Location Matters: Geospatial Policy Analytics over Time for Household Hazardous Waste Collection in California

Kustini Lim-Wavde[1], Robert J. Kauffman[1], Tin Seong Kam[1], Gregory S. Dawson[2]
[1] School of Information Systems, Singapore Management University
[2] School of Accountancy, W.P. Carey School of Business, Arizona State University

Abstract
By integrating mapping and geospatial data into a county-level dataset for exploratory analysis, we will demonstrate how to provide useful insights for waste managers and local governments regarding spatial patterns of household hazardous waste (HHW) collection and how it changes over time. We use map-based visualization to display patterns of spatial intensity and county locations for HHW collection in California from 2004 to 2015. We use exploratory spatial data analytics methods to characterize the spatial distribution of HHW collected per person. When we considered the spatial relationships, we were able to develop and estimate a geographically-weighted regression to explain how different regional factors influence the amount of HHW collected. These factors include demographic characteristics, HHW management policy instruments, and environmental quality enforcement and consideration of these factors are necessary to create a successful recycling program.

Keywords: Choropleth Mapping; Geospatial Policy Analytics; Geographic Weighted Regression; Household Hazardous Waste; Waste Management

Citation:

Acknowledgments: The authors wish to acknowledge California's Department of Resources Recycling and Recovery for kindly sharing data.
Contact: kustini.2012@phdis.smu.edu.sg

1 Introduction

The essence of *sustainability* is human society's capacity to control and manage the waste that it creates so as to not damage the ecosystems and environment that we exist within and which support us. *Household hazardous waste* (HHW) represents leftover household products containing corrosive, ignitable, or reactive materials that can create adverse impacts on the environment and people's health if they are not disposed of properly (U.S. EPA, 2014a).

This research contributes to *sustainability science*, which "*probes the interactions between global, social, and human systems, the complex mechanisms that lead to the degradation of these systems, and concomitant risks to human well-being*" (Springer, 2016). *Geographic information science* is essential in the study of sustainability to develop spatially-oriented models of nature and society. It enables us to bring geographical and other data together, to make sense of the data, and to communicate the results to relevant policymakers (Davis et al., 2011). This research aims to support policy-makers in formulating effective policies, particularly area-oriented environmental policies related to HHW. It also demonstrates the blending – an *analytical fusion* – between machine-based data visualization and explanatory empiricism, that increasingly characterizes computational social science (Chang et al., 2014).

We assess and evaluate the spatial patterns of HHW collection activities and the impacts of the related policies in one state, California, in the U.S. These activities involve state and local governments and participating households (U.S. EPA, 2000). State and local governments provide HHW permanent or temporary facilities to collect, transfer, process, recycle and dispose of waste safely (U.S. EPA, 2013). They perform load-checking for hazardous materials in trash before disposal in landfills and organize HHW collection events, door-to-door HHW collection programs, and HHW-related educational campaigns. The success of waste collection activity depends on the participation of households that identify, segregate, store and transfer the hazardous waste for collection. If the maximum number participate in waste collection programs, then most hazardous material can be diverted so it will not enter landfills with trash or mix with storm water. This will prevent HHW-led contamination.

Studies on recycling and waste management have been mostly done using household survey data, with cross-sectional data analysis or panel data analysis. Some authors have conducted economic work to understand the incentives related to the generation of HHW and recycling (Callan and Thomas, 2006; Morris and Holthausen, 1994; Richardson and Havlicek, 1978; van den Bergh, 2008). Others have examined its non-economic determinants (Bartelings and Sterner, 1999), and rates at which it accumulates and gets recycled (Abbott et al., 2011; Sidique, 2010).

Some studies were performed using county-level aggregated data (Lim-Wavde et al., 2016; Saphores, 2006). An exception is the work of Jenkins et al. (2003), who assessed recycling of waste for 20 U.S metro areas for households. These authors did not consider spatial autocorrelation, nor did they examine other spatial relationships. If there were differential spatial effects, regression models that only handled time-wise but not spatial autocorrelation would have produced less-than-best policy inferences (Voss, 2006). We consider spatial relationships.

Analysis using data with a geographical dimension yields useful insights for different localities. But spatial patterns may change over time as well due to shifts in influential factors or determinants. Thus, exploratory analysis of data with *space-time structure* may reveal interesting spatiotemporal patterns. Understanding them can provide useful input for policy-makers and government officials to evaluate the management of HHW programs.

We ask: (1) Were there key spatial patterns of HHW collection during the study period? (2) Did they change over time? (3) Why did the amounts of HHW collected differ spatially across counties and regions?

The main findings are: (1) There is evidence of spatial patterns in the distribution of HHW collection density across different California regions. We discovered clusters of counties with high HHW collection density in the north and spatial outliers for counties with lower density compared with neighboring counties in Central California. (2) County members of the spatial clusters and outliers changed over the years. (3) The factors for HHW collection had varied influences in the regions. HHW programs seemed to be more cost-effective in the northern than southern region. We attribute these findings to commonalities in large states: a population that is ethnically and economically partitioned into discrete areas in the state, causes substantive differences in recycling practices.

2 Background

Several studies applied spatial data analysis to examine waste management issues. We use them to triangulate toward methods that answer our research questions and support helpful policy analytics uses.

Zhang et al. (2015) analyzed the spatial patterns of *municipal solid waste* (MSW) generation and its determinants in the island city of Xiamen in China. They mapped and analyzed MSW generation quantities and densities using spatial autocorrelation tools in ArcGIS. They found global spatial autocorrelation for MSW generation quantities and densities, and detected MSW hotspots concentrated in industrial zones and downtown areas. They showed that commercial and vacant land areas influenced seasonal and MSW quantities and densities.

There has been other research that has focused on foreign direct investment, international trade, and the concentration of pollution and hazardous waste in places that are most advantageous for such production to occur and the related effluents and waste to be produced. This phenomenon is called the *pollution haven hypothesis* (Cole, 2004), and is associated with changes in the global spatial patterns of manufacturing production (Janicke et al., 1997). The interest in macro-spatial issues related to hazardous waste and the potential for environmental damage is also connected to free trade and the growth of global commerce (Antweiler et al., 2001; Cole, 2000).

