The use of GIS in employment impact assessments

Strengthen2: Employment impact assessment to maximise job creation in Africa

Author / Alina Game

Funded by the European Union

ISBN: 978-92-2-0351987 (web PDF)

Non-technical summary

STRENGTHEN2, a joint initiative of the European Union and the International Labour Organization (ILO) is focussing on maximising job creation through investments in sub-Saharan Africa (SSA). Part of the project is exploring the use of Geographic Information Systems (GIS) in employment impact assessments (EmpIA), to enable the measurement of secondary and long-term employment effects of infrastructure investments across a range of sectors. This note explores the potential of using GIS data and methods to assess the employment outcomes of selected infrastructure projects, tailored towards SSA. Investments in infrastructure are important for employment outcomes, either through the direct creation of jobs or secondary employment from increased opportunities that the infrastructure provide for the local population. For example, access to new or improved roads provide increased access to labour markets, which could allow people to reach more or better job opportunities.

GIS provides a framework for the collection, storage, processing and analysis of location data. GIS data are highly accessible and offer a valuable solution for observing potential impacts at scale, anywhere in the world. Combining GIS data with microdata that includes both a geographic element and employment information in economic models, offers the ability to understand how infrastructure impacts employment outcomes. This note focuses on three types of infrastructure investment, transport (roads), energy (electricity) and environment (water), to understand how GIS can be applied to measure the access to these types of infrastructure and the subsequent impacts this could have on employment for the affected population. GIS data and satellite imagery are highly applicable for the above three types of infrastructure investment. For transport infrastructure, spatial data and GIS methods can be used to measure the changes in access to labour or commodity markets for the population, providing insight into how new or improved roads can open up access to more and better jobs. Data sources including satellite imagery, can be compared over time to measure changes in agricultural productivity from water projects and products such as night-time lights (NTL) can be utilised to measure access to electricity from energy investments.

Using GIS in employment impact assessments focuses on examining the secondary and long-term employment effects and is dependent on available microdata containing employment information. Data availability at specific time periods of interest, before and after the infrastructure investment is required to understand the change in employment outcomes over time. Although there can be difficulty in obtaining such data, there is a range of survey datasets providing relevant information, including labour force and household surveys in SSA. Additional GIS data sources like NTL data can also be explored further for measuring economic and potential employment changes over time. Econometrics and statistical modelling techniques can be applied to identify the causal relationship between the GIS measures of infrastructure, such as access markets, electricity or impacts from irrigation and employment information from microdata. This can provide an estimate of the impact that access to infrastructure has had on employment outcomes, such as the number of jobs and the composition and quality of employment.

A lack of microdata containing relevant employment information at a suitable scale does not need to limit the applicability using GIS to try to understand employment impact from infrastructure investments. GIS can be used to combine a range of additional data layers, which can help provide quick insights into factors that could influence employment, for example, environmental data related to climate, or population composition, to understand who is most impacted by infrastructure investments and where they are located. Combining these multiple data sources should be tested to identify methods that can be operationalised for employment impact assessments and replicated for similar investments across SSA. GIS can also be valuable for the current or future planning of infrastructure project placement, which could enabled the prioritisation of areas to be served by infrastructure and plan the optimal location for where they should be placed. This makes GIS a flexible tool for monitoring, planning and analysing the potential employment impacts of investments in infrastructure.

▶ Contents

1. Introduction

GIS provides a framework for the collection, storage, processing and analysis of spatial data. Operationalising the use of GIS for employment impact assessments is relatively new; however, there has been some research into applying geospatial data for measuring the impacts of infrastructure projects. GIS holds great potential for measuring the employment impacts of infrastructure projects, due to the availability and accessibility of geospatial data. This is particularly relevant in low and middle-income countries where access to data and ground data collection can be more difficult. Spatial data sources, such as satellite imagery provide global availability at high temporal and spatial resolutions which makes them a valuable data source for investigating the impacts of projects at scale, anywhere in the world.

GIS can provide a method for measuring the secondary and long-term employment effects of infrastructure investments, investigating the impacts ex-post once a project is complete and operational. Integrating GIS measures with additional economic data, such as that from census or labour force surveys (LFS) that contains an employment dimension, can be used to gain insights into how infrastructure impacts employment. Combining data and applying econometric modelling techniques can provide estimates on the causal effects of employment outcomes. Utilising this information before and after investment in infrastructure can provide an evaluation of the impact of such investments. This can be important to not only look at increases in employment, but changes in the quality and composition of employment through structural transformation, with shifts to more skilled occupations and higher value sectors. This is particularly relevant in SSA, where investments may not result purely in the creation of new jobs but rather influence changes in the quality of employment, highlighting the importance of investigating job quality when assessing the impact of such projects.

There are several types of infrastructure investment where GIS could be applicable to EmpIA. The specific GIS data and methods applied are dictated by the infrastructure type that is being investigated. For the sake of this note, transport, energy and environment infrastructure are focussed on; however, similar methods could also be applied to a wider range of infrastructure that contain a spatial component, such as social infrastructure. Investments in transport infrastructure can provide improved access to labour markets which can positively impact development outcomes by allowing households to reach new and better employment opportunities. Access to other types of infrastructure, such as energy or social infrastructure can also impact labour market outcomes and alter the composition and quality of employment for the population. Investments in water and irrigation can positively impact agriculture, improving employment outcomes and livelihoods in this sector.

This note discusses the methodology for utilising GIS data and workflows for assessing the employment impacts of infrastructure investments, with a focus on SSA. Due to the spatial nature of data, the infrastructure projects that will be focussed on include roads, electricity and water, though the methods presented could be applicable to other developments. Section 2 provides a literature review on the previous work looking at the impacts of infrastructure on economic activity and employment. Section 3 focuses on the key research questions that GIS could be used to answer and further opportunities for incorporating GIS in employment impact assessments, based on the previous research reviewed. Section 4 presents the relevant data, GIS processing, analysis, econometrics that can be applied, along with additional options for operationalising GIS in EmpIA. Section 5 concludes the note.

▶ 2. Literature Review

This section provides a review of previous studies investigating the impacts of infrastructure investment on economic and employment outcomes. It focusses on the methods and data used to measure access to such infrastructure and opportunities for the incorporation of additional geospatial data. For the most relevant studies investigating the relationship between infrastructure and employment that are listed below, the main findings, data sources and methods used are further summarised in the table in Appendix 1.

2.1 Investment in transport infrastructure and market access

Transport infrastructure and economic impacts

Combining measures of access, calculated using GIS methods or taken from existing household surveys, with econometric models has previously been conducted to investigate the impacts of transportation infrastructure projects. Improved transport infrastructure can increase access to labour or commodity markets which can have a positive impact on economic development and employment for the surrounding population. Banerjee et al. (2020) investigated the economic impact of access to transportation networks in China over a 20-year period at the county level. They produced a hypothetical network of straight-line distance between Treaty Ports , which were ports opened to foreign trade and historical cities, using the distance from the centre of a county to one of these 'straight-lines' as the measure of access to transportation infrastructure. They applied regression techniques to investigate the impact that transportation infrastructure had on per capita GDP, manufacturing firms locations and profit and household income. Including province and fixed year effects, they discovered that proximity to transportation lines had a moderate positive causal effect on per capita GDP across sectors. Distance to line had no effect on per capita GDP growth or household income; however, they did discover that inequality is higher in better connected areas. The model used argues that relatively low mobility of capital is likely to be associated with higher income inequality. This assumes that the direction of capital movement occurs from less connected areas to better connected areas, which could have resulted in the higher income inequality closer to transportation lines.

Bird and Straub (2020) studied similar variables to look at the impact of the expansion of road networks in Brazil on the growth and allocation of economic activity including population, GDP and GDP per capita. They constructed transport lines that coincided with the countries' new radial highway system and calculated the distance to these constructed lines to instrument for the change in access to state capitals. They discovered that exogenous proximity[1] to roads increased the concentration of economic activity in terms of population and GDP growth around large urban areas due to better access to main centres, where average effects were strongest in urban centres within 200km of state capitals. Elasticities for GDP were 2.8 at less than 100km from a road, 2.4 between 100 and 200km and 1.1 between 200 and 500km and were statistically significant. They note that these results are further driven by the characteristics of the endpoint city, where cities with increased amenities such as access to water and schooling and a higher share of non-agricultural activities in GDP, produced stronger economic agglomeration effects. This highlights that increased GDP and population growth do not only depend on improved transport access through proximity to roads, but also on the characteristics of the location that road is leading to.

Research by Faber (2014) investigated the effects of transport infrastructure investment in the form of new highways in China. The study applied GIS techniques to produce hypothetical least cost spanning tree networks[2] that connected major Chinese cities. Using an instrumental variables (IV) strategy, it looked at the relationship between network connections and industrial gross value added, non-agricultural gross value added, local government revenue and GDP, where IV is an approach used to try and account for endogeneity or reverse causality. Results were statistically significant and showed that a connection to the transport infrastructure led to a reduction in local GDP growth in peripheral counties that were between targeted metropolitan centres relative to non-connected peripheral counties. This is driven by a reduction in industrial output growth. This indicated that such infrastructure projects do not diffuse economic activity into space as

[1] An exogenous measure of access was produced by constructing transport lines based on historical preconditions. The distance to these constructed lines was then calculated using buffer zones which produced the exogenous proximity to roads.

