Effects of Neighbourhood Demographic Shifts on Findings of Environmental Injustice: A New York City Case-Study

Author(s): Makram Talih, Ronald D. Fricker and Jr

Source: *Journal of the Royal Statistical Society. Series A (Statistics in Society)*, Vol. 165, No. 2 (2002), pp. 375-397
Published by: Wiley for the Royal Statistical Society
Stable URL: http://www.jstor.org/stable/3559933
Accessed: 02-02-2017 01:26 UTC

J. R. Statist. Soc. A (2002)
165, *Part* 2, *pp.* 375–397

Effects of neighbourhood demographic shifts on findings of environmental injustice: a New York City case-study

Makram Talih

Yale University, New Haven, USA

and Ronald D. Fricker, Jr

RAND, Santa Monica, USA

[Received March 2001. Revised December 2001]

Summary. We study the question of how changes in neighbourhood demographics affect findings of environmental equity. Many cross-sectional studies of association between neighbourhood racial and ethnic composition and the location of environmentally undesirable sites have been conducted. However, no evaluations have been conducted that examine how neighbourhood demographics change over time, and how those changes are related to the observed cross-sectional results. If the question is whether an observed association is the result of discrimination, it is crucial that the historical changes in neighbourhood structure are well understood. We develop some methods based on standard statistical techniques and illustrate their application by using the metropolitan New York City region as a case-study.

Keywords: Discrimination; Economic distance; Environmental equity; Exposure surface; Pollution; Spatial–temporal clustering

1. Introduction

Two distinct approaches have emerged in analyses of environmental justice. The first seeks to determine whether there is an association between racial demographics and the location of environmentally undesirable sites. The second examines how such an association may have occurred. We refer to the former approach as an *outcome-based* analysis and to the latter as a *process-based* analysis. Outcome-based analyses seek to describe the current state of a neighbourhood or region, to identify those populations that are currently most affected or at risk. Process-based analyses, in contrast, attempt to explain how a current state of association came to be. Often, those who conduct outcome-based analyses implicitly assume that a finding of association is evidence of discrimination. Similarly, those who conduct process-based analyses often implicitly assume that a historical lack of association demonstrates a lack of discrimination. Either assumption can be in error.

As in Fricker and Hengartner (2001), to distinguish clearly between the types of analysis, we define the study of environmental equity as the determination of whether existing environmentally undesirable sites and the distribution of various racial or ethnic populations is fair, in the sense that the sites (or their effects) are or are not systematically concentrated in (or imposed

Address for correspondence: Ronald D. Fricker, Jr, RAND, 1700 Main Street, PO Box 2138, Santa Monica, CA 90407-2138, USA.
E-mail: Ron_Fricker@rand.org

0964–1998/02/165375

on) one or more racial or ethnic minorities. We contend that the existence of environmental inequity does not imply the larger, and often unspoken, proposition that the minority communities are unfairly imposed on *because they are minority communities*. For such a result, we use the term 'environmental discrimination' to emphasize that such a determination is a conclusion of causation, not simply of association. Finally, we say that environmental justice encompasses both environmental equity and environmental discrimination.

With these distinctions, we say that an outcome-based analysis may result in a determination of environmental inequity but not discrimination. Only a process-based analysis can address the question of environmental discrimination. Most analyses in the literature are cross-sectional, linking the location of polluting sites to population demographics for one period in time. They are thus outcome based and can therefore only evaluate environmental equity. We consider a process-based analysis to be the natural extension of a finding of inequity in an outcome-based analysis. Therefore, in this work, we extend the previous environmental equity analysis of metropolitan New York City of Fricker and Hengartner (2001) in a process-based analysis of neighbourhood demographic shifts. The data that are analysed can be obtained from

`http://www.blackwellpublishers.co.uk/rss/`

2. Background and related research

We use metropolitan New York City in the USA as a case-study to illustrate our methods. Although the background literature focused mainly on studies conducted in the USA, the question of environmental justice is a global issue. We believe that the issue eventually must arise in any society that has had a historically segregated minority population that subsequently has acquired sufficient political power to demand redress of either perceived or actual societal inequities.

2.1. Theories of population change and environmental injustice

Liu (1997) described four theories of neighbourhood change: the classical invasion–succession model, the neighbourhood life-cycle model, the push–pull model and the institutional theory of neighbourhood change. The invasion–succession model is based on biological ecology and suggests that the composition of a neighbourhood changes as the density of a minority group increases until it becomes the majority or dominant group. This model implies that racial changes in a neighbourhood may not be related to environmentally undesirable sites. The model is best characterized by the 'white flight' phenomenon, in which a predominantly white neighbourhood's residents leave in response to a small influx of black residents. As a result, the community switches from being virtually all white to all black.

The neighbourhood life-cycle model suggests that communities go through natural cycles of development, transition, aging, decay and renewal. At each stage, the neighbourhood sees a change in the socioeconomic status of the residents and perhaps other demographic shifts. The model posits that the age of the housing stock and population density are the two most important community characteristics that determine neighbourhood change. As with the invasion–succession model, the assumption here is that the neighbourhood demographic changes may not be related to environmentally undesirable sites.

In contrast, the push–pull model assumes that there are two sets of forces that push and pull residents out of and into a neighbourhood. Under this model, the siting of an environmentally undesirable facility is assumed to push out various residents, probably affluent, and to pull in other residents, probably the less affluent, through their effects on neighbourhood desirability and economics (particularly property values). Alternatively, the institutional theory of neighbourhood change suggests that institutions play a large role in neighbourhood change, but that change does not necessarily have to be negative. It also suggests that the siting of facilities may be attracted to neighbourhoods in decline; thus they may not necessarily be the cause of decline but could be the result of a neighbourhood in decline.

Daniels and Friedman (1999) proposed three explanations for environmental injustice: racial discrimination, economic stratification and urban ecology. Racial discrimination means that minority communities were either directly or indirectly unfairly imposed on because they were minority communities. Economic stratification implies that unequal distribution of pollution occurs as part of market-driven processes. Urban ecology posits that urban development influences the geographic location of both residents and industries.

We tie these sets of theories together as follows. First, racial discrimination, in the form of overt discrimination in the siting of facilities, can only occur under the push–pull and institutional theory of neighbourhood change models because, as we previously discussed, the other two theories of neighbourhood change are not related to the siting of facilities. This implies, for example, that if invasion–succession is the dominant form of neighbourhood change, then it is less likely that explicit discrimination in siting facilities explains an observed case of inequity. In contrast, if push–pull is the dominant form of change, and the push–pull results from the siting of facilities, then some type of overt discrimination may have occurred.

Second, economic stratification may occur as a result of the neighbourhood life-cycle, push–pull and institutional theory of neighbourhood change models, all of which may or may not be related to or caused by indirect discrimination. However, we note, as Daniels and Friedman (1999) did, that the direction of causation in economic stratification is not clear without additional study. It is possible that sites are placed in less economically advantaged locations for appropriate business reasons, and it is also possible that sites cause changes in the region's economic status. Furthermore, the larger societal issue of existing compensation differentials by race may contribute to economic stratification and thus environmental inequity.

Third, urban ecology encompasses all the neighbourhood change models. Basically, under the urban ecology theory, businesses seek to locate in regions that are best suited to conducting business, so, for example, transportation-dependent or intensive businesses situate near transportation arteries etc. Similarly, neighbourhood populations will locate according to preferences such as newer, less dense suburbs over older, more dense urban areas. In general, urban ecology implicitly assumes that overt discrimination is not present, though the resulting interaction of business and population dynamics may represent a form of subtle discrimination.

2.2. Previous longitudinal environmental justice studies

Most environmental justice studies use simple univariate statistical tests (Sexton *et al.*, 1993) to evaluate the cross-sectional association between the location of environmentally undesirable sites and the racial demographics at and around the site locations. A few studies have conducted longitudinally evaluations of observed associations. These studies include Mitchell *et al.* (1999), Been and Gupta (1997), Liu (1997), Oakes *et al.* (1996) and Been (1994). The basic goal of these studies is to determine which came first, the sites or the minority populations. They generally proceed by conducting additional cross-sectional analyses in prior time periods to determine whether, for the sites and demographics of those periods, the currently observed associations continue to hold.

For example, using toxics release inventory (TRI) facilities in South Carolina, Mitchell *et al.* (1999) compiled census information about each area in which a site was contained for as long as the site was in existence. The historical demographic profiles of the site areas were then compared with average state-wide racial profiles. The TRI is a publicly available database, maintained by the United States Environmental Protection Agency (USEPA). It contains information on chemical emissions from over 20000 facilities in the USA, tracking over 600 toxic chemicals. It receives reports on chemicals from the largest manufacturers, as well as some Government-operated facilities. A facility that employs 10 or more full-time workers, and whose activities fall within the standard industrial classification codes 20–39, is required to report to the TRI if it uses at least 10000 lb per year of any listed chemical.

Other examples include a longitudinal study by Been and Gupta (1997) of the community demographics before and after the siting of 608 hazardous waste treatment, storage and disposal facilities (TSDFs), using simple univariate statistical tests in combination with linear and logistic regression. Liu (1997) appraised the siting of nine solid waste facilities in Houston, Texas, by

(a) comparing neighbourhood demographics before and after the siting of the facilities and
(b) evaluating neighbourhood demographic changes in the pre- and post-siting periods.

Comparisons were made between the tracts containing the solid waste facilities and 'control' tracts selected to have similar characteristics to the site tracts. Oakes *et al.* (1996) linked 1970, 1980 and 1990 census data to TSDFs operating in 1992. In their work, they conducted three types of analyses:

(a) cross-sectional tests for differences in demographic characteristics;
(b) longitudinal panel comparisons between TSDF tracts and selected non-TSDF tracts;
(c) longitudinal evaluations of tract characteristics before and after a facility had been sited.

