The Spatial Effects of Regional Poverty: Spatial Dependence, Spatial Heterogeneity and Scale Effects

Mengxiao Liu [1], Yong Ge [2,3,4], Shan Hu [5] and Haiguang Hao [1,*]

[1] Chinese Research Academy of Environmental Sciences, Beijing 100012, China
[2] State Key Laboratory of Resources and Environmental Information Systems, Institute of Geographic Sciences and Natural Resources Research, Chinese Academy of Sciences, Beijing 100101, China
[3] College of Resources and Environment, University of Academy of Sciences, Beijing 100049, China
[4] Jiangsu Center for Collaborative Innovation in Geographical Information Resource Development and Application, Nanjing 210023, China
[5] Department of Urban Informatics, School of Architecture and Urban Planning, Shenzhen University, Shenzhen 518060, China
* Correspondence: haohg@craes.org.cn

Supplementary information (SI)

1. Explanatory variables

A total of 21 variables that may affect or be related to poverty were assembled and summarized at the county and township-level and 19 at village level. These variables were divided into the first,second and third nature geography factors, and derived using existing publicly available data. The variables were chosen based on an extensive literature review and expert knowledge of poverty in Hubei province. Appendix Table 1 in the main text provides a summary of the assembled data.

On the whole, the index system is composed of seven first-class indicators: physical conditions, natural resources, geolocation and transportation, governance capability of local government, informatization level, public service, and human resources.

(1) Physical conditions. Composed of five second-class indicators: elevation, precipitation,

temperature, and natural disasters. The elevation in Hubei Province has a large spatial differentiation, and the poverty distribution roughly shows a trend of increasing with the rise of the elevation, The climate conditions: annual average temperature and annual cumulative precipitation could reflect the climate conditions of regional agricultural production and development. Natural disasters are represented by the proportion of the population affected by natural disasters within the county. The location conditions are mainly characterized by geographical location. The location conditions of each administrative region are replaced by the distance from each administrative region (represented by the location of its government residence) to the higher-level administrative region. The connection between cities is often carried out through road networks, rather than the actual straight distance between two points. Therefore, this study uses distance based on road networks to measure the distance between counties and their central cities.

(2) Natural resources. Composed of four second-class indicators: cropland resources, mineral resources and farmland production potential. Per capita cropland was selected to reflect the richness of cropland resources. Considering the importance of agriculture income for rural households, ground slope is also adopted to characterize agriculture conditions. When the slope is less than 15 °, it is more suitable for crop cultivation. Therefore, the ratio of cultivated land area with a slope above 15 ° to the total cropland is selected to characterize the planting conditions. Mineral resources are characterized by the potential value of per capita mineral resources in each region.

(4) Geolocation and transportation were represented by geographical location road construction, train station construction, The geographical conditions of each administrative region are replaced by the distance from each administrative region (replaced by the location of its government residence) to the upper administrative region. The connection between cities is often through the road network, rather than the actual straight-line distance between two points. Therefore, this study uses the distance based on the

road network to measure the distance between each county and its central city. Roads and train stations are the most commonly used indicators to reflect regional transportation construction. We used road network density to characterize road development, and the distance to the nearest train station to characterize the construction of regional train stations; The proportion of internet access households to the total population was calculated to represent the of internet development;

(5) Governance capability of local government is represented by per capita local financial revenue and the investment attracted by county government. The local fiscal revenue is expressed using per capita indicators.

(6) Public service was constituted with health, education, and social security. The health development is mainly measured by number of registered doctors per 10000 people. Education is measured by the level of teaching staff and reflected by the number of registered teachers per 10000 people. The social security development is characterized by the proportion of New Rural Co-operative Medical System participants.

(7) Human resource was represented by the scale of rural labor force,the employment index of rural labor force , the Proportion of aged 60 or above and educational attainment of people. The scale of rural labor force is reflected by the proportion of rural working population. This indicator not only characterizes the scale of rural labor force, but also indirectly reflects the proportion of the population due to illness, disability, and aging in the region. The employment situation of rural labor force is characterized by the proportion of migrant workers. The educational attainment of people was calculated by average educational attainment of people

At the township-level, in order to ensure the comparability of poverty factors at different scales, the selection of indicators for poverty factors at the township level should be consistent with that at the county level. Due to the scale effect of poverty factors themselves, the elements on the two scales cannot be completely consistent. Therefore, considering the data accessibility and feature representativeness, some adjustments have been made. There are 5 indicators same as that at the county scale, namely, per capita local financial revenue, the investment attracted, the number of beds per 10000 people, the number of registered teachers per 10000 people and proportion of the population having access to the internet.