[2] 'Least cost spanning tree networks' or 'minimum spanning trees' are methods to connect points via a set of lines or edges, using the minimum distance and shortest route possible.

may be expected, from metropolitan regions to the periphery; but instead, reduce industrial and total output growth within connected peripheral regions compared to those of non-connected regions.

In the absence of relevant microdata that contains economic or employment information, the use of satellite data such as NTL have been explored as a proxy of economic activity and compared to infrastructure projects to look at the effects over time. NTL data offer satellite images of human light emissions at night, at regular intervals over a significant time period. This makes them a valuable data source for identifying changes over time, where they have previously been explored as a proxy for population and economic activity in the form of gross domestic product (GDP). Bluhm et al. (2020) sought to measure the impact of Chinese investment in Africa over the period from 2000 to 2014, by measuring the spatial distribution of economic activity. They used locations of investment projects and NTL data to produce a spatial Gini coefficient in an attempt to identify shifts in economic activity. Results showed that investment projects reduced the concentration of economic activity within regions, which increased for completed projects and had the largest effects in poor regions that are most in need of infrastructure financing. NTL data offers an alternative, readily accessible source of data that can be used to identify changes in economic activity over time; however, further work needs to be done to investigate the applicability of this dataset for identifying changes in economic activity and converting this to measuring changes in employment.

Transport infrastructure and employment impacts

Further research and evaluations of projects have also looked at employment impacts including the composition, dimensions and shifts in sectoral employment as a result of investment in infrastructure. Baum-Snow (2013) investigated the relationships between urban highway construction and the decentralisation of jobs between 1960 and 2000 in the U.S. The study utilised census data on counts of workers and working residents by industry and applied economic models to estimate the effects of highway construction on the decentralization of central city jobs. The dependent variable of focus was the change in log of employment by industry, which he measured against the endogenous variable of interest which was the change in the number of highways between 1950 and 2000 that serve a primary cities' central business district, instrumented with the number of planned radial highways constructed prior to 2000. He found that manufacturing was the only industry that had a statistically significant response to highway construction, where each new highway caused 10 per cent of manufacturing jobs and working residents to leave a metropolitan area. Overall, discovering that the addition of one new highway caused 16 per cent of central city working residents to move to suburbs, where new highways allow for population decentralization.

The above studies concentrate on investigating impacts of transport infrastructure investment on the economic and employment impacts in urban areas. Further research has also gone into investigating the impacts of these investments in rural areas, particularly focussed on the impacts geared towards the agricultural sector. In rural Indonesia, Yamauchi (2016) investigated the impact of road quality on labour supply and wages, using two survey rounds from 2007 and 2010. The study used a fixed effects IV estimation to investigate the relationship between changes in village-level transport speed and monthly/daily incomes and working time, split by agricultural and non-agricultural employment. The number of road projects completed between 2007 and 2010 was used as an instrument for change in average transportation speed. The study found a positive effect between an increase in transport speed from road projects and an increase in wages in both agricultural and non-agricultural employment. An increase in speed was also associated with a decline in working time in agriculture, for household members who were more educated.

Asher and Novosad (2016) studied the impact of rural road investment in India on employment and economic outcomes. Access to rural roads was measured using a dummy variable representing if the village received a new road as part of the investment. They utilised economic, population and poverty census data to investigate the effects of access on several dependent variables including employment in agriculture, household earnings, income and included NTL data to investigate the effects on NTL luminosity. They included a number of controls at the village-level, including indicators for amenities (schools, medical centres, and electrification), literacy rate and distance from closest town along with district cut off fixed effects. They discovered that access to a new road for previously unconnected villages' leads to a reallocation of labour out of agriculture and into wage labour market participation, with a 10 per cent reduction in the share of household workers in agriculture and an 8 per cent increase in household earnings. These effects were most significant in villages that were closer to city centres where people can take advantage of commuting and short-term migration. When comparing to NTL data, they identified that luminosity grew faster following road construction compared to beforehand. They conclude that new road construction from investment facilitates the access of rural labour to external employment, rather than resulting in the growth of the nonfarm sector in rural areas.

Alternative measures of access

When investigating transport access, most of the research above applies straight-line distance, road length, or presence of a road project to a particular administrative area to define accessibility to transportation infrastructure. A limitation of this approach is that it does not consider information such as road quality, road type, topography or other physical obstacles that can influence accessibility. Much research also concentrates on measuring access to the infrastructure itself, rather than investigating the change in access the infrastructure creates to a labour market. To incorporate these, additional GIS data and analysis can be utilised to produce measures such as travel time that provide a more accurate depiction of access to markets. Previous work has investigated this, where Jacoby (2000) applied travel time obtained from household survey data to investigate household benefits from road investments in rural Nepal. The log of median travel time to a market centre was used as the independent variable of interest and regression techniques applied to investigate the relationship between this, land plot values and agricultural wages. The study identified a weakly significant relationship between wages in agriculture and market access defined by travel time. The OLS estimates show that wages declined as travel time increased, where a 10 per cent increase in travel time reduced agricultural wages by 0.5 per cent.

In SSA, Storeygard (2016) investigated transport costs and distance between a city and major port in 15 coastal countries on the role of city income. Data on road conditions (paved/unpaved) were incorporated when calculating distance using the shortest path along a country's road network and NTL data was used as a proxy for city GDP. To get a measure of transportation costs, network distance was inter-acted with oil prices. Results showed that a larger distance between sample cities and port cities coincided with lower GDP and an increase in transport costs. Cities with more paved roads were not so affected by changes in transport costs, whereas cities with less paved roads were more affected by transport costs to secondary cities compared to primary cities. When looking at the relationship between road condition and employed population working in manufacturing and controlling for distance to primary city, the study discovered that in cities where roads are more paved, there is a positive correlation with a larger fraction of adults working in the manufacturing sector in a subset of countries with available data (Ghana, Guinea, Senegal and Tanzania). Including this additional information is relevant when looking at the impacts from investments in transport infrastructure, as road quality and condition of existing roads is something that can be improved through such investments.

2.2 Access to electricity and employment impacts

Alternative measures of access

Access to electrification can increase economic development and result in shifts in sectoral employment, where electrification can provide new employment opportunities and increased productivity, for example by freeing up time previously spent on household activities such as cooking. Modelling these changes in electricity access and the subsequent economic and employment impacts can be difficult as investment in electricity infrastructure projects are often endogenous in their placement due to a range of economic and political factors. In rural South Africa, Dinkelman (2011) used an IV approach to attempt to produce unbiased estimates of the impact of electrification on the employment growth rate. Land gradient (slope) was utilised to instrument for project placement, as a higher gradient increases the cost of connection which makes it a factor when considering electricity project placement. This was complemented with a fixed effects strategy to investigate the relationship between a dummy variable of if a community received an electrification project and the dependent variables of male/female employment, hours of work, log hourly wage and log monthly earnings. Poverty rate, female-headed household and adult sex ratio were controlled for, where data on the indicators were obtained from household survey data and were combined with administrative and spatial data on electricity investments. Both methods discovered that employment increased in areas of rural electrification, with an increase in female employment of 9.5 percentage points within a 5-year period from the IV results, which translates to 15,000 more women participating in the labour force out of the baseline population. The study also discovered an increase in hours of work (for the average change in electrification rate, this fell between 1.3 and 1.9 hours more work per week) and a reduction in female wages and increase in male earnings. They suggest that this could be due to household electrification releasing female time from home to market work as well as enabling increased small-scale labour demand via self-employment and micro enterprises.

Similarly, Tagliapietra et al. (2020) sought to understand the effects of electricity on labour markets in Nigeria. The study also applied an IV approach to address endogeneity, exploiting a series of instruments including household distance to the grid, distance to the nearest power plant and land gradient of the location site. They applied probit, biprobit and propensity score matching to investigate the relationship between the dependent variables of the proportion of people of working age (split by male/female, agriculture/non-agricultural and urban/rural) and access to electricity. Access to electricity was measured using a dummy variable obtained from household survey data. They discovered that access to electricity led to a 7 per cent decrease in the proportion of household members employed in agriculture. Non-agricultural employment increased by 15 per cent, with stronger effects of electricity access for male employment. Access to electricity led to an overall 8 per cent increase in employment, with a reallocation of labour away from agricultural and towards non-agricultural work, resulting in an increase in labour market participation.

