilcoxon tests at a confidence level of 95%.

Table 4: Number of house units for vacancy

Year	For rent	Rented, not occupied	Sale	Sold, not occupied	Seasonal
2010	677	72	400	114	296
2011	707	90	305	114	348
2012	789	78	316	77	299
2013	717	102	216	39	246

Table 5: Levels of future effects for new move-ins.

Effect levels	For rent	Rented, not occupied	Sale	Sold, not occupied	Seasonal
Long-term impact	Yes	Uncertain	Yes	Uncertain	No
Short-term impact	Yes	Uncertain	Uncertain	Uncertain	Yes
Overall	Yes	Uncertain	Yes	Uncertain	Yes