At the village level, many of these variables could not be obtained directly. Therefore, we developed approximate variables derived from ancillary vector and raster datasets, and assembled and summarized these at the village level.

1.1 The first nature geography variables

The first nature geography factors that affect poverty are physical conditions and natural resources. Natural conditions include basic living conditions, such as topography, climate conditions, ecological conditions and geographical location, while natural resources refer to inorganic and organic natural capital, such as land, forests, water, mine resource and tourism resources.

Topography

Topographic features, including elevation and slope, were obtained from the Google Earth Engine (GEE) platform, which originated from the Advanced Land Observing Satellite (ALOS) Global Digital Surface Model. GEE is a cloud-based platform for analyzing publicly available, remotely sensed planet-scale imagery and other environmental data (Gorelick et al., 2017). ALOS World 3D - 30m is a global digital surface model (DSM) dataset with a horizontal resolution of approximately 30-meter (1 arcsec. latitude and longitude) (Tadono et al., 2014). Slope was calculated with the "ee.Terrain.slope" function in GEE, which computed local gradients using the four neighbors connected to each pixel. The administrative boundaries of the region were imported into GEE to calculate the average elevation and slope.

Climate

In this paper, we adopted temperature and precipitation to indicate climate conditions. The meteorological data called the "China surface climate data daily data set (V3.0)" were provided by China National Meteorological Network, which contains the daily data of air pressure, air temperature, precipitation, evaporation, relative humidity, wind direction, wind speed, sunshine hours and 0cm ground temperature elements of 699 basic meteorological stations in China from May 2014 to December 2019 （http://data.cma.cn/). The original data is daily data, and the annual average temperature and annual cumulative precipitation data of each station are obtained using R script, and the annual average temperature and annual precipitation are interpolated by using the collaborative Kriging method in ArcGIS 10.6 software combined with the station altitude data.

Land cover

The county and town-ship level land cover data were derived from the annual 100m global land cover data of Copernicus global land cover layers: cgls-lc100 collection (Buchhorn et al., 2020). The service provides a series of global biogeophysical products for the current situation and changes of the land surface. It is updated every year with 2015 as the base year. The data set contains 14-15 layers of data, mainly including land use data and various land use proportion data. Land use classification data includes 10 categories (bare / sparse vegetation, cultivated land, grassland, moss / lichen, shrub, forest, permanent snow / glacier, urban construction land, permanent water area and seasonal water area) and 23 sub categories. In this study, the land use proportion data are used to calculate the per capita cultivated land area, the proportion of cropland with a slope greater than 15^o, forest coverage.

While the village-level land cover data were obtained from the second National Land Survey, which was updated by the local government in 2015. These data are based on interpretation of high-resolution remotely sensed images and were validated by field surveys organized by the local government. They are the most accurate and detailed land use data for Yunyang. At the first level there are 13 land cover categories. These are cropland, garden, forest, grassland, built-up land, water, land for business, land for industry, land for public service, land for residents, land for special use (e.g., military), land for transportation and other lands. The first level classes are subdivided into 61 subclasses at the second level. For example, cropland is subdivided into paddy field and dry land. In our study, only some first levels which may be related to poverty were chosen: cropland, garden, woodland, grassland, and built-up land. The proportion of different land cover types in each village was extracted. Considering the important role of cropland in agricultural production, we also calculated the proportion of cropland with a slope greater than 15°, to reflect the quality of cropland in each village; slopes greater than 15° have limited suitability for crop plantation (Liu et al., 2015; Liu and Li, 2017). Using ArcGIS 10.6, cropland vector maps were converted to raster maps with the same resolution and boundary as the slope data exported from GEE (i.e., 30 m).

Natural disasters

The geological disaster data were downloaded from the geocloud platform managed by the Geological Survey of China, which includes only landslide information. Geological disasters were recorded with their GPS locations. The impact of geological disasters on each village was characterized as the proportion of households living within 1 km of a geological disaster.