NTL as a proxy for electrification

The latter studies (Dinkelman 2011; Tagliapietra et al. 2020) utilise household survey data to provide a measure of access to electricity. However relevant, up-to-date household survey data can often be unavailable in developing countries. The use of satellite imagery such as NTL has been shown to correlate with electrification and due to its high spatial and temporal availability, provides a method for measuring electricity access in any geographic location over large time periods. Research has investigated the use of NTL data as a proxy for electrification, where the following studies present examples of this, however are reviewed in less depth as they do not go on to look at the economic or employment effects from electrification.

When investigating the relationship between NTL and electrification, Min et al. (2013) examined the relationship between NTL data from the Defence Meteorological Satellite Program (DMSP) and village-level survey data on access to electrification in Senegal and Mali. They found high correlations between the NTL images and electrified villages and statistically significant results that electrified villages were consistently brighter than unelectrified villages in the NTL images. Results show that electrified villages appeared brighter due to the installation of streetlights, whereas the correlation between NTL and individual household electricity use was lower. This indicates that NTL are useful for identifying areas of electrification, however the brightness measures are less valuable for looking at changes in levels of household electricity use. In Vietnam, Min and Gaba (2014) also showed a similar correlation when looking at validating NTL data against survey data, discovering that brightness of electrified villages was both a function of the number of electrified homes and streetlights. They provide evidence that even electrified rural villages with no public street lighting can be observed from NTL imagery, highlighting the potential of using NTL data to measure electrification projects with or without streetlight illumination.

Incorporating NTL data as a proxy for access to electrification and comparing this to impacts on employment has been less researched, the most in-depth investigation on this has come from Burlig and Preonas (2016) when identifying the effects of a rural electrification programme in India. They applied a regression continuity strategy using an eligibility cut-off based on a population threshold that deter-mined whether a village was eligible for electrification. They investigated the relationship between changes in NTL brightness (where an increase in brightness indicates an increase in electrification) on employment, asset ownership, household income and education. They discovered a substantial increase in electricity use, however found that electrification did not yield significant changes in any of the outcomes. The research presents a methodology for investigating the impacts of electrification utilising NTL data at a national scale, where they emphasise the need for further work in this area.

2.3 Environmental infrastructure and employment impacts

Environmental infrastructure, such as water supply projects are another infrastructure type with the potential for being assessed using GIS and remote sensing techniques. Water projects often aim to increase water supply or irrigation for agriculture, which could positively impact agricultural productivity, in turn affecting agricultural employment. Much research has utilised remotely sensed satellite data for measuring the impacts of irrigation, but few have focussed on the subsequent impacts these measures could have on employment.

Using remote sensing to calculate spectral indices relating to irrigation has previously been investigated in SSA, however without the link to employment. Previous work includes the use of satellite imagery to measure irrigated areas and the change in these over time. For example, Gumma et al. (2011) applied Landsat and MODIS satellite imagery to map the change in irrigated areas in Ghana. They calculated the Normalised Difference Vegetation Index (NDVI)[3] which provides an indicator of vegetation health and cover, to identify irrigated and non-irrigated areas, in combination with land use/

[3] The NDVI provides an indication of vegetation health or 'greenness' which can provide insight into changes in vegetation quality, density and coverage. It is calculated using the spectral reflectance measured from satellites, where healthier vegetation reflects more near-infrared (NIR) light.

land cover raster[4] data. Understanding the change in size of irrigated area could be a useful indicator to compare to employment effects. Some research takes this a step further and utilises remote sensing for estimating agricultural productivity and crop yields, which could correlate with employment effects such as profits, earnings or hours of work. Wahab et al. (2018) applied remote sensing methods through the use of high-resolution imagery from unmanned aerial vehicles (UAVs) to measure maize crops in Ghana. Also applying the NDVI for green vegetation, they were able to estimate crop vigour and yields of maize and when comparing to in-field measurements, found a good performance of NDVI. This demonstrates the applicability for utilising high resolution imagery through the use of UAV's, or potentially commercial satellite options to measure crop yields which could be used to assess the impacts of irrigation schemes.

These variables, along with others related to soil, water and vegetation productivity could be combined with microdata on employment to try to understand the impact that water investments have on employment outcomes. This link with employment has been less researched, however along similar lines in Tanzania, Bhargava et al. (2018) investigated the impacts of environmental degradation using an indicator of soil health and examined the effects on agricultural profits. They utilised geo-located survey data on rural households from the World Bank Living Standards Measurement Study - Integrated Surveys on Agriculture (LSMS-ISA) and compared this to soil organic carbon content (SOC), which was obtained from a combination of ground-truthed[5] and remotely sensed satellite data. The key dependent economic variable of interest was plot-level agricultural profits. Using a household fixed effects econometric model and controlling for household characteristics, results showed an increase in agricultural returns of approximately 7-8 dollars per additional kilogram of SOC. This indicated that higher soil health increases agricultural profits and highlights the potential for matching high resolution remotely sensed soil data with household surveys at the national level, which could be replicated further for other remote sensing indices obtained from satellite imagery.

[4] Raster data are spatial datasets that are made up of a set of regularly spaced pixels or 'grid cells'. The size of the pixels determine the dataset's spatial resolution. Examples of raster data include land cover, elevation or satellite imagery.

[5] Ground-truth data can be defined as data collected from the field or location by direct measurement.

▶ 3. Research Questions

Previous work has highlighted the applicability of using GIS to measure the employment impacts of infrastructure investments that contain a spatial component. The remaining part of this note will focus on the methodological process required to assess the employment impacts for transport, energy and water projects. Conducting in-depth analysis of infrastructure projects will often be done in collaboration with national stakeholders to inform the basis of the study and questions of focus. Based on the previous literature, the following set of questions can begin to be answered. Combining this information with data availability and GIS methods, a set of research questions can be formulated and are explored further throughout the note.

- How can GIS be used to measure access to markets?

Accessibility can be inhibited by a range of physical access or other social, economic or cultural factors. GIS provides an efficient way to analyse physical accessibility, with the ability to combine multiple datasets to measure access via distance, speed or time across space. Previous research has measured access in terms of straight-line distance, where further work should be done on incorporating travel time, road condition data and additional geospatial datasets, such as elevation, land-cover and physical barriers such as rivers that could impede access. If information on road type and condition before and after investment are available, this data can also be fed into accessibility analysis which will provide a more realistic measure of access when investigating the impact of transport infrastructure investments. Variables such as distance to road, road speed, distance to markets and travel time to markets are examples that could be explored. Travel time measurements can cover a variety of transport modes, including car, motorcycle, walking or cycling based on average transport speeds. If data is available, public transportation such as buses can also be explored by incorporating stop locations, walking time to stops and routes. Access for both labour and commodity markets could be investigated, where labour markets can be taken as the centroid of built-up or populated areas, where the majority of employment opportunities may be based. Commodity markets can be defined as a specific market location where commodities are bought or sold.

- To what extent can NTL data be used as a proxy for electrification?

When looking at the impacts of electrification on employment, there has been less research done at larger scales, due to constraints on limited availability of household survey data that provides measure of access to electricity. GIS data on locations of electricity transmission lines or power plants can also provide an estimate of access to electricity, but this data is not always available and even if the population is within a certain distance to the transmission network, it does not necessarily guarantee access to electricity. To overcome this need for additional survey data with information on electricity access or electricity infrastructure data, NTL can provide a potential way to measure access to electricity and to examine the changes in access over time. NTL are available at high temporal resolution, globally and have been shown to be a good proxy for electrification. Further work could be done on incorporating NTL data with modelling techniques described above, such as IV estimation, to look at the impacts of electrification on employment in countries that may not have rich availability of household survey data. To test the relationship between NTL and electrification, microdata (e.g. census) that contains information on access to electricity could be used to investigate the relationship. Comparing NTL over two census periods (10 years) where there is known data on access to electricity, could provide an indication of how NTL can be used as a proxy for electrification. Once this relationship has been established, NTL could then be used as an indicator for access to electricity to investigate specific time periods before/ after energy infrastructure investments and help understand subsequent impacts on employment by using microdata such as LFS.

- How can satellite imagery be used to measure agricultural impacts of water projects?

There has been much research into utilising satellite imagery and remote sensing techniques to measure irrigation and the impacts of water supply on soil and vegetation. Satellite imagery can provide a means for assessing large areas without the need for survey. However, assessments could be further supplemented by higher resolution satellite imagery or the use of UAV's to monitor the impacts of irrigation in greater detail. Less work has then combined these indicators within econometric models to try to understand the potential impacts on employment. Further work should be done on utilising microdata related to employment, particularly those related to agricultural employment and variables such as profits or working hours which could be impacted by crop productivity from irrigation.

- How can NTL data be applied to measure economic activity?

When examining the economic impacts of infrastructure projects, NTL data has been widely applied as a proxy for economic activity in the form of GDP. Examining changes in NTL brightness over time can provide a multiplier to examine changes in GDP, on the basis that an increase in brightness corresponds with an increase in economic activity. NTL can be used to identify changes in specific areas, but also to examine the changes in the patterns of spatial allocation of economic activity in new locations, for example the dispersion of economic activity due to a new road being built. This can provide an indication of the economic impacts of infrastructure investments, without the need for specific microdata of a suitable time period. Although work has been done on establishing the relationship between NTL brightness and GDP, further research needs to look into the potential link that this has with employment outcomes.