Been (1994) revisited two previous studies, one by the US General Accounting Office (1983) of landfills in USEPA region IV and the other by Bullard (1983) of incinerators and landfills in Houston, Texas. Been evaluated whether market dynamics played a significant role in creating the observed association between sites and race. The analysis was motivated by the observation that almost a fifth of all US households move each year, so the demographics around sites can change significantly over the life of a facility. Therefore, if site location and market forces cause changes in the demographics surrounding the sites, then modifying the siting process to be more equitable will not be an effective long-term solution. Been's analysis consisted of simple comparisons of the percentage of the population in the immediate vicinity of each site that was black with the percentage of the county and state that were black, both before and after the siting of a facility. No formal statistical tests were done in the analysis.

Our work differs from previous researchers' in that we define new methods to capture general population trends and to describe the evolution of *neighbourhoods*. Designed to investigate processes, our methods shed light on whether observed changes in neighbourhoods are consistent or inconsistent with the various theories of change: invasion–succession, life-cycle, push–pull or institutional theory. The previous longitudinal analyses used arbitrary units of analysis, such as census tracts, failing to consider explicitly the actual neighbourhoods in the models and tests; thus they ignore the natural spatial structure in the data. We also differ from most other analyses in that we use manufacturing zones as surrogates for all types of environmentally undesirable sites, rather than focusing on a single type of site. This reflects a more robust definition of environmental justice, as discussed in Bullard (1996).

This work is an extension of Fricker and Hengartner (2001). Using an outcome-based approach, they studied the question of environmental equity in the metropolitan New York City in 1990. In particular, Fricker and Hengartner used population demographics for 2216 census tracts linked to 354 environmentally undesirable facilities, including TRI sites, TSDFs and other common urban problem sites such as landfills, incinerators, bus garages and sewage treatment plants. Fricker and Hengartner found that racial or ethnic demographics, in particular the percentage of Hispanics in a tract's population, were significantly associated with the presence of potentially environmentally adverse sites in 1990. They concluded that in 1990 both Hispanics and non-Hispanic blacks were living more closely around sites in Queens. In Brooklyn they found that tracts with a minority Hispanic population are more proximate whereas tracts that are predominantly Hispanic (more than 80%) may be less proximate.

3. A methodology for determining how neighbourhoods change over time

Here we study how the neighbourhood demographics have evolved between 1970 and 1990 in Queens and Brooklyn to understand how previously observed associations may have occurred. We evaluate the evolution of neighbourhoods in comparison with the location of manufacturing zones. The use of manufacturing zones allows us to capture the effects of living next to a whole host of undesirable sites, not just TRI sites or some other subset of undesirable sites.

In particular, we use the location of manufacturing zones to characterize the mechanism by which the location of environmentally unfriendly sites occurs. To justify this, we demonstrate that the observed locations of TRI sites in Queens and Brooklyn are consistent with this assumption. Ultimately, we derive a 'manufacturing index' for each census tract in Queens and Brooklyn that we interpret as a measure of how proximate a population is to various environmentally undesirable sites. Finally, using clustering to group census tracts by demographic time trends, we graphically display how neighbourhoods evolve over time and link the manufacturing index derived to the clusters of tracts to provide insights into the complex interactions of people, geography and industry. Thus, rather than using census tracts in and of themselves as units of analysis, as is done elsewhere in the literature, we argue that appropriately defined clusters of tracts—which we take as a proxy for neighbourhoods—provide the most appropriate unit to study.

Our approach is demonstrated using the Brooklyn and Queens boroughs of New York City as a case-study. Those two boroughs, indeed New York City as a whole, are unique in many ways. Two features of particular importance for the methods that we employ in this analysis are

(a) the population has changed greatly in the last few decades, while
(b) the industrial areas are known to have been relatively fixed for a long period of time, at least as far as the two boroughs are concerned.

For example, between 1970 and 1990 the combined five boroughs of New York City saw a decrease of 7% in the total population, from almost 7.9 million in 1970 to just over 7.3 million in 1990. Furthermore, this decrease was not gradual but was concentrated between 1970 and 1980 when the city experienced a decrease of more than 10% in population. Following that decrease, the city then saw a modest increase in population between 1980 and 1990.

Furthermore, the changes in population were not proportional by race nor by borough. From 1970 to 1990 the entire city experienced a drop of 36% in its non-Hispanic white population, a drop of 66% in its non-Hispanic black population and an increase of 22% in its Hispanic population. For the two boroughs that are the subject of this paper, Queens saw a drop of 43% in its non-Hispanic white population, a decrease of 5% in its non-Hispanic black population and an increase of 105% in its Hispanic population. In Brooklyn, the non-Hispanic white population decreased by 40%, the non-Hispanic black population by 80% and the Hispanic population by 3%, for an overall population decrease of 12%. Minorities other than Hispanics and non-Hispanic blacks were not analysed separately as they could not be determined consistently from the three decennial censuses between 1970 and 1990.

3.1. Zoning as a mechanism for siting: defining a manufacturing index

Almost all previous environmental justice studies have focused on a subset of environmentally undesirable sites. Generally these sites are either TRI sites or TSDFs. These sites are often used in analyses because they are readily available from administrative databases. However, the databases are subject to various criticisms, including the following.

(a) Data elements, such as the locations of facilities, are prone to significant errors. For example, latitude and longitude co-ordinates may be inaccurate, so much so that some of our New York facilities were seemingly located in the East River. Facility addresses may also be incorrect, in that they may be corporate mailing addresses and not the addresses of polluting facilities.
(b) Facilities only qualify for reporting to the USEPA if they exceed certain polluting thresholds. Thus, smaller facilities and facilities that do not produce materials in excess of reporting requirements do not appear in the data.
(c) Reporting requirements have not been in existence for very long (the mid-1980s at the earliest) so it is difficult to conduct a detailed historical process-based analysis that goes very far back in time. Further, gathering historical data about the operation of facilities on a large scale is difficult if not impossible. In fact, we tried for New York City and found that there were no reasonably accessible and consistent records with sufficient data that were relevant for this type of analysis from which to draw.

We overcame these difficulties by focusing on zoning. Zoning has remained fairly unchanged in Brooklyn and Queens since the 1960s, which allows us to study more easily the dynamics of neighbourhood change in comparison with a fixed reference, so that we can look for identifiable patterns of neighbourhood change in and around the traditional manufacturing districts. Furthermore, because zoning has remained fixed, we can reasonably dismiss the possibility that zoning was used as a mechanism for environmental discrimination. For other boroughs or municipalities in which zoning has significantly evolved over time, this may be more problematic.

The first New York City zoning resolution was created in 1916 in response to overwhelming development in Lower Manhattan (New York City Department of City Planning, 1990). Before zoning, industrial areas were developed as mixed use areas, with housing of lower calibre built around manufacturing plants for workers and their families. Owing to major shifts in land use and population migrations, the present resolution was adopted in 1961. On the basis of the lay-out of the city at the time that it was enacted, the 1961 resolution separated residential and commercial areas from manufacturing districts. However, as the base of manufacturing jobs has continued to decrease in the boroughs (240000 in 2001, compared with 750000 in the late 1960s), housing demand has increased drastically, putting pressure on the redevelopment of areas that were originally zoned for manufacturing. In spite of this, zoning has remained essentially fixed in the Brooklyn and Queens boroughs of New York City.

Using the latest version of the zoning maps (New York City Department of City Planning, 1999), we manually coded the census tracts in Queens and in Brooklyn (because they have not yet been digitized) as either 'manufacturing tracts' or 'non-manufacturing tracts'. We define a manufacturing tract as a census tract containing within its boundaries any part of one or more manufacturing districts. In Queens and in Brooklyn, about one in five tracts were coded

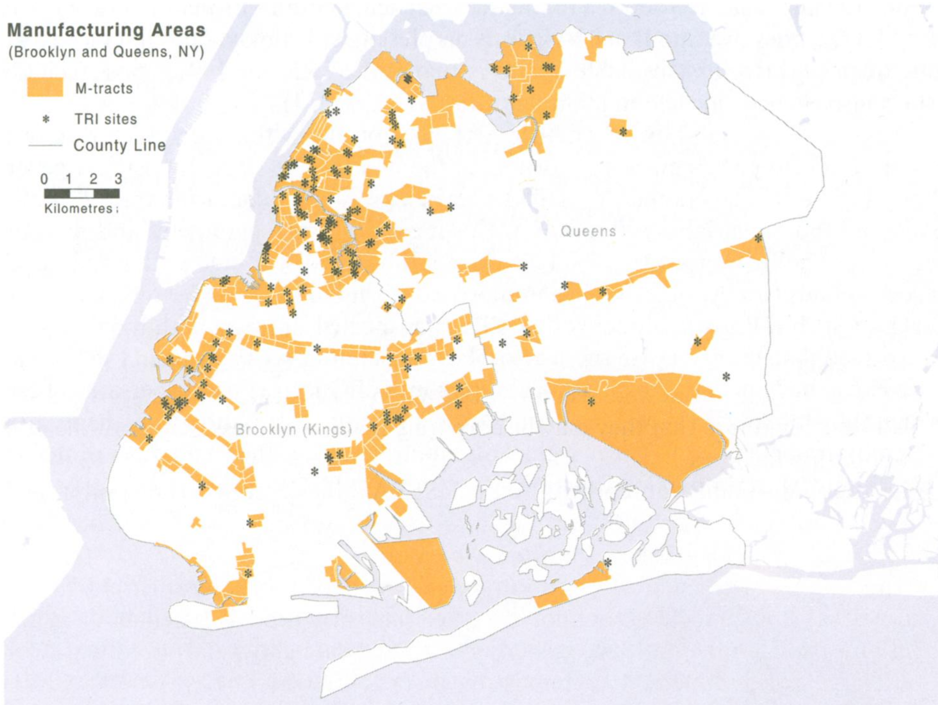


Fig. 1. Manufacturing areas in Brooklyn and Queens, New York: 270 out of 1436 tracts are manufacturing tracts; TRI facilities (a particular type of manufacturing facility) are given special attention in the environmental justice literature—here we see that they have almost exclusively been sited in manufacturing tracts, which is consistent with the New York City zoning resolution; in fact only five of the 150 TRI facilities are not located in manufacturing tracts and these have located in commercial districts where some sort of light manufacturing is allowed

