Power Outages, Hydropower and Economic Activity in Sub-Saharan Africa

by

Giovanni Occhiali

A thesis submitted to the University of Birmingham for the degree of

DOCTOR OF PHILOSOPHY

Department of Economics

Birmingham Business School

College of Social Sciences

University of Birmingham

September 2016

Abstract

Many Sub-Saharan African economies experienced high growth rates over the last decade, a welcome change from the depression and stagnation which characterized them during the 1980s and 1990s. However, improved economic performances were mostly driven by an increase of both price and demand for the natural resources of which the continent is rich, so that these growth rates were not associated with a significant increase in industrial diversification. The poor quality of the power infrastructure of many African economies represents one of the major obstacles to their structural transformation. In this thesis we investigate the effects of an unstable power supply on the profitability of Sub-Saharan African firms. To avoid estimation issues related to the possible endogeneity of the relationship between power supply and productivity we develop an instrument based on the water available for hydropower production. Our results show that frequent power outages are indeed a very significant drag on firms' performance, much more so for firms without access to back-up capacity than for the overall sample. The final part of the thesis also investigates the general relationship between hydropower production and economic activity in Sub-Saharan Africa through the use of night-light data.

Acknowledgments

To start with, my most sincere thanks go to my supervisors, Prof. Matthew Cole, Prof. Robert Elliott and Prof. Eric Strobl for all the advices, guidance and comprehension they have offered during the course of the PhD. I am also grateful for the generous funding provided by the Department of Economics of the University of Birmingham, and I want to thank Dr Marco Barassi, Dr. Kamilya Suleymenova, Dr. Marta Guerriero and Dr. Ceri Davies for the many suggestions, especially, but not only, on how to tackle class duties.

I would not have made it to the end of this route without the daily help and the many discussions that I had with many of my fellow PhD students. A special thank goes to the whole of Underwood Close 37, Marea, Dan, Allan and Mera, with whom I broke bread every day for two years: you have been my family away from home. The mention of honour goes to Enrico: the late hours and glasses for the road have been too many to be publicly acknowledged, but I know that without each and every one of them I would not be writing these words today. You can take part of the merit but none of the blame.

The friends of old times have also been there when I needed them most, as they always have been, be that in Ferrara, London, Sardegna or Rwanda. You are too many to be thanked one by one, and I consider myself lucky to have all of you by my side. For the special role played in the darkest of moments I have though to mention the most recent additions to this group and those who made them possible: Bianca, Lucia, Chiara and Jarek.

My deepest gratitude goes to my Mum, Dad and Brother: to you I owe almost all of my strengths and practically none of my weaknesses. Sorry for having bothered you so much, I have no doubts I will continue.

My final thanks are for Sara. Having found the courage to approach you in that elevator will always remain my greatest achievement in four years at the University of Birmingham.

List of Contents

List of Figures

List of Tables

List of Abbreviations.

2SLS – 2 Stages Least Square

ADS – Alby, P., Dethier, J.J. and Straub, S. (2011)

AICD – Africa Infrastructure County Diagnostic

DEM – Digital Elevation Model

DMSP OLS – Defense Meteorological Satellite Program Operational Linescan System

EROS – Earth Resource Observation and Science

FEWS – Famine Early Warning System

FS – Foster, V. and Steinbucks, J. (2009)

GDP – Gross Domestic Product

GeoSFM – Geospatial Stream Flow Model

GLCC – Global Land Cover Characterization

GRDC – Global Runoff Data Centre

GW – Gigawatt

ICT – Information and Communication Technology

i.i.d. – independently and identically distributed

LIML – Limited Information Maximum Likelihood

MW – Megawatt

NASA – National Aeronautics and Space Administration

NOAA – National Oceanic and Atmospheric Administration

OLS – Ordinary Least Squares

PC – Principal Component

PCA – Principal Component Analysis

SEDAC – Socioeconomic Data and Application Center

TFP – Total Factor Productivity

USGS – United States Geological Survey

VNIR – Visible and Near-Infrared

WEPP – World Electric Power Plants

WBES – World Bank Enterprise Survey

Introduction.

The prospects for a successful development of many economies in the African continent, especially in Sub-Saharan Africa (SSA), are definitely brighter today than they were at the beginning of the century. The gross domestic product (GDP) of the region doubled since 2000, experiencing an average growth rate of 5% during the period 2008-2015 and six of the fastest growing economies on the planet over the period 2014 to 2017 are prospected to be African (IMF 2015, World Bank 2016). Significant progresses have also been made in primary enrolment and completion rates; in reducing child mortality, in improving the quality of maternal health, in the incidence of HIV/AIDS and malaria and in the accessibility of information and communication technologies (IMF 2015, UNECA 2015). The combination of these factors, together with a slow-down of the economic performances of many high-income countries, which were more affected by the financial crises of 2008-2009 than most SSA economies, has led to a growing narrative of "Africa on the rise".

Although all the above achievements are not to be diminished and surely offer a more fertile ground for more to come, they do not represent the full picture of the socio-economic situation in the continent nor are easily sustainable in the long run, as the GDP growth of 3.4% in 2015 demonstrates (IMF 2016). The impressive growth rates of the last decade were mostly driven by a combination of high prices and strong international demand for many of the commodities in which the continent is rich. The excessive reliance of many SSA economies on their resource endowment has been a source of concern for a long time. Although primary commodities have played an important role in the development of many industrialised countries, strategies to develop industrial forward linkages in these sectors have often required intense government involvement and capability, and changes due to the increasing financialisation of commodity markets and to innovations in global value chains' structure have only complicated the picture (Hirschman 1958, UNECA 2013).

The nature of the sectors driving economic growth has deep implications for the characteristics of the development path which a country will follow. Primary commodity sectors tend to be more capital than labour intensive, especially mineral and oil extraction, and normally generate limited value added. The marked relevance of these sectors in bringing about the recent growth across SSA offers a first explanation of its almost jobless nature: between 2001 and 2013 the percentage of Sub-Saharan Africans in vulnerable employment decreased by only 2.3%, standing at a level of 77.4%. Given the demographic structure of the continent and the impressive population growth rates in most countries, which have led the population to increase from 642 million in 2000 to 912 million in 2013, the creation of more decent employment opportunities is vital (ILO 2014). An increasing population also contributes to explain the less impressive performance of the region in terms of per capita GDP growth, which averaged at 2.09% for the period 2001-2014 (which contrasts with an average of 7.71% in the East Asia-Pacific region, UNECA 2016b).

It is then clear that efforts towards a continuing structural transformation of most SSA economies have to be sustained, and indeed this has been the focus of many of the continents policy makers and of many international development organizations (UNECA 2011, 2013, 2014, 2016a). Historically, the sectors which offered the greatest contribution to sustained and economy-wide growth tend to have been within manufacturing, and surely it does not bode well that the contribution of manufacturing to economic output in the continent has declined by 1.5% between 1980 and 2013. Similarly, value added in most African manufacturing sectors is only half of what it was in East Asia when most of its countries reached lower-middle income status and in 2010 95% of SSA population, excluding South Africa, had a manufacturing value added of less than 100$ per capita (against 622$ in Brazil and 820$ in China, UNECA 2016b).

Although the specific industrial policies which might help SSA countries to achieve a more balanced economic structure necessarily differ along with the national conditions, a common constraint across the region is the precarious status of the energy sector. The fact that modern energy, and electricity specifically, underpins *all* industrial activities and that its continuous development from the late 19th century onwards has led to tremendous increases in human well-being hardly needs mentioning (Rosenberg 1998, Elias and Victor 2005). However, there are still almost 1.5 billion people on the planet who do not enjoy the contribution of electricity to daily life, and approximately half of them (620 million) are located in the African continent (IEA 2014). A quick exposition of a few energy statistics might help to illustrate the challenge which the fruitful development of the African energy sector still represents.

The installed capacity in the whole of SSA stands today at 90 gigawatt (GW, approximately half of which is located in South Africa alone), a significant increase from the 68 GW of 2000 but still significantly less than what would be required to meet demand (IEA 2014). To offer a comparison, the United Kingdom has an installed capacity of 80.8 GW. Moreover, a significant part of the installed capacity in SSA is not even operational, as many power stations have fallen into disrepair due to a lack of maintenance. Transmission and distribution losses, averaging at 18% when South Africa is excluded, are more than double the world average, adding a significant cost to electricity tariffs which are already amongst the highest in the world. Per capita electricity demand, almost unchanged since 2000, is 75% lower than in Asia, and the average residential electricity consumption is 317 Kwh (225 Kwh if South Africa is excluded), approximately half of that of China or a fifth of Europe. Electricity access varies widely across the continent, from a minimum of 3% of the population in Chad to a maximum of 85% in South Africa, with urban rates always consistently higher than rural ones. The overall access in the region increased from 23% in 2000 to 32% in 2012, with 145 million people being connected to the grid. However, as population growth outpaced electrification rates, there are today 100 million more people without electricity access than there were in 2000 (IEA 2014).

Finding a solution to the long-term under-investment in the power sector has been a priority of almost all governments in the continent for at least two decades. Many countries embarked in energy sector reforms during the end of the 1990s, as state-owned utilities were underperforming from both a technical and a financial perspective and states were unable to mobilize the required funds to expand the electricity sector. The most common reforms entailed the corporatization of the national utilities, so to improve their financial stance and at least theoretically detach tariffs setting decisions from the political process, and the opening of electricity markets to independent power producers, so to increase generation capacity, often accompanied by the institution of an independent power regulator to increase the investment appeal of the sector. Although it must be recognised that reforms have in some cases increased the efficiency of electricity markets and independent power producers have given some contributions to grid-capacity, neither the ability of the now-commercialized national utilities to charge cost-recovery prices nor the amount of transmission and distribution losses have changed considerably (AFREPREN 2004, Eberhard *et al.* 2011).

At the end of the last decade there was still a loud cry for increased infrastructure investment in SSA, exemplified by the call for 93 billion dollars per year of investment in the World Bank 2009 report "Africa's Infrastructure: A Time for transformation" (Foster and Briceño-Garmendia 2009). The situation started to improve soon afterwards because of a mixture of factors. First, thanks to the commodity boom, and generally to a more stable macroeconomic stance and thorough revenue collection, domestic resources increased substantially and government expenditure on infrastructure followed, reaching 51.4 billion dollars, or 63% of total infrastructure spending, in 2012. Second, the most relevant international and institutional investors also started to increase their annual commitment to infrastructure development, which grew from 5 billion dollar in 2003 to 30 billion in 2012. Private participation in infrastructure investment also increased by 9.5% per year between 2002 and 2012, and both the World Bank and the African Development Bank significantly stepped up their commitment. One of the most impressive changes has though been the increased relevance of Chinese investments, which now represent around 20% of infrastructure commitment in the continent (Gutman *et al.* 2015).

Energy infrastructure received a significant share of all the above investment, but this is especially true for the Chinese commitment, as energy infrastructure received 34% of their spending over the period 2005-2012. Chinese companies are involved in projects related to generation, transmission and distribution in 37 SSA countries and have secured more than 200 green-field projects between 2010 and 2020 (counting only those already financially secured), more than half of which have already been completed. Chinese contractors have been responsible for connecting more than 7 GW of new generation capacity between 2010 and 2015, or 30% of the capacity added in the period. Moreover, 56% of this new generation capacity exploits renewable energy sources, predominantly hydro-power, with Chinese companies having become the leaders of the sector in Africa, which now represents their most relevant export market (IEA 2016).

If the contribution of Chinese companies to the development of the hydropower sector represents a recent development, the relevance of this technology in the African portfolio of generation capacity is nothing new. Currently there are 20 GW of installed hydropower capacity in the continent, making it the most used renewable resource in the continent and the main source of power generation for many countries, including the Democratic Republic of Congo, Kenya, Mozambique and Uganda to cite just a few. Although hydro-power is characterized by a high face-capital requirement, especially in large-scale hydropower plants, it offers the lowest average cost of generation amongst all production technologies, renewable or not. Notwithstanding the social impact which characterises large dam construction, which has received considerable attention in the literature (World Commission on Dam 2000, Duflo and Pande 2007), the development of the largely untapped hydropower potential in the continent (estimated at 283 GW) is quoted by most funders as one of the main ways to guarantee the required boost to the generation capacity of the region without contributing to the ever-increasing carbon emissions, of which the energy sector is one of the main contributors.

However, while hydropower is seen as part of the solution to climate change, the generation of hydroelectricity in Africa is dependent upon a stable and predictable climate. The continent is already experiencing significant climate related stresses, amongst which reduction in surface run-off, increased competition for water resources and frequent floods coupled with an increase in the incidence and length of warm spells and droughts. The relevance of these factors varies across different SSA regions, but all of them are directly or indirectly related to the ability of any country to exploit its hydropower resources. Burundi, Ghana, Kenya, Rwanda, Tanzania and Uganda all experienced serious issues in meeting energy demand through hydropower production during the end of the last decade due to frequent droughts. In 2015 and 2016 rainfall failures led to a noticeable increase in the frequency of power outages in Zambia, and more generally in the Zambezi river basin, in which some hydropower plants are failing to meet their prospected return on investment. However, as the climate factors influencing hydropower are many and the modelling of future changes is subject to uncertainties connected with varying climate scenarios, it is unlikely that the role of hydropower in the energy mix of SSA will change drastically over the medium period (Niang *et al.* 2014, Cole *et al.* 2014, UNECA 2016a).

In this thesis we first investigate the effects of unreliable power supply on the sales performance of SSA firms, using hydropower availability for electricity production as an instrument to resolve possible problems of endogeneity involved in one-step estimation. Although the significant bulk of energy used in SSA is goes to the residential sector (more than 60%, versus 25% across other developing regions and 20% in OECD countries), the (un)availability of a stable electricity supply has a significant impact on all productive sectors. Across SSA, industry typically accounts for less employment than agriculture and contributes less to GDP than services but absorbs two thirds of the energy directed towards productive uses (with the three of them together accounting on average for 21% of total energy consumption). Frequent power outages lead to damages to production equipment and foregone sales, and their overall cost across SSA is estimated at as much as 2.1% of GDP and 4.9% of total sales (Eberhard *et al.* 2011, IEA 2014). Moreover, as firms have long identified the unreliability of electricity supply as one of the main obstacles to their expansion, the relevance of backup generation across the continent has increased significantly over the last two decades, representing a significant share of the installed capacity in all SSA regions.

The first chapter introduces the most recent World Bank Enterprise Surveys, one of the main data sources for the analysis, and develops the Ordinary Least Squares framework which will constitute the basis for the successive instrumental variables analysis. The chapter starts with a review of both the macroeconomic and the microeconomic literatures connecting infrastructure development to growth performance, covering the experience of both developed (Aschauer 1989, Canning and Pedroni 1999 amongst others) and developing countries (from Lee *et al.* 1996 to Allcott *et al.* 2014), including both infrastructure-wide (Escribano *et al.* 2009. Moyo 2012a) and sector specific studies (water, Davies *et al.* 2001, road network, Luo 2004, ICT, Lio and Liu 2006). The chapter then continues with a thorough exposition of the firm-level data on a continental and country level, confirming both the previous picture of endemic power outages (firms from 9 out of the 38 countries in the sample experienced on average more than a 1,000 hours of power outage per year) and the relevance of generator ownership across firms of all sizes and in all countries. The following analysis aims at obtaining a first approximation, still possibly influenced by endogeneity issues, of the elasticity of firm's output to power outages. Furthermore, two of the most recent models used in the literature to analyse the determinants of generator ownership in developing countries, that of Foster and Steinbucks (2010) and of

Alby, Dethier and Straub (2011), are also applied to the data, so to obtain a better understanding of the current trends in the continent. Our results confirm that power outages are indeed a considerable problem for firms in SSA, although particularly so for firms without access to backup options, which are unable to shield themselves from the bulk of the negative effects. Power outages also appear to be one of the main determinants in a firm's choice of investing in backup generation, although this is also influenced by its access to finance, its export status and, to a lesser extent, the relevance of electricity as an input in production.

The most recent literature does however raise serious and well-founded concerns about the assumption of an exogenous relationship between the performance of a firm and the quality of its energy supply. For example, a firm's decision of its initial plant location might be influenced by the quality of the nearby energy infrastructure, and more experienced and connected managers might gain access to prime locations. Another example of a possible source of endogenity are power holidays policies, adopted by many governments in order to ration the available electricity between different sectors, which inevitably take into account the relevance and contribution of the latter to the national economy. We are then in need of finding a suitable instrument, affecting the amount of power supplied but not firms' performance, and we individuate it in the amount of water available for the generation of hydroelectricity, which as previously noted is one of the most relevant technologies in the generation portfolio of many SSA countries.

The second chapter is therefore dedicated to the exposition of the Geospatial Streamflow Model developed by the US Geological Survey, the hydrological model used by the Famine Early

Warning System to assess the probability of occurrence of extreme weather events, the main cause leading to famine, constituting the basis for the construction of our instrument. After reviewing the most recent literature covering the hydrology of the African continent and how this is likely to be influenced by climate change, we move onto the assessment of why the scarcity of physical information about most African catchment areas makes non-data intensive hydrological models such as the one used in the chapter particularly suited. After a careful exposition of the internal mechanics of the model, the chapter continues with the presentation of its results for 8 of the 9 main continental basins in Africa[1] for the period 2001-2010, including a principal component analysis aimed at isolating common hydrological groups. The final part of the chapter constitutes a thorough assessment of the reliability of the model's results through their comparison with the historical records for 440 gauge stations, available through the Global Runoff Data Centre (GRDC) of the German Institute for Hydrology. This investigation, based upon both Copula functions, which have been receiving a growing attention in hydrology, and panel regressions, concludes that, although there are divergences in the model performance in different basins, the results are sufficiently reliable to be used as a starting point for the construction of the instrument. Anticipating the analysis of the following chapters, we also show how the instrument constructed on the GeoSFM estimate is indeed a significant predictor of hydroelectric production.

The third chapter concludes the analyses of the effects of power outages on firms' sales by connecting the economic analysis developed in the first chapter to the hydrological one developed in the second. To begin with, the chapter explains the procedure used to connect the

[1] Northern Africa has not been included in the analysis.

modelled streamflow with the data about power plants contained in the World Electric Power Plant (WEPP) database from PLATTS, a global provider of energy and commodity information. It then carefully discusses how we linked the power plants to the cities in which firms are located. Subsequently, summary statistics are presented for the two preferred forms of the instruments, which are either a single measure of the deviation of the yearly average streamflow from its long-term average or a series of 4 disaggregated indexes more directly accounting for the frequency of weaker and stronger negative and positive deviations from the long term average. The following analysis shows how the results from both 2 Stages Least Squares and Limited Information Maximum Likelihood are substantially different from the Ordinary Least Square estimates, suggesting that problems of endogeneity are indeed present in the relationship between a firm's performance and the quality of its electricity supply for firms without access to back-up generation. Specifically, it appears that the previous analysis was seriously under-estimating the detrimental effect of power outages on the latter, which incur much higher sales losses from unreliable electricity supplies. This result appears to be robust to a series of different specifications and combinations of instruments, while in the concluding part of the chapter we also investigate the possible selection bias of the estimates for firms with a generator.

Finally, the fourth chapter investigates a different link between economic activity in SSA and availability of water for hydropower generation, namely if it is possible to connect the latter with the luminosity of SSA countries as measured from space, which has been recently shown to be a valid proxy for economic activity. To do so, we rely on data collected by the Defense Meteorological Satellite Program – Operational Linescan System (DMSP OLS), whose satellites have been circling around the earth multiple time per days taking pictures in which each pixel, representing approximately a square kilometre, is given a value from 1 to 63 depending on the intensity and frequency of the light emitted in its area. These data, which have been collected from the 1970s and are digitally available from the early 1990s, have been receiving growing attention from social scientists in recent years and have been used in studies exploring a variety of topics (from population density and urban expansion to subnational GDP) as made clear by the literature review. After presenting the national level data, a methodology used to expand the city boundaries in the absence of reliable data on urbanization is introduced, as the ability to account for the growing relevance of urban centres in the continent might be relevant in obtaining meaningful estimates. The analysis, carried out through pooled OLS, fixed effects panel estimations and quantile regressions, does not validate the hypothesis of a direct link between hydroelectricity production and economic activity at neither the sub-national nor the national level. However, in the models in which the urbanization rates are allowed to vary by city and the sample is restricted to countries in which hydropower represents at least 30% of the installed capacity, we find some evidence that a higher availability of water for hydropower generation is, as expected, associated with more intense light emission, and the more so the higher the share of hydropower in the generation portfolio.

Chapter 1

Energy Infrastructure in Sub-Saharan Africa: a Firm Level Analysis.

1.1 Introduction.

The current level of economic diversification – and to a certain extent of economic development – of many Sub-Saharan African (SSA) countries is in many aspects comparable to that of 30 years ago. It is true that since the second half of the 1990s positive growth rates of the GDP in many SSA countries have led to renewed hope for the continent, but these growths were mostly due to trends in international commodity prices and to the unearthing of new mineral reserves. The contributions of industrial and manufacturing sectors remain low and just in a few cases account for more than 30% of a country's GDP. The share of the labour force employed in industry is also low, ranging from to 2% in Mauritania to a maximum of 30% in Swaziland[2] (CIA World Factbook).

[Figure 1.1 about here]

It is difficult to imagine that many SSA countries will be able to catch up with the standard of living of their economically more developed counterparts without undertaking and completing a process of structural transformation. The vast majority of the population of SSA still lives in rural areas, often maintaining themselves on a mixture of subsistence agriculture and income obtained through employment in the informal sectors. One of the main drivers of structural transformation for countries that managed to grow successfully has been industrialisation. It is then natural that the focus of economic policymakers has been on how to foster the development of the manufacturing sectors.

[2] For a recent review of African industrialization prospects see Page 2012, for analysis focusing on specific aspects see UNECA 2011, 2013, 2014, 2016a, and 2016b.

In the literature the factors deemed important for a propitious industrial development include the availability of capital for investment, a well-defined set of property rights, a good business climate, the liberalization of the labour market coupled with investment in workers skills and stable exchange rate policies, to name but a few. Another well recognized ingredient for success is an adequate level of infrastructure: energy, and often water, are vital for industrial production; and the presence of well-developed roads, rails, river networks and deep-sea ports is needed for the transport and commerce of goods and final products, both in a national and an international markets.

All of the above elements constitute somewhat of an obstacle for a fruitful industrial development of SSA economies, although to different extents in different countries. The poor quality of many African state institutions is well recognised: weak protection of property rights, scarce application of the rule of law, both of which increase transaction costs in often already inefficient markets, and low government accountability are all widespread in the region. The latter has often resulted in rent-seeking behaviour by government officials, and it is undeniable that high corruption levels have been, and still are, another common factor in many African economies (Ng and Yeats 1999, Sachs *et al* 2004, Fosu *et al* 2006, Collier 2007). While these are all generally accepted facts, there still is much disagreement on which type of institutional development would lead to better development outcomes (Hickey 2012, Khan 2012, Noman and Stiglitz 2012).

The lack of skills of the African workforce is also regarded as a hindrance to the diversification of the continent economic structure. After years of focus on primary and secondary education, which led to relevant increases in both enrolment rates, the attention has recently been moving towards the important role of tertiary education, which seems to have, contrary to previous evidence (Psacharopoulos and Patrinos 2004), a higher rate of return than basic education, indeed as high as 25-30%, especially when the risk of unemployment is taken into account (Colclough *et al.* 2010, Diagne and Diene 2011, Barouni and Broecke 2014). However, while the microeconomic link between education and private earnings is seldom questioned, the effect of a higher endowment of human capital on firms' productivity is less straight-forward and often depend on others firms' characteristics or on which aspects of workers skills are taken into account (Söderbom and Teal 2000, Teal 2010, Danquah and Ouattara 2014). Similar tenants hold with regard to the effects that a more skilled workforce will have on the FDI attractiveness of African states (Cleeve *et al.* 2015, Ssozi and Asongu 2016).

Finally, and central to this study, the level of infrastructure in Africa is often taught to play a large role in dampening structural transformations in the continent, as all SSA countries are united in facing some forms of infrastructural gap. The road network is scarcely developed (and rail networks virtually non-existent), which in vast and scarcely populated countries implies that development of internal commerce is difficult, mostly based on a few markets in bigger towns due to a long and costly movement of goods. Water infrastructure, mainly water storage facilities, are also particularly underdeveloped, with huge costs for the wellbeing of African citizens and agricultural output.[3]

[3] See for example the proceedings from the International Journal of Hydropower and Dams Conference Africa 2013.

In terms of energy infrastructure, SSA has the worst energy outlook in the world. Of the 122.6 Gigawatts (GW) of installed generation capacity in Africa, 39% is situated in Northern Africa (Morocco, Algeria, Tunisia, Libya and Egypt) and 36% in the Republic of South Africa alone. This leaves only 25% for the rest of SSA, slightly less than 31 GW[4]. Energy access figures are similarly daunting: only 32% of the population has access to electricity, and the level in rural areas is down to 14%. SSA also scores poorly in comparison with other developing regions: the installed capacity of 92.3 MW per million people is 20% of that of East Asia or Latin America. The per capita electricity consumption of 561 Kwh is five times lower than the world average, and it can be noted that Northern Africa and RSA account for 38% and 40% respectively (all the above figures come from Brown, Muller and Dobrotkova 2011, Muller, Marmion and Beerepot 2011, Sokona, Mulugetta and Gujba 2012, IEA 2014)[5].

