Farther on down the road: transport costs, trade and urban growth in sub-Saharan Africa

Adam Storeygard*

October 2011
Job Market Paper

Abstract

How does isolation affect the economic activity of cities? Transport costs are widely considered an important barrier to local economic activity but their impact in developing countries is not well-studied. This paper investigates the role of inter-city transport costs in determining the economic activity of sub-Saharan African cities. Sub-Saharan Africa is a relevant and important setting because of its high spatial concentration of manufacturing and low but increasing levels of urbanization. However, the lack of panel data on both local economic activity and transport costs has prevented rigorous empirical investigation of this question. I fill this gap with two new datasets. Satellite data on lights at night proxy for city economic activity, and new road network data allow me to calculate the shortest route between cities. Cost per unit distance is identified by plausibly exogenous world oil prices. The results show that an oil price increase of the magnitude experienced between 2002 and 2008 induces near cities to become 6 percent larger than otherwise identical cities one standard deviation farther from a major port. Combined with external estimates, this implies an elasticity of city economic activity with respect to transport costs of -0.2 at that distance. Moreover, the effect differs by the surface of roads between cities. Cities connected to the port by paved roads are chiefly affected by transport costs to the port, while cities connected to the port by unpaved roads are more affected by connections to secondary centers.

JEL classification: F15, O18, R11, R12, R4
Keywords: Urbanization, Transport costs, Infrastructure, Roads, Sub-Saharan Africa

---

*Department of Economics, Brown University. Email: adam_storeygard@brown.edu. I am grateful to David Weil, Nate Baum-Snow, and especially Vernon Henderson for their advice and support. I thank Katherine Casey, Uwe Deichmann, Andrew Foster, Delia Furtado, Jake Goldston, Walker Hanlon, Erik Hurst, Blaise Melly, Sriniketh Nagavarapu, Mark Roberts, Nathan Schiff, Michael Suher, Ishani Tewari, Matt Turner, Nicholas Wilson, and seminar participants at Brown, CIESIN, the North American Regional Science Council (NARSC) annual meeting, and EconCon for helpful comments and suggestions, Uwe Deichmann and Siobhan Murray for access to and guidance on the roads data and NARSC and the World Bank's Research Support Budget and Knowledge for Change Program for funding. All errors are my own.

1 Introduction

Sub-Saharan Africa has experienced substantial economic growth in the past decade, after a quarter century of stagnation and contraction.[1] It is perhaps too early to answer the question of why this period of growth has arrived. However, it is important to know where it has happened within countries and why these places have been favored.

In this paper, I argue that transport costs have played a critical role in determining the growth of cities. Sub-Saharan Africa has notoriously high transport costs compared to other major regions of the world. Population density is relatively low, with a substantial fraction of people residing far from the coast. Ocean-navigable rivers, which provide transport to the interior of most other regions, are virtually non-existent. And road networks are sparse and poorly maintained, on the whole. I ask whether periphery cities with lower transport costs to their country's main port grew faster than those farther away or with poorer road connections, in the context of dramatically rising oil prices over the 2000s decade. A typical problem with testing this kind of question in poor countries is that relevant data on cities and transport costs do not exist. This paper provides novel measures of both. First, night time lights satellite data (Elvidge et al., 1997; Henderson, Storeygard and Weil, forthcoming) are used to construct a 17-year annual panel of city-level measures of economic activity for 287 cities in 15 countries. Second, a new set of roads data provides information about route length and surface material. Transport costs are thus identified by the interaction between world oil prices and distance along these routes. Because I have data on many cities per country over a substantial time period, I can control for annual shocks separately for each country, as well as initial size and growth rates of individual cities.

Focusing on countries whose largest, or *primate*, city is also a port, I find that as the price of oil increases from \$25 to \$97 (as it did between 2002 and 2008), if city A is 465 kilometers (1 standard deviation) farther away from the primate than initially identical city B, its economy is roughly 6 percent smaller than city B's at the end of the period. At a differential of 2360 kilometers, the largest in the data, this rises to 32 percent. I then determine that this effect falls disproportionately on cities that are connected to the primate by paved roads, most likely because they are initially more engaged in trade. Cities connected to the primate by unpaved roads appear to be more affected by transport costs to secondary cities.

The majority of Africa's population growth is expected to be in cities over the next

[1] This paper generally defines sub-Saharan Africa as all countries of the African mainland with no Mediterranean coastline, plus Madagascar.

few decades (National Research Council, 2003). Indeed, of the approximately 2.5 billion net gain in global population expected by 2050, over 30 percent is expected to be in African cities (United Nations, 2008). Much less is understood, however, about which cities will experience the bulk of that growth. This is the first paper that has the data to systematically address the past growth of cities in sub-Saharan Africa. Which cities grow and which do not will have major implications for the future of the region. If growth is concentrated in large coastal cities, agglomeration economies may improve, but urban infrastructure needs, congestion, and the risks associated with sea level rise will all increase. Many countries have pursued decentralization policies at least in part because of these concerns. If growth is more balanced across a large group of cities, intercity infrastructure may be more important. If cities near international borders grow more than others, they may affect international migration and trade.[2] Regional cities that grow enough may become political power bases constituted along social or industrial lines.

Despite a great deal of attention paid to the largest cities like Lagos, Kinshasa and Nairobi, 61 percent of African urban residents lived in cities with a population of less than 500,000 as of 2000, and this percentage is expected to remain high over the next couple of decades. Only 15 percent are expected to live in cities of more than one million by 2015 (National Research Council, 2003). There are only 5 out of 43 sub-Saharan countries (Liberia, South Africa, Togo, and both Congos) in which the majority of 2005 urban residents were in urban agglomerations with a 2007 population of more than 750,000.[3] In most countries, smaller cities are also growing faster (United Nations, 2008).

These large cities may however have greater importance to national economies than their populations alone suggest. In many countries, a large city or core region, often a port, plays a very important role in the economy, as the largest domestic market, the chief manufacturing center, the primary trading connection with the rest of the world, and the seat of elites and often of the government. Ports have a special role because most African trade is transoceanic. Trade among the contiguous eight members of the West African Economic and Monetary Union (WAEMU) represented less than 3 percent of their total trade for each year in the 1990s (Coulibaly and Fontagné, 2006).[4] If anything, one would expect more trade among these countries than other sets of neighbors, because

[2]The center of some ten sub-Saharan African capital cities is within 20 km by land of an international border. Such proximity is extremely rare in other regions, except among countries less than 40 km in diameter. In six of these cases, the built-up area of the city directly abuts the border or border-river. More than half the countries in the region have capitals within 100 km of a border by air.

[3]All country counts were conducted prior to the independence of South Sudan.

[4]The eight countries are Benin, Burkina Faso, Côte d'Ivoire, Guinea-Bissau, Mali, Niger, Senegal and Togo.

they share a common currency and thus lack one important trade friction. Other cities in the periphery have relationships with their country's core that are potentially critical to their success. And countries spend to improve those links or simply to reverse decay. Almost $7 billion is invested per year on roads in sub-Saharan Africa, with a substantial portion funded by donors (World Bank, 2010). Worldwide, transport accounts for 15 to 20 percent of World Bank lending, with almost three quarters of that amount going to roads (World Bank, 2007a).

This paper relates primarily to four bodies of work. The first is on the effect of transport costs on the size and growth of cities and regions. This work has been done primarily using cross-country data (e.g. Limão and Venables, 2001) or the construction of very large national transport networks in the United States (Baum-Snow, 2007; Chandra and Thompson, 2000; Atack, Bateman, Haines and Margo, 2010; Duranton and Turner, 2011), India (Donaldson, 2010), and China (Banerjee, Duflo and Qian, 2009). However, little has been done in sub-Saharan Africa, which has worse roads, lower urbanization, lower income, and much less industry, and consists of many countries, as opposed to one unitary state.[5] Similarly ambitious transport infrastructure projects have not been carried out in post-independence sub-Saharan Africa. This paper instead relies on the plausibly exogenous annual changes in transport costs induced by world oil price fluctuations, which allow me to determine the impact of transitory shocks. This does not mean that the changes were small, however, as average annual oil prices varied by a nominal factor of 7.6 during the period of study (1992–2008).[6] These shocks are also of interest because they are more likely to be repeated in the future.

The second related literature is on the scope and drivers of urbanization and urban economic growth in Africa. This literature is almost exclusively cross-country in nature, so that unobserved country-level factors may be confounding results (Fay and Opal, 2000; Barrios, Bertinelli and Strobl, 2006). An exception is Jedwab (2011), who looks at districts within two countries, Ghana and Côte d'Ivoire, and argues that local production of cash crops, specifically cocoa, spurred urbanization outside of the few largest cities. In his setup, consistent with mine, these secondary towns form primarily as "consumption cities" where farmers sell their products and buy services and imported goods, as opposed to manufacturing centers as is often assumed in models of urbanization and city formation. Unlike all these papers, the present outcome of interest is a proxy for economic activity (lights) that is available for individual cities on an annual basis, as opposed to popula-

[5] An exception is Buys, Deichmann and Wheeler (2010), who consider the possible effects of road upgrading on *international* trade.

[6] This corresponds to a real factor of 5.8 using the United States Consumer Price Index (CPI-U).

tion, which is typically only available for censuses carried out at most every ten years. This allows me to observe short-run (annual) changes and to control for all potentially confounding country-level variation with country*year fixed effects.

In stressing the role played by the largest city in each country, this work also has implications for the study of urban primacy ((Ades and Glaeser, 1995; Henderson, 2002) and decentralization. I find that primate cities are growing slower than others on average net of the transport cost effect. Finally, in focusing on the importance of coastal cities, this work relates to the literature on geographic determinants of growth, including Gallup, Sachs and Mellinger (1999) and Collier (2007), which emphasize coastal access and the problems of being landlocked, respectively.

The remainder on the paper has the following structure. Section 2 provides a simple conceptual framework to facilitate interpretation. In section 3, I describe the lights and roads data and the methods used to integrate them. In section 4, I describe the econometric specification used, and in section 5, I report results. Section 6 concludes. An Appendix provides further details on the data and methods used.

2 Conceptual Framework

Economists often think about the role of intercity transport costs in city growth in the context of two-region New Economic Geography (NEG) models following Krugman (1991), or one of many variants, including one that adds agricultural transport costs (Fujita, Krugman and Venables, 1999), one that adds a foreign sector accessible from a port in one city (Behrens, Gaigné, Ottaviano and Thisse, 2006), and one that is tailored to African urban primacy (Pholo Bala, 2009). While changes in transport costs drive urban growth in these models, they do so by inducing manufacturing firms to change their location from the (ex-post) periphery city to the (ex-post) core, with its larger home-market effects.

This is unlikely to be the driving force behind the current growth of most cities in Africa, because manufacturing activity is already highly concentrated in the largest cities. For example, as of 2002, the Dar es Salaam administrative region contained 0.16 percent of mainland Tanzania's land area, and 8 percent of its population, but 40 percent of its manufacturing employment and 53 percent of manufacturing value added (National Bureau of Statistics, 2009).[7] Although Tanzania has a coastline of over 1,400 kilometers and three other ports, Dar es Salaam handled 95 percent of its port traffic as of 1993 (Hoyle and Charlier, 1995).

[7] These manufacturing statistics are based on establishments with more than 10 employees

If the periphery city already has minimal manufacturing and is more of a market center for exchanging manufactured goods imported from the core with agricultural products from nearby rural areas, then decreased transport costs will not bring increased competition with core manufacturing. Instead, they may increase exports of rural agricultural goods that are otherwise not sold on the market by farmers, and increase imports of manufactured goods, including inputs to agricultural production, in exchange, or at least change the relative consumption of rural and periphery city residents. Alternatively, it is possible that these hinterland cities are essentially in autarky: because they already have very high transport costs, they are largely insulated from changes in them. In sub-Saharan Africa, while some non-primate cities clearly have manufacturing, it is this trading margin that I expect is more relevant to understanding the growth of cities.

The following model embeds this intuition in a very simple framework in which changes in transport costs drive urban growth. The economy under study is one that exports agricultural goods and imports manufactures. Specifically, the periphery city under study exports its agricultural surplus and imports its manufactured goods, both via a core or port city. The key ingredients are tradable goods prices fixed in the core city by international trade but affected in the periphery city by intercity transport costs.

A representative farmer in the periphery has preferences $U(a, m)$ over an agricultural good (a) and a manufactured good (m), with a diminishing marginal rate of substitution (MRS) between the two. The farmer is endowed with a_0 units of the agricultural good, and can travel costlessly to a nearby periphery city, which consists of perfectly competitive trading firms, to buy the manufactured good.[8] From the perspective of the farmer, these trading firms sell the manufactured good and buy the agricultural good. Trading firms employ a variable factor, their only input, in a constant returns to scale (CRS) production function at a perfectly elastic price w per unit of m traded. Therefore profits are zero and the income of the city is just wm. There is no mobility between sectors or locations.

Manufactured and agricultural goods have prices p_m and p_a, respectively, in the core city, which is much larger than the periphery city and directly connected to world markets, so these prices are exogenous to activity in the periphery city. Both goods can be transported between the core and periphery cities for a per-unit transport cost τ. The farmer's budget constraint is then

$$(a_0 - a)(p_a - \tau) = m(p_m + \tau + w). \tag{1}$$

Revenue from sold agricultural goods is used to buy manufactured goods. It is assumed

[8] Adding a travel cost here would not substantively change any results.

that $p_a > \tau$. Otherwise the farmer would be in autarky, simply consuming his or her own produce, and there would be no trading sector. In the empirical work below, this autarky is the null hypothesis and it is rejected. Given certain consumer preferences, the cutoff τ for autarky could be lower than p_a.

The farmer maximizes $U(a, m)$ with the result that

$$\frac{U_m}{U_a} = \frac{p_m + \tau + w}{p_a - \tau}. \tag{2}$$

Denote the left hand side of Equation (2) $g(m)$ and the right hand side $h(\tau)$. A diminishing MRS implies that $g(m)$ is monotonically decreasing in m, or $g'(m) < 0$. Differentiating $h(\tau)$ with respect to τ yields

$$h'(\tau) = \frac{p_a + p_m + w}{(p_a - \tau)^2} > 0 \tag{3}$$

Since $g(m)$ is monotonically decreasing in m, we can uniquely define its inverse $g^{-1}(h)$ and know that $\frac{\partial g^{-1}}{\partial h} < 0$. Inverting Equation (2) as $m = g^{-1}(h(\tau))$, applying the chain rule and multiplying by the exogenous parameter w gives the testable prediction

$$\frac{\partial (wm)}{\partial \tau} < 0. \tag{4}$$

In words, increasing the transport costs of a periphery city decreases the trade there between agricultural goods and manufactured goods. Since this trade is the only activity in the periphery city, its income decreases as well. This is the comparative static tested below. While this is a very simple model, it has a couple of appealing features. First, the only assumption on the utility function is a diminishing MRS. Second, including different transport costs for the two goods yields similar results. This is useful because agricultural and manufacturing transport costs may differ depending on the season and the overall trade balance, since trucks must complete round trips, and the demand for export transport may be substantially lower than the demand for import transport, especially outside of harvest seasons (Teravaninthorn and Raballand, 2009). The analogous corner case of the Krugman (1991) model, in which all manufacturing has agglomerated in the core, delivers the same prediction on periphery city income, but with much more specific functional forms.

