Methodological Advancements in Exploring Spatial and Temporal Variability in Associations between Rural Outmigration and Natural Resource Availability in Resource-Dependent Communities

Stefan Leyk[1], Galen J. Maclaurin[1], Lori M. Hunter[2], Raphael Nawrotzki[2], Wayne Twine[3], Mark Collinson[4] and Barend Erasmus[3]

[1]Department of Geography, University of Colorado Boulder, CO, USA
[2]Department of Sociology/Institute of Behavioral Science, University of Colorado Boulder, CO, USA
[3]School of Animal, Plant and Environmental Sciences, University of the Witwatersrand, South Africa
[4]School of Public Health, University of the Witwatersrand, South Africa

Abstract

Migration-environment models tend to be aspatial, even though the associations between outmigration and environmental explanatory variables are likely to vary across the broader study site. Thus the issue of spatial non-stationarity and temporal variation of migration-environment associations remains unexplored to date. This research expands beyond current approaches by developing migration models at different nested spatial scales (i.e. global, village, and subvillage) to explore the relationships between outmigration and socioeconomic and environmental variables. Demographic survey data from rural South Africa, combined with indicators of natural resource availability from satellite imagery are employed to investigate the spatial and temporal variations in these associations. We show how this nested modeling approach brings out different spatial patterns at finer scales and provides more detail about the observed associations. This allows us to evaluate how a general increase in detail influences model performance and variations in modeled relationships.

Introduction

Models of the migration-environment association in resource-dependent regions continue to become more sophisticated through increased use of, for example, longitudinal and/or multi-level models (e.g., Yabiku et al. 2009). Still, fairly simplistic indicators of environmental resources are generally included such as estimated rainfall or general undifferentiated measures of natural resource availability (e.g., Henry et al. 2004). Such measures do not allow in-depth investigation of associations that can be assumed non-constant or, non-stationary, within a larger region.

This research expands upon these models by tapping into the potential of spatially explicit demographic surveillance data from a remote rural region of South Africa, combined with indicators of both spatial and temporal variation in natural resource availability across the study site. We make use of the Normalized Difference Vegetation Index (NDVI) derived from MODIS remote sensing data to measure natural resource availability and variability.

Although migration-environment models tend to be aspatial, the associations between outmigration and its socioeconomic and environmental explanatory variables are likely to vary across the broader study site. Further, this spatial variation is not likely random and, instead, possibly varies in ways driven by socioeconomic and climatological conditions which could result in clusters of target associations. It is this spatial heterogeneity we explore here, and also use methodologically, to add nuance to our understanding of the migration-environment connection within rural South Africa. Existing approaches for exploring variations in outcome-predictor associations usually rely on local estimation e.g., varying coefficient models (Cleveland et al. 1991; Hastie and Tibshirani 1993) or geographically weighted regression models (Brunsdon et al. 1996; Fotheringham and Brunsdon 2002). Limitations of such models are local overfitting effects as well as non-robust frameworks to model non-normally distributed outcome data e.g., count data.

In this study we develop an analytical framework that overcomes the above limitations and offers a methodological approach that allows:

(1) comparison of models estimated on a set of sub-populations of the same level (e.g., villages) within the study site in order to study spatial variation of model associations. This procedure is similar to the idea of varying coefficient models but here each single estimation will be for a whole sub-population and thus have sufficient statistical rigor for robust modeling while avoiding overfitting effects from local estimations.

(2) comparison of models across (nested) population-levels i.e., global, village and subvillage levels in order to examine the effect of varying population and region size on the target model associations. This approach will help identify meaningful (sub)populations with stable target associations and no or negligible within unit variation.

(3) comparison of models for two different points in time (2002 and 2007) in order to estimate the effect of changing environmental or climatological conditions on target model associations.