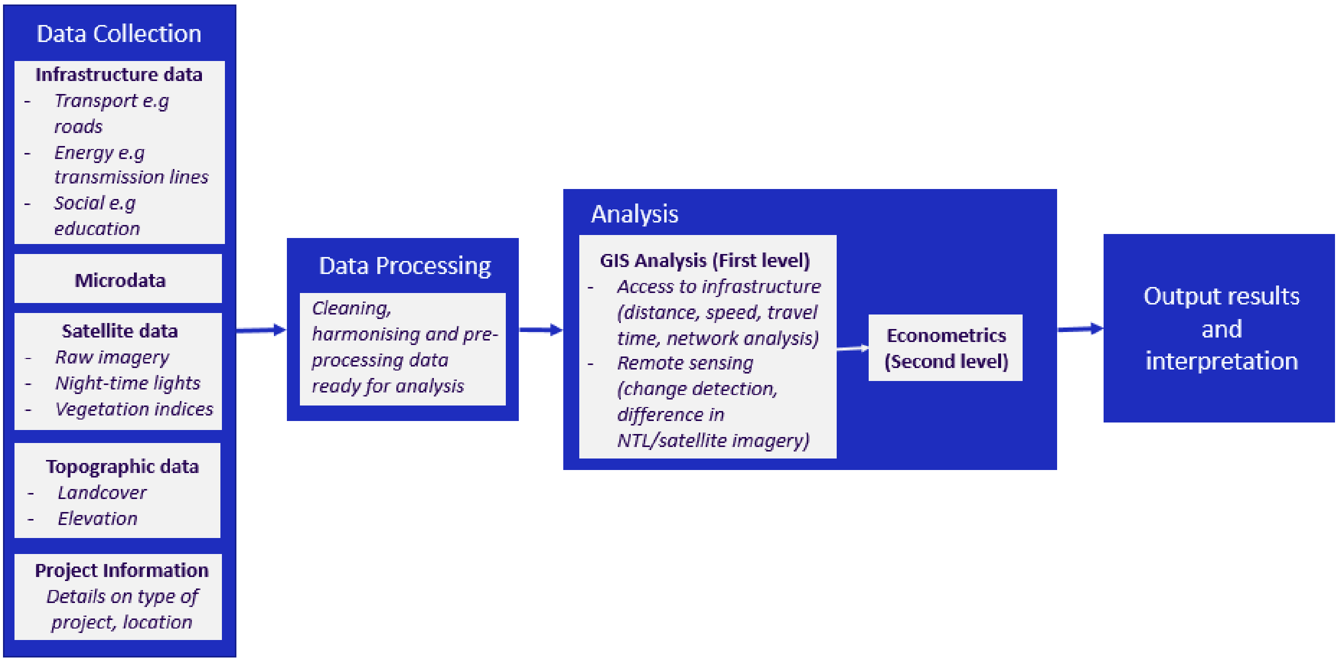
These initial learnings and further opportunities observed from the literature lead onto the following research questions to be investigated using GIS and employment data in order to understand the employment impacts related to infrastructure investments. Employment outcomes to be explored include not only looking at the number of jobs but also those related to the quality and composition of employment. The potential methodologies to explore these questions are presented in the below sections. Research questions include but are not limited to:

1. **What impact does access to labour or commodity markets from new or improved transport infrastructure have on employment outcomes?**
2. **How does access to electricity, measured using NTL, impact employment outcomes?**
3. **How do agricultural effects from irrigation, measured using remote sensing, impact agricultural productivity and employment?**
4. **In the absence of suitable microdata, how can NTL be used as a proxy for economic activity and what link does this have with employment outcomes?**

▶ 4. Methodology

This section of the note discusses the different geospatial datasets, processing and analysis required for assessing the employment impacts of infrastructure investments. Figure 1 presents a proposed workflow for using GIS to assess employment impacts of infrastructure investment, from obtaining input data from a variety of sources, to processing and analysing the data. The flowchart provides examples that could be further expanded or altered depending on the project type being assessed.

▶ **Figure 1. Potential workflow for using GIS for analysing the employment impacts of infrastructure investments**



Source: Own research.

The rest of the section on methodology focuses on four main areas:

1. Relevant GIS datasets and microdata that could be utilised for assessing employment impacts.
2. Processing of geospatial data and the creation of indicators that denote access to infrastructure, including market accessibility, night-time light brightness or environmental measures related to water projects.
3. Combining the GIS indicators with microdata and applying relevant econometric modelling techniques to identify the impact that each indicator has on employment outcomes.
4. Opportunities for operationalising GIS in EmpIA.

4.1 Data

In order to assess the employment outcomes of infrastructure investments, a range of spatial data and microdata including a geographic component is required. Specific data types will vary depending on the type of infrastructure project being assessed. Datasets can be split into the main categories of a) GIS data, consisting of a variety of raster and vector[6] datasets and b) microdata, which comes in the form of surveys but will often also contain spatial information. The majority

[6] Vector data are a type of spatial dataset that are comprised of points, lines or polygons. Examples of vector data include roads, points or co-ordinates of interest and administrative units.

of data required should be accessible and open-source; however, additional proprietary data such as higher-resolution satellite imagery could also be obtained to enhance certain analyses.

GIS data

Table 1 shows an example of some geospatial datasets that could be valuable for analysing the employment impacts of infrastructure investments. It lists some of the most popular sources for globally available datasets; however, there are a wide range of datasets available that should be explored on a country-by-country basis.

▶ Table 1. Descriptions of potential geospatial datasets and sources that may be valuable in employment impact assessments in SSA

Dataset	Description	Type	Example Sources	Year	Resolution
Infrastructure	A range of infrastructure data is available including roads, towns, schools, health facilities, energy, water	Vector	OpenStreetMap (OSM)	2021	N/A
Night-time lights	Night-time light brightness	Raster	Payne Institute	1992-2013 (DMSP) 2012-Present (VIIRS)	~2.7km (DMSP) ~750m (VIIRS)
Satellite imagery	Raw satellite images	Raster	Landsat, Sentinel, MODIS	Up to 2021	30m, 10m, 250m dependant on source
Land cover	Gridded land cover classification	Raster	European Space Agency (ESA)	2015	~100m
Elevation	Elevation in metres	Raster	NASA Shuttle Radar Topography Mission (SRTM)	N/A	~90m
Population	Estimated population count per grid cell	Raster	WorldPop	Up to 2020	~100m

Source: Own research.

Infrastructure data: Infrastructure data including transport, social, energy and environment data will be available in vector format as lines, points and polygons. For assessing the impacts of transport infrastructure, digitized roads will provide the basis for analysing market access. Points of interest, such as locations of infrastructure projects, urban centres and other amenities like education or healthcare can also be utilised. Available polygon data includes delineations of urban/rural areas, administrative boundaries or other areas of interest. Infrastructure data is freely available from a range of sources, such as OpenStreetMap (OSM) which provides data on roads, schools, water, health facilities, settlements and more. Such data is updated regularly, although coverage can often be better in urban areas, particularly the case for low and middle-income countries. Service providers, such as energy companies also provide geolocated data on infrastructure, such as electricity transmission lines. Data availability will be dependent on the country of focus and for specific projects, manual digitization of features may need to be undertaken, or further data sought from relevant national stakeholders.

Night-time lights: NTL data provide information on human emissions of lights at night at a global scale. The first NTL data were collected between 1992 – 2013 by the Defence Meteorological Satellite Program Operational Line Scanner (DMSP/OLS) at a spatial resolution of ~2.7km. The DMSP became unavailable after 2013 and a new global NTL dataset from the Visible Infrared Imaging Radiometer Suite (VIIRS) is currently available from 2012 to present day at a spatial resolution of ~750 metres. VIIRS NTL data are available in monthly or annual composites for the entire globe. NTL data is freely available up to a certain time period, where data of a higher temporal resolution data can be purchased. NTL datasets are valuable as a proxy for electrification, human settlements and economic activity. In the absence of relevant microdata, NTL could be used to measure changes in economic activity over time and further explored for translating this into employment.

Satellite imagery: Raw satellite imagery can provide a means for observing changes in infrastructure or the effects of infrastructure over time. Satellite imagery is a valuable data source as it is available at high spatial and temporal resolutions globally. There are a number of freely available sources of satellite imagery, the U.S Geological Survey (USGS) provides access to range of sources such as Landsat, MODIS and Sentinel. Sentinel-2 offers the highest freely available spatial resolution at approximately 10 metres, but satellite imagery of a finer detail can be obtained from commercial sources. Satellite imagery can be used for the digitization of features, such as new transport infrastructure like roads or points of interest. Satellite imagery is also valuable for assessing environmental projects, as it can be utilised to calculate a range of indices relating to climate, vegetation and water. Sources such as MODIS already provide a range of pre-calculated vegetation and climatic indices that are freely available for download and use in analysis.