The technical, financial and institutional capacity of many African states has been undermined by decades of forced structural adjustments and incomplete economic reforms. This creates a very different situation from the one in which energy networks have developed in many industrialized countries. Depending on different objectives, the resources required to fill the infrastructural gap vary widely: in 2007 the African Development Bank estimated that investments of 4 billion dollar per year were required to fill the existing demand gap. In 2008 the African Infrastructure Country Diagnostic (AICD) estimated that to achieve an overall electricity access of 35% by 2015[6] - a goal which has been missed - the necessary investments ranged from 13 billion dollars in the Eastern Africa Power Pool to 68 billion dollars in the Southern one (Ram 2007, Eberhard *et al.* 2008, Agbemabies, Nkomo and Sokona 2012). In a

---

[4] Compare it, by example, with the 80.8 GW installed in UK alone.

[5] The poor quality of available information for the continent might lead to discrepancies between the statistics of different agencies; however those differences are small compared with the magnitude of the figures.

[6] As aimed in most of the National Development Plans of SSA.

situation in which many African states rely on overseas development aid (ODA), several African countries spend half of their foreign reserves on fossil fuel imports.

Given the budget constraints faced by most SSA governments, public investment in energy infrastructure will divert a significant share of funds away from other destinations, such as education or healthcare, and the most transparent way to compare different public projects is through cost-benefit analysis. The difficulties existing in conducting the latter in developing countries, in which imperfectly competitive labour, financial and good markets might well be the norm and the relevance of the potential equity issues is different than in developed ones, is recognised in the literature (Dinwiddy and Teal 1996, Quah 2012). Both infrastructure and alternative investment recipients, such as education, are characterised by some degree of complementarity and by different types of externalities, which might be hard to quantify. Although some of the externalities arising from energy investment can be evaluated with regard to the increase in production that they generate, others relate to the expansion of the grid to previously unserved areas, where the outcome is unlikely to be as easily measurable in terms of output. Other types of social costs entailed might depend on the electricity-generation technology, such as the destruction of natural environment and local livelihood sometimes caused by hydropower dams. For the latter case, it is particularly hard to find cost and benefit assessments which will be deemed agreeable both by opponents and proponents of the project. At any rate, some evidence on the high social rate of return of energy infrastructure in developing countries exists (World Bank 1994, Canning and Bennathan 2000), and although extensive comparisons with other types of investment are missing, given the vital role that a stable electricity supply plays in every aspects of daily life (Rosenberg 1998, Elias and Victor 2005) we assume for the remainder of the work that this return is high enough to justify our attention.

It is with this background in mind that we move to the investigation of the economic effect of the quality of energy supply on firms' performance, estimating the elasticity of firms' output to electricity disservice. The structure of the chapter is as follows: section 2 provides a review of the literature, focusing on SSA; section 3 presents the data used in the study and summary statistics at continental and country level; section 4 covers the estimation of firms' elasticity to outages; in section 5 models for generator ownership are presented while section 6 concludes.

1.2 Literature Review.

This paper adds to the literature on the contribution of infrastructure to development and economic growth. This can be divided in to two main branches, one that looks at the effects at the macro-level, and one that looks at effects at the micro-level. The latter is becoming increasingly important thanks to the recent availability of firm level data for developing countries[7].

Two early reviews of the macro-level literature are provided by Aschauer (1989) and Gramlich (1994). Although this early literature considered mainly developed countries (mostly intra-state differences in the US) similar approaches have been employed in the study of developing countries. The main conceptual and econometric problems were already clear: common time trend across countries' expenditure in infrastructure investments; the difficulties of acquiring

[7] The papers reviewed in this section were selected from those found by searching for different combinations the words "Infrastructure", "Electricity", "Outage", "Shortage", "Africa", "Growth" and "Firm" in EBSCOhost and the bibliographies found therein.

data to proxy the many missing variables, and especially the role of information on energy prices; the direction of causality between growth in infrastructure and growth in GDP; etc. Gramlich (1994) concludes by encouraging a greater focus on micro-level studies, which he argues are more likely to produce policy relevant analysis.

One of the first papers to examine the firm-level effects of infrastructure deficiencies is Lee *et al.* (1996). Focusing on Nigerian, Indonesian and Thai manufacturing firms in the World Bank Surveys, they look at the extent of the costs that firms incurred because of low infrastructure quality; how they react through private investment and how public and private interests should be coordinated to achieve a second best solution when state budgets are constrained.

Using a Barro-type regression, Canning and Pedroni (1999) analyse the contribution of different infrastructure provisions to GDP per capita. Based on a 40 year panel for countries at different stages of development, they construct a growth model in which infrastructure investment is achieved by diverting private savings and public investments. Decisive support is found for an endogenous growth-model in infrastructure, where investments have a positive effect on short run growth but are insignificant in the long run. Finally, the necessity of dividing the effect of different types of infrastructure provision is underlined as the authors find no evidence of global shortage of telephone infrastructure but do of energy capacity.

In terms of infrastructure specifically, Davis *et al.* (2001) analyses the effect of improving water infrastructure on microenterprises using data for 137 Ugandan firms. The rationale for the study arose from the perception of business owners that obstacles related to water provision were amongst the most relevant in constraining their activities. The evidence gathered by the authors, who exploit differences in water infrastructure quality between two neighbouring villages, however, points toward the fact that water connections have higher economic benefits for households than those experienced by microenterprises, whose use of water does not differ that much from that of the former. The study further stresses the importance of gathering information on the real needs of local communities before committing budgets on the basis of unsubstantiated assumptions.

Returning to the macro-literature, Esfahani and Ramírez (2003) examine the role of institutions in mediating the economic effect of infrastructure investments. After developing a model similar to that of Canning and Pedroni (1999), the authors test their hypotheses with a panel of 75 countries over three decades. They conclude that once the simultaneity of GDP and infrastructure investment is taken into account, the contribution of the latter to the former is substantial. Moreover, they stress the relevance of institutional capacity, intended as more than pure project development and fund disbursement, for achieving the best result with investments in infrastructure[8].

An analysis of the role played by the road network development is provided by Luo (2004), who takes into account how the geographical location of firms interacted with infrastructure investment in the western region of China. Using a Solow-type growth model applied to a twenty year panel, the author concludes that the development of a good transport infrastructure is important for fostering intra-regional trade. She also finds evidence that, via a general reduction in transport costs, there are important spill-over effects for regions not directly in receipt of the investment.

[8] The authors also stress the need for improved data collection if the topic is to be thoroughly researched for developing countries.

The impact of information and communication technology (ICT) on agricultural productivity is examined by Lio and Liu (2006), who employ a 5 year panel for 81 countries. While the link might not appear clear at a first glance, the authors carefully analyse the role that ICT plays in distributing knowledge about the existence and use of new technologies, helping their adoption process. On the other hand, they also stress that ICT investments are not enough on their own, as their returns can be shown to be consistently different between countries with different level of human capital.

One of the first papers to directly analyse the effect of power outages is Fisher-Vanden *et al.* (2008). Incorporating the effect of electricity shortages into a trans-log cost function model, the authors examine TFP growth rates over 5 years for 1340 Chinese firms in eleven sectors. As proxy for the outages, the authors develop a scarcity measure reflecting the likelihood of shortages in any given year based on information on the ratio between thermal electricity generation and capacity from the China Electricity Yearbook. As the variable will probably suffer from measurement error, and furthermore there might be policies influencing both firms' productivity and their demand for electricity, they consider electricity scarcity as endogenous and instrument it with weather variables. Although some of the results are recognized as tentative, the authors conclude that there is some evidence of enterprises shifting away from energy/electricity intensive production into more labour and material intensive activities because of electricity disservice.

Another paper from 2008 is Eifert and Gelb (2008), who consider the effect of indirect costs on business performance indicator, specifically in the African case. After explaining the relevance of these inputs in production, the authors justify their focus on the continent due to the particularly high indirect costs share which African firms face, especially those infrastructure related. This is due to a series of factors, ranging from low-density of production networks, which hinder the emergence of economies of scale, to poor infrastructure stocks, which increases transportation costs that are already high for land-locked countries. The estimation of indirect-costs-augmented revenue and value functions for six industries in seventeen countries lead the author to conclude that many studies over-estimate value-added and factor productivity of African firms because they fail to take into account exactly this type of costs.

In occasion of the launch of the AICD initiative, a few papers focused on the role that infrastructure might play in fostering economic and social development in Africa. Calderon (2009) relies on an unbalanced panel of 136 countries over the period 1960-2005 and applies an instrumental variable generalised methods of moments to estimate the combined effect of the quantity and quality of ICT, electricity and road infrastructure – with different indexes obtained through PCA – on growth of real GDP per worker. His results show how both the stock of infrastructure and its quality matter for GDP per worker growth, although the effect of the former greatly outstrips that of the latter. Through a simulation exercise, the author shows that if all African countries were to catch up with the stock of infrastructure of Mauritius, the region leader, this will entail an average increase in GDP growth of 2.2% per year.

Another paper from the AICD series, Steinbucks and Foster (2010), looks instead at the determinants of generator ownership in the continent. Despite the fact that in-house generation is up to three times more expensive than acquiring energy from the grid, the authors claim that own-generation would remain high even if power supplies were to become much more reliable.

While benefits from generator ownership exist, especially in term of reduction of lost load, the difference between cost and benefits is not found to be statistically significant.

Escribano *et al.* (2009) explore the effects of different infrastructure provision in SSA on TFP growth of African manufacturing firms. They further distinguish between high/low income and quick/slow growing countries using an unbalanced panel of 26 African countries with data obtained from World Bank Enterprise Surveys (WBES)[9]. They find evidence that different types of infrastructure have different effects in countries at different stage of development, and that electricity problems are especially relevant for poorer countries. A similar study by Iimi (2011), which though looks at firms in Eastern Europe and Central Asia, reaches similar conclusions. The estimation of a trans-log cost function for 26 countries leads him to conclude that while ICT provision does not seem to have a significant effect on firm costs, frequent and lengthy power outages create a significant burden for many firms and so does the time required to restore water provision after a disservice.

Dinkelman (2011) exploits geographical and timing differences in the South-African roll-out plan for rural-electrification to estimate how big a role it plays in fostering employment. Looking at household level variables over a 5 year time period, the author finds that the strongest employment effect is for female workers in their thirties and forties (that is, when child care is less relevant) coming from middle-income households. The link between rural electrification and development is explored also by Cook (2011), who reviews the existing literature to assess feasibility and effectiveness of donor funded scheme. The conclusion is that many schemes have ended up providing cheaper access in rural areas to those who were already

---

[9] A description of the WBES questionnaires is furnished in the next section.

better off, so that the majority of poor households are still unable to afford the connection costs, in part because operators are not sufficiently incentivised to provide cheap access[10].

A different analysis on the effect of rural electrification is that of Peters and Vance (2010), who compare the profits of micro-enterprises in two areas of Northern Benin, one with and one without grid access, through propensity score matching. Their data, collected for the baseline study of a rural-electrification project implemented by the German development agency, cover 276 firms in five electrified and five non-electrified villages. Although the authors do not find any significant difference in the profitability between firms in non-electrified villages and those in electrified villages which existed before grid-expansion, a positive effect is identified for firms created after that moment, which tend to use more capital intensive and electricity-reliant production techniques.

Another paper which focuses on the role of electricity constraints for firms in developing countries is Alby, Dethier and Straub (2011). They develop a model, based on Holmstrom and Tirole (1997), to understand the extent to which size and sectoral distributions are influenced by poor electricity network. Using the WBES for 62 developing countries, the authors conclude that electricity intensive firms are those most negatively affected by poor energy infrastructure and that the initial level of assets and access to finance play a crucial role for firm survival and for firm growth.

Moyo (2012a and 2012b) also evaluates the effect of infrastructure deficiencies in SSA. In Moyo (2012a) the author focuses on export performance, assessing the different impacts of

[10] The extents to which cross-subsidization can be exploited are also discussed, and so it is the role of rural energy agency and of investments in complementary infrastructures.

disruption from telecommunications, water, electricity, transport-network and border customs. Moyo (2012b) concentrates instead on the effect of power cuts on Nigerian manufacturing firms. In both cases, the author concludes that there is enough evidence of a sufficiently large negative effect from disruptions to justify the redirection of government budget towards improving infrastructure.

Lipscomb *et al.* (2013) simulate instead different developments of the Brazilian electricity grid in the period 1960-2000, exploiting topographic placement of hydropower plants to determine how big the gains from proper targeting of infrastructure development sites are[11]. Their results suggest that classical general-equilibrium models fail to take into account the endogenous placement of infrastructure, and consequently massively underestimate the real return from electrification.

How varying electricity costs affect firms' performance in India is the target of Abeberese (2013). The author constructs an instrument based on the interaction between the wholesale price of charcoal to utilities and the share of thermal generation installed in a given Indian state. The results suggest that production technologies are influenced by the price of electricity: as the latter rises, firms move to less energy and machinery intense technologies, reducing both output and labour productivity, therefore distorting the industrial structure of the country.

Apeaning and Thollander (2013) provide a case study of the biggest Ghanaian industrial area focusing on a similar issue, which is barriers to energy efficiency. The results suggest two sets of factors being the main influences: the "lack of budget funding" and "access to capital" on

[11] I.e. one which takes into account the structural transformations which are likely to interest the region as opposed to purely geography based cost consideration.

the one hand, and volatility of energy prices, which dampens electricity use, on the other. In the same year, Andersen and Dalgaard (2013) try to estimate how relevant poor electricity infrastructure was in reducing African GDP per capita growth, exploiting variation in lightning density as an instrument for outages, concluding that it was indeed a drag on economic performance.

Oseni and Pollitt (2013) also look at connections between firms' development and quality of electricity supply in Africa, this time through the lenses of generator ownership. Given the high frequency of power outages in the continent, they are not surprise by finding that back-up capacity represents 20% of all installed capacity in SSA. Following the literature, the authors test two main hypotheses, namely that back-up capacity is able to mitigate the majority of outage costs and that larger firms integrated in international markets suffer fewer losses from outages. The first hypothesis is rejected by the data, as unmitigated losses represent the greatest share of outage cost due to incomplete back-up by the majority of firms. The second hypothesis is then also rejected: although it is true that bigger firms and exporters have a generally higher demand for generators, this is still not sufficient to achieve complete backup.

One of the best recent paper looking at the effect of frequent outages on firm productivity is Allcott *et al.* (2014). The authors, working on the Indian manufacturing sector, also suffering from frequent and prolonged outages, are able to develop a detailed model that integrates electricity as a Leonthieff input in the production function relying on extremely specific data from 22 textile plants in an industrial park, suffering from both scheduled and unscheduled power holidays. The model is then extended to the whole of India thanks to a long panel dating back to 1992 and to the availability of energy data from India Central Electricity Authority. To avoid endogeneity concerns, the level of outages is instrumented by the share of electricity demand met by hydro production for each state in each year, as these data are also available through the Indian central energy agency. The final effect of outages depends on firm characteristics: firms with back-up capacity face increased costs, as self-generating is more expensive than acquiring electricity from the grid, often leading to sup-optimal combination of inputs. The firms without generator are though those facing the blunt of the outage cost: not only need these firms to stop production altogether when the supply of electricity is interrupted but they also waste all non-storable inputs which have already been purchased.

Another recent work focusing on power outages using Indian data is Alam (2014). The author focuses on the heterogeneous firms' effects of infrastructure quality using detailed firm data for three important sectors of the Indian economy: one which is non-electricity dependent (brick making) and two for which electricity is a more vital inputs, differentiated by the availability of strategies other than acquiring a generator for adapting to poor electricity delivery (rice and still mils, with the former representing the sector with alternative strategies). As a proxy for outages the author develops a measure based on the exploitation of different cleaning procedures of nigh-light data images from the Defense Meteorological Satellite Program – Operational Linescan System, which permits her to identify variations at the district-level. Her model, points toward the irrelevance of power outages for non-electricity intensive sector and of their short-run changes on firms' decision to acquire a generator. Furthermore, strategies other than back-up capacity are also shown to be very relevant as rice mills, which can increase capital productivity at the expense of raw material costs, are able to completely shield themselves from outages' damage leading instead to a loss of productivity in the steel sector.

Finally, Geginat and Ramalho (2015) consider how lengthy and complex procedures to acquire an electricity connection might also be playing a role in constraining economic activities. The main scope of the paper is to construct an index for the ease of getting electricity, combining the number of procedures required, the average time before the connection is operational and the overall cost (excluding bribes). The authors rely on 600 interviews with respondents located in the capitals of 183 countries, all presented with the same general "case study" to obtain comparable data. While the number of procedures is fairly stable in their sample, the time required exhibit a much greater variation, especially in SSA where the range goes from the 91 days in the Mauritius to 455 days in Guinea Bissau. As the majority of the energy providers in the considered countries are state-owned, it is difficult to differentiate the overall quality of the service between the structures of ownership, although there are some signs that private providers are slightly cheaper.

To conclude, from this literature review we can see how the effect of infrastructure on macro and micro level variables has been a long standing question in the economic field. Most of the initial studies looked at how the historical rates of infrastructure investments affected national growth in developed countries, although it is also possible to find early examples of micro level studies investigating the connection between infrastructure provision and firms' performance in developing countries. The number of these latter studies has been growing over the last 15 years thanks to the increased availability of firm level data, which opened the possibility of considering a much wider set of questions. However, evolution in econometric techniques also led to a continuous reassessment of cross-county macro-relationships.

At the macro level, most of the literature agrees in individuating a positive effect of infrastructure investment on GDP growth, although with different qualifications. Some stresses the importance of analysing separately different types of infrastructure, while others recognise the need of dividing infrastructure quantity from its quality, or to specifically consider the role that a continuous institutional supervision plays in achieving the best outcome from those investments. These different perspectives reflect the main issues faced by this strand of the literature, namely how infrastructure investment is often endogenous to economic growth and how difficult it is to account for the heterogeneity of these investments, both in terms of quality and type.

Studies at the micro level investigate a vast array of relationships, and almost all find a significant effect of some type of infrastructure on the variable under consideration, be that the effect of cleaned water provision on house-hold and microenterprise or that of ICT investment on adoption of new agricultural technology. Most studies covered in this section have dealt with issues related to energy infrastructure, as that is the subject of the current analysis. Three of the papers reviewed investigated the consequences of expanding rural electricity network, both from the point of view of the different type of schemes involved and of their outcomes, concluding that although positive effects exist, they might interest only certain part of the population or some of the firms located in those areas. Other analyses were instead concerned with generator ownership, especially in the African context. From these studies we can see how, despite the fact that electricity generation from back-up capacity is more costly than its acquisition from the grid, or that firms seldom achieve complete back-up regardless of their size, the acquisition of a generator might be the only way for them to access credit and hence expand their operation.

The remaining group of studies reviewed is that covering the direct effect of energy variables, either energy price or frequency of blackouts, on firms' productivity and sales. Three of these studies use Indian panel data, while the other relies on a panel of Chinese firms, and all of them treat the energy variable as endogenous to productivity, leading to instrumental variable estimations. The reasons quoted for these endogeneity concerns are usually a mix between unobservable firms' characteristics, measurement error in the outage variable, the quality of the electricity infrastructure affecting the initial decision of plant location and the existence of policies influencing both firms' productivity and their energy demand.

Due to the scarcity of firm level and energy data for the Africa, this type of studies has never before been applied to the continent, despite the fact the electricity shortages are endemic in most of its economies. It is this research gap which we intend to fill, contributing, to the best of our knowledge, with the first instrumental variable estimate of this relationship for SSA, thereby providing African policymakers with a reliable estimate of the detrimental effect of power outages.

1.3 Data and Summary Statistics

1.3.1 Questionnaire structure, standardization and transformation.

The main source of data for this study are the WBES, carried out by the World Bank with the intention of gathering information on the business climate in developing countries. Information include classical balance sheet data as well as continent and country specific firms' characteristics[12]. To construct our dataset, we select the most recent available data for all non-island Sub-Saharan African states, which gave us 38 countries with data collected between 2006 and 2014. During this time, the World Bank out-sourced the study to two different companies that used two different questionnaires, although the main difference between the two was the question order.

However, there are other differences that we need to take into account. The sample for each country study was selected from the universe of eligible firms through a procedure of stratified random sampling with replacement. The form of questionnaire that was submitted to the firm is determined by the sector of activity: manufacturing firms receive a more detailed version than those in service sectors[13]. Unfortunately, some of the differences in the questionnaires given to firms tend to be on their cost structure and infrastructural quality, exactly the variables of interest for our study.

In the questionnaires given to manufacturing firms, the section that relates to infrastructure contains data on electricity, water and telecommunications services provided to the firm; while the service questionnaires excludes information on water infrastructure and it has inferior data on electricity provision. Of the 26 questions, 9 are on energy-related issues: *1) Has the firm ever applied for electricity connection; 2) How long did it have to wait for the connection; 3) Was a bribe necessary to obtain the connection; 4) Have you experienced power outages; 5) What is the average number of power outage per month; 6) What is the average length of a power outage; 7) What is the loss in production due to the outages; 8) Does the firm own a generator; 9) What percentage of electricity is generated in-house.*

[12] While some questions or sections, e.g. the effect of HIV on the workforce, are continent wide.
[13] This is the last stratum. The first two strata are geographic location and firm size.

Finally, there are differences in the way in which capital expenditure is treated in different questionnaires and strata. The older version of the questionnaire includes information on capital expenditure only for manufacturing firms but not for service and retail; versions of the questionnaire used between 2006 and 2013 contain information on capital expenditure for the whole sample, but only for economies above a given size; after 2013 there are no more questions about annual depreciation.

To work with cross country level data collected over a range of years it was necessary to perform various transformations of all the monetary data in the sample. First, all prices were deflated to 2005 levels using the GDP deflator from the World Bank. Second, they were transformed in international dollars using the Purchasing Power Parity (PPP) Index, again obtained from the World Bank.

We also constructed a number of variables using the information contained in the questionnaires. Regarding the explanatory variables, the average number of outages per year was obtained by multiplying the average number of outages per month by 12, while the yearly hours of outage were obtained multiplying the latter by the average length of an outage. The two measures should ideally pick up the different problems that an unexpected disruption of electricity causes to production: first, every time the service stops, a machine which was supposed to be turned off may be damaged by the sudden interruption of electricity (number of outages); second, without the presence of a generator every activity which requires electricity cannot be performed (hours of outage)[14].

[14] To reduce over-dispersion in both outage measures, the number of power outages has been capped at ten per day, while to the hours of outage the upper boundary of the number of hours in a year has also been imposed. While the second assumption is a logical requirement, we have dropped the first and the results remain almost unchanged.

Firms were considered exporters (0/1 dummy variable) if they exported at least 1% of their annual production, with no distinction being made amongst trade partners. Dummy variables were created for small (< 20 employees), medium (20-100 employees), large (100-300 employees) and very large firms (> 300 employees), for firms with access to credit and, finally, for publically traded companies. Finally, total sale is used as independent variable for all baseline results and all but one robustness checks, in which a measure of total factor productivity (TFP) was also created relying on the procedure used by Cui, Lapan and Moschini (2012)[15].

1.3.2 Continent wide summary statistics.

The complete sample is composed of 13,310 firms[16], of which 6,164 belong to manufacturing sectors (the remaining are services of various kinds). The firms are not equally divided amongst the 38 different countries: the country accounting for the highest number of firms is the Republic of South Africa (1,056), followed by Nigeria (1,029) and Senegal (625). Of the remaining 35 countries, 14 contribute approximately 150 firms each while all the others have between 200 and 650 firms. The first thing to notice is how, unsurprisingly, there are great differences between the countries in the sample, from the industrial structure to the reliability of their energy infrastructure. Just to give a couple of examples, the share of manufacturing firms ranges from 20% in Burkina Faso to 66% in South Africa, while the average number of employees (of the firms in the samples) ranges from 14.6 in Guinea Bissau to 184.5 in Malawi. This can also be seen by the diverse cost structure of the average firm in the different countries

[15] See section 4 for an exposition of their procedure.
[16] We have dropped the top one percentile by total sale to avoid the effect of outliers.

