Migration Clusters in Brazil: an Analysis of Areas of Origin and Destination

Ernesto Friedrich Amaral

Research question and data

The main goal of this research is to analyze whether the pattern of concentration of migrants in a specific area of destination is the same as that of the area of origin of those migrants. The 2000 Brazilian Census has migration data for regions of origin (municipality) and destination (group of census tracts). Migration flows from 415 municipalities in the state of Bahia to 875 groups of census tracts in the mesoregion (metropolitan region) of São Paulo are analyzed. Migrants were considered men with at least 23 years of age (N=4,553). Using this classification, this study avoids the inclusion of familial migration (women and children). Moreover, since the migration was done between 1995 and 2000, the age of 23 is chosen to get those men that migrated with at least 18 years of age.

Descriptive analysis

The descriptive analysis was done spatially, using the Figures 1 to 9, which were constructed using both ArcMap and GeoDA. The Figure 1 just illustrates the migration flow that is analyzed in this study. Because the main internal migration in Brazil is the movement from the State of Bahia to the mesoregion of São Paulo, with high levels of migration rates, this is the stream investigated.

Figures 2, 3, 4 and 5 have information for the out-migration rates from the State of Bahia to the mesoregion of São Paulo. More specifically, Figure 2 indicates that the lowest out-migration rates are located in the areas in the West and Southeast parts of Bahia. Moreover, the higher rates are spread out in the center areas of the state. However, looking in the results of Figures 3 and 4, a more detailed analysis using the "hot spots" (Gi*), which uses information on migration of surrounding areas, indicates the formation of clusters in the state of Bahia. The low out-migration cluster indicates that municipalities with lower flows into São Paulo, that are surrounded by the same kind of municipalities, are located in the most developed areas in Bahia, which are in the West, for those municipalities with higher agriculture production; and in the East, for those municipalities surrounding the state capital of Salvador. The high out-migration cluster is located in central areas of the state. This cluster is formed by municipalities that tend to send more migrants to São Paulo, and have surrounding municipalities that also have high out-migration rates. Those municipalities are well-known by their low socioeconomic and developmental indices, comparing to the rest of the state of Bahia. Figure 5 indicates the results for the Local Indicator of Spatial Autocorrelation (LISA) and Moran's I. In general, this map emphasizes the results showed by the "hot spots" analysis, in which the West and Northeast are characterized by low levels of out-migration, and are surrounding by municipalities also with low levels of out-migration (blue areas). The municipalities with high out-migration rates that are surrounding by the same kind of municipalities are located in the middle of the state, as well as in the Center-East area of the state.

Figures 6, 7, 8 and 9 show the spatial analysis for the mesoregion of São Paulo. Figure 6 illustrates that higher in-migration rates are more spread out in the periphery of the area in study. Lower in-migration rates are more concentrated in the core of the mesoregion of São Paulo, as well as in the East part of this urban area. Figures 7 and 8 show the cluster analysis done with the "hot spot" (Gi*) technique. Low in-migration cluster, which aggregates those groups of census tracts with low migration rates surrounded by other groups of census tracts with low migration, are located in the middle part of the mesoregion, but more predominantly in the Southeast part of the area. Moreover, the high in-migration cluster indicates that the West and the Northeast areas of the mesoregion of São Paulo are characterized by high levels of in-migration rates for the groups of census tracts and their neighbors. Figure 9 presents the results for LISA and Moran's I. For this specific case, the cluster analysis is more difficult to be understood. However, the map still illustrates that groups of census tracts in the West and the Northeast portions of this urban area have high in-migration clusters. Moreover, in the Northwest, some areas are characterized by groups of census tracts that have low in-migration rates, but are surrounded by groups of census tracts with high in-migration clusters (light blue).

Regression models

The software GeoDA was used to run regression models that could explain both the out-migration rates from Bahia to São Paulo, and the in-migration rates to São Paulo from Bahia. After the use of OLS regressions, the results suggested the need to use spatial error models (Lagrange multiplier).

In the analysis of the out-migration rates, the independent variables were: mean years of education in each municipality of Bahia; and proportion of population working in each municipality. The outcome (out-migration rates) was calculated in such a way that in its denominator was included the male population with at least 23 years of age in each municipality.

In São Paulo, for the analysis of in-migration rates, the mean years of education in each group of census tract and the proportion of population working in each group of census tract are used as independent variables to explain in-migration rates. In this case, rates were calculated using the male population with at least 23 years of age in each census-tract group in the denominator.

Four sets of independent variables

The regression models were built using four different groups of independent variables. First of all, databases were organized only for males aged 23 or more. Among those, two different sets of independent variables were constructed. The first set used the original independent variables such as explained above. The second set used independent variables weighted by estimates of neighbor areas.

After that, databases were constructed getting information for the whole population. Then, the third set of independent variables used data for the entire population and the original independent variables. The fourth set of independent variables used the same data for the population as whole, but using independent variables weighted by estimates of neighbor areas.

