TITLE

An analytical approach to assess spatial interplays among migration drivers: the case study of Ghana

ABSTRACT

Scholarly research has investigated how slow-onset climate factors are associated with population vulnerability for the creation or exacerbation of economic and violent conflicts. Yet, the lack of data makes difficult the measurement of interplays among contextual factors and population dynamics. Taking advantage of newly available net migration datasets, which provide estimates of migration changes at fine spatial resolution, we define a three-step methodological strategy. First, we use the Geographically-Weighted-Regression to assess spatial variations of the estimated local parameter effects by geographical unit. Second, we adopt the Multiscale-Geographically-Weighted-Regression, recently applied in different fields of social science research, to process flexibility in spatial heterogeneity at different geographic scales. Third, we adopt the variation of forest models, the Geographically-Weighted-Random-Forest-Regression, which disaggregates the traditional forest model into multiple local models to highlighting local variations. Our empirical analysis focuses on Ghana to explore how precipitation patterns, extreme temperatures, conflicts, and artisanal (small-scale-company) gold mining have influenced migration patterns over 30 years, from 1985 to 2014. By offering a comparative perspective to interpret results, our study contributes to assess at what extent geospatial regression techniques and machine learning models complement each other to depict spatial-temporal variations of the relationship among environmental conditions, conflicts and migration patterns.

LONG ABSTRACT

BACKGROUND

Studies have well documented that droughts and precipitation anomalies as having the potential to lead to a range of short- and long-term migration outcomes, being highly context specific (Gray and Wise, 2016). There is also research suggesting that environmental factors can have an indirect influence on migration through the creation or exacerbation of violent and/or non-violent conflict over access to land or natural resources (UN-Habitat, 2022). As Ferris (2020) notes, beside scholars and humanitarian actors mostly interested in the consequences of displacements linked to sudden-onset, an increasing number of development researchers recognize the assessment of long-term consequences of slow-onset events as a crucial strand to elucidate linkages between climate, conflict, and migration.

THE CASE OF GHANA

Environmental and climate factors have been identified by scholars as having effects on migration patterns in Ghana. It has been widely documented that households, depending on subsistence agriculture in the rural north regions, have experienced out-migration to urban and rural areas in the comparatively wetter central and southern parts of the country (Kuuire et al 2016, Luginaah et al 2009, Teye et al. 2015, Schraven and Rademacher-Schulz 2016, Sow et al 2014, van der Geest et al 2010, van der Geest 2011). Districts that have relatively higher levels of out-migration tend to be those with lesser vegetation cover and/or suffer from land degradation, with out-migrants from these areas often destined to the southwest regions, where cocoa plantations offer employment or to artisanal gold fields, mainly located in the west parts of the country (van der Geest et al 2010, Nyame and Grant 2014). Such moves have been increasingly initiated during the rainy season and differ from the circular patterns of out-migration experienced during the dry season (and its subsequent return) that have been practiced in the northern regions by generations, to maintain household food security (Rademacher-Schulz et al 2014, Schraven and Rademacher-Schulz 2016). Looking at slow-onset environmental conditions, there is usually a set of overlapping political and socioeconomic factors at play that shape changes in migration behaviours (Olaniyan et al., 2015). For instance, conflicts emerging from land tenure disputes have a history of displacing people in Ghana. Studies suggest that conflicts often ensued between pastoralists (such as Fulani herders) and various farming communities, arise from the crop destructions caused by stray cattle (Baidoo, 2014; Issifu et al. 2022). Various hotspots of farmer-herder conflicts are observed throughout Ghana, with violence in Agogo in the Ashanti Region (Appiah-Boateng and Kendie, 2022), and tensions in northern Ghana and the Volta basin (Tonah 2000, Tonah 2006) as herders seek viable lands for cattle.

