Environment International 119 (2018) 152–164

Contents lists available at ScienceDirect

Environment International

journal homepage: www.elsevier.com/locate/envint

Modeling spatial distribution of population for environmental epidemiological studies: Comparing the exposure estimates using choropleth versus dasymetric mapping

Weeberb J. Requia[a,*], Petros Koutrakis[b], Altaf Arain[c]

[a] Harvard University, School of Public Health, Department of Environmental Health, 401 Park Drive, Landmark Center 4th Floor West, Boston, MA 02115, United States
[b] Harvard University, School of Public Health, Department of Environmental Health, Boston, MA, United States
[c] McMaster University, School of Geography and Earth Sciences, Hamilton, Ontario, Canada

ARTICLE INFO

Handling Editor: Yong Guan Zhu

Keywords:
Population data
Air pollution
Exposure studies
Health effects
Choropleth mapping
Dasymetric mapping
Epidemiologic studies

ABSTRACT

Precise population information is critical for identifying more accurate environmental exposures for air pollution impacts analysis. Basically, there are two methods for estimating spatial distribution of population, choropleth and dasymetric mapping. While the choropleth approach accounts for linear distribution of population over area based on census tract units, the dasymetric model accounts for a more heterogeneous population density by quantifying the association between the area-class map data categories and values of the statistical surface as encoded in the census dataset. Environmental epidemiological studies have indicated the dasymetric mapping as a more accurate approach to estimate and characterize population densities in large urban areas. However, investigations that have attempted to compare the exposure estimates from choropleth versus dasymetric mapping in environmental health analysis are still missing. This paper addresses this gap and compares the impact of using choropleth and dasymetric mapping in different exposure metrics. We compare the impact of using choropleth and dasymetric mapping in three case studies, defined here as case study A (relationship between urban structure types and health), case study B ($PM_{2.5}$ emissions and human exposure), and case study C (distance-decays of mortality risk related to $PM_{2.5}$ emitted by traffic along major highways). These case studies represent previous investigations performed by our research group where spatial distribution of population was an essential input for analysis. Our findings indicate that the method used to estimate spatial distribution of population impacts significantly the exposure estimates. We observed that the choropleth mapping overestimated exposure for the case study A and B, while for the case study C the exposure was underestimated by the choropleth approach. Our findings show that the dasymetric model is a preferred method for creating spatially-explicit information about population distribution for health exposure studies. The results presented here can be useful for the environmental health community to more accurately assess the relationship between environmental factors and health risks.

1. Introduction

Several environmental health scientists have focused on methodological questions regarding the modeling of input data in epidemiologic studies. This has involved, for example, remote sensing for exposure assessment of air pollutants (Bechle et al., 2013; Guo et al., 2016; Liu et al., 2009) and ambient air temperature (Kloog et al., 2015; Shi et al., 2016, 2015); statistical models to assess exposure and response (Bind et al., 2015; Lee et al., 2015), and models based on Geographic Information System (GIS) to estimate spatial distribution of input variables such as environmental pollutants, land use, and population density (Requia et al., 2016; Svendsen et al., 2012). These investigations are contributing to the precision public health analysis (Khoury et al., 2016), which is a new emerging healthcare era focusing on approaches and technologies to more accurately assess the relationship between environmental factors and health risks (Collins and Varmus, 2015).

Among the methodological questions, of particular importance are the models to estimate accurate and detailed distribution of population in large urban areas. Most environmental studies consider population data based on census geographic entities. In situations where environmental health research is centered on geographic units different from the census tract (in most of the cases, units with finest scale), studies have considered a simple linear relationship between number of people and area under consideration. This is a limitation because each census

* Corresponding author.
E-mail address: wjrequia@hsph.harvard.edu (W.J. Requia).

https://doi.org/10.1016/j.envint.2018.06.021
Received 3 April 2018; Received in revised form 31 May 2018; Accepted 17 June 2018

Table 1
Previous epidemiological studies that used dasymetric model.

First author and year	Location	Type of analysis
Hay et al. (2005)	Kenya	Climate suitability for malaria transmission
Hu et al. (2008)	Florida, US	Association of stroke with air pollution, income and greenness
Maantay et al. (2013)	New York, US	Asthma hospitalization risk due to proximity to pollutant sources
Brantley et al. (2012)	US	Local pediatric hospital coverage (describe by geographic proximity the extent to which the US pediatric population (aged 0–17 years) has access to pediatric and other specialized critical care facilities)
Parenteau and Sawada (2012)	Ottawa, Canada	Land use regression model for the estimation of NO2 concentrations
Aubrecht et al. (2013)	Europe	Social vulnerability variation considering health-related climate change parameters particularly affecting elderly
Cleckner and Allen (2014)	Virginia, US	Spatial Modeling of Mosquito Vector Exposure
Barrozo et al. (2016)	São Paulo, Brazil	Relative risk of infant mortality
Shandas et al. (2016)	Oregon, US	Assess exposure to NO_2
Freeman et al. (2017)	Chicago, US	System-based method for estimating cancer rates

Note: studies are in chronological order.

tract aggregates populated and unpopulated spaces. This erroneous notion of homogeneous distribution over an area may generate inaccurate health risks that may have implications for the development of evidence-based policies in order to address socio-environmental determinants of health and for preventing diseases (Khoury et al., 2016).