3 Data and spatial methods

In order to test this model, attention is restricted to a set of 15 *coastal primate* countries in which the main city is also the main port, so inexpensive transportation to the primate is important for trade with both the largest domestic market and the rest of the world (Figure 1).[9] Counterclockwise from the northwest, these countries are Mauritania, Senegal, Guinea, Sierra Leone, Liberia, Côte d'Ivoire, Ghana, Togo, Benin, Nigeria, Cameroon, Gabon, Angola, Mozambique, and Tanzania. Further details about all data used are in the Appendix.

3.1 City lights

To date, very little economic data, especially for income and especially as a panel, have been available for individual African cities.[10] In most national household surveys, if any city is individually identifiable, it is only the largest city in a country. The largest program of firm surveys, the World Bank Enterprise Surveys, rarely collects data in more than four cities per African country, and it is typically not clear how the surveyed cities are selected. Censuses often report populations for many cities, but they are almost always carried out at intervals of at least ten years, which limits their usefulness. In order to fill this gap, I propose a novel data source as a proxy for city-level income: satellite data on light emitted into space at night.

Satellites from the United States Air Force Defense Meteorological Satellite Program (DMSP) have been recording data on lights at night using their Operational Linescan System sensor since the mid-1960s, with a global digital archive beginning in 1992.[11] Since two satellites are recording in most years, 30 satellite-years worth of data are available for the 17-year period 1992–2008. Each 30-arcsecond pixel in each satellite-year contains a digital number (DN), between 0 and 63, that represents an average of lights in all nights after sunlight, moonlight, aurorae, forest fires, and clouds have been removed algorithmically, leaving mostly human settlements.[12] Figure 2 shows the lights data for

[9]Five other countries in sub-Saharan Africa fit this criterion but are not included in analysis. Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia are excluded because they lack (at least) roads data. Using the city definitions below, The Gambia has only one city, and therefore it provides no information in the presence of country*year fixed effects.

[10]Some administrative data on economic indicators such as employment are collected for smaller regions in some countries, including Tanzania, but assessing their comparability is a challenge, and they are typically not available for multiple years.

[11]These sensors are designed to collect low light imaging data for the purpose of detecting moonlit clouds, not lights from human settlements.

[12]A 30-arcsecond pixel has an area of approximately 0.86 square km at the equator, decreasing proportionally with the cosine of latitude. The data are processed and dis-

one satellite-year for Tanzania. No lights are visible in the overwhelming majority of the land area. In Figure 3, a closer view of Dar es Salaam, Tanzania's largest city, shows a contiguously lit area 20–30 km across, extending farther in a few directions along main intercity roads just as the city's built up area does.

Henderson, Storeygard and Weil (forthcoming) show that light growth is a good proxy for income growth at the national level. Annual changes in gross domestic product (GDP) are correlated with changes in DN, with an elasticity of approximately 0.3 for a global sample as well as a sample of low and middle income countries. In both samples, the lights explain about 20 percent of the variation in log GDP net of country and year fixed effects.

The chief strength of the lights lies in their geographic specificity—they are highly local measures. To proceed with lights as a measure of city-level GDP, it must first be shown that the strong national relationship holds for *subnational* regions. This is problematic because of a mismatch in data availability. Rich countries tend to have good local economic data, but the lights data are heavily topcoded in their cities. Lights topcoding is less of a problem in most poorer countries, and especially in sub-Saharan Africa, where almost no pixels are topcoded (15 per 100,000, or 3 per 100,000 outside of South Africa and Nigeria). However, good local economic data are rarely available. China and South Africa represent good compromises, with relatively little topcoding but relatively high quality panel income data.

China has panel GDP data for two relevant types of subnational regions: cities proper and prefectures.[13] In Table 1, columns 1 and 2 show, at the city proper and prefecture level, respectively, that the elasticity of GDP with respect to light is significantly positive in a 1990/1992–2005 long difference specification.[14] The point estimate is very similar to the one for the global sample.

South Africa also has household income data in its 1996 and 2001 censuses that can be used to form an income panel for administrative units called magisterial districts (MD).[15] When lights and income are summed within 284 MD, the elasticity of changes in income

tributed by the United States National Oceanic and Atmospheric Administration (NOAA), http://www.ngdc.noaa.gov/dmsp/downloadV4composites.html. Accessed 22 January 2010

[13]I am grateful to Vernon Henderson and Qinghua Zhang for providing the China evidence based on their work in progress with Nathaniel Baum-Snow, Loren Brandt, and Matthew Turner.

[14]The lights are from 1992 but the GDP data are from 1990—the closest year with good data.

[15]Produced by Statistics South Africa and distributed as: Minnesota Population Center. Integrated Public Use Microdata Series (IPUMS), International: Version 6.1 [Machine-readable database]. Minneapolis: University of Minnesota, 2011. https://international.ipums.org/. Accessed 28 April 2011. The 2007 census also includes household income data, but the smallest comparable geography in the IPUMS sample is the province, and there are only nine of them. The IPUMS data for both 1996 and 2001 are 10 percent samples of the full census.

with respect to changes in light is 0.17 (Table 1, column 3). This elasticity stays in the range 0.169 to 0.171 whether six province-level aggregates of small MDs (used to preserve confidentiality) are included or excluded, and whether topcoded household incomes are modeled at the topcoded value or 1.5 times that amount (not shown). The analogous t-statistic ranges from 2.96 to 3.08. In the analysis below, I use the cross-country elasticity to convert changes in light to income changes.

For the present study, several steps were taken to convert the pixel-level lights data into cities. Figures 4–6 document these steps for Tanzania. The 30 satellite-years of lights data were first combined into one binary grid encoding whether a pixel was lit in at least one satellite-year (Figure 4). These ever-lit areas were then converted to polygons: contiguous ever-lit pixels were aggregated, and their digital numbers were summed within each satellite-year. Polygons not corresponding to a known city, based on census populations with latitude-longitude pairs, were dropped.[16] Figure 5 shows all the lit polygons and city points. For those light polygons that did contain one or more census cities, the population of all such cities were summed to obtain a population. Most lights correspond to at most one census city.[17]

Figure 6 shows those lights that corresponded to known cities. In most countries, census information about cities with populations as small as 10 thousand was available, but in some, the cutoff was higher. For all regressions below, I restrict to cities with combined population over 20 thousand and lit in at least 2 years. The dropped lights most likely correspond to forest fires or random noise not flagged by NOAA's algorithm, or smaller towns, and contain 13 to 16 percent of total DN in the 15-country sample. The total DN was recorded for each city polygon for each year, averaging across multiple satellite-years where necessary.[18] The light in each country with the largest associated population in 1992 is designated the primate.[19]

[16] City Population, http://www.citypopulation.de. Accessed 19 March 2010.

[17] Light pixels for a given satellite-year actually represent the average light from several slightly larger overlapping pixels from many orbits within the satellite-year. Because of this, the lit area of a given city tends to be somewhat larger than its actual size. Among densely populated high and middle income countries, this means, for example, that the majority of land in the United States east of the Mississippi River or in continental Western Europe is contiguously lit, so that cities cannot be defined purely based on light contiguity. In Africa, this is much less of a problem because of sparser light overall. Snow also tends to increase the footprint and magnitude of lights. Again, this is less of a problem in Africa than elsewhere. And even if the area of a given city is overestimated, the light summed for that city in still presumably coming from that city or its outskirts—it may just be partially displaced a pixel or two from where it actually originates.

[18] Lights arising from gas flares, as delineated by Elvidge et al. (2009) were also removed. These affected only 4 populated lights in the 15-country sample.

[19] In practice, the primate designation does not change over the course of the sample period in any sample country.

3.2 Transport costs

Trade and transport cost data are also not widely available for Africa.[20] In the international trade literature, trade costs are sometimes estimated from a gravity equation based on trade flows (Anderson and van Wincoop, 2004), or price dispersion (Donaldson, 2010) but trade flow data between cities and city-level price data are also not widely available. Furthermore, city growth may endogenously decrease transport costs. Among other reasons including the allocation of paved roads (discussed below), more transport companies are likely to compete on a route to a growing city than on a route to a stagnant one.

I deal with this by decomposing variable transport costs into two components: 1) the world price of oil, which varies across time but not across cities, and 2) the road distance between a city and its country's primate, which varies across space but not time.[21] Figure 7 shows the evolution of oil prices during the study period. In general, they were relatively steady until a consistent rise beginning in 2002. However, there was some movement in the previous period, including substantial decreases (as a fraction of the initial price) in 1992–1994, 1996–1998, and 2000–2001.

Oil is a convenient proxy for transport cost per distance because no countries in the sample are individually capable of influencing its price substantially. However, motorists consume refined petroleum products, mostly gasoline and diesel, not oil, and some countries, especially oil producers, subsidize their prices. Country-specific diesel prices, surveyed in November in the main city, are available for most countries roughly every two years (Deutsche Gesellschaft für Technische Zusammenarbeit, 2009). As shown in Figure 7, diesel prices averaged over a balanced panel of 12 countries from the main estimation sample generally rise in parallel with oil prices. Nigeria, Gabon, and Angola, the three sample countries for which oil production represents the largest fraction of GDP, show similar, though somewhat noisier time trends (Appendix Figure A.1) despite the fact that they typically had lower prices than average in most years, most likely because of subsidies. Using data from a survey of truckers in several African countries, Teravaninthorn and Raballand (2009) estimate that fuel represented roughly 35 percent of transport costs for trucks in 2005, when oil prices were roughly the mean of the minimum and maximum annual price for the period.

Rudimentary national statistics like road density and percentage of roads paved, which are typically used in cross-national studies, fail to capture the role of roads in connecting

[20] Teravaninthorn and Raballand (2009) provide figures for several important routes from landlocked country capitals to the ports that serve them.

[21] Specifically, the oil price used is the annual average Europe Brent Spot Price FOB, in dollars per barrel, from the United States Energy Information Administration (http://tonto.eia.doe.gov/; accessed 5 Jul 2010).

cities, and are subject to a great deal of error.[22] A recent World Bank project on infrastructure in Africa has improved the state of georeferenced roads data for the continent so that they can be used to assign infrastructure to specific cities and routes between cities (World Bank, 2010).[23] The resulting dataset combines information on road location and surface assigned to specific (and recent) years from each country's Transport Ministry or equivalent, or a consultant specific to the project. It contains information on over a million kilometers of roads in 39 countries.[24] For over 90 percent of this length, a measure of the surface type is recorded. The comprehensiveness of the coverage varies by country, but only in that some countries contain more minor roads. Intercity roads are available for all countries. Figure 8 shows these roads data for Tanzania. Roads go through all the populated cities shown. Most roads are unpaved, and most paved roads are found along a few major corridors.

The shortest path along the road network was calculated between the centroid of each city-light and three destinations: (1) its country's primate city, defined as the light with the largest associated population in 1992, (2) the nearest city in the same country with a 1992 population of at least one hundred thousand, and (3) the nearest city in the same country with a 1992 population in the top quintile of the population distribution for that country. Plausible primate city routes were found for 287 out of 299 cities in the 15-country sample.[25] Figure 9 shows all roads and primate routes for Tanzania. Descriptive statistics are in Table 2.

Limiting attention to road transport costs might be a poor source of variation if rail played a major and independent role. However, roads dominate transport in Africa, carrying 80 to 90 percent of passenger and freight traffic (Gwilliam, 2011). In most

[22]Countries such as Canada, Australia, and Botswana have low road density relative to their economic peers. But their road systems are not particularly inadequate. In each case, a contiguous region containing half or more of total land area has very low population density, so that the marginal benefit of an additional road there is very low. The chief problem with the percentage of roads paved is that the denominator is affected by the coverage of national roads data systems, which can vary substantially. The World Development Indicators (WDI) reports both of these measures. But of the 120 annual changes in road density available for the 255 country-years in the present sample, 66 are zero, while another 8 are, implausibly, over 10 percent in absolute value. When data are missing for a period of one or more years for a given country, the annual growth rates implied by the values before and after the data gap are even more implausible. Similar statements can be made about the percent paved data. These large changes are likely due to reclassification rather than actual road construction.

[23]The most comprehensive previous spatial database, Vector Map Level 0 (VMAP0, formerly known as Digital Chart of the World, DCW), is a declassified US military product combining data of unknown quality from 4 decades, with little metadata. In some countries, there are clear gaps in coverage. Most strikingly, the most densely populated areas of Bangladesh, surrounding the capital Dhaka, have essentially no roads.

[24]Of the 43 countries in sub-Saharan Africa, only Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia do not have data.

[25]The remaining 12 cities include 4 on islands or in exclaves.

countries, rail only exists along a few corridors that are also served by roads. Of course, regardless of rail's importance, for the purposes of this paper, it is still a transportation form that uses energy generated from fossil fuels, so including rail would likely have little effect.[26]

4 Empirical specification

My baseline specification testing the effect of transport costs in Equation (4) is:

$$\ln y_{it} = \beta p_t x_i + \lambda_{ct} + \gamma_i + \omega_i t + \epsilon_{it} \tag{5}$$

where y_{it} is light output for city i in year t, p_t is the price of oil, x_i is the distance between city i and its destination city along the road network, λ_{ct} is a country-year fixed effect (FE), γ_i is a city fixed effect, and $\omega_i t$ is a linear city-specific time trend. Standard errors are clustered at the city level.[27] The regression sample is limited to cities with a 1992 population of at least 20,000, lit in more than one year, because populations and locations of cities of less than 20,000 are not available for several countries, and cities lit in only one year add no intensive margin information because of the city fixed effects. The time period is limited to 1992–2008 because of the lights data availability. Summary statistics for the resulting sample of 287 cities in 17 years are in Table 2. Distances are measured in kilometers, and prices are in dollars.

As oil prices increased over the course of the last decade, I expect that transport costs increased more for cities farther away from their country's core. Thus, I can use static distance and road surface measures interacted with the exogenous oil price increase to identify the differential change in transport costs faced by near and far cities. I expect that among most periphery cities without significant manufacturing, the less-connected will experience a relative loss in economic activity. This effect may be mitigated or even reversed in cities that do have manufacturing, if they have enough of a home market that its protection outweighs decreasing access to the primate and international markets. However, I have no systematic information on manufacturing. To the extent that this effect is present, it is driving my estimate of β toward zero.

Country-year fixed effects control for any national-level time-varying economic conditions. In the context of the model, the relevant factors are prices in the primate city.

[26]Rail is also less likely to matter than roads because of its higher fuel efficiency and greater dependence on parastatals with long term contracts.