Waste mitigation includes reducing the adverse impacts of environmental waste by separating, collecting, processing, and recycling waste, particularly recyclable and hazardous waste. Agovino et al. (2015) analyzed spatiotemporal data in Italy on differentiated kinds of waste that were separated and collected for recycling (organic, packaging, bulky, electronic, and textile waste). They studied *pro-environmental behavior* characterized by a higher rate of differentiated waste to total waste collected. Using exploratory spatial data analysis, they observed an uneven distribution of annual average differentiated waste collection rates, specifically, a cluster of provinces with high rates in northern Italy and another cluster of provinces with low rates in the central and southern parts. They also modeled the spatial and time dynamics of *spatial spillover effects* between contiguous provinces using a spatial Markov transition analysis model, with one region's performance influencing another's. Provinces with best or less-than-best observed pro-environmental behavior influenced those nearby.

Tong and Wang (2005) examined the different spatial patterns of recycling for electronic waste in the rural areas of China. They selected rural districts that were especially influenced by the creation of new infrastructure and manufacturing capabilities to produce electronic products for export in global trade. They noted that the issues with electronic waste are not isolated to China, but are a by-product of the global flows of international trade, and the different kinds of waste associated with them.

Others focused on different forms of hazardous waste in different municipal and geographic settings. Abbott et al. (2011) found that there was substantial variation in household-level recycling of waste in different geographical regions in the U.K. They discovered how different recycling collection policies affected waste recycling rates in 2006 to 2008 in 434 municipal management areas. Another example is the causal and explanatory work of Lasaridi et al. (2009) on regional and municipal differences in waste management costs in Greece in the 2000s. The authors had difficulty capturing accurate data on the economic geography of waste production.

3 Context and Data

For this research, we collected spatial data and public data from government agencies and websites. The study period covers 2004 to 2015 based on California's Department of Resources Recycling and Recovery records (CalRecycle, 2015a, 2015b). It handles hazardous waste collection and recycling program-related data across the state's 58 counties.

3.1 Context: California

California is the third largest state in the U.S., and it has highly diverse urban and rural counties, with varied demographics, income and wealth levels. The Northern California region is comprised of 48 counties and the Southern California region has just 10. Central California includes Fresno, Kings, Madera, Mariposa, Merced, Monterey, San Benito, Stanislaus, Tulare and Tuolumne counties. The level of environmental awareness of its people for recycling practices with solid and hazardous waste is quite advanced.

Many have argued, however, that California is most appropriately seen as a collection of sub-states rather than as a single state. Indeed, a ballot initiative was proposed for a November 2016 vote to divide California into six different states ranging from Jefferson (the northern part of what is currently California) down to Southern California which borders Mexico (Chausee, 2014). Sponsored and funded by venture capitalist, Tim Draper, the argument is that California is simply too diverse and too large to be efficiently governed as a single state. This is not the first effort to subdivide California into smaller states: rather this is simply the latest in over 220 proposals to divide California since its admission to the U.S. in 1850, and at least 27 of the proposals were seen as serious ones (Wood, 2011). In short, California is an excellent natural laboratory to examine the differential effects of recycling decisions on demographically and economically diverse populations.

Public and county-level agencies each year submit data to CalRecycle covering HHW collection and recycling for the period July 1 to June 30. They do this via "Form 303" (CalRecycle, 2015). It captures the HHW collection and recycling quantities for various material categories, collection program types, and disposal methods. We aggregated the data by summing the county collected waste weights, by report year. We also used direct HHW collection costs from fiscal years 2007/2008 to 2012/2013 with county aggregation.

3.2 Data and Variables

The mapping and geospatial data of California and its county boundaries were acquired from GADM (2015) Version 2.8. This dataset covers 58 counties. For the map projections, we used North American Datum of 1983 (NAD83) that was officially adopted by California (Ericksen, 1994).

The counties varied by wealth, geographic area, and the extent to which they were mostly urban or rural (USDA, 2013). Our dataset for the present spatiotemporal research covers only 39 counties, which was dictated by the sufficiency of demographic data during the study period on HHW output. To supplement the data from CalRecycle, we collected county and demographic data from the U.S. Census Bureau's (2012). They include population, household rental rate, household income, and high school graduation rate, among other variables. We collected HHW grant awards data from CalRecycle's (2016) Grants Database. We obtained the number of Resource Conservation and Recovery Act (RCRA)[1] penalties from the Environmental Protection Agency (U.S. EPA, 2016) Enforcement and Compliance History Online (ECHO)[2] and aggregated the data by county and year. (See Tables 1 and 2.)

VARIABLES	DEFINITIONS
Pop	Population in county (people)
MeanHHInc	Mean household income in county (US$ 0,000s)
EduHS%	% population over 25 with high school diploma
Rental%	% of household who rented a house
MCLViol#	# contaminant-level violations in drinking water
Penalty#	# Resource Conservation Recovery Act penalties
HHWGrant$	HHW grant(s) awarded (US$ millions)/person
HHWProgram$	HHW collection costs (US$ / person)
HHWCollD	Quantity HHW collected (in lbs/person)

Table 1. Variables for counties and HHW-related observations

VARIABLES	MEAN	STDDEV	MEDIAN
Pop	982,072	1,716,439	432,637
EduHS%	83%	8%	86%
Rental%	42%	7%	43%
HHInc	$85,750	$26,094	$81,995
HHWGrant$	$0.04/person	$0.16/person	$0/person
HHWProgram$	$0.54/ person	$0.91/person	$0.04/person
MCLViol#	95.28	121.83	41
Penalty#	0.18	0.56	0
HHWCollD	5.28 lbs./person	5.07 lbs./person	3.12 lbs./person
Notes. #Obs: 468, 39 counties, 2004-2015. *HHWProgram$* only includes data from 2007-2012. Some variables were computed via multiple sources.			

Table 2. Description of the county-level variables

4. Analysis

We next discuss the two main methods employed in this research: *spatial autocorrelation analysis* and *geographically-weighted regression.* They constitute the explanatory empiricism portion of this research.