(see Figure 1.2):

[Figure 1.2 about here]

In light of these differences, the only figures which will be presented on a continental scale are those relative to the diverse levels of performance of energy infrastructure and those relative to generator ownership, while a brief description of the different countries sampled shall follow. It must also be noted that all the following figures represent a description of the overall sample (in this section) or of that of each country (in the next section), i.e. they are not population weighted and cannot be taken as representative for the country or the continent.[17]

The country which suffer the lowest number of outages is Malawi (12.44 per year on average), while the maximum is reached in Nigeria (more than a 1,000 outages per year on average). If we consider the average length of an outage, the range is from 1.35 hours in Niger to more than a day in the Republic of Congo. The combination of frequent and long outages means that 9 countries in our sample have firms who face on average more than a thousand hours of outages each year. The reasons for such a frequent failure of the electricity networks are various: the improved economic performance of most countries over the last 15 years led to a quick growth in energy demand, which far outstripped the investment in energy supply, further increasing the already existing gap; installed generation capacity is often not properly maintained, and so are transmission and distribution networks, leading to high transmission losses; frequent droughts in certain regions often lead to the inability of effectively exploit the hydropower capacity on

[17] Although the WBES offer specific weights for each country, we did not possess the information for reconstructing the weight at the supra-national level, which is our level of interest.

which many countries rely for a significant share of their electricity[18] (Karekezi 2009, IEA 2014, IEA 2016).

[Figure 1.3 about here]

The perception of the damage caused by poor energy infrastructure is also varied. In each questionnaire the business owner was asked to estimate the loss due to outages in terms of percentage of their output and how relevant the electricity provision situation was in constraining their expansion. The correlation between losses and different measures of outage are not as strong as one might expect, suggesting that the industrial sector to which a firm belongs is important to determine how detrimental it is to have an unstable energy provision[19]. From an examination of the data it seems probable that the figures reported suffer from severe inaccuracies, which might be due to lack of time in compiling the questionnaires, difficulties in measuring output or political reasons (see Oseni and Pollitt 2013 for a discussion). For example, the average hours of outage for a firm without generator reporting no losses due to outages are 364, and given these firms' inability to keep up production in these periods (equivalent to a whole month assuming a 12-hours working-day) this seems highly unlikely. For this reason we have decided to present the numbers reported in the questionnaire but we will base our analysis on the direct estimation of the effect of outages on sales, without trying to validate the respondents' figures. As it can be seen in Figure 1.4, countries which are hit the most in terms of lost output are the Republic of Congo (27.7% of output lost due to outages) and the Central

[18] For recent examples of such occurrences see http://www.bbc.com/news/world-africa-34491984 , http://www.nytimes.com/2016/04/13/world/africa/zambia-drought-climate-change-economy.html?_r=0 or http://www.forbes.com/sites/riskmap/2016/02/04/drought-in-southern-africa-threatens-social-unrest-power-supply-challenges/#1b74c0be4fdd

[19] The correlation ranges from a minimum of 0.18 between losses and number of outages and a maximum of 0.32 between losses and hour of outages.

African Republic (26.1%)[20], while from Figure 1.5 the damages sustained seems to be linearly related to the hours of disservice.

[Figure 1.4 and 1.5 about here]

The correlations between outage measures and the percentage of firms that find electricity the main obstacle for expansion are positive but very weak, ranging from 0.05 with respect to the number of outages and 0.15 with respect to the hours. The three countries in which more firms claim to be constrained mainly by electricity are Senegal (42.7% of firms in the sample), Central African Republic (41.5%) and Tanzania (32.3%), while the countries for which energy-related issues are the least relevant are Mauritania (0.8%), Sudan (0.7%) and Guinea (0.04%)[21].

[Figure 1.6 about here]

The final set of figures that we report at a continental level are those for generator ownership. As is pointed out by Foster and Steinbucks (2009), SSA shares of in-house generation are not particularly high when considered on a continental level. However, there is a great deal of variance across countries. While Foster and Steinbuks were able to track the relevance of in-house generation at the firm level as a proportion of the installed capacity from the information contained in an older version of the WBES, this is no longer possible with the current questionnaire. Nevertheless, looking at the continental level of ownership (48%) still hides big

[20] It is though worth remembering that all of the estimates on lost output are self-reported and therefore likely to be somehow biased.

[21] Given the way in which the question was phrased, *"Which of the element of the business environment ... currently represents the biggest obstacle for the establishment? ",* low percentages do not imply that electricity is not a major obstacle, just not *the* major obstacle.

differences amongst countries like Ivory Coast, Mozambique or South Africa, in which the ownership is less than 20%, and others like Chad, Congo or Nigeria in which it is above 80%. The picture is similar if we look at the percentages of electricity used from own-generation by the firm. The continental average of 32% is mostly driven by a couple of countries characterized by a high generator ownership and very unreliable electricity supply (e.g. Guinea Bissau, Liberia or Nigeria, all above 60%) and hides deep differences across SSA.

Figure 1.7 presents the breakdown of generator ownership and electricity generated in-house by incidence of outages. As expected, both the percentage of electricity generated in-house and the percentage of generator ownership increase in the hours of outage. In Figure 1.8 we show instead the data relative to generator ownership and the share of electricity self-generated across size. As it can be seen, the share of firms owning a generator increases with size, but the share of electricity self-generated decreases. This implies that while larger firms are more likely to have back up option the relevance of self-generation is higher for smaller firm.[22].

[Figure 1.7 and 1.8 about here]

1.3.3 Country specific summary statistics.

In the following section a brief analysis of the industrial structure shall be provided for each country in the sample.

[22] Probably because of easier access to credit, more than 40% of firms above 100 employees owns a generator versus an average lower than 20% for firms with less than 20 employees.

Angola.

In the sample there are 326 Angolan firms, 58% of which are small, 31% medium, 7% large and 4% very large. The majority of the firms in the sample (274) are located in Luanda, the country capital, while the remaining is divided between Benguela and Huambo. Non-manufacturing firms predominate (60%) - with retail and wholesale sectors accounting for 35% - while the main manufacturing sector is food processing (14 %). The average number of outages in the country is 73.1 per year, while the average number of hours is 783.4. There is little geographical variation in the relevance of the outage phenomenon, and similarly comparable is the incidence across firm size.

[Figure 1.9 about here]

Benin.

Benin contributes to the sample with 146 firms, almost all located in the economic capital of Cotonou (125), while unfortunately there is no information on the location of the remaining 21. The vast majority of the sample is composed of small firms (72.6%); of the remaining, 19.2% are medium sized, 5.4% large and 2.8% very large. The sample is almost equally divided between manufacturing and non-manufacturing firms (68 the former, 78 the latter) and there are no information on sectoral division. The national average of hours of outage is a 1361, but it is worth noticing that the rate of reply to the energy infrastructure questions is pretty low in the country (25%); moreover, there appears to be a strong difference across size in the relevance of outage, with small firms experiencing more than twice as many outages as medium sized ones.

Botswana.

Botswanan firms in the sample are 233, divided between the capital Gaborone (200) and Francistown (33). In this case, the share of non-manufacturing firms is almost twice as high as that of manufacturing (145 vs. 88), with retail accounting for 30% of firms in the sample (the main manufacturing sectors are garments, 6%, and metals and fabricated metal products, 4.3%). The sample is composed by 52.3% of small firms, 31.8% of medium firms, 10.3% of large firms and 5.6% of very large firms. The average number of hour of outage per year is 154, and while it does not appear to vary much with size (a part from very large firms which are those more hardly hit) it shows quite a strong geographical variation: the average is 75 hours in Francistown and 167 in Gaborone.

[Figure 1.10 about here]

Burkina Faso.

Burkina Faso is another example of a country with a low industrial base. Of the 375 firms in the sample, only 70 belong to a manufacturing sector, while retail and construction sector accounts for almost 50% of the firms. The biggest manufacturing sector are food process (5.54%) and metals and fabricated metal product (4.66%), not surprisingly if we consider the relevance of the gold mining for the country. In term of geographical diversification, the vast majority (294) of the firm comes from Ouagadougou, the state capital, with the remaining being situated in Bobo Dioulassou, the second city. There is a noticeable and high geographical variation in outage relevance, with the two cities having respectively an average of 273 and 413 hours. Similarly, there is not much variation in outage across firm dimension, again excluding very large firm which appears to be much less exposed (115 hours against 383 hours for smaller ones).

[Figure 1.11 about here]

Burundi.

Burundi's industrial structure is strongly predominated by small and medium firms (94%), with almost no firms above the 100 employees threshold (8 out of 152), with a small majority of manufacture (51%). Firms involved in food processing activity represent a fifth of the sample, followed in importance by wholesale (12.5%) and hotel and restaurant (11.2%). The capital Bujumbura hosts two thirds of the firm (the remainder are equally split between Gitega and Ngozi), and there is little geographical variation in outage incidence. There is some variation across firm size but no clear pattern is identifiable.

[Figure 1.12 about here]

Cameroon.

Cameroon contributes to the sample with 347 firms, 207 of them located in Douala, the biggest city in the country and the richest of the whole Central African community (106 of the remaining are in the capital Yaounde, 34 in Bafoussam). The retail sector accounts for 42% of firms in the sample, and as in many other cases the main manufacturing sector is food and beverage processing (8.3%). Almost half of the sample is composed of small firms (48%), followed by medium (32.6%). There is little geographical variation in the relevance of outages, which stand at a national average of 135 events equal to 366 hours without provision per year.

Again, for very large firms the outage phenomenon appears to be starkly less relevant (97 hours per year) than for the rest of the sample (376 hours).

[Figure 1.13 about here]

Central African Republic (CAR).

CAR presents one of the worst energy outlooks of the whole continent, with the fourth highest average outages per year (2671 hours). Almost all of the firms are located in the capital Bangui (136 out of 142), while the remaining 6 are in Berberati, the third largest city of the country (and oddly enough none of these 8 firms ever experienced an hour of outage). As for Benin, there is no information on sectoral division of firms apart from manufacturing (24%) and services. The majority of the sample is composed of small firms (70%), with just 8 large firms (5.6%) and 2 with more than 300 employees, which in contrast with many other countries in the sample appear to be the more hardly hit by outages (3420 hours).

Chad.

Chad is the third country in the sample for duration of outages, with a national average at a staggering 2585 hours per year. All 150 firms are located in the capital N'djamena, so that no information on geographical variation is available. Equally unavailable are information on the sectoral division of the firms, apart from the usual between manufacturing (49%) and services. Chad is another example of a country in which the absolute minority of very large firms (2, 1.3%) is the group most hardly hit by outages (5526 hours), while the majority of small firms (64%) is much less affected by the phenomenon (2689 hours).

Republic of Congo.

The Republic of Congo closes the group of central African countries incurring in the heaviest electricity problem, as it is the only country in the sample presenting an average of hour of outages higher than 4000 hours (4145.6). The sample (149 firms) is composed for more than a half by small companies (61%), followed by medium (26.9%), large (4.7%) and very large (7.4%). Despite the high GDP contribution of oil and timber sector, 4 of the top 5 sectors in the sample are in service (72%), with food processing being the main manufacturing one (8.7%). There is little size variation in the relevance of the outage phenomenon, while geographical variation between the capital Brazzaville and the second city Pointe-Noire is more relevant as the latter experiences 40% more hours of outages than the former (4822 hours vs. 2991).

[Figure 1.14 about here]

Cote d'Ivoire.

Cote d'Ivoire contributes to the sample with 508 firms, 75.6% of which have less than 20 employees. The majority of the sample belongs to service sectors (62%) – which account for a little more than 50% of the GDP – with the most relevant sectors being retail (23.4%), while the main manufacturing one is food and beverage processing (10.8%). The sample shows quite a strong variation of outages incidence across both size and geographical location. The vast majority (384) of the firms are located in Abidjan (economic centre and former capital) which faces 280 hours of outage per year, less than a half of those faced by the 41 firms in San Pedro (603 hours), while the 2 firms in the current capital of Yamoussoukro face less than 20. Similarly, while small and medium firms are hit by approximately the same amount of electricity disservice, large and very large firms are much more severely hit.

[Figure 1.15 about here]

Democratic Republic of Congo (DRC).

The DRC accounts for 573 firms, with a slight majority in the manufacturing sector (51.5%, with food and beverages being the most relevant at 15.9%), while the retail sector alone accounts for 24%. As usual, the vast majority of the sample is composed of small firms (74%), with large and very large adding up to 5%. The country energy outlook is as dire as that of many other central African states, with the national average of outage at almost 1000 hours. Kisangani and Matadi experience on average half the hours of outage of Kinshasa and Lubumbashi, despite the status of political capital and of biggest mining centre of the latter. As in other countries, small and medium companies face at least twice as many hours of outage as large and very large ones.

[Figure 1.16 about here]

Eritrea.

The Eritrean sample is composed of 135 firms divided across the three main cities of Asmara, Mendefera and Massawa. The country belongs to the group of those with no sectoral information available, apart from the division between manufacturing (54%) and services. Eritrea seems to be one of the countries less affected by outages, with a national average of 180 hours per year. While the hours of outage seem to increase with size (with large firms having almost twice as many hours of outages than small firms), the only noticeable geographical difference is that between firms in Asmara/Mendefera and those in Massawa, with the formers facing less than 120 hours of outage per year and the latter more than 400.

Ethiopia.

Ethiopia participates in the sample with 522 firms, with a little less than a half belonging to manufacturing sectors (48%). As for most countries in the study, the retail sector accounts for the majority of firms in the sample (27.6%), followed by food and beverages processing (9.6%) and wholesale (7.5%). Again similarly to other countries, the sample is predominantly composed by small (53.4%) and medium (29.3%) firms. The national average of hours of outage is 570, and the phenomenon is equally spread across the country, maybe excluding the province of Tigray which shows a much smaller incidence (173 hours). The variation across size is very small, a part from very large firms which appear to be hit slightly more.

[Figure 1.17 about here]

Gabon.

Standing at 22.3%, the manufacturing firms' share of Gabon is the third lowest across the SSA sample. Of the top 5 sectors, which account altogether for 75% of firms in its sample, only one is in manufacturing (and that is unspecified manufacturing). This stands somehow in contrast with the GDP composition of the country, which enjoys a big contribution from extractive industries (oil and timber). It is though worth noticing that the recent growth they fuelled had scarce redistributive effect and 60% of the labour force is still employed in agriculture. Unsurprisingly, the majority of the sample is composed of small and medium enterprises (87% of the sample cumulatively).The national average of 570 hours of outage per year is not amongst the highest in the continent, even though it shows a pretty strong variation across firm size (with large firms experiencing the bulk of it) and some geographical variation (Port Gentil, second city in the country and main seaport/site of extractive industries, experiences less disruptions than the national average).

[Figure 1.18 about here]

Ghana.

Despite a national average of 874 hours of outage per year, Ghana is far from having the worst energy outlook in Western Africa. The variation across size – in an industrial structure vastly predominated by small firms (71%) – is fairly small, with large firms experiencing less disruption than the country average. Its geographical variation across the 3 different productive centres included in the sample (Accra, Takoradi and Tamale) is also limited, with firms located in Takoradi being hit slightly harder. Of the 555 firms, 59.6% belongs to manufacturing sectors, with both "food and beverage", "fabricated metal products" and "publishing, printing and recorded media" in the top 5 sectors.

[Figure 1.19 about here]

Guinea.

Guinea has the third highest average of hours of outage in Western Africa after Nigeria and Sierra Leone with more than 2000 hours of electricity dysfunction per year. At the same time it has one of the highest share of manufacturing firms in the sample (60.5% of the 223), with medium sized firms (7.1% of the sample, while small firms are almost 90%) appearing to be those facing the highest incidence of outages, with an average well above the national one (2820 hours). The vast majority of the firms sampled come from the capital Conakry, so that it is difficult to understand the relevance of geographical variation of outages (the only other city sampled is Kindia, which shows a lower than the average outage mean, possibly due to its role of military headquarter of the country).

Guinea Bissau.

The small country of Guinea Bissau contributes to the sample with 157 firms, 68.15% of which belongs to service sectors. Its industrial structure is predominated by small and medium firms, which account together for 98.7% of the sample, with just 2 firms with more than 100 employees and none with more than 300. The national average of hours of outage is 1276 and it shows little variation across size, while it is impossible to analyse its geographical variability given that all firms come from the capital Bissau.

Kenya.

Kenya has a national average of hours of outage per year equal to 535 hours, in line with the averages for Eastern Africa. The variation across firm size is low, with very large firms being the only category showing a noticeable departure from the national average (784 hours). Out of the 4 productive sites sampled, two shows a relevantly different average: Nakuru – fourth city of the country – with a sensibly higher level (884 hours) despite its status of main agricultural and diary production centre; Kisumu, second city in the Lake Victoria Basin, with a much lower average at 218 hours . The sample is almost equally divided between non-manufacturing (49.8%) and manufacturing firms, with the most relevant sector amongst the latter being as usual “food and beverage processing”.

[Figure 1.20 about here]

Lesotho.

Lesotho has one of the highest preponderance of large and very large firms (17% together) in the sample. 59% of the firms belong to service sectors, with the main manufacturing sectors being garments, food and beverage processing and textiles (10.8%, 6.5% and 5% respectively). All the firms sampled are based in the capital Maseru, so that it is impossible to estimate the geographical variation of outages relevance (which stands at a national average of 365 hours per year). Size wise, the only noticeable difference is with regard to very large firms, which seems to experience a much higher incidence of outage (646 hours on average).

Liberia.

Liberia presents a very peculiar situation. The survey has been carried out in 2009, which is 3 years after the elections of 2006 that de facto concluded the peace process that began in 2003. The 149 firms in the sample are almost evenly divided between manufacturing (49%) and service (51%), although little information about further sectoral division is available as half of them were reported as either general manufacturing or service. The industrial structure is dominated by small firms (82%), with just one firm with more than 300 employees having being sampled. The low incidence of outage (251 hours per year) is surely dependent on the incredibly low access to public electricity (which is somewhere around 1% of the population), available only in the capital Monrovia (while overall access to electricity including self-generation is 10% of population in urban areas and 2% in rural). The aim of understanding the geographical variability of the phenomenon is further impeded by the incredibly high fragmentation of the sample across small villages, which is probably due to survey implementation problems and enumerators training issues.

[Figure 1.21 about here]

Malawi

Malawi presents a more varied industrial structure than many other countries, at least under the point of view of size diversification, with large and very large firms accounting together for 30% of the sample. The division between manufacturing and service firms is almost even (49% and 51% respectively), but we have unfortunately no information about the sectoral division. The national average of 74 hours of outage per year is the third lowest in SSA, even though it shows quite a strong variation across firm size (standing at a mean value of 49 hours for small firms and 130 for large firms). Unfortunately, even for Malawi it is hard to perform a meaningful geographical analysis of differences in intensity of outage events due to high fragmentation of the sample.

Mali.

The structure of the Malian economy is similar to that of other Western African countries, with strong preponderance of small and medium firms (94% of the sample); of non-manufacturing sectors (57%) and of "food and beverage processing" and garments amongst the manufacturing. The national average of 555 hours of outage per year is still in the bottom half of the overall distribution; it shows considerable geographical variation, with an average of 1631 hours in Sikasso and one of 148 in Segou. Similarly varied is the incidence across sizes, with very large firms being the most hardly hit.

[Figure 1.22 about here]

Mauritania.

Mauritania is a further example of an industrial structure almost completely composed of small and medium firms (83% and 16% respectively). As in the majority of other countries in the sample, non-manufacturing sectors are more relevant than the manufacturing ones, the latter representing 34.3% of the sample. The average level of outage is 165 hours per year (nationally), with that of Ivory Coast one of the lowest in Western Africa. Firms located in Nouadibhou (second city of the country and only other included a part from the capital Nouakchott) seems to be more hardly hit, and the incidence of the outages phenomenon also increases with the size (with a mean value of 128 for small firms and of 684 for large ones).

[Figure 1.23 about here]

Mozambique.

The 597 Mozambican firms were sampled across the three main cities of Beira, Maputo and Nampula (even though some firms are located in Matola, which lies 12 km from Maputo). The majority of them are either small or medium firms (73% and 22% respectively), with only 4 firms passing the threshold of 300 employees. While manufacturing firms make up more than half the sample - retail alone accounts for 32%, there are 3 manufacturing sectors in the top 5 (food and beverages processing; metals and fabricated metal products; garments). The national average of hour of outage is 162 hours, and while it shows very little geographical variability it appears to be of decreasing importance in size of the firm.

[Figure 1.24 about here]

Namibia.

Namibia contributes to the sample with 355 firms, with the vast majority being in service sectors (74%) and retail and construction accounting for more than 50% alone. The size distribution of firms is similar to that of many other African economies, with small and medium firms accounting together for 95% of those in the sample. The national average of hours outage stands at 192 hours, with quite a strong difference in average disruption between the cities sampled, as the capital Windhoek experiences more than twice as many hours as the other two cities (Walvis Bay and Oshakati). Its relevance appears also to vary across size, although no clear pattern can be identified.

[Figure 1.25 about here]

Niger.

Niger belongs to the group of small economies for which no information about sectoral distribution of firms is available. Of the 137 in the sample, 95% is either small or medium sized, with 57 firms (41.6%) belonging to manufacturing sectors. While the average of 482 hours of outage per year is not low in absolute terms, it is still in one of the lower in the region, even though it shows a noticeable geographical variation, with the average increasing to 743 hours if only the firms located in the market town of Maradi, a big agricultural hub, are considered. The phenomenon also appears to increase in relevance with the increase in size.

Nigeria.

Nigeria has the second biggest sample, with 1,029 firms divided across nine cities. The industrial structure is though fairly similar to that of the other Western African countries, with a very high incidence of small and medium firms (73% the former, 18% the latter). Even under the point of view of the sectoral division, Nigeria does not depart too strongly from other sampled countries: manufacturing firms account for slightly more than half of the sample (63%), with “food and beverage processing” and “publishing, printing and recorded media” being the two biggest sectors. The national average of 3609 hours of outage is the second highest in the continent, and it shows little variation across size of firms (excluding the large firms which appear to be more interested by the phenomenon). On the other hand, the geographical variability is extremely high, ranging from an average of 2409 hours in the capital Abuja to one of 5037 hours in Kano.

[Figure 1.26 about here]

Rwanda.

The small Eastern African economy has the lowest average of hour of outagess in the region (323 hours per year), which appears to be of decreasing importance in firm size (geographical diversification is hard to appreciate as only 3 firms are located outside the country capital). Unfortunately there is no information about the sectoral division of the 191 firms sampled (51% of small size and 36% of medium) a part from the classic division between manufacturing sector (36%) and services.

Senegal.

Senegal contributes to the sample with 625 firms, 56% of which belong to services. The main manufacturing sectors are "food and beverage processing" (accounting for 13% of firms sampled) and garments (accounting for 8%). As in almost all the other countries in the region, the highest share of the sample is composed by small and medium firms (88% and 9% respectively). The mean number of hours of outage stands at 852 per year, and appears to be decreasing in intensity with the size of the firm (while showing also some degree of geographical variation, as the mean for Saint-Louis and Kaolack, 2 of the 4 cities sampled, is well below the national one).

[Figure 1.27 about here]

Sierra Leone.

Sierra Leone presents the 5th worse energy outlook in the whole of SSA, with a national average of 2046 hours of outage per year. The variation is almost inexistent both across size and geographically, demonstrating how the civil war left the whole of the country with very poor energy infrastructure (fact which might also be noted looking at the share of electricity expenditure in the cost structure of firms in the country, much higher than the continental average). The vast majority of the firms in the sample is either of small (76%) or of medium (17%) size, with 46% of them belonging to manufacturing sectors (the most relevant of which are garments and "food and beverage processing").

[Figure 1.28 about here]

Republic of South Africa (RSA).

The RSA has the biggest sample (1,062 firms) in the study and the 4th lowest mean of outages (118 hours per year). The share of small and medium firms is in line with that of Southern Africa (82% cumulatively), while that of manufacturing is one of the highest (66%). Four of the top five sectors are in fact manufacturing, with the chemical sector accounting for 8% of the firms (RSA hosts 24% of the chemical firms in the whole SSA sample). There is some variability across size in the incidence of outages, with smaller firms being hit less and very large firms more, but not much geographical variation if we exclude Port Elizabeth (which has an annual mean of 19 hours of outage, possibly in relation to its status as main hub of the automobile industry in RSA).

[Figure 1.29 about here]

Sudan

The first ever World Bank Enterprise Survey was run in Sudan in 2014. Of the 263 firms sampled, 60% are in service sectors, with the more important manufacturing sectors represented by chemicals (10.6%) and food and beverages processing (9.9%). As common throughout the study, the majority of firms are either small or medium (96.5% cumulatively), with only one firm passing the threshold of 300 employees. The average level of outage is 118 hours, the lowest in East Africa, with the two very large firms seemingly those hit the hardest (234 and 288 hours). It is not possible to analyse geographical variation as all firms are located in the capital Khartoum.

[Figure 1.30 about here]

Swaziland.

Swaziland's economy is heavily dependent on South Africa. Of the 302 firms sampled, almost 50% of them belong to the retail sector, with manufacturing sectors accounting for only 14% of firms, almost all in garments (6%) or "food and beverage processing" (5%). As common, the majority of firms are either small (73%) or medium (18%) sized. The latter appears to be slightly more hit by outages, experiencing on average 96 hours of disservice per year against a national mean of 68. Geographical variation is almost completely absent a part from a lower incidence in the capital Mbabane.

Tanzania.