Land cover and elevation data: Global raster datasets providing topographic information can be a useful input in accessibility analysis, particularly for calculations of travel time as they impact travel speed and direction that can inhibit accessibility. Land cover data are available globally in raster format, where each grid cell is classified based on the most prominent land cover class. An example of a global land cover dataset from the European Space Agency (ESA) available at approximately 100m spatial resolution. For elevation, digital elevation model (DEM) data is available for most of the world from a number of sources such as the Shuttle Radar Topography Mission (SRTM) at a spatial resolution of approximately 90m. Overlaying these topographic datasets with transport infrastructure data such as roads can be used to simulate a range of transport scenarios to further understand access.

Population data: Population data containing information on the number of people, disaggregated by age/sex is a useful input when analysing accessibility, to understand the total number and composition of the population impacted. Knowing where people are can provide valuable insights into the population with access to a road, electricity or other infrastructure. Census data provides the most globally accepted and accurate source of population statistics, although it has limited temporal availability (occurring every ten years), where this time period can often be longer for countries in SSA. Alternative geospatial datasets containing population estimates are available from a variety of sources, which are produced using a range of geostatistical and modelling techniques to predict the number of people in raster format (e.g the number of people in each 100m grid square). Popular sources of raster population data include that of WorldPop, who produce approximately 100m x 100m gridded estimates of population data globally, including age and sex breakdowns in their datasets.

Microdata

To provide the information to understand the impact that infrastructure has on employment outcomes, microdata containing an employment dimension will be required. Often, geographic information is collected during the process of conducting a household survey, such as an LFS or census. For surveys conducted more recently, GPS coordinates of households are often obtained in the field and if exact coordinates are not collected, sampling units or enumeration areas[7] used to design the survey are often geo-referenced. Although specific locations of households are not made publicly available due the need to protect the confidentiality of participants, data can be aggregated up to administrative boundaries or anonymised to a certain degree to be utilised in analysis. Household surveys that do provide geospatial coordinates publicly, follow anonymization techniques such as off-setting locations by a certain distance, where the offset is usually larger for rural areas with lower population, to reduce the chances of re-identification. For use in research, working with any sensitive information at a high spatial resolution needs to adhere to guidelines and protocols set out by the data producer.

LFS data provides a valuable source of information as it is entirely related to employment and labour market outcomes. When investigating the employment impacts of infrastructure projects using GIS, the key employment indicators of focus are going to be those related to secondary or long-term effects that occur after a project is completed and is operational. Employment by occupation, sector and industry, and information such as wages, hours of work and earnings are examples of variables that may be of interest to explore how investments contribute to structural transformation through shifts to

[7] Enumeration areas or sampling units are geographic areas that are used when splitting a country or survey location into smaller areas to be surveyed.

more skilled, increased wage occupations. Comparisons across age groups and gender will also be interesting dimensions to examine. LFS are often conducted on a semi-regular basis and there are a number of SSA countries that have a regular temporal availability of LFS data, where data is collected annually or even quarterly. Although the sample size of these surveys varies from country to country, a representative sample of the population will be selected for the survey. Additional sources of microdata, such as census can also be utilised, although these are only conducted every ten years, making a geospatial component less likely on older censuses. Reliability, regularity and availability of census data can often be limited for countries in SSA.

Although there is availability of microdata across many countries in SSA, accessing the data can be difficult due to confidentiality constraints and take significant time. One option for overcoming this is using data aggregated up to enumeration areas or small administrative units. This will still provide information on employment but at a lower resolution, grouped to a set geographic area, however if these areas are small enough they can still be used in analysis to compare to infrastructure investments. If data is still unavailable, additional, readily available spatial datasets have been explored to provide information on changes in economic activity as discussed previously. The use of NTL data, has been previously explored to monitor economic changes, where an increase in NTL brightness is associated with an increase in economic activity. Although this will not give as much insight into precise numbers of changes in employment as survey data would, it can be useful to identify increases/decreases, shifts and the reallocation of economic activity across space.

4.2 Spatial Data Processing and Analysis

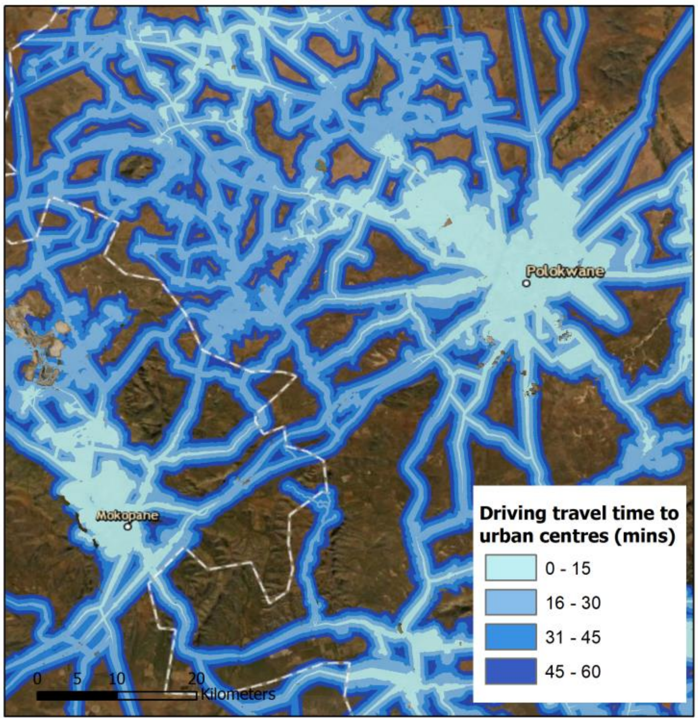
This part of the note covers the stages of processing and analysis of the relevant geospatial data for use in econometric modelling. Depending on the type of infrastructure project being assessed, such as whether the project is a road or electrification project, will depend on the GIS data processing and analysis that is undertaken. The GIS workflows can be split into categories depending on the project, where for transport infrastructure, market access will be calculated using GIS accessibility analysis methods. For energy and environment projects, satellite imagery will be processed and analysed using remote sensing techniques.

Accessibility analysis

To measure the impact of transport infrastructure projects, such as new or improved road construction, bridges, or transport corridors, accessibility analysis can be applied to calculate the time it takes or the distance required to travel to reach labour or commodity markets. In the absence of information that defines the location of labour markets, market locations are often adopted as the nearest urban centre or the closest urban centre over a certain population threshold. Much previous work has measured access as the distance of a location from the nearest road, or alternatively by denoting access if a location falls within a certain distance from a road. The prior are relatively straight forward measures of access, but with some additional information, GIS can be used to provide a more in-depth and accurate measure of accessibility. The least-cost distance path calculates the shortest route along roads to a destination and provides a more accurate measure of access compared to calculating the Euclidean (straight-line) distance from an origin to destination. This method only requires digitized road network data and should be applied in the absence of additional information.

For a more in-depth calculation of access, GIS can also be used to calculate travel time, by combining additional raster datasets including elevation and land-cover that can impede access. This topographic information can be overlaid with road data to calculate travel time to or from a location. For best results, information on the type of road and relevant road speeds for the country of interest should be collected and input in the model. Travel time can be calculated for a range of transport modes such as walking, driving or by motorcycle, which will be reflected in the relevant speeds of transport applied to each road. Public transport scenarios can also be modelled if sufficient information is available on stop locations and transport routes. The exact method adopted for measuring access should be selected based on the transport modes most used in the study area of focus, such information could come from survey data, additional data collection or in-country expertise. Although this provides a more in-depth view of access, it should be noted that this still provides only a snapshot of accessibility and does not take into account personal characteristics affecting travel or other extenuating factors such as traffic that could impact access. Figure 2 shows an example of driving travel time calculations for part of Limpopo, South Africa, to the nearest urban centre. Driving travel time was calculated using land cover, elevation and road data.