DATA

To carry out the analysis, we assemble: i) net migration estimates; ii) environmental and climate conditions; iii) conflict events; iv) observations of economic activities. All variables have been harmonized at 25km2 grid resolution. i) The net migration estimate dataset (Alessandrini, Ghio, Migali, 2020) is overlaid across Ghana. The spatial grid includes cells that are centered within the country of Ghana and do not overlap with cells centered within neighboring countries. National geographical regional boundaries of Ghana are used to identify and describe areas in addition to harmonized data; although, regional boundaries changed in 2018, hence our analysis and visualizations refer to the previously established 10 regional borders (Ghana Statistical Service, 2010). ii) Ghana exhibits unique environmental conditions, which include both arid zones and tropical rainforests, enabling the study of different climate variables in relation to net migration (Dinerstein et al., 2017). Using the Climate Hazards Group InfraRed Precipitation with Station (CHIRPS) (Funk et al., 2015) and the Climate Hazards Group Daily Temperature (CHIRTS) (Funk et al., 2019), we derive the following climate change indicators: a) Precipitation, meaning 5-year mean annual precipitation (PREC); b) Maximum length of consecutive dry days in one year, where rain is less then 1 mm, calculated from CHIRPS total gridded rainfall, and averaged over five years (MLDS); c) Maximum length of consecutive wet days in one year, where rain is greater than or equal to 1 mm, calculated from gridded rainfall, averaged over five years; d) Average daily maximum temperatures, averaged to 5-year intervals; e) Average daily minimum temperatures, averaged to 5-year intervals. iii) We rely on significant conflict events and related fatalities recorded at geo-point locations by the Armed Conflict Location and Event Data Project (Raleigh et al., 2010) and Uppsala Conflict Data Program (UCDP) Georeferenced Event Dataset (Sundberg et al., 2013). iv) For the periods 2005-2009 and 2010-2014, we use the database provided by the College of Earth Sciences, Chengdu University of Technology and the School of Geography, Earth and Atmospheric Sciences, Faculty of Science, University of Melbourne (Tang & Werner, 2023). The datasets are derived from 2008-2013 high-resolution satellite imageries to identify the areas of economic activities, such as the artisanal small scale gold mining activities in Ghana.

METHODS

A three-step approach is defined as follows. First, the Geographically Weighted Regression (GWR) is used to model the spatial variations of the estimated local parameter effects by geographical unit. Second, we adopt the Multiscale Geographically Weighted Regression (MGWR), proposed by Oshan and colleagues (Oshan et al. 2019). The model has been recently adopted in different fields of social science research due to its flexibility to process spatial heterogeneity at different scales (Fotheringham 2017). The main advantage of using MGWR consists of removing the spatial non-stationary condition and differentiating the local effects by each dependent variable. This implies the estimates of covariate-specific bandwidths (Yu and Fortheringham, 2020), as formally explained by the following equation:

$$y_i = \sum \beta_{bwj} (u_i+v_i) x_{ij} + \epsilon_i$$

β_{bwj} corresponds to the bandwidth resulting from the application of a spatial weighting kernel for approximating x_{ij} , namely the predictor j of the variable x_i.The bandwidths are selected through a back-fitting algorithm to optimize the expected log likelihood (Oshan et al., 2019). We apply an adaptive Gaussian function to deal with the spatial distribution of observations. A Monte Carlo approach is used to check the nonstationary conditions for each variable (Fotheringham, Brunsdon, and Charlton, 2002). Third, we adopt a recent variation of forest models (Georganos et al., 2021, Georganos and Kalogirou, 2022), the Geographically Weighted Random Forest Regression (GWRF). The GWRF disaggregates the traditional forest model (Breiman, 2001) into multiple local models to highlighting important local variations and enables studies on the spatial heterogeneity of the data (Nduwayezu et al., 2023). Random Grid Search (RGS) is used to find the optimal hyperparameter, and the optimal bandwidth is determined by minimizing the Out of Bag (OOB) error and the mean-squared error (RMSE) to check and rank the endogenous explanatory factors underlying the net migration (Schutte et al., 2021). The relevance of each variable describes the associated increase in RMSE. Generally, a high relevance denotes the variable's explanatory power in the model, while a low relevance indicates that the variable weakens the model's prediction power. Specifically, the relative importance (RI) indicates the rank of how well a given factor predicts net migration in relation to the highest one (RI=1). To compare models' results and assess their complementary to depict the spatial variation of the relationship among the selected variables (Fotheringham, Brunsdon, and Charlton, 2002), we use the Akaike Information Criterion (AIC) (Akaike, 1974) and the goodness-of-fit measures (R2).