Data and data processing

Using census data from 2002 and 2007 from over 12,000 households in 21 villages at the Agincourt Health and Demographic Surveillance Site (Fig. 1), an impoverished rural area in the northeast of South Africa, we derive household-level temporary outmigration counts of residents older than 15 years as our outcome variable. We also derive socio-economic and demographic attributes as household-level explanatory/control variables. In this analysis we derive an asset index, household socio-economic status (SES), and use it as a key central variable. This variable, which was constructed as an additive scale combining modern assets, livestock assets as well as information about power supply, water and sanitation, and dwelling structure, was identified as an important explanatory variable in our preliminary analysis as well as in recent research (Mberu 2006). We use only this one socio-economic control variable since our focus is on the methodological advancement in modeling environment-migration associations. Thus limiting the number of control variables keeps the complexity of the analysis at a moderate level.

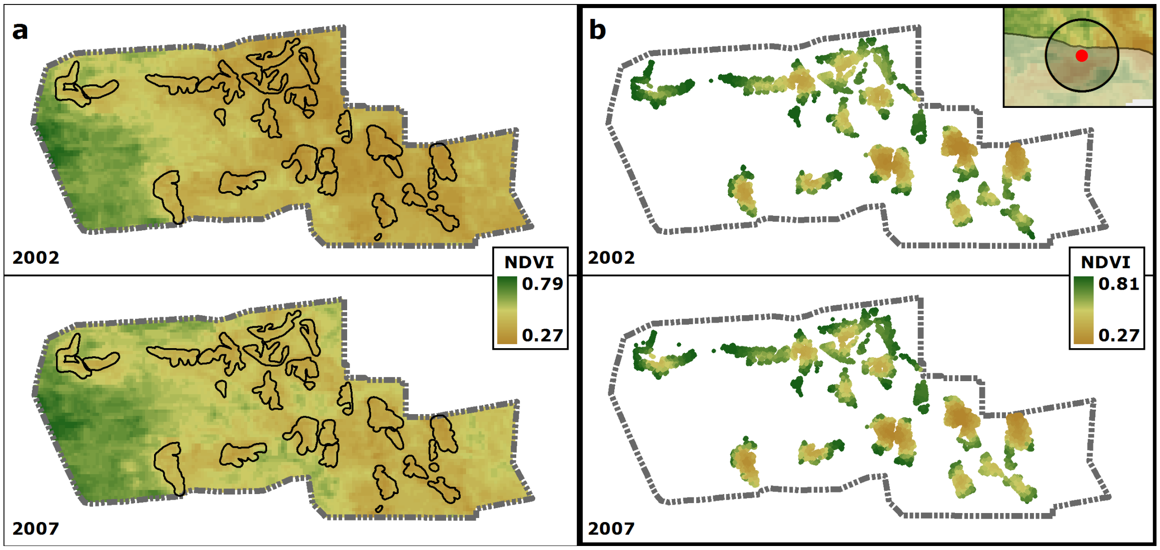

Fig. 1: Mean relative greenness, 2002 and 2007, for (a) the Agincourt study area; (b) extracted within household collection zones *(displayed at household level within the original village-level polygons).* Inset (b) shows an example of how natural resource availability was calculated for each household

We used the Normalized Difference Vegetation Index (NDVI) to derive a *greenness* metric to quantify natural resource availability for residents of the Agincourt study area on the household level (Fig. 1). NDVI has been used to monitor plant growth (vigor), density of vegetation cover and biomass production (Foody et al. 2001, Wang et al. 2004). It is therefore an effective measure of the availability of natural resources used in livelihood strategies (e.g. firewood, seeds, wild foods, fencing materials, etc.).