Results

The analysis of regression models indicated that there was no significant difference between models using only men 23+ and whole population in covariates. Moreover, the OLS results indicated the need to use spatial error models (Lagrange multiplier).

The models for the State of Bahia indicated better estimates for spatially weighted years of education. In general, years of education are inversely correlated with out-migration. In addition, the proportion of population working is also inversely correlated with out-migration, but not statistically significant.

The models for the mesoregion of São Paulo indicated better estimates for original covariates (non-spatially weighted). Such as observed above, years of education are inversely correlated with in-migration. Finally, results suggest that proportion of population working is positively correlated with in-migration.

Comments on Spatially Weighted Covariates

In Bahia, migrants are more likely to leave municipalities with low levels of education, which are surrounded by municipalities with the same levels of education. In the case of São Paulo, migrants are more likely to move into areas with low education, and high proportion of population working, which are surrounded by areas with different levels in covariates. This is suggesting that for larger areas (municipalities), mean education and employment tend to have same levels of neighbors.

Future Improvements

Low-skilled migrants are moving into areas with higher opportunities of jobs, but with low-skilled population. However, the question on whether this is an indicative that those migrants are working on low-skilled jobs in São Paulo is not yet answered by those regression models. Even using spatially weighted covariates and spatial error models (Lagrange multiplier), Figures 10 and 11 indicate that regression residuals still do not have a randomly spatial distribution. One possible improvement to this study would be the inclusion of occupation variable in the model to get better estimates.

FIGURE 1
Migration Flow from the State of Bahia to the Mesoregion of São Paulo

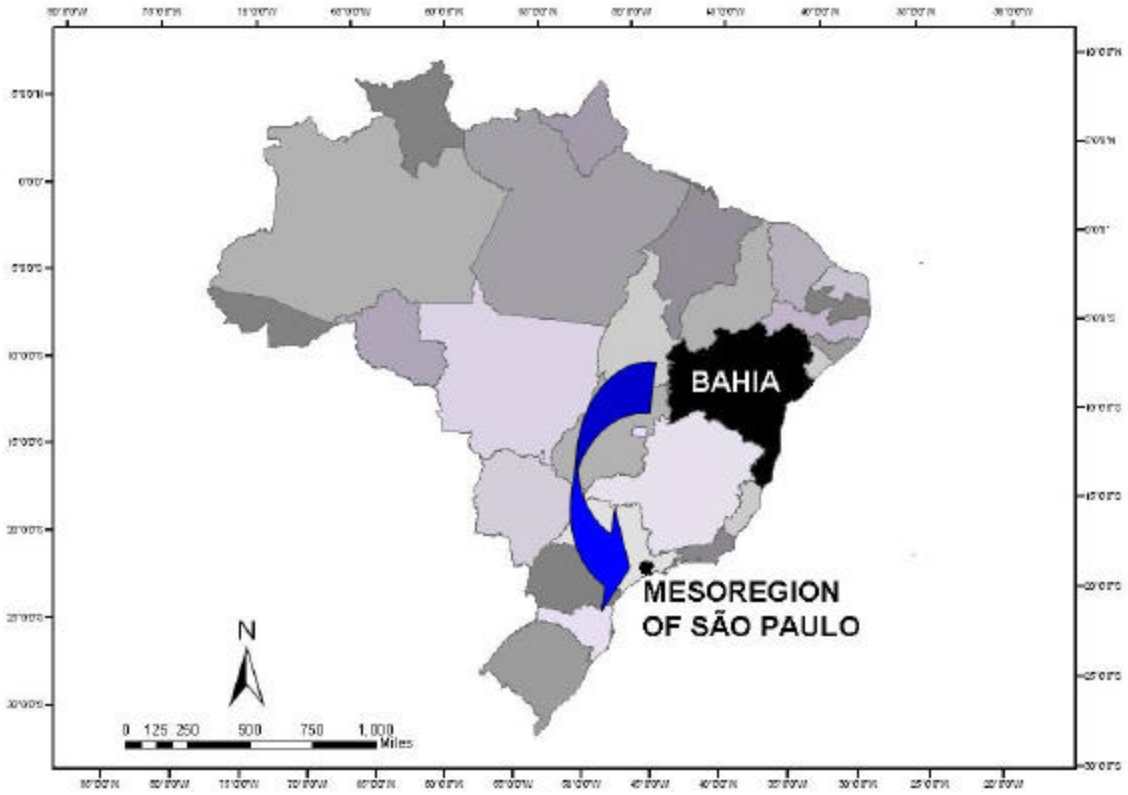

FIGURE 2
Out-migration Rates from the State of Bahia to the Mesoregion of São Paulo