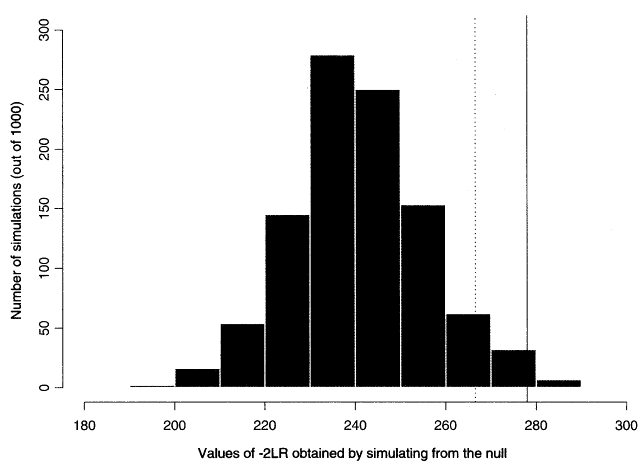
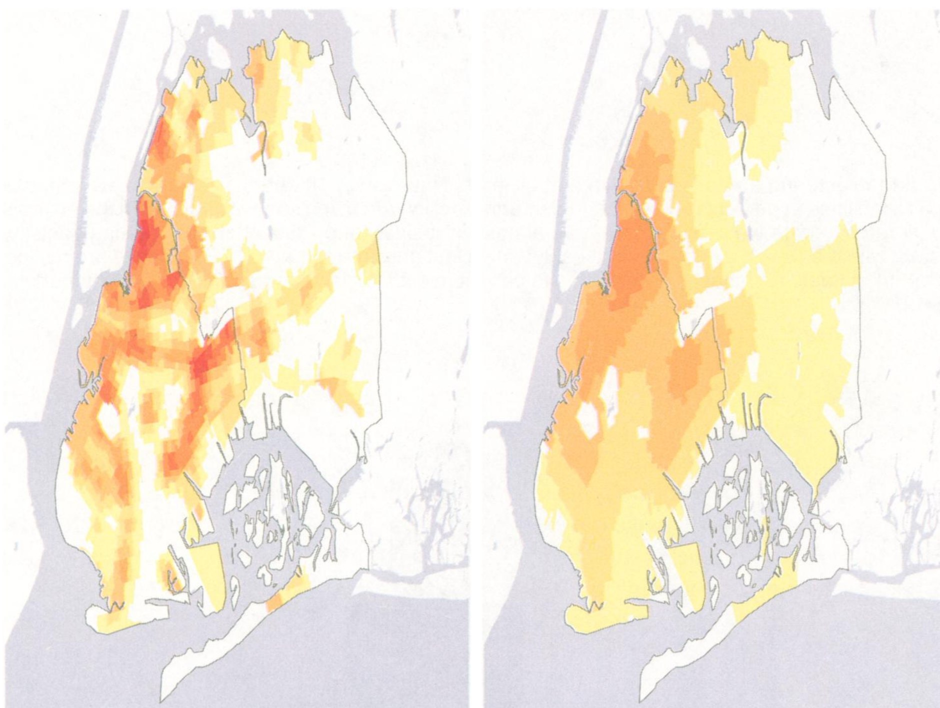


Fig. 2. Result of the generalized likelihood ratio test for the null hypothesis that TRI sites are uniformly distributed within manufacturing tracts according to $N_i | \lambda_i, A_i \sim \text{Poisson}(\lambda_i A_i)$, and $\lambda_i \sim \text{gamma}(\alpha, 1)$: the test rejects the null hypothesis (⋮, 95th percentile; |, observed value)

as manufacturing tracts. However, looking at a particular brand of manufacturing facilities, namely TRI facilities, we find that 150 have been present in Brooklyn and in Queens, at some point between 1987 (the first available reporting year in the TRI) and 1997. Those facilities have almost exclusively been located in manufacturing tracts (Fig. 1).

We investigated the question of whether TRI sites located within manufacturing tracts are distributed uniformly therein, in proportion to the area of the tract, by using a generalized likelihood ratio test. The number N_i of TRI sites in a given manufacturing tract T_i is modelled as a Poisson random variable with mean $\lambda_i A_i$, where λ_i is a rate parameter and A_i is the area of each tract. The hypothesis that TRI sites are uniformly distributed within manufacturing tracts corresponds to a hypothesis of a common underlying rate: $\lambda_i = \lambda$ for all i. One important drawback of such a Poisson model is that TRI sites located in tracts with large areas appear more dispersed than in tracts with small areas. For an example, see the tract in Fig. 1 containing the John F. Kennedy international airport in the south of Queens. Overdispersion is best taken into account by looking on each λ_i as arising from a gamma distribution with shape parameter α_i and scale parameter $\beta = 1$. The marginal distribution of N_i is then a member of the negative binomial family, with probability function



(a) (b)

Fig. 3. Manufacturing index in Brooklyn and Queens, New York: (a) levels of the manufacturing index resulting from a Gaussian kernel smoother with bandwidth 500 m; (b) levels of the index from a 1.5 km bandwidth (the larger the bandwidth, the more gradual is the decrease in potential exposure as we move away from the manufacturing poles; the maximum height of the surfaces is 1.6–1.8% (red), and the minimum height is 0–0.2% (yellow); the range 0–1.8% is divided into nine disjoint intervals of length 0.2% each, and levels smaller than 0.05% have been omitted; the heights shown are the values of the manufacturing density at the centroid of each tract)

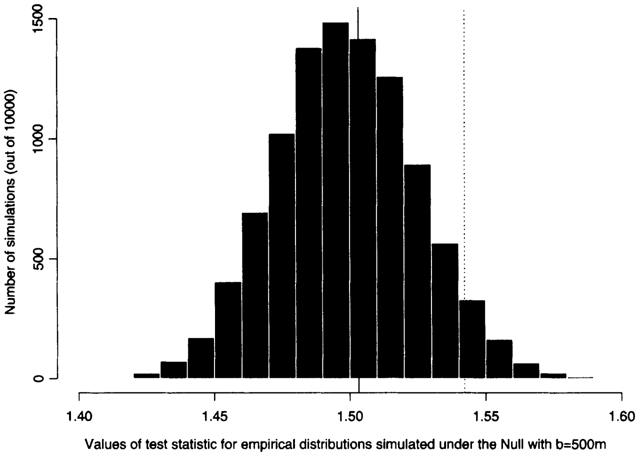


Fig. 4. Result of the maximum variation distance test, for the null hypothesis that $\mathbb{P}(A$ contains a TRI site$\,|\,H_0) = P_b(A)$, where $P_b(A) = \mathbb{P}(A$ contains a manufacturing district$\,|\,b)$: the test statistic is $\tau = \sup_{A\in\mathcal{A}} |P_b(A) - P_n(A)|$; the significance level of τ is estimated by simulating values $\tau_1^*, \ldots, \tau_K^*$, corresponding to $\tau_k^* = \sup_{A\in\mathcal{A}} |P_b(A) - P_n^{*k}(A)|$; the simulated empirical distributions $P_n^{*k}(A)$ are obtained by sampling from $P_b(A)$ with replacement n times (where n is the total number of TRI sites; the test supports the null hypothesis (⁝, 95th percentile; |, observed value)

$$p(n_i\,|\,\alpha_i) = \frac{\Gamma(\alpha_i + n_i)}{\Gamma(\alpha_i)n_i!}\left(\frac{A_i}{1 + A_i}\right)^{n_i}\left(\frac{1}{1 + A_i}\right)^{\alpha_i}.$$

Thus $\mathbb{E}(N_i) = \alpha_i A_i$ whereas the variance is $\alpha_i A_i + (\alpha_i A_i)^2/\alpha_i$, which allows for extra-Poisson variation due to variables other than area alone. Dean and Lawless (1989), as well as references therein, discussed overdispersion and developed tests for detecting extra-Poisson variation. On the basis of the statistic $\Sigma_i N_i$, we have instead calculated very crude method-of-moments estimates of the mean and variance, and have found it reasonable to adopt the negative binomial model.

The hypothesis that TRI sites are uniformly distributed within manufacturing tracts now corresponds to a hypothesis of a common underlying *expected* rate: $\alpha_i = \alpha$ for all i. Maximum likelihood estimates $\hat{\alpha}$ of α under hypothesis H_0 (and of each α_i under H_1) are not available in closed form; however, they can certainly be obtained numerically. Also, as the asymptotic χ^2-approximation for minus twice the log-likelihood ratio statistic LR does not hold, we calculated an approximate p-value by simulating the test statistic under the null hypothesis, first generating the λ_i according to a gamma$(\hat{\alpha}, 1)$ distribution, and subsequently the N_i^{*k}, $k = 1, \ldots, K$, according to a Poisson(λ_i) distribution. Fig. 2 shows that the generalized likelihood ratio test rejects the null hypothesis at the 5% significance level. Rejection occurs because the observations are spatially clustered: the large value of the test statistic $(-2\,\mathrm{LR} \approx 278)$ indicates that TRI facilities are clustered within certain parts of the manufacturing areas. This is consistent with Fricker and Hengartner (2001) who showed that TRI sites were not distributed as a thinned Poisson process of all the sites that they considered, owing to similar clustering.