NDVI values for each year were calculated by taking the annual mean of 16 day composites from MODIS satellite imagery (250 meter resolution). From these annual means, we took the mean of the year of analysis and the two years prior to create greenness grids for 2002 and 2007 (Fig. 1a). Including the two years prior to the years of analysis takes into account the temporal variation in natural resource availability leading up to 2002 and 2007. The first time period was characterized by relatively high but slightly decreasing greenness values (between 0.53 – 0.47 in average); the latter time period showed similar mean greenness but higher variation across the years with an increasing trend from 2005 to 2007 (between 0.43 – 0.54). Thus in average the mean greenness values where very similar in both time periods but were based on different "histories" of resource availability. As a result the spatial distributions in 2002 and 2007 showed some visible differences (Fig. 1). From the two greenness grids, we excluded areas within village boundaries as these are not communal lands and therefore are not used for collecting resources. The next step was to create buffer zones of 2000 meters around each household (but excluding the area within that village and within neighboring villages) (Fig. 1b). The buffer distance was based on the distance that residents tend to travel to access natural resources. Finally, the sum of NDVI values within this buffer zone was calculated for each household and used as a surrogate for the amount of natural resources available to the members of each household.

Methods

Defining Regions for varying population levels

We created three different nested sets of underlying (sub)populations ("population levels") that were used subsequently for model estimation. First, we fit models on the *global level* taking the total number of households in all 21 villages as input. Second we developed *village-level* models (i.e., a model is fitted for each individual village). These village-level models are then compared with each other. Finally, we derived *subvillage-level* regions. Since there are no census units defined within a village, we developed a procedure to randomly simulate subvillage regions (Fig. 2). The intention was to carry out increasingly "localized" model estimations of outmigration that are still based on underlying populations of sufficient size and variability to develop robust full statistical models. This approach is in contrast to local estimations such as in Geographically Weighted Regression (GWR) models (Brunsdon et al, 1996; Fotheringham and Brunsdon, 2002) which have the disadvantage of overfitting effects due to the methodological setup.

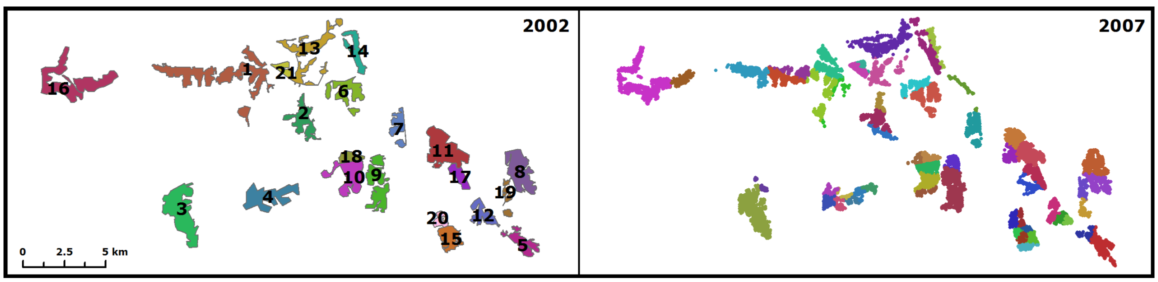

Fig. 2: Different population (or region) levels used in this analysis in order to subdivide the global population: (a) village level, and (b) subvillage level (one exemplary random regionalization outcome)

Nested subvillage regions were randomly generated such that they subdivide villages into smaller exclusive (non overlapping), contiguous, nested areas. Essentially this is a spatial bootstrap sampling method. Our approach enforced subvillage regions to contain a minimum of 51 and a maximum of 548 households. These thresholds enforce that the smallest and largest villages are split at least in two regions while still maintaining a statistically robust sample size. The number of households in each region varied randomly between these two thresholds. Spatial contiguity in this regionalization process was achieved by randomly selecting seed points (households) within each village. Regions were then generated by joining all the remaining households with the closest seed point to create the subvillage units. We performed this simulatation 500 times to assess the average model structure and performance over all regionalization runs in the subsequent modeling process.

Modeling and coefficient derivation

At each population-level (i.e., global, village and sub-village), we fit Generalized Linear Models (GLM) for Poisson-distributed household-level temporary outmigration counts. For each model we derived coefficient estimates and their corresponding p-values. We tested model residuals for spatial autocorrelation using Global Moran's I (Moran, 1950) as well as for spatial clustering using Local Moran's I (Anselin 1995).