Tanzania is one of the few countries in the sample in which the highest share of firms surveyed belongs to manufacturing sectors (59%), with the main sectors being "food and beverage processing" (12%); "wood and furniture" (16%) or garments (10%). As usual, small and medium firms are the majority of the sample (66% and 18% respectively). The mean number of hours of outage per year at a national level stands at 870, even though the cities of Arusha and Mbeya show a much lower incidence (447 hours and 455 respectively).

[Figure 1.31 about here]

Togo.

Togo is the last country for which no information on sectoral distribution is available. Of the 147 firms sampled, all located in the capital city of Lome, 37% belong to manufacturing sectors; with small and medium firms accounting respectively for 64% and 23% of the sample. The average level of outage of 773 hours per year is slightly lower than the regional average and the phenomenon appears to be more relevant for small and medium firms than it is for larger ones.

Uganda.

Uganda contributes to the sample with 505 firms divided across 6 cities. The majority of them is either small or medium (92% cumulatively), and manufacturing sectors account for 56%, with "food and beverage processing"; "metal and metal products" and "garments" being the most important of them. The national level of outage is of 1254 hours per year on average and shows some size variation, as large firms are those hit the most (2853 hours) and very large ones the least (591). The only significant geographical variation is the city of Mbale, in which outages seems to be much less relevant (258 hours on average).

[Figure 1.32 about here]

Zambia.

The Zambian sample (531 firms) is predominantly composed by small and medium firms (97% cumulatively), with only two firms passing the threshold of 300 employees. Manufacturing firms have a slight majority (51.4%) event though retail shops, hotels and restaurants accounts for 35% of the sample. The mean level of outages in the country is 255 hours per year and it is driven mainly by firms located in the capital Lusaka (52% of sampled firms), which experience a higher than average level of disservice, while the main mining centre Ndola is characterized by a much lower incidence (86 hours).

[Figure 1.33 about here]

Zimbabwe.

The last country included in the sample, Zimbabwe, has the highest incidence of outages (585 hours per year on average) in Southern Africa. This shows almost no size variation and little geographical variation, excluding firms located in Manicaland which experience almost twice as many hours of outage (1092). The sample (594 firms) is composed for 62% by manufacturing firms (mostly in "food and beverage processing" or "garmets") and shows a similar size structure to that of other countries, with small and medium sized firms accounting cumulatively for 78% of the sample.

[Figure 1.34 about here]

1.4. Productivity Analysis.

As previously described in the literature review, only a small number of studies focus on infrastructure quality in SSA at the firm level. This is partially due to the scarcity of existing data and the difficulties in obtaining new ones: many SSA countries experience frequent instability; political and institutional turmoil are often a constant, so that the environment cannot be considered the most conductive for interviews (cost consideration a side). At the same time, these same factors have reduced the institutional capacity of many national statistical offices over the years, so that it is sometimes hard even to obtain a master list of existing firms. Scarcity and/or unreliability of data imply that it is hard to work on a panel data basis at a continental level.

1.4.1 Methodology.

The main interest of the study is then to provide a first set of estimates of the firms' cost of outages in term of sales considering the vast majority of SSA countries. The first step shall be the decision of which measures of performance can be used given the available information. The main productivity indicator used in the economic literature is TFP, the calculation of which though requires information about capital expenditure[23] and almost always the availability of panel data which is not present in our case (notwithstanding other still unresolved issues with this type of indicator, see Cohen 2003). Cui, Lapan and Moschini (2012) have though developed a method to calculate TFP without information about capital expenditure: relying on the assumption that all firms in a given industry can be characterized by the same homogeneous production function, they are able to separate the contribution of the labour input from that of all the others, observables and unobservables. If we are then willing to assume that all firms in the same industry face the same price for all inputs, from the two assumptions follow that profit maximizing firms in the same sector will choose the same inputs' combination and therefore the unobservable component can be approximated by industry specific terms[24]. The methodology was developed for and applied to a United States' firm-level dataset, for which these assumptions were not excessively unrealistic. Given the summary statistics presented in the previous part it is though clear that the same cannot be said about the SSA context. Even though this measure shall be included in the study, the result obtained will be mostly analysed as robustness check.

[23] Available just for a minority of firms in the sample, see section 3.
[24] It is worth remembering that none of the above assumptions implies that firms in the same sector have the same productivity.

With the available figures, the most sensible decision is then to use as dependent variable the revenue from total sales - used in logarithmic form - which stands as a proxy for total output. While this might be a somehow crude measure, it has the advantage of being available for almost every firm in the sample[25].

The explanatory variables of main interest for the study are the two different measures of power outages obtainable from the questionnaires, which are the number of outages per year and the hours of outage per year. These will be included in logarithm form to take into account possible non-linearities in electricity disservice.[26]

All the classical control variables have also been included: size of the firm (as a series of dummies); exporter status, access to credit; age of the firm; percentage of foreign ownership; property structure (as dummy variable assuming a value of one if the firm has publicly traded share) and country dummies. The last set of information which shall be included as control is the 2-digit SIC code, which is however unavailable for firms in 12 countries[27]. A manufacturing sector dummy (we have always information about manufacture/service division) is then included alternatively to the SIC code to check if increasing the sample leads to changes in the results.

[Table 1.1 about here]

[25] As previously stated, all the monetary figures in the study have been deflated to 2005 level and then transformed into international dollars using PPP indexes.
[26] Versions of the model including the two different measures in both linear and quadratic forms have also been tested as robustness check.
[27] Benin, CAR, Chad, Eritrea, Guinea, Guinea Bissau, Malawi, Mauritania, Niger, Rwanda, Swaziland and Togo.

This leads us to estimate an empirical specification of the form:

$$Y_i = \alpha + \beta_1 X_i + \beta_{ji} Z_i + \varepsilon_i$$

where Y_i stands for the logarithm of total sale; X_i is the logarithm of one of the two different measures of outages; Z_i is a j-th dimension vector of control variables which includes all of the above. The model is estimated via OLS with robust standard errors.

As all the models are going to be estimated via OLS it is surely worth mentioning that some endogeneity concerns exist. These are due to four main causes: the first is that there might be error in the reported incidence of outages; the second is that there might be some unobservable firm characteristics which we cannot control for as panel data are available only for a small sub-sample of countries for which the majority of firms are not matched; the third that the initial decision of plant location might have been influenced by the quality of the electricity infrastructure; the fourth that policies which affect firm's performance might also affect black out levels. While the first one is inherent with the use of reported surveys and the second cannot be tackled in a cross-country setting, the third and fourth are more serious. We recognize the relevance of the issue, but it is unfortunately impossible to explore it further with the available data.

1.4.2 Results.

The benchmark model for the estimation is presented in Table 1.2.[28] As it can be seen, only one

[28] While the two different measures for outages are related to two different sets of effects on production they are correlated by construction and therefore always included one at a time.

of the outage measures, the logarithm of the number of outage, is significant, while all control variables have the expected signs and significance. It appears then that outages are indeed a drag on firms' activity, although not a huge one as an increase of 10% in the number of outages reduces firms' sales of only 0.8%. With regards to the other control variables, smaller firms show lower sales, which increase in size; exporters and firms with higher percentage of foreign ownership also show higher total sales; finally it can be noted that both having access to credit from a bank and a public share property structure have a positive effect on sales (although it must be remembered that all these variables have been included simply as controls as our main interest lies in the effect of the outages).

It is also interesting to investigate if we shall obtain different results dividing the sample between firms which own a generator and firms that do not, as the previously reported figure indicate the very high share of back-up ownership. Table 1.3 presents then the coefficients for firms which do not own a generator, while Table 1.4 presents those relative to firms with a generator.

[Table 1.3 and 1.4 about here]

As can be noted looking at Table 1.3, the picture for firms without a generator remains fairly consistent. Now both measures are negative and significant, even though only at 10% for the coefficient of the hours of outage. Moreover, the point estimates are always higher for firms without generators than they are for the overall sample, exactly as one would imagine. With regards to firms that own a generator, as in the overall sample only the number of outages is significant, and this time only at 10%. The low and partial significance for the sub-sample might be due to the fact that the availability of backup generation does not automatically coincide with its use. Adenikinju (2003) shows for the Nigerian case how electricity production with back-up generation can be up to three times more expensive than buying electricity from the grid, so that generators are not always used and even when used they will lead to a higher production cost.[29].

Moreover, comparing the estimates from tables 3 and 4 there are other differences that can be noted: the benefit of growing in size up to the threshold of 300 employees seems to be stronger for firms without a generator, while the opposite is true after that threshold, as point estimates for large firms are higher for firms without a generator, while the reverse is true for very large firms; the positive effect of access to credit is stronger for firms without generators, while the reverse is true for access to export markets, foreign ownership and permanence in the market (although even in this case it must be born in mind that the above variables were included as controls).

Firms that cannot afford a generator appear then to start off from a worse situation that than those who can: they are more exposed to electricity disservice, especially the inability to produce during blackout periods. Access to credit also seems to lead to different consequences between the two sub-samples, but those are not so easily interpretable[30]. While the coefficient on the hours of outage might seem low, it is useful to bear in mind the dismal infrastructure condition that firm in the sample face: on average, a firm without a generator is afflicted by 562 hours of outage per year. For it, a reduction to the average hours of outage in the Republic of

[29] The models for every dependent variable have been compared using the McFadden-Hausman specification tests and the null hypothesis of statistical equivalence between the coefficients has always been soundly rejected.
[30] This might be because ownership of a generator helps access to credit as it can be used as collateral but equally it might be easier to afford a generator once access to credit is obtained.

South Africa (118 hours), arguably the country with the best infrastructure in the sample, will entail an increase in sales of 3.2% (roughly 700,000 dollars at 2005 PPP).

There is though an important qualifications for these results: as stated in the previous section, the OLS framework in which we have worked does not allow us to tackle neither endogeneity nor unobserved heterogeneity. While concerns about the latter are unavoidable given the nature of the data, a solution for the former shall be presented in the following chapters of the thesis.

1.4.3 Robustness checks.

Apart from ensuring that the results are not driven by any particular country, from the deflators and PPP indexes used or on dropping the top percentile of sales and outages, two main robustness checks have been performed[31]. The first is to drop the industry code and use instead a dummy indicating if the firm belongs to a manufacturing sector, increasing the sample between 1,300 firms ca. and 250 firms ca. depending on the specification.[32]

[Table 1.5 about here]

In Table 1.5 we present the results for the whole sample and for both sub-samples of firms with and without a generator. The estimates show a very similar trend to the previous: the coefficients on the (log) hours of outage are never significant, those on the number of outages

[31] Linear and quadratic specifications have also been tried, and so have quintile regressions. While the latter are generally unstable, the former yield comparable results, excluding the outage variables which are statistically significant but economically irrelevant.
[32] It is worth noting that the share of firm owning a generator is 52.08% if we consider only countries for which we have sectoral distribution and 53.04% if we consider the manufacture/service sample, which is almost the same.

are always significant and point estimates for firms without generator are higher than those for firms which own one. Some differences can also be noted in the magnitude of the size variables: expanding the sample to these smaller economies reduces the bonus from being very large and increases the deficit from being small.

The last set of results utilise the TFP measure as in Cui, Lapan and Moschini (2012). Given that the construction of the measure requires information about sectoral distribution, this automatically excludes all the countries just taken into account.

[Table 1.6 about here]

Despite the aforementioned differences in input prices across the different countries in our sample (and very likely the same stands true for production technologies), the coefficients on the number of outages in Table 1.6 remain highly statistically significant, while that on the hours of outages is also significant at 10% for the sub-sample of firms without generator. All other covariates maintain the expected sign and significance.

1.5. Generator Ownership Analysis.

As the previous analysis makes clear, it is of great interest to understand what drives firms' decision to acquire a generator, given that it appears that those who suffer more from a poor infrastructure quality are those without such generators. The second part of the study shall then concentrate on exploring the determinants of generator ownership in the SSA context.

1.5.1 Methodology.

Two different models will be taken into account, moving from the simpler to the more complex; while their empirical specification is similar, the underlying economic logics are not the same, so that it is interesting to see if both find confirmation in the data.

Firstly, we shall replicate Foster and Steinbucks (2010), FS from now on, to see if we obtain different results via an expansion of the sample with newer data[33]. They re-adapt the model developed by Reinikka and Svensson (2002) to analyse firms' behaviour in situations of inadequate public service provision. In this setting, a risk neutral firm makes a capital investment in period 1 which return in period 2 depends on the stock of complementary capital available, which might be public or private in nature. The availability of public complementary capital is uncertain in period 2 but is observed in period 1, in which the firm might also decide to invest in a private substitute, incurring in a fixed cost. If it does so, it will ensure a given positive return from the investment in the second period, while if it does not it will face the uncertainty: the return will be the same as if the private substitute was acquired if public capital is available in sufficient quantity, lower if the required amount is not available. The decision to invest in substitute capital depends then on the firm characteristics (how relevant is complementary capital) and on what it observes in the first period. The authors consequentially adopt the following specification:

$$Y_i = N\ (X_i, Z_i)$$

[33] Foster and Steinbucks also worked on a SSA sample but used older rounds of the WBES.

where Y_i represents a firm owning a generator with probability one; N stands for the standardized normal distribution; X_i is one of the two outages measure and Z_i is a vector of controls including firm size, exporter status, firm age, firm sector and country dummies. The model is estimated as a probit function with robust standard error, and the marginal effect are also reported.

Secondly, we shall consider a comparable model, which was developed by Alby, Dethier and Straub (2011), ADS from now on, based on Holmstorm and Tirole (1997). As the previous one, it is aimed to explore the nexus between generator ownership, incidence of the outages phenomenon and firms' size distribution across different sectors.

The theoretical model developed is one of continuous moral hazard, in which an entrepreneur wanting to undertake an investment more costly than her initial endowment has to borrow the difference between the two. The net return on the investment depends on the quality of the electricity supply, measured by the frequency of outages; furthermore, how much outages affect the investment depends on how reliant on electricity is the sector in which the entrepreneur operates[34]. The probability of a successful investment increases monotonically with the amount of effort exerted by the entrepreneur, which cannot be verified by the lender, who only observes the final outcome. The viability of the project depends on its net present value per unit of investment, assumed to be always negative when no effort is exerted but possibly positive when some is; given that the quality of electricity of supply affects the returns on the investment, there is a threshold level of outages above which the project is never viable. The model is set so that for sectors more reliant on electricity the threshold level is lower than for sectors less

[34] Only two types of sectors are assumed, a strongly reliant one and a weakly reliant one, with the effect of outages always more relevant for the former.

reliant, i.e. the same number of outages leads to a higher failure rate.

The maximum amount pledgeable to the lender depends on the possible return on the project, on the benefit incurred by the borrower when no effort is exerted and how much the effort exerted influences the probability of success. Furthermore, the lender must at least break even and, as the credit market is assumed to be perfectly competitive, the borrower appropriates all surplus and invests the highest possible amount. Finally, there is always an alternative investment – a generator – that can be performed at any point. While investing in a generator reduces available assets for other investments, it also insures the entrepreneur against electricity disservice, securing a minimum return when effort is exerted[35].

This setting gives rise to two different scenarios for investment in back-up generation. As long as the shortcomings of the energy infrastructure are not excessive (i.e. the net present value per unit of investment is positive for some level of effort), only big firms (that is, with high initial endowment) autonomously invest in generator – obtaining easier access to credit – while smaller firms still obtain credit and find profitable to produce without acquiring back up capacity. Increases in the frequency outages, as long as their level remains below the threshold for which no investment is ever viable, lead to increases in generator ownership as more firms will find profitable to invest in generators. The alternative scenario is the one determined by very poor electricity infrastructure (i.e. the net present value per unit of investment is negative regardless of the level of effort without generators): only entrepreneurs with enough initial assets to invest autonomously in a back-up capacity can access the credit market and continue

[35] The model is set in such a way that the return in case of generator ownership is always lower than that which could be obtained with an infrastructure of good quality, so to reflect the higher cost of in-house electricity generation.

production, while for those not sufficiently endowed from the beginning returns from production are insufficient to obtain credit.

Two main hypotheses stem from the model. The first is that generator ownership is related to the outage phenomenon, but only if the electricity infrastructure is sufficiently developed (i.e. the net present value per unit of investment is positive for some level of effort), otherwise electricity disservices are too relevant to lead to further investment in generator when not autonomously available. In the latter case only firms with sufficient initial endowments to acquire back-up capacity from the beginning access the credit market, so that increases in outages are not connected with increases in generator ownership. To test the hypothesis, the authors employ the following specification:

$$Y_i = \theta_j + \theta_c + \alpha \log(N_i) + \gamma X_i + \varepsilon_i$$

where Y_i represents a dummy variable equal to 1 if the firm i owns a generator; θ_j is a set of industry dummy; θ_c a set of country dummy; N_i is the number of outages faced by the firm and X_i is a set of control variables. The latter include a dummy equal to one when the firm is in a sector considered sensitive to the quality of the electric service provided[36] and an interaction term between this dummy and the outage measure. The electricity-sensitive dummy has been based on the electricity share of the total cost for the average firm in the sector, and it is equal to one if it is above the mean across sectors. The model is estimated as a probit with standard error clustered at the industry-country level.

[36] According to the model, electricity reliant sectors are more likely to own a generator to start with, so that increases in outages lead to lesser increases in their probability of owning a generator.

The second hypothesis follows from the first and it is related to size distribution across sectors. In an electricity-sensitive sector, where return on investments is critically linked to the quality of the electricity supply, a higher incidence of outages will lead to fewer small firms, as the lack of initial endowments to invest in back up capacity leads to low profitability, eventually pushing them out of the market. Accordingly, in the remaining medium and large firms the proportion of generator ownership should be higher than in sectors which are not sensitive to electricity. This is the specification used to test the hypothesis:

$$Z_{jc} = \theta_j + \theta_c + \beta O_{jc} + \delta\left(O_{jc} * S_h * C_h\right) + \gamma X_{jc} + \varepsilon_{jc}$$

where Z_{jc} is the share of micro (less or equal to 5 employees), very small (between 5 and 10 employees) or small (between 10 and 20 employees) firms in sector j in country c; θ_j is a set of industry dummy; θ_c a set of country dummy; O_{jc} is the average number/hours of outages in sector j and country c; the interaction term in bracket includes the former, a dummy variable S_h equal to one if the firm is in an electricity-sensitive sector and a dummy variable C_h equal to one if the firm is located in a country with a number/hour of outage above the median across countries; X_{jc} represents the usual vector of control variables grouped at the sector-country level. The model is estimated as a tobit with standard error clustered at the industry-country level.

1.5.2 Results.

Table 1.7 presents our results for the FS model. They appear extremely similar to the originals, showing that little has changed in the industrial structure of SSA over the last 5 years. The main differences between their results and ours is the relevance of the interaction term relative to small firms – insignificant in our sample while significant in theirs; also our magnitude of the exporter coefficient is a third of theirs. Overall though, the picture is very similar: while outages are definitely one of the determinants of generator ownership, looking at the elasticities they do not appear to be the only nor the main, although they are the second. Big firms and exporters are those more likely to own a generator, probably reflecting easier access to credit or, as already pointed out by FS, necessity to follow more stringent regulation so to enter export markets. Interestingly, it appears that increases in outages for big and very big firms reduce their likelihood to own a generator, regardless of the measure used, fact not easily explainable in FS framework.

[Table 1.7 about here]

The main limitation of the FS approach, as recognized also by the authors, is that their analysis is not able to properly explore the effect of access to finance. Without a theory which explicitly takes this factor into account it becomes hard to disentangle the link: do firms which own a generator have a higher probability of obtaining a line of credit because of the major productivity which derives from it? Or does the connection run in some other direction?

This issue is openly addressed by ADS, who model exactly this type of relation. The first step involved is that of dividing the sample between firms in electricity-sensitive sector (defined as sectors in which the share of electricity cost is above the inter-sectoral median) and those below[37]. Both Table 1.8 and Table 1.09 report the summary statistics of the two sub-samples

[37] Sector reliance on electricity in the original paper was calculated considering the cost structure of sectors in more advanced economies, while we calculate it over the whole sample. The decision is motivated in our eyes by

(overall and across size distribution), the latter when sector reliance on electricity as an input has been compared only to that on labour, the former when fuel and raw material costs are also taken into account.

[Table 1.8 and 1.09 about here]

The immediate difference that can be noted comparing the two tables is in the distribution of firms, skewed towards the non-electricity sensitive sectors in Table 1.8 and much more even when only labour and electricity costs are considered, suggesting complementarity between the two inputs in our sample. The measure used also influences, although marginally, the shares of generator ownership and of average electricity cost, the former almost always higher for non-electricity-intensive sectors, the reverse for the latter. Incidence of outages also seems to be consistently higher for non-electricity intensive firms, suggesting that there might be more interest in initial plant location when electricity is a more relevant input.

The first of the two tables is the one presenting the nearest picture to that of ADS, but it is hard to determine the reasons for it. First, we ignore which costs they have considered in calculating electricity reliance, so that differences might arise due to different measurement. Secondly, size distribution in our sample is very different from the one used in their study: in their case, medium and large sized firms amount respectively to two-third and a half of the small sized ones; in our case, they amount respectively to a fourth and a tenth. Especially this second point highlights the relevance of small sized firms for the contexts of many SSA economies. Let us

two main factors: first, ADS criteria for defining more advanced economies yield much less reasonable results in our context; second, while technological differences do certainly exist amongst countries in our sample, they are likely to be much less relevant than in that of ADS which was covering the whole world.

now focus on the estimation results, where the first two columns use the sectoral division which considers all production costs but capital depreciation and the last two the one relying only on labour and electricity costs; furthermore, columns 2 and 4 include an interaction term between electricity intense sector and number of outages.

[Table 1.10 about here]

The first thing that can be noticed is that the coefficient on the number of outages is remarkably similar to the one obtained through the application of the FS model, suggesting that despite ignoring factors which are now taken into account the main effect was already consistently estimated. The variable measuring the impact of possible financial constraints is relevant and with the expected sign; furthermore it strengthens the effect of having access to credit. The coefficient on electricity-intensive sectors is negative and significant when electricity reliance is calculated relying only on labour and electricity cost and insignificant otherwise, implying that, contrary from what expected from the model, firms in those sectors are less likely to possess a generator to start with (or equally likely, depending on significance). Finally, the interaction term is never significant, representing another main difference in our results. In ADS these were negative and significant, as a consequence of the already higher likelihood of owning a generator for firms in electricity-reliant sectors. Given that this is not the case in our sample, the latter finding should not be a surprise. The first hypothesis of the model does not then seem to be confirmed in the SSA context, and we control further for this as they did, i.e. dividing the sample between the sectors which are electricity reliant and those which are not.

[Table 1.11 about here]

As expected, while there are differences across covariates signs and relevance, the coefficient on the outage variable is equally relevant in the two sub-samples and far from being 36% lower for the electricity-reliant sector as found by ADS, on the contrary it is slightly higher for one of the two measure (and identical for the other). The differences probably stem from the definition of an electricity-reliant sector and from the average level of outage: as previously noted, a definition based on electricity share will always depend on the proportion and cost of electricity self-generated[38].

We shall now concentrate on the last part of the analysis, namely the one which tests the second prediction of the model, which is whether the quality of energy infrastructure has an influence on firm size distribution across sectors. We shall follow ADS exposition and start introducing some summary statistics on firm size distribution, comparing sectors more and less reliant on electricity and countries more and less hit by the outages phenomenon. As earlier, Tables 1.12 and 1.13 present an overview of the sample following both sectoral division – ignoring and considering raw material and fuel costs.

[Table 1.12 and 1.13 about here]

The picture emerging from the analysis of the summary statistics is not as clear as in the ADS sample, although there are some similarities. In their sample, ADS found that countries more hardly hit by electricity disservice had a lower share of small firms in both electricity reliant and non-electricity reliant sectors. However, because of sector specific characteristics, sectors more strongly reliant on electricity had a higher share of small firms in both types of countries.

[38] This explanation though goes against the higher relevance of access to credit in sector less reliant on electricity, which already has a higher share of generator ownership.

In our sample, regardless of the measure of electricity reliance employed, countries suffering more from outages have a lower share of small firms in electricity-sensitive sectors, as in ADS, but a higher one in those non-electricity sensitive. Also in contrast with their sample, in our case sectors more sensitive to electricity always have a lower share of small firms than those less sensitive to it.

[Table 1.14 about here]

Table 1.14 presents then the results of the Tobit estimation aimed at checking the hypothesis in a more rigorous way. This time the hypothesis that there is a relation amongst the relevance of outages and sectoral and size distribution finds some support in the data. The coefficients on the outage variable is always positive and significant, pointing towards a negative general effect of outage on firms' ability to grow in size. The interaction term isolating the effect for electricity intensive firms is insignificant for micro firms but negative and significant for small and very small firms (although only for one of the two electricity intensity measures for the latter) as predicted by the theory. This seems to indicate that the quality of the electricity supply does indeed play an effect on firm's survival capacity for particular sectors, but that this becomes relevant only after the firm has already grown above a given threshold (which at 5 employees is though relatively small).

