Geospatial Analysis of Safe Delivery App Events Based on Geographically Weighted Regression Tool

AGILE: GIScience Series, 3, 47, 2022. https://doi.org/10.5194/agile-giss-3-47-2022
Proceedings of the 25th AGILE Conference on Geographic Information Science, 2022.
Editors: E. Parseliunas, A. Mansourian, P. Partsinevelos, and J. Suziedelyte-Visockiene.
This contribution underwent peer review based on a full paper submission.

Geospatial Analysis of Safe Delivery App Events Based on Geographically Weighted Regression Tool

Masoume Mahboubi[1], Astrid Grønbæk[2], Ana Fernandes[3], Casper Fibæk[3,4], Paolo Dabove[1]

[1] Department of Environmental Engineering (DIATI) - Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Turin, Italy.
[2] Maternity Foundation, Forbindelsesvej 3, 2. floor 2100, Copenhagen Ø, Denmark.
[3] NIRAS Consulting, Sortemosevej 19, 3450 Allerød, Denmark.
[4] Aalborg University, Department of Planning, A.C. Meyers Vænge 15, 2450 Copenhagen SV, Denmark.

Correspondence: Masoume Mahboubi (masoume.mahboubi@studenti.polito.it)

Abstract. This study explores the spatial relationship between the number of recorded events of the Safe Delivery App, which is a mobile learning tool to train midwives in developing countries, during three months, and several independent variables, including the number of health facilities, pregnancies, number of women of childbearing age, number of infants aged 0-1 years, mobile network coverage data and total population density. The study aims to identify and analyse the reach of the Safe Delivery App at the district level in Ghana country, considering the correlation between dependent and independent variables. The geospatial analysis of App usage events layered with several related explanatory variables is based on Geographically Weighted Regression. The explanatory variables were able to predict and explain the number of events with of accuracy score of 90 % at the district level. The results have provided valuable insights into the further roll-out of the App and helped to highlight the districts that need further support to roll out the Safe Delivery App considering the analysed independent variables.

Keywords. Geographically Weighted Regression, Safe Delivery app, Ghana, Spatial Relationship

1 Introduction

Geographic Information System (GIS) and geospatial analysis based on GIS tools play an essential role in the provision of healthcare issues. Decision support systems developed by GIS analysis and integration of spatial, geographic, and health-related data can significantly assist policymakers in improving service coverage and performance (McLafferty, 2003). GIS tools enable users to efficiently discover the geographical reach of healthcare service, its coverage, and adequacy, and determine whether allocating new resources is necessary for certain locations (Wang, 2020; Avtar et al., 2020). The availability of a wide range of spatial data strengthens the relevance of data science and geospatial analysis in achieving Sustainable Development Goals (SDGs) (Nabiyeva and Wheeler, 2020). Furthermore, several studies have deployed GIS tools to perform analyses of the health of pregnant women (Muhammad et al., 2021; Sotomayor-Beltran et al., 2018). The Safe Delivery App (SDA) is a mobile Learning application for smartphones developed by the Maternity Foundation (MF), the University of Copenhagen, and the University of Southern Denmark. The SDA provides skilled birth attendants with direct and instant access to evidence-based and up-to-date clinical guidelines on Basic Emergency Obstetric and Neonatal Care. The SDA provides life-saving information and guidance through easy-to-understand animated instruction videos, action cards, and drug lists. It can serve as a training tool both in pre- and in-service training and equips healthcare workers even in remote areas with a clinical reference tool (Safe Delivery App, 2022). The SDA was first introduced to Ghanaian midwives in a study in 2014 and a scientific study was conducted exploring the effect on health outcomes (Klokkenga et al., 2019). The Ghanaian SDA was formally launched in

Accra, Ghana in December 2017. Two main approaches were deployed to introduce the SDA to Ghanaian birth attendants. Promotion of self-use through broader dissemination and Facilitated use. Since its launch, MF has been involved in efforts to introduce the SDA in training delivered to at least 400 healthcare workers and trainers in Ghana. To explore the reach of the Safe Delivery App, we are interested in the intensity of usage in different districts in Ghana and whether district-level usage is correlated with higher birth rates, the number of health clinics, and other relevant factors that we expect to be correlated with the number of events in the SDA. The analysis works under the hypothesis that the more births a midwife faces, the more likely she is to use the SDA. If this assumption is correct, GWR analysis can help identify areas of low usage compared to the expected levels based on the number of births, health facilities, and other relevant factors. If such areas exist, targeted efforts to support training and roll out of the SDA can be conducted to ensure midwives in these areas have access to clinical guidelines and instructional videos. Fig. 1 shows the interface of the SDA. The geospatial analysis of app usage events layered with population data alongside other datasets provides insights into the further roll-out and helps highlight if certain districts with high rates of estimated pregnant populations need further support. To determine the spatial relationship between independent and dependent variables, a global Ordinary Least Square (OLS) and then a Geographically Weighted Regression (GWR) regression tools, are considered (Wu D., 2020). The goal of this study is to probe the correlation and spatial relationship between a dependent variable and independent variables; to analyse if there is a spatial relationship between the number of events of SDA and some other related variables and if variations in the number of events can be explained by the independent variables or not.