[27]If, alternatively, the methods of Conley (1999) are used to account for spatial and temporal autocorrelation, the resulting standard errors are smaller.

Empirically, they also include the level of industrialization, oil production, and terms of trade, as well as policies, including gasoline subsidies and preferential trade pacts with developed countries like the American Growth and Opportunity Act (AGOA) and the European Union's Everything but Arms (EbA) policy. They also control for global macroeconomic fluctuations, including commodity prices, as well as differences across satellites in the lights data. City fixed effects control for initial size and all other fixed city characteristics. City-specific time trends allow each city to be on its own growth path.

The identifying assumption for β is thus that there is no other time-varying within-country variation net of linear growth correlated with network distance to the primate times the change in oil price that affects city growth, or more specifically,

$$E(\epsilon_{it}|p_s x_i, F_i s, \lambda_{cs}, \gamma_i, \omega_i s) = 0, \qquad s, t = 1992, 1993, ..., 2008 \tag{6}$$

In specification checks below, distances to other large cities are tested in combination with distance to the primate to determine whether it is actually the cost to the primate city that matters, as opposed to other correlated transport costs.

In order for these regressions to pick up these effects, it must be the case that transport costs substantially affect contemporaneous economic activity, and that oil prices affect transport costs. On the first point, Gollin and Rogerson (2010) find that in Uganda, internal transport costs for crops can easily exceed their farmgate price. It is hard to imagine that this does not affect cropping decisions. Using national trade flows data, Limão and Venables (2001) find that transport costs affect international trade substantially, with an elasticity of around -3. The World Bank Enterprise Surveys of establishments ask respondents whether "transportation of goods, supplies, and inputs...present any obstacle to the current operations of your establishment?" In the most recent (2006–2009) round, in all 15 countries studied, over half of respondents said that transportation was an obstacle, and in 11 countries, at least a quarter said that it was a major or very severe problem.[28]

On the second point, Teravaninthorn and Raballand (2009) report a breakdown of transport costs for truckers and trucking companies along several international corridors from a port to the capital city of a landlocked country. Because these are international journeys, I expect costs to be somewhat higher than for those journeys that remain in the coastal country, but as a fraction of total distance, these routes are overwhelmingly in the coastal country.[29] On the Accra-Ouagadougou route, over 80 percent of which by distance

[28]Data are available from http://www.enterprisesurveys.org/. Accessed 22 January 2010.

[29]Furthermore, it is not obvious that the international nature of these journeys increases distance-

is in Ghana, variable costs are 9 times the size of fixed costs, and fuel represents 74 percent of variable costs. Tires, which also use petroleum products, represent another 15 percent. For the Mombasa-Kampala route, over 80 percent of which is in Kenya, the analogous numbers are 0.45, 79 percent and 13 percent. For the Douala-N'Djamena route, nearly all of which is in Cameroon, the analogous numbers are 4.8, 60 percent and 17 percent.

Measured light is not strictly the same as light output. Most noticeably, some 5 percent of city-years have a DN value of zero.[30] Roughly half of these are from years before the city was ever lit ("pre-entry"). This is consistent with continuous city growth that eventually passed a threshold above which lights were detectable. The other half are years in which a city has no lights after having been lit in one or more previous years ("late zeroes").

In the Appendix, I present a model of the satellite-pixel-year level data-generating process that could in principle be estimated using maximum likelihood methods. However, the relationship of interest and all of the regressors are at the city level. Rather than perform this estimation on a dataset with approximately 5.6 million pixel-satellite-years, I instead simply sum lights across pixels within a city, average across satellites within a year, and run tobit regressions with a cutoff value of 5.5, because 6 is the smallest nonzero value found in the data, and the smallest increment is 0.5.[31]

5 Results

Table 3 column 1 reports regression estimates of Equation (5). In this and all subsequent tables, distances are measured in thousands of kilometers, and oil prices in hundreds of dollars. The negative coefficient of -0.68 on $distance(Primate) * P_{oil}$ implies that if the price of oil increased from \$25 to \$97 per barrel (as it did between 2002 and 2008), if city A is 465 kilometers (1 standard deviation) farther away from the primate than initially identical city B, its lights are 23 percent smaller than city B's at the end of the period. Applying the light-income growth elasticity $\epsilon_{GDP,light} = 0.277$ from Henderson, Storeygard and Weil (forthcoming), this implies a city product differential of 6.1 percent.

dependent variable costs more than fixed costs, as some components of the latter like insurance and customs logistics may also be higher.

[30]This statistic is calculated after cities that are never lit and cities lit in only one year have already been removed from the sample, as described above.

[31]Fixed effects tobits are biased for short panels, but this panel is 17 years long and a small percentage of observations are censored. The smallest increment in city DN is 0.5 because satellite-year pixel values are integers but there are up to two satellite per year. Averaging across two satellites sometimes produces half-integer values. Using a tobit cutoff of 1 results in estimates of the coefficient of interest with larger absolute values.

This is consistent with the model above. Far cities see their transport costs increase more than near cities, so their income falls more.

The coefficient in column 1 can be interpreted as a semi-elasticity in the context of the model above. An elasticity of city product with respect to transport costs is in some respects a more intuitive measure, but since $ln(p_t x_i)$ is equal to $ln(p_t) + ln(x_i)$, it is collinear with the country-year and city fixed effects and cannot be estimated separately. However, a distance-specific oil price elasticity can be calculated. Column 2 reports the coefficient of interest when $p_t * x_i$ is replaced with $ln(p_t) * x_i$. It is again negative and significant as expected. This can be translated into a distance-specific elasticity using three additional parameters:

$$\epsilon_{GDP,\tau} = \frac{\epsilon_{GDP,light}\epsilon_{light,P_{oil}}}{\epsilon_{\tau,P_{diesel}}\epsilon_{P_{diesel},P_{oil}}}. \tag{7}$$

A simple regression of $ln(P_{diesel})$ on $ln(P_{oil})$ using the (Deutsche Gesellschaft für Technische Zusammenarbeit, 2009) data for the fifteen sample countries provides an estimate of $\epsilon_{P_{diesel},P_{oil}} = 0.6$. Treating the (Teravaninthorn and Raballand, 2009) average fuel share as the marginal fuel share implies $\epsilon_{\tau,P_{diesel}} = 0.35$. Combining these estimates implies $\epsilon_{GDP,\tau} = -0.2$ at the median distance from the primate, 439 km, and -0.4 one standard deviation (465 km) farther away. This calculation is meant to be illustrative, as it may suffer from several potential biases, including upward (toward zero) bias from substitutability of oil in the production of transport and downward (away from zero) bias from substitutability of transport in the production of city activity.

5.1 Robustness

5.1.1 Other functional forms

Table 4 reports the results of several alternate specifications. In column 1, 5.5 is added before the DN is logged and an OLS specification is used instead of a tobit.[32] Results are very similar.

Country size varies dramatically within the estimation sample. So for example, the farthest city in Sierra Leone is only 310 kilometers away from the primate, whereas in Mozambique, the farthest is over 2000 kilometers away. In order to ensure that not all variation is coming from the largest countries, column 2 shows a specification in which $p_t x_i$ is replaced by $ln(p_t)ln(x_i)$.[33] The coefficient of interest is negative and significant.

[32]This specification uses 5.5 because of the integer nature of the DN data in combination with the censoring at 6 described above.

[33]The distance from the primate to itself is arbitrarily redefined as 1 kilometer in the log-log specifica-

In all results reported so far, the shortest route to the primate was calculated assuming that travel along unpaved roads is just as fast as travel along unpaved roads. In columns 3 and 4, routes were calculated assuming that travel on unpaved roads takes 50 percent and 100 percent longer, respectively, than travel along paved roads. Although the calculated routes are slightly different, the coefficient of interest is essentially unchanged.

5.1.2 Other effects of oil prices

Oil prices could be driving these results for reasons not directly related to intercity transport costs, but only in a way that is correlated with distance to the primate city. Table 5 provides evidence against three potential mechanisms.

Because some sample countries are oil producers, oil prices could also affect the within-country pattern of economic activity on the oil supply side. When oil prices rise, cities near production or exploration areas are more likely to benefit from increased employment and wages. Because oil wells in this set of countries are largely near the coast, oil well proximity is correlated with distance to the primate. However, while transport costs to the primate increase continuously away from the coast, it is unlikely that local oil industry effects persist throughout the country.[34] In Table 5 columns 1 and 2, I report results for the baseline specification when all cities at least partially within 50 and 100 kilometers of an oil or gas field are excluded.[35] They are very similar to the baseline specification results. This suggests that local oil industry effects are not driving my results.

It is also possible that the price of oil (and gas and coal, whose prices tend to co-vary with oil's) is directly reducing the size of distant city lights, because light is produced by electricity, and some electricity is produced by fossil fuels. It is unlikely that this is driving any results, for two reasons. First, most countries have national grids, and many are connected to international ones. Second, to the extent that this is not true, or that transmission costs proportional to distance matter, more than a third of electricity in the region is produced from hydropower, with the remainder produced primarily by thermal (oil, gas, or coal) plants (World Bank, 2010). If expensive oil is increasing the price of electricity within countries, it should do so less where hydro is the most likely source. World Bank (2010) also reports the location of power plants, by type. Column 3 restricts attention to the 251 cities in those countries that have both hydro and thermal power plants and adds to the baseline specification a term interacting $distance(Primate) * P_{oil}$

tion.

[34]The oil industry could also have national effects that are correlated with the oil price, but these are removed by the country*year fixed effects.

[35]Oil and gas field centroid locations were manually georeferenced from Persits et al. (2002). There are 24 cities within 50 kilometers, and 50 cities within 100 kilometers.

with an indicator that the closest plant to the city is a hydro plant. The new interaction is small and insignificant and has very little effect on the coefficient on $distance(Primate) * P_{oil}$.[36] The interaction term is small and insignificant, suggesting that proximity to a hydro plant has no effect on the relationship between transport costs and lights growth.

A related concern is that cities far from the primate might not be on the power grid, and therefore might be more likely to rely on non-grid electric lights fueled by diesel generators. High oil prices could reduce diesel generator use, lowering lights more in faraway cities than near ones. Data on the location of existing electrical transmission lines are available from World Bank (2010) for 13 of 15 sample countries. Transmission lines pass through 184 of 260 cities (71%) in these 13 countries. In column 4, when the sample is restricted to these 184 cities least likely to rely on diesel generators, the results are very similar to those in column 1. This suggests that diesel generators are not driving my results.[37]

As noted above motorists use gasoline and diesel, not oil, so as a further check on my result, I can use the price of diesel instead of the price of oil for the subset of countries and years for which it is available. However, countries often subsidize diesel, and this introduces potential reverse causality because countries may subsidize in part to prevent the isolation of hinterland cities. The oil price is a valid instrument for the diesel price, because it is a very strongly predictor and is set on world markets in which no sample country holds sway.[38] In column 5, results for the main specification are similar when the sample is restricted to country-years with a known diesel price. In column 6, the OLS specification using the instrumented diesel price also has a negative and significant semi-elasticity.

[36] This analysis excludes three plants, one in Nigeria and two in Tanzania, characterized as neither thermal nor hydro. All three are part of sugar or paper mills.

[37] Kerosene lamps are even less likely to be driving my results. Data on household electricity are available in 24 Demographic and Health Surveys for nine sample countries during the sample period. Weighting by urban population within and then across countries using data and projections from United Nations (2008), 75 percent of urban households have electricity. A simple unweighted average of the 24 surveys and a compound average of the nine countries after averaging across surveys in each country give 62 and 63 percent, respectively. According to Mills (2002), locally made kerosene lamps produce 5 to 10 lumens, while store-bought models produce 40 to 50 lumens. Electrical light tends to be cheaper than kerosene, so households with electrical connections are unlikely to use kerosene for lighting. A 60-watt incandescent light bulb produces 800 lumens. So even if all households without electricity had the most advanced kerosene lamp and other households had a single 60-watt bulb, only 2 percent of household light would be from kerosene. This is almost certainly a substantial overstatement, because outdoor light is more likely to come from public or commercial establishments that are less likely to use kerosene.

[38] Of course, the oil price also affects the gasoline price, in a very similar way, so the diesel price is best considered as a proxy for diesel and gasoline prices together in this context

5.1.3 Alternate city growth specifications

Table 6 explores alternate time trends. In Column 1 the main effect is about half as large after all city-specific trends are removed from Equation (5), but still significant. This suggests that cities are on different growth paths, with those nearer to the primate growing faster during this period net of transport costs, so that the transport cost effect gets attenuated when they are not accounted for. Column 2 adds a single common trend for all primate cities, the interaction between the *primate* indicator and *year*. It enters negatively and highly significantly, implying that holding constant the transport cost effects, the overall growth trend for primate cities is slower than that for the rest of the sample. This is consistent with Henderson, Storeygard and Weil (2011), who show that summing across sub-Saharan Africa as a whole, primate cities grew more slowly than other areas during the same time period.

5.1.4 Distances to other cities

All results so far have considered transport costs only to the primate city. Transport costs to medium-sized cities may also drive economic activity. Table 7 reports results controlling for transport costs to these other cities. Column 1 repeats column 1 of Table 3. In column 2, route distance to an alternate destination, the nearest city with a 1992 population of at least 100 thousand, has a magnitude comparable to primate distance, but with a much larger standard error.[39] It does not impact the primate distance coefficient substantially. Column 3 refines column 2's specification slightly by only considering this alternate distance in the case of cities whose nearest city of at least 100 thousand is not the primate, to reduce the correlation between the two measures. The results are similar. Columns 4 and 5 are analogous to columns 2 and 3, with the intermediate destination now the nearest city in the top quintile (by 1992 population) of sample cities in the country. The effect of the primate distance is somewhat reduced, but still significant, and the effect of the top quintile city is twice as large or more. This result will be explored further when road surface is considered explicitly in Table 9. Still, no two coefficients on $p_t x_i$ in this table are significantly different from each other.

5.2 Road Surface

The results shown so far have not used the available information on road surface. However, road surface helps to explain under what circumstances transport costs to intermediate cities might matter more than transport costs to the primate. The roads dataset includes

[39] About a third of the cities the sample (94 out of 287) have a 1992 population of at least 100 thousand.

(static) information on road surface type, so each route can be characterized by the fraction of its length that is paved. For simplicity, this measure is converted to an indicator denoting whether a city's route is more paved than the route of the median city in its country. If road surface were randomly assigned, in the short run we might expect a less negative β for the more paved routes, because driving on paved roads is cheaper, in fuel, time, and maintenance costs, than driving on unpaved roads. In a study on South African roads, du Plessis, Visser and Curtayne (1990) find that the fuel efficiency of a 12-ton truck traveling 80 kilometers per hour is 12-13 percent lower on a poor unpaved road (Quarter-car Index, QI=200) than the same truck at the same speed on even a poor paved road (QI=80). This is almost certainly an underestimate, because trucks are unable to maintain high speeds on unpaved roads, and fuel efficiency tends to rise with speed in this range. However, road surface is clearly not randomly assigned, as governments and donors are more likely to pave a road to a city that is economically important or expected to grow.[40] Even if roads were initially assigned randomly, after assignment better-connected places are more able to engage in trade.