There is one final point which needs to be stressed, as it is probably playing a role in driving the results: the industrial structure of the SSA countries in the study is much less diversified than that in ADS. This can be quickly understood comparing the numbers of observations of the Tobit model in our study and in theirs: 484 for ours; 1,777 for theirs. As these numbers stand for size shares in sector-country (and we uses 2 digit industry code from the same data source), if we divide them by the number of countries we obtain the average number of sectors per country, which can be seen as a crude measure of industrial diversification: it is 19 in our study and 28 in theirs.

1.6. Conclusions.

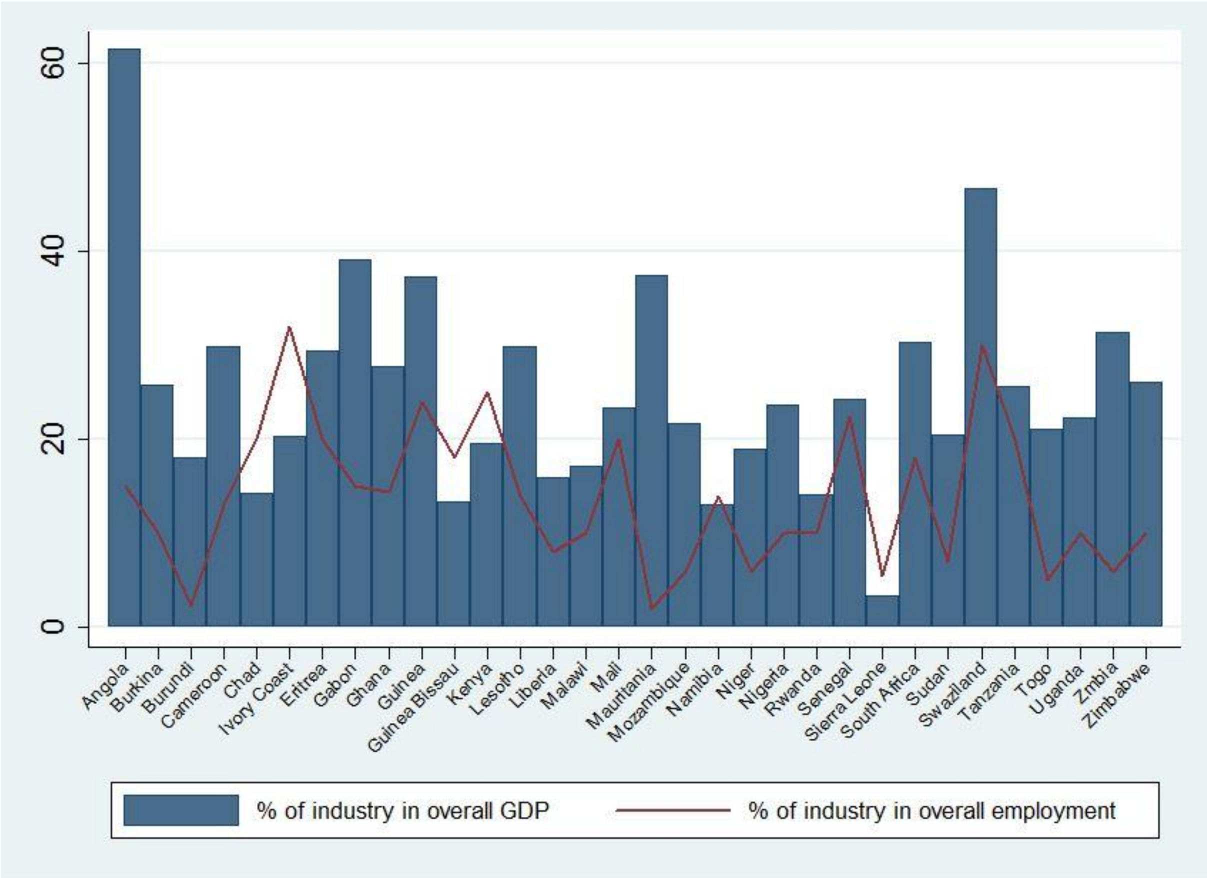
Sub-Saharan African economic development is surely constrained by a number of different factors, and the dismal situation of the power infrastructure of almost every country in the continent is likely to be one of the main elements. The effect of an inadequate generation capacity is clear to see, ranging from excessive reliance on biomass from households, with huge consequent health and environmental costs; to frequent and extensive outages constraining firms. It is especially on the effects on the latter that this study focuses. After reviewing the most recent literature – both on a macro and on a micro level – and presenting extensive summary statistics of the World Bank Enterprise Survey for every country included in the analysis, we use revenue functions to explore how much of a burden outages are for firm profitability. In light of the results obtained we then focus on the analysis of the determinants of generator ownership in the Sub-Saharan context, applying to our data two of the models recently developed in the literature.

Considering the whole sample, the first main conclusion is that power outages are indeed a drag on firm profitability. When we further differentiate analysis between firms with and without a generator, the effects appear much stronger for the latter and just about perceivable for the former. These results are robust to different specifications, although the estimates might suffer from the presence of endogeneity as they have been obtained in an OLS framework in which neither endogeneity nor unobserved heterogeneity can be tackled. Ignoring the many economic changes which will necessarily be entailed in a decisive improvement of energy infrastructure, the reduction of the amount of hours faced by an average SSA firm without a generator to that of its Southern African counterpart will imply a ceteris paribus increase in its sales of 3.2%, which is roughly 700.000 2005 international dollars.

It is also clear from both models of generator ownership considered that the level of outage is an important determinant of firms' decision to invest in back-up generation capacity. While it appears that firms in sectors that rely more on electricity for production are more likely to opt for this type of investment to start with, there is no particular evidence in the data that their decision is differently influenced by changes in electric provision quality, even though this might depend on the way in which electricity reliance has been defined. There also are some signs that the growth paths for firms in these sectors are more heavily influenced by the frequency of outages than that for firm in sectors less reliant on electricity.

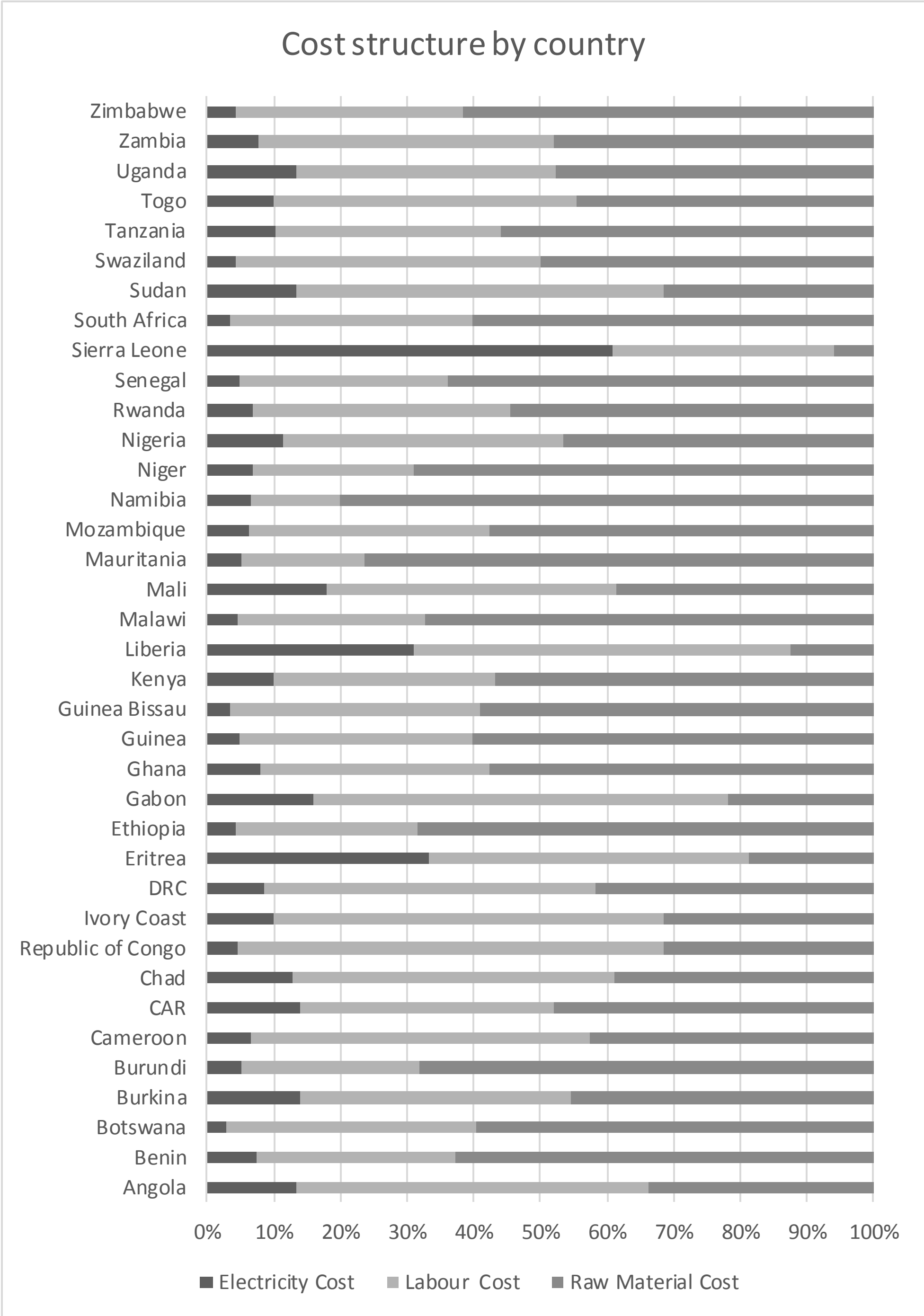
Overall then, the chapter points towards the following policy suggestion: improvements in energy infrastructure will be favourably met by increases in profitability across firms of all sectors and size in SSA, and even more so for firms which do not have access to back-up capacity. While these improvements are not likely to reduce importantly the weight of in-house generation, as there are other relevant factors behind the demand of generators, many are the probable economic progress that will follow. At the same time, it is unlikely that these improvements alone will determine particular changes in the economic structure of many SSA states, both under the point of view of sectoral and size distribution. To meet the challenge of diversifying their economies, those countries will probably need to develop a more complex set of industrial policies, part of which should surely be the easing of credit constraints which appear to be playing an important role in preventing firm expansion.

Figure 1.1 – GDP and employment share of industry.



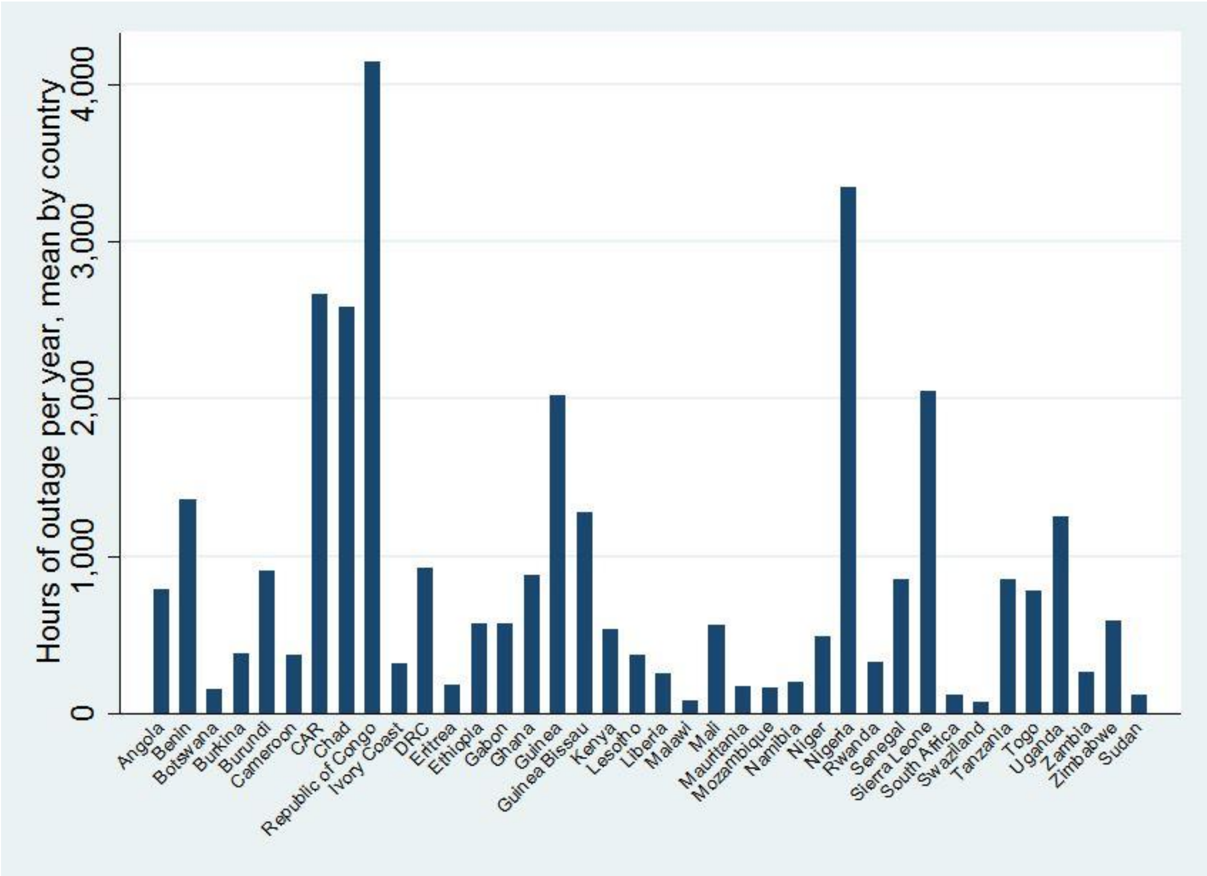
Share of GDP and employment of the industrial sector for selected countries, CIA World Factbook.

Figure 1.2 – Average cost structure across SSA.



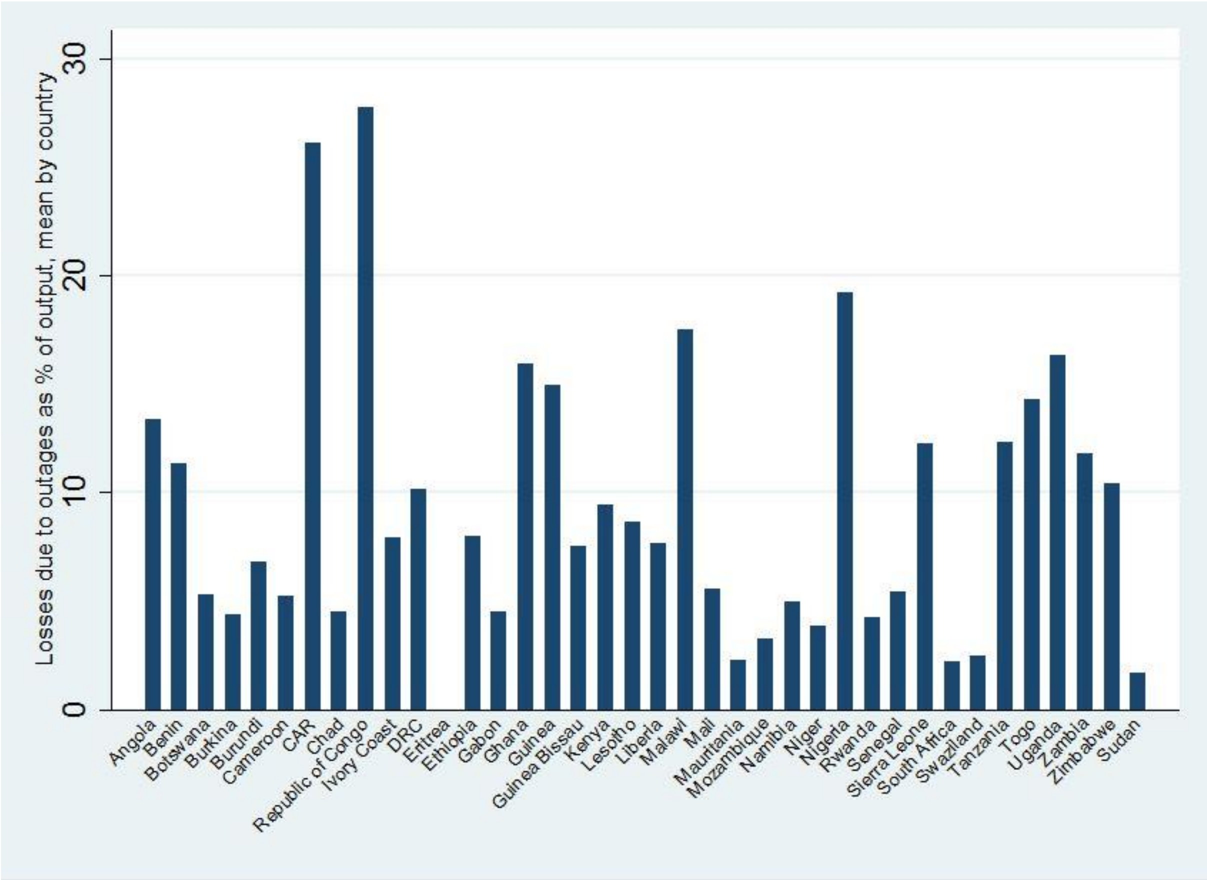
Average firm cost structure across SSA countries, excluding fuel costs and capital depreciation.

Figure 1.3 – Average hours of outage per year.



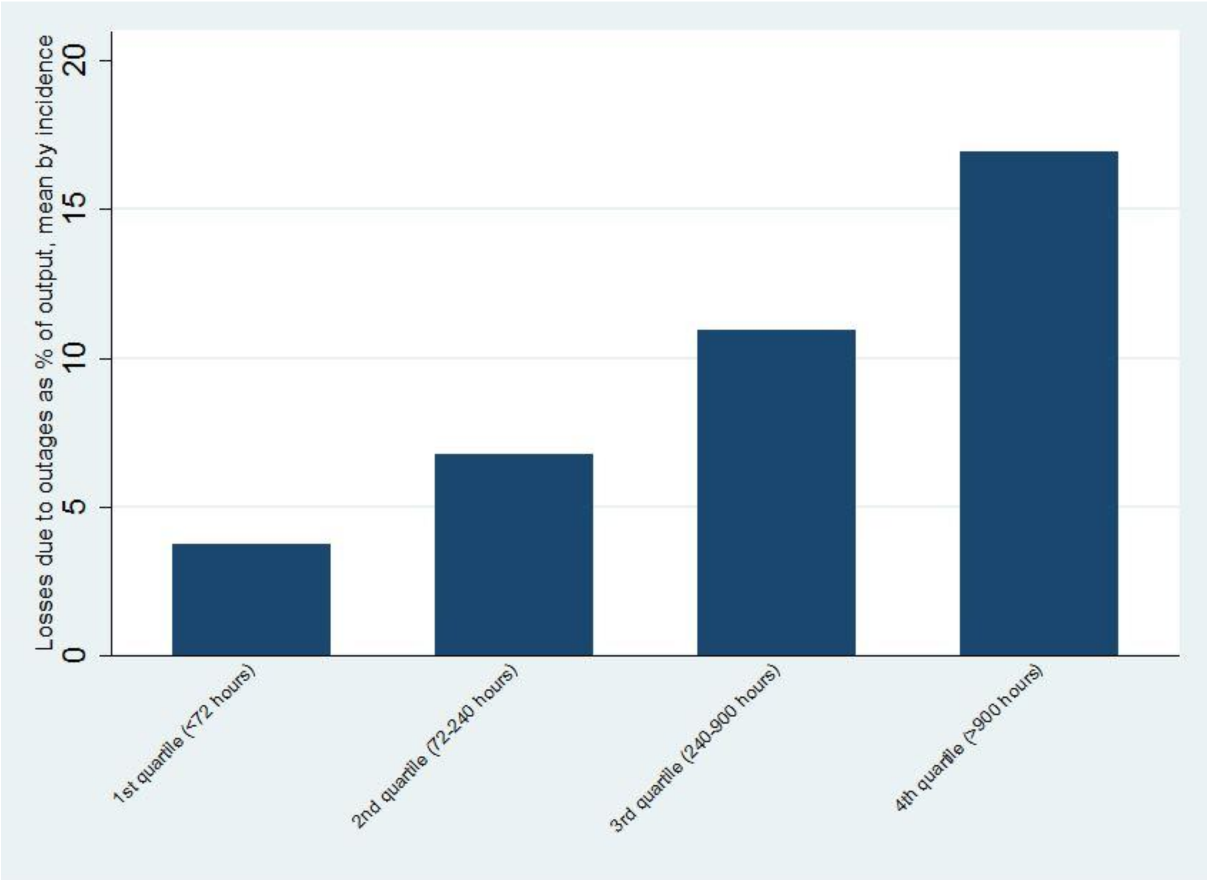
Hours of outage per year, mean by country.

Figure.1.4 – Average losses to power outages as share of output.



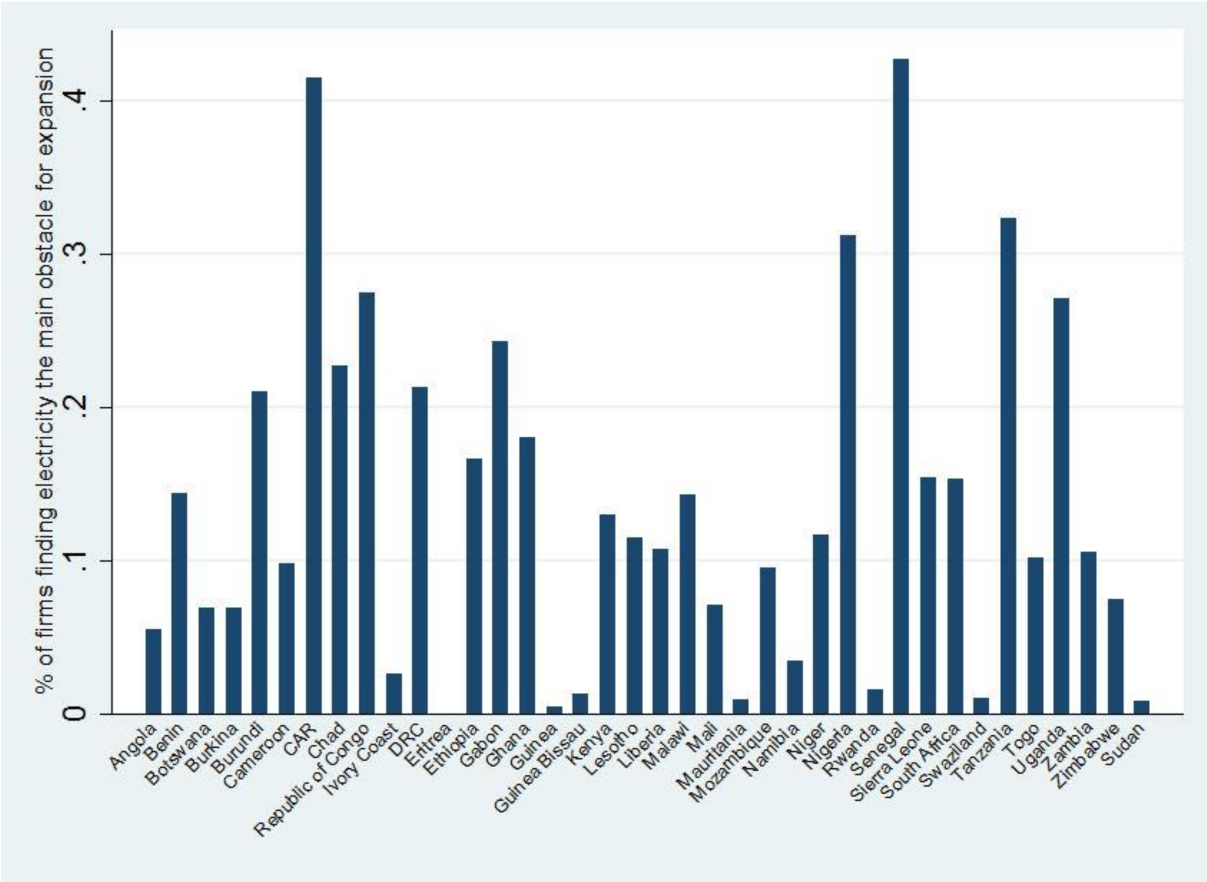
Average loss of output due to outages by country.

Figure 1.5 – Losses to power outages by their incidence.