Figure 1. The interface of the Safe Delivery App

2 Methods

In this project, two regression methods have been implemented. First, a global OLS regression is implemented on the variables, and then in order to detect the spatial autocorrelation condition in the results of OLS residuals, a Moran‘s I index is calculated, to see if the hypothesis of randomness in data is valid or it can be rejected (Comber, et al., 2022).

GWR is a methodology for spatial analysis. This tool is a regression tool that takes the heterogeneity in location into account, meaning that it calibrates the regression model in each location and produces a local model instead of one global model (OLS). In this study, in both regression models, the dependent variable is the number of events (count of events) recorded by the SDA in June - August 2021 in Ghana. The independent variables are extracted from different sources.

2.1 Data and Software Availability

The study area is the country of Ghana, and independent and dependent variables and their sources are summarized in Tab. 1. Administrative boundaries represent the district level. Population density with resolutions of 10 and 100 meters is based on the results of a previously performed study (Fibaek, 2022). Data of women of childbearing age consisting of ages 20-40 with a resolution of 100 meters is downloaded from Worldpop freely available dataset, pregnancy raster with a resolution of one kilometre is from Worldpop dataset. Infants with a resolution of 100-meter is extracted from sex and age data. Usage data are recorded with Safe Delivery App in a three-month period and can be available upon request. For processing, ArcMap 10.7.1 was used with the student license. For map visualisation, the open-source software QGIS 3.22.3 was utilised.

Table 1. Used dataset and their sources

Data	Name of Source	Source
Administrative boundaries	HDX	(Administrative Boundries, 2022)
Recorded events	Safe Delivery App	(Safe Delivery App, 2022)
Population	NIRAS A/S	(Fibaek, 2022)
Women of childbearing age	Worldpop	(Worldpop Datasets, 2022)
Pregnancy	Worldpop	(Worldpop Datasets, 2022)
Infants (0-1 years old)	Worldpop	(Worldpop Datasets, 2022)
Network (2G, 3G, 4G)	Coverage Data	(Collins Bartholomew and GSMA, 2021)

2.2 Data Preparation

The count of recorded events of the SDA in the period of June to August is the dependent (response) variable. In the processing, only events with location data are considered. The App allows for offline usage, which is not tracked and therefore excluded from the analysis. Some users decide not to have GNSS-location enabled, and for the selected period, 37 percent of the recorded events included GNSS coordinates. Population density for each district is produced based on the zonal statistics tool from the population raster (Fibaek, 2022). Women of childbearing age distribution is produced with combining age structure of 20 until 40 years old, female sex, which is combined and analysed with zonal statistics tool. Pregnancy count raster is also available, so with the zonal statistics tool, their distribution is available. To make sure that live births have been considered, infants' distribution from age structure data of 0 to 1 years old, female and male sexes have been considered. Mobile coverage data of 2G and 3G is available considering the number of users provided by the operators and tower locations, but for the 4G network, only tower locations are available; for each district, the network with a maximum number of users has been considered (Collins Bartholomew and GSMA, 2021). The number of health facilities data is from the open street map platform.
As an example of the used dataset, total population data at the district level is visualized in Fig. 2 (Fibaek, 2022).

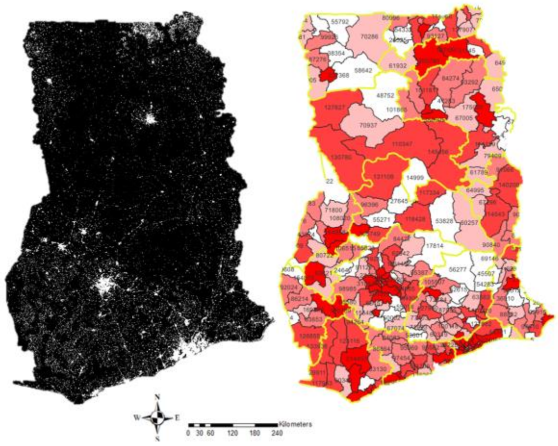

Figure 2. Left: total Population raster in Ghana, right: total population distribution at the district level