.4.1 Spatial Autocorrelation

We use the *global Moran* (1950) measure for the characteristics of the overall spatial distribution

[1] RCRA governs disposal of solid and hazardous waste. The Environmental Protection Agency (EPA) controls hazardous waste generation, transport treatment, storage, and disposal.
[2] We use RCRA's last penalty date for penalties in a county in a year.

of HHW collection density in California. It is a spatial autocorrelation index:

$$I = \frac{N}{\sum_i \sum_j w_{ij}} \frac{\sum_i \sum_j w_{ij}(y_i - \bar{y})(y_j - \bar{y})}{\sum_i (y_i - \bar{y})^2} \quad (1)$$

where N is the number of spatial analysis units (in our case, the number of counties), y_i (or y_j) is the HHW collection density in county i (or j), $\bar{y}$ is the mean of all observed HHW densities, and w_{ij} is the spatial weight matrix. The value of the index ranges from -1 to 1. Its interpretation is: (1) $I > 0$ indicates a clustering pattern; and (2) $I < 0$ suggests a dispersion pattern. Also (3), if $I \approx 0$, then a random spatial pattern is indicated.

To identify clusters of HHW collection density, we calculated the *local Moran index* for each county:

$$I_i = \frac{y_i - \bar{y}}{S_i^2} \sum_{i=1, j \neq i}^{N} w_{i,j}(y_j - \bar{y}) \quad \text{and} \quad S_i^2 = \frac{\sum_{i=1, j \neq i}^{N} (y - \bar{y})^2}{N-1} - \bar{y}^2 \quad (2)$$

As in Eq. 1, N is the number of spatial analysis units (in our case, the number of counties), y_i (or y_j) again is the HHW collection density in county i (j), $\bar{y}$ is the mean of all observed HHW densities, and w_{ij} is the spatial weight matrix. The interpretation of this index is: (1) if I is significant and positive, county i is associated with the surrounding counties with high densities of HHW, but if (2) I is significant and negative, then county i is associated with surrounding counties with low densities of HHW.

The spatial weight matrix (w_{ij}) is a matrix that defines the relationships among the features in the dataset.[3] California has diverse geography and land areas. Northern California consists of counties with smaller land areas, while the south has counties with larger land areas. If we use contiguity-based weighting, then some counties will have many neighbors and some others will have very few. We decided to use an adaptive distance-based weight matrix. This is calculated using a *bi-square distance function* with an *adaptive distance limit* but a fixed number of neighboring counties. This ensures the same number of neighboring counties when calculating the spatial weight of a pair of counties. The function is discontinuous and excludes observations beyond distance b; the weights decrease as the distance between observed points (d_{ij}) increase.[4]

4.2. Geographically-Weighted Regression

When spatial autocorrelation exists in a dataset, modeling the data using an *ordinary least squares* regression model will not capture the spatial variation across the different areas of study. For better policy inferences with data on the geography dimension for our explanatory empirical approach, *geographically-weighted regression* (GWR) is used to capture spatially-varying relationships between the dependent and independent variables (Brunsdon et al. 1996):

$$y_i = \beta_{i0} + \sum_{k=1}^{m} \beta_{ik} x_{ik} + \epsilon_i \quad (3)$$

The dependent variable y_i is for location or county i, x_{ik} is the value of independent variable k at county i, m is the number of independent variables, and ϵ_i is the error for county i. We estimate the βs for each variable k for county i using a weighted least squares approach. The parameter estimate is:

[3] Generally, there are three different types of weights: contiguity-based, distance-based, and kernel weights. *Contiguity-based weights* only consider the other counties that share the same boundary as their neighbors. *Distance-based weights* are specified using a function of the distance separating the counties; the neighbors can be determined using the k-nearest neighbor criterion or distance bands (or thresholds). *Kernel weights* combine the distance based thresholds together with continuously-valued weight functions, such as bi-square, tri-cube, exponential, or Gaussian functions. We performed sensitivity analysis by calculating the weight matrices using these functions, and we found that bi-square function allowed us to detect more counties in the spatial clusters compared to the other functions.

[4] The function is: $w_{ij} = \left\{\left(1 - \left(\frac{d_{ij}}{b}\right)^2\right)^2 \quad if\ |d_{ij}| < b,\right.$ and 0 otherwise. For a fixed number of neighbors, we needed a large enough sample size to calculate the spatial autocorrelation. We tested the results using 20, 25, 30, and 35 neighbors. Finally, we selected 30 because this enabled us to detect the highest number of counties in the spatial clusters. Distance between counties is by population centroid coordinates. Compared with the geographic county centers, centroids capture spatial autocorrelation and uneven distributions of population in counties of household waste collection better.

$$\hat{\beta}_i = (\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{W}(u_i, v_i)y$$

$\boldsymbol{X}$ is the independent variables matrix with a column of 1s for intercepts; y is the dependent variable vector; $\hat{\beta}_i = (\beta_{i0}, \ldots, \beta_{im})^T$ is the vector of $m + 1$ regression coefficients; and $\boldsymbol{W}$ is the diagonal matrix for geographically weighting the observed data for county i at location (u_i, v_i), which is the county seat coordinate. The weighting is via a kernel weighting function. We use equal spatial weights for all counties for the global regression, and then we apply the weight matrix computed using an adaptive bi-square distance function (as in spatial autocorrelation analysis).

We model HHW collection density as linear in HHW-related policy, policy enforcement, and demographics. We include county demographic factors: the percentage of high school graduates (*EduHS%*), household income (*HHInc*), and the percentage of rental households (*Rental%*) as independent variables. Households with higher incomes have more time to participate in HHW programs. More educated people should be more knowledgeable about HHW risks and can obtain information related to HHW programs. Rental household percentage represents community attachment. Renters will be less attached than owners. More renters imply less HHW collection.

