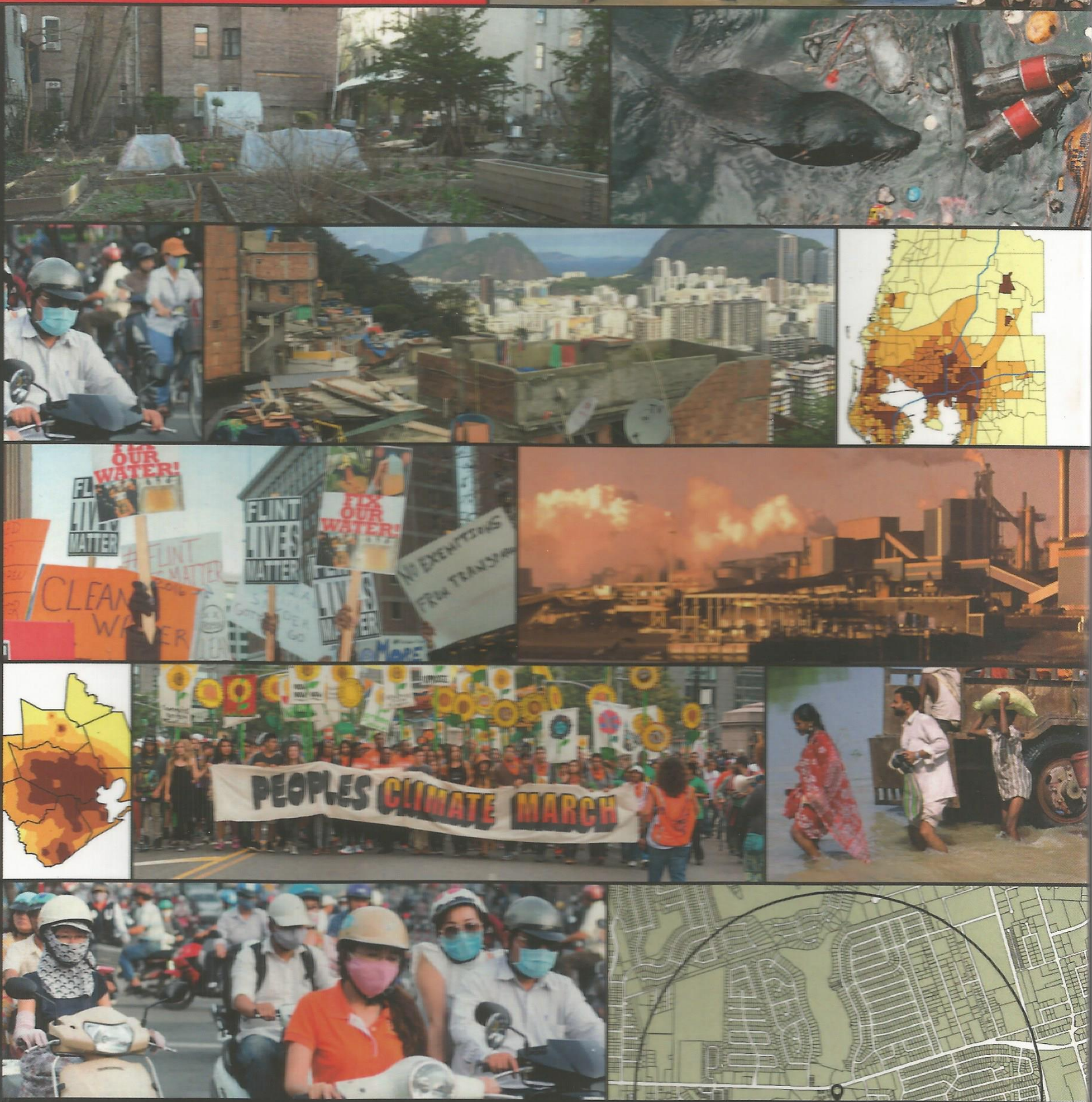
ROUTLEDGE

ROUTLEDGE HANDBOOKS



The Routledge Handbook of Environmental Justice

Edited by Ryan Holifield, Jayajit Chakraborty and Gordon Walker

16

ASSESSING POPULATION AT RISK

Areal interpolation and dasymetric mapping

Juliana Maantay and Andrew Maroko

Introduction

Today, the need for visualization of population data is increasingly crucial, not just for descriptive purposes – to show the geographic extent and density of populations – but also for spatial analytical and predictive modelling purposes, in order to inform risk assessments and public policy formation on many urban issues, including environmental justice concerns (Lwin and Murayama 2011; Maantay et al. 2009; Moon and Farmer 2001).

This chapter provides a description and comparison of techniques used to estimate population counts and socio-demographic characteristics in areas potentially impacted by environmental and other hazards, or to ascertain which populations do not have adequate access to health-promoting land uses and facilities. These techniques can be used to facilitate environmental justice and health equity analyses. Environmental impacts are often spatially represented by discrete bounded areas (e.g. circular, plume, or network buffers) or continuous surfaces (e.g. outputs from dispersion modelling or land use regression). Because these impact exposure areas will normally not conflate geographically with population data and their units of aggregation, it is necessary to devise a method whereby the disparate units of analysis can be harmonized and reconciled. The challenge is to determine the optimal method to estimate the affected population, either within discrete boundaries or within a certain pollutant concentration, as derived from a continuous surface, for example.

To that end, we examine and compare the relative merits of methods including spatial coincidence/selection (such as centroid containment), and disaggregation techniques (such as areal interpolation, filtered areal weighting, and cadastral-based dasymetric mapping), using an example of a potentially impacted population living near a controversially sited waste water treatment plant in Harlem, New York.

Population estimation for environmental health and justice analysis

Geographic Information Science (GISc) methods have been used in environmental justice research primarily to analyse the spatial relationships between sources of pollution (and other environmental burdens and hazards) and the socio-demographic characteristics of potentially affected populations. More recently, they have also been used to estimate and locate populations affected by lack of geographic access to health-promoting activities and facilities, such as open spaces for physical recreation, establishments selling healthy food options, and medical providers. With GISc, we try to map instances of environmental injustice, usually by plotting the locations of facilities or land uses suspected of posing an environmental and human health hazard or risk, and then determining the racial, ethnic, and socioeconomic characteristics of the potentially affected populations, as compared to a reference population. Many of the techniques that are described below would either not be feasible to do without utilizing GISc or would be extremely cumbersome and time-consuming. Quantitative environmental justice and health equity analyses have essentially been made possible because of GISc methods.

In order to be able to estimate the counts of vulnerable sub-populations who may be adversely affected by noxious land uses or events or who suffer from lack of access to health-promoting facilities, we need to have an accurate count of people potentially impacted by environmental hazards, within the geographic extent of the hazardous conditions or at-risk locations. This can be challenging to calculate because population information is most frequently based on data aggregated by census tracts, postal ZIP codes, or counties, in the US, and equivalent administrative or government jurisdictional boundaries in other countries. It is rare that environmental and health conditions coincide with these arbitrary and usually artificial boundaries. Hence, there is a need to figure out how to parse the data, and arrive at the best representation of population distribution within non-coincidental areas.