2.3 Geographically Weighted Regression

If the global OLS regression tool fails to be fitted on the whole dataset, and the pattern of clustering is present in its results, then a GWR regression can be implemented. In other words, if the spatial variation of a variable considering a given neighbourhood is important for the study and heterogeneity plays a role, GWR is a comprehensive tool (Wu and Zhang, 2021). This is based on the first law of geography, that spatial autocorrelation for closer neighbours is stronger (Li and Griffith, 2022). The theory behind GWR is to fit an equation and a model based on a dependent variable and several explanatory variables (Brunsdon et al., 1998). Then, this model is used for several purposes: analyse the correlation between dependent and explanatory variables, discover the key variables and, to predict if the actual incidents are following the predictions of models or if the actual value of the dependent variable should be more or less than predicted value by the fitted model, and how much is the predicted number of events in each district (Wu D. , 2020). There are several ways to analyse the performance of the model; R-squared (R^2) and R-squared adjusted (R^2 adjusted) are two assessors for the goodness of the model (Wu D. , 2020), and high values of R^2 and R^2 adjusted, indicate a stronger correlation and fit between variables (Yu, et al., 2020). Another way to analyse results is to inspect the standard residuals maps (Hu and Xu, 2019). GWR results can assist in illustrating how much is the over or under-estimation and where are the significant areas by the GWR model (Hu and Xu, 2019); and we can see, based on each independent variable, it is expected to have more SDA events in a given area, or there are already more events than the model expected. The distance and neighborhood to be considered in this analysis are determined by the bandwidth parameter; in other words, the scale of the relationship between response and predictors is determined by bandwidth. In this research, to avoid the overfitting of the GWR model, the optimal bandwidth parameter is based on a trial and error, a manual bandwidth parameter of 30 neighbors. In this analysis, the number of events in each district is considered as a response variable, predictor (or independent) variables include total population, infants population, women of childbearing age, network coverage data, number of health facilities, and number of pregnancies.

3 Results and Discussion

Two regression models of OLS and GWR have been produced. Tab. 2 illustrates the R^2 adjusted, and Moran's I autocorrelation results of OLS model; the OLS model achieves the poor results of 13% on the whole dataset and fails to have a strong fit on the data. Besides, Moran's I autocorrelation results show the presence of spatial autocorrelation and shows statistically significant values, and reject the null hypothesis of randomness in the OLS model.

Table 2. Numerical results of OLS model.

R^2 adjusted	P-value	Z-score	Moran's I index
0.13	0.00007	-3.8	-0.36

Since, based on Moran's I index on OLS standard residuals, the presence of dispersion is approved in the data, the GWR model is considered (Comber, et al., 2022). The bandwidth parameter is set to 30 neighbours. It is worth mentioning that GWR is a local model, so local R^2 values are available for each district. Tab. 3 shows the average results of R^2, R^2-adjusted, and Moran's I index for standard residuals of GWR. Based on the GWR model, the number of events is explained by the six independent variables with a R^2 value of 90%, which represents a high regression score. Fig. 3 represents the local R2, standard residuals, and intercept coefficient raster for the study area. In Fig. 4, the distribution of the local coefficient estimates from the GWR model and their spatially varying relationships between the number of events and each predictor variable are presented. From the results, the significant variables and significant districts can be detected; for instance, the population coefficient raster is highly significant in most of the districts, but the network coverage coefficient raster represents a random behaviour in almost all districts, except for a few of them. Moran's I calculated for the GWR standard residual is non-significant, meaning that the null hypothesis cannot be rejected here.

The standard residuals map in Fig. 3 shows the App's performance inside each region for various districts. As an example, two districts of one region can have completely different behaviour inside a given region; one of them can be positively significant meaning that it has a high number of recorded events considering the independent variables, however, the other one can be negatively significant, meaning that the model expects more events from that district. By knowing this about different districts, MF and partners can focus efforts on the blueish significant areas to attract more users and broaden the usage of the App. Like any other regression model, GWR encounters some problems. For instance, if we introduce correlated variables or in the presence of local collinearity, it faces design problems and fails (Wheeler and Tiefelsdorf, 2005), but still is one of the best models for regression analysis.