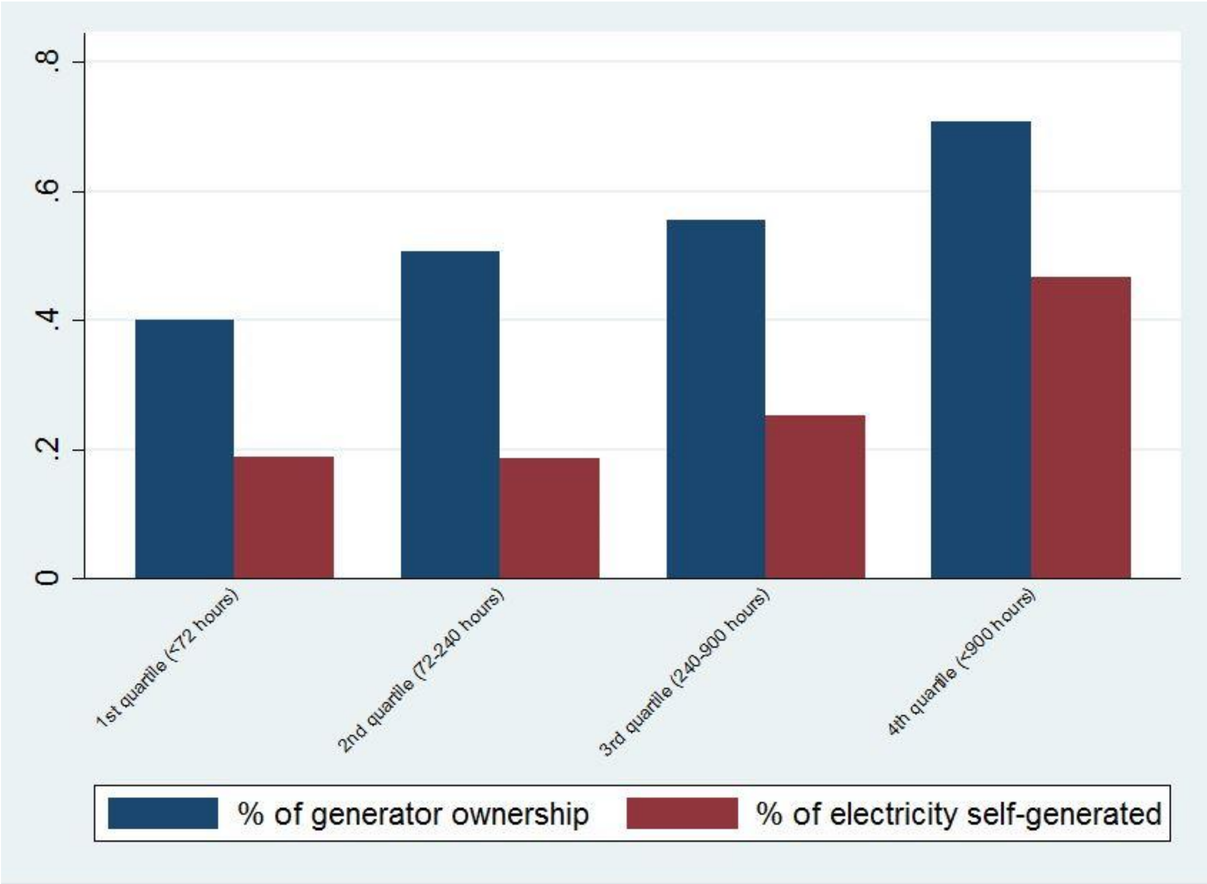
Losses due to outages by outages incidence

Figure 1.6 – Electricity as main obstacle to expansion.



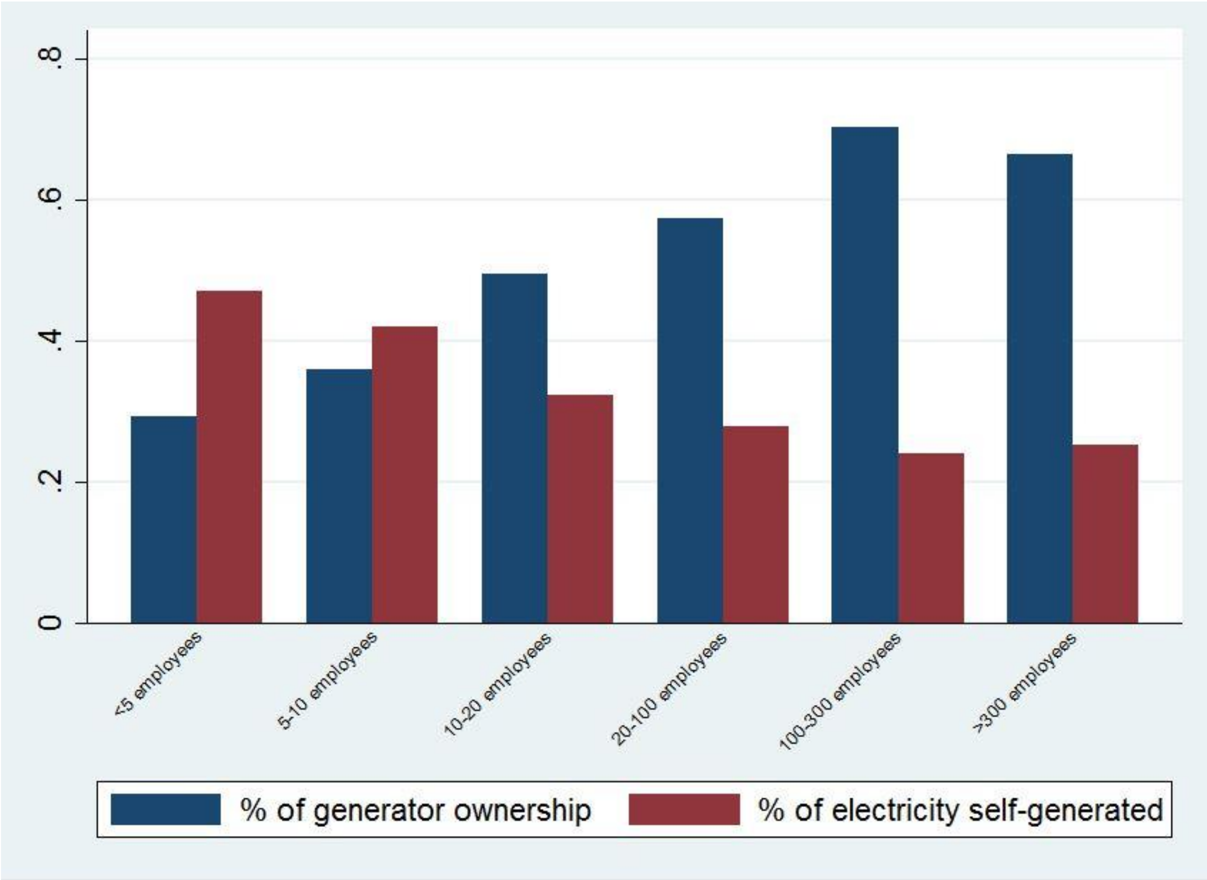
Percentage of firm declaring that electricity is their main obstacle by country.

Figure 1.7 – Generator ownership and self-generation by power outage incidence.



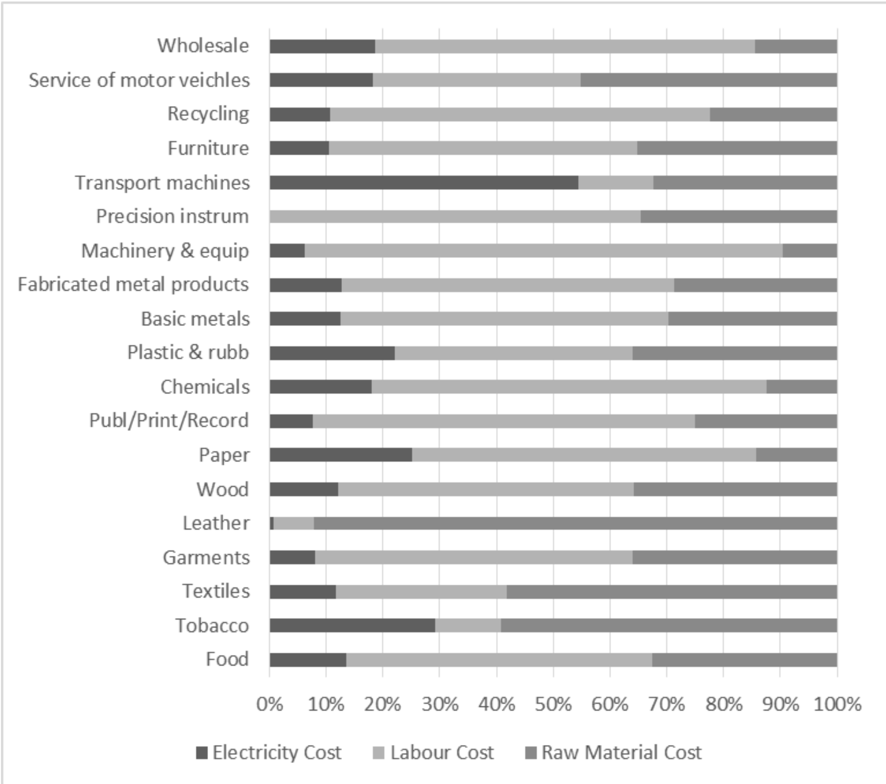
Percentage of generator ownership and of electricity self-generated, by outage incidence.

Figure 1.8 – Generator ownership and self-generation by firm size.



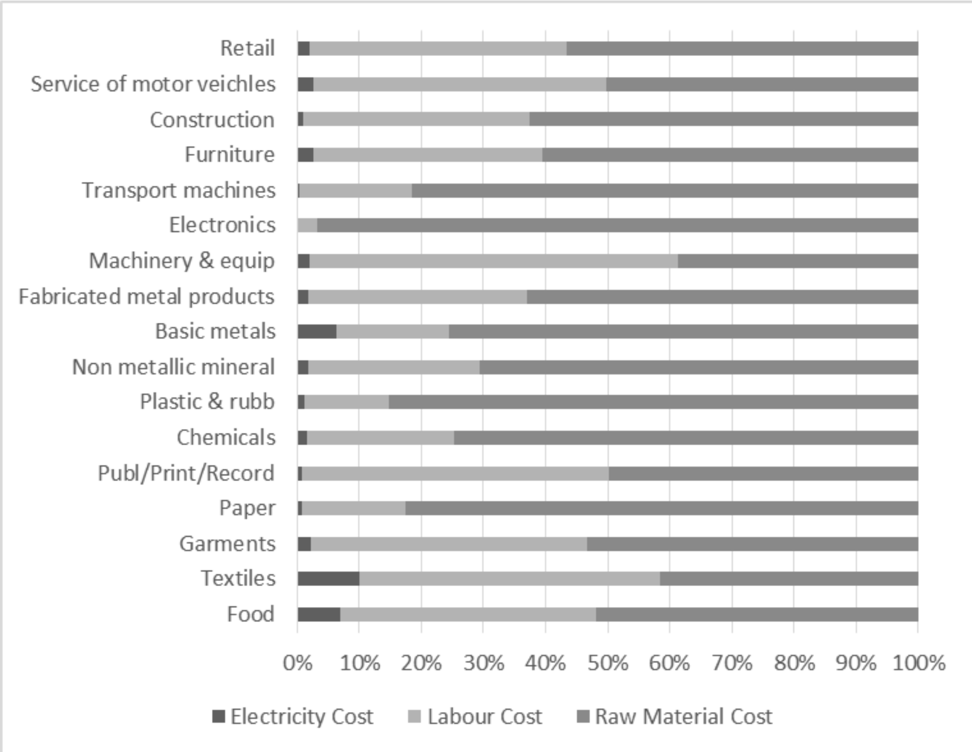
Percentage of generator ownership and electricity self-generated, by firm size.

Figure 1.9 – Average cost structure across industries, Angola.



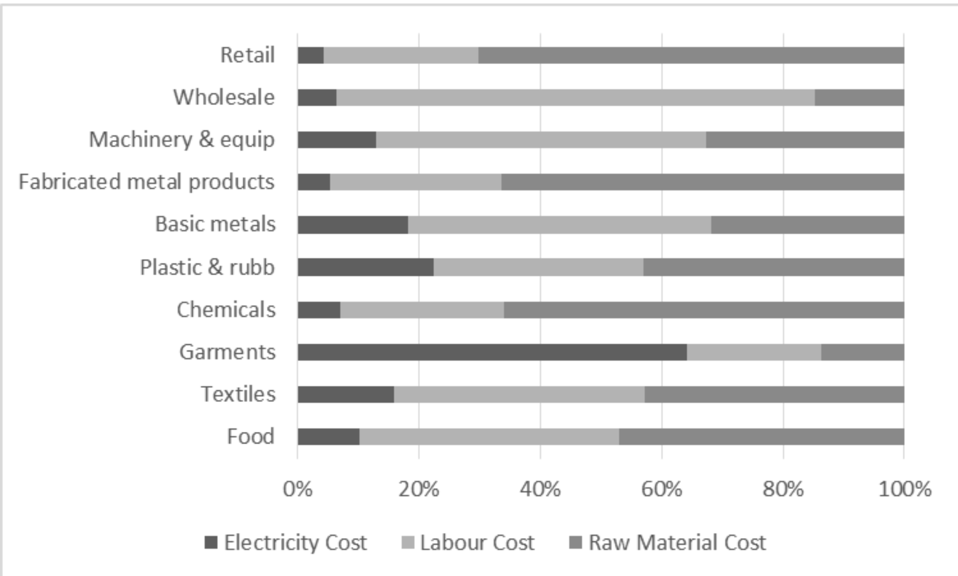
Cost structure of average Angolan firm, by industry.

Figure 1.10 – Average cost structure across industries, Botswana.



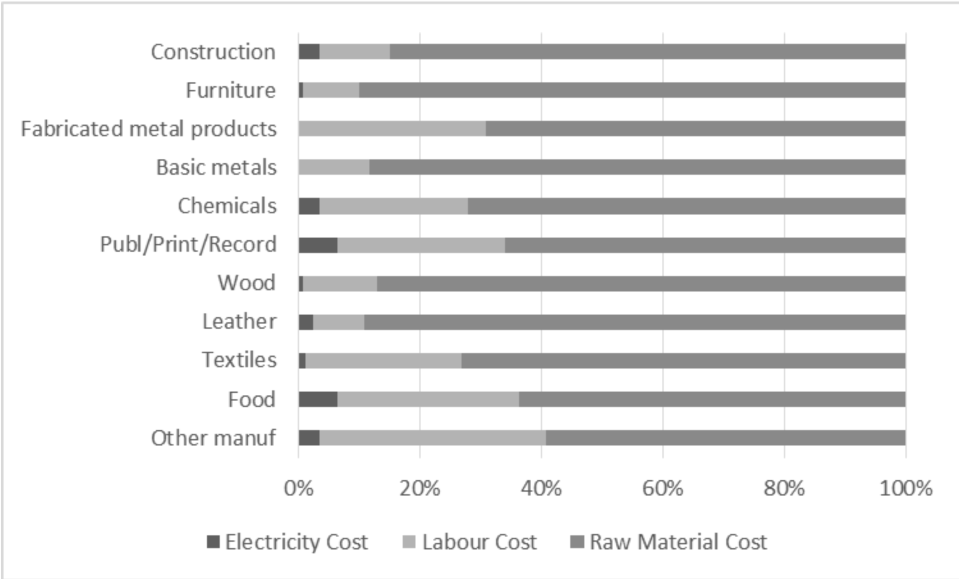
Cost structure of average Botswanan firm, by industry.

Figure 1.11 – Average cost structure across industries, Burkina Faso.



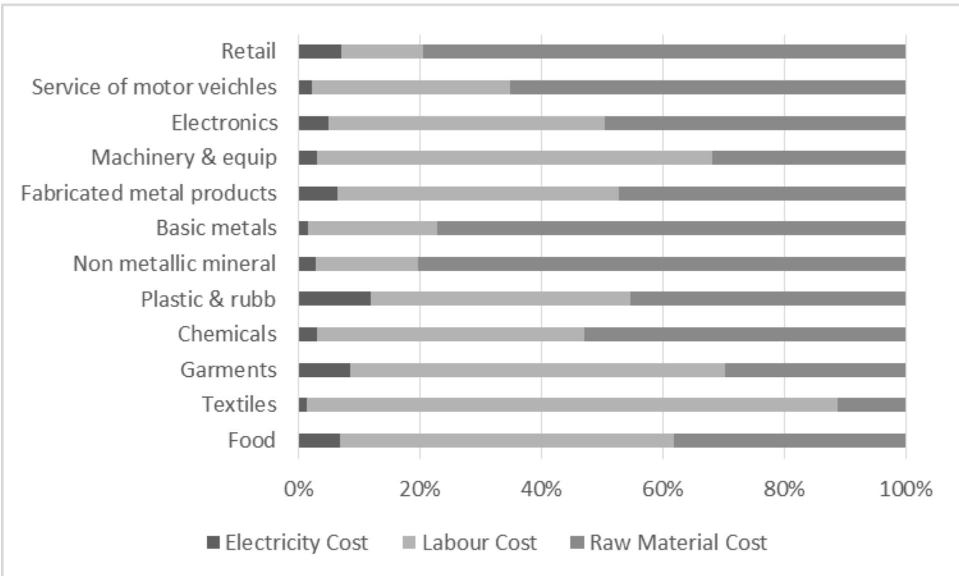
Cost structure of average Burkinan firm, by industry.

Figure 1.12 – Average cost structure across industries, Burundi.



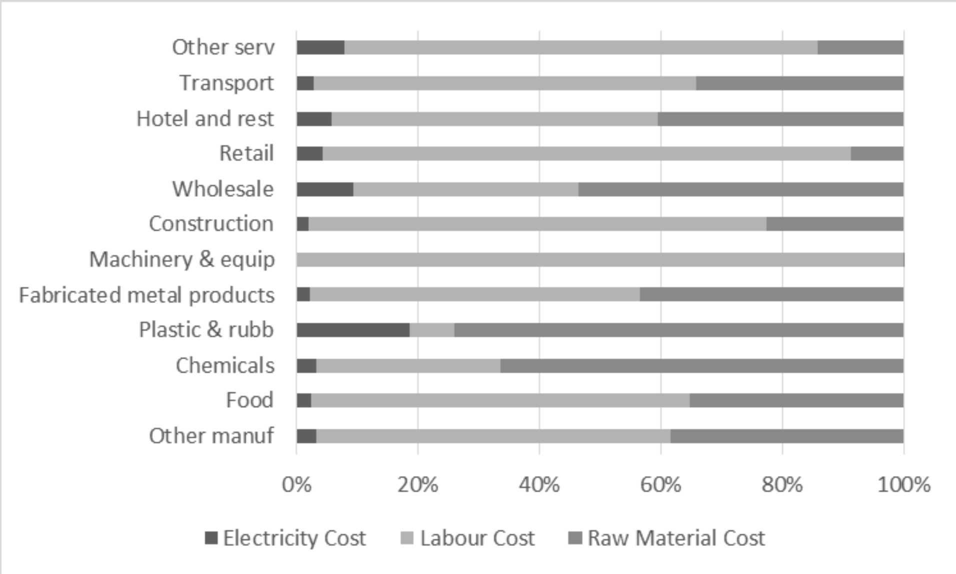
Cost structure of average Burundian firm, by industry.

Figure 1.13 – Average cost structure across industries, Cameroon.



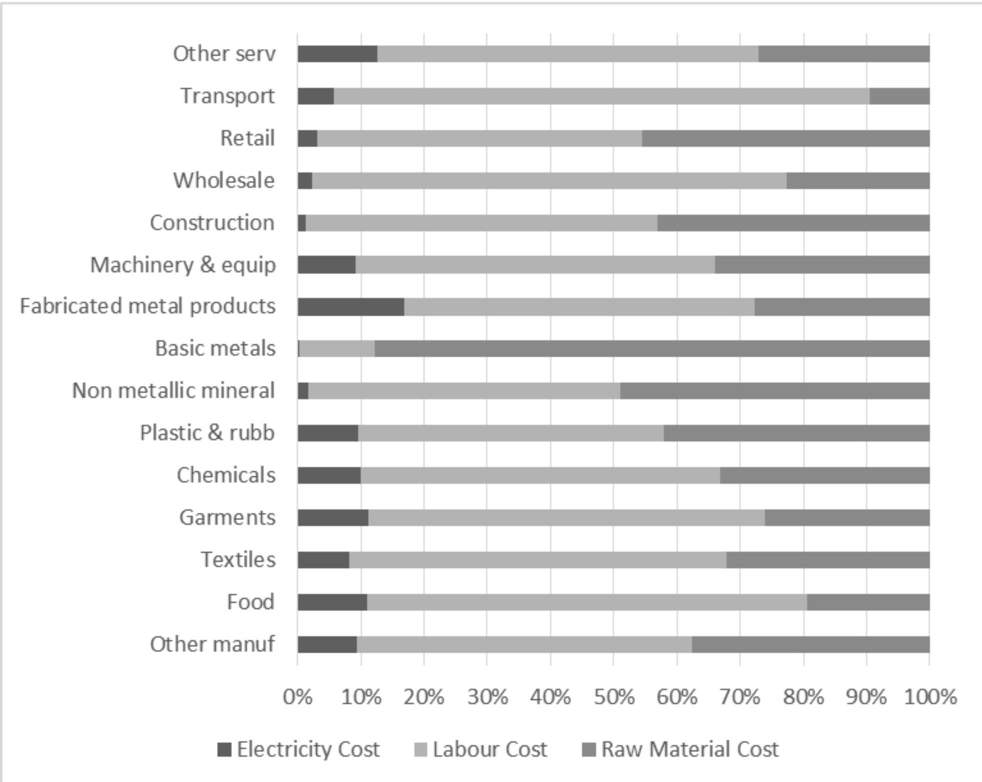
Cost structure of average Cameroonian firm, by industry.

Figure 1.14 – Average cost structure across industries, Republic of Congo.



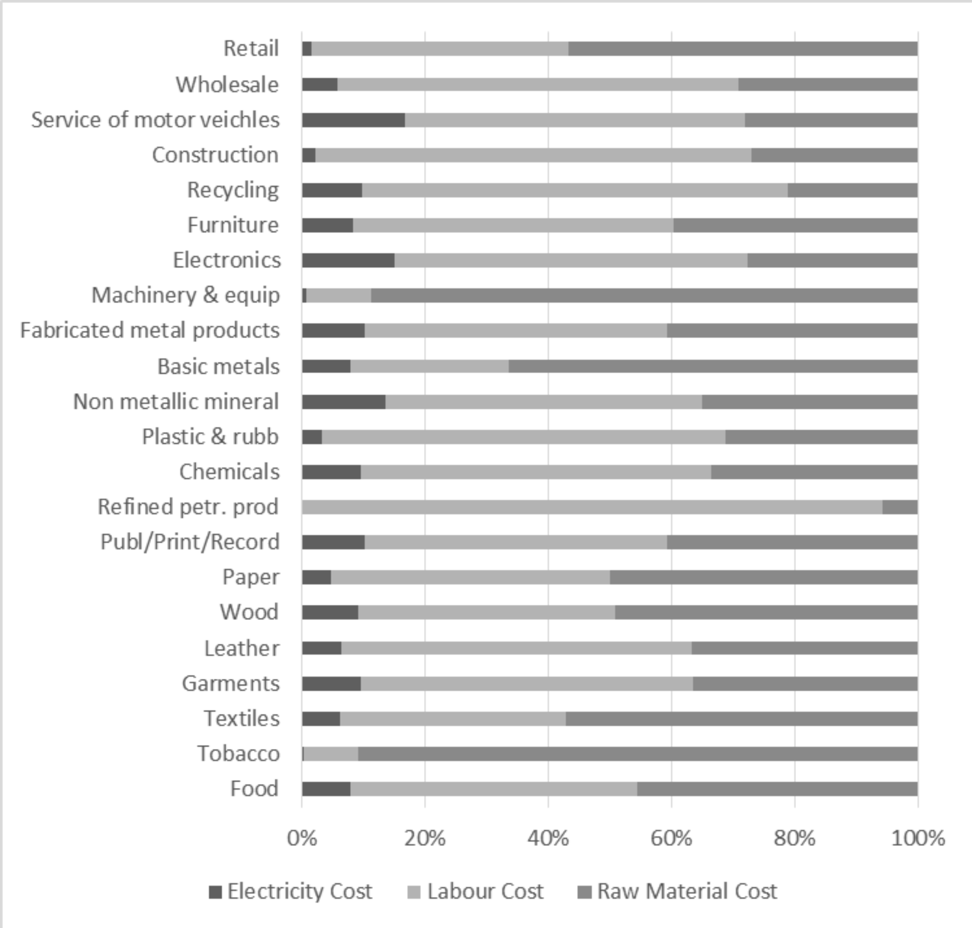
Cost structure of average Congolese (Republic of) firm, by industry.

Figure 1.15 – Average cost structure across industries, Ivory Coast.