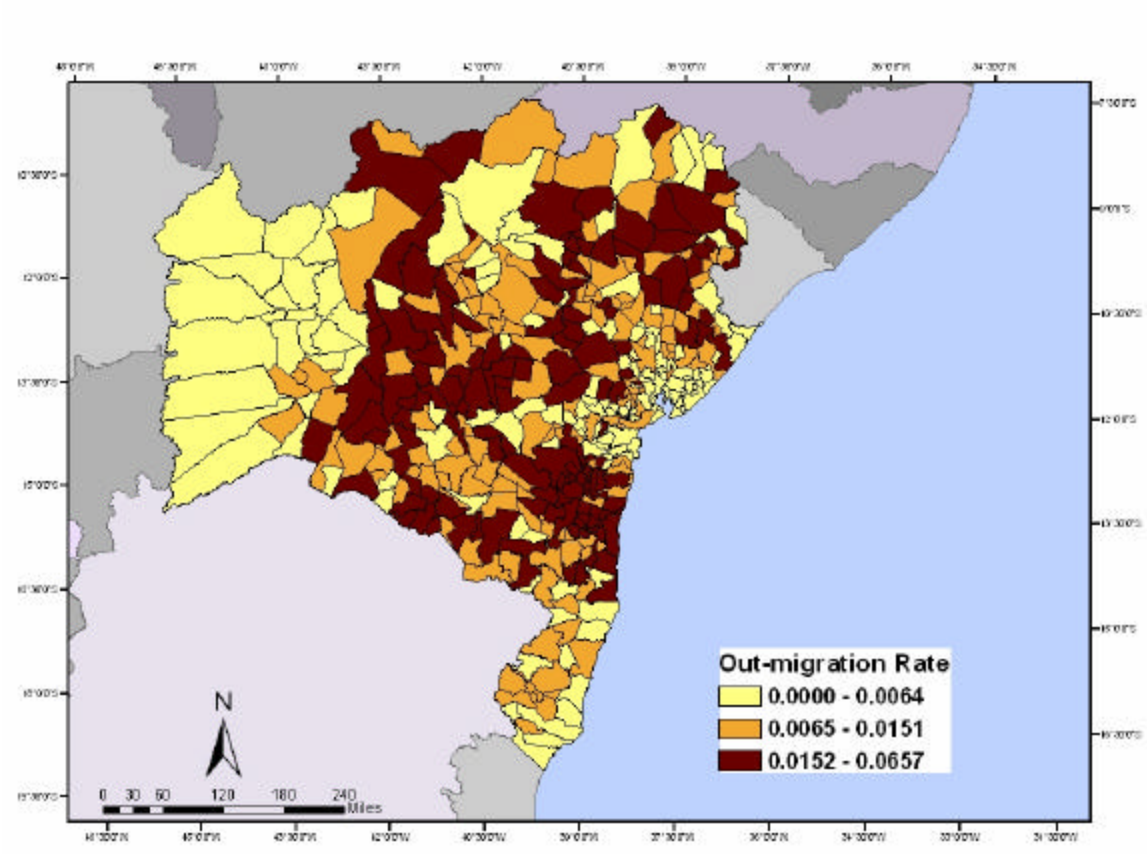

FIGURE 3
Hot Spots (Gi*) using Inverse Distance for the State of Bahia

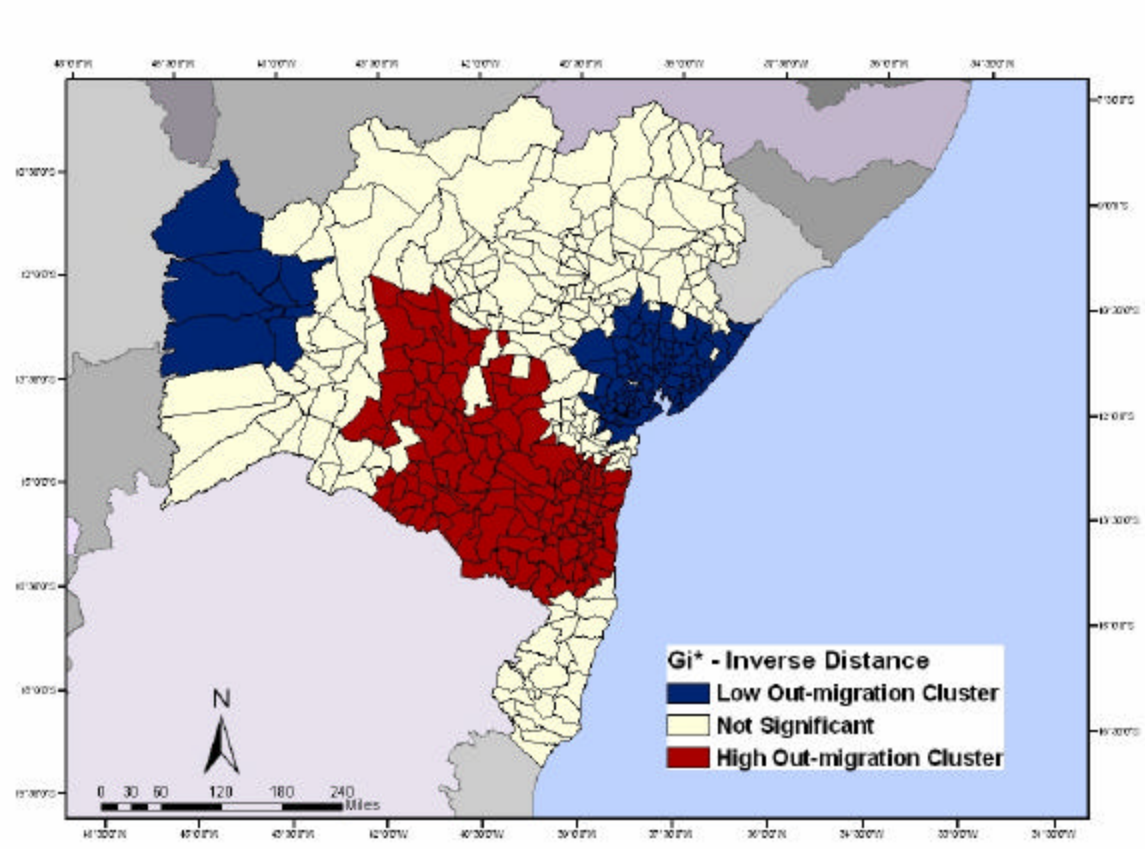