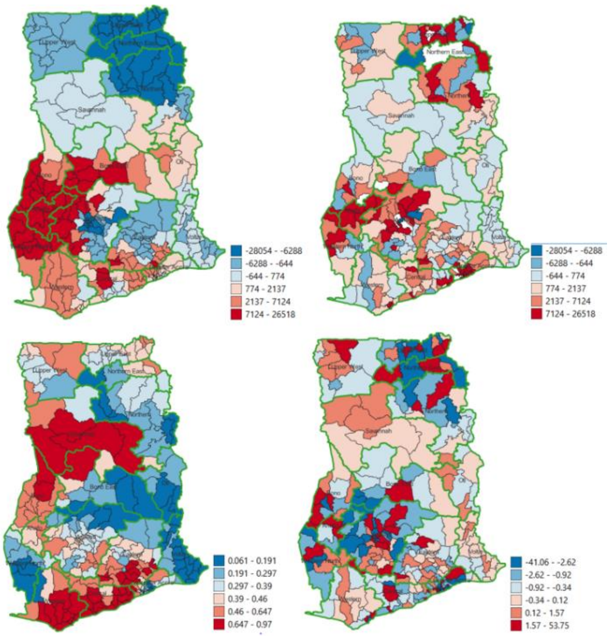

Figure 3. Top left: intercept coefficient raster, top right: prediction map, bottom left: local R^2, bottom right: standard residuals.

4 Conclusions

This study explores a regression tool to analyse the spatial relationship between the number of events recorded by the SDA in three months and independent variables of a number of health facilities, pregnancy, women of childbearing age, infants, mobile network coverage, and total population at the district level in Ghana. Because of the presence of dispersion in the results of the OLS model, the GWR model is considered as final regression tool, based on which, explanatory variables predicted and explained the number of events with a degree of accuracy of about 90%, and the coefficient rasters for six explanatory variables and intercept is presented to detect the significant variables and districts. Furthermore, standard residual maps assisted in discovering districts with an immense level of usage based on their needs and detecting districts that need to promote their usage considering independent variables. The results provided a valuable overview into the further roll-out of the app and helped to detect the districts that need further support for using the SDA. Future studies can focus on expanding the model by integrating a wider period of app usage and providing a generalized overview of the usage of SDA.

Table 3. Numerical results of GWR model.

R^2	R^2 adjusted	P-value	Z-score	Moran's I index	AICc	Sigma
0.9	0.83	0.06	-1.83	-0.05	121	2770

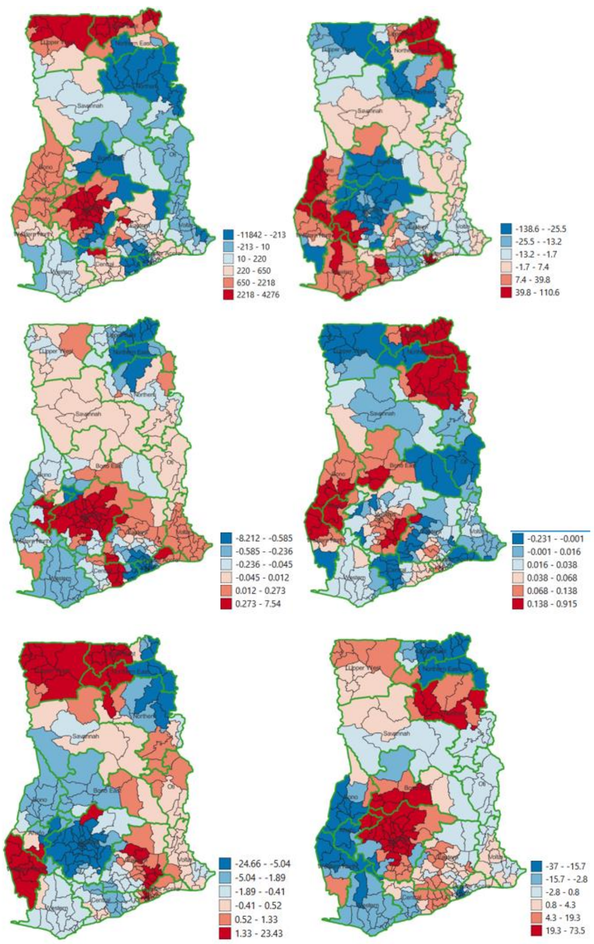

Figure 4. Local coefficient rasters of Top left: number of health facilities, top right: infants, middle left: network coverage, middle right: population, bottom left: pregnancy, bottom right: Women of childbearing age.

References

Retrieved from Safe Delivery App: https://www.maternity.dk/safe-delivery-app/ last access: 2022, March 31.