Spatial representation of impact areas

Areas affected by environmental impacts are generally depicted in one of two ways: either as discrete areas, bounded by some distance-based buffer zone (Chakraborty and Maantay 2011); or as a continuous surface, based on pollutant concentration or dispersion. These main types of spatial representation are listed and illustrated below, and discussed in more detail in Chapter 15.

Buffers

Constant-distance buffers (fixed-distance buffers)
Example: A buffer of 2500 feet around a polluting facility, representing the likely distance that specific smokestack emissions will travel through the air (Figure 16.1a).

Concentric buffers (multiple-ring buffers)
Example: Concentric buffers around points, representing the smokestacks of industrial plants, with several different potential impact distances, which allows for a sensitivity analysis to be performed (Figure 16.1b).

Variable buffers
Example of variable point buffer: The smokestacks of polluting facilities are shown as points, and different sized impact buffers might be based on the different toxicity levels of the pollution mix emitted from the stacks of each of these noxious facilities, with the more toxic ones having larger impact buffers. The buffers in this case would not necessarily reflect the actual geographic extent of the impact, but rather be an indication of the magnitude/hazardousness of the impact (Figure 16.1c).

Example of variable line buffer: In developing a realistic buffer to indicate potential impacts from roadway vibration or noise, variability could be based on vehicular count.

Wider, busier roads would have a larger impact buffer than secondary or local roads, and the buffer around each type of road would vary accordingly (Figure 16.1d).

Plume buffers
Example: A plume buffer based on modelled outputs would show an approximation of the flow of air pollutants from a smokestack, using information about pollutant type, prevailing wind direction and wind speed at the site, exit velocity of the emissions from the stack, and other features of the pollutant and landscape morphology that would affect the geographic distribution of the pollutant (Figure 16.1e).

Network buffers
Example: A network buffer of streets would show the catchment area (or "walking-shed") of population having access to a healthy food store (or a park entrance), if we assume a normal optimal walking distance of no more than a quarter of a mile. By contrast, a fixed-distance circular buffer of a quarter of a mile around the store or park entrance would not as realistically capture the population having good access, since it would not accurately reflect actual walking distance along sidewalks or streets (Figure 16.1f).

Continuous surfaces

Environmental conditions also can be shown as a continuous surface, such as can be obtained from dispersion modelling, land use regression, or interpolated ambient environmental indicators. Example: a continuous surface of air pollution values is aggregated to the units that will be analysed, such as by taking an average of the pollutant concentrations across a census tract, which can then be linked with socio-demographic data for each tract (Figure 16.1g).

Methods of population estimation

Spatial coincidence

In this simplest method of determining the affected or impacted population (Figure 16.2a), the geographic unit either contains an environmental "good" or "bad," or it does not. If it does, then all the population within that unit is considered "affected" by the facility or land use. This is a binary analysis, and of course is quite inexact, since population that is close to the facility or land use, but on the other side of the containing boundary line would be considered not impacted, or not exposed, or without access, whereas a population within the geographic unit but far away from the facility or land use would be considered to be impacted. This method is adequate for broad-brush analyses, for instance a nation-wide study at the county level, but is not sufficiently nuanced for a more fine-grained city- or community-scale analysis.

Selection by proximity buffers (distance-based analysis)

In utilizing proximity buffers (Figure 16.2b–d), there are a few important decisions to make in order to estimate affected population within a buffer zone. The underlying population data is generally obtained from a census or other demographic database, and the data is typically aggregated by census enumeration district (tract, block group, etc.), or by postal ZIP code, county, or state, depending upon the geographic extent of the study and its level

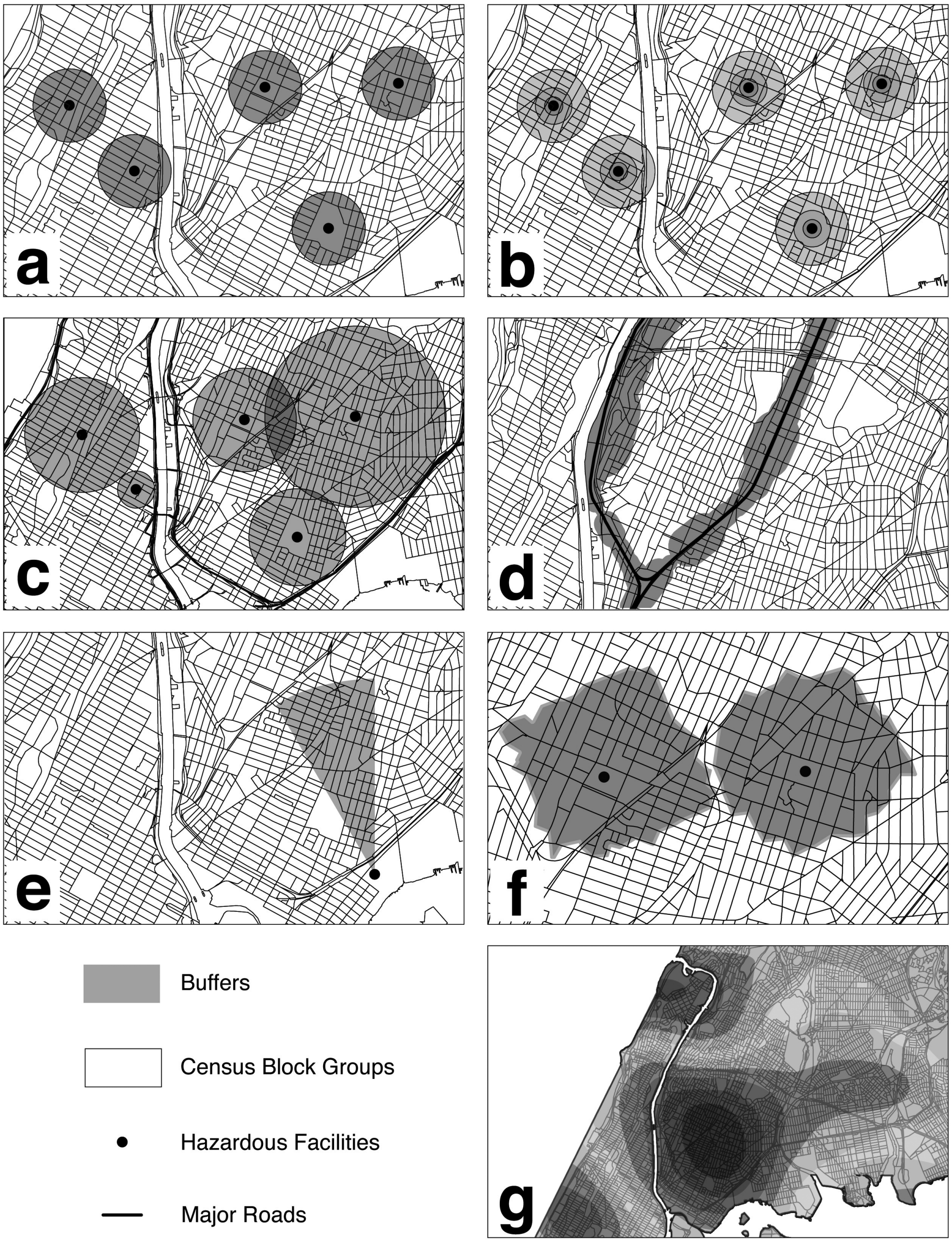


Figure 16.1 Methods of delineating impact areas: a. Constant-distance buffers (fixed-distance buffers); b. Concentric buffers (multiple-ring buffers); c. Variable point buffer; d. Variable line buffer; e. Plume buffer; f. Network buffers; g. Continuous surface