FIGURE 4
Hot Spots (Gi*) using Inverse Distance Squared for the State of Bahia

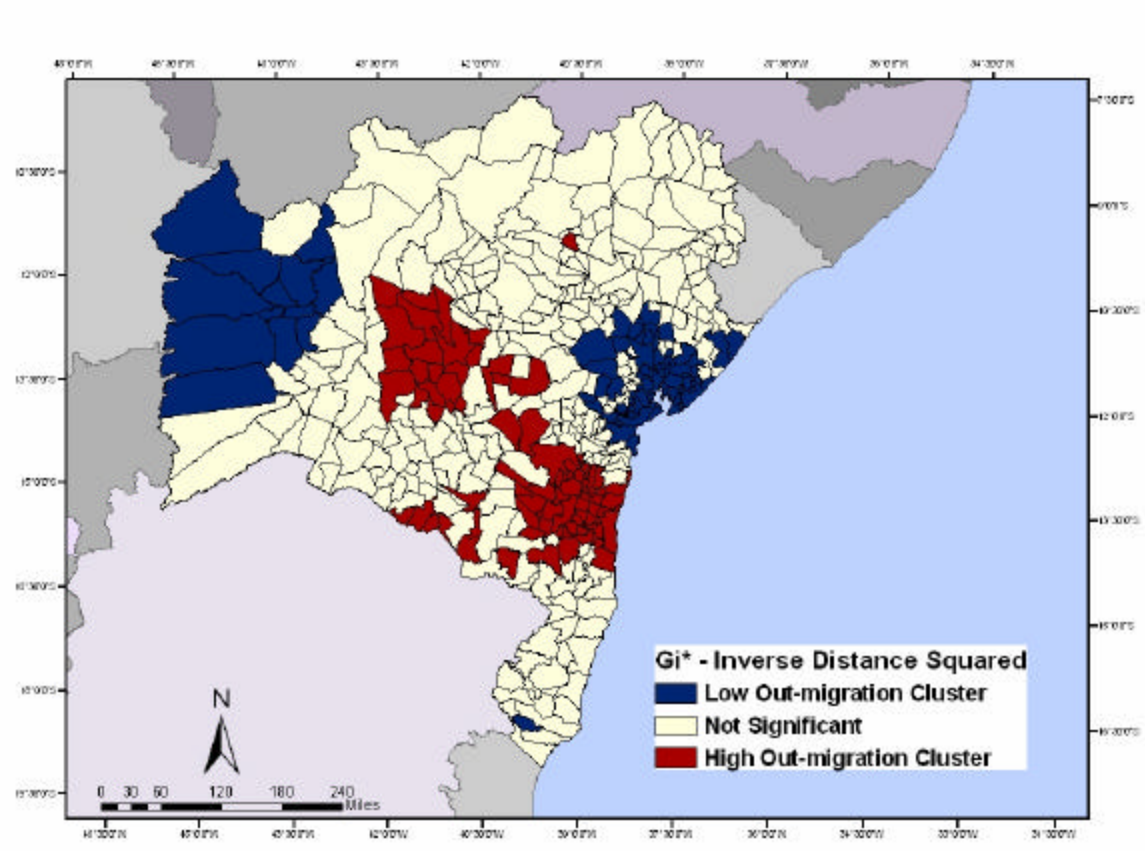

FIGURE 5