We contend that TRI facilities cluster as follows: the number of TRI sites in any given region is proportional to the degree of industrialization in and around that region. We calculate our measure of industrialization in a region as a function of the density of manufacturing districts thereabouts by using kernel smoothing. Kernel smoothing is a convenient method that allows each tract's industrialization measure to be influenced by its neighbours', both in terms of the number of manufacturing tract neighbours and in terms of how close those manufacturing tracts are to the tract in question. In general, tracts that have many close manufacturing tract neighbours will have a higher industrialization measure than those that do not have many manufacturing tract neighbours. Further, a non-manufacturing tract in the middle of a heavy manufacturing region can have a higher measure than an isolated manufacturing tract. Thus, kernel smoothing results in a measure that incorporates spatial relationships and that better reflects the way that residents would actually judge the degree of industrialization in an area.

Note that New York City zoning distinguishes between light (M1), medium (M2) and heavy (M3) manufacturing districts. The M1 districts are typically buffer zones isolating the heavier manufacturing activity from residential areas. In aggregating those districts to the tract level, we do not distinguish between the different types of manufacturing districts in the designation of manufacturing tracts. By suitably smoothing over manufacturing tracts, however, we tend to recover central manufacturing areas because the manufacturing index that we construct is higher at the manufacturing poles (which are typically M2 or M3) and gradually decreases outwards from the poles (corresponding to the idea of M1 districts).

From an exposure assessment perspective, the gradual decrease of the manufacturing index corresponds to a gradual decrease in potential exposure to manufacturing activities. Of course, our measure of exposure here is determined by proximity alone, as the specifics of TRI or other activities within the manufacturing districts (such as quantities produced or released and types of chemicals used) are very difficult to determine. Furthermore, the aggregation of all the individual pollutants, even if they were known, into a coherent single measure of exposure would be difficult, if not impossible, though such aggregate measures warrant further research.

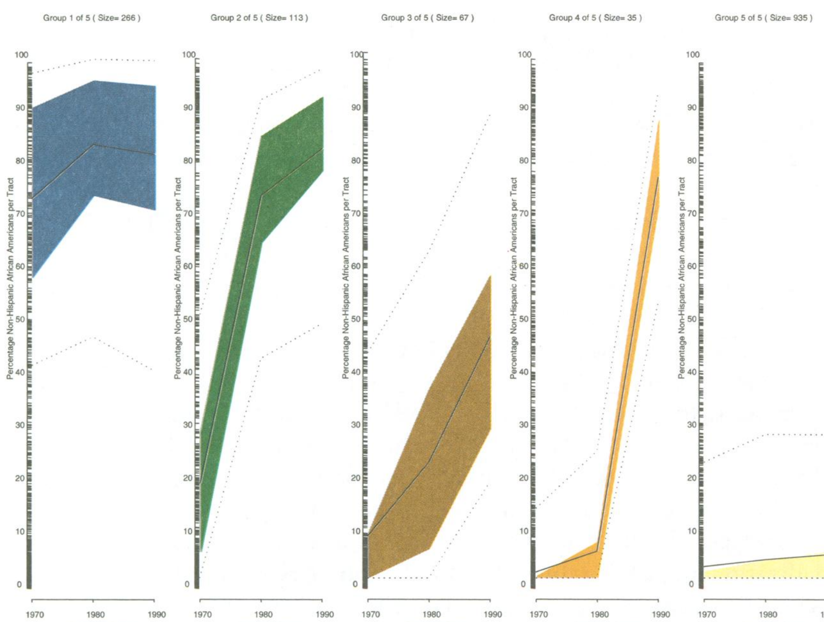


Fig. 5. Tract clustering in terms of changes in percentage of non-Hispanic blacks: five groups of tracts are obtained; the figure is divided into five panels; the census years are on the *x*-axis; the percentage of non-Hispanic blacks, ranging from 0% to 100%, lies on the *y*-axis; each panel contains information about the size of the cluster, as well as the cluster means (———), interquartile ranges (shaded areas), fifth and 95th percentiles (·······)

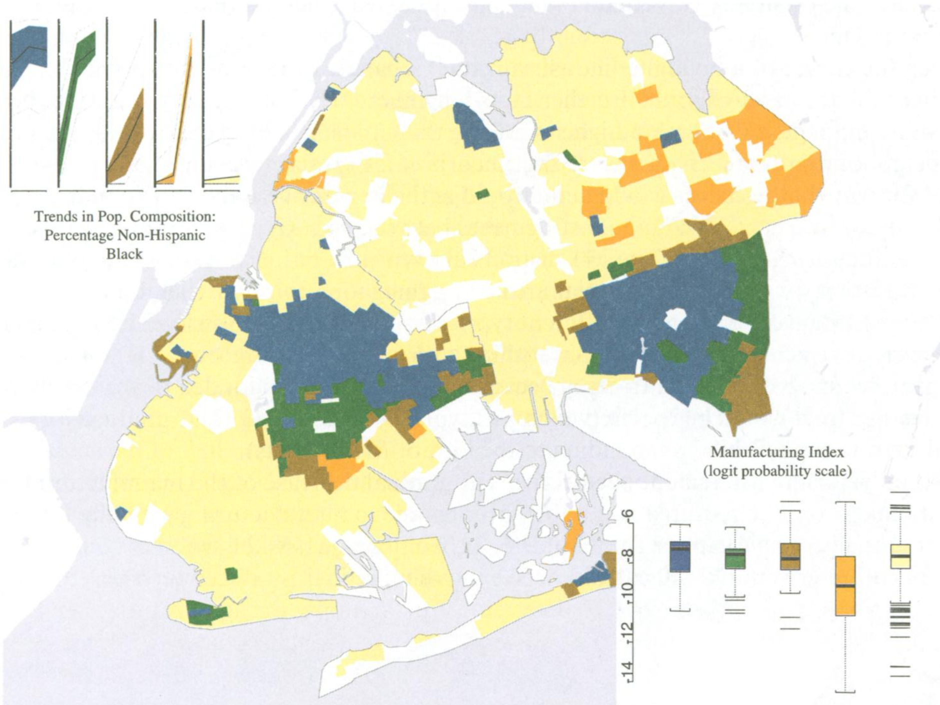


Fig. 6. Tract clustering in terms of changes in percentage of non-Hispanic blacks: the major time trends map to meaningful parts of the city, namely the Bedford-Stuyvesant, Brownsville and Crown Heights quarters in Brooklyn and the Jamaica, Hillside and Saint Albans quarters in Queens; note that non-Hispanic black neighbourhoods are expanding *away* from manufacturing poles, towards Kensington in Brooklyn and towards Cambria Heights in Queens; furthermore, some affluent non-Hispanic black neighbourhoods are beginning to develop in wealthier areas—even further from the manufacturing regions—such as Oakland Gardens and Glen Oaks in the north-east of Queens

To calculate the kernel smoother, we centre a Gaussian kernel at each manufacturing tract. Each kernel is defined over the grid of points formed by the 1436 tract centroids. A finer grid of points, consisting of block centroids, provides 40–60 points on average within each tract, but to determine manufacturing blocks *versus* non-manufacturing blocks from the zoning maps has proved prohibitive, since those have not yet been digitized. The axes of each kernel are aligned with the east–west and north–south axes obtained via the projection of the longitude–latitude co-ordinate system onto the US state plane (North American Datum (of 1983)) rectangular co-ordinate system for the New York Long Island zone (Federal Information Processing Standards zone 3104). We assume independence of the east–west and north–south axes as far as siting is concerned, since we do not have additional information to suggest that direction is important in this regard.

By renormalization, the support of each Gaussian kernel is restricted to the subset of the plane spanned by the Brooklyn–Queens landmass; areas such as public parks and cemeteries, as well as water areas, where there is neither population nor manufacturing, are removed. Renormalization is achieved by redistributing any excess mass uniformly within the region of interest. In particular, manufacturing tracts on the boundary will contribute more excess mass than manufacturing tracts that are in the centre of the region. By redistributing that excess mass equally among all the other tracts, we seek to reduce the bias that kernel smoothing tends to introduce at the boundary. Silverman (1986) has discussed other alternatives for bias reduction at the boundary.

Given the choice of a bivariate Gaussian kernel, a bandwidth b must be specified. Selecting b involves the trade-off of appropriately smoothing over the whole region to capture the main areas of manufacturing, while not oversmoothing so that the resulting measure also reflects the local neighbourhood of a given tract (Silverman, 1986). Furthermore, the choice of bandwidth should fundamentally reflect how people would actually perceive proximity to manufacturing areas. The contour plots in Fig. 3 show the result of Gaussian kernel smoothing on the manufacturing tracts with bandwidths of 500 m and 1.5 km. (Surfaces corresponding to intermediate bandwidths that we also evaluated are not shown.) To fix ideas, we note that it takes about 5–10 min walking to travel 500 m, whereas it takes about 5–10 min driving to travel 1.5 km in the city.