PRELIMINARY RESULTS

Descriptive results are presented by Figure 1. Three main migration patterns are detected from the analysis of 5-year net migration in Ghana from 1985 to 2014: i) urban pathways in the south coastal areas; ii) persistence of the north-south regional differential over time; iii) changes in migration dynamics in the Western and Volta regions over 2000-2004, 2005-2009 and 2010-2014 intervals.

INTERPLAYS BETWEEN THE CHANGES IN NET MIGRATION AND ENVIRONMENTAL FACTORS

When the changes in net migration and environmental variables are modelled, results from GWR depict no significant association in 1990-1994 and 1995-1999 periods. We detail results for subsequent periods. In 2000-2004, the mean of annual maximum temperatures is significantly associated with the changes in net migration, but coefficients vary across territories. We map the significant coefficients of co-variants over the period 2000-2004 (Figure 2) to appreciate the variability across Ghanaian territories. In 2005-2009, both GWR and MGWR models confirm the dominance of the small company gold mining effects in the interplays with environmental, climate conditions, conflict fatalities and net migration. Nevertheless, results from the two models differ when examining the spatial variability. Based on the MGWR model results, in the Upper East region, the small company gold mining effects are negatively related to the change of net migration, whereas in the Western and Central regions the effects are positive, while the GWR model extends the area of incidence to the Greater Accra and Eastern coastal areas. By examining the latest period 2010-2014, both GWR and MGWR models identify a significant correlation between the length of consecutive wet days with the changes in net migration. As final step, we apply the GWRF, using Gaussian model fitting and fixed kernel bandwidth estimations to run robustness checks of results derived from GWR and MGWR. The GWRF is here used for modelling the intervals 1990-1999 and 2000-2014, separately. As a result, the GWRF estimated model fits for 1990-1999 have very small explanatory power, which is in line with the outcomes from the GWR and MGWR models. Consequently, GWRF models are trained with data for the 5-year intervals 2000-2004, 2005-2009 and 2010-2014. Over the intervals, 2010-2014 reports the highest R-square value and lowest optimized bandwidth, which indicates that variables show more localized importance within the model.

EXPECTED FINDINGS

From a comparative perspective, the MGWR outperforms the other models in terms of R-square and AICs values. In 2000-2004, the model records the highest R-square values, implying that the model has a good explanatory power, followed by the period 2010-2014 (R-spare values are 71% and 68% respectively).

TABLE 1
Summary of models' indicators, 2000-2014

Time Interval	Geographic Weighted Regression **GWR**		Multiscale Geographic Weighted Regression **MGWR**		Geographically Weighted Random Forest Regression **GWRF**	
	R-square	AIC	R-square	AIC	R-square	AIC
2000 - 2004	0.056	1085	**0.718**	870	0.138	1829
2005 - 2009	0.152	1068	0.432	1006	0.124	2230
2010 - 2014	0.492	998	**0.677**	815	0.200	1503

From a methodological perspective, both GWR and MGWR provide standardization of values, which could not be replicated in the GWRF to predict and estimate values of the explanatory variables. Nevertheless, using GWRF, the relative importance of variables can be established to complement the GWR and MGWR measures of their local significance. All three methods contribute to gauging the relevance of selected contextual variable within their respective models. Maximum temperature is consistently showed to be relevant throughout all models for the interval 2000-2004. The MGWR is the only model in 2000-2004 that envisages a local significance of the length of consecutive wet days. The artisanal gold mining showed relevance in all models for 2005-2009. The length of consecutive wet days shows significance in both local and global regression models but no importance in the estimates from the GWRF. Precipitations are significance in the global geographic weighted regression but not enough power to be significant in the local MGWR model nor the GWRF. None of the models predict the length of consecutive dry days, minimum temperature, and conflict variables as relevant. The here proposed three-model strategy allows us to deeply examine the links between migration and contextual factors, mapping, developing and testing a new analytical approach to explain the spatial interplay among migration drivers. Findings would contribute to refining the conceptualisations of environmental and climate conditions as drivers of migration and population changes across territories, giving insights for rethinking mobility in the frame of a more complex, nuanced processes where environmental factors are related to both mobility and immobility population behaviours.