Figure 2. Driving time to urban centre in minutes, for an area in Limpopo, South Africa

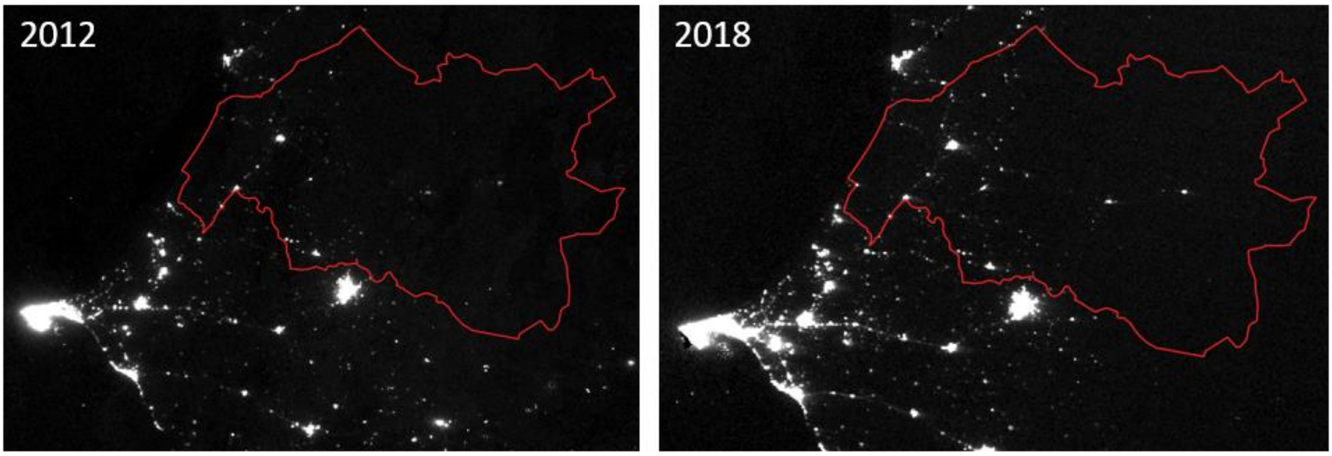


Data Source: Travel time produced using Roads, OpenStreetMap (2021), Land-cover, European Space Agency Landcover CCI (2015); Elevation, NASA Shuttle Radar Topography Mission (SRTM) (2013).

Remote sensing techniques

When investigating the impacts of energy projects, NTL data can be utilised to measure access to electrification. Depending on the time periods required, the NTL data may require harmonisation due to the change in data source from DMSP to VIIRS that occurred in 2012/2013. The more recent VIIRS NTL observations are of a higher quality compared the DMSP, with a higher spatial resolution of ~750m and improved radiometric properties (Schueler et al. 2013). To measure change in access to electricity, NTL images can be obtained before and after an electrification project to identify areas that benefitted. An increase in brightness or luminosity can denote areas that were recently electrified. This indicator of 'access to electricity' can then be further utilised in economic models to examine the impacts on employment outcomes. Figure 3 shows night-time lights over Senegal, between 2012 and 2018, which benefitted from the Senegal Rural Electrification project in the Louga region (highlighted in red), an African Development Fund project completed in 2018.

▶ **Figure 3. Night-time lights over Senegal between 2012 and 2018, with the Louga region highlighted in red.**

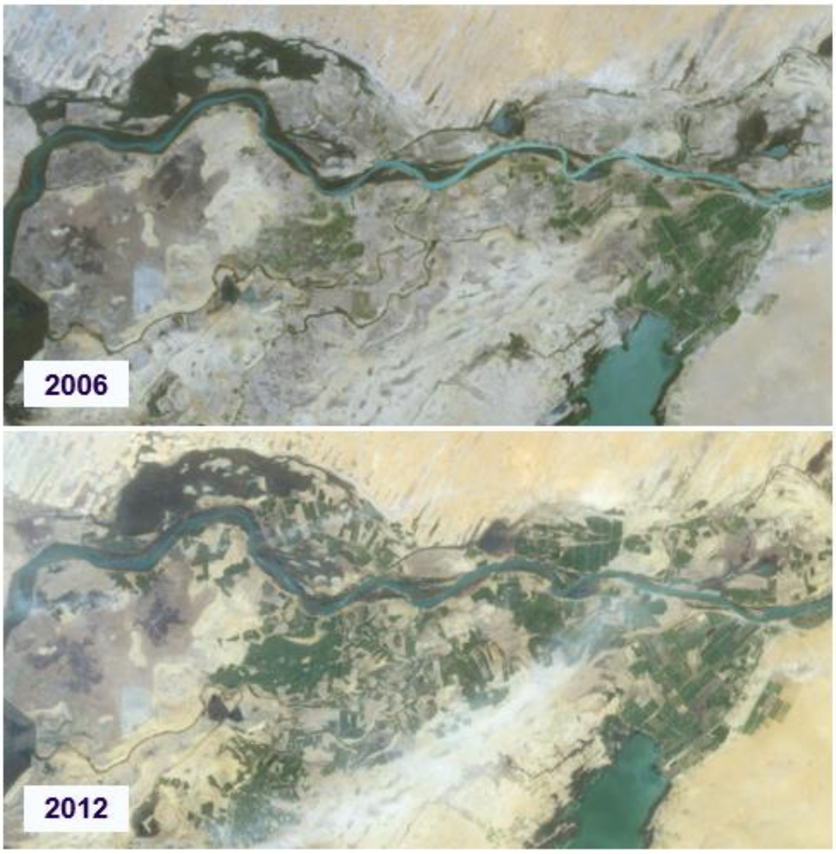


Data source: VIIRS Night-time lights, Earth Observation Group, Payne Institute for Public Policy; Global Administrative Boundaries, GADM.

Remote sensing techniques can be applied to measure the impacts of water programmes, such as irrigation schemes that could have a positive impact on agricultural employment and livelihoods. Using satellite images, a range of indices can be obtained or calculated to help understand variables that may impact agricultural productivity which could in turn impact employment outcomes. Relevant spectral indices include vegetation indices such as the NDVI, soil indicators relating to health and moisture and estimators of crop yields, which could all provide valuable indications of the effects of irrigation schemes. Comparing satellite images and the subsequent indices over different time periods, before and after irrigation can enable one to identify populations and the size of the area that has benefitted from water projects. Limitations of using satellite imagery for such analysis include cloud-cover which can inhibit the ability to pull out meaningful insights from the data, however this can be overcome to some extent due to the high temporal availability of satellite imagery. Using remote sensing to measure agricultural impacts may be less accurate compared to traditional on-the-ground survey techniques, however these can be time-consuming, expensive and difficult to implement so the use of satellite imagery to replace or enhance this methodology is valuable.

Figure 4 presents Landsat satellite imagery for a region of the Senegal River, which benefitted from the World Bank Senegal River Basin Multi-Purpose Water Resources Development Project (MWRD) between 2006 and 2012. The project involved improving irrigation for residents along the river and has led to an increase in agricultural activities, which can be seen across the satellite imagery. This provides an example of the level of change that can be detected from utilising satellite data, where measuring the change in agriculture could be combined with microdata to provide insight into how water investments impact agricultural employment.

▶ **Figure 4. Landsat 7 satellite imagery for part of the Senegal River basin between 2006 and 2012, highlighting the increase in agricultural activities (darker green areas) over time.**



Data source: Landsat-7 imagery courtesy of the U.S. Geological Survey (USGS).

4.3 Econometric Modelling

To try to understand the impacts that investments have on employment and the potential causality, econometric models can be applied. The geospatial indicators described above, such as access to markets and electricity, or agricultural impacts can be combined in models with georeferenced information on employment from relevant microdata, such as from LFS or census. Utilising a range of regression analysis techniques will allow the potential impacts that infrastructure projects have on employment to be examined.

The key dependent variables of interest that will be explored include those related to secondary and long-term employment effects. These could include variables related to income and employment, including looking at the composition of employment and changes in occupations and sectors. Other effects related to employment quality, such as earnings, wages and hours of work will be useful to explore, which will help understand if changes in employment are characterised by better jobs. Where possible, these variables should be broken down by age and gender to try to understand the groups that have been impacted the most. Independent variables of interest will be those that denote 'access' to the infrastructure investment. Based on the project that is being assessed, for transport infrastructure, the access measure will be based on access to labour or commodity market (using least-cost distance paths, speed or travel time). For electrification, access will be denoted by change in night-time light brightness and for water projects, the independent variable could be a measure of a change in agricultural productivity or irrigation. A range of control variables including household characteristics that could influence employment outcomes should also be explored.

One limitation to take into account when investigating infrastructure projects is the potential endogeneity of the infrastructure investments that are being assessed, where project placement may be endogenous due to a range of economic or political factors, or to meet an existing demand in growing places. To overcome this issue of reverse causality, an IV approach could be implemented as demonstrated in the literature. Previous work on electrification has used instruments including gradient of the project location site, distance to the electric grid and/or power plant and for roads, instruments related to historical transport routes, or physical geography provide examples of variables that could be explored and created using GIS. Accounting for this endogeneity will help enable the estimation of causal employment impacts of infrastructure projects.

4.4 Operationalising GIS in EmpIA

Considering access to relevant microdata with a spatial component can sometimes be difficult to obtain and combining this data in econometric models could be time consuming with a requirement of a certain skill level, there are intermediary steps that can be taken to utilise GIS in EmpIA. These options will not give insight into the potential causality of infrastructure investments on employment outcomes, but could be valuable to gain quick insights into new infrastructure placement and for use as a planning, decision-making and communication tool.