The road network of a country can change endogenously, in both an extensive and an intensive sense. On the extensive margin, entirely new roads can be built. While this occasionally happens, the overwhelming majority of road improvements take place in the location of existing roads, because this is so much cheaper that purchasing/appropriating, clearing, and grading new land. In rich countries, it is sometimes the case that limited access roads are built away from the existing route between two cities, because the existing road serves a local purpose that would be destroyed by access limitations. But limited access roads are extremely rare in sub-Saharan Africa outside of South Africa.

The intensive margin is a somewhat thornier problem. Road surfaces can be improved or widened, and they can also deteriorate. However, I expect that the oil price changes in this time period, which include a nominal increase by 760 percent between 1998 and 2008, are large enough that they overwhelm more modest changes in road infrastructure. While the $7 billion annual regional roads investment may be a substantial portion of regional annual GDP, it does not necessarily buy a large length of new or maintained roads. By comparison, China, which has less than half the land area, spent about $45 billion per year between 2000 and 2005 on highways alone (World Bank, 2007b), presumably with higher efficiency.

In Table 8 we see that empirically, hinterland cities with routes to the coastal primate that are more paved than the median route in that country are 0.295 and 0.493 log points larger on average than places with routes less paved, in terms of population and lights,

[40] For an exception, see Gonzalez-Navarro and Quintana-Domeque (2010).

respectively. In Column 3, even after controlling for distance to the primate, more paving is correlated with a larger fraction of adults working in the manufacturing sector, in a sample of districts in 4 countries (Ghana, Guinea, Senegal, and Tanzania) for which census data are available from IPUMS.[41] These results are all consistent with the idea that cities connected by more paved roads could be more hurt by higher oil prices because they are more economically connected to the primate, whereas cities that are connected by mostly unpaved roads are smaller and closer to autarky.

The regressions in Table 9 exploit the paving information by including two terms of the main effect $\beta p_t x_i$ in Equation (5), one for cities with routes to the primate more paved than the median and one for cities with less paved routes. Column 1 demonstrates that the transport cost effect is similar in the two categories of cities. Paving is endogenous to local economic activity, so no causal interpretation is possible here. It is likely that transport costs affect the two sets of cities in slightly different ways. Routes to some cities were paved for any number of reasons (early manufacturing promise, political or military importance, corruption), and then this paving helped these cities to grow more, at least in part because of transport-sensitive firms that were then penalized by increases in oil prices. On the other hand, unpaved roads require slower and more fuel intensive travel, so given the same demand for transport services, cities along them are penalized more per mile.

However, without a mostly paved road to the primate, firms in a city may seek alternate trading connections, relying on intermediate cities instead. Column 2 adds the distance to the nearest city in the top population quintile if that city is not the primate, separately based on the paving status (high or low) of the route to the primate. As in Table 7, higher transport cost to a top quintile city decreases output. However, this effect is limited to cities with relatively unpaved routes to the primate. This suggests that these cities, relatively unconnected to the primate, are essentially consumer cities as in the formulation of Jedwab (2011). Conversely, among cities that are relatively well-connected to the primate, it is the primate distance that matters, not the top quintile city distance. Not surprisingly, the intermediate (top quintile non-primate) cities are themselves 20 percent more likely than other non-primate cities to have their connection to the primate mostly paved.

The results in column 2 are summarized graphically in Figure 10. Three cities, A,

[41] Produced by Ghana Statistical Services (2000), National Statistics Directorate (Guinea; 1983), National Agency of Statistics and Demography (Senegal; 1988) and National Bureau of Statistics (Tanzania; 2002), respectively, and distributed as: Minnesota Population Center. Integrated Public Use Microdata Series, International: Version 6.1 [Machine-readable database]. Minneapolis: University of Minnesota, 2011. https://international.ipums.org/. Accessed 28 April 2011.

B, and C, identical except for their locations, are connected to a primate city P and a secondary (i.e. top quintile) city S, with the distance relationships $d_{SA} = d_{SB} < d_{SC} < d_{PC} = d_{PB} < d_{PA}$. When oil price rise, if roads PA, PB and PC are paved (left figure), A will grow slower than B, which will grow about as fast as C. If these three roads are unpaved (right figure), A and B will grow at the same rate, faster than C.

6 Conclusion

This paper provides evidence that transport costs impact urban economic activity in sub-Saharan Africa, by making access to critical core cities more expensive, with recent increases in oil prices easily removing several percentage points from the size of far hinterland cities in countries where the largest city is on the coast. This is consistent with a simple model in which hinterland farmers are constrained to buy fewer manufactured goods from the core when transport costs are high, and also with a core-periphery model in which all manufacturing has clustered in the primate city. Despite being larger and likely facing smaller absolute changes in costs, cities with more paved routes are no more or less sensitive to changing transport costs, most likely because they are more integrated with national and global markets. However, cities with less paved routes seem to be less affected by transport costs to the primate city than they are by transport costs to a nearer secondary city.

While previous work has shown that improvements in transport infrastructure can increase local activity and growth, most of it is based on very large construction projects, and none has been in an African context where industry is highly concentrated. The nature of the variation in the current work, provided simply by changes in oil prices interacted with distance, means that the results are unlikely to be driven by changes in long term investment in non-transport sectors. Instead, they provide clearer evidence of the direct short run effect of transport costs on urban economic activity.

Annual city-level measures of economic activity provide evidence net of the country-year level variation used in previous comprehensive work on urbanization, urban growth, and coastal access in sub-Saharan Africa. More generally, this city-level variation opens up exciting new possibilities for future research.

References

Ades, Alberto F. and Edward L. Glaeser (1995) "Trade and Circuses: Explaining Urban Giants," *The Quarterly Journal of Economics*, Vol. 110, No. 1, pp. 195–227, February.

Anderson, James E. and Eric van Wincoop (2004) “Trade Costs,” *Journal of Economic Literature*, Vol. 42, No. 3, pp. 691–751, September.

Atack, Jeremy, Fred Bateman, Michael Haines, and Robert A. Margo (2010) “Did Railroads Induce or Follow Economic Growth?: Urbanization and Population Growth in the American Midwest, 1850-1860,” *Social Science History*, Vol. 34, No. 2, p. 171.

Balk, Deborah, Francesca Pozzi, Gregory Yetman, Uwe Deichmann, and Andy Nelson (2004) “The Distribution of People and the Dimension of Place: Methodologies to Improve the Global Estimation of Urban Extents,” working paper, CIESIN, Columbia University.

Banerjee, Abhijit, Esther Duflo, and Nancy Qian (2009) “On the Road: Access to Transportation Infrastructure and Economic Growth in China,” mimeo, MIT.

Barrios, Salvador, Luisito Bertinelli, and Eric Strobl (2006) “Climatic change and rural-urban migration: the case of sub-Saharan Africa,” *Journal of Urban Economics*, Vol. 60, No. 3, pp. 357–371.

Baum-Snow, Nathaniel (2007) “Did Highways Cause Suburbanization?” *The Quarterly Journal of Economics*, Vol. 122, No. 2, pp. 775–805, 05.

Behrens, Kristian, Carl Gaigné, Gianmarco I.P. Ottaviano, and Jacques-François Thisse (2006) “Is remoteness a locational disadvantage?” *Journal of Economic Geography*, Vol. 6, No. 3, pp. 347–368.

Buys, Piet, Uwe Deichmann, and David Wheeler (2010) “Road network upgrading and overland trade expansion in sub-Saharan Africa,” *Journal of African Economies*, Vol. 19, No. 3, pp. 399–432.

Chandra, Amitabh and Eric Thompson (2000) “Does public infrastructure affect economic activity?: Evidence from the rural interstate highway system,” *Regional Science and Urban Economics*, Vol. 30, No. 4, pp. 457–490, July.

Collier, Paul (2007) *The bottom billion: why the poorest countries are failing and what can be done about it*, Oxford: Oxford University Press.

Conley, Timothy G. (1999) “GMM estimation with cross sectional dependence,” *Journal of Econometrics*, Vol. 92, No. 1, pp. 1–45, September.

Coulibaly, Souleymane and Lionel Fontagné (2006) "South–South Trade: Geography Matters," *Journal of African Economies*, Vol. 15, No. 2, pp. 313–341, June.

Deutsche Gesellschaft für Technische Zusammenarbeit (2009) *International Fuel Prices 2009*, Eschborn, Germany: Deutsche Gesellschaft für Technische Zusammenarbeit, 6th edition.

Donaldson, Dave (2010) "Railroads of the Raj: Estimating the Impact of Transportation Infrastructure," NBER Working Papers 16487, National Bureau of Economic Research, Inc.

Duranton, Gilles and Matthew A. Turner (2011) "Urban growth and transportation," mimeo, University of Toronto.

Elvidge, Christopher D., Kimberly E. Baugh, Eric A. Kihn, Herbert W. Kroehl, and Ethan R. Davis (1997) "Mapping of city lights using DMSP Operational Linescan System data," *Photogrammetric Engineering and Remote Sensing*, Vol. 63, No. 6, pp. 727–734.

Elvidge, Christopher D., Jeffrey M. Safran, Ingrid L. Nelson, Benjamin T. Tuttle, V. Ruth Hobson, Kimberly E. Baugh, John B. Dietz, and Edward H. Erwin (2004) "Area and position accuracy of DMSP nighttime lights data," in Ross S. Lunetta and John G. Lyon eds. *Remote Sensing and GIS Accuracy Assessment*, Boca Raton, FL: CRC Press, pp. 281–292.

Elvidge, Christopher D., Daniel Ziskin, Kimberly E. Baugh, Benjamin T. Tuttle, Tilottama Ghosh, Dee W. Pack, Edward H. Erwin, and Mikhail Zhizhin (2009) "A Fifteen Year Record of Global Natural Gas Flaring Derived from Satellite Data," *Energies*, Vol. 2, No. 3, pp. 595–622.

Fay, Marianne and Charlotte Opal (2000) "Urbanization without growth: a not-so-uncommon phenomenon," Policy Research Working Paper Series 2412, The World Bank.

Fujita, Masahisa, Paul R. Krugman, and Anthony Venables (1999) *The Spatial Economy*, Cambridge, MA: MIT Press.

Gallup, John Luke, Jeffrey Sachs, and Andrew Mellinger (1999) "Geography and economic development," *International Regional Science Review*, Vol. 22, No. 2, pp. 179–232.

Gollin, Douglas and Richard Rogerson (2010) "Agriculture, Roads, and Economic Development in Uganda," NBER Working Papers 15863, National Bureau of Economic Research, Inc.

Gonzalez-Navarro, Marco and Climent Quintana-Domeque (2010) "Street Pavement: Results from an Infrastructure Experiment in Mexico," Working Papers 1247, Princeton University, Department of Economics, Industrial Relations Section.

Gwilliam, Kenneth M. (2011) *Africa's transport infrastructure*, Washington, DC: World Bank.

Henderson, Vernon (2002) "Urban primacy, external costs, and quality of life," *Resource and Energy Economics*, Vol. 24, No. 1-2, pp. 95–106, February.

Henderson, Vernon, Adam Storeygard, and David N. Weil (2011) "A Bright Idea for Measuring Economic Growth," *American Economic Review*, Vol. 101, No. 3, pp. 194–99, May.

——— (forthcoming) "Measuring Economic Growth from Outer Space," *American Economic Review*.

Hoyle, Brian and Jacques Charlier (1995) "Inter-port competition in developing countries: an East African case study," *Journal of Transport Geography*, Vol. 3, No. 2, pp. 87–103.

Jedwab, Remi (2011) "African Cities and the Structural Transformation: Evidence from Ghana and Ivory Coast," mimeo, Paris School of Economics.

Krugman, Paul R. (1991) "Increasing returns and economic geography," *Journal of Political Economy*, Vol. 99, No. 3, pp. 483–499.

Limão, Nuno and Anthony Venables (2001) "Infrastructure, geographical disadvantage, transport costs, and trade," *The World Bank Economic Review*, Vol. 15, No. 3, pp. 451–479.

Mills, Evan (2002) ""The $230-billion Global Lighting Energy Bill," in *Proceedings of the 5th International Conference on Energy-Efficient Lighting*.

National Bureau of Statistics (2009) *Tanzania Statistical Abstract 2006*, Dar es Salaam: National Bureau of Statistics.

National Research Council (2003) *Cities Transformed: Demographic Change and its Implications in the Developing World*, Washington, DC: The National Academies Press.

Persits, Feliks, Thomas Ahlbrandt, Michele Tuttle, Ronald Charpentier, Michael Brownfield, and Kenneth Takahashi (2002) "Map Showing Geology, Oil and Gas Fields, and Geologic Provinces of Africa, version 2," Open File Report 97-470A, version 2.0, United States Geological Survey.

Pholo Bala, Alain (2009) "Gates, hubs and urban primacy in Sub-Saharan Africa," CORE Discussion Papers 2009039, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE).

du Plessis, H. W., A. T. Visser, and P. C. Curtayne (1990) "Fuel Consumption of vehicles as Affected by Road-Surface Characteristics," in Wolfgang E. Meyer and Jean Reichert eds. *Surface characteristics of roadways: international research and technologies*, Philadelphia: American Society for Testing and Materials, pp. 480–496.

Teravaninthorn, Supee and Gael Raballand (2009) *Transport prices and costs in Africa: a review of the main international corridors*, Washington, DC: World Bank.

United Nations (2008) *World Urbanization Prospects: The 2007 Revision*, New York: United Nations.

World Bank (2007a) *A Decade of Action in Transport: An Evaluation of World Bank Assistance to the Transport Sector, 1995-2005*, Washington, DC: World Bank.

——— (2007b) "An Overview of China's Transport Sector - 2007," EASTE working paper 15, Transport, Energy and Mining Sector Unit, Sustainable Development Department, East Asia and Pacific Region, The World Bank.

——— (2010) *Africa's infrastructure: a time for transformation*, Washington, DC: World Bank.