At the subvillage-level, the approach was the same yet models were fit for random regions across the 500 simulations. For each subvillage model from each simulation, the coefficient estimates and their p-values were stored for all households that were part of that region. Thus for each household model estimates were stored from 500 different configurations of subvillage regions. Finally we took the mean coefficient estimates for each household across all simulations and calculated the proportion of the simulations where those coefficients were significant ($p < 0.05$). Thus a spatial distribution of varying mean model coefficients at the household level was created based on a series of statistically robust sub-village models.

Mapping spatial/ temporal variation of target associations at different population levels

At the village level we created maps that show coefficient values and statistical significance using village boundaries (polygon feature data). At the subvillage level, we mapped mean coefficient estimates and the proportion of significant tests over 500 model runs for each household location (point feature data). This mapping allowed us to observe and visualize (1) changes in model structure across different population (or region) levels for the whole study area, and (2) at each population level the spatial variation or spatial heterogeneity in our two target associations (i.e., SES-outmigration and NDVI-outmigration).

These maps were created for both points in time (2002 and 2007) in order to compare the model outcomes and spatial structure in the generated resultant data when using the exact same methodological setup.

In order to better understand the spatial structure of the model performance, and thus to identify regions of potential under and over prediction, we also created maps of Local Moran's I, a class of Local Indicators of Spatial Association (LISA) (Anselin 1995) on the model residuals.

Results and Discussion

Global model outcomes

The global level model diagnostics indicated that both explanatory variables, SES and NDVI, are highly significant in 2002 and 2007 ($p < 0.01$). The model coefficient for SES is positive in both years but considerably smaller in 2002. The coefficient for NDVI is positive and very similar in both years.

The residuals of the global model show significant spatial autocorrelation based on Moran's I ($p < 0.05$), which suggests that the error structure is not random. This test result, which in conventional procedures would indicate a need for model approaches that account for spatial dependence in the error structure, can be further examined based on maps of clusters of local spatial association (Fig. 3). As can be seen there are significant local clusters of high residual values that are well separated from clusters of low residual values. This translates to spatially clustered over- and under-predictions, respectively, and suggests the need for approaches that account for spatial non-stationarity in order to better understand the target associations. In this analysis, we do this by investigating spatially refined subpopulations or subregions as described in the methods. The spatial locations of the identified LISA clusters do not vary considerably between 2002 and 2007 (Fig. 3).

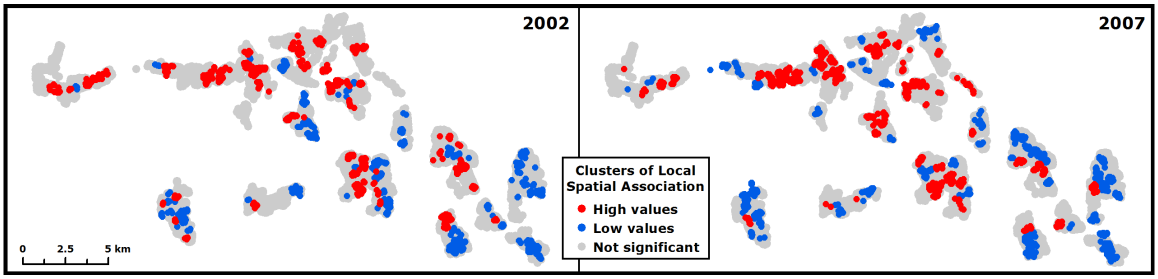

Fig. 3: Maps of statistically significant ($p < 0.05$) local clusters of high and low residual values of the global population level model computed using LISA tests.