An effective HHW-related policy maximizes household participation in collection programs. We identified two factors that are crucial in the operation of HHW collection programs: HHW collection costs and HHW grant awards. HHW collection costs cover costs required for collection programs from permanent, temporary, and recycle-only facilities, door-to-door collection and curbside programs, load checking, and collection events. The costs are financed by waste parcel and garbage fees, grants from CalRecycle, shared cost with retail partners, hazardous waste surcharges, and sales of recyclable waste. These costs, *HHWProgram$*, should have a positive association with HHW collected per person.

HHW grant awards, *HHWGrant$* per person, are from CalRecycle to help local governments, cities, counties, and local waste agencies responsible for HHW management to establish and expand programs (CalRecycle, 2016). They provide permanent collection facilities with effective HHW public educational programs and campaigns. This funding targets new programs for rural and under-served areas, and expansion of programs in small cities and regions. These grant awards should have a positive association with HHW collected per person.

Policy enforcement influences people's environmental awareness and their actions. In some countries and regions, HHW is banned from the trash. Enforcement influences household behavior in managing their HHW. We could not obtain data to directly measure HHW ban enforcement. So we used proxies for enforcement that may influence household waste management: contaminant level violations in drinking water (*MCLViol#*); and Resource Conservation and Recovery Act (RCRA) penalties (*Penalty#*). More of these represent more enforcement effort. When households perceive policies are enforced, they are more likely to dispose of HHW properly.

For waste generation in system-wide materials management, we must consider consumer product purchases with hazardous content as determining HHW collection. However, it was difficult to obtain this data from public sources. So we rely on the demographic characteristics instead.

5 Results

We next present our spatial visualization and geographically-weighted regression results. These methods demonstrate how machine-based identification of spatial clusters plus explanatory empiricism deliver policy analytics. Note that we used data collected from 58 counties for the global and local Moran's I statistics estimates. This is because we don't need to use demographic data in this analysis. For the geographically-weighted regression analysis, we used data from 39 counties, as described in Table 1 and 2.

5.1 Spatial Visualization

In 2007, the amount of HHW collected in California was about 97 million pounds, the highest from 2004 to 2012. The choropleth map in Fig. 1 shows the *HHW density distribution*. This was created by normalizing HHW quantity in pounds relative to the size of the population in the county. The shaded color class intervals were calculated using the *natural breaks classification method* (Jenks and Capsall, 1971) with R classInt (Bivand, 2015).

From the choropleth maps of HHW collection density from 2004 to 2015, we observed that medium to high densities (8-28+ lbs./person) of HHW collected occurred in Northern California and parts of Central California, particularly the Sierra and Coastal Mountains Regions (Sierra, Nevada, Yuba, Placer, Trinity, Lassen, Inyo). Some counties in the North and Central Coast Regions (Mendocino, Marin, Sonoma) collected moderate HHW amounts (4-13 lbs./person). Counties that collected less than 2 lbs/person were in Central and Southern California (Alameda, Butte, Fresno, Humboldt, Imperial, LA, Madera, Modoc, Mono, Riverside, San Diego, Stanislaus, Tulare). Collection density varied dramatically across areas. In 2014, Alpine County, which did not report any HHW collection before then, collected about 538 lbs./person. This was a very high density of HHW amount compared to other counties.

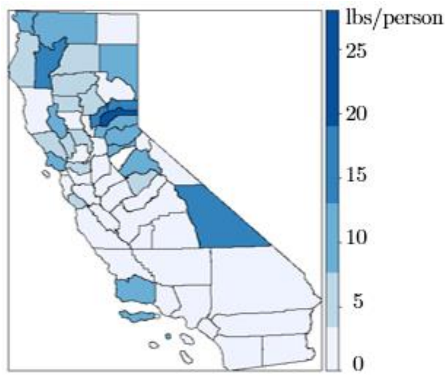

Fig. 1. Choropleth map of HHW collected, 2007

The global Moran statistics confirmed that there were clustering patterns in HHW collection density in California from 2004 to 2015. Table 3 suggests that Moran's I was > 0 for all of our study years. A larger value of the statistic represents a larger clustering pattern, and I was largest in 2011. The size of clusters decreased from 2004-2005, increased in 2005-2006, decreased in 2006-2009, increased in 2009-2011, and then decreased again from 2011-2015.

YEAR	MORAN'S I
2004	0.068***
2005	0.044**
2006	0.179***
2007	0.176***
2008	0.155***
2009	0.129***
2010	0.178***
2011	0.216***
2012	0.131***
2013	0.098***
2014	0.107***
2015	0.108***
Note. Signif.:*** $p < 0.01$, ** $p < 0.05$	

Table 3. Global Moran's I test results for HHW collection density, 2004-2015

The local Moran's I statistic allows us to calculate the index for each county. After calculating them, we created a *local indicator of spatial association* (LISA) *cluster map* (Anselin, 1995). We did so to visualize and identify the locations of spatial patterns and hotspots based on the magnitude of spatial correlation. Fig. 2 gives the LISA clusters for HHW collection density in 2011. We only mapped local Moran's Is if they were at the 5% significance level.

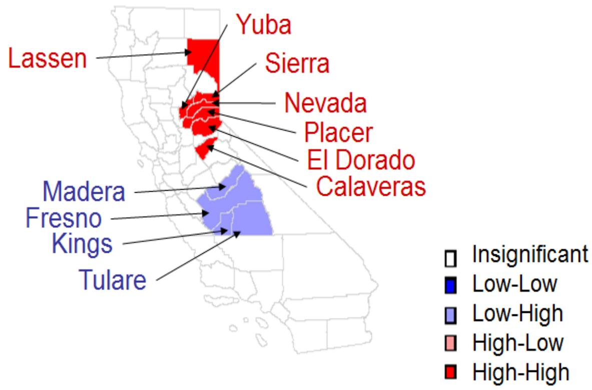

Fig. 2. Indicators of spatial association, 2011

The cluster map is divided into 4 quadrants: High-High (H-H), High-Low, Low-High (H-L), and Low-Low. We observed only H-H and L-H clusters in California. The *H-H quadrant* represents spatial clusters of counties that had high HHW collection density and most neighboring counties had this too. The *L-H quadrant* represents spatial outliers that had low HHW collection density, with most neighboring counties showing higher HHW collection density. The county members of the spatial clusters and outliers changed over the years. Table 4 summarizes the counties in these quadrants.