LISA Clusters for the State of Bahia

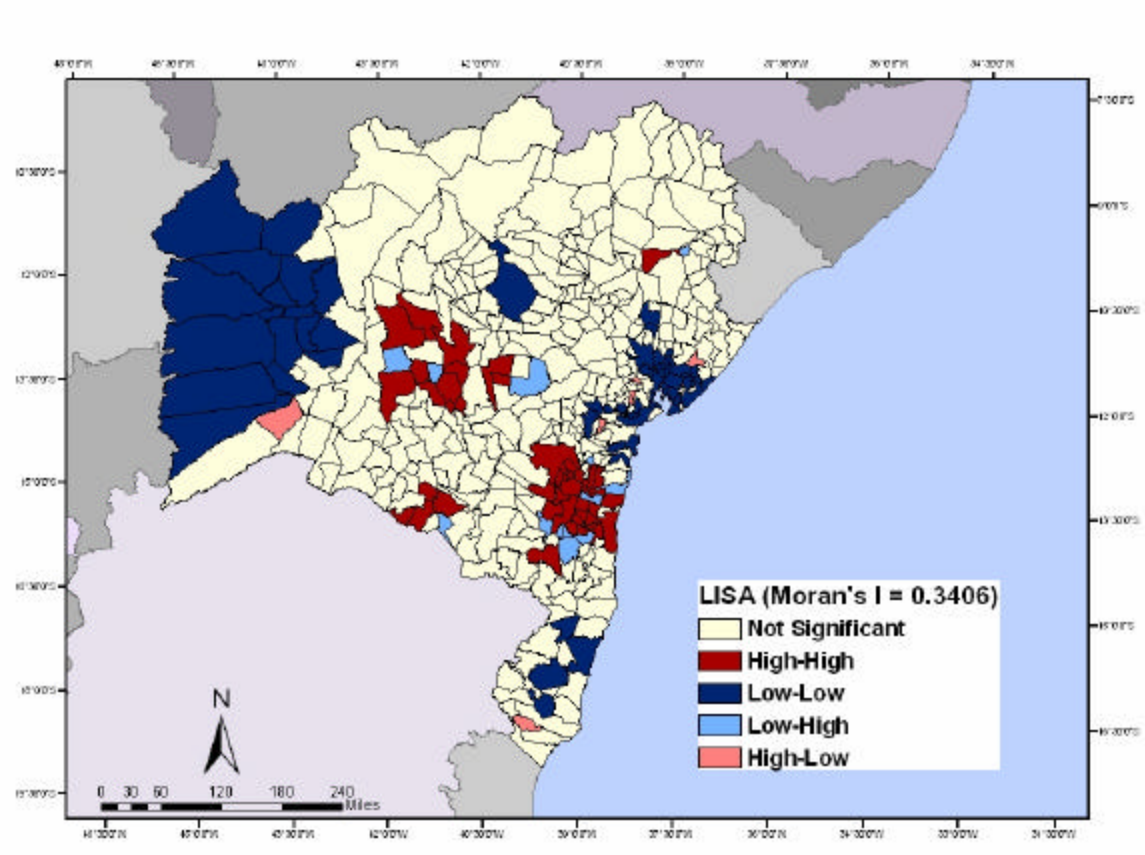

FIGURE 6
In-migration Rates to the Mesoregion of São Paulo from the State of Bahia