	Global Level				Village Level				Subvillage Level			
	SES Coefficients		NDVI Coefficients		SES Coefficients		NDVI Coefficients		SES Coefficients		NDVI Coefficients	
Year	**Est.**	**Std. Error**	**Est.**	**Std. Error**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**	**Mean**	**Std. Dev.**
2002	0.082	0.025	0.290	0.081	-0.016	0.111	-0.079	0.353	0.008	0.282	0.623	4.526
2007	0.166	0.026	0.321	0.074	-0.172	0.152	0.127	0.187	0.178	0.228	-0.061	3.210

Table. 1: Summary of model coefficients (SES and NDVI) for different population (or region) levels used in this analysis: global population level, village level, and subvillage level.

Village level model outcomes

When running the models on the village level, we discover considerable variation in coefficient estimates across villages and interesting patterns of change between the two years (Fig. 4). We found significant residual autocorrelation in only three villages, which suggests that the degree of residual spatial dependence decreases considerably at the village level.

The overall SES-outmigration association on the village level is negative on average in both years as opposed to positive associations found on the global level (Table 1). The associations appear to be more stable (less spatial variation) in the later time period (2007) and shows more variation in the first time period (2002) (Fig. 4a). The originally computed values of the SES variable show higher mean values over the whole study site in 2007 (in a year when higher mean greenness was found suggesting higher availability of natural resources). The coefficient values of SES decrease in most villages in 2007 (i.e., change signs or become more negative). This represents an interesting finding since while there is higher variation in the migration-SES association in the early time period, this relationship is predominantly negative in most villages and of less variation in 2007. This indicates that households with lower SES made decisions to migrate more likely in 2007 than in 2002. The observed change in spatial patterns of model relationships between the two years is remarkable and suggests a possible influence of changing environmental conditions.

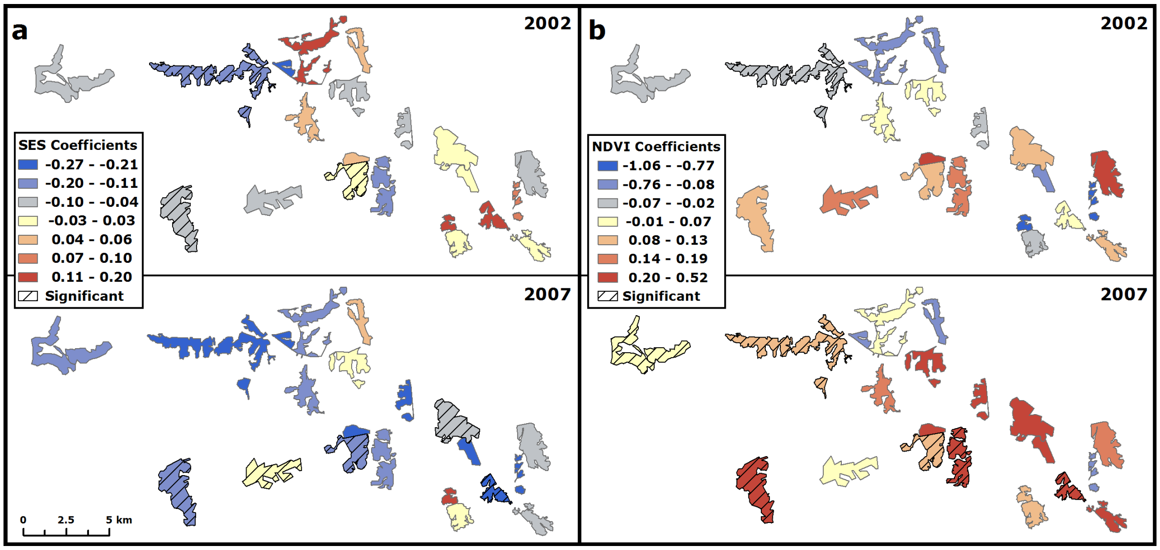

Fig. 4: Village level model coefficients for target associations (a) SES-outmigration and (b) NDVI-outmigration for 2002 and 2007. If coefficients were tested significant ($p < 0.05$) they appear hashed in the figure.