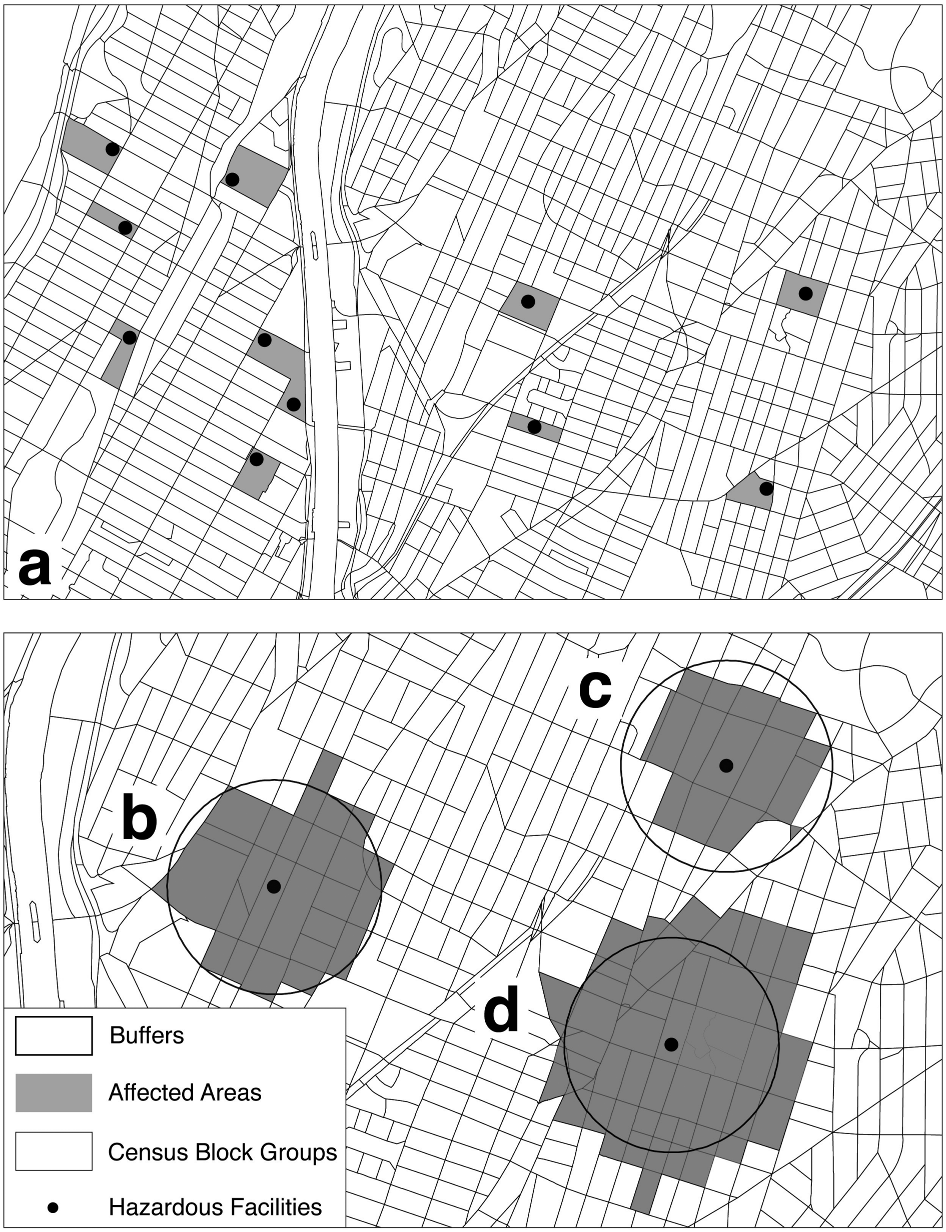


Figure 16.2 Selection of "impacted" population: a. Spatial coincidence; b. Centroid containment; c. Complete polygon containment; d. Polygon intersection

of detail. When the proximity buffers are overlaid on the unit of aggregation, they will not conflate – they will not be spatially co-incident. Some units will be partially within the buffers and only a few may lie entirely within the buffer. In order to mitigate this spatial incongruity, it will be necessary to decide which of the units (census tracts, for instance) to include in calculating the affected population. The typical way of doing this in a GIS is to select the tracts that meet certain criteria. We could select 1) Complete polygon containment: all the tracts that fall completely within the buffers; 2) Centroid containment: all the tracts whose centroid (the geometric centre of the unit) falls within the buffers; or 3) Polygon intersection: all the tracts that intersect with or are contained within the buffer. The 50 per cent areal containment method is also frequently used in EJ analyses, which entails including a polygon in the selection set if at least 50 per cent of its area is within the buffer. However, this method generally yields a very similar result to that obtained by centroid containment. Selection choice #1 is generally the most exclusive (containing the fewest tracts and therefore the least population), while #3 is generally the most inclusive (containing the most tracts, and the maximum population). The buffers can be fixed distance, concentric (multiple-ring) buffers, variable distance buffers, or plume buffers, or based on a continuous surface, but the problem of calculating the population within the affected areas remains the same.

Areal interpolation

The problem of non-coincident spatial boundaries and the frequent need to transfer data from one set of zones to another (e.g. population within census block groups versus population within a flood zone) is a long-standing dilemma, and is not addressed well by choropleth mapping, which simply distributes the population or other data evenly throughout the spatial unit (Maantay and Maroko 2009). Typically, the issue of rectifying attribute data from different sets of spatial units is handled by a procedure called *areal interpolation* (Figure 16.3a). This very simple method redistributes the source data (e.g. population) based solely on area proportions.

In areal interpolation, we again are using buffers or a continuous surface to delineate the potential areas of exposure, access, or impact, and census data or similar for the underlying population data. But, as opposed to the selection method, we will try to apportion the population in each tract according to the amount of the tract that is within the buffer. So, for example, if the buffer intersects the tract in such a way that 20 per cent of the tract is within the buffer and 80 per cent is outside the buffer, then we would make the assumption that 20 per cent of the tract's population was within the buffer, and therefore "affected." One major deficiency of this method is that it is based on the erroneous assumption that population is distributed evenly and equally throughout the unit of aggregation. This is rarely the case, and tracts may have population primarily in one part of the tract and not in other parts.

Dasymetric mapping

The underlying concept of dasymetric mapping involves the process of disaggregating spatial data to a finer unit of analysis, using additional (or "ancillary") data to help refine locations of population or other phenomena being mapped (Holt et al. 2004; Maantay et al. 2007; Mennis 2003). Dasymetric methods range from basic to more complex in order to achieve greater accuracy and a more realistic portrayal of population distribution, and are discussed below in order of ascending complexity.