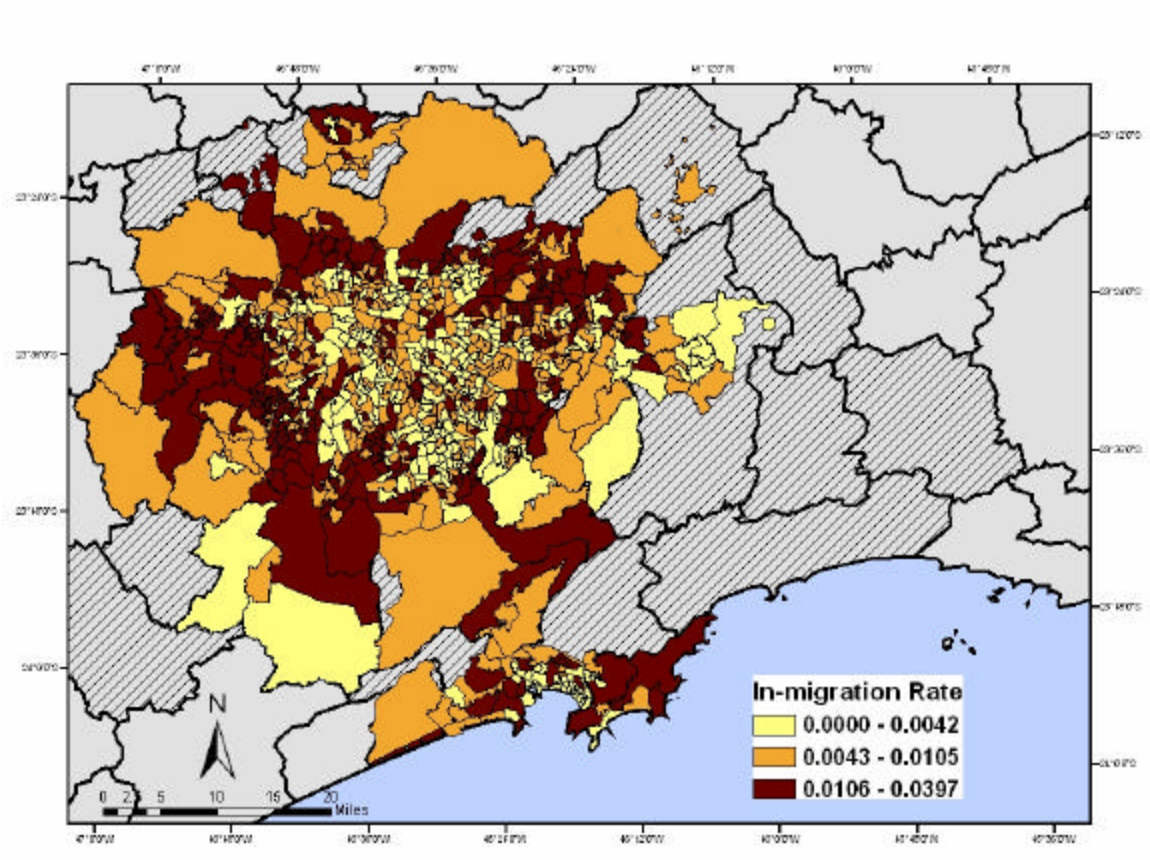

FIGURE 7
Hot Spots (Gi*) using Inverse Distance for the Mesoregion of São Paulo

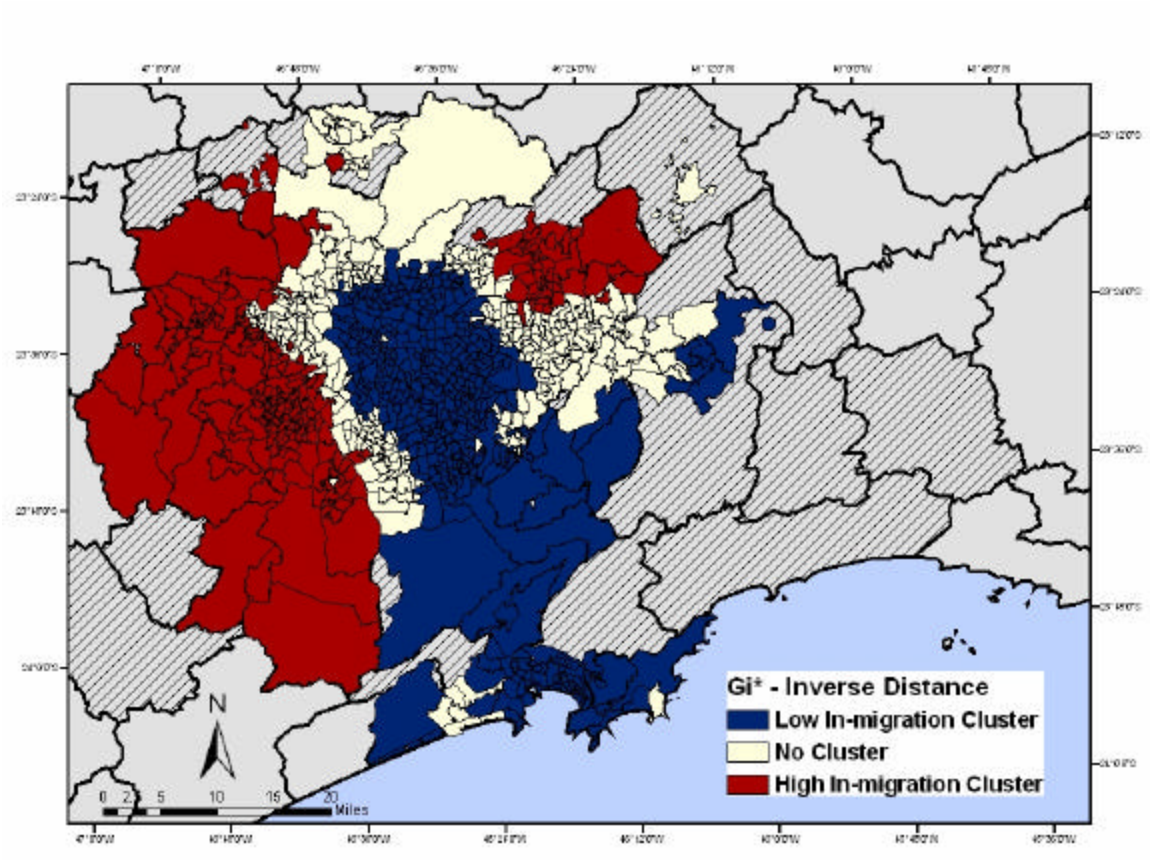

FIGURE 8
Hot Spots (Gi*) using Inverse Distance Squared for the Mesoregion of São Paulo