The migration-NDVI association on the village level, which is on average slightly negative in 2002 and positive in 2007, also shows a high degree of variation across villages in the first time period and is non-significant for most villages (Fig. 4b top). For the later time period a higher number of villages show significant (positive) migration-environment relationships (Fig. 4b bottom). This change in spatial association patterns suggests that on the village level NDVI becomes a more important explanatory variable under changing environmental conditions (i.e., higher natural resource availability in 2007).

In summary, the maps in Fig. 4 illustrate considerable variation in the spatial patterns of model associations on the village level as well as differences between the two time periods which would remain completely hidden when using global modeling approaches only.

Sub-village level model outcomes

The same model procedure applied to the subvillage level over 500 simulations shows more refined spatial patterns of model associations (Fig. 5). Within village boundaries there is considerable spatial variation in the mean coefficient estimates for both SES and NDVI. This level of analysis shows that the relationships observed at the village level break down significantly and coefficient estimates even switch signs in some of the villages (Fig. 4a and 5a). At this level of analysis, new spatial patterns in the target associations can be discovered. This indicates that there is considerable local variation of SES and NDVI associations within villages such that village-level models cannot capture this variation sufficiently. Clusters of similar coefficient estimates could be used to identify within-village regions of strong and similar migration-SES and migration-NDVI associations. The existence of such clusters would demonstrate that such associations can change drastically at finer spatial scales and could be used to refine our understanding of the processes at work. The described modeling approach ensures that these areas of similar associations are not created by chance as they are based on statistically robust subvillage models, using randomly generated regions over 500 simulations.

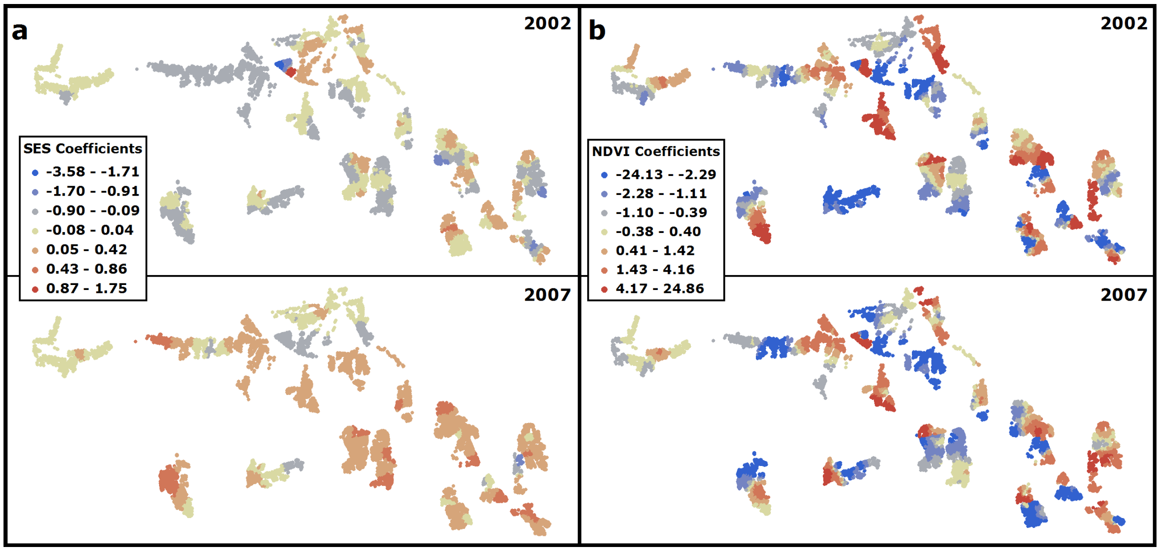

Fig. 5: Sub-village level average model coefficients for the two target associations (a) SES-outmigration and (b) NDVI-outmigration for 2002 and 2007 over 500 simulations and thus based on 500 model runs.

The proportions of coefficients tested significant over 500 model runs (Fig. 6b) provide additional information about the average strength of the considered association in explaining household outmigration. The spatial patterns of the migration-NDVI association show overall similarity but some interesting differences can be discovered such as in villages 4 or 11.