On average, tract centroids are separated by a distance of about 350 m. If, for the sake of argument, we suppose that the tracts are square, then a disc of radius 525 m around the centre of an average tract would be required to cover the four neighbouring tracts that would share its boundaries. Indeed, Table 1 shows that a disc of radius 500 m around half of the actual Queens and Brooklyn tracts will include their three or four nearest neighbours. However, a radius larger than 500 m will be required to smooth over enough adjacent tracts, especially for large or elongated tracts. As we can see from Table 1, 22% of the tracts will have either only one or no nearest neighbour within a 500 m window. We thus choose in this work to set $b = 1.5$ km, which

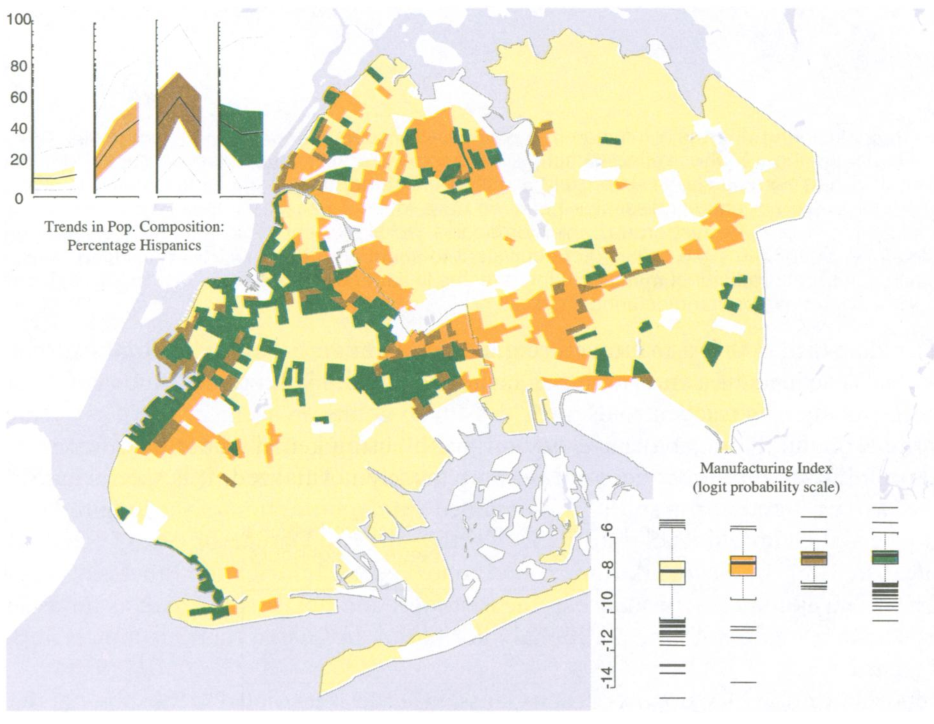


Fig. 7. Tract clustering in terms of changes in percentage of Hispanics: Hispanic communities are more integrated than non-Hispanic black communities, yet they remain on the rim of those communities; whereas the Hispanic communities in Brooklyn have been relatively stable, at about 35–40% of the tract population (group 4), those in Queens are much newer, and they have expanded from about 10% to about 40% between 1970 and 1990 (group 2); Hispanic communities have tended to develop near the traditional manufacturing regions in the city, which means that *a posteriori* they appear to bear the burden of hosting undesirable land uses

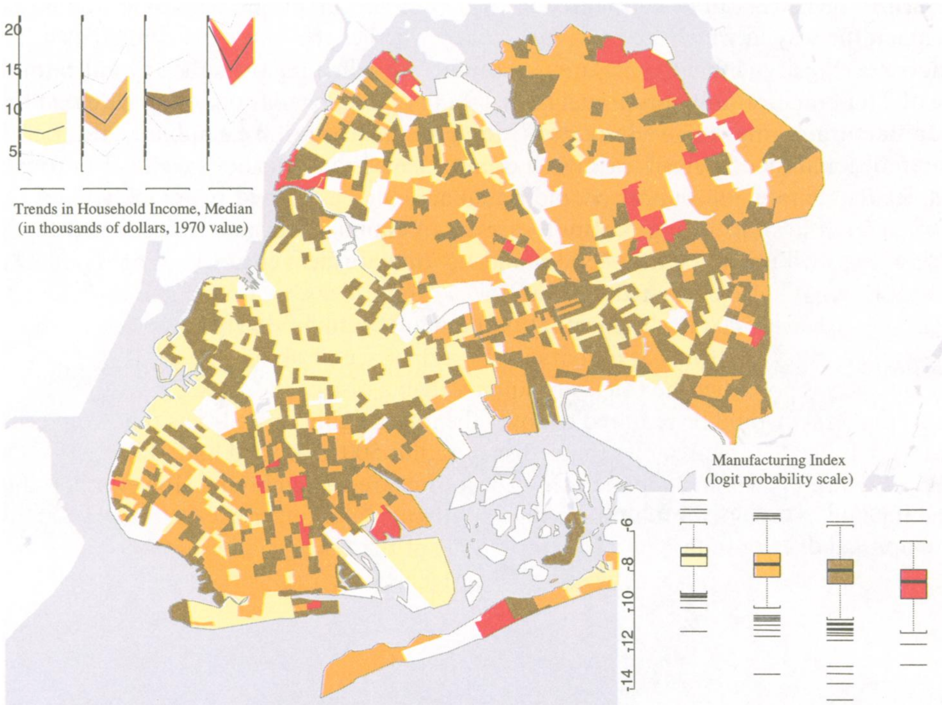


Fig. 8. Tract clustering in terms of inflation-free changes in median household income: the economic downturn at the end of the 1970s is reflected across all income brackets, in varying degrees of intensity; the clustering of curves here has been essentially on the levels, rather than on the slopes; although the second and third groups are not entirely dissimilar here, we have found it necessary to take $K = 4$ in our clustering procedure to isolate the last group, which embodies the higher end of the income distribution (with a *lower bound* of \$20000 in 1970 dollars, i.e. equivalent to about \$80000 now); it is no surprise that income remains an important predictor of manufacturing rate, the lower end of the income distribution being typically associated with the higher manufacturing index

reflects both the idea that a manufacturing district within a 5–10 min drive will be perceived as 'close' by residents of a tract and that such a bandwidth will provide sufficient smoothing between neighbouring tracts.

Having defined a manufacturing index surface by using kernel smoothing, which can also be interpreted as a probability density if appropriately normalized, it is then natural to ask

Table 1. Percentage of tracts around which a 500 m window includes k nearest neighbours

Number k of tracts within disc	*Frequency (%)*
0	10
1	12
2	21
3	27
4	21
5	7
6	2

whether the TRI sites can be considered to be random observations according to this distribution, much the way in which we previously asked whether they were distributed as a function of tract area. We argue that the degree of manufacturing in a region is the key indicator for the siting of TRI facilities. Of course, such is the role of zoning regulations: to ensure that the bulk of manufacturing activity is well isolated from residential areas. We emphasize this point here, as the traditional focus of the literature has been the siting of facilities *per se*. In probabilistic terms, let $B \subset \mathbb{R}^2$ denote the closed and bounded region delimited by Brooklyn and Queens. The tracts $T_1, \ldots, T_{1436}$ form a partition of B into disjoint areas, and subsets of B that are of interest when working at the tract level of aggregation are elements of $\mathcal{A} = \sigma(\{T_1, \ldots, T_{1436}\})$. For $A \in \mathcal{A}$,

$$P_b(A) = \mathbb{P}(A \text{ contains a manufacturing district} \mid b)$$

is the probability distribution of manufacturing districts, estimated from the manufacturing tracts with a particular choice of bandwidth b, as discussed earlier. Our null hypothesis H_0 is thus

$$\mathbb{P}(A \text{ contains a TRI site} \mid H_0) = P_b(A).$$

To test the null hypothesis, we compare the hypothesized null distribution of the TRI sites with their empirical distribution

$$P_n(A) = \frac{\#(\text{TRI sites in } A)}{n},$$

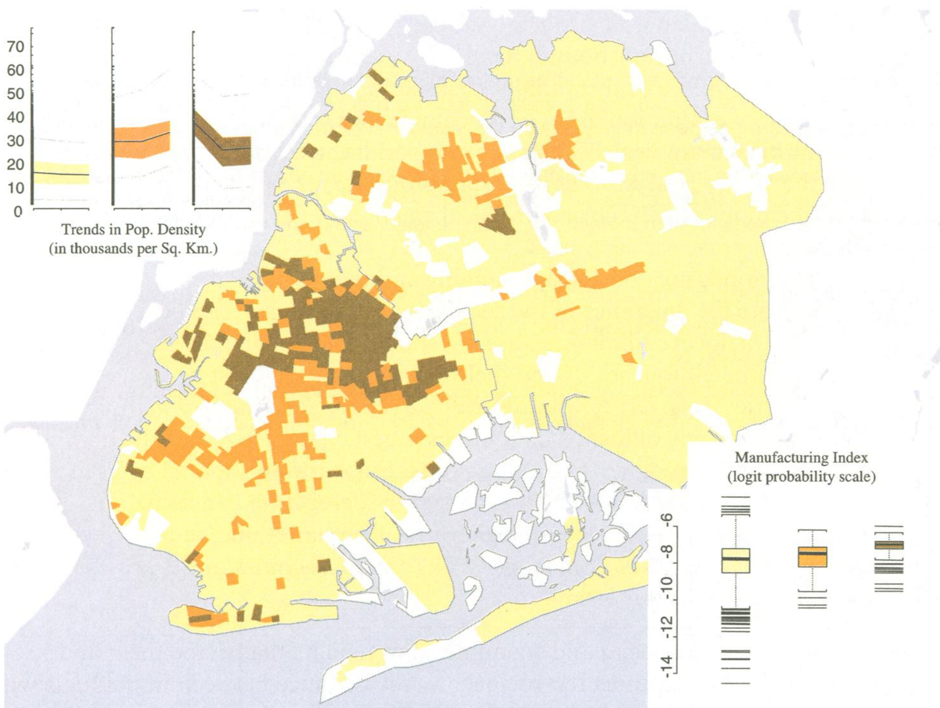


Fig. 9. Tract clustering in terms of changes in population density: according to the life-cycle model, a decrease in population density may be a precursor of neighbourhood change; the third group of tracts, which experienced a sharp drop in population density in 1970–1980 (from about 30000 to about 20000 per km^2) is located in a high manufacturing region

where n is the total number of TRI sites. We choose the test statistic to be the maximum variation distance, or L_1-distance, given by

$$\tau = \sup_{A \in \mathcal{A}} |P_b(A) - P_n(A)| = 2 \sum_{j=1}^{1436} \{P_b(T_j) > P_n(T_j)\}\{P_b(T_j) - P_n(T_j)\}.$$

The significance level of τ is estimated by simulating values $\tau_1^*, \ldots, \tau_K^*$ corresponding to

$$\tau_K^* = \sup_{A \in \mathcal{A}} |P_b(A) - P_n^{*k}(A)|,$$

where the simulated empirical distributions $P_n^{*k}(A)$ are obtained by sampling from $P_b(A)$ with replacement n times. For large K (say 10000), the result is the histogram in Fig. 4, which supports the null hypothesis at the approximate 5% significance level.