Tables and Figures

Table 1: Relationship between lights and economic activity in subnational units of developing countries

	(1) Δ ln(GDP)	(2) Δ ln(GDP)	(3) Δ ln(HH income)
Δ ln(light)	0.265**	0.268**	0.171***
	[0.101]	[0.119]	[0.0573]
Observations	204	146	284
R-squared	0.054	0.053	0.028
unit	city proper	prefecture	magisterial district
sample	China	China	South Africa
years	1990/92-2005	1990/92-2005	1996-2001

Each column is a separate OLS long difference regression for the years shown. The independent variable is the log of the lights digital number, summed across all pixels in the unit shown, and averaged across satellite-years within a year when applicable. In columns 1 and 2, the lights are from 1992 but the administrative GDP data are from 1990, the closest year with good data. In column 3, household incomes are aggregated are summed from a 5 percent sample of the census, and topcoded households are assigned a value 1.5 times the topcode value. Robust standard errors are reported in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 2: Descriptives

Variable	*Mean*	*StdDev*	*Min*	*Max*
lights digital number (DN)	3153	10992	0	152239
ln(light+5.5)	6.279	1.908	1.705	11.93
distance(Primate)	560.7	464.9	0	2360
distance(pop=100k)	136.7	168	0	898.7
distance(poptop20%)	162.3	169.7	0	1166
fraction of path to Primate paved	.682	.26	.115	1

The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. All distances are measured in kilometers. Distance(Primate), distance(pop=100k), and distance(poptop20%) are the road network distance to the largest city in the country, the nearest city (in the same country) with a population of at least 100 thousand, and the nearest city in the top quintile of the country's 1992 city population distribution, respectively.

Table 3: Main results

	(1) ln(light)	(2) ln(light)
distance(Primate)*P_{oil}	-0.683***	
	[0.223]	
distance(Primate)*$ln(P_{oil})$		-0.344***
		[0.108]
Observations	4,879	4,879
left censored cases	263	263

Each column is a separate tobit regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Distance(Primate) is the road network distance to the largest city in the country, measured in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 4: Robustness of functional form

	(1)	(2)	(3)	(4)
	ln(light+5.5)	ln(light)	ln(light)	ln(light)
distance(Primate)*P_{oil}	-0.616***		-0.653***	-0.640***
	[0.215]		[0.213]	[0.207]
ln(distance(Primate))*$ln(P_{oil})$		-0.0866***		
		[0.0194]		
Observations	4,879	4,879	4,879	4,879
R-squared	0.598			
sample	all	all	all	all
model	OLS	tobit	tobit	tobit
left censored cases		263	263	263
dirt factor			1.5	2

Each column is a separate regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. Dirt factor is the ratio of the time required to traverse a given length of dirt road and the time required to traverse the same length of paved road, used in calculating shortest routes. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 5: Exclusion of other channels by which oil prices could be affecting lights

	(1) ln(light)	(2) ln(light)	(3) ln(light)	(4) ln(light)	(5) ln(light)	(6) ln(light+5.5)
distance(Primate)*P_{oil}	-0.626***	-0.606**	-0.616***	-0.699***	-0.448**	
	[0.243]	[0.255]	[0.231]	[0.263]	[0.180]	
distance(Primate)*P_{oil}*1(hydro closest)			0.0810			
			[0.270]			
distance(Primate)*P_{diesel}						-0.505***
						[0.126]
Observations	4,471	4,029	4,267	3,128	2,226	2,226
R-squared						0.579
sample	oil>50km	oil>100km	power plant	electrified	diesel	diesel
model	tobit	tobit	tobit	tobit	tobit	IV:oil-unclustered
left censored cases	247	216	230	127	87	

Each column is a separate regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year. Distance(Primate) is the road network distance to the largest city in the country, in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. P_{diesel} is the price of diesel in the county's capital city in November of the given year, in dollars per liter. In columns 1 and 2, cities at least partially within 50 and 100 km, respectively, of an oil well are excluded. Column 3 includes only cities in countries with both hydro and other power plants. 1(hydro closest) is a dummy indicating that the nearest power plant to the city is a hydro plant. Column 4 includes only cities with electric transmission lines passing through them. Columns 5 and 6 are limited to country-years for which P_{diesel} is available. In column 6, P_{oil} is the instrument for P_{diesel}. The tobit cutoff is light=5.5. Robust standard errors, clustered by city except in column 6, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 6: Results with alternate timing assumptions

	(1) ln(light)	(2) ln(light)
distance(Primate)*P_{oil}	-0.302**	-0.396***
	[0.139]	[0.140]
primate*year		-0.0338***
		[0.00909]
Observations	4,879	4,879
left censored cases	263	263

Each column is a separate tobit regression that includes country*year and city fixed effects. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Distance(Primate) is the road network distance to the largest city in the country, measured in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 7: Distances to other cities

	(1) ln(light)	(2) ln(light)	(3) ln(light)	(4) ln(light)	(5) ln(light)
distance(Primate)*P_{oil}	-0.683***	-0.623***	-0.573**	-0.503**	-0.443*
	[0.223]	[0.226]	[0.225]	[0.228]	[0.237]
distance(pop=100k)*P_{oil}		-0.535			
		[0.501]			
distancepop=100k)*P_{oil}*1(nearest pop=100k isn't primate)			-0.888		
			[0.736]		
distance(poptop20%)*P_{oil}				-1.205**	
				[0.485]	
distance(poptop20%)*P_{oil}*1(nearest poptop20% isn't primate)					-1.240**
					[0.610]
Observations	4,879	4,879	4,879	4,879	4,879
left censored cases	263	263	263	263	263

Each column is a separate tobit regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Distance(Primate), distance(pop=100k), and distance(poptop20%) are the road network distance to the largest city in the country, the nearest city (in the same country) with a population of at least 100 thousand, and the nearest city in the top quintile of the country's 1992 city population, respectively. Distances are measured in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. The dummy variables interacted with these distances in columns (3) and (5) are one if the nearest large city (of 100k in column 3, or in the top quintile in column 5) is not the primate. The tobit cutoff is light=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 8: Paving and city size

	(1) ln(population)	(2) ln(light+5.5)	(3) fraction manufacturing
1(paving > median)	0.295***	0.463**	0.0151***
	[0.104]	[0.183]	[0.00521]
distance(Primate)			-0.0196**
			[0.00943]
Observations	4,624	4,624	293
R-squared	0.244	0.176	0.506
unit	city	city	district
sample	coastal primate	coastal primate	IPUMS

Each column is a separate OLS regression, with country fixed effects. The primary independent variable is a dummy indicating that the unit's path to its country's primate city is more paved than the average within that country. In columns 1 and 2, the unit of analysis is the city-year, for a balanced annual panel of 272 non-primate cities in 15 coastal primate countries over the period 1992–2008. In column 1, the dependent variable is the log of population, while in column 2 it is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. In column 3, the sample is census administrative units in Ghana (2000), Guinea (1983), Senegal (1988), and Tanzania (2002), and the dependent variable is fraction of the employed population over age 10 working in manufacturing. Distance(Primate) is the road network distance to the largest city in the country, measured in thousands of kilometers. Robust standard errors, clustered by city except in column 3, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

Table 9: Results by paving status

	(1)	(2)
	ln(light)	ln(light)
dist(primate)*P_{oil}*1(primate route paving < median)	-0.693***	-0.272
	[0.247]	[0.283]
dist(primate)*P_{oil} *1(primate route paving > median)	-0.663***	-0.671**
	[0.255]	[0.276]
dist(poptop20%)*P_{oil}*1(primate route paving < median)*1(nearest poptop20% not primate)		-1.889**
		[0.820]
dist(poptop20%)*P_{oil}*1(primate route paving > median)*1(nearest poptop20% not primate)		0.116
		[1.031]
Observations	4,879	4,879
left censored cases	263	263

Each column is a separate tobit regression that includes country*year and city fixed effects, and city-specific linear time trends. The unit of analysis is the city-year, for a balanced annual panel of 287 cities in 15 coastal primate countries over the period 1992–2008. The dependent variable is the log of the lights digital number, summed across all pixels in the city, and averaged across satellite-years within a year when applicable. Dist(Primate) and dist(poptop20%) are the road network distance to the largest city in the country and the nearest city (in the same country) in the top population quintile, respectively. Distances are measured in thousands of kilometers. P_{oil} is the price of oil (specifically the annual average Europe Brent Spot Price FOB) in hundreds of dollars per barrel. 1(primate route paving < median) is a dummy indicating that a city's route to the primate is more paved than the route of the median city in that country to the primate. 1(nearest poptop20% not primate) indicates that the nearest city in the top quintile is not the primate. The tobit cutoff is DN=5.5. Robust standard errors, clustered by city, are in brackets. *, **, *** mean significance at the ten, five, and one percent level, respectively.

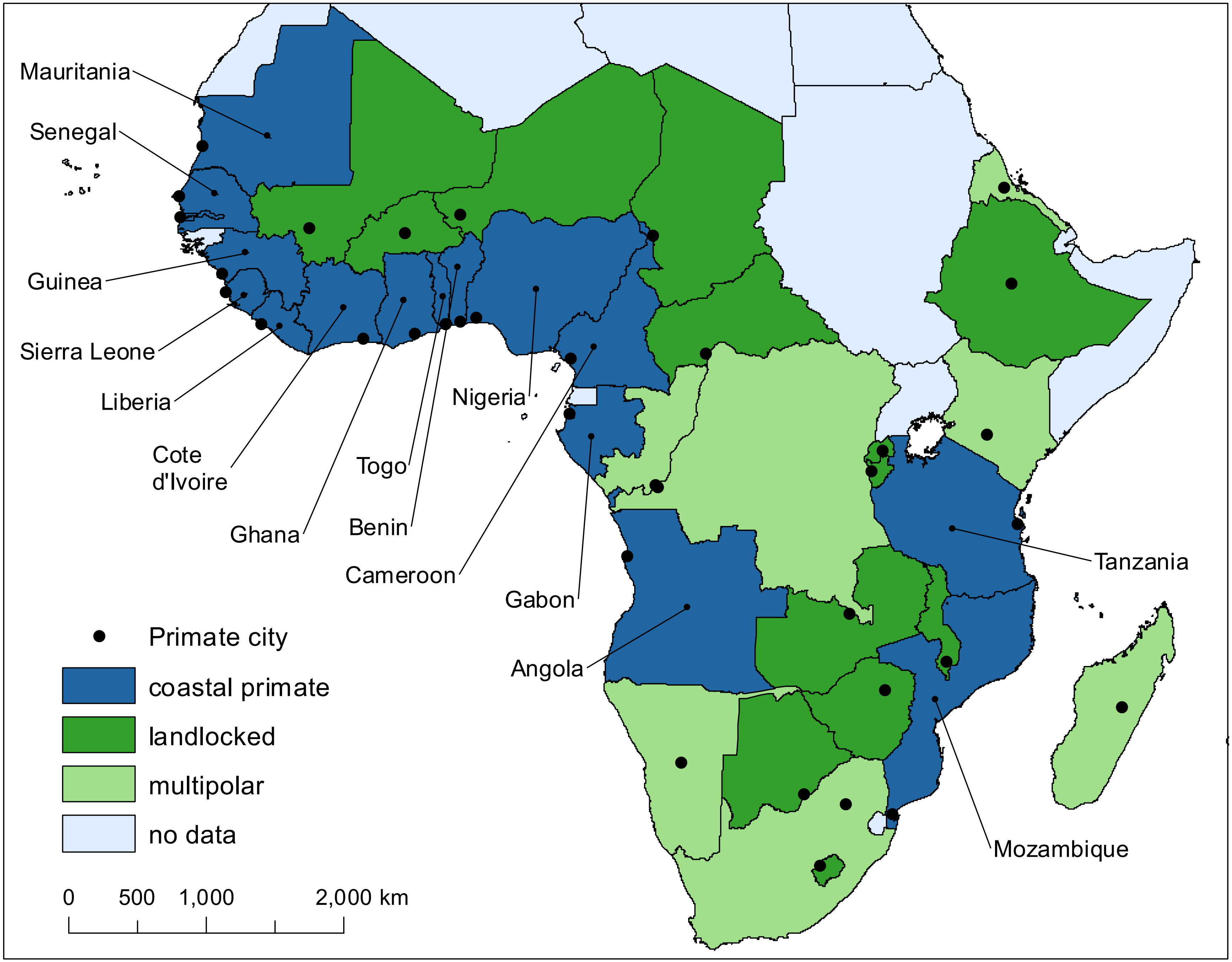

Figure 1: African countries classified by their coastal access, the coastal access of their primate city, and availability of comparable roads and lights data