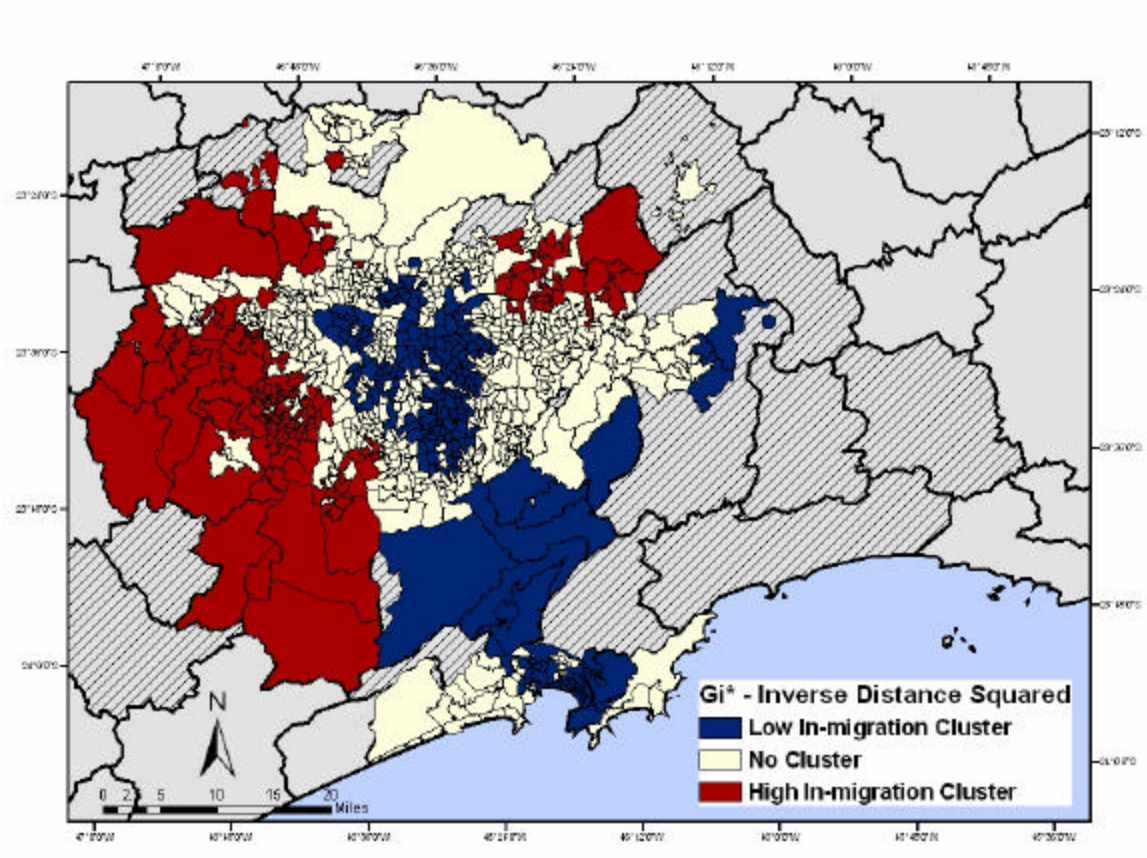

FIGURE 9
LISA Clusters for the Mesoregion of São Paulo

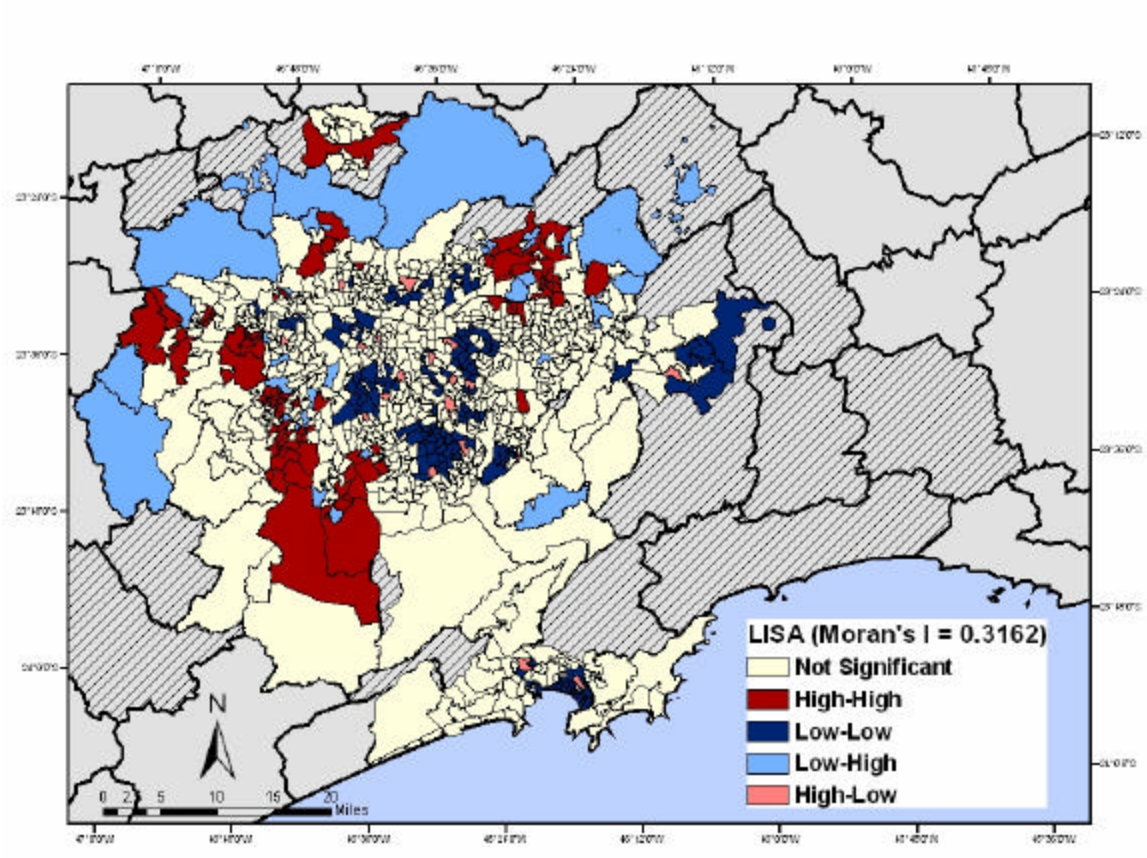

FIGURE 10