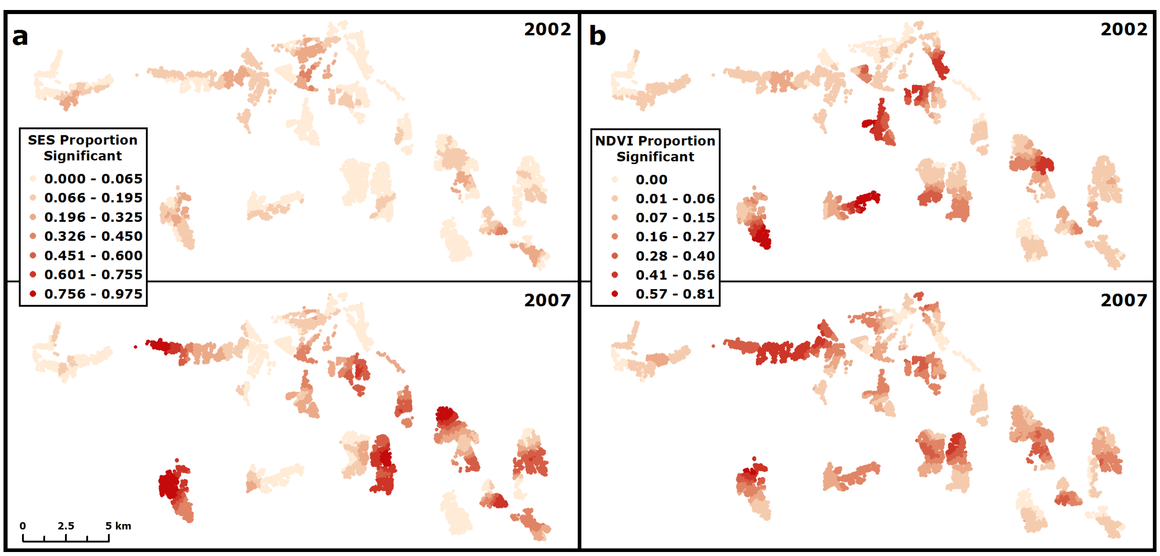

Fig. 6: Sub-village level proportions of model coefficients that have been tested significant over 500 simulations (regionalizations) and thus based on 500 model runs for the two target associations (a) SES-outmigration and (b) NDVI-outmigration for 2002 and 2007.

The migration-SES association shows less spatial variation within villages in both years but there are unexpected differences when the two years are compared (Fig. 5a). In 2002 the average coefficient values show positive and negative values, whereas in 2007 the associations are predominantly positive. This represents a very different picture when compared to the village level coefficients (Fig. 4a). The subvillage coefficient estimates in 2007 are supported by high proportions of significant associations (Fig. 6a). Thus again the finer scale analysis shows different spatial patterns from that at the village level, and in doing so provides more detail about the observed associations.

Concluding Remarks and Outlook

This research addresses an urgent need in migration research to investigate the impacts of environmental variables on temporary outmigration in addition to commonly used socio-economic variables on the household level in an impoverished region in rural South Africa. The household level survey data available for this area allow a very detailed investigation of such research questions; the existence of spatial identifiers of single households provides a unique opportunity to investigate the spatial detail of migration patterns and relationships.

The presented analysis provides unique insights into the spatial variation or spatial heterogeneity of migration-environment associations in resource-dependent regions in rural South Africa. In particular, we investigate model associations between temporary household-level outmigration and independent variables including NDVI-based environmental measures as well as commonly used socio-economic factors. The described observations allow for much more in-depth investigation and understanding of spatial patterns of migration-environment associations but also of associations between migration and commonly used attributes such as SES. Such analysis outcomes remain hidden when using global statistical models.