Figure 1

Net migration in Ghana by 5-year interval, 1985-2014

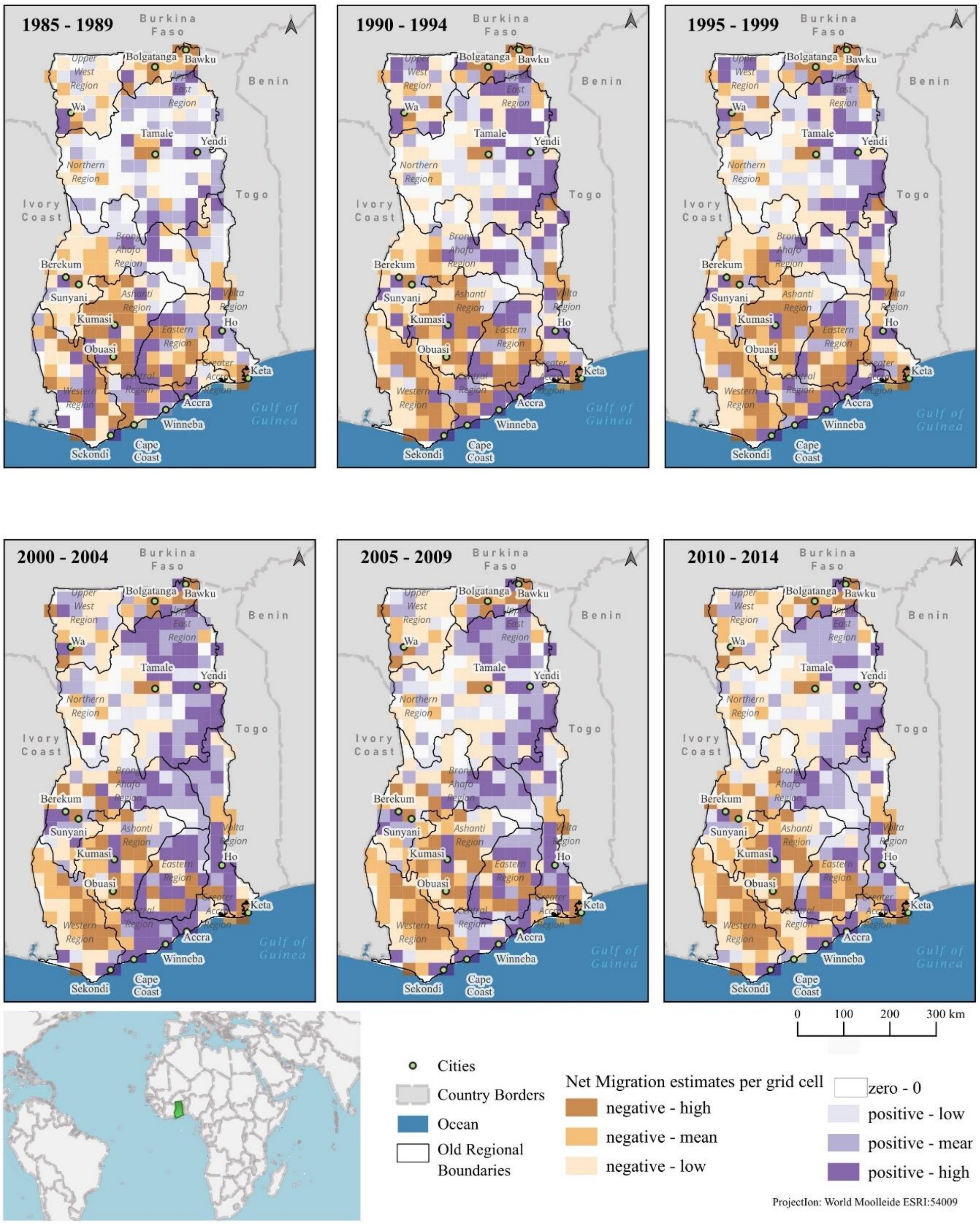

Figure 2

GWR and MGWR models' results in 2000-2004