Retrieved from Administrative Boundries: https://data.humdata.org/dataset/cod-ps-gha, last access: 2022, March 31.

Retrieved from Worldpop Datasets: https://www.worldpop.org/ last access: 2022, March 31.

Avtar, R., Aggarwal, R., Kharrazi, A., Kumar, P., and Kurniawan, T. A.: Utilizing geospatial information to implement SDGs and monitor their Progress. *Environ Monit Assess 192*, No 35, https://doi.org/10.1007/s10661-019-7996-9, 2020

Brunsdon, C., Fotheringham, S., and Charlton, M.: Geographically Weighted Regression. . Journal of the Royal Statistical Society: Series D (The Statistician), 47, 431-443. https://doi.org/10.1111/1467-9884.00145, 1998.

Comber, A., Brunsdon, C., Charlton, M., Dong, G., Harris, R., Lu, B., . . . Harris, P.: A Route Map for Successful Applications of Geographically Weighted Regression. *Geogr Anal.*, https://doi.org/10.1111/gean.12316, 2022.

Fibaek, C.:Area and Population Estimates from Sentinel Data - Ghana and Mediterranean Egypt. Mendeley Data, V1, , doi: 10.17632/gf8v525tm6.1, 2022.

Hu, X., and Xu, H.:Spatial variability of urban climate in response to quantitative trait of land cover based on GWR model. . Environ Monit Assess, 191, 194, https://doi.org/10.1007/s10661-019-7343-1, 2019.

Klokkenga, C. M., Enemark, U., Adanu, R., Lund, S., Sørensen, B. L., and Attermann, J.:The effect of smartphone training of Ghanaian midwives by the Safe Delivery application on the incidence of postpartum hemorrhage: A cluster randomised controlled trial. Cogent Medicine, 6:1, DOI: 10.1080/2331205X.2019.1632016, 2019

Li, B., and Griffith, D. A.:The Moran Spectrum as a Geoinformatic Tupu: implications for the First Law of Geography. Annals of GIS, 28(1), 69-83, DOI: 10.1080/19475683.2022.2026473, 2022.

McLafferty, S. L.:GIS and Health Care. Annual Review of Public Health, 24(1), 25-42, https://doi.org/10.1146/annurev.publhealth.24.012902.141012, 2003.

Muhammad, A. D., Masni, R. N., Indarty, A., A, M. T., Mallongi, A., Basir, M., Ansyari, A.: Mapping of high-risk detection of women pregnancy on antenatal care in Talise Health Center, Palu City, Indonesia. Gaceta Sanitaria, 35(12), S152-S158, https://doi.org/10.1016/j.gaceta.2021.10.015, 2021.

Nabiyeva, G., and Wheeler, S.:Geographic Information Systems as a Tool to Support the Sustainable Development Goals. World Academy of Science, Engineering and Technology, Open Science Index 159, International Journal of Computer and Information Engineering, 14(3), 78 - 85, 2020.

Sotomayor-Beltran, C., Segura, G. W., and Tarazona, D.:Anemia During Pregnancy in Peru in 2017: A Geographic Information System Study. EEE 38th Central America and Panama Convention (CONCAPAN XXXVIII), 1-5, doi: 10.1109/CONCAPAN.2018.8596336, 2018.

Wang, F.:Why public health needs GIS: a methodological overview. Annals of GIS, 26(1), 1-12, DOI: 10.1080/19475683.2019.1702099, 2020.

Wheeler, D., and Tiefelsdorf, M.:Multicollinearity and correlation among local regression coefficients in geographically weighted regression. . J Geograph Syst 7, , 161–187, https://doi.org/10.1007/s10109-005-0155-6, 2005.

Wu, D.:Spatially and temporally varying relationships between ecological footprint and influencing factors in China's provinces Using Geographically Weighted Regression (GWR). Journal of Cleaner Production, 261, 121089,ISSN 0959-6526, https://doi.org/10.1016/j.jclepro.2020.121089, 2020.

Wu, X., and Zhang, J.:Exploration of spatial-temporal varying impacts on COVID-19 cumulative case in Texas using geographically weighted regression (GWR). Environ Sci Pollut Res, 28, 43732–43746, https://doi.org/10.1007/s11356-021-13653-8, 2021.

Yu, H., Gong, H., Chen, B., Liu, K., and Mingliang, G.: Analysis of the influence of groundwater on land subsidence in Beijing based on the geographical weighted regression (GWR) model,. Science of The Total Environment, 738, 139405, ISSN 0048-9697,https://doi.org/10.1016/j.scitotenv.2020.139405, 2020.