Our framework allows for comparing model associations across different nested population levels (i.e., global, village and sub-village) in order to evaluate how a general increase of detail influences model performance and variations in modeled relationships. This demonstrates how such model associations can be impacted by the size of the underlying population or the populated area/region.

More importantly, we were able to quantify the variation in the target associations across the sub-populations or sub-regions of the same level (e.g., across all villages) within the study site. The strength of the presented procedure lies in the development of full models with sufficient statistical power which makes it more reliable for interpreting the results. The estimated model associations on the sub-village level show considerable variation even within villages which might indicate that community-level dynamics are very important and influential for migration decisions on the household level. This addresses important questions regarding the degree of variation in migration related associations that can be expected within political or administrative units such as villages.

In our example we show that the associations between migration and explanatory variables, SES and NDVI, produce different degrees of spatial variation across the study site. Interestingly, we discovered considerable differences in these patterns between the two years of interest suggesting that target associations can change under varying environmental conditions. In other words, there is some indication that the change in environmental conditions can impact such model associations and thus the relevance of single variables in explaining outmigration on very local scales.

Within the months between now and the PAA meetings, we will focus on the refinement of this methodological framework by using more control variables and including interactions between socio-economic factors and environmental measures. In the longer term we will attempt to include a higher number of points in time in order to analyze changes over time and better understand the impact of changing conditions in the environment or general socio-economic conditions. Finally, we will delve into the nuance in substantive interpretation afforded by application of these methodological advancements to population-environmental modeling within the Agincourt study site.

Acknowledgements

Supported by NIH 1R03 HD061428, "Environmental Variability, Migration, and Rural Livelihoods." The work has also benefited from the NICHD-funded University of Colorado Population Center (grant R21 HD51146) for research, administrative, and computing support. The content is solely the responsibility of the authors and does not necessarily represent the official views of the CUPC, NIH, or NICHD.

References

Anselin, L. 1995. "Local Indicators of Spatial Association – LISA." *Geographical Analysis* 27(2):93-115.

Brunsdon, C. F., A. S. Fotheringham, and M. E. Charlton. 1996. "Geographically Weighted Regression: A Method for Exploring Spatial Nonstationarity." *Geographical Analysis* 28(4):281-298.

Cleveland, W. S., E. Grosse and W. M. Shyu. 1991. "Local regression models." In *Statistical Models in S* (Chambers, J. M. and Hastie, T. J., eds), 309–376. Wadsworth & Brooks, Pacific Grove.

Foody, G. M., M. E. Cutler, J. McMorrow, D. Pelz, H. Tangki, D. S. Boyd, and I. Douglas. 2001. "Mapping the Biomass of Bornean Tropical Rain Forest from Remotely Sensed Data." *Global Ecology and Biogeography* 10(4):379-387.

Fotheringham, A. S., C. Brunsdon, M. Charlton. 2002. *Geographically Weighted Regression: The Analysis of Spatially Varying Relationships*. Wiley, New York, 269 pp.

Hastie, T. and R. Tibshirani. 1993. "Varying-Coefficient Models." *Journal of the Royal Statistical Society*. Series B 55(4):757-796.

Henry, S., B. Schoumaker, et al. 2004. "The Impact of Rainfall on the First Out-Migration: A Multi-level Event-History Analysis in Burkina Faso." *Population and Environment* 25(5): 423-460.

Mberu, B.U. 2006. "Internal migration and household living conditions in Ethiopia." *Demographic Research* 14:509-539.

Moran, P. A. P. 1950. "Notes on continuous stochastic phenomena." Biometrika 37:17–33.

Wang, J., P. M. Rich, K. P. Price and W. D. Kettle. 2004. "Relations between NDVI and Tree Productivity in the Central Great Plains." *International Journal of Remote Sensing* 25(16):3127-3138.

Yabiku, Scott, Jennifer E. Glick, Elizabeth A. Wentz, Steven A. Haas and L. Zhu. 2009. "Migration, health, and environment in the desert southwest." *Population and Environment*. 30(4-5):131-158.