Filtered areal weighting

Filtered areal weighting (FAW; Figure 16.3b) takes into account that some parts of a census tract or other geographic unit do not have any population, and therefore when apportioning population according to the amount of the tract that is within the buffer, these unpopulated areas should be excluded. Conventionally, the areas that are excluded are parks, open spaces, and water bodies. The buffered areas are "filtered" by an ancillary data set or sets, which are used to mask out the non-populated areas. In the simplest version, a binary redistribution is used: areas are deemed either inhabited or non-inhabited. If, from the land cover data, one can infer that an area is uninhabited, no population from the population source layer would be assigned to that polygon or pixel, leaving all of the population to be distributed to the remaining areas. In essence, the ancillary data set in this example acts to mask the census tract data so that the uninhabited land is left devoid of population (Eicher and Brewer 2001). This type of mapping often utilizes land cover data from satellite images to create the filtering or masking information, in combination with areal interpolation, and is usually considered a simple form of dasymetric mapping.

A further refinement of the filtered areal weighting method is to exclude, in addition to parks, open spaces, and water bodies, those areas which are typically uninhabited or very sparsely inhabited, such as industrial or commercial areas. The locations of these areas can be ascertained by either land cover data from satellite imagery or land use or zoning maps from municipal or county sources, and then used to mask the population, as in basic areal weighting.

Cadastral-based dasymetric mapping

Cadastral-based dasymetric mapping (Figure 16.3c) is a still further refinement of filtered areal weighting. In addition to excluding areas that are typically uninhabited, it uses additional ancillary data, such as property lot (cadastral) data, containing, for instance, the number of residential units or residential square footage per lot, to redistribute spatial data in a more accurate and logical manner. Building volume as derived from building footprint data has also been used to develop a more accurate picture of population distribution (Wu et al. 2005). Details such as housing tenure, ownership, and values can also be used to assess socioeconomic characteristics of proximate households.

This is considered an improvement over the disaggregation methods used in filtered areal weighting, which can be of limited utility in urban areas, due to, among other reasons, the lack of distinction typically made between land cover and land use, and spatial resolution that is too low for the application. Complete information about land use, and more importantly, population density, is not reflected clearly in the land cover data (Forster 1985).

Moreover, in a spatially heterogeneous urban area such as New York City, simply knowing whether or not an area comprises residential land use is insufficient for calculating population, since residential buildings range from containing one household to several hundred households on one lot, and population density can vary widely even within the relatively small area of a census block group, or the even smaller city block. This type of cadastral data is now available for most urban areas in the United States and other developed countries, since its primary purpose is for property tax collection.

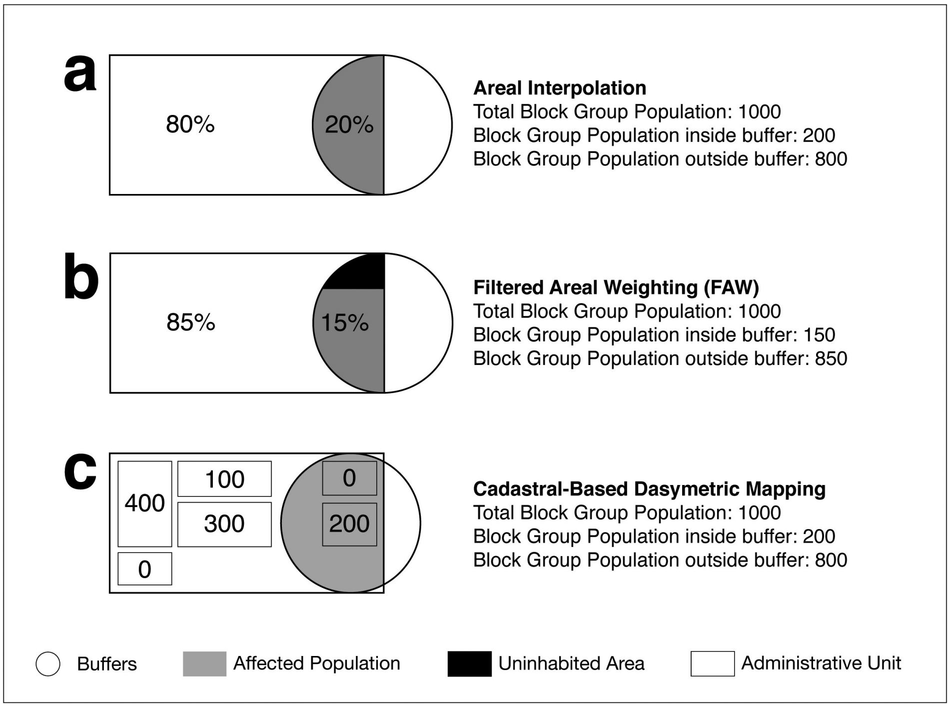


Figure 16.3 How "impacted" population numbers are estimated by various methods: a: Areal interpolation; b. Filtered areal weighting; c. Cadastral-based dasymetric mappping

Case study of the North River Waste Water Treatment Plant, New York City

In this case study, we use the environmental justice implications of the North River Waste Water Treatment Plant as an illustration of the various methods to calculate populations and sub-populations potentially affected by a specific land use (Figure 16.4). The siting of the North River Waste Water Treatment Plant (WWTP), on the Hudson [North] River in Harlem, NYC, was very contentious, and more than 20 years later is still considered to be an example of environmental injustice to the largely minority community nearby. The plant was completed in 1994, but planning for the facility began nearly a century earlier, in 1914, when the city decided that a WWTP was needed to handle the waste water of the entire west side of Manhattan. At that time, most of Manhattan's waste water was released as raw sewage into the Hudson [North] River or the East River. Site selection for the WWTP began in earnest in the late 1950s, due largely to the burgeoning population of Manhattan, accompanied by a big push in the nascent environmental movement to clean up the polluted Hudson River. A number of possible sites were considered, including one at West 72 Street, in a neighbourhood that was at that time just beginning to gentrify. There was a huge outcry from the middle- and upper-class residents against this facility being located in their neighbourhood, and a politically connected mobilization to prevent the plant from being sited there. Robert Moses, infamous as New York's "Power Broker," also helped to scuttle the plans for the West 72 Street WWTP.