We observed that Sierra County was the hotspot of HHW collection in 2004 and later. Nevada, Yuba, Placer and a few other neighboring counties also were in the H-H quadrant since 2006. From the choropleth maps, these counties collected at least 8 lbs./person a year. These hotspots were located in the north with a few were in the Central region. However, Fresno, Madera, Tulare, and Kings were outliers. They had lower HHW collection amounts, but their neighboring counties collected somewhat larger amounts per person.

YEAR	HIGH-HIGH	LOW-HIGH
2004	Sierra, Nevada	
2005	Sierra	
2006	Sierra, Nevada, Yuba, Placer, Del Norte, Glenn, Mendocino, Trinity	Fresno, Madera
2007	Sierra, Nevada, Yuba, Placer, Amador, Calaveras, Trinity	Fresno, Tulare
2008	Sierra, Nevada, Yuba, Placer, El Dorado, Amador, Del Norte	Tulare
2009	Sierra, Nevada, Yuba, El Dorado, Calaveras, Del Norte	
2010	Sierra, Nevada, Yuba, Placer, El Dorado, Amador, Lassen, Calaveras	Fresno, Madera, Tulare
2011	Sierra, Nevada, Yuba, Placer, El Dorado, Lassen, Calaveras	Fresno, Madera, Tulare, Kings
2012	Sierra, Nevada, Lassen, Calaveras, Amador, Mono	Tulare
2013	Sierra, Nevada, Lassen, Mono, Yuba	
2014	Sierra, Alpine	
2015	Alpine, Mono	

Table 4. Summary: counties in LISA cluster maps of California's HHW collection density, 2004-2015

5.2 GWR Results

We estimated the GWR model for each year from 2004-2015. Before obtaining the estimates, we calibrated the model based on the datasets in that year because some independent variables in Eq. 3 were unavailable in specific years. For example, HHW grants were not awarded in 2011; program cost reporting was only available in 2007-2012; and there was not much variation in the number of MCL violations in some years. In this way, we were able to assess the effects of all of the influential factors, though not for all years in the sample.

We selected variables in a systematic way using a pseudo-stepwise procedure (Gollini et al., 2015).[5] We used the model specification with the lowest AIC_c for each year. We then ran another procedure to find the optimum bandwidth for the selected model for each year.[6] We estimated the model using equal spatial weights for all counties to get the global regression results. Then, we applied the weight matrix computed using the estimated optimum bandwidth and adaptive bi-square distance function. Appendices 1 and 2 show the regression and GWR estimates.

The global regression results estimate the effects of the influential factors on the HHW collection density without considering the spatial correlations. We found no significant effects for the environmental policy enforcement factor, except for *MCLViol#* at 0.10 ($p < 0.05$) in 2005. We found a positive association for HHW-related policies with the amount of HHW collected per person for some of the years. An example is the estimated effects of HHW grants at 2.33 ($p < 0.01$) in 2008. The coefficients of HHW collection cost (*HHWProgram$*) were 1.35 ($p < 0.10$) in 2007 and 1.28 ($p < 0.10$) in 2010. For the demographics, we found positive effects for education level (*EduHS%*) in 2004, 2007-2008, and 2012-2013, though. *HHInc* was positive and significant in 2005, but was negative and weakly significant in 2007. *RentalHH%* was negative and weakly significant in 2006, but more significant in 2009, 2011 and 2014.

The GWR estimation results gave more insights than the global regression results. We included an F_3 test that indicated whether the range of estimates for each independent variable were significant (varied over counties) (Gollini et al., 2015; Leung et al., 2000). We will only discuss the significant coefficients ($p < 0.01$) as a result.

Fig. 3. *HHWProgram$* effects (2007, 2009, 2010 – left to right)

Fig. 3 shows that HHW program cost effectiveness was better in Northern than Southern California in 2007, 2009 and 2010. Then, they were positive in the north but mostly negative in the south. Fig. 4 shows that *HHWGrant$* in the GWR estimation was positive in 2008 and 2010. The positive influence of HHW grant awards was stronger in Northern California in 2008, but it was much stronger in the south in 2010. The changing spatial relationships may have occurred because of the diverse types of projects proposed and implemented in 2008 and 2010.[7] More projects were proposed for new facilities in the Northern Region in 2008, but more in the Southern Region in 2010. In both years, the positive influence of HHW grant awards in Central California was relatively low.

[5] We ran the *model.selection.gwr* function of the GWmodel package in R. After it finds the best model with the least AIC_c, it resets the independent variables, and introduces a variable for the corrected AIC_c values for each possible model. The *corrected AIC* is similar to the *basic AIC,* but is a function of sample size, and dependent on bandwidth (Hurvich et al. 1998).
[6] We used the *bw.gwr* function from the GWmodel package in R. The optimum bandwidth in the calibrated model was 37 nearest neighbors for all the years, except in 2015, which was 27 neighbors.
[7] In 2008, counties in the north (Butte, Santa Clara, Yuba) and a few counties in the south (Imperial, Los Angeles) proposed new HHW facilities. More drop-off and curbside sites were proposed in both regions. A program was proposed in rural counties (Butte, Glen, Colusa) for 33 new sites for sharps collection, and a partnership for 11 new Universal Waste collection, HHW facility improvements, mobile collection and public outreach programs. In 2010, some counties in the north (Calaveras, El Dorado, Santa Clara) proposed to expand existing HHW facilities. Some other northern counties (Alameda, Shasta, Yuba) proposed partnerships with retail sites to obtain fluorescent lamps and batteries, improve collection, and expand reuse, door-to-door HHW, and sharps programs.

Fig. 4. *HHWGrant$* effects (2008 left, 2010 right)

The influence of environmental policy enforcement, based on the number of RCRA penalties (*Penalty#*), had an unexpectedly negative association with HHW collection density in some years. The impact of penalties was negative in 2005, 2006, 2007 and 2013. (Fig. 5 gives the impacts distribution.) The negative association between this factor and HHW collection density was stronger in Northern than Southern California.