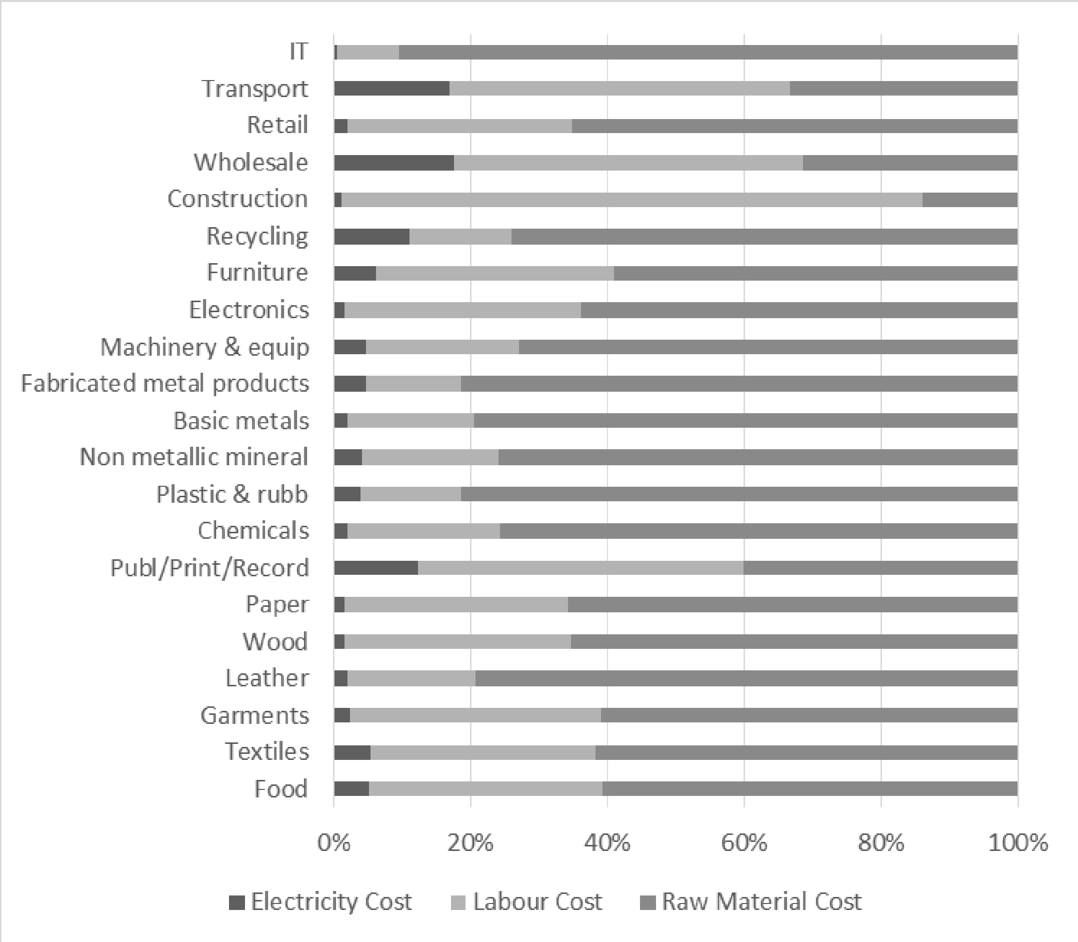
Cost structure of average Ivorian firm, by industry.

Figure 1.16 – Average cost structure across industries, Democratic Republic of Congo.



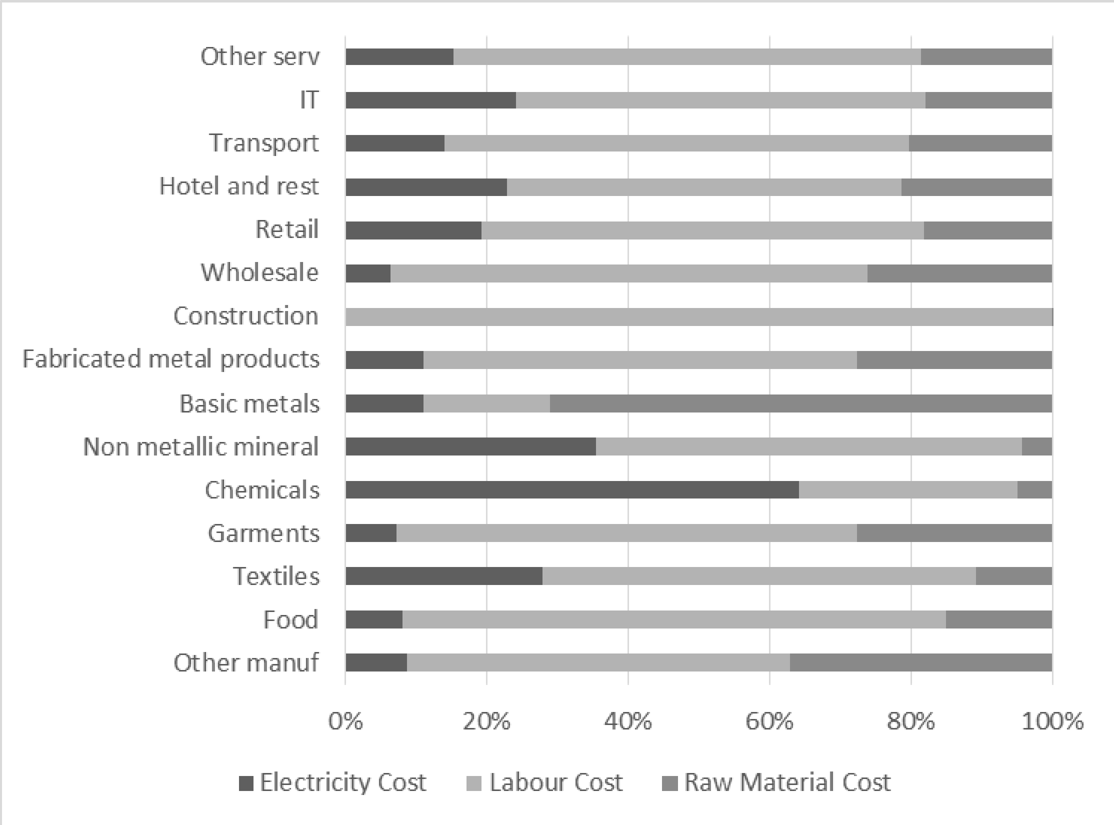
Cost structure of average Congolese (Democratic Republic of) firm, by industry.

Figure 1.17 – Average cost structure across industries, Ethiopia.



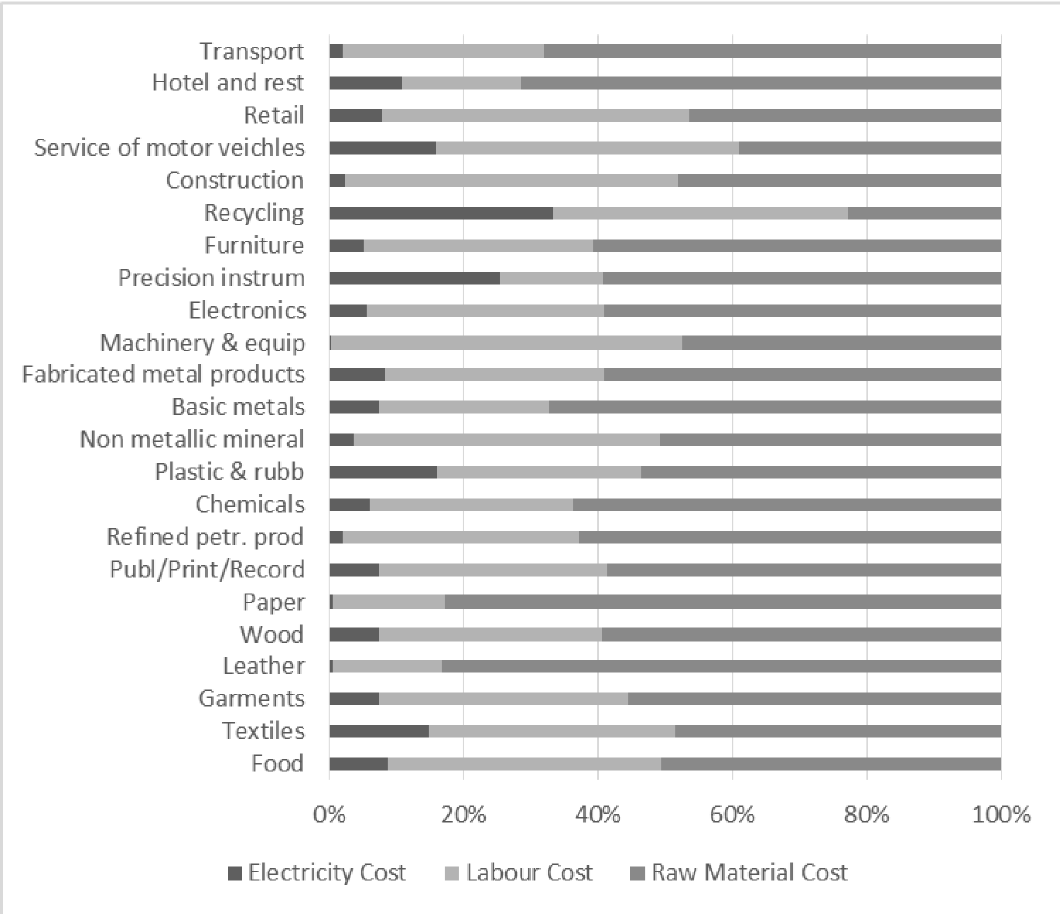
Cost structure of average Ethiopian firm, by industry.

Figure 1.18 – Average cost structure across industries, Gabon.



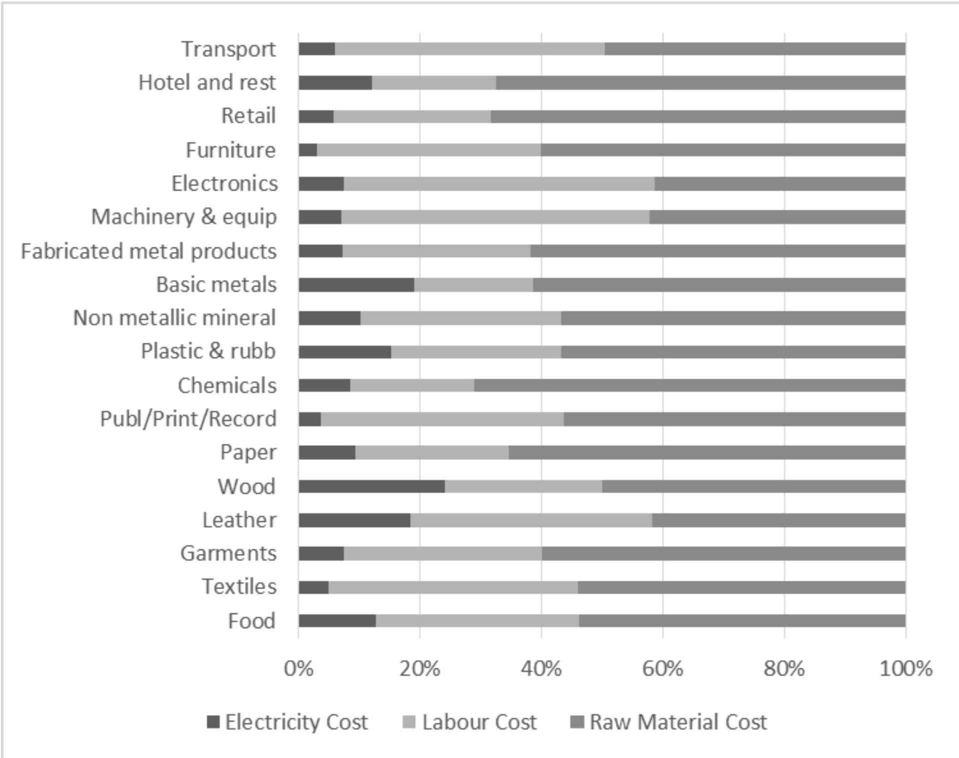
Cost structure of average Gabonese firm, by industry.

Figure 1.19 – Average cost structure across industries, Ghana.



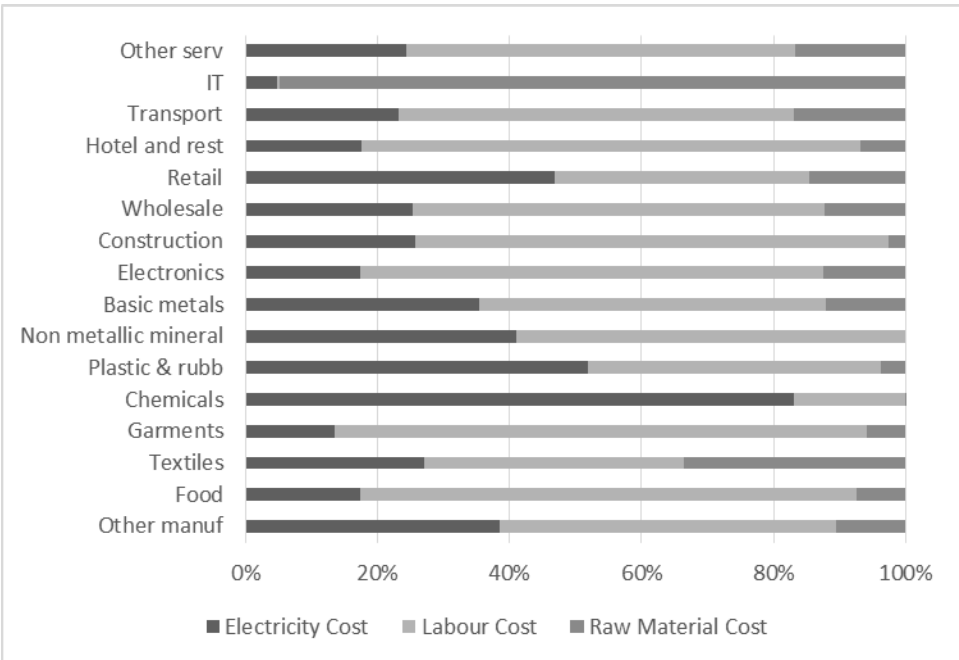
Cost structure of average Ghanaian firm, by industry.

Figure 1.20 – Average cost structure across industries, Kenya.



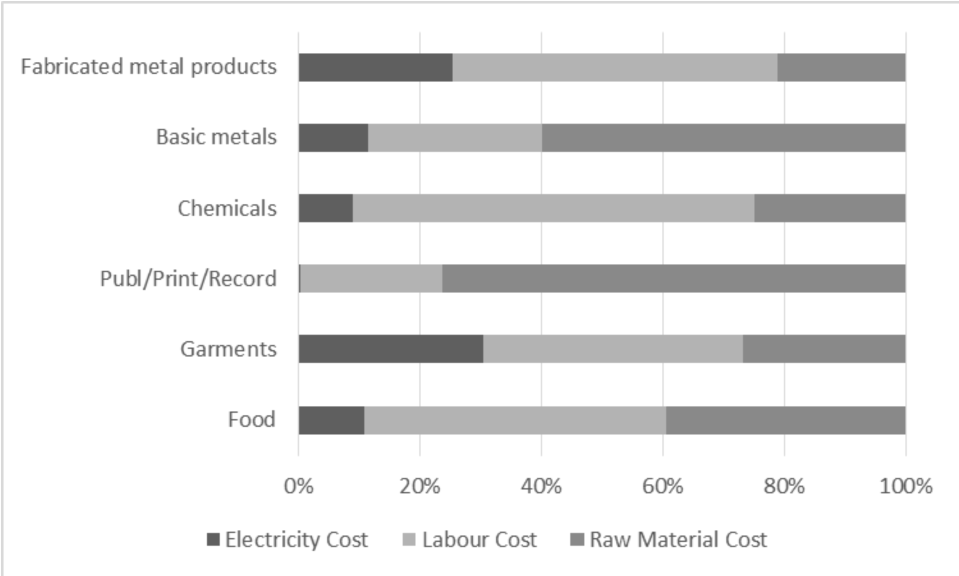
Cost structure of average Kenyan firm, by industry.

Figure 1.21 – Average cost structure across industries, Liberia.



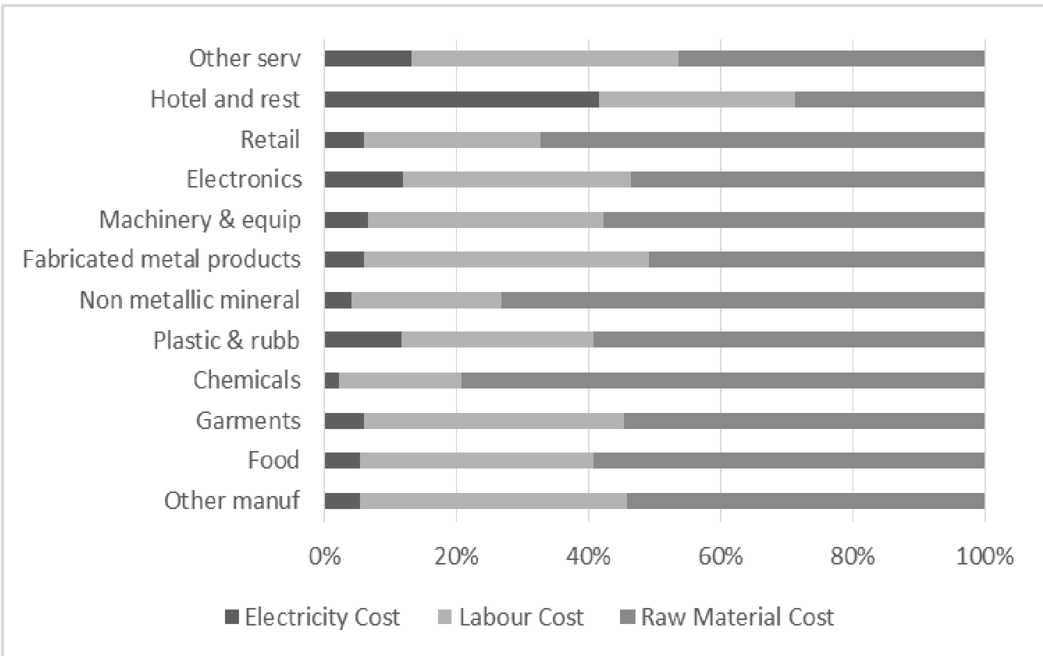
Cost structure of average Liberian firm, by industry.

Figure 1.22 – Average cost structure across industries, Mali.



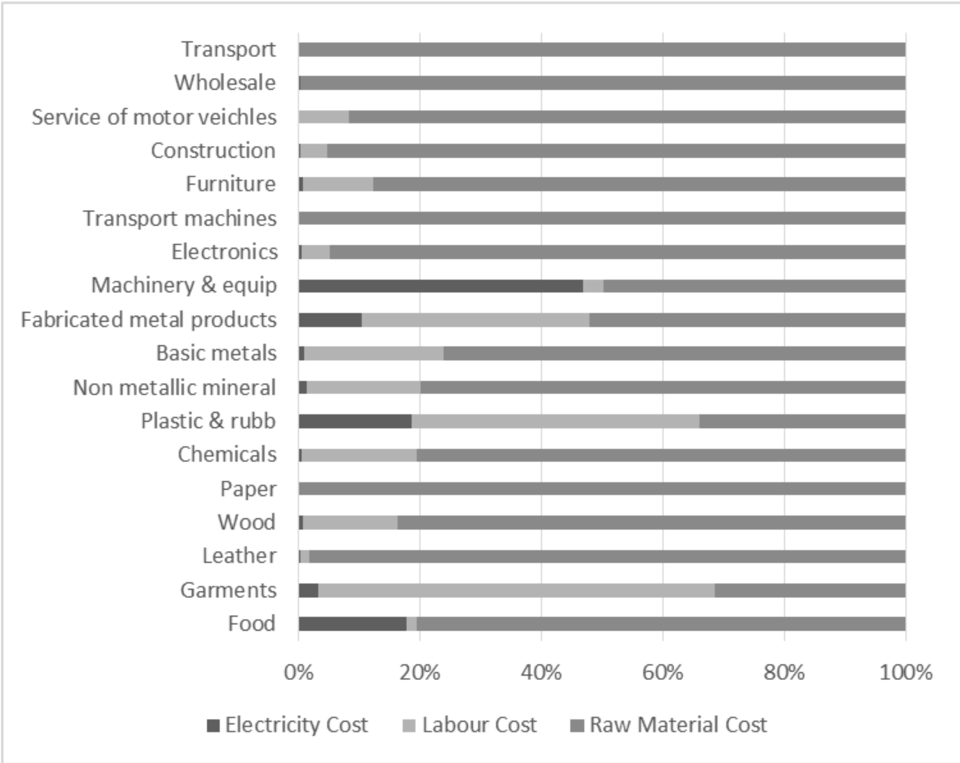
Cost structure of average Malian firm, by industry.

Figure 1.23 – Average cost structure across industries, Mozambique.



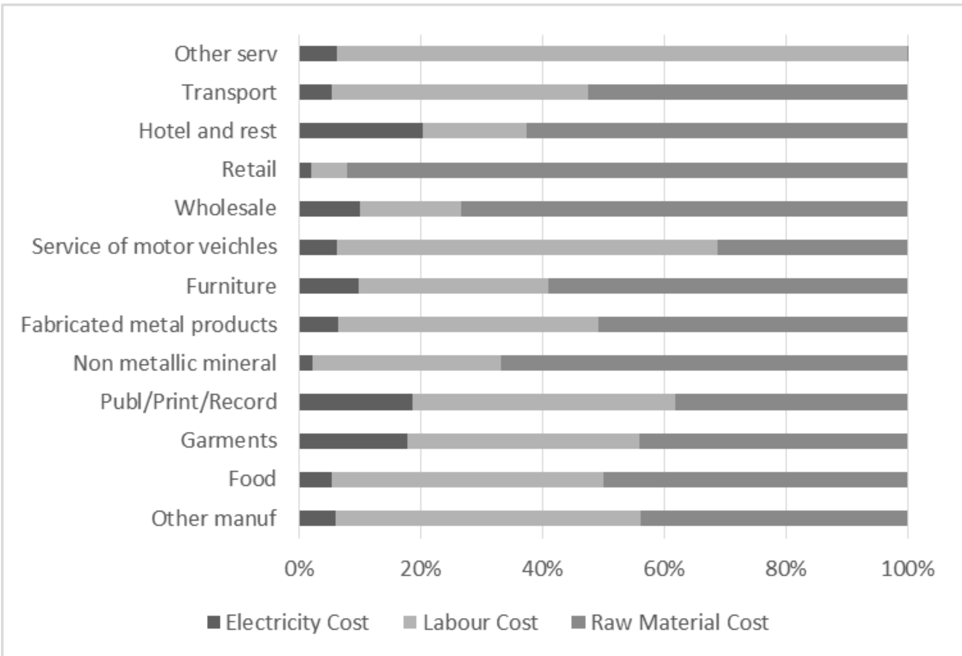
Cost structure of average Mozambican firm, by industry.

Figure 1.24 – Average cost structure across industries, Namibia.



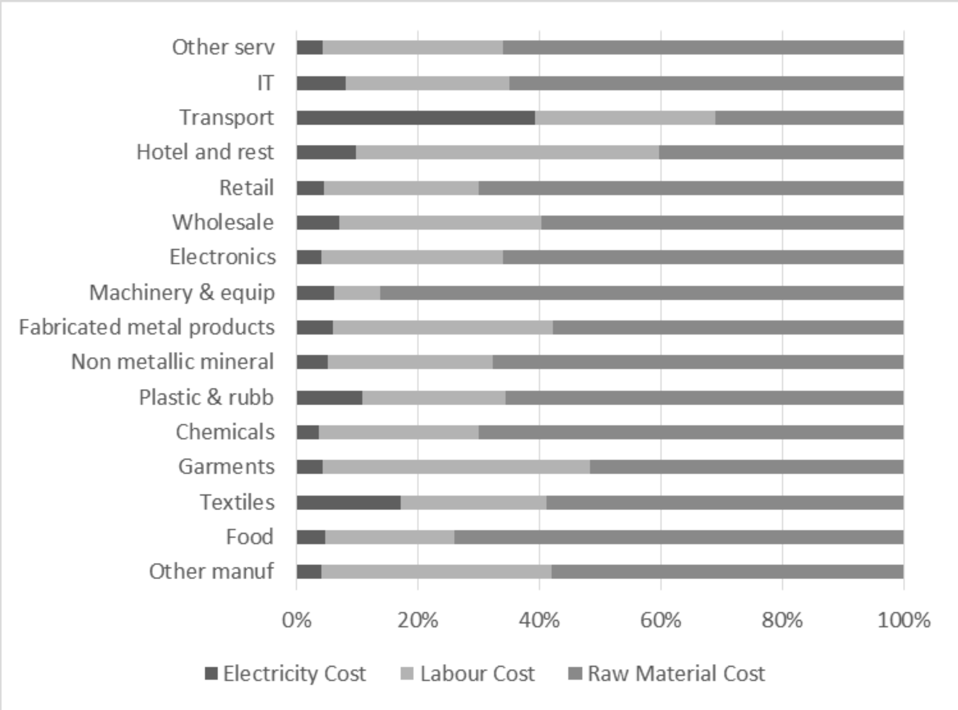
Cost structure of average Namibian firm, by industry.

Figure 1.25 – Average cost structure across industries, Nigeria.