The city then moved their prospects to a 28-acre, 8-block stretch between 137 and 145 Streets in the largely black community of Harlem. Design studies were started, and detailed

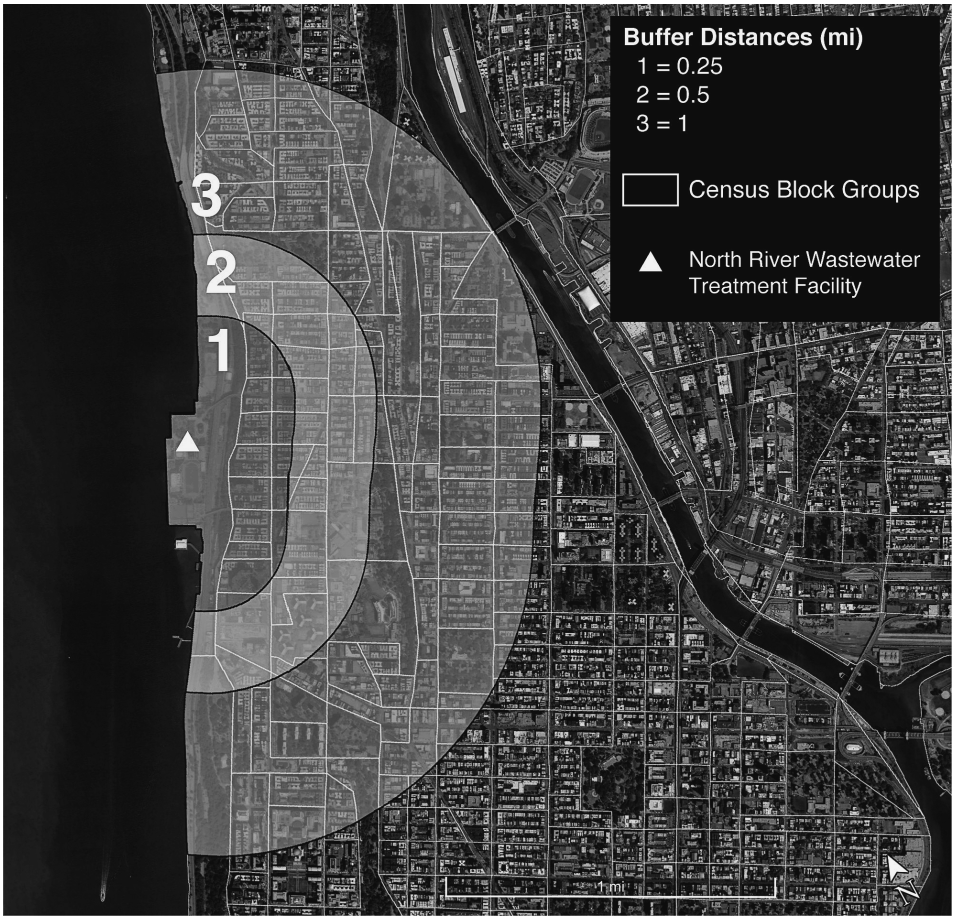


Figure 16.4 Aerial image of case study area, showing the location of the North River WWTP with concentric buffers indicating potential impact areas

plans were completed by 1971. The City Planning Commission approved of the plans long before any public hearings were held, and there was no legal requirement at that time for environmental impact assessments or public input process (Bunch 1994; Gladwell 1993; Mandell 1993; NACCHO 2011a, b). A look at the 1990 census data bears witness to the stark difference in socio-demographic characteristics between the neighbourhoods surrounding the two possible sites: The population around the originally proposed West 72 Street site was 84 per cent white and 8 per cent black, with an average household income of $123 000 per year, and 8.5 per cent of its people below the poverty line, while the population around the eventually built West 145 Street site was 15 per cent white and 60 per cent black, with an average household income of $26 000 per year, and 34 per cent of its people below the poverty line. Although New York City's government and private development interests have been dumping burdensome facilities in poor and minority communities probably since the city's 17th century founding, this was one of the first such decisions in modern times to be such a blatant and highly visible example of discriminatory siting, and one whose controversies received so much attention from the public and the press.

The plant, which is one of 14 WWTPs in NYC, handles 125 million gallons of waste water per day in dry weather, and is designed to handle 340 million gallons in wet weather (storm water is also processed in addition to waste water and sewage from buildings). To help mitigate the adverse impacts on the surrounding community, a new sports and recreation centre was built on top of the WWTP. Since the completion of the plant and the park, complaints from nearby residents about sickening odours and increased respiratory illness have continued without abating. The air quality problems stemming from the plant beneath the park have made using the park less than ideal for passive and active recreation.

> Shortly after construction, community members complained about the overpowering smell of rotten eggs coming from the plant. The smell became more potent during the summer months and travelled as far down as 120th Street and up to 157th Street—almost a two-mile distance. Residents had been forced to keep windows closed and stay inside to avoid the stench. Many have also reported itchy eyes, shortness of breath, and respiratory symptoms.
>
> *(NACCHO, 2011b)*

Comparison of methods in the case study

In order to demonstrate the potential differences in results that can be a product of population estimation techniques, concentric (multi-ring) fixed-distance buffers were created around the North River Waste Water Treatment Plant (at a quarter of a mile, half a mile, and 1 mile) and different estimation techniques were used, which included three selection methods, areal weighting, filtered areal weighting, and cadastral-based dasymetric disaggregation (Figure 16.5). Total population as well as major racial/ethnic breakdowns were estimated as "impacted" based on the methods and each method's output are compared. Demographic, planimetric, and cadastral data were from as near 2000 as possible, in order to approximate the affected populations soon after the plant came online.

Data used

Population demographics were collected from the 2000 US census at the block group level for New York County, NY (borough of Manhattan). *Block Groups* (BGs) are statistical divisions of *census* tracts, are generally defined to contain between 600 and 3000 people, and are used to present data (www.census.gov/geo/. . ./gtc_bg.html). In size, they are roughly intermediate between a census tract and a census block. Table P004 from the summary file 1 (100 per cent sample) was used to create variables for total population, Hispanic/Latino, non-Hispanic white, non-Hispanic black, non-Hispanic Asian, and non-Hispanic other (which represents residents who do not self-identify as any of the previous population groups).

Parks and open space data were downloaded from NYC open data and represent spatial features such as cemeteries, courts, and tracks as well as state and city parks. (https://data.cityofnewyork.us/Recreation/New-York-City-Open-Spaces)

Cadastral data used was from the MapPLUTO dataset, which is a "spatialized" version of Primary Land Use Tax Lot Output (PLUTO) data, provided by the NYC Department of City Planning (www.nyc.gov/html/dcp/html/bytes/archive_pluto_mappluto.shtml). This information was used for defining the number of residential units (e.g. number of apartments) in each tax lot.

The waste water treatment plant was delineated using the MapPLUTO data above and manually modified using aerial photos in order to create as accurate a boundary as possible.