Thus, having shown that the distribution of TRI facilities is subordinate to that of manufacturing districts, we focus on zoning as the general mechanism for siting potentially hazardous facilities. In doing so, we incorporate in the analysis a wider variety of undesirable land uses than just TRI facilities or commercial TSDFs alone. Zoning data have additional advantages such as being much less prone to administrative errors and omissions, not subject to irregular reporting patterns like the TRI and other facilities data and being complete as compared with TRI data in which facilities are only required to report uses or releases in excess of certain thresholds.

3.2. *Identifying the natural units of analysis: neighbourhoods*

Virtually all past environmental justice analyses have used aggregated demographic data collected in units based on arbitrary boundaries such as census blocks or tracts. These analyses then treat the arbitrary units as substantially meaningful and independent, neither of which is true. The natural unit of evaluation in environmental justice is the *community* or *neighbourhood*, which usually comprises many census blocks or tracts and whose boundaries are likely to change over time. Thus the unit of analysis that is most appropriate for assessing environmental justice is a local unit that captures both the geographical extent of a neighbourhood and its temporal changes.

In fact, neighbourhoods are constantly changing and their boundaries are being perpetually redrawn to incorporate the dynamics of the socioeconomic and geographical components that shape them. Simply put, local communities are not static networks of individuals. Thus, both the natural boundaries of neighbourhoods and their temporal nature must be factored into studies of environmental justice, recovering what we call a *dynamic local unit of analysis*. Indeed, we shall show that the geography of local communities is best characterized by the time changes of their socioeconomic and demographic determinants. Furthermore, spatial agglomeration is a well-known fact of urban environments, and complete segregation can sometimes result from only a weak preference by individuals for neighbours of similar ethnic, racial or even socioeconomic background (Shelling, 1971). Social networks and local communities are not determined by geography alone; thus it is important

(a) to identify the relevant social and economic components that shape them and
(b) to find a metric that captures the socioeconomic distance between individuals within a group and individuals outside that group.

To recover the geography of neighbourhoods from that of census tracts, we group tracts according to their demographic and social characteristics. (We use tract level data because for the 1970 census it is the finest level of data that is available.) The distance measure that we use between tracts is an instance of the notion of economic distance. The novelty of our approach is in that we incorporate *time evolution* of different socioeconomic and demographic measures in assessing that distance. We show that it is the change in socioeconomic and demographic characteristics around a neighbourhood that best delineates it. The strength of our approach is demonstrated when, using census tract level as our basic geographic unit, we can recover the meaningful neighbourhood structures in Brooklyn and Queens.

We start with 1970, 1980 and 1990 census data (US Department of Commerce, Bureau of the Census, 1970, 1980a, 1990a), preprocessing demographic and socioeconomic variables as necessary to ensure consistency and comparability between the three decennial censuses. For example, some 1970 tracts were split or merged in 1980, and some 1980 tracts were split or merged in 1990. To account for this, we carefully aggregated any tract that had split at any time so that we had consistent tract boundaries across all three census decades (US Department of Commerce, Bureau of the Census, 1980b; 1990b) though, because of the age and relatively static population density of New York City, there were actually very few changes in tract boundaries. Thus, for each $t = 1, 2, 3,$ corresponding to the 1970, 1980 and 1990 censuses respectively, we extracted the following variables: the population density per square kilometre (D_t), Hispanics as a percentage of the total population (H_t), non-Hispanic blacks as a percentage of the total population (B_t), non-Hispanic whites as a percentage of the total population (W_t), median household income, in 1970 constant value dollars (I_t), the median value of owner-occupied housing units (O_t), in 1970 dollars, and the median value of rents (R_t), in 1970 dollars. Economic variables were calculated from interval range aggregates for each tract. Medians were computed in most cases, adjusting the dollar values for inflation. However, in the 1980 census, only the mean value of owner-occupied housing units was given for each tract.

The greatest difficulty with the census data was that definitions of race and ethnicity had changed between 1970 and 1980, the most significant of which was that Hispanic ethnicity was defined differently in 1970. To recover the Hispanic ethnicity category that was used in both the 1980 and the 1990 censuses from the 1970 data, we created a Hispanic indicator, referring to populations identified in 1970 as being either of Puerto-Rican descent (by birth or parentage) or Spanish speaking (the language spoken at home): the sum of those two categories was the most comparable with 'Hispanic' as defined in 1980 and 1990 for New York City. Once the Hispanic ethnicity category had been consistently defined, counts of non-Hispanic blacks and non-Hispanic whites were calculated. Minorities other than blacks or Hispanics could not be determined consistently from the three censuses; thus they were left out of the calculation.

Although citywide or even national forces, such as changes in income and in population, are common drivers of neighbourhood change, once such overall trends have been taken into account neighbourhood-specific trends clearly emerge. It is those local patterns of change that are important: when grouped together, tracts that share such particular time trends constitute the geographical extent of local communities. For a given variable V_t in $\{D_t, H_t, B_t, W_t, I_t, O_t, R_t\}$ we assume a piecewise linear trend curve, interpolating linearly between the values $V_t^{(i)}, t = 1, 2, 3,$ for each of the 1436 tracts T_i covering Brooklyn and Queens. Having done so, each tract T_i can be identified with a particular pattern describing the changes in $V_t^{(i)}$ that occur over that tract during the 1970s ($\Delta V_2^{(i)} = V_2^{(i)} - V_1^{(i)}$) and during the 1980s ($\Delta V_3^{(i)} = V_3^{(i)} - V_2^{(i)}$).

Thus, the whole trend curve for variable $V_t^{(i)}$ in a tract T_i is entirely determined by the vector of values $G_i = (V_1^{(i)}, \Delta V_2^{(i)}, \Delta V_3^{(i)})$. On the basis of the Euclidean distance $\|G_i - G_j\|$ between the grouping vectors for tracts T_i and T_j, we can then ascertain how similar or dissimilar T_i and T_j are in terms of V_t. After standardizing each component of G_i for each variable V_t, we obtain a *bona fide* distance matrix expressing the closeness of the tracts in terms of that variable. Tracts are then grouped in terms of level and slopes for each variable V_t, and typical trends are determined by using the K-means clustering algorithm (Hartigan, 1975).

We emphasize the use of the derivative information for each curve for clustering, rather than the curves themselves. The goal is to create groups that have similar slopes in the trend curves, rather than adding a new group that only differed from the others owing to the level of the curves. Without standardizing those components, the clustering vectors G_i are equivalent to the vectors of values $(V_1^{(i)}, V_2^{(i)}, V_3^{(i)})$. Although those vectors will still be representative of the piecewise linear curve for each tract, the resulting groups will fail to isolate the changes in demographics and only tend to capture mean demographic levels.

The cluster centres that result from the K-means algorithm are chosen to minimize the within sum of squares in each cluster, given a prespecified number of clusters. Thus two clusters are well separated if their between sum of squares is larger than their respective within sum of squares. An approximate F-test (Hartigan, 1975) helps to justify adding a new cluster to an existing set of K clusters, based on the assumption of normality of the population, as well as on the assumption that the new cluster will be a subset of an existing cluster. The latter assumption need not hold in general; nevertheless large relative values of the mean-square ratio

$$(m-K-1)\left(\frac{\sum_{k=1}^{K} \mathrm{SSW}_k^{(K)}}{\sum_{k=1}^{K+1} \mathrm{SSW}_k^{(K+1)}}-1\right)$$

indicate that it is acceptable to increase the number of clusters from K to $K+1$. Here,

$$\mathrm{SSW}_k^{(K)}=\sum_{j=1}^{m_k}\|G_j-\bar{G}_k^{(K)}\|^2,$$

where $\bar{G}_k^{(K)}$ is the centre of the kth group in the partition of size K, m_k is the size of group k and $m=\Sigma_{k=1}^{K} m_k$. Thus $\Sigma_{k=1}^{K} \mathrm{SSW}_k^{(K)}$ is the partition error for the K-groups clustering, measuring a total misclassification rate for the partition.

As an example, Table 2 presents the analysis of variance in the case of the trend clustering with respect to the percentage of non-Hispanic blacks in each tract. The high reduction in within-cluster variance between $K=4$ and $K=5$ suggests selecting $K=5$ in favour of $K=4$ for this variable.