Mapping infrastructure investments and overlaying additional GIS datasets such as population, environment, and infrastructure relating to social, environment, and transport sectors could provide insight into the potential impacts of new infrastructure placement. Understanding the coverage of new investments can provide some idea of the number of people who will be impacted by a project, for example, calculating travel time for a new road and overlaying this with population data will give an indication of the number of people whose access has been improved, by how much and the areas they can now reach. These new locations that are now accessible could be linked to employment, such as labour or commodity markets, factory locations or other valuable infrastructure like health centres or schools. These could directly relate to employment, where having access to a labour market can open up opportunities for more or better jobs, or for other factors like access to education, could improve employment outcomes for the population in the longer-term. Employment data that is more readily available, for example from census, at lower spatial resolutions of administrative units, could also be overlaid to observe the spatial variation in different employment types, such as formal/informal or sectoral employment. By visualizing this breakdown of employment with other geospatial variables on access to infrastructure or those relating to social and environment factors, can provide additional insights on what impacts employment composition.

This gives an example of using GIS to gain quick insights from new infrastructure investments, but GIS could further be utilised as a planning tool for identifying opportunities for new projects. Understanding the existing accessibility of the population can enable the prioritization of new road investments to areas that are currently underserved. Utilising remote sensing data to look at access to electrification, can identify areas that should be prioritized for energy investments. Using satellite imagery to measure environmental impacts relating to water, soil or to study areas impacted by environmental degradation can inform where to tailor decisions relating to new environmental infrastructure projects or mitigation measures. GIS and remote sensing can offer a means to identify areas that could be prioritized, which can feed into the decision making process on where to place new infrastructure investments or the monitoring of existing infrastructure.

5. Conclusion

The application of GIS data and methods to measure the impacts of infrastructure projects has great potential for EmpIA. The high availability and accessibility of geospatial data, including satellite imagery makes it applicable to measure infrastructure in a variety of settings. GIS analysis can be applied to provide measures of access to investments in infrastructure, which can be valuable to investigate the impacts of transport and energy, such as electrification projects. The use of satellite data offers a means for measuring the effects from water and irrigation projects on agriculture and cultivation. Combining these measures with data that contains an employment dimension like from LFS or census, which also include geographic information, can be used to investigate the relationship that infrastructure investments have on employment outcomes.

Econometric modelling techniques can be applied to identify the causal relationship between the GIS measures of infrastructure, such as access to labour or commodity markets, access to electricity or impacts from irrigation and employment. Limitations around accessibility of microdata with geographic information, of a suitable time period, particularly in SSA can be a problem. However, alternative geospatial datasets such as NTL can provide a means for investigating changes in economic activity, where the potential links between this and employment should be further explored. GIS can also be used to combine and overlay a range of datasets, which could provide quick insights into current or future planning of infrastructure project placement, the composition of population and employment which could be valuable for decision-making for a variety of stakeholders. This makes GIS a flexible tool for monitoring, planning and understanding the potential employment impacts of infrastructure investments in a range of sectors. This highlights the high applicability and potential for operationalising the use of geospatial data and technologies in EmpIA in the context of SSA.

▶ References

Asher, Sam, and Paul Novosad. 2016. 'Market Access and Structural Transformation: Evidence from Rural Roads in India'. *American Economic Review* 76.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian. 2020. 'On the Road: Access to Transportation Infrastructure and Economic Growth in China'. *Journal of Development Economics* 145 (June): 102442.

Baum-Snow, Nathaniel. 2013. 'Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies', Brown University & NBER.

Bhargava, Anil K., Tor Vagen, and Anja Gassner. 2018. 'Breaking Ground: Unearthing the Potential of High-Resolution, Remote-Sensing Soil Data in Understanding Agricultural Profits and Technology Use in Sub-Saharan Africa'. *World Development* 105 (May): 352–66.

Bird, Julia, and Stéphane Straub. 2020. 'The Brasília Experiment: The Heterogeneous Impact of Road Access on Spatial Development in Brazil'. *World Development* 127 (March): 104739.

Bluhm, Richard, Axel Dreher, Andreas Fuchs, Bradley Parks, Austin Strange, and Michael J. Tierney. 2020. 'Connective Financing: Chinese Infrastructure Projects and the Diffusion of Economic Activity in Developing Countries'. CESifo Working Paper No. 8344.

Burlig, Fiona, and Louis Preonas. 2016. 'Out of the Darkness and into the Light? Development Effects of Rural Electrification'. Energy Institute at Haas Working Paper 268.

Dinkelman, Taryn. 2011. 'The Effects of Rural Electrification on Employment: New Evidence from South Africa'. *American Economic Review* 101 (7): 3078–3108.

Gumma, Murali Krishna, Prasad S. Thenkabail, Fujii Hideto, Andrew Nelson, Venkateswarlu Dheeravath, Dawuni Busia, and Arnel Rala. 2011. 'Mapping Irrigated Areas of Ghana Using Fusion of 30 m and 250 m Resolution Remote-Sensing Data'. *Remote Sensing* 3 (4): 816–35.

Jacoby, Hanan G. 2000. 'Access to Markets and the Benefits of Rural Roads'. *The Economic Journal* 110 (465): 713–37.

Faber, Benjamin. 2014. 'Trade Integration, Market Size, and Industrialization: Evidence from China's National Trunk Highway System'. *The Review of Economic Studies*, 81(3): 1046-1070.

Min, Brian, and Kwawu Gaba. 2014. 'Tracking Electrification in Vietnam Using Nighttime Lights'. *Remote Sensing* 6 (10): 9511–29.

Min, Brian, Kwawu Mensan Gaba, Ousmane Fall Sarr, and Alassane Agalassou. 2013. 'Detection of Rural Electrification in Africa Using DMSP-OLS Night Lights Imagery'. *International Journal of Remote Sensing* 34 (22): 8118–41.

Schueler, Carl F., Thomas F. Lee, and Steven D. Miller. 2013. 'VIIRS Constant Spatial-Resolution Advantages'. *International Journal of Remote Sensing* 34 (16): 5761–77.

Storeygard, Adam. 2016. 'Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa'. *The Review of Economic Studies* 83 (3): 1263–95.

Tagliapietra, Simone, Giovanni Occhiali, Enrico Nano, and Robert Kalcik. 2020. 'The Impact of Electrification on Labour Market Outcomes in Nigeria'. *Economia Politica* 37 (3): 737–79.

Wahab, Ibrahim, Ola Hall, and Magnus Jirström. 2018. 'Remote Sensing of Yields: Application of UAV Imagery-Derived NDVI for Estimating Maize Vigor and Yields in Complex Farming Systems in Sub-Saharan Africa'. *Drones* 2 (3): 28.

Yamauchi, Futoshi. 2016. 'The Effects of Improved Roads on Wages and Employment: Evidence from Rural Labour Markets in Indonesia'. *The Journal of Development Studies* 52 (7): 1046–61.

Annex 1: Summary tables of infrastructure studies

Transport infrastructure

Title	Author	Year	Summary and Findings	Location	Variables	Methods
Access to Markets and the Benefits of Rural Roads.	Jacoby	2000	Estimated household benefits from road investments, focussing on rural areas and distance to agricultural markets. Findings show that providing road access to markets produces the most benefits for poorer households. Discovered that a 10 per cent increase in travel time resulted in a small increase in agricultural wages by 0.5 per cent.	Nepal	Independent variable: Median travel time (log) Dependent variables: Plot value, agricultural wages (log daily wages)	Use median travel time (taken from household survey) to market centre to look at relationship between travel time, plot values and agricultural wages. Included dummy variables on household characteristics including age, sex, and education level. OLS and IV techniques were applied. Input data was obtained from household survey.
Trade Integration, Market Size and Industrialization: Evidence from China's National Trunk Highway System	Faber	2014	Investigate the effects of transport infrastructure investment (Chinas national trunk highway system) on trade integration, markets and industrialization. Discovered that connections led to a decrease in GDP growth among non-targeted, peripheral counties due to reduced industrial output growth.	China	Independent variable: Distance to the nearest NTHS route (distance measured using least cost path spanning tree and nearest Euclidean distance spanning tree). Used a binary indicator if a location had a connection to the highway. Dependent variables: total GDP, industry gross value added,	Applied an IV approach based on the construction of a hypothetical least cost path spanning tree network connecting major cities. Included province fixed effects and controlled for log distance between county & nearest targeted metropolitan city centre, political status, education attainment, share of agricultural employment. Input data was obtained from statistical yearbooks and census.

Title	Author	Year	Summary and Findings	Location	Variables	Methods
					agricultural gross value added, population.	
The Brasília experiment: The heterogeneous impact of road access on spatial development in Brazil	Bird and Straub	2020	Investigate the impact of road network expansion in Brazil between 1960s-2000s, by looking at the impact on spatial allocation of population and economic activity. Findings show better access to urban centres as a result of proximity to roads generated agglomeration effects in population and GDP growth. These effects were strongest in in proximity to urban centres up to 200km. GDP elasticities were 2.8 at less than 100km from a road, 2.4 between 100 and 200km and 1.1 between 200 and 500km and were statistically significant.	Brazil	Independent variable: Log of distance to transport lines Dependent variables: GDP (split by agriculture, industry, services sectors) and population	Constructed transport lines that coincided with the countries new radial highway system and calculated the distance to these constructed lines to instrument for the change in access to state capitals. Applied an OLS and IV approach where they controlled for distance to state capital, area, slope and elevation. GDP and population data was obtained from census.
Access to transportation infrastructure and economic growth in China	Banerjee et al.	2020	Investigate the effects of access to transport networks on economic growth over ~20 years (1987-2005). Discovered that proximity to transportation networks have a positive causal effect on per capita GDP levels across sectors however have no effect on per capita GDP growth. Found that distance from	China	Independent variable: log distance to historical transport line (Straight line distance to the line) Dependant variables: per capita GDP, manufacturing firms/profit and household income for agricultural households.	They applied regression analysis to investigate the impact that transportation infrastructure had on per capita GDP, manufacturing firms locations and profit and household income. Including province and fixed year effects, they discovered that proximity to transportation lines had a moderate positive causal effect on per capita GDP across sectors.

