Future Directions in Spatial Demography

TSE-CHUAN YANG
GIA Core, Population Research Institute
Pennsylvania State University
Email: tuy111@psu.edu

Trained as a macro-demographer with a strong training background in spatial statistics and mortality research, I am particularly interested in the questions of what new methodological developments in spatial analysis are possible in the near future, and how these new developments evolve from current mainstream spatial demography. I prepared this document to briefly discuss the limitations in current spatial analysis methods and development in spatial data, respectively. Drawing on these discussions, I will answer the two questions above by elaborating on the future challenges facing macro-demographers.

The essence of demography is to study the collective, rather than individual, behaviors [1]. As such, demographic changes and questions have traditionally been studied using data aggregated to various geographic units (e.g., counties and states), which is considered a macro-demographer’s perspective [2]. However, this paradigm has been affected as a better understanding of ecological fallacy has made it clear that results gained from areal data analysis cannot be used to make any inferences about individuals [3]. Indeed, such conclusions may also vary considerably based solely on the levels of aggregation [4]. Micro-demographers, who emphasize individual behaviors, have since come to outweigh macro-demographers in the demographic literature [2]. More recently however, the resurgence of spatial demography has reinvigorated macro-demographers [5, 6], fueled by growth in the diversity and availability of spatial analysis tools and data [7].

Since the late 1990s, macro-demographers have benefited from the development of user-friendly spatial analysis tools [8, 9] to conduct complex regression models where spatial structure is considered. Numerous research projects were focused on enhancing the capability to conduct explanatory statistical analyses, such as GeoDa [10], packages in R [11, 12], and GeoBUGS [13]. These software programs share two methodological features. First, influenced by the first law in geography [14], the explanatory regression models offered in these programs attempt to tackle the issue of spatial homogeneity (dependence) and overlook spatial heterogeneity (non-stationarity). The emergence of geographically weighted regression (GWR) by Fotheringham et al. [15] has given macro-demographers cause to reflect on the conventional one-model-fits-all approach. Second, the ability to incorporate a temporal dimension into regression models is limited. While some developers have endeavored to address this issue, most space-time analyses are exploratory or descriptive, such as STARS by Rey and Janikas [16] and disease mapping techniques [17].

1 It should be noted that CrimeStat by Levine, is able to conduct some spatiotemporal analyses, but its focus is on geographic locations, rather than the features associated with the locations. This is why CrimeStat is not discussed.
The availability of spatial data at different geographic levels also contributes to the expansion of spatial demography and allows researchers to connect social conditions with a range of demographic phenomena [5]. Since 2010, the US Census Bureau has begun to release the American Community Survey (ACS) five-year estimates, a rolling survey that provides the most current information on population and communities. More explicitly, the demographic, economic, housing, and other important social variables will be available from the ACS on an annual basis and the geographic resolutions can be broken down as low as the block-group level. Many Federal agencies have already maintained their demographic data in a similar fashion. For example, the National Center for Health Statistics provides mortality data that can be summarized into county-level every year [19], and the Federal Bureau of Investigation [20] released crime records at various geographic levels (e.g., counties, cities, and states). Before the ACS five-year data are available, the major source of quantitative information on the social conditions of an area is the decennial census. Many population studies are constrained by being forced to use the data around census years (e.g., 1980, 1990, and 2000). It is foreseeable that macro-demographers will attempt to combine the information from ACS five-year estimates and various demographic phenomena (e.g., mortality) to construct a spatiotemporal dataset and hence the demand for a space-time explanatory data analysis tool would arise.

The discussion above reveals two methodological challenges in spatial demography. The more pressing of these is to develop a user-friendly program that implements spatiotemporal dynamic models. Specifically, the space-time data structure will lead macro-demographers to explore how the variable of interest (e.g., fertility) varies across time and space simultaneously, or to investigate whether there is a spatiotemporal lagged effect of a certain independent variable on the demographic outcome. The spatiotemporal process underlying the new types of data will not only help macro-demographers to better understand how a demographic phenomenon is associated with a set of social variables, but also provide micro-demographers insight into what spatiotemporal context matters most. While the spatiotemporal dynamics have increasingly drawn spatial statisticians’ attention [21, 22], macro-demographers have not prepared, or at least not equipped, for a comprehensive investigation on space-time interactions in population studies. The proliferations of data that are spatially and temporally indexed would inevitably call for sophisticated spatiotemporal dynamic modeling.

The other challenge is to further embrace spatial heterogeneity into the spatial analysis landscape that is currently dominated by spatial homogeneity. While both spatial non-stationarity and dependence are unique characteristics of spatial data, relatively little attention has been paid to capturing the influence of non-stationarity in model implementation. These two features are not mutually exclusive [23]. Contrarily, spatial dependence and non-stationarity inform each other methodologically and empirically. Using the kernel density function, local spatial modeling (for heterogeneity) assigns more weights to the observations that are closer to the focal point. This is the process by which spatial dependence helps to explore non-
stationarity. On the other hand, the non-stationary process may be found by visualizing the residuals obtained from the global modeling approach (e.g., OLS). If ignoring the non-stationarity and using global spatial models to account for homogeneity, the analytic results may be biased due to model misspecifications and the spatial process would be misinterpreted [24]. While some efforts have been made to create a general spatial analysis framework [25], the development in this area is still in its infancy. As the goal of macro-demography and spatial analysis is to account for all possible spatial processes in the data, taking both spatial dependence and non-stationarity into account would be necessary.

In sum, from a macro-demographer’s perspective, in the near future, the spatiotemporally dynamic data on population structure and social conditions will challenge spatial demographers to integrate time into research questions and to develop useful tools to address these space-time issues. Moreover, the past methodological development in spatial analysis heavily relies on spatial homogeneity. The recent growth in local modeling, such as GWR and spatially varying models [15, 26], has led spatial demographers to rethink the existing spatial analysis structure and the need for a balanced approach to both spatial dependence and non-stationarity.

References: