




Measuring Neighborhood Impacts on Labor Out-Migration from Fanjingshan National Nature Reserve, China

Yanjing Liu^{1,2,7}  · Jie Dai^{1,2,3,4} · Shuang Yang^{1,3} · Richard Bilborrow⁵ · Minjuan Wang⁶ · Li An^{1,2}

Accepted: 12 May 2023 / Published online: 28 June 2023
© The Author(s), under exclusive licence to Springer Nature Switzerland AG 2023

Abstract

Neighborhood impacts on decisions about out-migration, though less explored and understood than individual- and household-level impacts, can be significant; the integration of these impacts in decision-making analyses may reveal mechanisms undetectable otherwise. However, detecting these impacts can be difficult, especially when prior theorization is lacking. In this paper, we compare three methods of measuring and reducing neighborhood impacts: multilevel modeling, eigenvector spatial filtering (ESF) based on Euclidean distance, and ESF based on topological distance. The second ESF method, in particular, is developed to accommodate the elevation profile of our study site at the Fanjingshan National Nature Reserve of Guizhou Province, China. Our previous work identified a suite of socioeconomic factors at individual and household levels that influence out-migration decisions, to which we apply the aforementioned methods to identify and control for neighborhood impacts. While the non-spatial and multilevel models generated nearly identical results, the results from the ESF models present several considerable differences. The Moran's I statistics for each non-binary variable show that spatial autocorrelation is present in some variables. Among the spatially autocorrelated variables, there are different degrees of change in significance levels when compared to those in the non-spatial model. Although most changes detected are small, we identify an additional significant variable—in our case area farmed—that was not observed before we apply the ESF. Changes in the significance levels of several other independent variables are also more significant after we applied the topological distance definitions. Methodologically, the new results suggest using the topological ESF approach may allow other studies to take into account spatial autocorrelation, especially in more rural areas where elevation differences are significant.

Keywords Neighborhood effects · Rural out-migration · Eigenvector spatial filtering · Multilevel model · Fanjingshan National Nature Reserve · China

1 Introduction

Geography plays an important role in the establishment and growth of communities. People tend to interact more when they live in proximity, and these interactions in turn determine the social norms they are likely to form and follow (Coleman, 1990; Bendor & Switstak, 2001; White & Johnson, 2016). The concept of neighborhood or community, and its effects on attitudes and values, is fundamental to the fields of human ecology, sociology, and geography—indeed all areas of social science. A “neighborhood” is defined as an area where the residents are “interrelated and integrated with reference to its daily requirements” (Hawley, 1950, p. 257). These intra-community relationships can generate correlated data when individuals from multiple communities are studied. Thus, ignoring neighborhood effects, or spatial autocorrelation in general, may lead to biased results from models investigating the effects of individual and household variables on behavior, including migration decisions (Bilsborrow, 2016; Bilsborrow et al., 1984; Chen et al., 2009; Sullivan et al., 2017; Zhang et al., 2021; Zvoleff et al., 2013).

Spatial processes in migration and their impacts on populations have long been examined in spatial demography (Rogers, 1968, 1975; Wilson, 1974; Howell & Frese, 1983). Thus migration can have significant impacts on population compositions in both origins and destinations, especially when persons migrate from poorer to wealthier countries or from rural to urban areas, which is well-known to be usually primarily for better economic opportunities (e.g., Ravenstein, 1885; Sjastad, 1962; Lee, 1966; Bilsborrow et al., 1984; Chandrasekhar & Sharma, 2015; Vega & Brazil, 2015; Mazza et al., 2018; Raymer et al., 2020). Interest in examining neighborhood effects inherent in migration decisions and consequences has mostly evolved more recently (Findley, 2019; Massey et al., 1990). Migrant flows have been found to be highly differentiated by citizenship and nativity (e.g., Lichter & Johnson, 2009; Raymer et al., 2013). Social networks created by geographical proximity and shared experiences have also been observed to be among the major factors affecting migration (and non-migration, from the psychic costs of leaving family, friends, and one’s local community) (Curran & Rivero-Fuentes, 2003; Lee et al., 1994; Lichter & Johnson, 2009; Massey, 1990; White & Johnson, 2016). The likelihood of migration, and the way it is funded and facilitated correlate highly with the spatial patterns of migrant origins (Bell & Ward, 2000; Crowder & South, 2008). For our study area, the Fanjingshan National Nature Reserve (FNNR), out-migration from the rural reserve area to distant cities is part of the major trend of rural–urban migration ongoing in China since the “opening up” in the 1980s. In addition to socioeconomic factors, migration decisions are also impacted by spatial location and interactions with neighbors.

In order to detect and reduce spatial autocorrelation from which neighborhood effects arise, studies in geography and ecology have sometimes utilized various forms of models, including multilevel models and eigenvector spatial filtering (ESF: see Beisner et al., 2006; Rangel et al., 2010). While multilevel models

require substantive information about the spatial units, ESF is less demanding and uses a form of data mining, in which alternative spatial definitions can be tested until the optimal one is found (Getis, 1992; Griffith, 2000).

In this paper, we compare the effectiveness of both multilevel and ESF methods in reducing spatial autocorrelation in the out-migration model. Using two different ways to measure neighborhoods (i.e., defined by Euclidean vs. topological distance) in ESF, we evaluate the potential for using topologically based distance matrices in an area with large elevation variations.

2 Background

2.1 Eigenvector Spatial Filtering (ESF)

Traditional regression analyses that do not account for neighborhood effects may be compromised due to violating the essential assumption that regression residuals are independent. And although individuals or elements in multilevel statistical models can be organized hierarchically to deal with neighborhood effects (Goldstein, 2011), such multilevel models require prior knowledge about the sizes of the classes or clusters (e.g., schools for studies of students, hospitals for patients, political districts for voters). However, in many if not most situations, the most relevant geographic area or size for such classes or clusters is not known, if even identifiable; it varies greatly with the prevalence and quality of transportation linkages and the specific type of behavior or decision under consideration (Bilborrow et al., 1984; Hawley, 1950).

In cases when we have little or no a priori knowledge of the appropriate neighborhood, an effective way to examine neighborhood impacts is to apply spatial filtering methods (Getis, 1992; Griffith, 2000). Other spatial models, including the simultaneous autoregressive (spatial error) model and the autoregressive response (spatial lag) model, are restricted to OLS regression models (Chun et al., 2016). The ESF method decomposes key variables in normal multiple regression models into spatial and non-spatial components to thereby eliminate spatial autocorrelation. The non-spatial components can then be analyzed in *any* standard regression model—this offers considerable potential to detect mechanisms that might be overlooked in models that ignore the spatial component. It defines an $n \times n$ spatial weights matrix C (n being the number of observations or data records), which is comprised of 1's and 0's, each representing a pair of objects being or not being considered spatial neighbors. In our data mining approach, we assign households at varying distances from the household under investigation as neighbors (i.e., these households are assigned a 1 and all the rest a 0), to account for different yet unknown neighborhood sizes. Transforming the C matrix, we get:

$$MCM = (I - 11^T/n)C(I - 11^T/n) \quad (1)$$

where I is the $n \times n$ identify matrix, 1 indicates a $n \times 1$ matrix (column vector with n rows of 1), and T represents the operation of transposing the matrix. It has been

shown that all the eigenvectors of MCM, i.e., $E_1, E_2 \dots E_n$ (representing the eigenvectors associated with eigenvalues ranked in descending order as $\lambda_1 > \lambda_2 \dots > \lambda_n$, where λ_n represents the eigenvalue for the n^{th} eigenvector), are orthogonal. Through some selection procedures (e.g., stepwise regression or choosing the top k eigenvectors where $k < n$), a subset of eigenvectors can be chosen and used as regressors in multivariate regression analysis. Adding these eigenvectors to regression models can remove or at least reduce the contribution of spatial components and generate less biased estimates of regression coefficients (Chun & Griffith, 2011; Griffith, 2000; Tiefelsdorf & Griffith, 2007).

ESF has been widely used in ecology (Beisner et al., 2006; Rangel et al., 2010) and social sciences for topics such as studies of land prices (Seya & Tsutsumi, 2013) and car ownership (Hankach et al., 2022). A few recent studies on migration also utilize this approach, but they focus on network autocorrelation, where the migration of someone from a given origin to a given destination is correlated with that of other persons from the same origin to the same destination (Liu et al., 2017; Gu et al., 2020). Gu et al. (2020), for example, examine the intentions of migrants to transfer their *hukou* from the origin to the destination (that is, their legal household registration determining permanent residency, conferring residency benefits such as free education and healthcare). Their research found significant network autocorrelation in migration intentions, whose impact on the regression model was filtered out by ESF. It has therefore been argued that the inclusion of variables to control for network autocorrelation can significantly improve models of the determinants of migration (Chun & Griffin, 2011).

In the context of our study, we are interested in out-migration decisions from FNNR and therefore have coupled the ESF method with the Cox model (see details in Sect. 3.2) to examine neighborhood effects. Households close to each other are more likely to make similar decisions and are also subjected to the same local government policies and environmental conditions. We hypothesize that neighborhood effects exist in the model at three different distance ranges (short, medium, and long distances; see details in Sect. 3.2) based on our prior knowledge of the site. We also hypothesize that models using topological distances can detect more neighborhood effects compared to the Euclidean-distance-based models because topological distances capture the traveled distance better in the landscape of our study site. The data-mining approach of ESF allows us to determine which neighborhood definitions have the most significant impacts on the variables we are interested in. In addition, in social analyses, ESF has been mostly applied in urban contexts where there is no meaningful variation in elevation. In a prior study, neighborhoods were identified in the same FNNR area with a similar approach but for a different model (Zhang et al., 2021), in which only Euclidean distances were used. Here, we add a topological distance matrix to the Euclidean distance matrix, which may shed light on how ESF could be applied in areas with large elevation variations, as in vast rural parts of the world.

2.2 Rural-to-Urban Migration in China

In the past 40 years, China has undergone profound socioeconomic changes following the "reform and open-door policy" initiated in 1978. Modernization and urbanization have expanded from coastal provinces to inland regions, as well as from big cities to many remote rural areas. Meanwhile, over 200 million adult migrants have left their rural homes and towns for metropolitan areas such as the Yangtze and Pearl River Deltas, seeking better opportunities for personal development and higher incomes to support their families (Zhao, 1999; Liang, 2016). Rural–urban migration is important in most developing countries (it was so earlier in present-day developed countries as well) as it is linked directly to the processes underlying socioeconomic development. Rural families may migrate as units or may send off a household member to earn a higher income and then benefit from receiving the migrant's remittances. In the latter case, and consistent with the theory of the New Economics of Labor Migration (Stark, 1991; Stark & Bloom, 1985; Stark & Taylor, 1991), the migrant is expected to earn more in the destination area and share it with the household of origin by sending (or bringing) back money or goods, increasing and diversifying the income sources of the origin household.

This type of migration appears to have accelerated in China following the large, national payments for ecosystem services (PES) programs, such as the Grain to Green Program (GTGP) and the Natural Forest Conservation Program (NFCP) (Liu et al., 2008; Zhang et al., 2018). The GTGP program has encouraged the transformation of farmland (or pasture) to secondary forests (or grassland), compensating landowners for withdrawing land from cultivation or pasturing. In addition, the Chinese government had already loosened long-standing restrictions on migration from rural to urban areas, such as the *hukou* system, which, as previously mentioned, was used in part to deter migrants from becoming permanent residents in destination cities. This institutional change released farm labor from cultivating farmland or pasturing and allowed them to out-migrate, especially to metropolitan or coastal areas (Zhang et al., 2018; Zhao, 1999).

The goal of most migrations is to seek a better life, but the paths and fruits of the pursuits vary. In the remote villages of the Fanjingshan Nature National Reserve (FNNR) of south-central China, over half of the households surveyed had at least one outmigrant, but their socioeconomic circumstances differed considerably. Some families still relied mainly on agriculture for their livelihoods, while others had already diversified their sources of income, via participating in tourism or other non-agricultural employment. Households in this study area also differed in ethnicity (most being minorities), age composition, education level, and even access to natural resources. Finally, they were participating in the two aforementioned programs (GTGP and NFCP) to varying degrees, receiving different amounts of subsidies based on how much farmland and forest land, if any, they had enrolled in the two programs (Yost et al., 2020).

This research examines the potential impacts of spatial neighborhood size on out-migration from households in the FNNR. Building on the basic Cox statistical model of household migration developed before for the study area (of Yang, 2019), we examine differences in model results before and after incorporating

neighborhood effects. In addition to adding to the literature on the spatial effects in migration, our methodology can also inform other forms of processes subjected to neighborhood effects.

3 Data and Methods

3.1 Study Site and Data Collection

Located in the Wuling mountain range in the Guizhou province, China, the FNNR is a highly biodiverse area and home to many endangered species such as the Guizhou snub-nosed monkey (*Rhinopithecus brelichi*), the Chinese giant salamander (*Andrias davidianus*), and the forest musk deer (*Moschus berezovskii*). The altitude in the reserve ranges from 700 to 2600 m, encompassing a variety of ecosystems (Yang et al., 2002). At the time of the survey in 2014, there were 3256 households residing within or near the boundaries of the FNNR. In addition to resource collection in the forest and agricultural practices, remittances from outmigrants are an important source of income. Following the rural-to-urban migration trend discussed previously, an increasing number of residents have chosen to migrate to cities with more job opportunities.

We conducted a household survey in 58 randomly selected natural villages from all the 123 natural villages located in or near the boundary of FNNR (Fig. 1). We interviewed the household head, if present at the time of the survey, or another knowledgeable adult, usually the spouse, if the head was not present. We collected data on each household's agricultural land area and land use, sources of household income, and the household's enrollment in and value of subsidies received from the GTGP and NFCP programs in the previous 14 years (2001–2014). For migration, we considered all persons aged 15–59 in the household as the persons of interest (number of laborers in Table 1), who could be making migration decisions (whether or not to migrate) in the year of migration. In every household, we also collected data on each adult's age, gender, education, marital status, etc., along with changes over time each year since 2000, notably in education, marital status, residence (whether an outmigrant or return migrant), main work activity (on the farm, off-farm, managing non-farm businesses, none). The main source of household income was also obtained for each year.

In households with an outmigrant still living away at the time of the survey (summer of 2014), the data above were collected *for each year* using an individual event history table (if more than one adult migrant existed in the household, one was selected randomly). For all households with or without an outmigrant, one (non-migrant) member aged 15–59 was also selected to obtain the same event history since 2000. All such individuals selected were considered the "population at risk of migration."

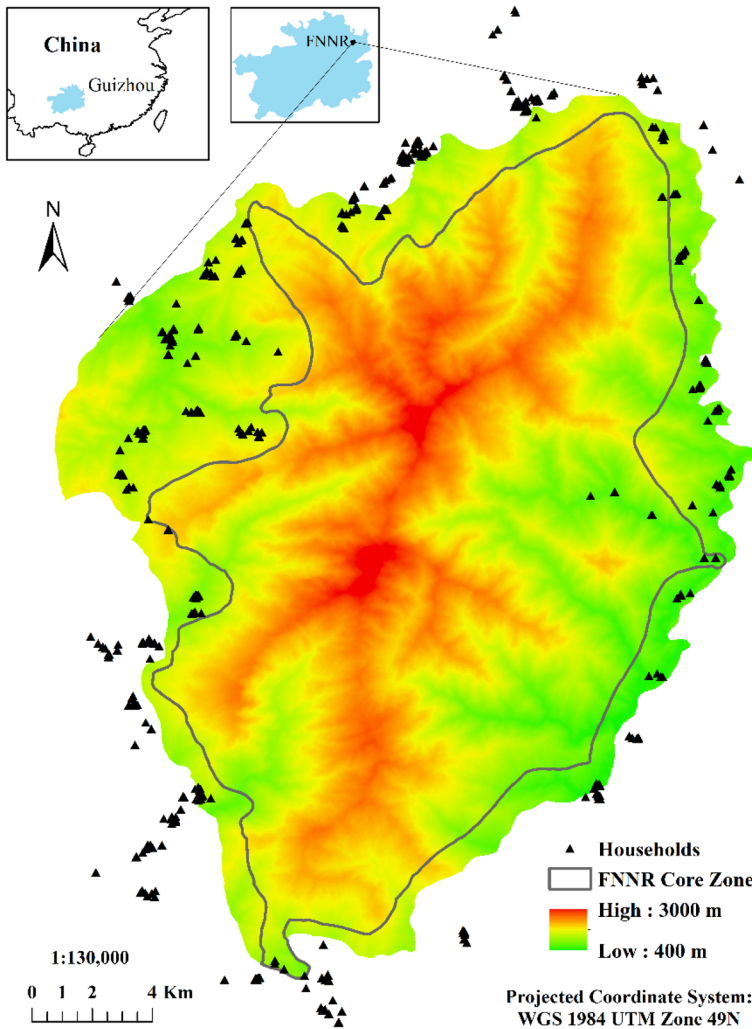


Fig. 1 Fanjingshan National Nature Reserve, showing the locations of interviewed households in this study

3.2 Cox Model

The dependent variable in the model is whether the selected individual was an out-migrant from the household in any year during the reference period (2001–2014)—an outmigrant being defined as a member (or former member) of the household who lived outside the county for more than 6 months and was living away at the time of the survey. Based on the migration literature and visits to the survey households, we identify a suite of independent variables at both individual and household levels to predict out-migration decisions (Table 1; drawing on Yang, 2019). We include

Table 1 Variable descriptions and summary statistics

Level	Variable	Description	Type	Mean	S.D.	Min	Max
Individual	Age	Age at time of interview	Continuous	37.39	15.36	15	86
	Gender	Gender	Dichotomous; 1 = male, 0 = female	0.79	0.41	0	1
	Education	School years completed	Continuous	7.54	3.79	0	18
	Marital status	Currently married?	Dichotomous; 1 = yes, 0 = no	0.88	0.33	0	1
	OffFarmWork	Individual engages in off-farm work	Dichotomous; 1 = yes, 0 = no	0.16	0.36	0	1
Household	HH Size	Total people living in household	Continuous	3.02	1.50	1	9
	NumLabor	Number of household members aged 15–59	Continuous	2.13	0.96	0	6
	Area of Farm	Area of farmland owned (in mu; 1 mu = 0.067 ha)	Continuous	5.67	5.17	0	60
	GTGP	Household participates in GTGP	Dichotomous; 1 = yes, 0 = no	0.56	0.50	0	1
	NFCP	Household participates in NFCP	Dichotomous; 1 = yes, 0 = no	0.71	0.45	0	1
	Income Ag	Agriculture is main source of household income	Dichotomous; 1 = yes, 0 = no	0.42	0.49	0	1
	Mig Network	Household has migration network	Dichotomous; 1 = yes, 0 = no	0.27	0.44	0	1

individual-level variables, such as age, gender, years of school completed, marital status, main work activity, as well as household-level variables, *i.e.*, household size, main source of income, farmland area, participation in GTGP and NFCP, etc. A household is defined as having a migration network if the household respondent reported having a close relative (defined as a parent, child, or sibling) living outside the local county in the year prior to migration, or for a household with no migrant in the reference years, having such a relative living outside the county five years before the survey. Note that most variables were measured with a time-sensitive perspective (*i.e.*, are time-varying), so changes during the modeling period (2001–2014) are incorporated in the survival analysis model, as will be explained later. Thus, some young persons aged into the main "population at risk of migration" (15–59), while others aged out during the reference period.

To include contextual variables in the model, we recorded the village ID of each household, which links the household to its village cluster¹ (23 in total). We also used GPS devices to record each household's exact geographical location, which was then used to calculate neighborhood metrics. Further details on the survey design, implementation, variable selection, and the multilevel Cox model specification are available in Yang (2019) and Yost et al. (2020). In this study, we identify 513 households (16% of total population)² in the study area where all the variables needed for the model are available (Fig. 1).

Survival analysis, a proportional hazards statistical estimation model, is an appropriate technique for examining the occurrence and timing of events (Allison, 2010; An & Brown, 2008; Klein & Moeschberger, 2003). In particular, Yang (2019) and Yost et al. (2020) utilize a multilevel Cox hazard model to predict the determinants of an individual's out-migration. The dependent variable in survival analysis is the "hazard," expressed as the binary result of out-migration in a year, with 1 indicating out-migration of that individual and 0 non-migration (Therneau, 2018):

$$y(t) = y_0(t)e^{X\beta+Zb} \quad (2)$$

where y_0 is the baseline hazard function, X and Z are the design matrices for the fixed and random effects, respectively, and β and b are vectors of regression coefficients. In our analysis, we start with this basic non-spatial model, and then incorporate dummy variables in the basic model, using the Village cluster ID to capture overall contextual effects on migration decisions. To better account for spatial effects, we integrate into this basic model the ESF method (An et al., 2016; Chun &

¹ We re-grouped the 58 sampled natural villages into 23 village clusters based on their geographic locations to ensure that each village cluster had at least 20 interviewed households for the Cox analysis. The village clusters consist of neighboring natural villages, not the official administrative villages.

² Out of the 3256 households living in the reserve, we conducted interviews with 605 scientifically randomly selected households (18.6% of total households). After determining the variables needed for the model, we were left with 513 sample households with complete data, which was still a fully statistically representative sample of the population. To check for possible bias in households excluded, we produced cross-tabulations to compare characteristics of the 513 included households with those of the 92 excluded ones, finding no major differences.

Griffith, 2011; Griffith, 2000). We calculate eigenvectors for each household at several predefined neighborhood sizes.

To do this, we first define neighborhoods based on Euclidean distance; all households within a certain fixed distance are identified as "neighbors" of the household of interest. We use the "spdep" package in R to generate the neighborhood matrices, starting with 0.04 km—which is the minimum distance that allows more than half of the data points to have at least one neighbor—and ending at 10 km, which covers around a third of the spatial extent of the entire area. In addition to ensuring each specification includes enough neighbors, these neighborhood definitions are also selected based on previous studies and theorization. Thus, they have been found to have significant impacts on another model at the same study site (Zhang, 2021). While the short-distance definitions (0.04 km to 0.1 km) capture interactions between close neighbors who might see each other on a daily basis, neighbors based on the moderate-distance definitions (0.1 km to 1 km) would not interact with each other as frequently, but are likely to share key infrastructure (e.g., access to schools, roads, and markets) and similar environmental conditions. Finally, the long-distance definitions (1 km to 10 km) are included because such more remote neighbors may belong to the same village, or other larger administrative unit, and be subjected to similar local government policies.

We also calculate the topological distance (see definition below) between each household with the "topoDistance" package in R and replace the Euclidean distance matrix with the resulting topological distance matrix. For this method, we use the same fixed-distance definitions from 0.04 to 10 km. The topological distances are defined as distances that also take into account terrain differences. Taking an additional digital elevation model (DEM) raster layer, the tool overlays the household locations on the DEM to find the elevation for each point. Then, it calculates the topological distances by finding the shortest topographic path between points, thus better representing real distances traveled. The differences between the Euclidean distances and the topological distances range from 0.03 to 3.78 km, which are significant in our neighborhood definitions.

To find the exact neighborhood size appropriate for our model under the ranges mentioned above, we use a data mining approach: choosing the one distance (out of the multiple ones tried) that matches certain thresholds of some indicators—to uncover or approach the optimal size. We then calculate the top 5 eigenvectors based on each distance specification (i.e., the ones with the highest eigenvalues: An et al., 2016; Chun & Griffith, 2011; Sullivan et al., 2017), and attach them to the original Cox model developed by Yang (2019) to re-estimate the regression. As the dependent variable is binary, the calculation of the residuals d follows from the following equations:

$$\ln \frac{p_i}{1-p_i} = b_0 + \sum_1^j b_j X_{ij}, \text{ and } d_i = \begin{cases} \sqrt{2|\ln(p_i)|}, & y_i = 1 \\ -\sqrt{2|\ln(1-p_i)|}, & y_i = 0 \end{cases} \quad (3)$$

where p is the probability of migration calculated from the statistical model, y the dependent variable, X the set of independent variables, and b the coefficients of the independent variables.

We then calculate the Moran's I statistic³ and the related z scores of the regression residuals of both the spatial and non-spatial models in R using the `spdep` package. For the non-spatial model, we calculate the Moran's I statistic with each neighborhood definition used in the spatial models in order to compare the results with those from the spatial models. For the spatial models, we calculate the statistics with the same neighborhood definition (e.g., the Moran's I results of the 5-km spatial model are calculated with a 5-km neighborhood definition). Akaike information criterion (AIC) scores⁴ are also calculated for each model and used to select models with better fit. Finally, we calculate the Moran's I statistics for all non-binary variables to identify specific variables that could be more spatially autocorrelated than others and thus causing more bias.

4 Results

4.1 The Basic Cox Model Results

The difference between the random-intercept and fixed-intercept multilevel models is found to be insignificant in ANOVA analysis ($p > 0.05$), so we report the results only for the fixed-intercept models.

Individual's age and the household having agriculture as its main source of income are consistently significant and negatively linked to out-migration, while gender (male), being married, number of working-age adults in the household, and the household having migration networks are also significant positive predictors (Table 2). Education and the area farmed are only weakly positively linked to out-migration, while the other variables left in the model for theoretical reasons (expected to be important) do not have statistically significant results when included in the full multivariate model. Overall, the results for all the statistically significant variables are consistent with theoretical expectations.

We then compare the results from the two non-spatial multilevel models, first the basic one and then including village dummy variables to capture the overall effects of village factors on individual out-migration. The significance levels of the independent variables are nearly identical in the two models, and their coefficients do not change much (Table 2). Five village clusters out of 23 have statistically significant effects, and three more marginal effects. In this case, however, no significant changes are observed in the individual or household variables—only

³ An overall indicator for spatial autocorrelation based on how similar/dissimilar neighboring features are. It ranges from -1 to 1 , -1 being perfectly dispersed, 1 perfectly clustered, and 0 perfectly random (Ord & Getis, 1995).

⁴ Calculated from the number of variables in the model and the sum of squared errors (SSE). Smaller AIC scores indicate better fit (Burnham & Anderson, 2002).

Table 2 Results of basic non-spatial model and dummy variable model on factors affecting determinants of out-migration

Variable	Basic model		With dummy variables	
	Coef	SE	Coef	SE
Age	-0.027***	0.005	-0.028***	0.005
Gender	0.318*	0.130	0.311*	0.138
Education	0.032~	0.017	0.023	0.017
Marital status	0.605***	0.178	0.578**	0.182
Off-Farm Work	-0.182	0.141	-0.091	0.153
HH Size	-0.007	0.052	0.030	0.055
Num Labor	0.174*	0.080	0.166*	0.084
Area Farm	0.015~	0.009	0.019~	0.010
GTGP	0.165	0.106	0.170	0.119
NFCP	0.042	0.118	-0.049	0.139
Income Ag	-0.470***	0.117	-0.443***	0.127
MigNetwork	0.394***	0.117	0.445**	0.121
Dummy 1			0.647	0.102
Dummy 2			0.561~	0.086
Dummy 3			0.650~	0.077
Dummy 4			0.134	0.698
Dummy 5			-0.059	0.891
Dummy 6			0.275	0.434
Dummy 7			0.083	0.819
Dummy 8			0.324	0.381
Dummy 9			0.019	0.958
Dummy 10			0.298	0.530
Dummy 11			0.316	0.383
Dummy 12			0.028	0.943
Dummy 13			0.466	0.202
Dummy 14			1.081**	0.005
Dummy 15			1.053**	0.003
Dummy 16			0.454	0.245
Dummy 17			0.707~	0.073
Dummy 18			0.745*	0.027
Dummy 19			0.207	0.576
Dummy 20			0.089	0.831
Dummy 21			0.912*	0.014
Dummy 22			0.691*	0.037

N = 513 HHs. *** $p < 0.001$, ** $0.001 < p < 0.01$, * $0.01 < p < 0.05$, ~ $0.05 < p < 0.1$

a very slight weakening of the effects of education and marriage at the individual level, of the household having agriculture as its main income source at the household level, and tiny increases in the importance of household farm area and migration networks.

Table 3 Moran'I test and AIC results for the ESF models where decrease in residual spatial autocorrelation is significant

Model	Neighborhood definition	Moran I statistic	<i>p</i> value	Z score	AIC
Model 1	Euclidean distance: 3 km	-0.016	0.944	-1.589	4278.312
Model 2	Euclidean distance: 4 km	-0.014	0.94	-1.557	4277.54
Model 3	Euclidean distance: 7 km	-0.01	0.947	-1.616	4276.751
Model 4	Euclidean distance: 8 km	-0.007	0.874	-1.144	4276.881
Model 5	Topological distance: 3 km	-0.001	0.081	1.3962	4278.111
Model 6	Topological distance: 4 km	-0.001	0.092	1.3306	4279.150
Model 7	Topological distance: 5 km	0.000	0.075	1.4393	4276.576
Model 8	Topological distance: 7 km	0.000	0.044	1.7053	4277.451
Model 9	Topological distance: 8 km	0.001	0.037	1.7884	4276.427
Model 10	Topological distance: 9 km	0.000	0.097	1.3015	4276.155

When ESF is incorporated into the model, spatial autocorrelation in the regression residuals (see 3.2 for details) is significantly reduced (Table 3). While the significance levels of most variables remain the same, several do change (Table 4).

Thus the household labor availability variable, significant (at the $p=0.05$ level) in the non-spatial model, becomes insignificant in some of the ESF models (Models 1, 2, 4, 5, and 6). The p -values of the area farmed are also greatly reduced in ESF models, switching from being insignificant (at the $p=0.05$ level) to significant in several models (Models 1, 5, and 6) after filtering out neighborhood effects. Third, the education variable has a decrease in significance level, its p -value changing from 0.055 in the non-spatial model to greater than 0.1 in most ESF models (Models 1, 2, 3, 5, 7, 8, 9, and 10). All of the coefficient values are similar to those in the non-spatial model.

4.2 Spatial Autocorrelation

For the Euclidean distance definitions, the Moran's I results from the non-spatial model show that in 5 out of 15 neighborhood definitions examined (3-km, 4-km, 6-km, 7-km, and 8-km), model residuals are significantly spatially autocorrelated ($|z| > 1.96$; Table 5). We observe a decrease in the absolute values of the z score in 11 out of 15 of the definitions, the exceptions being under the 0.5-km, 1-km, 2-km, and 5-km neighborhoods, suggesting an overall reduction in the spatial autocorrelation of residuals after applying the ESFs. The AIC score also decreases as we increase the measure of distance, becoming smaller than that of the non-spatial model (4281) at 2 km and then rising again at 9 km. This means that between the neighborhood definitions of 2 km and 9 km in this study area, ESF is more effective in reducing spatial autocorrelation and providing a better model fit. Among these models where the spatial autocorrelation is reduced, we identify 4 models (3-km, 4-km, 7-km, and 8-km) where the reduction is significant ($|z|$ decreases from more than 1.96 to less than 1.96). The AIC scores of these models are also optimal among all the ESF models.

Table 4 Cox model coefficients for non-spatial and ESF Models with significant improvements in z values, as determined from the Moran'I test and AIC score in Table 3

	Non-spatial model		Model 1		Model 2		Model 3		Model 4					
	Coefficient	p	Euclidian distance: 3 km		Euclidian distance: 4 km		Euclidian distance: 7 km		Euclidian distance: 8 km					
			Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p				
Age	-0.027	<0.001	-0.029	<0.001	-0.029	<0.001	-0.029	<0.001	-0.028	<0.001				
Gender	0.318	0.015	0.303	0.021	0.306	0.019	0.289	0.027	0.313	0.016				
Edu	0.032	0.055	0.027	0.104	0.027	0.105	0.026	0.117	0.028	0.088				
Married	0.605	<0.001	0.615	<0.001	0.618	<0.001	0.636	<0.001	0.610	<0.001				
OffFarm	-0.182	0.199	-0.184	0.202	-0.174	0.229	-0.199	0.170	-0.157	0.281				
HH_size	-0.007	0.890	0.020	0.703	0.019	0.717	0.009	0.862	0.011	0.834				
Labor_num	0.174	0.029	0.145	0.070	0.157	0.050	0.172	0.035	0.155	0.057				
Area_farm	0.015	0.092	0.018	0.045	0.017	0.055	0.017	0.079	0.016	0.085				
GTGP_part	0.165	0.119	0.112	0.319	0.125	0.268	0.147	0.199	0.131	0.257				
NFCP_part	0.042	0.724	0.058	0.637	0.049	0.695	0.077	0.535	0.063	0.614				
Income_agri	-0.470	<0.001	-0.441	<0.001	-0.460	<0.001	-0.480	<0.001	-0.486	<0.001				
Mig_net	0.394	0.001	0.391	0.001	0.396	0.001	0.394	0.001	0.374	0.001				
V1			3.574	0.003	-3.707	0.002	-2.588	0.042	-1.593	0.239				
V2			1.237	0.304	-0.504	0.693	1.463	0.250	0.789	0.516				
V3			1.116	0.370	1.586	0.183	2.183	0.077	-2.560	0.038				
V4			2.071	0.106	1.803	0.141	-1.850	0.122	-1.384	0.239				
V5			-0.567	0.669	-0.435	0.733	1.648	0.146	-0.798	0.524				
Non-spatial Model														
			Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
			Topological distance: 3 km		Topological distance: 4 km		Topological distance: 5 km		Topological distance: 7 km		Topological distance: 8 km		Topological distance: 9 km	
	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p	Coefficient	p
Age	-0.027	<0.001	-0.029	<0.001	-0.029	<0.001	-0.029	<0.001	-0.029	<0.001	-0.029	<0.001	-0.029	<0.001
Gender	0.318	0.015	0.301	0.021	0.305	0.020	0.303	0.021	0.289	0.027	0.296	0.024	0.288	0.027

Table 4 (continued)

	Non-spatial Model		Model 5		Model 6		Model 7		Model 8		Model 9		Model 10	
	Coefficient	<i>p</i>	Coefficient	<i>p</i>	Coefficient	<i>p</i>	Coefficient	<i>p</i>	Coefficient	<i>p</i>	Coefficient	<i>p</i>	Coefficient	<i>p</i>
Edu	0.032	0.055	0.027	0.104	0.028	0.096	0.027	0.106	0.027	0.111	0.025	0.127	0.026	0.116
Married	0.605	<0.001	0.617	<0.001	0.618	<0.001	0.629	<0.001	0.638	<0.001	0.626	<0.001	0.628	<0.001
Offfarm	-0.182	0.199	-0.188	0.191	-0.186	0.198	-0.183	0.205	-0.195	0.177	-0.204	0.160	-0.225	0.122
HH_size	-0.007	0.890	0.018	0.725	0.018	0.731	0.017	0.748	0.011	0.835	0.009	0.861	0.008	0.879
Labor_num	0.174	0.029	0.144	0.073	0.149	0.064	0.165	0.040	0.171	0.036	0.172	0.035	0.164	0.044
Area_farm	0.015	0.092	0.018	0.045	0.018	0.050	0.017	0.057	0.016	0.081	0.016	0.083	0.017	0.071
GTGP_part	0.165	0.119	0.112	0.318	0.119	0.291	0.135	0.233	0.144	0.209	0.146	0.202	0.139	0.227
NFCP_part	0.042	0.724	0.067	0.585	0.062	0.616	0.055	0.661	0.076	0.540	0.078	0.527	0.085	0.489
Income_agri	-0.470	<0.001	-0.446	<0.001	-0.453	<0.001	-0.469	<0.001	-0.480	<0.001	-0.483	0.000	-0.490	<0.001
Mig_net	0.394	0.001	0.391	0.001	0.396	0.001	0.399	0.001	0.396	0.001	0.391	0.001	0.381	0.001
V1			3.507	0.003	3.632	0.002	-3.276	0.008	-2.801	0.025	-2.540	0.047	-2.051	0.119
V2			1.269	0.291	1.045	0.398	-0.579	0.652	0.950	0.464	1.417	0.264	1.597	0.200
V3			-2.224	0.065	0.121	0.921	2.666	0.023	1.881	0.129	2.327	0.061	2.657	0.032
V4			1.109	0.399	1.951	0.117	-0.355	0.767	-2.429	0.036	-1.777	0.141	-1.395	0.247
V5			-0.319	0.810	0.230	0.856	-1.523	0.220	0.823	0.481	1.868	0.098	2.344	0.042

p values that increase across criterion significant levels (from <0.05 to >0.05 or from <0.1 to >0.1 and becoming less significant) are in red, while those that decrease across criterion significance levels (from >0.05 to <0.05 or from >0.1 to <0.1 and becoming more significant) are in green

Table 5 Moran'I test and AIC results for ESF models and the non-spatial model based on model residuals

Neighborhood dis- tance (km)	ESF models based on Euclidean distance						Improvement		
	Non-spatial model			Moran I statistic					
	Moran I statistic	p value	Z score	AIC	Moran I statistic	p value		Z score	AIC
0.04	0.007	0.434	0.166	4281.375	-0.002	0.492	0.021	4288.690	Marginal
0.06	0.019	0.319	0.471	4281.375	-0.002	0.489	0.027	4286.067	Marginal
0.08	0.015	0.326	0.450	4281.375	0.001	0.461	0.099	4285.674	Marginal
0.1	0.010	0.364	0.348	4281.375	-0.022	0.717	-0.574	4281.421	Marginal
0.5	0.004	0.374	0.321	4281.375	-0.027	0.915	-1.371	4281.962	No
1	0.003	0.347	0.392	4281.375	-0.029	0.976	-1.980	4281.114	No
2	0.012	0.095	1.313	4281.375	-0.027	0.988	-2.246	4276.093	No
3	0.016	0.023	2.004	4281.375	-0.016	0.944	-1.589	4278.312	Yes
4	0.016	0.008	2.406	4281.375	-0.014	0.940	-1.557	4277.540	Yes
5	0.009	0.049	1.659	4281.375	-0.014	0.965	-1.809	4277.991	No
6	0.010	0.017	2.132	4281.375	-0.013	0.974	-1.942	4276.482	Marginal
7	0.010	0.008	2.407	4281.375	-0.010	0.947	-1.616	4276.751	Yes
8	0.010	0.006	2.510	4281.375	-0.007	0.874	-1.144	4276.881	Yes
9	0.005	0.044	1.702	4281.375	-0.006	0.808	-0.869	4281.863	Marginal
10	0.004	0.082	1.392	4281.375	-0.006	0.831	-0.957	4282.581	Marginal
Neighborhood topo distance (km)	ESF models based on topological distance						Improvement		
	Non-spatial model			Moran I statistic					
	Moran I statistic	p value	Z score	AIC	Moran I statistic	p value		Z score	AIC
0.04	-0.026	0.628	-0.328	4281.375	-0.002	0.150	1.0359	4288.584	Moderate
0.06	0.007	0.423	0.193	4281.375	-0.002	0.271	0.6086	4283.597	Moderate
0.08	0.015	0.329	0.443	4281.375	-0.002	0.452	0.1194	4284.658	Moderate
0.1	0.007	0.399	0.256	4281.375	-0.002	0.198	0.8471	4282.783	Moderate
0.5	0.007	0.324	0.455	4281.375	-0.001	0.053	1.6189	4280.604	Moderate
1	0.002	0.390	0.279	4281.375	-0.001	0.004	2.6352	4278.256	No

Table 5 (continued)

Neighborhood topo distance (km)	Non-spatial model				ESF models based on topological distance				Improvement
	Moran I statistic	p value	Z score	AIC	Moran I statistic	p value	Z score	AIC	
2	0.012	0.112	1.216	4281.375	0.000	0.011	2.2926	4275.773	No
3	0.018	0.016	2.154	4281.375	-0.001	0.081	1.3962	4278.111	Yes
4	0.015	0.017	2.125	4281.375	-0.001	0.092	1.3306	4279.150	Yes
5	0.015	0.006	2.538	4281.375	0.000	0.075	1.4393	4276.576	Yes
6	0.010	0.027	1.929	4281.375	0.000	0.051	1.6318	4277.991	Moderate
7	0.009	0.021	2.042	4281.375	0.000	0.044	1.7053	4277.451	Yes
8	0.009	0.011	2.276	4281.375	0.001	0.037	1.7884	4276.427	Yes
9	0.009	0.010	2.318	4281.375	0.000	0.097	1.3015	4276.155	Yes
10	0.005	0.054	1.605	4281.375	0.000	0.171	0.9492	4281.833	Moderate

Z scores are in green ($|z| < 1.96$) if spatial autocorrelation is not statistically significant and in red ($|z| > 1.96$) if spatial autocorrelation is significant

For the topological distance definitions, the model residuals of the ESF models are significantly spatially autocorrelated ($|z| > 1.96$) based on using 6 neighborhood definitions (3-km, 4-km, 5-km, 7-km, 8-km, and 9-km). Similarly, there is an overall reduction in $|z|$, except for two definitions (1-km and 2-km). The AIC scores follow the same pattern as the Euclidean distance definitions, decreasing till 10-km. We identify 6 models (3-km, 4-km, 5-km, 7-km, 8-km, and 9-km) where there is significant reduction in $|z|$.

Finally, we conduct Moran's I calculations on each non-binary independent variable at each of the ten neighborhood definitions (Moran's I calculation is not appropriate for binary variables). While age and education are not spatially autocorrelated under *any* neighborhood definitions, the other statistically significant variables (number of working-age adults, area farmed, and household size) are highly spatially autocorrelated ($|z| > 1.96$). Again, by comparing the change in the p -values of their coefficients, we can see the effect of incorporating ESF on specific spatially autocorrelated variables. As discussed above, the number of laborers and area farmed are two independent variables that undergo significant changes in p -value. Both variables have larger z -scores from the Moran's I statistics, suggesting that they are spatially autocorrelated. The other spatially autocorrelated variable, household size, however, do not experience a significant change after the application of ESF, but is insignificant in any model anyway. Interestingly, one variable that is not spatially autocorrelated, education, also changes its significance level in the spatial models (Table 6).

5 Discussion

5.1 Effects of ESFs on Cox Model

While the multilevel Cox model with dummy variables does not capture significant neighborhood effects, the Moran's I results show that spatial autocorrelation is still present in the model residuals under several neighborhood definitions (Table 5). The application of ESF significantly reduces this spatial autocorrelation in model residuals, resulting in changes in the significance levels of three variables (education, number of laborers, and area of farmland). The Moran's I tests on individual variables also confirm spatial autocorrelation in some of the variables.

Education is weakly linked to migration in the non-spatial model ($0.05 < p < 0.1$), perhaps due to limited educational opportunities beyond primary in the rural study area—migrants usually leave to seek low-skilled work. Adding eigenvectors into the model consistently increases the p values in all neighborhood definitions. With exception of model 4 and model 6, the p values cross from below 0.1 to over 0.1, indicating that the impact of neighborhood is significant for the variable (Table 4). The Moran's I test on the education variable, however, does not indicate spatial autocorrelation at any neighborhood definition (Table 5). The change in significance level of education is therefore likely due to the change in significance of other variables that are somewhat correlated with level of education. This may also be due to the size of the dataset, which might be too small to capture its neighborhood effect.

Table 6 Moran's I results based on individual variable values at different neighborhood definitions

Model	Neighborhood definition	Moran's I statistics	Moran's I p value	Moran's I Z score	Coeff (non-spatial model)	p value (non-spatial model)	Coeff (ESF at same distance)	p value (ESF at same distance)
<i>Age</i>								
Model 1	Euclidian distance: 3 km	0.001	0.382	0.301	-0.027	<0.001	-0.029	<0.001
Model 2	Euclidian distance: 4 km	-0.006	0.692	-0.501	-0.027	<0.001	-0.029	<0.001
Model 3	Euclidian distance: 7 km	-0.004	0.672	-0.444	-0.027	<0.001	-0.029	<0.001
Model 4	Euclidian distance: 8 km	-0.005	0.776	-0.759	-0.027	<0.001	-0.028	<0.001
Model 5	Topological distance: 3 km	0.004689	0.240067	0.706088	-0.027	<0.001	-0.0289	<0.001
Model 6	Topological distance: 4 km	-0.00456	0.62599	-0.32125	-0.027	<0.001	-0.02861	<0.001
Model 7	Topological distance: 5 km	-0.00619	0.731656	-0.61783	-0.027	<0.001	-0.02857	<0.001
Model 8	Topological distance: 7 km	-0.00387	0.63675	-0.34979	-0.027	<0.001	-0.02868	<0.001
Model 9	Topological distance: 8 km	-0.00457	0.70196	-0.53004	-0.027	<0.001	-0.02907	<0.001
Model 10	Topological distance: 9 km	-0.00582	0.801806	-0.84809	-0.027	<0.001	-0.02893	<0.001
<i>Edu</i>								
Model 1	Euclidian distance: 3 km	-0.009	0.773	-0.749	0.032	0.055	0.027	0.104
Model 2	Euclidian distance: 4 km	-0.014	0.951	-1.658	0.032	0.055	0.027	0.105
Model 3	Euclidian distance: 7 km	-0.005	0.739	-0.641	0.032	0.055	0.026	0.117
Model 4	Euclidian distance: 8 km	-0.005	0.722	-0.59	0.032	0.055	0.028	0.088
Model 5	Topological distance: 3 km	-0.00801	0.73995	-0.64319	0.032	0.055	0.027072	0.10424
Model 6	Topological distance: 4 km	-0.01029	0.847454	-1.02558	0.032	0.055	0.027575	0.096093
Model 7	Topological distance: 5 km	-0.0142	0.963	-1.78662	0.032	0.055	0.02669	0.106295
Model 8	Topological distance: 7 km	-0.01031	0.9362	-1.52363	0.032	0.055	0.026532	0.110796
Model 9	Topological distance: 8 km	-0.00509	0.736796	-0.6335	0.032	0.055	0.025363	0.127073
Model 10	Topological distance: 9 km	-0.00583	0.801881	-0.84836	0.032	0.055	0.026059	0.116081

Table 6 (continued)

Model	Neighborhood definition	Moran's I statistics	Moran's I <i>p</i> value	Moran's I Z score	Coeff (non-spatial model)	<i>p</i> value (non-spatial model)	Coeff (ESF at same distance)	<i>p</i> value (ESF at same distance)
<i>Labor_num</i>								
Model 1	Euclidian distance: 3 km	0.038	<0.001	4.377	0.174	0.029	0.145	0.07
Model 2	Euclidian distance: 4 km	0.036	<0.001	5.014	0.174	0.029	0.157	0.05
Model 3	Euclidian distance: 7 km	0.022	<0.001	4.637	0.174	0.029	0.172	0.035
Model 4	Euclidian distance: 8 km	0.016	<0.001	3.871	0.174	0.029	0.155	0.057
Model 5	Topological distance: 3 km	0.037535	1.34E-05	4.198698	0.174	0.029	0.144429	0.072514
Model 6	Topological distance: 4 km	0.039657	1.51E-07	5.122111	0.174	0.029	0.149051	0.063899
Model 7	Topological distance: 5 km	0.027926	6.42E-06	4.362679	0.174	0.029	0.165479	0.040315
Model 8	Topological distance: 7 km	0.02343	1.83E-06	4.629757	0.174	0.029	0.170619	0.036257
Model 9	Topological distance: 8 km	0.019318	8.49E-06	4.301422	0.174	0.029	0.171744	0.035025
Model 10	Topological distance: 9 km	0.015938	4.41E-05	3.920952	0.174	0.029	0.163806	0.044068
<i>HH_Size</i>								
Model 1	Euclidian distance: 3 km	0.036	<0.001	4.214	-0.007	0.89	0.02	0.703
Model 2	Euclidian distance: 4 km	0.034	<0.001	4.714	-0.007	0.89	0.019	0.717
Model 3	Euclidian distance: 7 km	0.025	<0.001	5.191	-0.007	0.89	0.009	0.862
Model 4	Euclidian distance: 8 km	0.02	<0.001	4.755	-0.007	0.89	0.011	0.834
Model 5	Topological distance: 3 km	0.034462	5.42E-05	3.870896	-0.007	0.89	0.018434	0.724685
Model 6	Topological distance: 4 km	0.040057	1.17E-07	5.169929	-0.007	0.89	0.018042	0.731027
Model 7	Topological distance: 5 km	0.030714	9.28E-07	4.768514	-0.007	0.89	0.016871	0.747565
Model 8	Topological distance: 7 km	0.026119	1.54E-07	5.118891	-0.007	0.89	0.011032	0.834905
Model 9	Topological distance: 8 km	0.022697	3.13E-07	4.983298	-0.007	0.89	0.00929	0.861085
Model 10	Topological distance: 9 km	0.020007	7.49E-07	4.811621	-0.007	0.89	0.00808	0.879028

Table 6 (continued)

Model	Neighborhood definition	Moran's I statistics	Moran's I <i>p</i> value	Moran's I Z score	Coeff (non-spatial model)	<i>p</i> value (non-spatial model)	Coeff (ESF at same distance)	<i>p</i> value (ESF at same distance)
<i>Area_farm</i>								
Model 1	Euclidian distance: 3 km	0.053	<0.001	6.225	0.015	0.092	0.018	0.045
Model 2	Euclidian distance: 4 km	0.039	<0.001	5.56	0.015	0.092	0.017	0.055
Model 3	Euclidian distance: 7 km	0.034	<0.001	7.235	0.015	0.092	0.017	0.079
Model 4	Euclidian distance: 8 km	0.036	<0.001	8.262	0.015	0.092	0.016	0.085
Model 5	Topological distance: 3 km	0.055962	1.38E-10	6.311344	0.015	0.092	0.017957	0.044615
Model 6	Topological distance: 4 km	0.044804	1.83E-09	5.898638	0.015	0.092	0.017754	0.049703
Model 7	Topological distance: 5 km	0.036139	6.09E-09	5.697307	0.015	0.092	0.017425	0.056695
Model 8	Topological distance: 7 km	0.034417	5.41E-12	6.795176	0.015	0.092	0.016344	0.080766
Model 9	Topological distance: 8 km	0.035823	2.58E-15	7.823002	0.015	0.092	0.016348	0.082773
Model 10	Topological distance: 9 km	0.037757	2.52E-19	8.911512	0.015	0.092	0.016996	0.070902

z scores in green ($|z| < 1.96$) are not statistically spatially autocorrelated while those in red ($|z| > 1.96$) are spatially autocorrelated. Sizes of coefficients and *p*-values are also indicated (as in Table 4); *p* values that change from less than 0.05 to over 0.05 or from over 0.05 to less than 0.05 are in bold

At the household level, our results demonstrate that as the farm area variable became more significant, the number of laborers in the households became less significant, after the application of ESF. Their significance levels change likely due to the high spatial autocorrelation in the data (Table 5); once the appropriate ESF is used to filter out the negative impacts of spatial autocorrelation, its "hidden" impacts on migration are recovered: thus farm area has a more significant, positive impact on migration decisions (coefficient p -values decreased in all ESF models, with the p -value at model 1, 5, and 6 less than 0.05), while the number of laborers only appears to have a mild significance.

The positive relationship between farmland area and out-migration may seem contradictory, since in other studies on developing countries, access to land, either cultivated or non-cultivated (forest in this case), tends to be a key factor that *reduces* out-migration (e.g., Shaw, 1975; Bilsborrow et al., 1984, Chapters 2, 10; Massey et al., 1993) as having more land provides more opportunities to engage in agricultural production. But there are cases in which the ownership of more household assets, including agricultural land, *facilitates* out-migration (Bilsborrow et al., 1987; Davis et al., 2016). First, households with more farmland are likely to have more household income from land. In the case of China, they have an additional advantage: they are more likely to have land to enroll in the GTGP, which provides a modest cash compensation each year and could help fund out-migration to an urban destination (Davis et al., 2016; Yost et al., 2020). Apart from the household's capacity for funding out-migration, a person's willingness to migrate also relates to the household's food security in his/her absence. Households with more farmland and thus more grain production are more likely to have enough land and crops to meet their basic subsistence needs even after GTGP enrollment of land and out-migration. This is because farmers may intensify agricultural production on the remaining land, made possible by remittances from the outmigrants which can be used to purchase better farming equipment (Xu et al., 2006); in the FNNR, the labor-to-farmland ratio is indeed high. Finally, greater availability of farmland is often associated with remote, poorer places, where farmers are more likely to leave for higher-paying jobs (Zhang et al., 2018).

Interestingly, the number of laborers in the household becomes insignificant when ESF is applied, which might also arise from spatial autocorrelation in the data (Table 5). Given the one-child policy in China over the past four decades (terminated in 2016), the numbers of children in rural households are very similar, rarely being different from one or two, making the variable relatively stable (mean=2.13; max=6; min=0; standard deviation=0.96; Table 1). Such low variation in the number of laborers makes it difficult to detect any effects of the variable: while there is still the expected positive relationship between number of laborers in the household prior to migration and out-migration (even after the ESFs are applied), p values are consistently lower than 0.1).

5.2 Euclidean and Topological Distance Definitions

Comparing the Moran's I results of model residuals under the same distance from the Euclidean and topological neighborhood definitions, we can see that the topological

distance models have lower $|z|$ values from 3 to 10 km (Table 5; e.g., at 3 km, model 1 (Euclidean) has a $|z|$ of 1.589 while model 5 (topological) has a $|z|$ of 1.396). The topological models also better reduce the level of spatial autocorrelation than the Euclidean models—there are two additional distances (5-km and 9-km) where $|z|$ values become significantly lower in the ESF models using topological distances. In Table 4, we can also see that the changes in p values in variables are slightly larger in the topological models (e.g., for area of farmland at 4-km definition, the variable had a p -value of 0.50 in model 2 (Euclidean) and a p -value of 0.064 in model 6 (topological). This shows that the topological models are better at detecting and eliminating neighborhood effects. Although the topological models do not identify additional variables that underwent significant changes after the ESF process, they do generally highlight the strength of neighborhood effects in the dataset. Therefore, the topological models appear better suited for our study area, where elevation differences are quite significant (elevation ranges from 484 to 1632 m for the data points). This is especially true for larger distances because the difference between the Euclidean distance and the topological distance is small when the households are very close, i.e., only 0.2 or 0.5 km away from each other. But when the distance becomes larger, it can involve a significant hike up or down a mountain.

For both the Euclidean and topological models, greater neighborhood effects are detected using the longer-distance definitions, being especially significant for 3 km and 4 km. These distances are roughly consistent with the area of an administrative village, which may implement local policies (e.g., land or forest management) or carry out agricultural education programs, making households within the area more similar to one another. But there are also likely to be neighborhood effects that exist even outside the shared administrative boundaries; e.g., people near such boundaries may interact with one another or share natural resource conditions, but will still be classified into separate administrative units. Therefore, dummy variables for the administrative village will not fully capture all neighborhood effects. Even though there is a significant reduction in spatial autocorrelation in model residuals at longer distances (7 km, 8 km, and 9 km) as well, definitions at those distances do not show the same changes in significance levels of independent variables as they do at 3 km and 4 km. As the distance became bigger, a large portion of the dataset becomes categorized as neighbors, so the ESF method becomes less effective in reducing spatial autocorrelation. Even though our results demonstrate that neighborhood effects are more statistically significant based on using longer-distance definitions, this does not explain why some distances capture neighborhood effects better than others—for example, in the Euclidean definitions, the model residuals are spatially autocorrelated at 3 km, 4 km, 6 km, 7 km, and 8 km, but not at 5 km (Table 5). Such inconsistencies could result from our relatively small sample size or a few outliers (e.g., extreme values) in the data, but our results still demonstrate how a data mining component can be useful to detect neighborhood effects and refine prior theorization.

6 Conclusions

The neighborhood context is recognized as a potentially important predictor of individual-level behaviors as well as socio-demographic outcomes for individuals and for origin households (Bilsborrow et al., 1984; Hawley, 1950; Lee & Cubbin, 2002; Pickett & Pearl, 2001; Sampson, 2003). In addition to sharing similar social-economic characteristics, individual agents (e.g., persons or households) in the same neighborhood or community interact closely with each other, and thus tend to have many similarities in behaviors, values, and decision-making processes.

Nonetheless, identifying what constitutes a "neighborhood" has proved to be a challenge, both in theory and in practice. Due to difficulties in defining "neighborhood" and the cost of collecting detailed substantive community-level data (e.g., from interviewing community-level leaders directly to seek specific data on population size and characteristics, presence of infrastructure of various types, transportation facilities, wage and price levels, etc.: see Bilsborrow et al., 1984 regarding migration), researchers often resort to using existing administrative or political boundaries, creating artificial sampling clusters (grouping data from natural villages into artificial clusters), or otherwise creating "neighborhoods" arbitrarily, rather than on substantive or environmental grounds. For example, the boundary between administrative districts may pass through the middle of a valley where inhabitants of villages naturally interact regularly. Our study has experimented in one rural area of China in defining "neighborhoods" according to Euclidean distance or topological distance using various neighborhood sizes, and finds—in rural areas of China at least—it is only at certain sizes that we can observe significant neighborhood effects, i.e., differences in results compared to those of non-spatial models. Although the neighborhood sizes used (3-km and 4-km, in our case) might correspond to administrative units such as village, simply adding dummy variables for each village in a multilevel model cannot necessarily capture these effects. Thus, relationships may remain hidden in the non-spatial and multilevel models, or observed where they do not exist. Further study is needed to better understand the relationships between a "true" neighborhood size and the ones based on our data mining approach.

It is important to recognize that the neighborhood effects observed here are not very significant, as indicated by the consistency of most coefficients' significance levels (p -value; Tables 4) in the spatial and non-spatial models. The sizes of the coefficients of all variables also remain similar after ESF is applied. The Moran's I and $|z|$ scores of the spatial models' (with ESF) regression residuals, however, do undergo significant changes under some of the neighborhood definitions after eigenvector filters are applied. By calculating Moran's I statistics for the independent variables, we also find that several appear to have spatial autocorrelation. Thus, we select ESF models to test for spatial autocorrelation based on model residuals from the non-spatial models. The results identify three out of five non-binary variables to be spatially autocorrelated, two of which (number of laborers and area farmed) undergo changes in significance level after the spatial

filtering was implemented, although no such change is observed in other spatially autocorrelated variables. Household size, for example, is spatially autocorrelated ($|z|$ scores > 1.96) under all four selected neighborhood definitions, but has little change in significance level from incorporating ESF. In addition, a variable (education) with little spatial autocorrelation changes its significance level after ESF is applied (i.e., loses its statistical power). This may mean that the population diversity of our data set is *not sufficient* to capture neighborhood effects well due to the small area and fairly homogeneous population.

However, our results *do* suggest that employing appropriate modeling methods and including tests on the effects of various measures of neighborhood size can help to identify and capture the impacts of "neighborhood" and thereby generate more accurate estimates of the coefficients of variables that are often subject to neighborhood effects. At the same time, the methods demonstrated here will often, at minimum, increase the accuracy of measurement of effects of *some* variables (e.g., at the individual or household levels) even if they are not directly subject to statistically significant neighborhood effects. As shown in many previous studies, spatial correlation among variables is often present and should not be ignored. After including the spatial filters (i.e., eigenvectors), the independent variable, household farmland area, changed from being statistically insignificant to significant at several definitions of neighborhood distances, while the number of household members of working age changed from significant to insignificant. It is important to note that when a data-mining approach is used, neighborhood effects (e.g., distances at which it is the most prominent, and its levels of significance) can alter depending on the dependent variable and model under question. For example, Zhang et al. (2021) detected neighborhood effects in 0.002 km, 0.1 km, and 0.5 km in their model on household participation in PES programs using the same data. Both analyses illustrate that ignoring neighborhood effects can lead to misleading and even incorrect conclusions (An et al., 2016; Sullivan et al., 2017). The difference in neighborhood effects of the two models also suggests that different distances should be tested in different models even with the same dataset.

Finally, topological-distance-based neighborhood definitions might generate even more accurate results for areas with large elevation variations such as our study site, and many other rural sites around the world. Definitions of neighborhood generated with eigenvector filtering may be more relevant than, or at least can be complementary to, those captured by traditional approaches, e.g., using dummy variables for each community or cluster controls. Therefore, we recommend that the ESF approach be tested in other, especially larger or more diverse geographic conditions, to better determine an appropriate size for "neighborhood" in different contexts, according to the particular topic or decision process being studied, which can then be incorporated in the model to correct for spatial autocorrelation and thereby lead to better results in investigations of factors influencing people's decision-making and behavior. Replicating this method in many other different contexts and for different variables can further enhance understanding of what an appropriate "neighborhood" is for all manner of variables of human behavior, thereby facilitating the use of the ESF or other methods described here to effectively control for the neighborhood effects when trying to understand human behavior.

Acknowledgements This research was funded mainly by the National Science Foundation under its Dynamics of Coupled Natural and Human Systems program [Grant DEB-1212183 and BCS-1826839]. We also received financial and research support from San Diego State University, and are grateful to the Carolina Population Center and its NIH/NICHD population center grant [P2C HD050924] for general support for Bilborrow's collaboration. Finally, we thank Fanjingshan National Nature Reserve and the Research Center for Eco-Environmental Sciences at the Chinese Academy of Sciences for logistic support.

Author Contributions All authors contributed to the study's conception and design. Material preparation and data collection were performed by SY, LA, RB, and ML, and undergraduates students from Tongren University. Analysis was performed by JD and YL. The first drafts were written by JD and YL. All authors commented on previous versions of the manuscript and approved the final manuscript.

Funding This research was funded by the National Science Foundation under the Dynamics of Coupled Natural and Human Systems program [Grant DEB-1212183 and BCS-1826839]. This research also received financial and research support from the Center for Complex Human-Environment Systems, San Diego State University. Finally, we are grateful to the Carolina Population Center and its NIH/NICHD population center grant [P2C HD050924] for general support to Bilborrow for collaboration in this research.

Data Availability Uploaded as supplementary material.

Code Availability Uploaded as supplementary material.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

Ethics Approval Ethical approval for this study was obtained from the Institutional Review Board at San Diego State University (Protocol Number: 1732093).

Consent to Participate Informed consent was obtained from all individual and household participants in the study.

Consent for Publication The participants consented to the data being used for research, leading to this submission to the journal.

References

- Allison, P. D. (2010). *Survival analysis using SAS®: A practical guide*. SAS Institute.
- An, L. (2016). CNH eigenvector instructions. Available at http://complexities.org/Photo&PDF/CNH_Eigvec_Instructions.pdf
- An, L., & Brown, D. G. (2008). Survival analysis in land change science: Integrating with GIScience to address temporal complexities. *Annals of Association of American Geographers*, 98, 323–344. <https://doi.org/10.1080/00045600701879045>
- An, L., Tsou, M.-H., Spitzberg, B. H., Gupta, D. K., & Gawron, J. M. (2016). Latent trajectory models for space-time analysis: An application in deciphering spatial panel data. *Geographical Analysis*, 48, 314–336. <https://doi.org/10.1111/gean.12097>
- Anselin, L., & McCann, M. (2009). OpenGeoDa, open source software for the exploration and visualization of geospatial data. In *Proceedings of the 17th ACM SIGSPATIAL international conference on advances in geographic information systems* (pp. 550–551). <https://doi.org/10.1145/1653771.1653871>

- Barbieri, A., Carr, D., & Bilborrow, R. (2009). Migration within the frontier: The second generation colonization in the Ecuadorian Amazon. *Population Research and Policy Review*, 28, 291–320. <https://doi.org/10.1007/s11113-008-9100-y>
- Beisner, B. E., Peres-Neto, P. R., Lindström, E. S., Barnett, A., & Longhi, M. L. (2006). The role of environmental and spatial processes in structuring lake communities from bacteria to fish. *Ecology*, 87(12), 2985–2991. [https://doi.org/10.1890/0012-9658\(2006\)87\[2985:TROEAS\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2006)87[2985:TROEAS]2.0.CO;2)
- Bell, M., & Ward, G. (2000). Comparing temporary mobility with permanent migration. *Tourism Geographies*, 2, 87–107. <https://doi.org/10.1080/146166800363466>
- Bendor, J., & Swistak, P. (2001). The evolution of norms. *American Journal of Sociology*, 106, 1493–1545. <https://doi.org/10.1086/321298>
- Bilborrow, R. E. (2016). Concepts, definitions and data collection approaches. In M. J. White (Ed.), *International handbook of migration and population distribution* (pp. 109–156). Springer.
- Bilborrow, R. E., McDevitt, T., Kossoudji, S., & Fuller, R. (1987). The impact of origin community characteristics on rural-urban out-migration in a developing country. *Demography*, 24, 191–210. <https://doi.org/10.2307/2061629>
- Bilborrow, R. E., Oberai, A., & Standing, G. (1984). *Migration surveys in low-income countries: Guidelines for survey and questionnaire design*. Croom-Helm, for the International Labor Office.
- Burnham, K. P., & Anderson, D. R. (2002). *Model selection and multimodel inference: A practical information-theoretic approach* (2nd ed.). Springer.
- Chandrasekhar, S., & Sharma, A. (2015). Urbanization and spatial patterns of internal migration in India. *Spatial Demography*, 3(2), 63–89. <https://doi.org/10.1007/s40980-015-0006-0>
- Chen, X., Lupi, F., He, G., & Liu, J. (2009). Linking social norms to efficient conservation investment in payments for ecosystem services. *Proceedings of the National Academy of Sciences*, 106, 11812–11817. <https://doi.org/10.1073/pnas.0809980106>
- Chun, Y., & Griffith, D. A. (2011). Modeling network autocorrelation in space-time migration flow data: An eigenvector spatial filtering approach. *Annals of Association of American Geographers*, 101, 523–536. <https://doi.org/10.1080/00045608.2011.561070>
- Chun, Y., Griffith, D. A., Lee, M., & Sinha, P. (2016). Eigenvector selection with stepwise regression techniques to construct eigenvector spatial filters. *Journal of Geographical Systems*, 18(1), 67–85. <https://doi.org/10.1007/s10109-015-0225-3>
- Coleman, J. S. (1990). *Foundations of social theory*. Harvard University Press.
- Crowder, K., & South, S. J. (2008). Spatial dynamics of white flight: The effects of local and extralocal racial conditions on neighborhood out-migration. *American Sociological Review*, 73, 792–812. <https://doi.org/10.1177/000312240807300505>
- Curran, S. R., & Rivero-Fuentes, E. (2003). Engendering migrant networks: The case of Mexican migration. *Demography*, 40(2), 289–307. <https://doi.org/10.1353/dem.2003.0011>
- Davis, J., Sellers, S., Gray, C., & Bilborrow, R. (2016). Indigenous migration dynamics in the Ecuadorian Amazon: A longitudinal and hierarchical analysis. *The Journal of Development Studies*, 53, 1849–1864. <https://doi.org/10.1080/00220388.2016.1262028>
- Findley, S. E. (2019). Rural development and migration: A study of family choices in the Philippines. *Routledge*. <https://doi.org/10.4324/9780429305023>
- Getis, A., & Ord, J. K. (1992). The analysis of spatial association by use of distance statistics. *Geographical Analysis*, 24, 189–206. <https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>
- Goldstein, H. (2011). *Multilevel statistical models* (4th ed.). Wiley.
- Griffith, D. A. (2000). A linear regression solution to the spatial autocorrelation problem. *Journal of Geographical Systems*, 2, 141–156. <https://doi.org/10.1007/PL00011451>
- Gu, H., Liu, Z., & Shen, T. (2020). Spatial pattern and determinants of migrant workers' interprovincial Hukou transfer intention in China: Evidence from a national migrant population dynamic monitoring survey in 2016. *Population, Space and Place*. <https://doi.org/10.1002/psp.2250>
- Hankach, P., Gastineau, P., & Vandanjon, P.-O. (2022). Multi-scale spatial analysis of household car ownership using distance-based Moran's eigenvector maps: Case study in Loire-Atlantique (France). *Journal of Transport Geography*. <https://doi.org/10.1016/j.jtrangeo.2021.103223>
- Hawley, A. H. (1950). *Human ecology: A theory of community structure*. Ronald Press.
- Howell, F. M., & Frese, W. (1983). Size of place, residential preferences and the life cycle: How people come to like where they live. *American Sociological Review*, 48(4), 569–580. <https://doi.org/10.2307/2117723>
- Klein, J. P., & Moeschberger, L. M. (2003). *Survival analysis: Techniques for censored and truncated data*. Springer.


- Lee, B. A., Oropesa, R. S., & Kanan, J. W. (1994). Neighborhood context and residential mobility. *Demography*, 31, 249–270. <https://doi.org/10.2307/2061885>
- Lee, E. S. (1966). A theory of migration. *Demography*, 3(1), 47–57. <https://doi.org/10.2307/2060063>
- Lee, R. E., & Cubbin, C. (2002). Neighborhood context and youth cardiovascular health behaviors. *American Journal of Public Health*, 92, 428–436. <https://doi.org/10.2105/ajph.92.3.428>
- Liang, Z. (2016). China's great migration and the prospects of a more integrated society. *Annual Review of Sociology*, 42(1), 451–471. <https://doi.org/10.1146/annurev-soc-081715-074435>
- Lichter, D. T., & Johnson, K. M. (2009). Immigrant gateways and Hispanic migration to new destinations. *International Migration Review*, 3, 496–518. <https://doi.org/10.1111/j.1747-7379.2009.00775.x>
- Liu, J., Li, S., Ouyang, Z., Tam, C., & Chen, X. (2008). Ecological and socioeconomic effects of China's policies for ecosystem services. *Proceedings of the National Academy of Sciences*, 105, 9477–9482. <https://doi.org/10.1073/pnas.0706436105>
- Liu, Y., & Shen, J. (2017). Modeling skilled and less-skilled interregional migrations in China, 2000–2005. *Population, Space and Place*. <https://doi.org/10.1002/psp.2027>
- Massey, D. S. (1990). Social structure, household strategies, and the cumulative causation of migration. *Population Index*, 56(1), 3–26. <https://doi.org/10.2307/3644186>
- Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., & Taylor, J. E. (1993). Theories of international migration: A review and appraisal. *Population and Development Review*, 19, 431–466. <https://doi.org/10.2307/2938462>
- Mazza, A., Gabrielli, G., & Strozza, S. (2018). Residential segregation of foreign immigrants in Naples. *Spatial Demography*, 6(1), 71–87. <https://doi.org/10.1007/s40980-017-0036-x>
- Migali, S., & Scipioni, M. (2019). Who's about to leave? A global survey of aspirations and intentions to migrate. *International Migration*, 57(5), 181–200. <https://doi.org/10.1111/imig.12617>
- Ord, J. K., & Getis, A. (1995). Local spatial autocorrelation statistics: Distributional issues and an application. *Geographical Analysis*, 27(4), 286–306. <https://doi.org/10.1111/j.1538-4632.1995.tb00912.x>
- Pickett, K. E., & Pearl, M. (2001). Multilevel analyses of neighborhood socioeconomic context and health outcomes: A critical review. *Journal of Epidemiology and Community Health*, 55, 111–122. <https://doi.org/10.1136/jech.55.2.111>
- Rangel, T. F., Diniz-Filho, J. A. F., & Bini, L. M. (2010). SAM: A comprehensive application for spatial analysis in macroecology. *Ecography*, 33(1), 46–50. <https://doi.org/10.1111/j.1600-0587.2009.06299.x>
- Ravenstein, E. G. (1885). The laws of migration. *Journal of the Statistical Society of London*, 48(2), 167–227. <https://doi.org/10.2307/2979181>
- Raymer, J., Bai, X., Liu, N., & Wilson, T. (2020). Reconciliation of Australian demographic data to study immigrant population change across space and time. *Spatial Demography*, 8(2), 123–153. <https://doi.org/10.1007/s40980-020-00060-9>
- Raymer, J., Wiśniowski, A., Forster, J. J., Smith, P. W. F., & Bijak, J. (2013). Integrated modeling of European migration. *Journal of the American Statistical Association*, 108(503), 801–819. <https://doi.org/10.1080/01621459.2013.789435>
- Rogers, A. (1968). *Matrix analysis of interregional population growth and distribution*. University of California Press.
- Rogers, A. (1975). *Introduction to multiregional mathematical demography*. Wiley.
- Sampson, R. J. (2003). The neighborhood context of well-being. *Perspectives in Biology and Medicine*, 46, S53–S64. <https://doi.org/10.1353/pbm.2003.0059>
- Seya, H., & Tsutsumi, M. (2013). Application of model averaging techniques to spatial hedonic land price models. In *Econometrics: New research* (pp. 63–88). Scopus.
- Shaw, R. P. (1975). *Migration theory and fact*. Regional Science Research Institute, University of Pennsylvania.
- Sjaastad, L. A. (1962). The costs and returns of human migration. *Journal of Political Economy*, 70(5), 80–93.
- Stark, O. (1991). *The migration of labour*. Harvard University Press.
- Stark, O., & Bloom, D. (1985). The new economics of labor migration. *The American Economic Review*, 75(2), 173–178.
- Stark, O., & Taylor, J. (1991). Migration incentives, migration types: The role of relative deprivation. *The Economic Journal*, 101(408), 1163–1178. <https://doi.org/10.2307/2234433>

- Sullivan, A., York, A. M., An, L., Yabiku, S. T., & Hall, S. J. (2017). How does perception at multiple levels influence collective action in the commons? The case of *Mikania micrantha* in Chitwan, Nepal. *Forest Policy and Economics*, 80, 1–10. <https://doi.org/10.1016/j.forpol.2017.03.001>
- Therneau, T. (2018). *Mixed effects Cox model*. R Package.
- Tiefelsdorf, M., & Griffith, D. A. (2007). Semiparametric filtering of spatial autocorrelation: The eigenvector approach. *Environment and Planning A*, 39, 1193–1221. <https://doi.org/10.1068/a37378>
- Van Dalen, H. P., Groenewold, G., & Schoorl, J. J. (2005). Out of Africa: What drives the pressure to emigrate? *Journal of Population Economics*, 18(4), 741–778. <https://doi.org/10.1007/s00148-005-0003-5>
- Vega, A., & Brazil, N. (2015). A multistate life table approach to understanding return and reentry migration between Mexico and the United States during later life. *Demographic Research*, 33, 1211. <https://doi.org/10.4054/DemRes.2015.33.43>
- White, M. J., & Johnson, C. (2016). Perspectives on migration theory—Sociology and political science. In S. White (Ed.), *International handbook of migration and population distribution* (Vol. 5, pp. 69–89). Springer.
- Wilson, A. G. (1974). *Urban and regional models in geography and planning*. Wiley.
- Xie, W. (2019). Left-behind villages, left-behind children: Migration and the cognitive achievement of rural children in China. *Population, Space and Place*. <https://doi.org/10.1002/psp.2243>
- Xu, Z., Xu, J., Deng, X., Huang, J., Uchida, E., & Rozelle, S. (2006). Grain for green versus grain: Conflict between food security and conservation set-aside in China. *World Development*, 34, 130–148. <https://doi.org/10.1016/j.worlddev.2005.08.002>
- Yang, S. (2019). Exploring reciprocal relationships between payments for ecosystem services (PES) and the associated coupled human and natural systems (CHANS): A case study of Fanjingshan National Nature Reserve (FNNR), China. Doctoral Dissertation, Joint Program in Geography, San Diego State University and University of California, Santa Barbara.
- Yang, Y. Q., Lei, X. P., & Yang, C. D. (2002). *Fanjingshan research: Ecology of the wild Guizhou Snub-nosed monkey*. Guizhou Science Press. (in Chinese).
- Yost, A., An, L., Bilsborrow, R., Shi, L., Chen, X., Yang, S., & Zhang, W. (2020). Mechanisms behind concurrent payments for ecosystem services in a Chinese nature reserve. *Ecological Economics*, 169, 106509. <https://doi.org/10.1016/j.ecolecon.2019.106509>
- Zhang, H., An, L., Bilsborrow, R., Chun, Y., Yang, S., & Dai, J. (2021). Neighborhood impacts on household participation in payments for ecosystem services programs in a Chinese nature reserve: A methodological exploration. *Journal of Geographical Sciences*, 31(6), 899–922. <https://doi.org/10.1007/s11442-021-1877-1>
- Zhang, Q., Bilsborrow, R. E., Song, C., Tao, S., & Huang, Q. (2018). Determinants of out-migration in rural China: Effects of payments for ecosystem services. *Population & Environment*. <https://doi.org/10.1007/s11111-018-0307-5>
- Zhang, Q., Wang, Y., Tao, S., Bilsborrow, R. E., Qiu, T., Liu, C., Sannigrahi, S., Li, Q., & Song, C. (2020). Divergent socioeconomic-ecological outcomes of China's conversion of cropland to forest program in the subtropical mountainous area and the semi-arid Loess Plateau. *Ecosystem Services*. <https://doi.org/10.1016/j.ecoser.2020.101167>
- Zhao, Y. (1999). Leaving the countryside: Rural-to-urban migration decision decisions in China. *The American Economic Review*, 89, 281–286. <https://doi.org/10.1257/aer.89.2.281>
- Zvoleff, A., An, L., Stoler, J., & Weeks, J. R. (2013). What if neighbors' neighborhoods differ? The influence of neighborhood definition on health outcomes in Accra. In J. R. Weeks & A. G. Hill (Eds.), *Spatial inequalities: Health, poverty and place in Accra, Ghana*. Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

Authors and Affiliations

Yanjing Liu^{1,2,7}  · **Jie Dai**^{1,2,3,4} · **Shuang Yang**^{1,3} · **Richard Bilsborrow**⁵ · **Minjuan Wang**⁶ · **Li An**^{1,2}

✉ Yanjing Liu
yliu739@ucsc.edu

¹ Department of Geography, San Diego State University, San Diego, CA, USA

² Center for Complex Human-Environment Systems, San Diego State University, San Diego, CA, USA

³ Department of Geography, University of California, Santa Barbara, CA, USA

⁴ Center for Global Discovery and Conservation Science, Arizona State University, Tempe, AZ, USA

⁵ Department of Geography and Carolina Population Center, University of North Carolina, Chapel Hill, NC, USA

⁶ Learning Design and Technology, San Diego State University, San Diego, CA, USA

⁷ Department of Environmental Studies, University of California, Santa Cruz, CA, USA