Fig. 5. *Penalty#* effects, 2005, 2006, 2007 – left to right

For the demographics, the effects of *HHInc* were negative to zero with slightly stronger negative influence in the north in 2006. It seems that households with higher income in the north delivered less HHW. *RentalHH%* was negative or positive for all years, except in 2005, 2009 and 2014, when they were all negative. In all the years, the coefficients were higher in Southern California than in Northern California. It may be that the higher percentage of rental households, the lower the HHW collection amount per person in the north. However, it was interesting that it was the opposite in the South.

6 What We Learned

In this section, we first discuss the spatial analysis results for each year during our study period. We then extend our discussion to include the implications of our findings for HHW policy.

6.1 Results and Implications

Our visual and spatial analysis results showed that some Northern California counties had more HHW collected per person. For example, Sierra County exhibited high HHW collection density. The clusters of high-density HHW collection changed over time though, both increasing and decreasing. We found a number of high-density clusters in some counties in the north and a few counties in Central California (Nevada, Placer, Yuba, El Dorado, Lassen, Trinity, Calaveras, Del Norte, Glenn, Mendocino, Inyo).

Consistent with the discovery of clusters of counties with high HHW collection density in Northern California, we found that HHW program cost had a stronger positive influence in the north in 2007, 2009 and 2010 for the GWR results. In contrast, HHW program cost had a negative influence in most counties in the south. So it seems high HHW collection density in the counties in the north was related to more cost-effective HHW programs.

The GWR results also showed HHW grant awards had a positive association with HHW collection density. However, the effects varied across the counties depending on the type of projects were implemented using the grants. RCRA penalties may have discouraged households from participating in HHW collection programs, in contrast to what our intuition suggested. Perhaps enforcement efforts in the form of RCRA penalties were not effective tools to maximize household participation in HHW collection programs. Otherwise, we may need to find other proxies to represent policy enforcement-related HHW collection.

The spatial outlier counties (of relatively low density HHW collection compared to the surrounding neighboring counties), which we discovered in Central California, changed during the study period; earlier it was Fresno and Madera in 2006, then Fresno and Tulare in 2007, and Kings County joined these spatial outliers in 2011. This finding indicates that these counties were different from the neighboring counties. Governments in these counties would have had to make more effort to increase participation in HHW collection programs. They could have initiated waste agency officer exchanges with nearby counties that had high density HHW collection, for example.

6.2 Policy Implications

HHW programs in Northern California counties were more successful in encouraging households to deliver their HHW than in the south. There are several explanations for the cost-effectiveness levels of programs in the regions.

First, people in Northern California may have been more willing to separate, sort and deliver waste to HHW collection programs and facilities than the people in Southern California. Perhaps they had higher environmental awareness and more opportunities to participate in HHW collection programs. Second, the counties in the Southern Region experienced higher population density. So it may have been more difficult for the local governments in the south to enforce an HHW ban. Third, the HHW programs and facilities in Northern California may have had better technologies to collect and process HHW; this requires further investigation.

These findings indicate differences between the counties in Northern and Southern California in terms of program effectiveness, environmental policy enforcement, and household behavior. State-level policy-makers should consider these differences when developing new HHW-related policies. The influential factors also seem to be differentially stronger or weaker in region-specific ways and this neatly aligns with the concept of multiple sub-states existing within the single state of California. Because of differences that exist across California, our findings are useful for regional governments to evaluate local impacts of their policies and approaches to HHW collection to promote sustainability and better waste management.

7 Conclusion

Our study is the first that attempts to analyze spatial patterns of HHW collection density in a state across multiple years. We discovered spatial patterns of HHW collection density that changed over time in California from 2004 to 2015. Our analysis results indicated some clusters of counties in Northern California with high HHW collection density. We found a few outlier counties in Central California; their amount of HHW collected per person was lower compared with nearby counties in the north.

We assessed the impact of HHW-related policies, policy actions and demographic factors on the HHW collection density across the counties in California using a geographically weighted regression model. These factors had stronger or weaker associations with HHW collection density in the north or south regions. Environmental policy enforcement had a stronger negative influence in the north, while HHW program costs had more positive influence in the north. HHW grant awards had stronger positive influence, but depended on project type.

One limitation is the aggregated HHW collection data so we could not consider the variability in the HHW categories. Studying the spatial patterns of the HHW collection densities by material type may suggest different spatial clusters or county outliers for each category. We didn't investigate the influences of production, consumption of products with high hazardous content, trading activities, or product sales on HHW recycled. Another limitation is that we have not modeled the lags and local transitions for different levels of HHW collection densities over time. We can extend the GWR model to represent spatial and temporal simultaneity too. We also have considered the possibility of a hidden Markov model (machine learning) for spatial patterns, and predict collection density patterns.

We demonstrated how computational social science fusion analytics methods that are specialized for geographic information science, such as spatial clustering and geographical weighted regression, allowed us to assess the spatial-varying effects of environmental policies and demographic factors in sustainability practices, such as HHW collection. Even in a state as advanced in HHW collection and recycling as California, we were able to find low, medium and high levels of HHW collection density spread across the regions. This shows county location matters.

These findings are applicable to other states or regions with differences across their geographies. By focusing on regional impacts, policy-makers can more effectively target the programs and incentives to the specific dimensions of the area rather than by applying a one-size-fits-all approach. The outcome from this targeted examination should lead to better management of HHW and environmental sustainability.

8 References

Abbott, A., Nandeibam, S., & O'Shea, L. (2011). Explaining the variation in household recycling rates across the U.K. *Ecol. Econ.*, 70(11), 2214–2223.

Agovino, M., Crociata, A., & Sacco, P.L. (2016). Location matters for pro-environmental behavior: a spatial Markov chain approach to proximity effects in differentiated waste collection. *Ann. Reg. Sci.*, 56(1), 295-315.

Anselin, L. (1995). Local indicators of spatial association: LISA. *Geog. Anal.*, 27(2), 193-115.

Antweiler, W., Copeland, B.R., & Taylor, M.S. (2001). Is free trade good for the environment? *Amer. Econ. Rev.*, 91(4), 877-908.