Regression Residuals for the State of Bahia

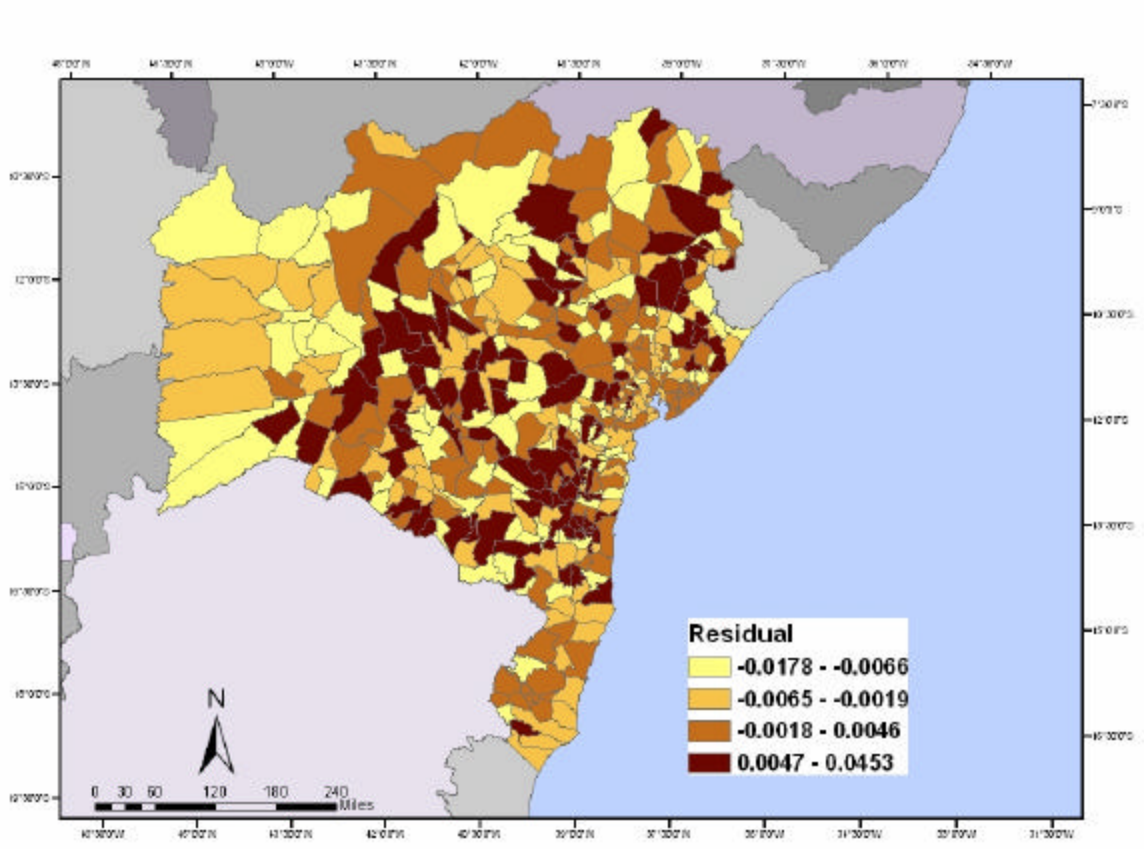

FIGURE 11
Regression Results for the Mesoregion of São Paulo