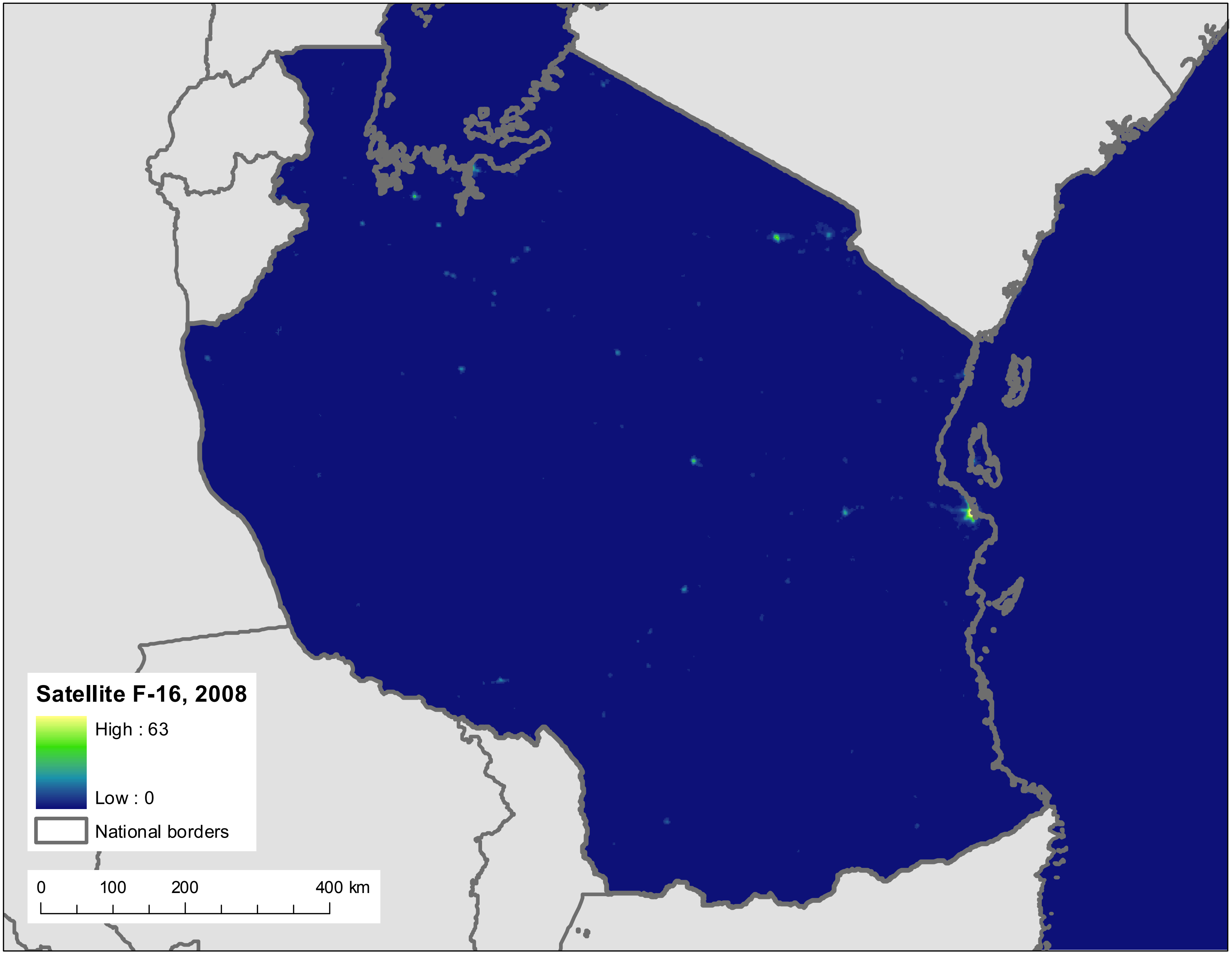

Figure 2: Lights digital number in Tanzania from satellite F-16, 2008

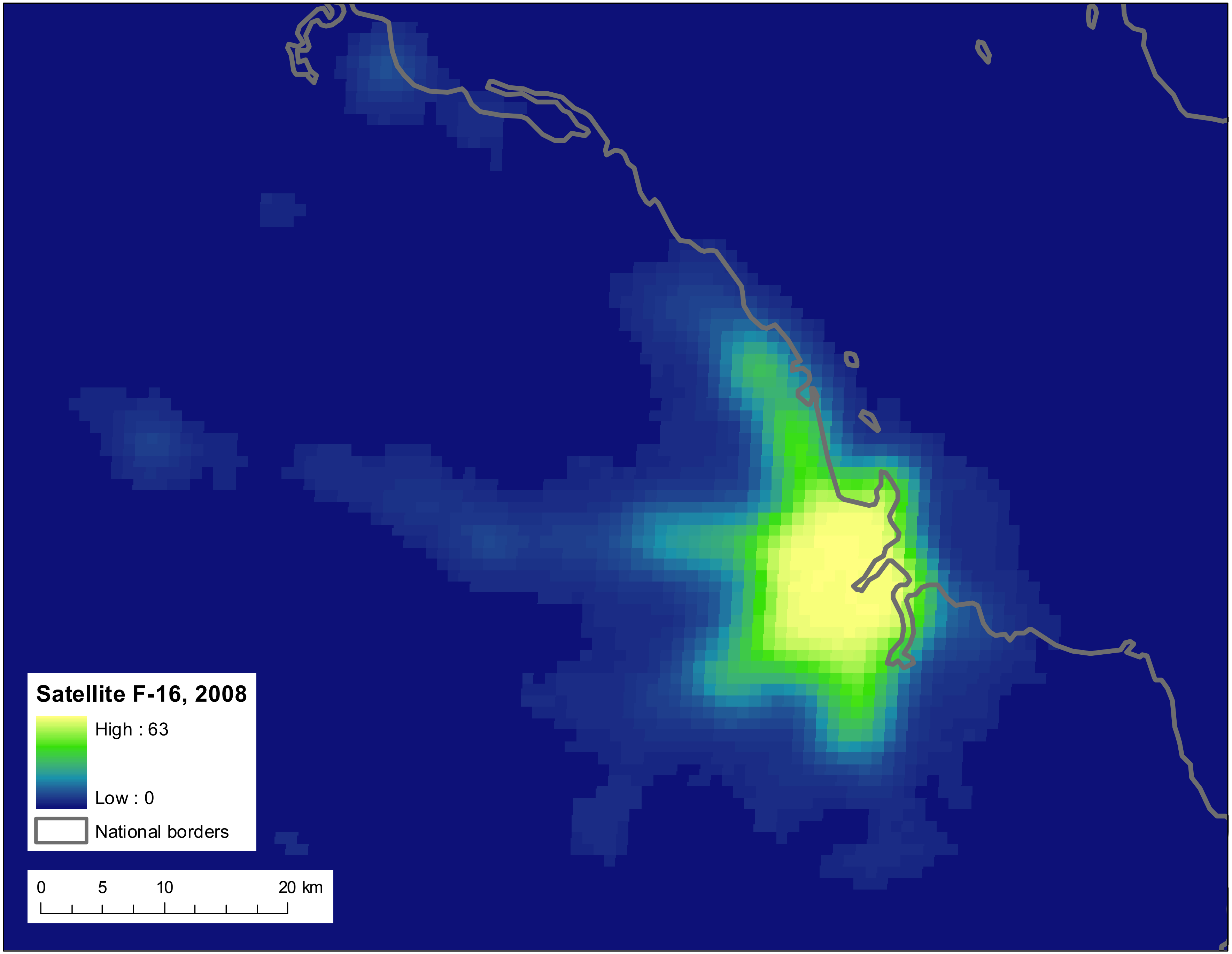

Figure 3: Lights digital number in and around Dar es Salaam, Tanzania from satellite F-16, 2008

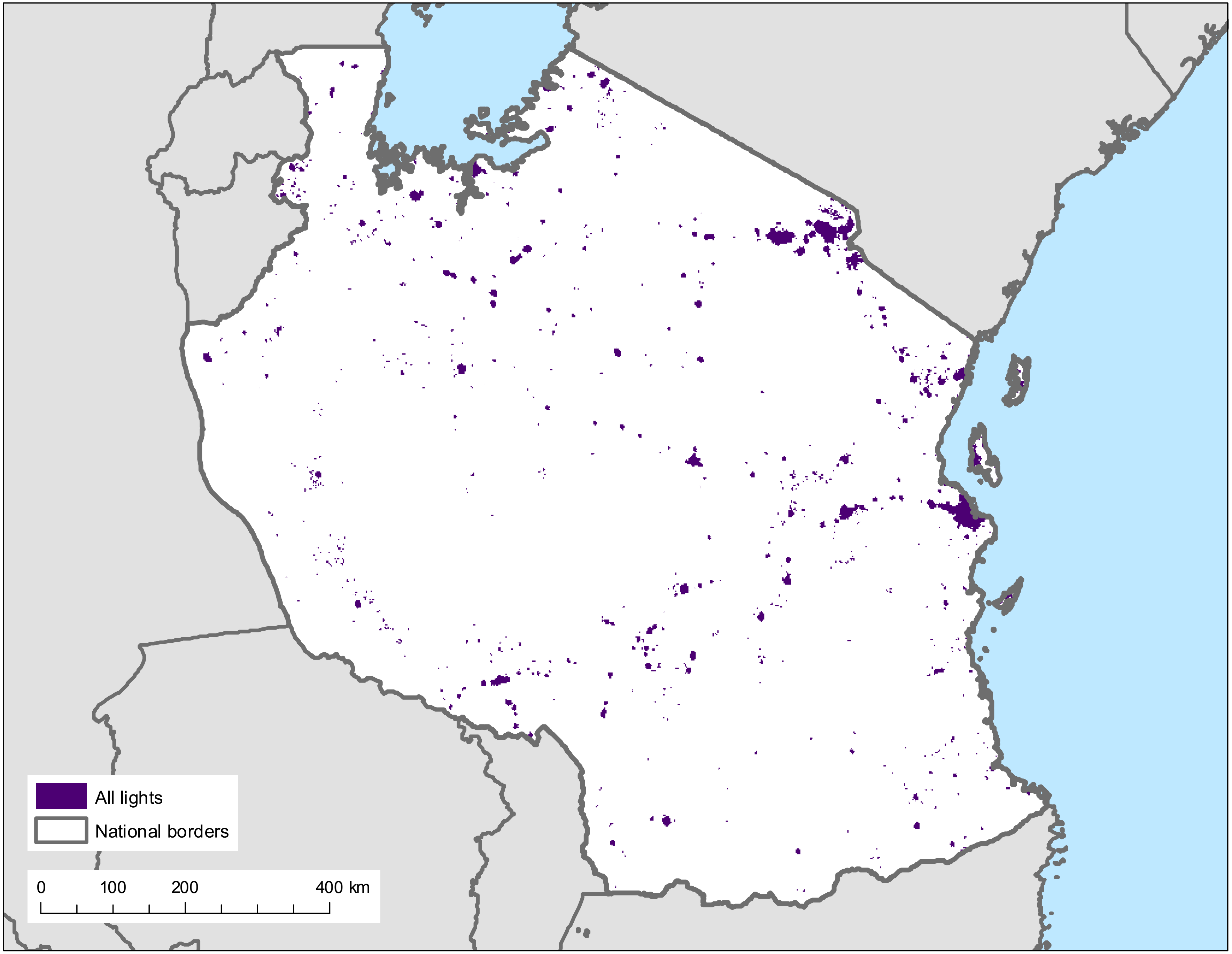

Figure 4: Binary lights (pixels lit in at least one satellite-year, 1992-2008), Tanzania