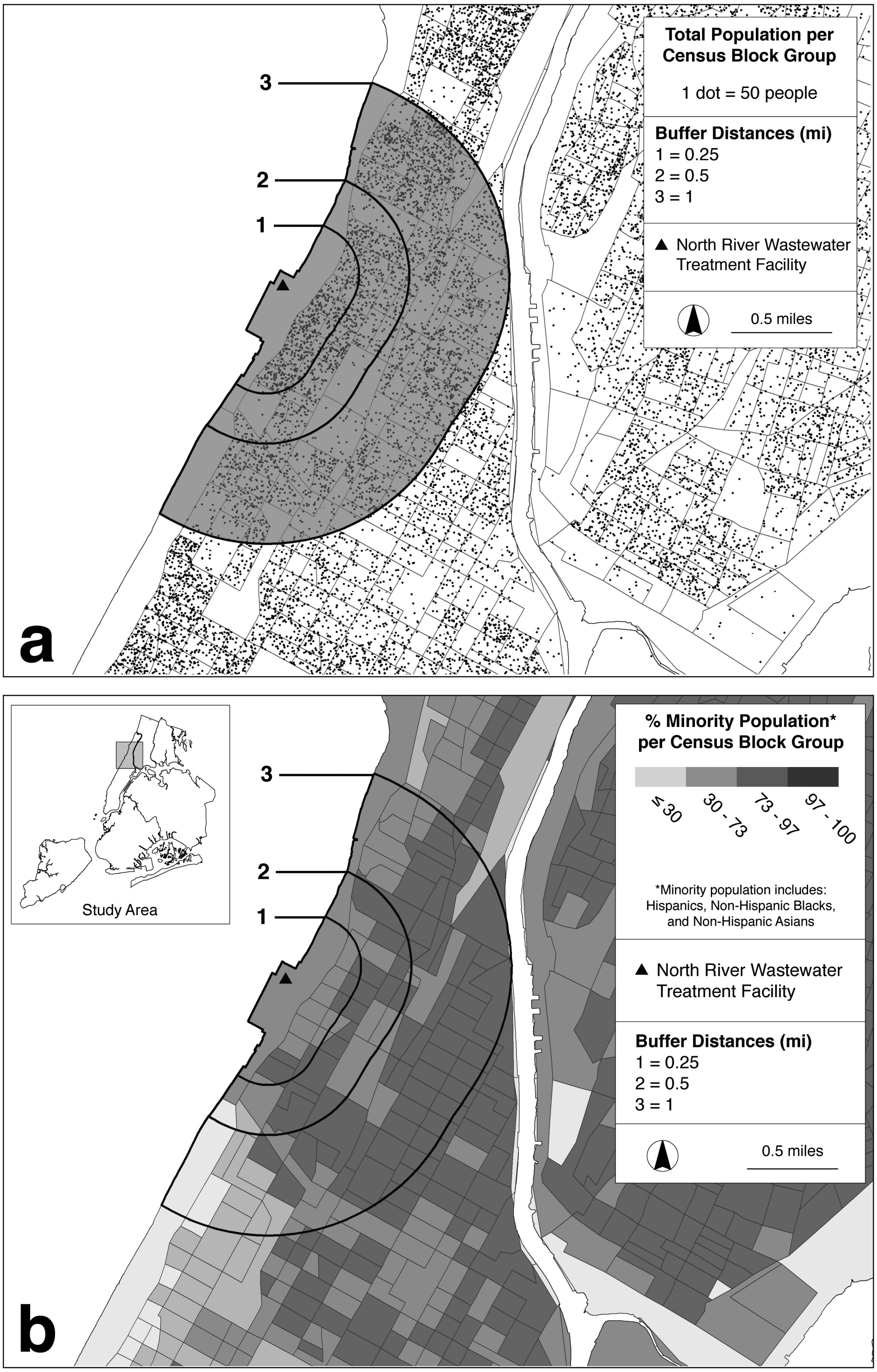


Figure 16.5 a. Case study area dot density map of total population distribution by census block group; b. Per cent minority population by block group

Selection by proximity buffers (distance-based analysis)

Simple selection is generally the fastest method for estimating populations within specified distances from a location (in this case the water treatment plant polygon). Three selection definitions were used: Complete polygon containment: the census block groups which fall completely within the buffers (most exclusive selection method); Centroid containment: those BGs which have their centroids (geographic centres) within the buffers; and Polygon intersection: those BGs which intersect the buffers (most inclusive selection method). To demonstrate the inclusive and exclusive nature of each of these methods, the half-mile buffer resulted in 23, 29, and 39 block groups being selected using the complete polygon containment, centroid containment, and polygon intersection methods, respectively.

Once the impacted block groups are identified, the demographic groups are simply summed, which results in estimates of the populations of interest which meet the criteria as described above. Not surprisingly, population estimates range with respect to method. Total population within half a mile was estimated as approximately 48 000, 57 000, and 71 000 for the complete polygon containment, centroid containment, and polygon intersection methods, respectively.

Areal interpolation

In order to perform areal interpolation in this case study, a number of steps were taken to prepare the data. First, the areas of each block group were calculated ($AREA_0$). Next, the block group polygons were combined with the buffers using the "union" function in ArcGIS. This results in the block group polygons which are intersected by the buffers being split, however the block groups which are either completely within the buffers, or completely outside of the buffers, will be unaffected. Areas were then recalculated to represent the potential new shapes ($AREA_1$). Finally, a simple formula is used which redistributes the original population (P_0) to new population estimates (P_1) based on the area ratios. It can be written as

$$P_1 = P_0 \star A_1/A_0$$

Using this method, if $A_1 = A_0$, then $P_1 = P_0$. However, if A_1 is, for instance, half as large as A_0, then P_1 will also be half as large as P_0. When compared with the selection methods above, the estimates generated by areal interpolation tend to be most similar to the centroid method, with outputs falling in between the more inclusive and exclusive techniques.

Filtered areal weighting

Filtered areal weighting is similar conceptually to simple areal interpolation; however, selected unpopulated areas are removed in order to get a more accurate estimate of where the population actually resides. In this case study, parks and open spaces were masked (filtered) from the census block group polygons using geoprocessing functions in ArcGIS before the area of the block groups ($AREA_0$) is calculated. What follows is functionally identical to simple areal interpolation, with the (filtered) block groups being split by the buffers, new areas being calculated ($AREA_1$), and finally the new populations being estimated by multiplying the original block group population by the ratio of the filtered areas. In this case study, the outputs are similar to the simple areal interpolation, with slightly larger estimates at the half-mile buffer. However, depending on the characteristics of the region being studied, the units of aggregation used in the analysis, and the buffer size, the results could show much larger differences between the methods.

The refined areal weighting method was also undertaken for the case study area, but in this instance the results were very similar to the cadastral-based dasymetric disaggregation method (see below). It should be noted that these two methods do not always yield similar results, depending upon the data used in the particular case, and the two methods might show quite different results.

Cadastral-based dasymetric disaggregation

The final estimation technique that will be demonstrated in this case study is also the most labour- and data-intensive. By assuming that population distribution is biased by an ancillary variable other than land area, in this case number of residential units, we should be able to achieve a more robust and nuanced estimate, and one which more accurately reflects actual conditions. In order to achieve this, cadastral data from MapPLUTO aggregated to the tax lot (cadastre) was overlain with the block group polygon layer as well as the buffer layers. Tax lots which fall within the block group and buffer layers were given their attribute information (buffer distances, block group identifiers, demographics, etc.). The total number of residential units in each block group was then calculated and appended to each tax lot. Next, a population estimate was created for each lot by applying a formula similar to that of simple areal interpolation, where the estimated tax lot population (PL_1) is equated to the original block group population (P_0) multiplied by the ratio of residential units in the tax lot (RU_1) over the total number of residential units in the block group (RU_0). These data can now be manipulated to create aggregates based on exposure or impact (e.g. inside a buffer). The results from the cadastral-based dasymetric disaggregation in this case study tend to be slightly more inclusive than those of areal interpolation or filtered areal weighting for the half-mile buffer. This will vary case by case, depending upon the particular way the population is distributed in any given study area and within the units of aggregation.