Bartelings, H., & Sterner, T. (1999). Household waste management in a Swedish municipality: determinants of waste disposal, recycling and composting. *Envir. Res. Econ.*, 13, 473–491.

Bivand, R., Ono, H., Dunlap, & R., Stigler, M. (2015). classInt: choose univariate class intervals. R version 0.1-23.

Brunsdon C., Fotheringham, A., & Charlton, M. (1996). Geographically-weighted regression: a method for exploring spatial nonstationarity. *Geog. Anal.*, 28, 281-289.

Callan, S.J., Thomas, J.M. (2006). Analyzing demand for disposal and recycling services: a systems approach. *East. Econ. J.*, 32, 221–240.

CalRecycle (2012). Permanent household hazardous waste collection facility profiles. Sacramento, CA.

CalRecycle (2015a). Home page. www.calrecycle.ca.gov. Sacramento, CA.

CalRecycle (2015b). Form 303: historical HHW collection data and voluntary survey responses. Sacramento, CA.

CalRecycle (2016). Household Hazardous Waste Grant Program. Sacramento, CA.

Chang, R., Kauffman, R.J., Kwon, Y. Understanding the paradigm shift to computational social science in the presence of big data. *Dec. Supp. Sys.*, 63, 67-80.

Chaussee, J. (2014). Billionaire's breakup plan would chop California into six states. *Reuters,* July 14.

Cole, M.A. (2000). Air pollution and dirty industries: how and why does the composition of manufacturing output change with economic development? *Envir. Res. Econ.* 17(1), 109-123.

Cole, M.A. (2004). Trade, the pollution haven hypothesis, and the environmental Kuznets curve: examining the linkages. *Ecol. Econ.*, 48, 71-81.

Davis, C.A., Jr., Fonseca, F.T., Camara, G. (2011). Environmental sustainability: the role of geographic information science and SDI in the integration of people and nature. *Handbk. GIS & Soc.*. Sage, London, UK, 123-137.

Ericksen, D.E. (1994). NAD 83: what is it and why you should care. SBE Natl. Conv., Los Angeles, CA.

GADM (2015). Global admin. areas. V.2.8. Davis, CA.

Gollini, I., Lu, B., Charlton, M., Brunsdon, C., & Harris, P. (2015). GWmodel: an R package for exploring spatial heterogeneity using geographically weighted models. *J. Stat. Softw.,* 63(17), 1-50.

Hurvich C.M., Simono. J.S., & Tsai, C.L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *J. Royal Stat. Soc. Ser. B*, 60(2), 1998, 271-293.

Janicke, M., Binder, M., & Monch, H. (1997). Dirty industries: patterns of change in industrial countries. *Envir. Res. Econ.*, 9, 467-491.

Jenkins, R.R., Martinez, S.A., Palmer, K., & Podolsky, M.J. (2003). Determinants of household recycling: a material-specific analysis of recycling program features and unit pricing. *J. Envir. Ecol. Mgmt.*, 45, 294–318.

Jenks, G.F., & Capsall, F.C. (1971). Error on choroplethic maps: definition, measurement, reduction. *Ann. Assoc. Amer. Geog.*, 61(2), 217-244.

Lasaridi, K.E., Rovolis, A., & Abeliotis, K. (2006). Waste management costs in Greece: spatial patterns and causal factors. *WIT Trans. Ecol. Envir.*, 98, 55-64.

Leung, Y., Mei, C., & Zhang W. (2000). Statistical tests for spatial nonstationarity based on the GWR model. *Envir. Plan. A*, 32, 9–32.

Lim-Wavde, K., Kauffman, R.J., Dawson G.S. (2016). Policy analytics, household informedness and the collection of household hazardous waste. *MATEC Web of Conferences*, 68 (14003), 1-5.

Moran, P. (1950). Notes on continuous stochastic phenomena. *Biometr.*, 37, 17-23.

Morris, G.E., & Holthausen, D.M., Jr. (1994). The economics of household solid waste generation and disposal. *J. Envir. Econ. Mgmt.*, 26, 215–234.

Richardson, R.A., & Havlicek, J., (1978). Economic analysis of the composition of household solid wastes. *J. Envir. Econ. Mgmt.*, 5, 1978, 103–111.

Saphores, J.D. (2006). Household willingness to recycle electronic waste: an application to California. *Envir. and Beh.*, 38, 183–208.

Sharma, J., & Agarwal, A. (2016). The value of spatial analytics in business intelligence. Oracle, Redwood City, CA.

Sidique, S.F., Joshi, S.V., Lupi, F. (2010). Factors influencing the rate of recycling: an analysis of Minnesota counties. *Res. Conserv. Recyc.*, 54, 242–249.

Springer (2016). Sustainablity science. Homepage. Available at http://link.springer.com/journal/11625.

Tong, X., & Wang, J. (2005). Transnational flows of e-waste and spatial patterns of recycling in China. *Euras. Geog. Econ.*, 45(8), 2005, 589-602.

U.S. Census Bureau (2012). American community survey. Washington, DC.

U.S. Dept. of Agric. (USDA) (2013). Rural-urban continuum codes. Economic Research Service, Washington, DC.

U.S. EPA (2000). Safe hazardous waste recycling. Washington, DC.

U.S. EPA (2014a). Household hazardous waste | common wastes & materials. Washington, DC.

U.S. EPA (2014b). Hazardous waste treatment, storage, and disposal facilities regulations: a user-friendly reference document for RRCA Subtitle C permit writers and permittees. Washington, DC.

U.S. EPA (2016). Enforcement and compliance history online (ECHO). Washington, DC.

van den Bergh, J. (2008). Environmental regulation of households: an empirical review of economic and psychological factors. *Ecol. Econ.*, 66, 559–574.

Voss, P., Long, D., Hammer, R., & Friedman, S. (2006). County child poverty rates in the U.S.: A spatial regression approach. *Pop. Res. Pol. Rev.*, 25(4), 369-391.

Wood, D.B. (2011). 51st state? small step forward for long-shot 'South California' plan. *Christ. Sci. Mon.,* July 12.