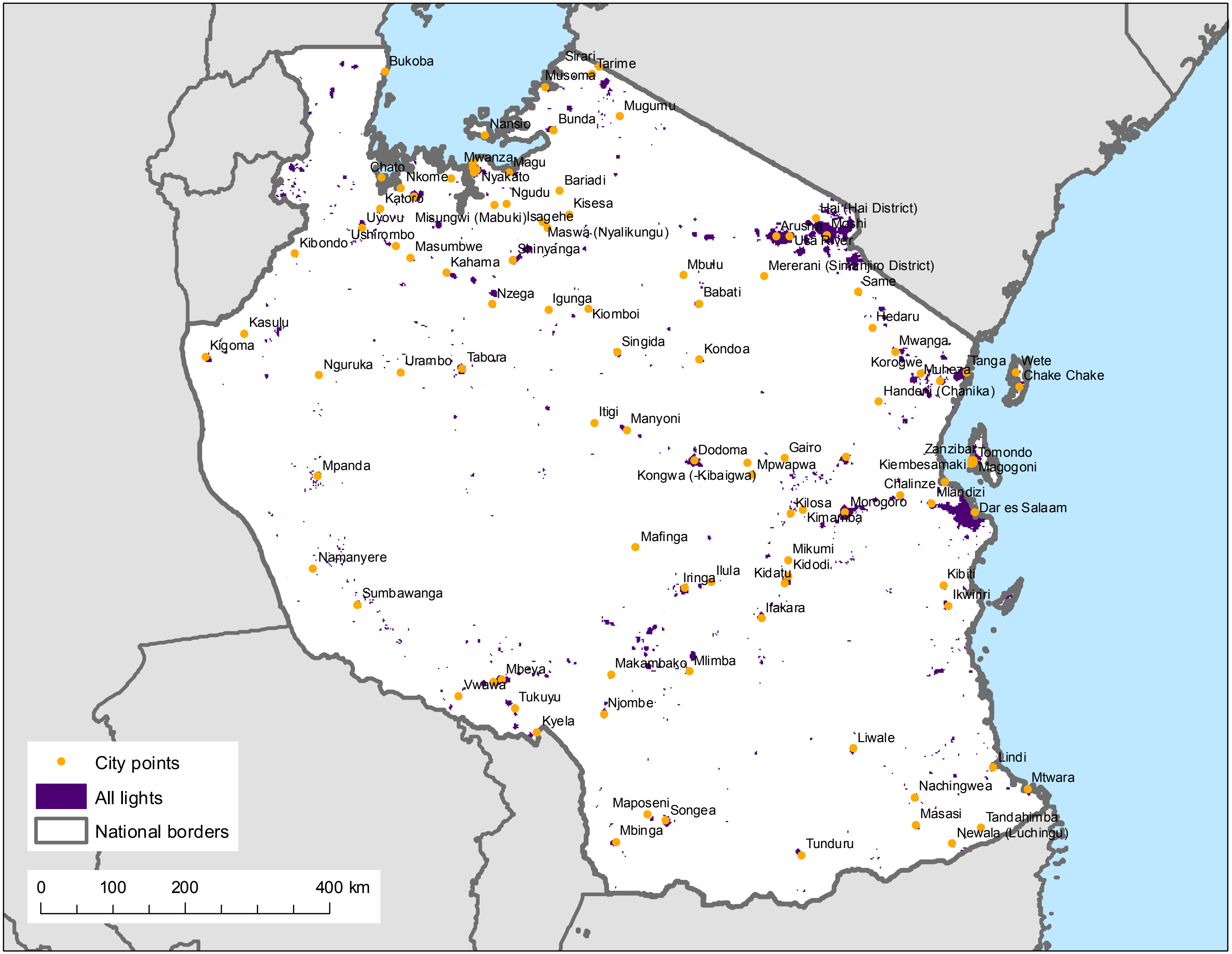

Figure 5: Binary lights and cities with known populations, Tanzania

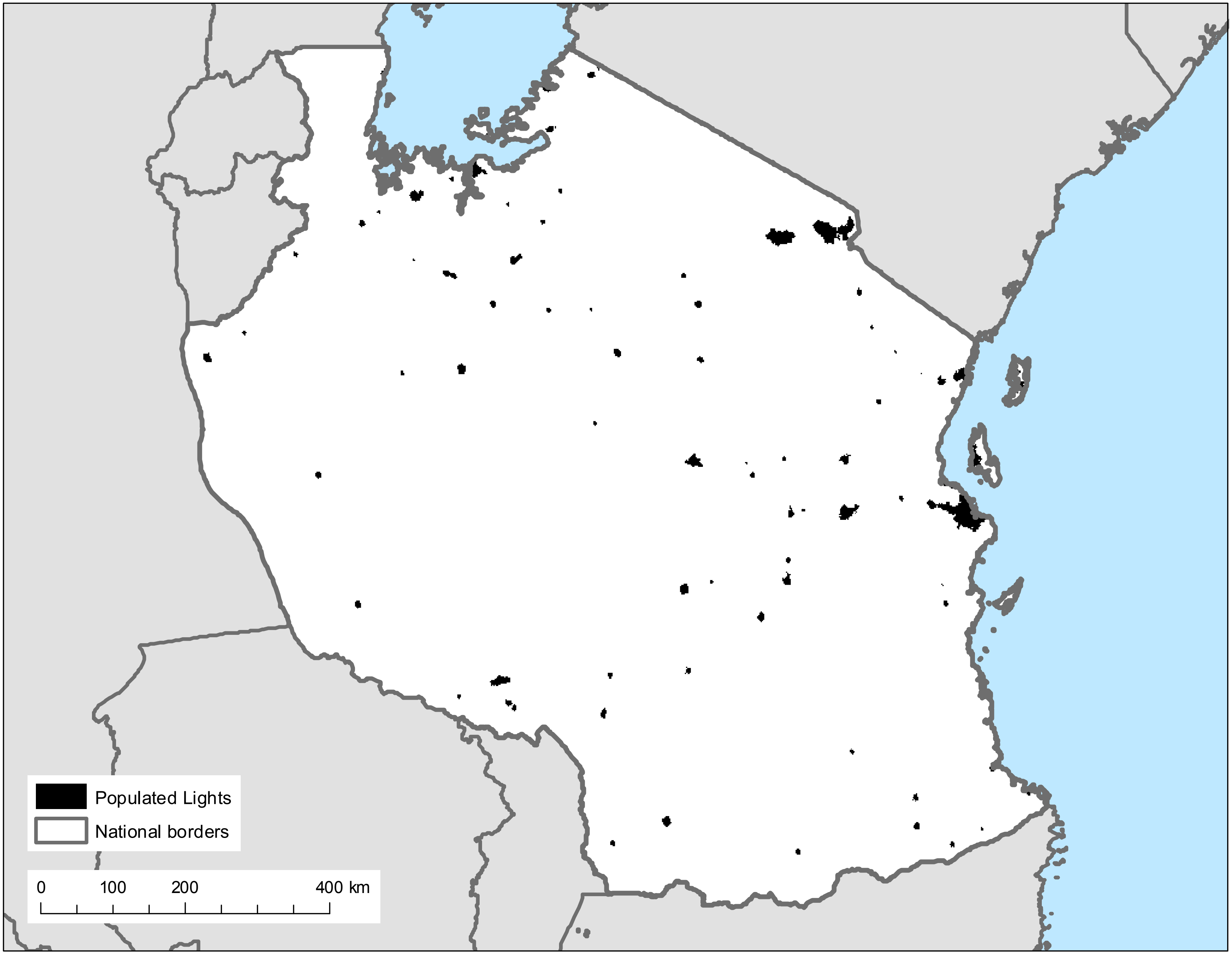

Figure 6: Binary lights, restricted to cities with known populations, Tanzania

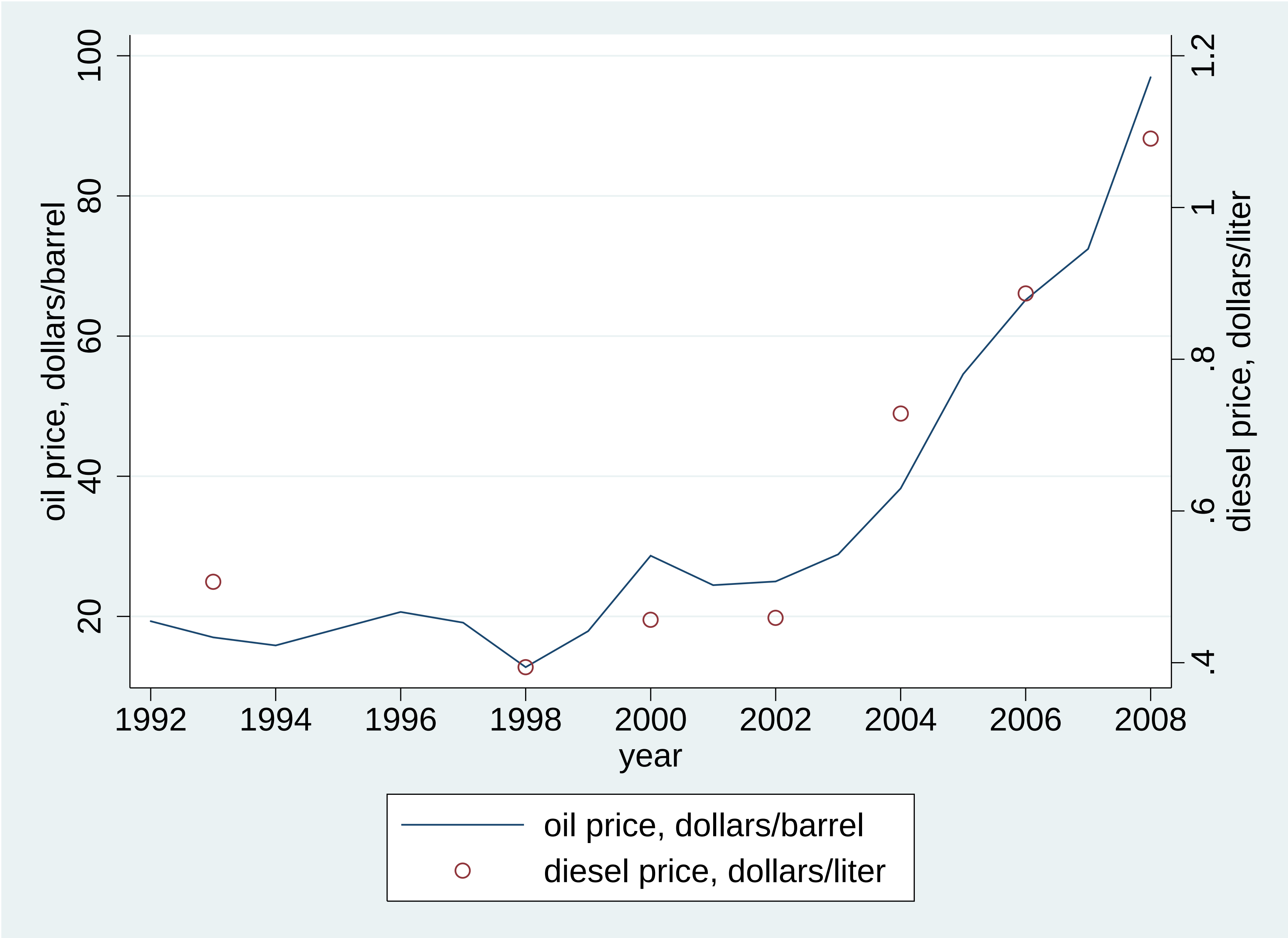

Figure 7: Oil and diesel prices (averaged across the 12 countries in the sample with data for all 7 years shown), 1992–2008. Diesel prices were surveyed in November, while oil prices are averaged across the whole year.

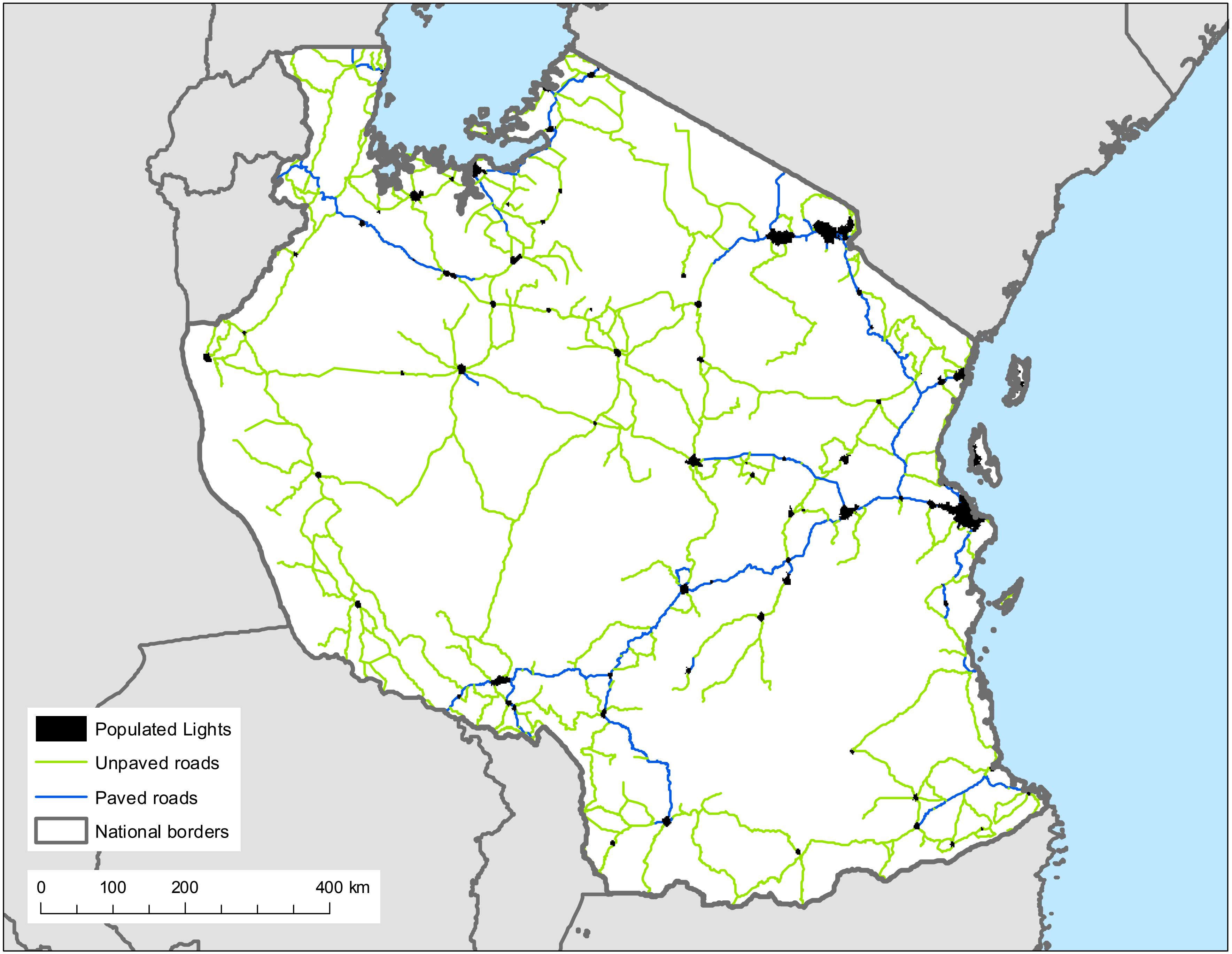

Figure 8: Roads in Tanzania

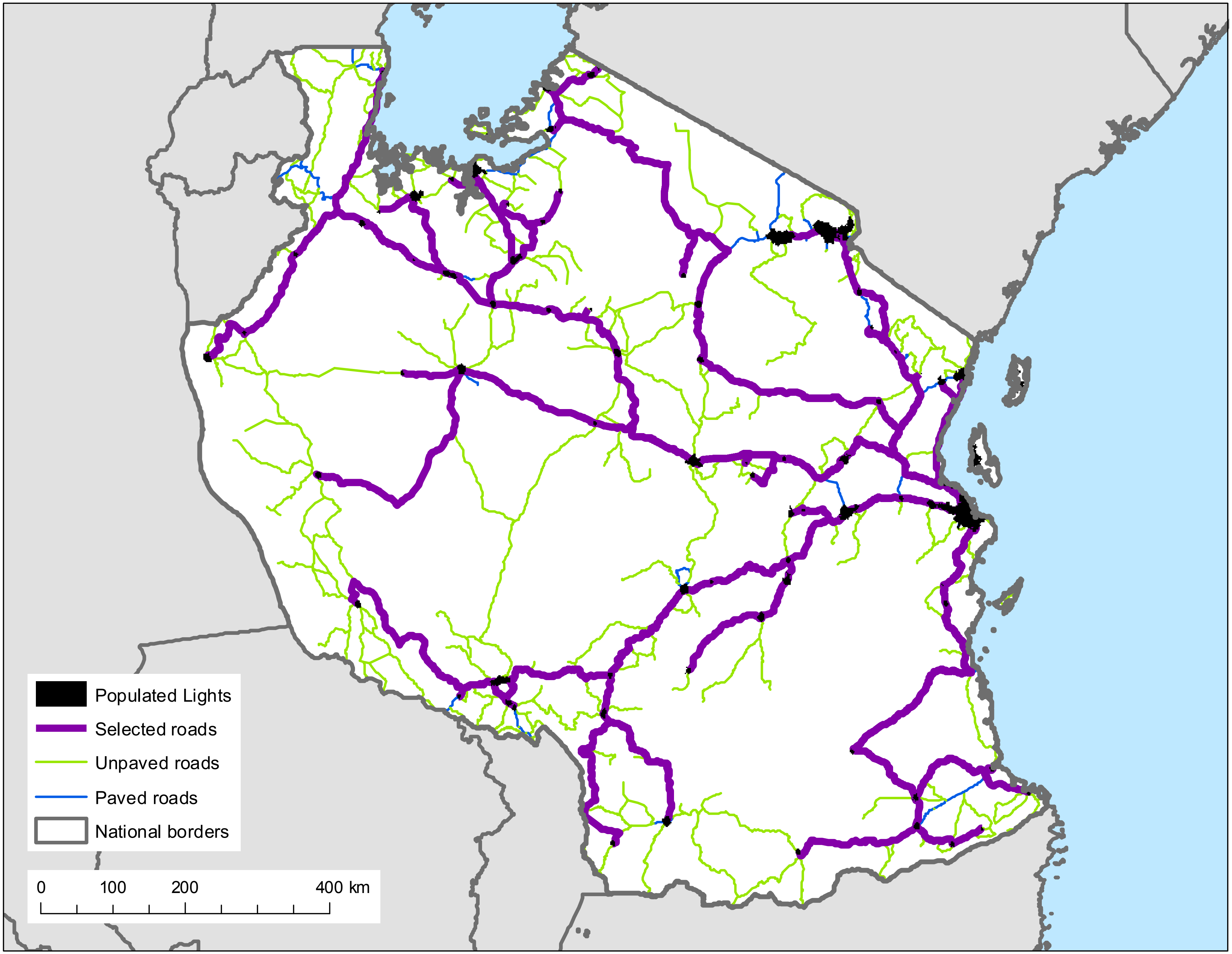

Figure 9: Shortest road routes from cities with known populations to Dar es Salaam, Tanzania