Case study results

Results from the case study are shown in Table 16.1, with both the absolute number population estimates, as well as comparisons of each method to the cadastral-based dasymetric disaggregation as per cent difference. The cadastral-based method was used as the standard for comparison amongst the methods because it is considered the most robust and is based on more realistic distribution of the population data (Eicher and Brewer 2001; Holt et al. 2004; Maantay et al. 2007; Martin 2006; Mennis 2003; Poulsen and Kennedy 2004). It can be seen that as the buffer distances increase, the relative differences in the estimates decrease. This suggests that if the impact areas of interest are quite large with respect to the units of aggregation that house the socio-demographic data (in this case, census block groups), the potential error introduced by the various methods becomes less marked. For instance, the polygon intersection selection method (most inclusive) is has a difference of 82 per cent when compared to the dasymetric method used, whereas the complete polygon containment method has a difference of -42.7 per cent. However, when examining the outputs for the one-mile buffer, those differences drop to 6.5 per cent and 22.1 per cent for the polygon intersection and complete polygon containment methods, respectively. It is also important to note that although many of the demographic groups examined follow the total population trends, there is still a fair amount of variation. For instance, when looking at the centroid

Table 16.1 Comparison of population estimates. Counts of estimated impacted populations as well as per cent difference of each estimate from the cadastral-based dasymetric disaggregation for each buffer size.

Estimation Method	*Buffer Dist.*	*Population Count Estimates*						*Per cent Difference from Cadastral–based Dasymetric Disaggregation*					
		Total Population	*Hispanic/ Latino*	*Non–Hispanic White*	*Non–Hispanic Black*	*Non–Hispanic Asian*	*Non–Hispanic Other*	*Total Population*	*Hispanic/ Latino*	*Non–Hispanic White*	*Non–Hispanic Black*	*Non–Hispanic Asian*	*Non–Hispanic Other*
Polygon Containment	1/4 mile	13,010	9,953	538	2,039	240	240	–42.7	–38.6	–30.4	–59.1	–27.1	–44.3
Centroid Containment		20,741	14,803	775	4,444	334	385	–8.7	–8.6	0.2	–10.8	1.5	–10.6
Polygon Intersection		41,515	29,047	1,290	9,801	477	900	82.8	79.3	66.8	96.7	44.9	109.0
Areal Interpolation		23,092	16,593	780	4,952	327	441	1.7	2.4	0.8	–0.6	–0.7	2.4
Filtered Areal Weighting		23,161	16,652	786	4,952	328	442	2.0	2.8	1.7	–0.6	–0.3	2.7
Cadastral-based Dasym. Disag.		22,713	16,198	773	4,983	329	431	N/A	N/A	N/A	N/A	N/A	N/A
Polygon Containment	1/2 mile	47,853	31,599	1,434	13,134	575	1,111	–20.8	–14.1	–37.7	–30.5	–39.7	–24.5
Centroid Containment		56,687	35,188	1,643	17,802	684	1,370	–6.2	–4.4	–28.6	–5.8	–28.3	–6.8
Polygon Intersection		71,630	40,540	3,494	24,364	1,418	1,814	18.6	10.2	51.8	29.0	48.6	23.4
Areal Interpolation		57,491	35,242	2,286	17,621	938	1,404	–4.8	–4.2	–0.7	–6.7	–1.7	–4.5
Filtered Areal Weighting		58,009	35,401	2,410	17,786	988	1,424	–4.0	–3.8	4.7	–5.8	3.5	–3.2
Cadastral-based Dasym. Disag.		60,408	36,792	2,302	18,889	954	1,471	N/A	N/A	N/A	N/A	N/A	N/A
Polygon Containment	1 mile	143,944	69,855	8,361	59,222	2,707	3,799	–22.1	–14.2	–26.9	–29.1	–26.3	–20.9
Centroid Containment		187,486	82,413	11,780	84,687	3,752	4,854	1.5	1.3	3.0	1.4	2.1	1.1
Polygon Intersection		196,714	84,413	13,071	89,925	4,193	5,112	6.5	3.7	14.3	7.7	14.1	6.5
Areal Interpolation		180,889	80,109	11,570	80,782	3,714	4,714	–2.1	–1.6	1.2	–3.2	1.1	–1.8
Filtered Areal Weighting		180,727	80,081	11,573	80,649	3,712	4,712	–2.2	–1.6	1.2	–3.4	1.1	–1.8
Cadastral-based Dasym. Disag.		184,774	81,389	11,433	83,478	3,674	4,801	N/A	N/A	N/A	N/A	N/A	N/A

containment method, there is a tendency to undercount all population when compared to the dasymetric method. However, the underestimation of non-Hispanic black residents is more severe than the undercounting of non-Hispanic white residents which could result in a mischaracterization of the exposed groups and the obfuscation of potential environmental justice implications. As such, it is extremely important when conducting EJ analyses to attempt to produce the most reliable and robust population estimates.

Conclusions

Accurate estimates of impacted populations can have profound influences on research findings, and ultimately on the political process and in decision-making. This case study has demonstrated the differences amongst various methods in an "after-the-fact" siting situation, as a worked example of the relative merits of the methods. Clearly, using these techniques as a pro-active analysis during the environmental impact assessment process, other public review procedures, or as part of a preliminary planning initiative would be much more beneficial.

The comparison of population estimation methods used in the case study finds that although spatial coincidence/selection is typically the quickest and most frequently used method to calculate at-risk population, this tends to result in more general estimates, depending upon the size of the geographic unit of aggregation in relation to the impact exposure area. On the other hand, population disaggregation techniques such as the cadastral-based dasymetric method are more computationally intensive, but often yield a more robust and reliable estimate.

This method will be useful in many disparate fields and serve many purposes, for instance emergency management operations and implementation; police operations; criminal justice; fire and ambulance services; utility providers; and any other crucial public support systems dependent upon population information. Additionally, urban planning applications will benefit from better spatial data on potentially impacted sub-populations, since understanding the locational characteristics of target populations would allow for more equitable resource allocation in spheres such as community infrastructure development, provision of open space and recreational opportunities, transportation access, and necessary environmental facilities (Maantay et al. 2007).

For EJ analyses, more precise estimations of populations within impact extents and better locational information on population and sub-population distribution can be beneficial in "making the case" to decision-makers and policy experts. The more realistic and accurate results obtained through these methods can contribute to more confidence in crafting equitable solutions, and potentially increased political support for EJ policies and remedies.

There are a number of limitations inherent in population estimations and assessing populations at risk. With all of the methods discussed here, even dasymetric techniques, assumptions are made regarding the distribution of population (e.g. the assumption that where there are more residential units, there is greater population).

Secondly, the population numbers rely on census data, which are known to often be inaccurate, with serious problems of undercounting certain populations, especially in large urban areas. Thirdly, census data only take into account the residential location of population, who may not be at home as many hours as they are in school or at work, thereby overestimating some of the population that might be impacted (Maantay et al. 2008). There have been a number of recent studies exploring the daily spatial mobility of individuals, to demonstrate the heterogeneity of many people's "activity spaces" (Chaix et al. 2012, 2013; Harrison et al. 2014; Matthews 2011). This is an important topic for future study, since only then we will be able to base exposure, access, and impact assessments on the actual locations that people occupy throughout their typical day, rather than relying solely on their residential locations.