Zhang, Q. Lin, T., Chen, S., Xiao, L., Wang, J., & Guo, Y. (2015). Spatial characteristics of municipal solid waste generation and its influential spatial factors in Xiamen. *J. Mat. Cyc. Waste Mgmt.*, 17, 399-409.

Appendix. Regression Results

VARIABLES	2004 Coef. (SE)	2005 Coef. (SE)	2007 Coef. (SE)	2008 Coef. (SE)	2009 Coef. (SE)	2010 Coef. (SE)	2011 Coef. (SE)	2012 Coef. (SE)	2013 Coef. (SE)	2014 Coef. (SE)
Intercept	-5.96 (4.76)	1,29 (2.66)	-9.39 (9.75)	-5.80 (9.20)	***13.01****** (4.70)	2.21 (8.63)	***12.52****** (3.93)	-5.05 (6.04)	-11.05 (6.96)	***13.14****** (2.51)
MCLViol#	-0.34 (0.36)	***0.10***** **(0.05)**				-0.01 (0.03)			0.00 (0.01)	**-0.0*** **(0.00)**
Penalty#	-0.18 (0.19)	-0.15 (0.15)	-0.08 (0.21)	0.01 (0.27)	-0.18 (0.27)	-0.03 (0.30)	-0.22 (0.47)		-0.15 (0.28)	
HHWGrant$			0.67 (1.66)	***2.33****** (0.70)	-1.27 (3.82)	1.10 (0.94)		-1.08 (3.21)	-2.73 (2.09)	
HHWProgram$			1.35* (0.75)	-0.37 (0.65)	0.74 (0.72)	1.28* (0.68)	1.40* (0.81)			
PEduHS%	***11.92***** (4.61)		***28.95****** (10.26)	***19.44***** (8.34)		8.71 (8.87)		***13.70***** (5.52)	***21.67***** (9.21)	
HHInc		***0.61***** (0.21)	-0.79* (0.40)		0.09 (0.36)	-0.03 (0.39)			0.00 (0.00)	
RentalHH%	-1.73 (5.39)	-7.06 (5.30)	-10.83 (10.46)	-15.83 (10.25)	***-25.82****** (9.20)	-13.69 (9.28)	***-21.52***** (9.23)	-6.24 (6.23)		***-20.37****** (5.88)
Adj. R^2 (%)	16.8	23.2	27.1	33.2	14.8	22.4	17.0	16.6	13.9	30.3

Notes. Model: GWR with global weight; obs. = 39 each year. Dep. var.: *HHWCollD*. Coefficients with $p < 0.10$ are highlighted in gray; coefficients with $p < 0.05$ are in bold and italics. Estimates for 2006 and 2015 are not included due to low adj. R^2. Signif.: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A1. Global regression results

VARIABLES	2004 Min	2004 Max	2005 Min	2005 Max	2006 Min	2006 Max
Intercept	-7.99	-4.11	-1.87	2.16	***-0.56******	***16.80******
MCLViol#	-0.45	-0.25	0.10	0.11		
Penalty#	-0.56	-0.14	***-0.85******	***-0.10******	***-1.46*****	***-0.07*****
HHWGrant$					0.62	1.35
PEduHS%	8.95	11.55				
HHInc			0.57	1.00	***-0.42******	***0.00******
RentalHH%	-5.26*	6.56*	-16.96*	-5.86*	***-21.67******	***0.76******
Adj. R^2	13.8%		22.2%		13.0%	
VARIABLES	**2007 Min**	**2007 Max**	**2008 Min**	**2008 Max**	**2009 Min**	**2009 Max**
Intercept	-15.31	-7.76	-10.99	-0.36	***-3.75******	***17.64***
Penalty#	***-1.44*****	***-0.05*****	-0.97	-0.05	-0.74	-0.15
HHWGrant$			***0.25******	***2.71******	-3.36	-0.14
HHWProgram$	***-0.58******	***1.93******	-0.59	0.47	***-1.13******	***1.16******
PEduHS%	14.89	35.96	10.75	19.30		
HHInc	-1.09	0.00			0.08	1.00
RentalHH%	***-19.81******	***7.06******	***-26.43******	***8.76******	***-37.02******	***-11.93******
Adj. R^2	28.2%		40.2%		23.5%	
VARIABLES	**2010 Min**	**2010 Max**	**2011 Min**	**2011 Max**	**2012 Min**	**2012 Max**
Intercept	***-10.98*****	***10.59*****	***-5.17******	***22.72******	***-12.54*****	***-0.10*****
MCLViol#	-0.02	-0.01				
Penalty#	-1.15	-0.07	-2.13	-0.13		
HHWGrant$	***0.94******	***7.67******			-1.38	6.16
HHWProgram$	***-0.64*****	***1.55*****	0.30	1.70		
PEduHS%	1.05	22.48			10.84	16.10
HHInc	-0.49	0.00				
RentalHH%	***-19.79******	***6.58******	***-43.02******	***17.73******	***-11.96******	***7.54******
Adj. R^2	25.0%		37.9%		15.5%	
VARIABLES	**2013 Min**	**2013 Max**	**2014 Min**	**2014 Max**	**2015 Min**	**2015 Max**
Intercept	-14.47*	-1.99*	***5.41******	***14.98******	***3.21*****	***6.13*****
MCLViol#	***-0.01******	***0.00******	-0.01	-0.01	-0.02	0.00
Penalty#	-1.06*	-0.09*				
HHWGrant$	***-3.51******	***9.02******			***-2.16*****	***26.15*****
HHWProgram$						
PEduHS%	9.71	25.37				
HHInc	0.00	0.00				
RentalHH%			***-23.72******	***-5.06******		
Adj. R^2	17.8%		33.3%		15.8%	

Notes. Model: GWR with weight matrix computed using an adaptive bi-square distance function; obs. = 39 each year. The optimum bandwidth in the calibrated model was 37 nearest neighbors for all years, except in 2015, which was 27 neighbors. Dep. var.: *HHWCollD*. Coef. with F_3 test $p < 0.10$ highlighted in gray; coef. with F_3 test $p < 0.05$ are in bold, italics. Signif.: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A2. GWR results