1.2 The second nature geography variables

Geographical location

The geographical conditions of each administrative region are replaced by the distance from each administrative region (replaced by the location of its government residence) to the upper administrative region. The connection between cities is often through the road network, rather than the actual straight-line distance between two points. Therefore, this study uses the distance based on the road network to measure the distance between each county and its central city.

Transportation development

Road density is a relevant indicator of transportation conditions. It was calculated by dividing the area of each village by the sum of the road lengths within the village. The road network data of counties are obtained from the national 1:250000 road data released by the National Bureau of surveying, mapping and geographic information. The road network data was released in 2018, but it actually reflects the road network in 2015. The road network includes national roads, provincial roads, county roads and township roads. In Yunyang, the local government provided road data at a scale of 1:1,200,000 as well as detailed information about rural roads information from 2016.

Railway stations development is another most commonly used indicator to reflect regional traffic construction. The train station data comes from 12306 website (https://www.12306.cn/index/).The distribution data of high-speed rail stations are from the 2016 national high-speed rail station distribution data of Harvard University world map (https://dataverse.harvard.edu/dataverse/chgis).

The proportion of Internet access households in the total population is used to characterize the regional internet development, and was collected from Annual assessment report on county economic work in Hubei Province No data was available at village-level in Yunyang county.

1.3 The third nature geography variables

Public services

It has been widely suggested that access to infrastructures and public services are important contributors to poverty, as better access can lead to more efficient agricultural production, better market opportunities, diversification of rural economies and improvement of living conditions(Gibson and Rozelle, 2003; Pozzi et al., 2009;Watmough et al., 2016). We consider the public service from health, education and social security. The health service is mainly measured by the regional medical level, which is characterized by the number of health technicians per 10000 people and the number of beds. Education is measured by the level of teachers and reflected by the number of teachers per 10000 people. Social security undertakings are characterized by the coverage of regional social insurance. At county-level, these indicators were obtained from the public dataset and their source are presented in Appendix Table.1. While at village-level, there is almost no publicly released annual socio-economic statistical data. Therefore, a list of accessibility indicators representing public services based on remote sensing data products and network big data were developed.

Accessibility can be simply defined as the shortest distance from any demand point in a study area to the closest facility (Weiss et al., 2018). However, Euclidean distances are inappropriate in this regard because they do not consider road quality and topographic conditions. Hence, the cost-distance algorithm which is a widely used methodology to calculate accessibility had been adopted (Juliao, 1999; Nelson, 2000; Longley et al., 2005; Pozzi and Robinson, 2008; Pozzi et al., 2009; Watmough et al., 2016). The cost-distance algorithm estimates the time required to travel from each grid to the destination along the least costly route. Travel time was calculated over a surface with friction, taking into account the road network and land cover. Roads and land cover were classified based on their estimated travel speeds.

Access to different facilities was obtained based on point of interest (POI) data collected from the open application programming interface (API) provided by a Chinese map company, Gaode. Each POI was recorded along with its class and location. Based on the classification system, we extracted industries (factories, enterprises, companies, industrial zones), schools (kindergartens, primary schools, secondary schools), medical care centers (hospitals, clinics, health centers), banks and markets.

To avoid edge effects, a 10 km buffer beyond the administrative boundary was implemented before calculating accessibility. Roads were divided into provincial, county, and rural roads, converted to a raster with 30 m spatial resolution, and integrated with the land cover map. According to the National Technical Standards for Highway Engineering (JTG B01-2014) in China, the traffic speed is 60 km/h on provincial roads, 35 km/h on county roads, and 15 km/h on rural roads. The average speeds for different land use types (Table S1) were based on previous studies. It was assumed that residents travel on roads with motor vehicles and travel on foot through other types of land cover. Water was classified as a travel barrier, but it can be passed with a boat or bridge. Thus, we included extra time to account for crossing water.