A cluster centre corresponds to the trend obtained by interpolating linearly between the group means of the 1970, 1980 and 1990 values for V_t. Thus the cluster centres correspond to mean curves for each cluster. In like manner, cluster medians, interquartile ranges and fifth and 95th percentiles correspond to curves that interpolate linearly between the corresponding quantities for each of the 1970, 1980 and 1990 values. We can thus draw a two-dimensional equivalent of a box plot for each cluster, comparing the typical shape of each cluster's curves relative to their variability and relative to the typical shapes of curves in the other clusters.

Fig. 5 shows the five main groups of tracts obtained by clustering with respect to changes in the percentage of non-Hispanic blacks. Fig. 5 is divided into five panels. The decennial census years are on the x-axis. The percentage of non-Hispanic blacks, ranging from 0% to 100%, lies on the y-axis. A jittered rug plot of the distribution of the percentage in 1970 is shown along the y-axis, which we see is bimodal: almost two-thirds of the tracts have virtually 0% non-Hispanic blacks in 1970, whereas the main bulk of the rest are very close to 90%. Each panel contains

Table 2. Analysis-of-variance table for *K*-means clustering in terms of the trend in percentage of non-Hispanic blacks†

K	*Partition error*	*Mean-square ratio (F-value)*	*Degrees of freedom*
1	4245.0		1415‡
2	3265.2	424.3	1414
3	2164.3	718.8	1413
4	1997.5	117.9	1412
5	729.0	2455.2	1411
6	596.0	314.6	1410

†The mean-square ratio measures the reduction of within-cluster variance between two consecutive partitions. $K = 5$ was selected.
‡20 tracts out of the total 1436 have no associated demographic information, as the population count therein is 0.

information about the size of the cluster, as well as the cluster means (the full black lines), interquartile ranges (the shaded region) and fifth and 95th percentiles (the dotted lines).

In Fig. 5, there are 266 tracts with a high percentage of non-Hispanic blacks from 1970 to 1990 that tend to remain relatively stable (group 1). 113 tracts with a minority of non-Hispanic blacks in 1970 saw a dramatic increase in percentage during 1970–1980 (group 2), bringing them to full majority during the 1980s. However, a group of 35 tracts (group 4) has seen a dramatic shift from almost no non-Hispanic black residents to an almost majority of non-Hispanic black residents recently during 1980–1990. Group 3 contains tracts that are still in transition, where the percentage of non-Hispanic blacks has increased steadily from 1970 to 1990 but is still below 50% in 1990. The bulk of the tracts (935), however, remains at a relatively low percentage of non-Hispanic blacks (group 5) from 1970 to 1990.

The remarkable feature of these clusters, as shown in Fig. 6, is that they clearly correspond to actual geographic neighbourhoods, and they show how those neighbourhoods have evolved over the 20 years from 1970 to 1990. Changes in neighbourhoods, such as expansion or contraction, are well captured by this method. For example, Fig. 6 shows that there were two large predominantly non-Hispanic black neighbourhoods in 1970, one in Brooklyn (the Bedford-Stuyvesant–Brownsville–Crown Heights region) and one in Queens (the Jamaica, Hillside and Saint Albans areas). These neighbourhoods evolved over the next two decades by growing primarily in south and south-easterly directions (towards Kensington in Brooklyn and towards Cambria Heights in Queens) and almost always expanding outwards at the boundaries. This is clear as the tracts that became predominantly non-Hispanic black in the 1970s were often at the boundaries of the 1970 neighbourhoods, and the tracts that became predominantly non-Hispanic black in the 1980s were often at the boundaries of the 1980 neighbourhoods.

Looking at the equivalent plots for non-Hispanic whites (which are not shown here), it becomes readily apparent that the invasion–succession model accounts for much of this neighbourhood change. As discussed earlier, this model suggests that the composition of a neighbourhood changes as the density of a minority group increases until it becomes the majority or dominant group. The changes here are clearly characterized by the white flight phenomenon: a neighbourhood's white residents leave in response to a small influx of black residents; as a result the community switches from being virtually all white to all black. Indeed, plots of the typical trends for tracts in terms of the percentage of non-Hispanic whites are a virtual mirror image of those in Fig. 6 for non-Hispanic blacks.

3.3. Putting it all together: comparing neighbourhood trends in relation to the manufacturing areas

To understand the evolution of neighbourhoods relative to the manufacturing areas of the city, we need to gauge the degree of industrialization within each group of tracts resulting from the curve clustering procedure described above. A simple visual method would be to compare the manufacturing index plots in Fig. 3 with the various demographic plots, such as Fig. 6, but such visual comparisons can sometimes be difficult to make and subsequently to summarize.

An alternative is to interpret the manufacturing index (appropriately normalized) as the probabilistic mechanism that generates sites; thus the probability that a tract contains one or more manufacturing districts is the integral of the manufacturing index surface over the area of the tract. We can calculate the probability mass that falls within each census tract (which we approximate by multiplying the area of the tract by the height of the density surface at the tract's centroid) and compare the distribution of these probabilities between clusters.

A cluster of tracts that tend to have higher probabilities can be interpreted as being closer to sites, i.e., the higher the probability that a particular group of tracts contains one or more manufacturing districts, the higher the degree of industrialization (and, by proxy, the extent of potential exposure) within that group of tracts. Side-by-side box plots, such as those in the lower right-hand side of Fig. 6, can then be used to compare the distributions.

For our comparisons, we use the logit transform of the probabilities. This is equivalent to a one-way logistic analysis of variance of the manufacturing probabilities, by cluster membership, where the response is a precalculated vector of probabilities and the explanatory variable is the group membership. The result is a single graphic in which the evolution of neighbourhoods, both spatially and with reference to manufacturing areas, is readily apparent.

3.4. Results: how neighbourhoods have changed around manufacturing zones

Figs 6–9 summarize our findings for Brooklyn and Queens, New York. Fig. 6 for non-Hispanic blacks (and the associated plot for non-Hispanic whites, which is not shown here) clearly shows that black and white residents have historically been highly segregated in these two boroughs. This is clear for two reasons. First, neighbourhoods tend to have either a high or low percentage of non-Hispanic blacks, and the majority black tracts are clustered into two main neighbourhoods. Second, the two main non-Hispanic black neighbourhoods tend to expand on their peripheries, so little or no integration occurs. The one major exception to this periphery expansion pattern is in a relatively more affluent region in Queens, north of the Jamaica–Hillside–Saint Albans region. However, even there we see segregation at the tract level.

Focusing on how these neighbourhood changes relate to the manufacturing regions, we see in the side-by-side box plots in the lower right-hand corner of Fig. 6 that the heart of the two traditional non-Hispanic black communities (group 1) has a slightly higher median manufacturing index than the others. However, in general the box plots show that the distribution of manufacturing indices for the various groups are very similar to each other. Indeed, other than group 4, they are also similar to the box plot for group 5, which is essentially the manufacturing index for the predominantly non-Hispanic white areas. This result is consistent with Fricker and Hengartner (2001) who did not find a statistically significant association between the percentage of non-Hispanic black residents in a tract and the location of sites for the combined metropolitan New York City region.

However, focusing on Queens and Brooklyn separately, we find significantly different neighbourhood evolution patterns. Queens has fewer manufacturing sites and lower manufacturing indices overall, and here we see that the peripheral expansion of the historically black neighbourhood is disencumbered by any nearby manufacturing zones. Thus, the boundary of this community is generally expanding *away* from manufacturing areas. Also, as discussed above, a smaller group of tracts (group 4) has become predominantly non-Hispanic black in a more affluent area (see Fig. 8 for a measure of regional affluence), indicating that some non-Hispanic black residents have bought into more desirable areas of Queens that are well isolated from manufacturing. This is, however, a recent trend, having occurred between 1980 and 1990. Brooklyn, in contrast, has a higher concentration of manufacturing tracts, and the majority of the peripheral expansion of the historically black neighbourhood has been in the south-east direction *towards* a region with higher manufacturing indices. These results are somewhat at odds with Fricker and Hengartner's (2001) borough level models that showed a positive association between the percentage of non-Hispanic black residents and sites in Queens but not in Brooklyn.

We interpret Fig. 6 to show that any inequity for these communities is the result of a neighbourhood evolution process. This neighbourhood evolution seems to result from an invasion–succession process of neighbourhood change. In the case of Queens, the invasion–succession process is resulting in a decreased association over time, whereas in Brooklyn it is resulting in an increased association over time. However, owing to the fixed nature of the zoning during 1970–1990, what these plots show for the two boroughs is that 'natural' neighbourhood evolution in and of itself can result in environmental inequities that are not likely to be the result of environmental discrimination.

Of course, the plots show a different type of discrimination that manifests itself in segregated residential areas. Furthermore, this analysis does not consider neighbourhood evolution before 1970 where such discrimination may have occurred. Certainly the plots show that the oldest predominantly non-Hispanic black neighbourhoods in Brooklyn were closer to manufacturing regions. This question requires further research.

Hispanics, in contrast, tend to be less segregated, in that there is no single, well-defined, homogeneous neighbourhood. However, tracts with a higher percentage of Hispanics tend to occur around the edges of non-Hispanic black communities and tend to concentrate within the manufacturing regions. The Hispanic neighbourhoods in Queens are generally younger communities that have developed mainly during the 1970s and that have continued to grow during the 1980s (Fig. 7).