The health geography community has been strongly recommending the dasymetric method to accurately map population data (Maantay et al., 2013; Mennis and Hultgren, 2006; Poulsen and Kennedy, 2004). In contrast to choropleth maps, which account for linear distribution of population over an area, the dasymetric model accounts for a more heterogeneous population density or distribution by quantifying the association between the area-class data categories [e.g., land use] and values of the statistical analysis as encoded in the primary dataset [e.g., census tract] (Langford and Unwin, 1994; Mennis, 2009). Here, regions containing open water, parks, and other unpopulated spaces are removed from census tracts to better estimate where people live by redistributing the population to the remaining areas (Maantay et al., 2013).

To our knowledge, only a few environmental epidemiological studies have used dasymetric model to estimate population data. We searched in PubMed using the following keywords: “dasymetric methods”, “dasymetric modeling”, “dasymetric mapping” “health analysis”, “public health”, “health effects”, and “human exposure”. We found only 10 studies, which are presented in Table 1. The first study was published in 2005 and most of the studies were conducted in US (5 studies). The type of analysis varies between these studies, which included analysis of health-related climate change, human exposure to air pollutants, hospital coverage, mosquito vector exposure, infant mortality, and cancer studies.

Such studies provide an important discussion of the dasymetric method as a robust approach to estimate spatial distribution of population in environmental health analysis. However, investigations that have attempted to compare the exposure estimates from choropleth versus dasymetric mapping are still missing. Our research addresses this gap, by comparing the impact of using choropleth and dasymetric mapping in different exposure metrics.

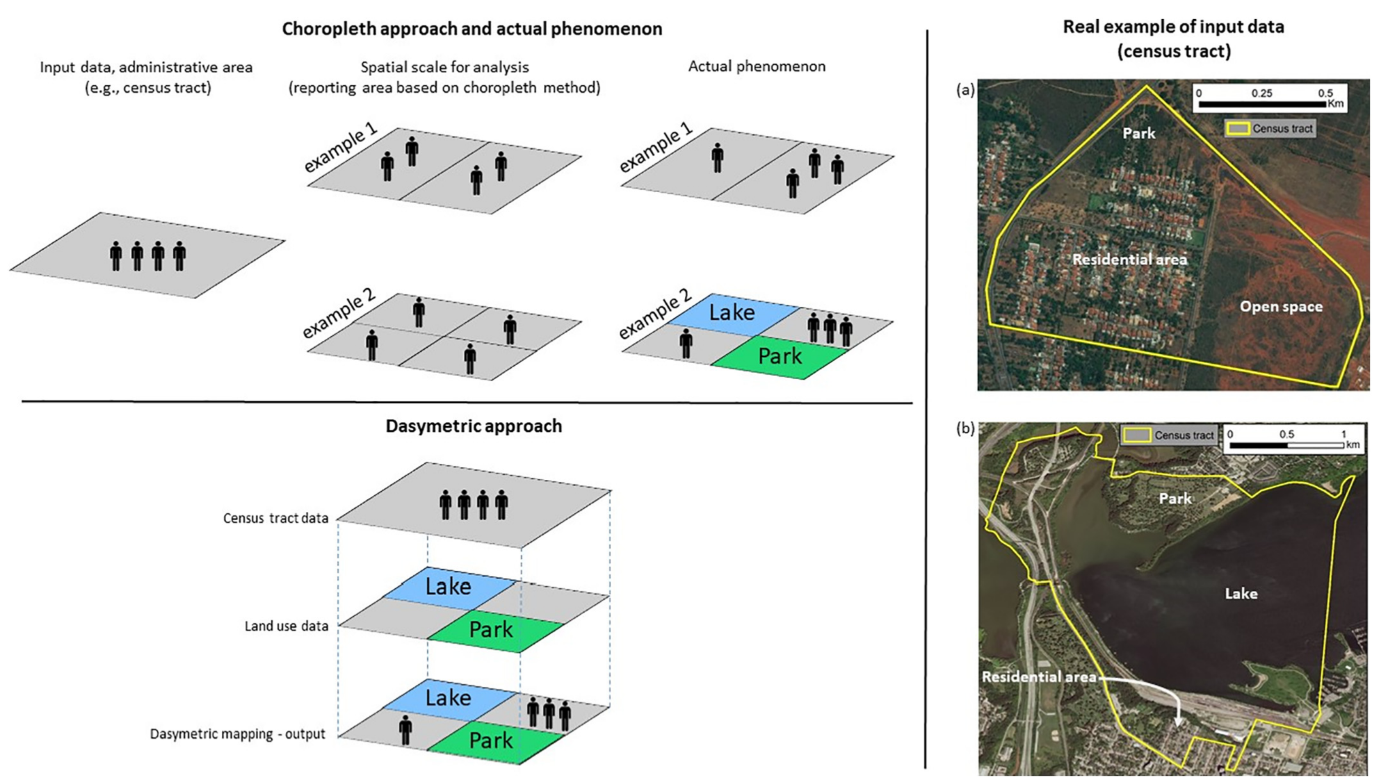

Fig. 1. Overview of the difference between choropleth and dasymetric mapping.
Note: the satellite images on the right side represent a specific area in the Federal District, Brazil (a) and in Hamilton, Canada (b).