Accessibility surfaces were generated using the cost-distance function in ArcMap 10.6 (ESRI, 2018) to estimate the time (in minutes) required to travel to different infrastructures. The least costly path was calculated using the cost-distance function for each cell in the analysis window to the target that will be the least costly to reach, based on iterative allocation (Pozzi and Robinson, 2008). Then an estimate of the time required to reach the most accessible services was assigned to each grid cell. POI data were set as targets and accessibility surfaces were generated using the cost-distance function in ArcMap 10.6. Friction surfaces based on land cover and road data were created according to Formula 1, and the final cost-distance model was built using Formula 2.

$$trcost = [merge(border, roads, land\ cover] \tag{1}$$

$$Access = costdistance(< source\ grid >, < cost\ grid >) \tag{2}$$

The spatial distributions of different services and their accessibility maps are presented in Fig. 2 of the main text. Finally, mean accessibility data were calculated on a rural settlements vector map, extracted from land cover data, to represent the accessibility of different public services in each village.

Table S1. Transportation speed and time cost to cross a 30-m grid cell.

Cover/road type	Estimated passing speed (km/h)	Time cost 30 m grid cells (s)
Province road	60	1.8
County road	35	3.06
Rural road	15	7.2
Built-up	5	21.6
Bare land	5	36
Woodland	3	27
Grassland	4	27
Cropland	4	108
Water	1	1.8

Hunan Resources

Some demographic variables were used to reflect human resources in a region, including proportion of labor force in the total population and proportion of migrant labor force in the total population. At county level, the two factors could be obtained from public dataset, the dataset and their source are presented in Appendix Table.1. At the village level, they were derived from statistical data provided by the Yunyang government and were available only for the year 2017.

References

1. Buchhorn, M., Lesiv, M., Tsendbazar, N. E., Herold, M., Bertels, L., & Smets, B. (2020). Copernicus global land cover layers—collection 2. Remote Sensing, 12(6), 1044.
2. Environmental Systems Research Institute (ESRI). (2018). ArcGIS Help Documentation
3. Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., Moore, R., (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. Remote Sens. Environ., 202, 18–27
4. Gibson, J., Rozelle, S. (2003). Poverty and access to roads in Papua New Guinea. Econ. Dev. Cult. Change., 52(1), 159-185
5. Juliao, R.P. (1999). Measuring accessibility using GIS. in Proceedings of GeoComputation '99. Available at http://www.geovista.psu.edu/sites/geocomp99/. Accessed 20 May 2020
6. Liu, Q., An, J., Wang, L., Wu, Y., Zhang, H.J.S., Research, T., (2015). Influence of ridge height, row grade, and field slope on soil erosion in contour ridging systems under seepage conditions. Soil Tillage Res., 147, 50–59
7. Liu, Y., Li, J., (2017). Geographic detection and optimizing decision of the differentiation mechanism of rural poverty in China. Acta Geogr Sin., 72(1), 161-173
8. Longley, P.A., Goodchild, M.F., Maguire, D.J. and Rhind, D.W. (2005). Geographic Information Systems and Science (second edition). Chichester: Wiley. pp. 517
9. Nelson, A. (2000). Accessibility, transport and travel time information. CIAT Hillsides Project. International Center for Tropical Agriculture (CIAT), Cali
10. Pozzi, F. , T. Robinson, (2008). Accessibility mapping in the Horn of Africa: Applications for livestock policy. IGAD LPI working paper No.11-08. Available at http://www.igad-

lpi.org/publication/working_papers.htm. Accessed 20 May 2020

11. Pozzi, F., Robinson, T., Nelson, A. (2009). Accessibility mapping and rural poverty in the horn of Africa. : Applications for Livestock Policy. Working Paper No 11–08. 2008, IGAD LPI, 35

12. Tadono, T., Ishida, H., Oda, F., Naito, S., Minakawa, K., Iwamoto, H.J., (2014). Precise global DEM generation by ALOS PRISM. ISPRS Annals. 2(4), 71

13. Watmough, G.R., Atkinson, P.M., Saikia, A., Hutton, C.W., （2016）. Understanding the evidence base for poverty–environment relationships using remotely sensed satellite data: an example from Assam, India. World Dev. 78, 188–203. https://doi.org/10.1016/j.worlddev.2015.10.031

14. Weiss, D.J., Nelson, A., Gibson, H., Temperley, W., Peedell, S., Lieber, A., Fullman, N.J.N., (2018). A global map of travel time to cities to assess inequalities in accessibility in 2015.Nature., 553(7688), 333.