What jumps out of a visual comparison of Fig. 7 and Fig. 1 is that the Hispanic communities have tended to develop along the traditional manufacturing regions in the city, which means that *a posteriori* they appear to bear the burden of hosting undesirable land uses. However, again since the manufacturing regions have been fixed over the time frame that we studied, the traditional assumption that the positive association implies that sites were imposed on the communities is not correct. Fig. 7 shows that most of the historically Hispanic tracts (those predominantly Hispanic in 1970) existed in manufacturing areas and the newer Hispanic communities have grown around the manufacturing areas (particularly the group 2 tracts). These results are consistent with the inequity finding of Fricker and Hengartner (2001) but the continued expansion of Hispanic communities in manufacturing regions casts doubt on the possibility that the association is the result of environmental discrimination.

However, the side-by-side box plots in Fig. 7 show that those tracts that have most recently seen an increase in Hispanic populations (group 2) have a smaller median logit compared with the tracts that were historically Hispanic (groups 3 and 4). Thus there is some evidence that newer Hispanic populations are less close to manufacturing and it is possible that a push–pull process is at work here, where older Hispanic communities are pulled towards the manufacturing regions whereas newer Hispanic communities tend to be pushed away.

These results naturally lead to the question of whether economics might be playing a role in neighbourhood development patterns. Similarly to the race and ethnicity plots, Fig. 8 shows four clusters of tracts for median household income. The box plots clearly show that lower incomes are associated with higher manufacturing indices. Said another way, those with higher incomes live in areas removed from the manufacturing regions. Furthermore, comparisons between Figs 6, 7 and 8 show that both Hispanic and non-Hispanic black neighbourhoods tend to be in areas with lower median household incomes. The association between race or ethnicity and economic prosperity is well known in the USA. Whereas Fricker and Hengartner (2001) found inequity by race or ethnicity after accounting for economic prosperity in their models, these results suggest that the stratification of income by race or ethnicity may be responsible for much of the observed inequity.

The neighbourhood life-cycle model posits that the age of the housing stock and population density are the two most important community characteristics that determine neighbourhood change. Under this model, a decrease in population density might have been a precursor for important demographic changes. Fig. 9 shows that the population density in the down-town Brooklyn area dropped sharply between 1970 and 1980 there (from about 30000 to 20000 per km^2), while remaining relatively constant elsewhere. Further, the tracts that had the greatest decrease in population density also have the highest manufacturing rates, so it is possible that higher manufacturing acted as a push force there, accelerating neighbourhood change. The median household income has remained very low in that neighbourhood between 1970 and 1990 (Fig. 8), and so has the median value of rent (that figure is not shown), which indicates that the aging process of that neighbourhood had been well on its way before the population density had dropped in 1970–1980.

In summary, for the period 1970–1990, we find that the observed environmental inequities in Brooklyn and Queens were consistent with Daniels and Friedman's (1999) economic stratification and urban ecology explanations, not overt discrimination. Of course, economic stratification might be the result of more subtle discrimination, and we cannot address the question of whether there was discrimination before 1970 owing to the limitations of our data. However, in these data we see some evidence that as incomes rise residents tend to migrate to areas further from manufacturing zones. This suggests that the advocacy of minority income parity might be employed as part of an effective approach to achieve environmental equity.

4. Conclusions

We have demonstrated a method that graphically recovers spatial–temporal neighbourhood constructs from historical demographic data and connects these neighbourhoods to (proxy) summary measures of exposure. For Queens and Brooklyn, New York, we have shown that neighbourhood change is a complex phenomenon that may be the result of many causes. In our data we have seen evidence of invasion–succession, push–pull and life-cycle effects that are consistent with the economic stratification and urban ecology explanations of neighbourhood change.

In the case of New York City, using zoning that has been fixed for the period of available demographic data, this work cannot result in a determination of environmental discrimination. Yet, we have shown that inequity can occur even when overt (aggregate) discrimination could not have occurred because zoning was fixed.

Our methods can be modified to work in cases where either zoning changes over time or if specific site birth and/or death data are available and thus can be used to investigate the question of environmental discrimination. For areas in which zoning has changed over time, manufacturing indices can be constructed by using kernel smoothing for each period for which demographic data are available (e.g. each decennial census). Then these indices can be summarized for neighbourhood clusters in terms of how the indices have changed for each cluster over time. This method might be used to suggest discrimination if, for example, manufacturing indices tend to increase over time in those neighbourhoods that are either historically minority communities or if the indices increase as the percentage of minority residents increases.

Our techniques can also be used when specific site birth and/or death data are available. First, the manufacturing index provides a probabilistic framework in which to judge whether new sites open in likely or unlikely locations. Second, the neighbourhood clusters can be used to define communities—the appropriate units of analysis—from which it can be determined whether the number of (new) sites within minority communities is greater or less than in the rest of the neighbourhoods.

An important issue with environmental justice is the assessment of exposure in a disadvantaged group. Our measure of exposure here has been determined by proximity alone, as specifics of hazardous substance releases, resulting from manufacturing activities, have proved difficult to determine. Future work should relate changes in ambient air (or stream water) quality to the location of residences within each neighbourhood. Although the experience of residents on one side of a stack (downstream) may be completely different from that of residents on the other side (upstream), we think that the neighbourhood as a whole, with its shared ethnic or racial history, as well as its shared socioeconomic status, forms the community of concern.

We conclude by stressing that more research into the structure and evolution of neighbourhoods is warranted. Future work should account for spatial and temporal correlation—after all, as this work has demonstrated, tracts are not independent. We are also interested in proposing more formal definitions for neighbourhoods, as well as exploring a clustering procedure that penalizes for geographic proximity, so that the socioeconomic clusters are more homogeneous. In more advanced work we plan to consider a (Gaussian) random field model for the differences ΔV_t, so that the ΔV_t are specified conditionally on the underlying graph structure, and where the strength of the links between nodes will be allowed to vary over time.

Acknowledgements

This work was partially supported by National Science Foundation grant SBR-9816560, funding from RAND, as well as a doctoral scholarship for the first author from the Quebec 'Fonds pour la formation de chercheurs et l'aide à la recherche'. The authors would like to thank Nicolas Hengartner for his advice on parts of this paper and Jennifer Pace for her outstanding programming assistance. We would also like to thank the referees for their valuable suggestions.

References

Been, V. (1994) Locally undesirable land uses in minority neighborhoods: disproportionate siting or market dynamics? *Yale Law Rev.*, **103**, 1383–1422.

Been, V. and Gupta, F. (1997) Coming to the nuisance or going to the barios?: a longitudinal analysis of environmental justice claims. *Ecol. Law Q.*, **24**, 1–56.

Bullard, R. D. (1983) Solid waste sites and the Black Houston community. *Socl Inq.*, **53**, 273–288.

——(1996) Environmental justice: it's more than waste facility siting. *Socl Sci. Q.*, **77**, 493–499.

Daniels, G. and Friedman, S. (1999) Spatial inequality and the distribution of industrial toxic releases: evidence from the 1990 TRI. *Socl Sci. Q.*, **80**, 244–262.

Dean, C. and Lawless, J. F. (1989) Tests for detecting overdispersion in Poisson regression models. *J. Am. Statist. Ass.*, **84**, 467–472.

Fricker, R. D. and Hengartner, N. W. (2001) Environmental equity and the distribution of toxic release inventory and other environmentally undesirable sites in Metropolitan New York City. *Environ. Ecol. Statist.*, **8**, 33–52.

Hartigan, J. A. (1975) *Clustering Algorithms.* New York: Wiley.

Liu, F. (1997) Dynamics and causation of environmental equity, locally unwanted land uses, and neighbourhood changes. *Environ. Mangmnt*, **21**, 643–656.

Mitchell, J. T., Thomas, D. S. K. and Cutter, S. L. (1999) Dumping in Dixie revisited: the evolution of environmental injustices in South Carolina. *Socl Sci. Q.*, **80**, 229–243.

New York City Department of City Planning (1990) *Zoning Handbook: a Guide to New York City's Zoning Resolution.* New York: New York City Department of City Planning.

———(1999) *Zoning Maps.* New York: New York City Department of City Planning. (Available from `www.ci.nyc.ny.us/html/dcp/html/zonedex.html`.)

Oakes, J. M., Anderton, D. L. and Anderson, A. B. (1996) A longitudinal analysis of environmental equity in communities with hazardous waste facilities. *Socl Sci. Res.*, **25**, 125–148.

Sexton, K., Olden, K. and Johnson, B. L. (1993) "Environmental justice": the central role of research in establishing a credible scientific foundation for informed decision making. *Toxicol. Indust. Hlth*, **9**, 685–727.

Shelling, T. (1971) Dynamic models of segregation. *J. Math. Sociol.*, **1**, 143–186.

Silverman, B. W. (1986) *Density Estimation for Statistics and Data Analysis.* London: Chapman and Hall.

US Department of Commerce, Bureau of the Census (1970) *Census of Population and Housing 1970, Summary Tape File 3: Technical Documentation.* Washington DC: US Department of Commerce, Bureau of the Census.

———(1980a) *Census of Population and Housing 1980, Summary Tape File 3: Technical Documentation.* Washington DC: US Department of Commerce, Bureau of the Census.

———(1980b) *Census of Population and Housing 1980, 1980-1970 Tract Comparability File: Technical Documentation.* Washington DC: US Department of Commerce, Bureau of the Census.

———(1990a) *Census of Population and Housing 1990, Summary Tape File 3: Technical Documentation.* Washington DC: US Department of Commerce, Bureau of the Census.

———(1990b) *Census of Population and Housing 1990, 1990-1980 Tract Comparability File: Technical Documentation.* Washington DC: US Department of Commerce, Bureau of the Census.

US General Accounting Office (1983) *Siting of Hazardous Waste Landfills and Their Correlation with Racial and Economic Status of Surrounding Communities.* Washington DC: US General Accounting Office.