Title	Author	Year	Summary and Findings	Location	Variables	Methods
			line had little effect on income growth.			Calculated distance to historical transport lines using straight-line distance. They applied regression analysis to look at the relationship between distance and the dependant variables, controlling for the length of road and distance to city/river/coastline and included province fixed effects and year fixed effects. County-level economic data from China was used.
Market Access and Structural Transformation: Evidence from Rural Roads in India	Asher and Novosad	2016	Investigates the impact of rural road construction on rural employment and economic outcomes (2000-2012). Discovered that road construction in previously unconnected villages lead to a 10 percentage point reduction in the share of household workers in agriculture, with an equivalent increase in wage labour market participation. Found that reallocation to wage labour was strongest in locations closer to major cities, with the average effect driven by villages less than 63km from cities.	India	Independent variable: Presence of rural road construction (dummy) Dependent variables: household income source (by occupation), household earnings (income) and assets, NTL luminosity.	Looked at the relationship between road on construction and employment in agriculture, household earnings, NTL (as proxy for economic activity). Used a fuzzy regression discontinuity design, including district cut-off fixed effects and village-level controls. Baseline village characteristics include education, medical centre, electrified, distance from town. Data was obtained from population and socioeconomic census.
The Effects of Improved Roads on Wages and Employment: Evidence from Rural Labour Markets in Indonesia	Yamauchi	2016	Investigates the impact of road quality on labour supply in rural Indonesia. Findings show that an increase in transport speed raised wages in employment and decreased working time in agriculture. They found that more educated	Indonesia	Independent variables: log average transport speed at village-level (transport speed was determined by road quality)	Used both IV and OLS estimation to look at the relationship between working time, income and change in transport speed as a result of road projects between 2007-2010 that improved road quality. Measured at the village level and included village dummies to absorb village-level shocks. Data on household

Title	Author	Year	Summary and Findings	Location	Variables	Methods
			households experience an increase in wages and total income for non-agricultural employment when road quality improves.		Dependant variables: working time, monthly and daily incomes for agricultural and non-agricultural employment	characteristics were obtained from household survey data and village census. Instruments used included indicators for the number of road projects completed in the time period interacted with years of schooling, age and household size.
Urban Transport Expansions, Employment Decentralization, and the Spatial Scope of Agglomeration Economies	Baum-Snow	2013	Investigated the relationship between highway construction and the decentralization of jobs between 1960-2000. Discovered that new highways resulted in residential decentralization, displacing 16 per cent of the working population and 6 per cent of jobs to the suburbs. Using a central city radius between 2 and 9km from the CBD, the addition of each radial highway was estimated to cause about 20 per cent decentralization of city workers.	US	Independent variable: change in the number of highways between 1950 and 2000 that serve a primary cities' central business district. Dependent variables: change in log of employment, with outcomes by industry (manufacturing, services, trade)	Applied OLS regression and IV estimation to look at the average treatment effects between highway construction and employment by industry. Included control variables on size of the central city (as larger cities receive more planned highways) and the log of SMSA employment + residents. The endogenous variable was instrumented with the number of planned highways constructed prior to 2000. Input data came from census.
Farther on down the Road: Transport Costs, Trade and Urban Growth in Sub-Saharan Africa	Storeygard	2016	Investigated the effects of transport costs and the income of SSA cities, based on distance to primary city. Found that oil prices increase induces the income of cities near that port to increase by 7 per cent relative to otherwise identical cities 500 km farther away.	SSA	Independent variable: Transport costs, road condition (dummy variable denoting whether a city's route is more paved than the route of the median city in the selected country) Dependant variable: the log of lights (as a proxy for GDP), for a sample of some countries with	Calculated distance using the shortest path along a country's road network, applying alternative speeds based on road condition (paved vs unpaved roads). Use the changes in prices of oil interacted with distance to look at the impacts of transport costs. Applied tobit regressions including country -year and city fixed effects. NTL data was used as a proxy for GDP, road data was sourced from World Bank,

Title	Author	Year	Summary and Findings	Location	Variables	Methods
					available census data, included the fraction of the employed population over age 10 working in manufacturing	and employment in manufacturing data was sourced from census for a subset of countries.

Electricity infrastructure

Title	Author	Year	Summary and Findings	Location	Variables	Methods
Out of the Darkness and Into the Light? Development Effects of Rural Electrification	Burlig and Preonas	2016	Estimated the effects of India's national electrification programme by looking at the impact of rural electrification (using NTL to measure electrification) on economic development. Findings show that eligibility for the electrification programme led to an increase in NTL brightness and access to electricity. However, they did not discover any meaningful impact on increase in brightness and the economic/development outcomes of interest.	India	Independent variable: change in NTL brightness (where an increase is an indication of increase in electrification) Dependant variables: employment, household income, asset ownership, education	Used a regression discontinuity design to look at the village level effects of electrification on employment, asset ownership, household wealth. Applied an eligibility cut off based on population threshold for eligibility of a village receiving investment in electrification. Controlled for 2001 population as the running variable, state fixed effects and the 2001 levels of the dependant variables. NTL data was used from DMSP and socioeconomic variables were taken from census data.
The Effects of Rural Electrification on Employment: New	Dinkelman	2011	Estimated the impact of electrification on employment growth in South Africa. Findings show that electrification projects raise female employment within a 5-	South Africa	Independent variables: Presence of Eskom electrification project (dummy)	Used an instrumental variables strategy and fixed effects approach. Performed OLS and IV regression looking at the relationship between male/female employment rate and the presence of an electrification project.

Evidence from South Africa.			year period (an increase of 9.5 percentage points), increase hours of work (falls between 1.3 and 1.9 hours more work per week), reduced female wages and increased male earnings.		Dependent variables: male/female employment, hours of work, log hourly wage, log monthly earnings	Included ten district fixed effects and instrumented for average land gradient. Control variables included poverty rate, female headed household and adult-sex ratio. Data was obtained from census and LFS.
The impact of electrification on labour market outcomes in Nigeria	Tagliapietra et al.	2020	Sought to understand the effect of electricity access on labour market outcomes, by applying an IV approach similar to Dinkelman (2011). Discovered that electrification resulted in a shift out of agricultural employment of 7 per cent and into non-agricultural employment of about 15 per cent. Access to electricity led to an overall 8 per cent increase in employment.	Nigeria	Independent variable: Access to electricity (dummy) Dependent variables: proportion of people of working age who are employed, split by female/male and agricultural/non-agricultural	Applied an instrumental variable approach to look at the relationship between access to electricity and employment. Used a bivariate probit model. Instruments included household distance to the grid, distance to the nearest power plant and gradient of household location site. Other covariates of interest included gender and age of household head, household\ size, highest education obtained, distance to nearest road and population centre. Data was obtained from general household survey (GHS).

Water infrastructure

Title	Author	Year	Summary and Findings	Location	Variables	Methods
Breaking Ground: Unearthing the Potential of High-resolution, Remote-sensing Soil Data in Understanding Agricultural Profits	Bhargava et al.	2018	Investigated how changes in land quality, using soil organic carbon (SOC) content, affects agricultural profitability and agricultural technology adoption for farmers in Tanzania. Findings show an increase of US$7–8 per acre per year for each	Tanzania	Independent variable: soil health (measured using soil organic carbon content) Dependant variables: agricultural profits, land value	Applied fixed effects regression to look at the relationship between soil health (SOC) and agricultural profits and land value. They included dummy variable for the main crop grown on the plot. Controlled for observable and unobservable household-level factors and for observable plot-level variables. The

and Technology Use in Sub-Saharan Africa			gram per kilogram change in SOC. SOC and fertilizer use highlight a positive interaction with plot-level agricultural profits.			first includes household income, farmer ability, behavioural aspects, and household-level market imperfections. The latter includes plot characteristics, such as slope, proximity to household, and irrigation, as well as various inputs into the production process, such as chemical use, seed type, and family and hired labour by task. Input microdata came from WB LSMS.

Contact details

International Labour Organization
Route des Morillons 4
CH-1211 Geneva 22
Switzerland

T: +41 22 799 7449
E: game@ilo.org

This publication was produced with the financial support of the European Union. Its contents are the sole responsibility of the ILO and do not necessarily reflect the views of the European Union.