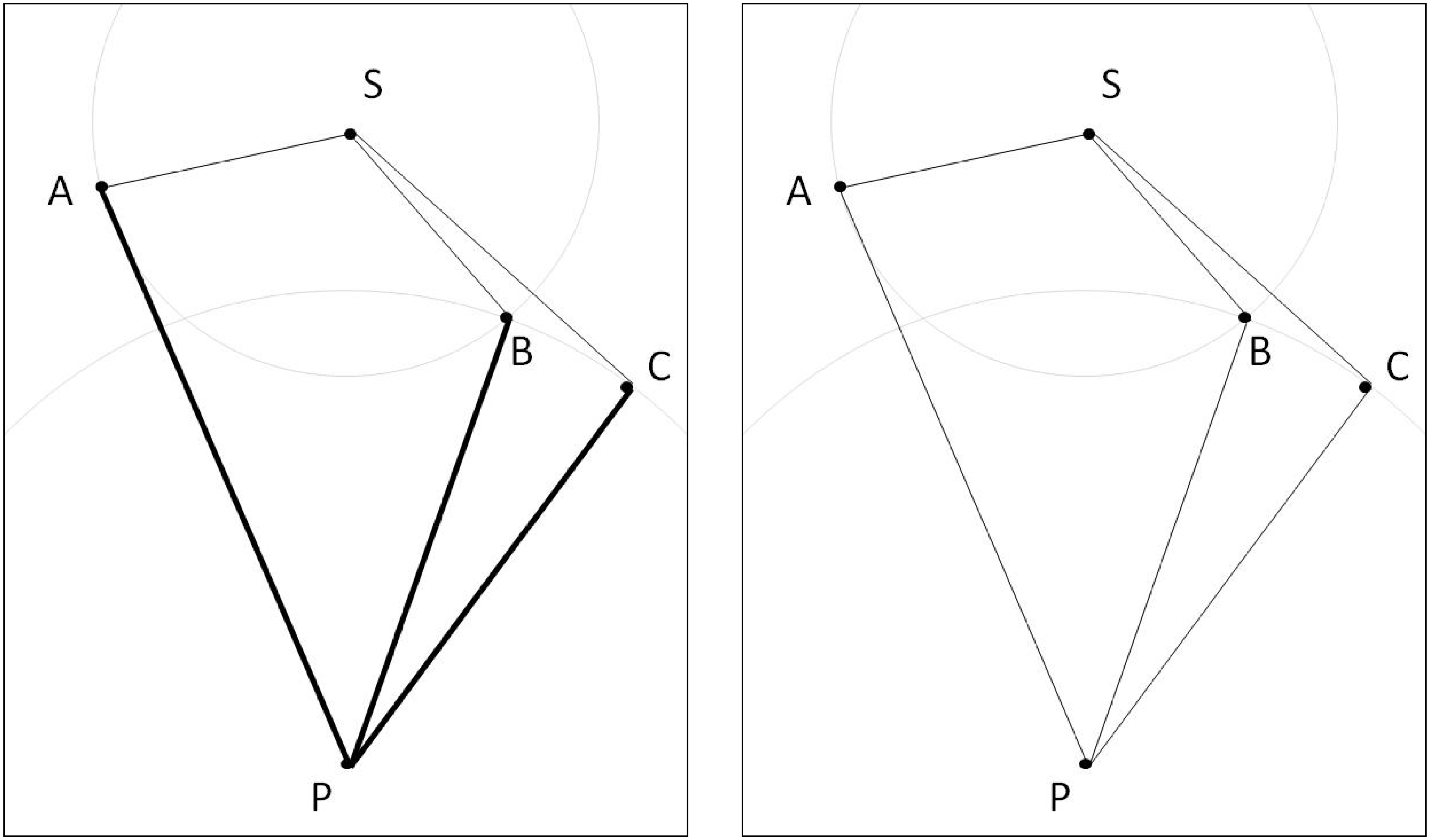

Figure 10: Diagram of the paving results. Three cities, A, B, and C, identical except for their locations, are connected to a primate city P and a secondary (i.e. top quintile) city S, with the distance relationships $d_{SA} = d_{SB} < d_{SC} < d_{PC} = d_{PB} < d_{PA}$. When oil price rise, if roads PA, PB and PC are paved (left figure), A will grow slower than B, which will grow about as fast as C. If these three roads are unpaved (right figure), A and B will grow at the same rate, faster than C.

A Data and spatial methods

A.1 City points

City locations (latitude and longitude) and census populations were collected from the City Population website and spot-checked with official sources where available.[42] Using city-specific growth rates based on multiple censuses where available, or national urban growth rates from United Nations (2008) otherwise, I generated a population for all years for each city.

For 14 of the 15 countries in the ultimate sample, City Population claims to list all cities above a given (country-specific) population, typically 5,000, 10,000, or 20,000.[43] However, it does not explicitly cite the year for which this claim is made. Of the 738 cities with location and population information in these 15 countries, 9 city points fall below this cutoff for all years 1990 to 2008. These are included in the sample until explicit population cuts are made.

A.2 Lights

The lights data are described in Henderson, Storeygard and Weil (forthcoming). For the present study the 30 satellite-years of lights data were first combined into one binary grid encoding whether a pixel was lit in at least one satellite-year. The resulting contiguous ever-lit areas were converted into polygons, and split by national borders. Only lights within 3 kilometers of one of the city points described above with a known census population were kept in the sample. The 3-km buffer is used because of georeferencing error in both the points and the lights (Balk et al., 2004; Elvidge et al., 2004). While some of the other lights are likely to be small settlements, some are noise from the sensor or from fires lasting for too long to be excluded by the data cleaning algorithm, mines or other facilities. In general, they are smaller and weaker lights as well.

The resulting sample is 485 lights in 15 countries. Populations for each point were summed across all points assigned to each light. In 49 lights, more than one city was present; in 26 of these, exactly two were present. In 13 cases, a point fell within 3 km of multiple lights. In such cases, the point's population was only retained by the light to which it was closest. The light with the largest 1992 population within each country was designated the primate. In most countries, this corresponded to the historical

[42]http://www.citypopulation.de/. In 9 cases where coordinates or population were unavailable from City Population, coordinates from Google Earth or World Gazetteer (http://www.world-gazetteer.com) were used.

[43]For Angola, there is no explicit cutoff.

political capital. The only exception is Douala, Cameroon, which is larger than Yaounde. The historical political capitals that are not current formal political capitals are Dar es Salaam, Tanzania, which was replaced by Dodoma, Abidjan, Côte d'Ivoire, replaced by Yamoussoukro, and Lagos, Nigeria, replaced by Abuja.

A.3 Roads

The African Infrastructure Country Diagnostic database contains comparable roads data for all countries of the African mainland with no Mediterranean coastline, plus Madagascar, except for Djibouti, Equatorial Guinea, Guinea-Bissau, and Somalia.[44] Two datasets were produced for Sudan: one for the north and one for the south. The datasets generally have comparable metadata on road surface, quality, and hierarchy (primary/secondary/tertiary), as well as estimates of traffic. Still before they could be used for the current analysis, several changes had to be made. All steps below were carried out using ArcGIS 9.3 software, except for some tabular cleaning done in Excel and Stata. Nearly all steps in ArcGIS, and all steps in Stata, were automated in Python, Arc Macro Language (AML), or Stata scripts.

The roads data were cleaned tabularly, to ensure that the relevant fields were coded consistently, and projected to a sinusoidal projection, which is more conducive to distance calculations than their native plate carrée (latitude and longitude).[45] Next, roads from all countries were combined into one large dataset, and a topology was built with the rule "no dangles". This means that every dead end was flagged. In most cases, dead ends are legitimate features of the road network. In other cases, however, they are artifacts of a data generation process in which some segments that are connected in the real world are not connected in the dataset. This is critical in the network analysis to follow.

Problematic dangles were fixed in several ways. First, using the topology "Extend" tool, dangles were extended up to 100 meters if that would cause them to no longer be dangles. In theory, the topology "Trim" tool could be used for the opposite task to dangles less than 100 meters long. However, a bug in ArcGIS made this infeasible. But extra dangles only affect final results to the extent that they cause additional "spiders" to be created (see below).

The Extend operation does not close all gaps of less than 100 meters. To see this, imagine the forward slash and backslash characters typed with a space between them: / \. Extending either character individually, even by doubling its length, would not make it touch the other, because they are pointed in the wrong direction. To deal with cases

[44] http://www.infrastructureafrica.org/

[45] The projection's central meridian is 15 degrees east longitude.

like these, "bridges" were created as follows. All dangles were paired with the closest other dangle if it was within 100 meters using the Spatial Join tool, and connecting lines were created between these pairs of dangles. These bridges were added to the rest of the roads.

The AICD roads database was gathered with explicit reference to inter-city roads. Unfortunately, this means that in many cases, information on roads within cities was not collected, greatly reducing the connectivity of the dataset in many countries. "Spiders" were created to model missing city roads. For every dangle falling within a city, a road was created between the city centroid and the dangle. The implicit assumption is that radial road travel within cities is comparatively easy. Spiderlegs are assumed to be paved. The resulting spiders were added to the roads, and all spiderlegs and any roads that intersected them were "planarized". Before planarizing, the topology of the network was such that a spiderleg could cross a road without being connected to it. Planarizing ended this.

Manual edits were necessary for several reasons. Recall that Extend did not close holes if a dangle was not pointed at another road, and that bridges were only created between two nearby dangles, not a dangle and a non-dangle. So extra segments of less than 10 meters each were created in 6 other places to fix dangles affecting routes to 16 cities. Recall also that spiders were created between a centroid and any dangles within a city. So once a spider is created, the spiderleg is the network location of the centroid, so if it connects to a dead end, other nearby roads cannot be reached even if they are very close. In these cases, deleting one or more spiderlegs fixed the problem. Five spiderlegs affecting 4 cities were deleted.

A.4 Route calculation

To prepare for building the network, the roads were intersected with all land borders, so that the resulting border posts could be used as barriers—non-traversable points on the network. Coastlines were not treated as borders in this operation, because the only reason a road would cross a coastline is because of misalignment—the resulting route is most likely legitimate.

A network dataset was built using the roads dataset. The "Closest Facility" solver was used with the following settings. All light centroids were used as the "Incidents", centroids of primate cities were used as "Facilities", and the intersections of the roads and the land borders were used as "Barriers". Each city was assigned a network location on the closest road within 5 kilometers of its centroid.

Unfortunately, because of a quirk in the program, the total calculated length is a true geodesic distance, while distances by paving status are projected distances. However, this never causes a discrepancy of more than a few percent, and because the same projection is used for the whole continent, these errors are highly correlated within countries.

Of 485 populated lights, 464 (96 percent) received plausible routes. Of the remaining 21, 6 were in exclaves or islands, and 2 had centroids more than 5 kilometers from the nearest domestic road. Three received no routes because they were on road segments disconnected from the primate by a gap of at least 100 meters. The remaining ten received implausible routes (because of suspicious gaps of longer than 100 meters in the road network) and were removed. To the extent that these cities are in fact less connected than others, or that government officials have not mapped their roads correctly or at all, they are more likely to be excluded from traditional data sources like censuses and surveys as well.

A.5 Pixel-level data-generating process

The pixel-level data-generating process can be modeled as follows:

$$y_{jist} = \begin{cases} 0 & \text{if } y^*_{jist} < 2.5 \text{ or } \sum_{k \in i} \mathbf{1}\{y^*_{kist} >= 2.5\} < 4 \\ 63 & \text{if } y^*_{jist} > 62.5 \\ int(y^*_{jist} + 0.5) & \text{otherwise} \end{cases} \tag{8}$$

where j indexes pixels, which nest in cities, s indexes satellite-years within a year, y_{jist} is measured pixel-level light, and y^*_{jist} is true (latent) pixel-level light. Two nonlinearities appear here, in addition to rounding to the nearest integer. Processing by NOAA converts to zero nearly all (1) individual pixel values of 1 or 2 and (2) clusters of less than 4 nonzero pixels. In both cases, NOAA's processing algorithm interprets these patterns as random noise.

The relationship of interest is at the city level, as are all of the regressors, but the lights data are generated nonlinearly at the pixel level. Rather than estimate Equation (8) via maximum likelihood with approximately 5.6 million pixel-satellite-years, I instead simply sum lights across pixels and satellites within a city:

$$y_{it} = (1/S_t) \sum_{s=1}^{S_t} \sum_{j \in i} y_{jist} \tag{9}$$

where S_t is the number of active satellites (always 1 or 2), and run a tobit regression with a cutoff value of 5.5. The theoretical minimum non-zero city-year has a DN value of 6: in one satellite-year it is unlit, while in the other satellite-year, it consists of 4 pixels, each with a DN of 3. In practice, this is also the minimum non-zero city-year DN value in the main estimation sample.

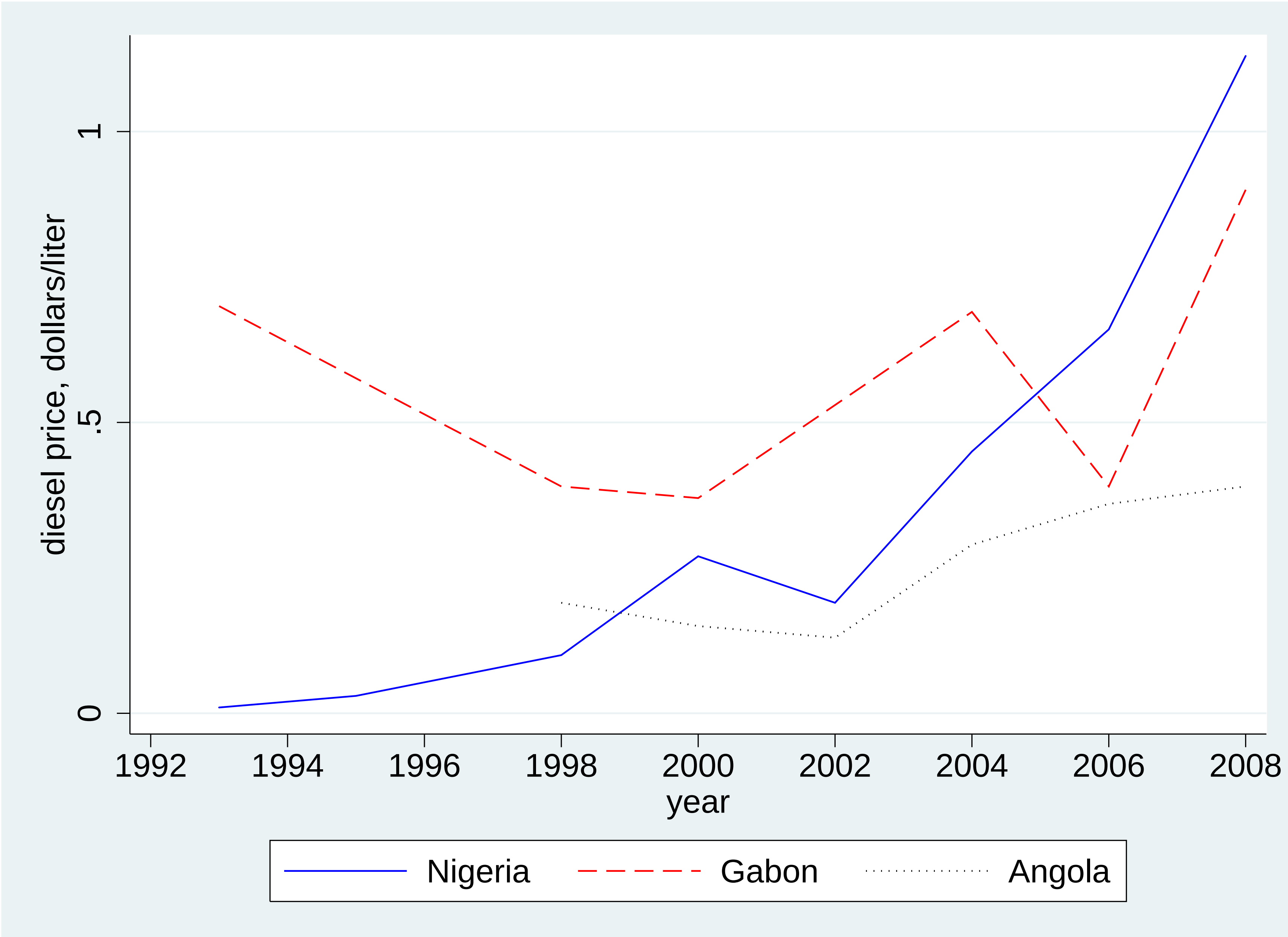

Figure A.1: Diesel prices for the oil producing countries, 1993–2008