Although we believe that the cadastral-based dasymetric disaggregation technique will generally yield the most reliable results for population estimation, it is a much more labour- and data-intensive method, and therefore may not be appropriate for every analysis. The extra effort involved may not provide a concomitant increase in reliability or accuracy in every situation, and the difference in the complexity of the methods might not produce significantly different results. This will also depend upon the degree of exactitude required by the specific analysis being undertaken. If time permits, very likely the best and most thorough approach is to conduct exploratory spatial data analysis (ESDA), whereby several methods are used to view the data from as many different vantage points as possible, and then selecting the one that most faithfully reflects reality (or a combination/averaging of results from various techniques). Often, presenting the reader with the results of several different methods is the most transparent and honest approach. Exploring your data as much as possible in the initial phase of a research project is always the best way to start, where feasible. Data exploration and visualization through mapping and spatial analysis often provides a more robust understanding of the data, as well as improved clarity in viewing the phenomena under study, which will lead to better design of further analyses and additional hypothesis generation, in an iterative fashion (Maroko et al. 2011).

Every "place" will be different: how population is distributed within the study extent and within each unit of aggregation, for instance, can radically affect the outcome of the analysis, as well as influence whether the selection of the various methods described in this chapter makes a marked difference in the results. The researcher or analyst must understand how different methods of population estimation might obscure their findings for the specific place and phenomenon they are investigating, dependent on the data and time available for the analysis.

Acknowledgements

The National Oceanic and Atmospheric Administration's Cooperative Remote Sensing Science and Technology Center (NOAA-CREST) provided critical support for this work under NOAA grant number NA17AE162. The statements contained within this chapter are not the opinions of the funding agency or the US government, but reflect the authors' opinions.

The authors would also like to thank Adam Jessup, cartographer and graphic designer extraordinaire (and a graduate student in Lehman College's MS-GISc Program), for his work on creating the maps, diagrams, and other figures accompanying the text.

References

Bunch, W. (1994). "$52 Million won't flush stink." *New York Newsday*, June 30.

Chaix, B., Kestens, Y., Perchoux, C., Karusisi, N., Merlo, J., and Labadi, K. (2012). "An interactive mapping tool to assess individual mobility patterns in neighbourhood studies." *American Journal of Preventive Medicine*, 43(4): 440–50.

Chaix, B., Meline, J., Duncan, S., Merrien, C., Karusisi, N., and Perchoux, C. (2013). "GPS tracking in neighbourhood and health studies: a step forward for environmental exposure assessment, a step backward for causal inference?" *Health and Place*, 21: 46–51.

Chakraborty, J. and Maantay, J.A. (2011). "Proximity analysis for exposure assessment in environmental health justice research." In Maantay, J.A. and McLafferty, S. (eds.) *Geospatial Analysis of Environmental Health*, Dordrecht, NL: Springer-Verlag, pp. 111–138.

Eicher, C. and Brewer, C. (2001). "Dasymetric mapping and areal interpolation: implementation and evaluation." *Cartography and Geographic Information Science*, 28: 125–138.

Forster, B.C. (1985). "An examination of some problems and solutions in monitoring urban areas from satellite platforms." *International Journal of Remote Sensing*, 6(1): 139–151.

Gladwell, M. (1993). "Letter from the park: play area hailed at first flush now appears ripe for criticism." *Washington Post*, May 28.

Harrison, F., Burgoine, T., Corder, K., van Sluijs, E., and Jones, A. (2014). "How well do modelled routes to school record the environments children are exposed to?: a cross-sectional comparison of GIS-modelled and GPS-measured routes to school." *International Journal of Health Geographics* 13(1): 5.

Holt, J.B., Lo, C.P., and Hodler, T.W. (2004). "Dasymetric estimation of population density and areal interpolation of census data." *Cartography and Geographic Information Science*, 31: 103–121.

Lwin, K.K. and Murayama, Y. (2011). "Accuracy assessment of GIS based building population estimation algorithm." In Murayama, Y. and Thapa, R. (eds.) *Spatial Analysis and Modelling in Geographical Transformation Process: GIS-based Applications*, Dordrecht, NL: Springer-Verlag, pp. 99–112.

Maantay, J.A. and Maroko, A.R. (2009). "Mapping urban risk: flood hazards, race, and environmental justice in New York." *Applied Geography*, 29: 111–124.

Maantay, J.A., Maroko, A., and Herrmann, C. (2007). "Mapping population distribution in the urban environment: the Cadastral-based Expert Dasymetric System (CEDS)." *Cartography and Geographic Information Science*, 34(2): 77–102.

Maantay, J.A., Maroko, A.R., and Porter-Morgan, H. (2008). "A new method for mapping population and understanding the spatial dynamics of disease in an urban area: asthma in the Bronx, New York City." *Urban Geography*, 29(7): 724–738.

Maantay, J.A., Maroko, A.R., and Culp, G. (2009)."Using geographic information science to estimate vulnerable urban populations for flood hazard and risk assessment in New York City." In Showalter, P., and Lu, Y. (eds.) *Geotechnical Contributions to Urban Hazard and Disaster Analysis*, Dordrecht, NL: Springer-Verlag, pp. 71–97.

Mandell, J. (1993). "A state park in Harlem: nature in a surreal setting: a sewage plant by any other name still stinks." *New York Newsday*, May 27.

Maroko, A.R., Maantay, J.A., and Grady, K. (2011). "Using geovisualization and geospatial analysis to explore respiratory disease and environmental health justice in New York City." In Maantay, J.A. and McLafferty, S. (eds.) *Geospatial Analysis of Environmental Health,* Dordrecht, NL: Springer-Verlag, pp. 39–66.

Martin, D. (2006). "An assessment of surface and zonal models of population." *International Journal of Geographical Information Systems*, 10(8): 973–89.

Matthews, S.A. (2011). "Spatial polygamy and the heterogeneity of place: studying people and place via egocentric methods." *Communities, Neighborhoods, and Health*. Dordrecht, NL: Springer-Verlag, pp. 35–55.

Mennis, J. (2003). "Generating surface models of population using dasymetric mapping." *The Professional Geographer*, 55(1): 31–42.

Moon, Z.K. and Farmer, F.L. (2001). "Population density surface: a new approach to an old problem." *Society and Natural Resources*, 14: 39–49.

National Association of City and County Health Officials (NACCHO). (2011a). *A Neighborhood Fights Back: Building the North River Waste Water Treatment Plan in West Harlem*. www.rootsofhealthinequity.org

National Association of City and County Health Officials (NACCHO). (2011b). *Polluting Sites in Northern Manhattan*. www.rootsofhealthinequity.org/polluting-sites-in-manhattan.php

Poulsen, E. and Kennedy, L.W. (2004). "Using dasymetric mapping for spatially aggregated crime data." *Journal of Quantitative Criminology*, 20(3): 243–62.

Wu, S., Qiu, X., and Wang, L. (2005). "Population estimation methods in GIS and remote sensing: a review." *GIScience and Remote Sensing*, 42(1): 80–96.