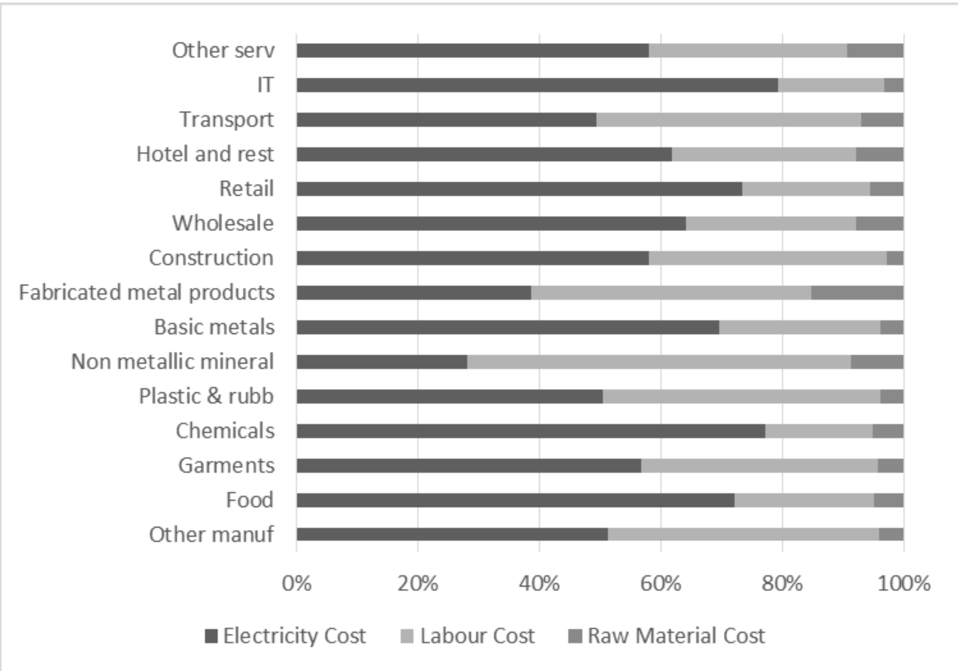
Cost structure of average Nigerian firm, by industry.

Figure 1.26 – Average cost structure across industries, Senegal.



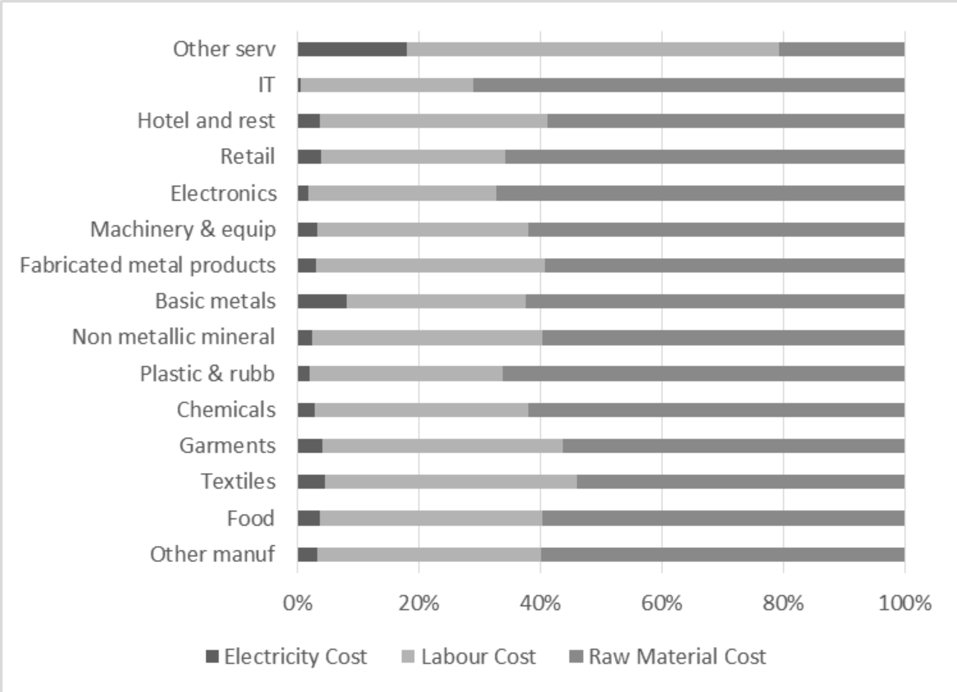
Cost structure of average Senegalese firm, by industry.

Figure 1.27 – Average cost structure across industries, Sierra Leone.



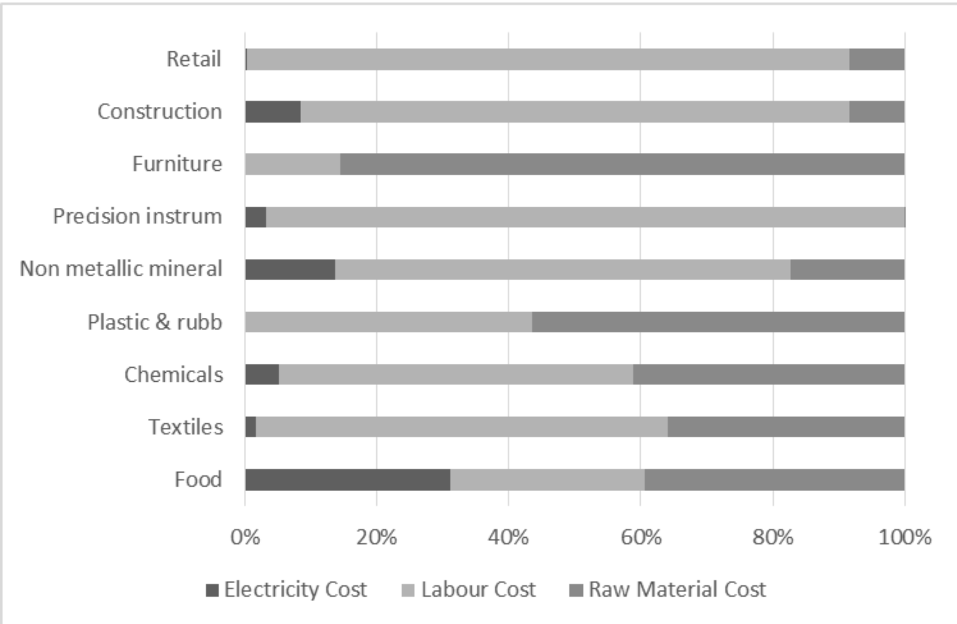
Cost structure of average Sierra Leonese firm, by industry.

Figure 1.28 – Average cost structure across industries, Republic of South Africa.



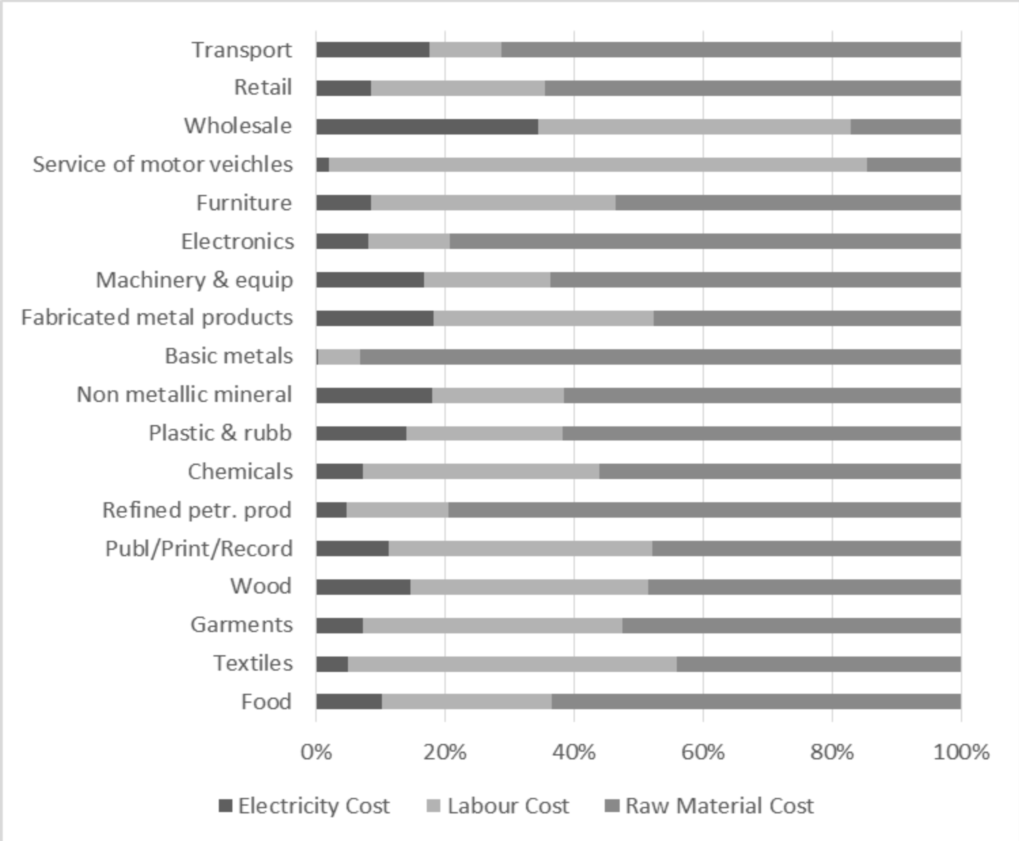
Cost structure of average South African firm, by industry.

Figure 1.29 – Average cost structure across industries, Sudan.



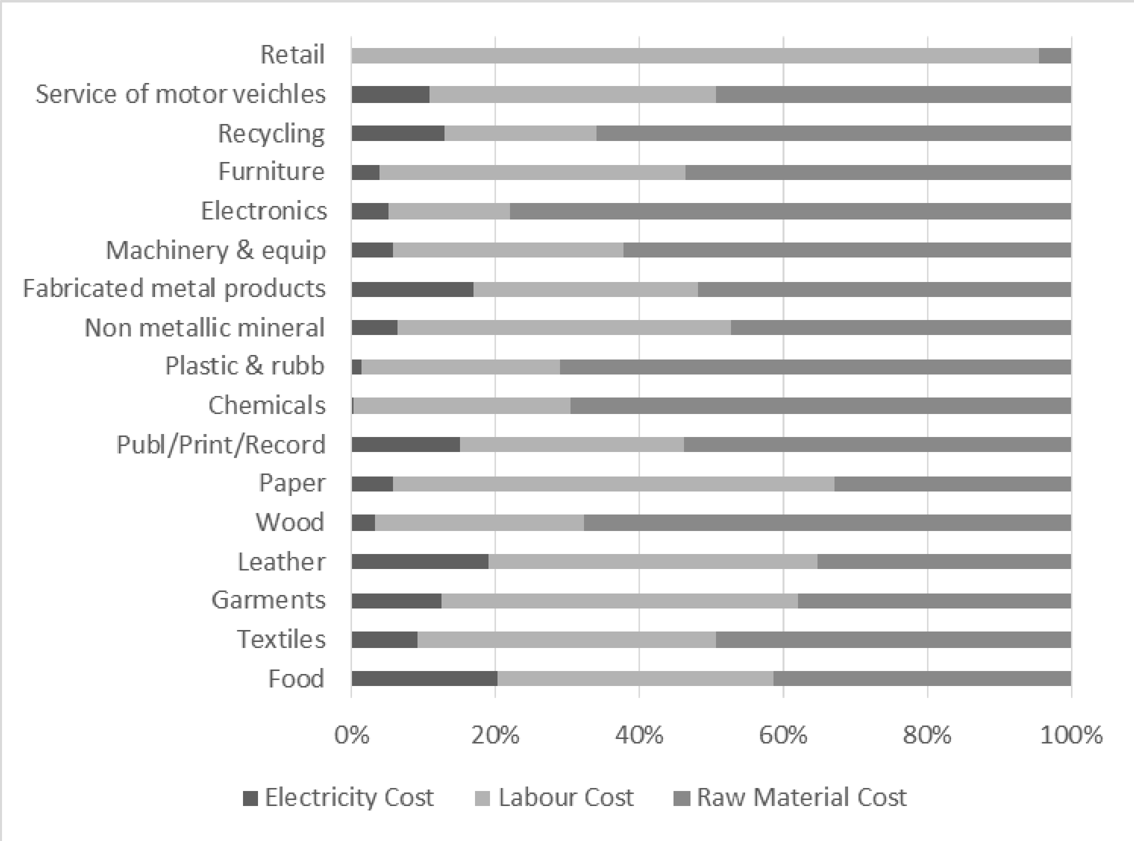
Cost structure of average Sudanese firm, by industry.

Figure 1.30 – Average cost structure across industries, Tanzania.



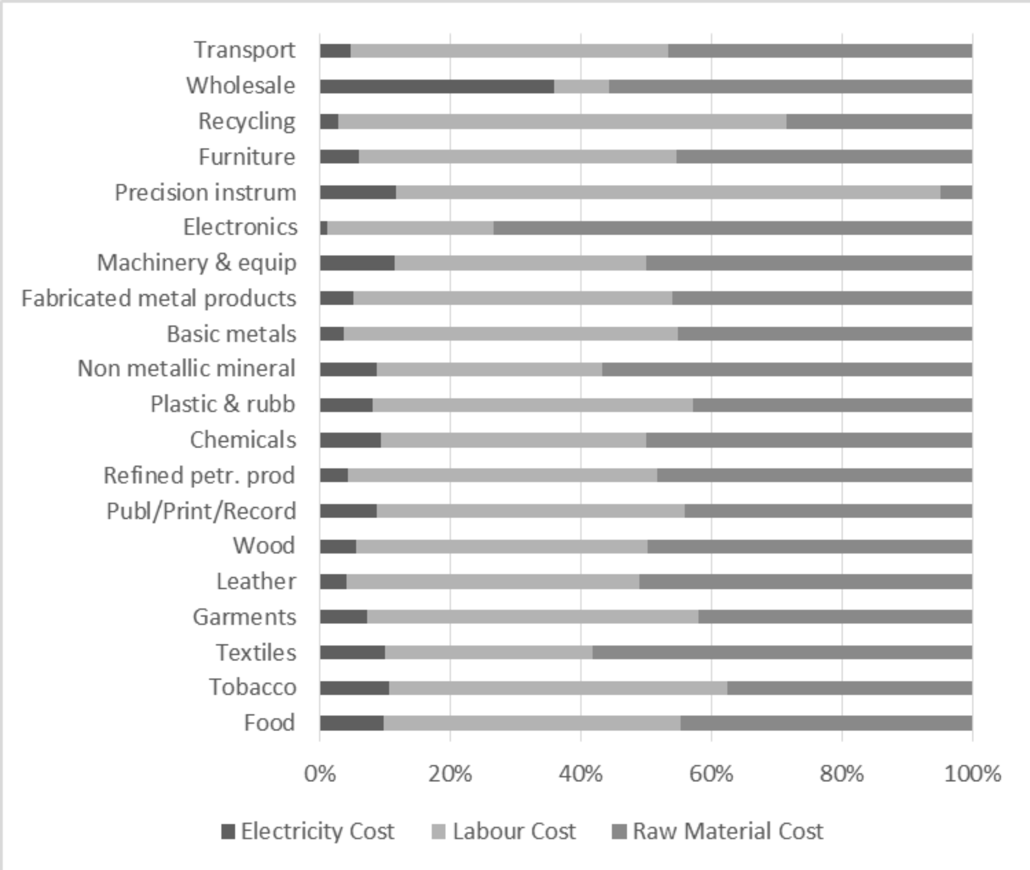
Cost structure of average Tanzanian firm, by industry.

Figure 1.31 – Average cost structure across industries, Uganda.



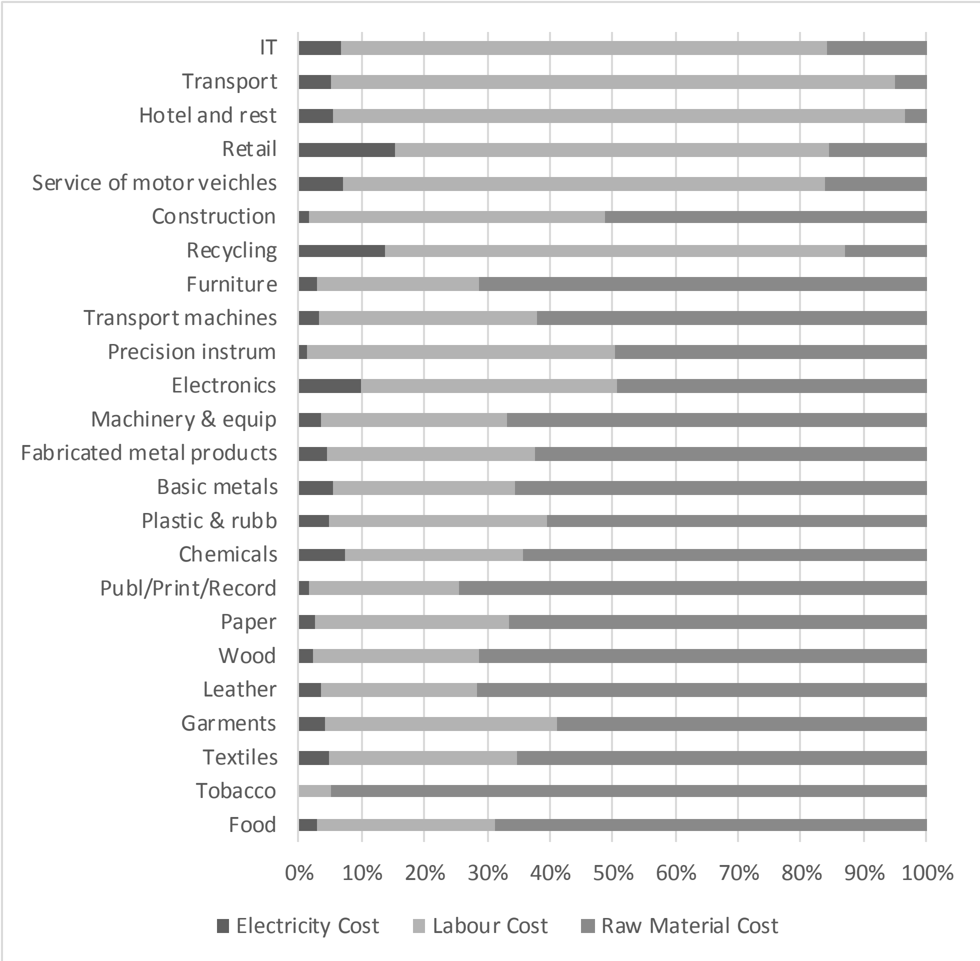
Cost structure of average Ugandan firm, by industry.

Figure 1.32 – Average cost structure across industries, Zambia.



Cost structure of average Zambian firm, by industry.

Figure 1.33 – Average cost structure across industries, Zimbabwe.



Cost structure of average Zimbabwean firm, by industry.

Table 1.1 – Covariates summary statistics.

Variable	Mean	S.D.	min.	MAX.
Number of outage	150.95	259.16	0	3720
Hours of outage	881.05	1626.08	0	8765
Small	0.63	0.48	n.a.	n.a.
Medium	0.26	0.44	n.a.	n.a.
Large	0.07	0.26	n.a.	n.a.
Very large	0.04	0.19	n.a.	n.a.
Exporter	0.1	0.3	n.a.	n.a.
Access to credit	0.21	0.41	n.a.	n.a.
Publicly traded share	0.07	0.26	n.a.	n.a.
Foreign ownership	0.12	0.31	0	1
Firm age	14.13	13.84	0	172
Manufacturing	0.48	0.5	n.a.	n.a.
Number of employees	50.92	215.22	0	6500

Covariates summary statistics.

Table 1.2 – Effect of outages on firms' revenue.

	(1) Total Sale	(2) Total Sale
Number of PO	-0.08***	
	(0.03)	
Hours of PO		-0.03
		(0.02)
small	-1.40***	-1.40***
	(0.06)	(0.06)
large	1.10***	1.09***
	(0.10)	(0.10)
very large	1.87***	1.87***
	(0.16)	(0.16)
exporter	0.60***	0.60***
	(0.09)	(0.09)
Credit	0.49***	0.49***
	(0.06)	(0.06)
Share	0.40***	0.40***
	(0.08)	(0.08)
Foreign ownership	0.92***	0.92***
	(0.09)	(0.09)
Firm age	0.34***	0.34***
	(0.03)	(0.03)
Constant	13.78***	13.63***
	(0.30)	(0.28)
Number of obs.	6677	6677
R^2	0.58	0.58

OLS estimation with robust standard errors in parenthesis. The dependent variables is the logarithm of total sale, expressed in PPP 2005$. Both regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.3 – Effect of outages on revenue of firms without a generator.

	(1) Total Sale	(2) Total Sale
Number of PO	-0.13***	
	(0.04)	
Hours of PO		-0.04*
		(0.02)
small	-1.41***	-1.41***
	(0.09)	(0.09)
large	1.38***	1.37***
	(0.16)	(0.16)
very large	1.23***	1.22***
	(0.29)	(0.29)
exporter	0.47***	0.48***
	(0.13)	(0.13)
Credit	0.56***	0.56***
	(0.09)	(0.09)
Share	0.46***	0.47***
	(0.14)	(0.14)
Foreign ownership	0.70***	0.70***
	(0.15)	(0.15)
Firm age	0.25***	0.25***
	(0.04)	(0.04)
Constant	14.64***	14.36***
	(0.59)	(0.57)
Number of obs.	2723	2723
R^2	0.68	0.68

OLS estimation with robust standard errors in parenthesis. The dependent variables is the logarithm of total sale, expressed in PPP 2005$. Both regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.4 – Effect of outages on revenue of firms with a generator.

	(1) Total Sale	(1) Total Sale
Number of PO	-0.09*	
	(0.05)	
Hours of PO		-0.04
		(0.03)
small	-1.21***	-1.21***
	(0.09)	(0.09)
large	0.78***	0.77***
	(0.13)	(0.13)
very large	2.06***	2.05***
	(0.19)	(0.19)
exporter	0.58***	0.59***
	(0.12)	(0.12)
Credit	0.33***	0.33***
	(0.09)	(0.09)
Share	0.20	0.20
	(0.11)	(0.11)
Foreign ownership	0.87***	0.87***
	(0.13)	(0.13)
Firm age	0.40***	0.41***
	(0.05)	(0.05)
Constant	13.67***	13.51***
	(0.42)	(0.39)
Number of obs.	2982	2982
R^2	0.54	0.54

OLS estimation with robust standard errors in parenthesis. The dependent variable is the logarithm of total sale, expressed in PPP 2005$. Both regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.5 – Robustness checks: Manufacturing dummy.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.07***		-0.11***		-0.08*	
	(0.03)		(0.04)		(0.04)	
Hours of PO		-0.02		-0.03		-0.03
		(0.02)		(0.02)		(0.03)
small	-1.50***	-1.50***	-1.53***	-1.54***	-1.30***	-1.30***
	(0.05)	(0.05)	(0.08)	(0.08)	(0.08)	(0.08)
large	1.19***	1.19***	1.36***	1.35***	0.90***	0.89***
	(0.09)	(0.09)	(0.16)	(0.16)	(0.12)	(0.12)
very large	1.88***	1.88***	1.10***	1.09***	2.05***	2.05***
	(0.15)	(0.15)	(0.26)	(0.26)	(0.18)	(0.18)
exporter	0.54***	0.55***	0.48***	0.48***	0.49***	0.50***
	(0.08)	(0.08)	(0.12)	(0.12)	(0.11)	(0.11)
Credit	0.54***	0.54***	0.62***	0.62***	0.34***	0.34***
	(0.05)	(0.05)	(0.09)	(0.09)	(0.08)	(0.08)
Share	0.52***	0.52***	0.55***	0.55***	0.28**	0.28**
	(0.07)	(0.07)	(0.13)	(0.13)	(0.10)	(0.10)
Foreign ownership	0.98***	0.98***	0.74***	0.74***	0.96***	0.95***
	(0.07)	(0.07)	(0.14)	(0.14)	(0.11)	(0.11)
Firm age	0.33***	0.33***	0.25***	0.25***	0.41***	0.41***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Manuf dum	-0.17***	-0.17***	-0.04	-0.04	-0.24**	-0.24**
	(0.04)	(0.04)	(0.07)	(0.07)	(0.08)	(0.08)
Constant	14.03***	13.88***	14.46***	14.20***	13.95***	13.83***
	(0.27)	(0.25)	(0.56)	(0.54)	(0.34)	(0.32)
Number of obs.	7954	7954	2981	2981	3379	3379
R^2	0.57	0.57	0.67	0.67	0.53	0.53

OLS estimation with robust standard errors in parenthesis. The dependent variables is the logarithm of total sale, expressed in PPP 2005$. Columns 1 and 2 present results for the whole sample, columns 3 and 4 for firms without a generator, columns 5 and 6 for firms without. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.6 – Robustness checks: TFP.

	(1) All TFP	(2) All TFP	(3) No generator TFP	(4) No generator TFP	(5) Generator TFP	(6) Generator TFP
Number of PO	-0.08***		-0.14***		-0.05	
	(0.03)		(0.04)		(0.05)	
Hours of PO		-0.03		-0.04*		-0.02
		(0.02)		(0.02)		(0.03)
small	0.27***	0.27***	0.21**	0.21**	0.41***	0.41***
	(0.06)	(0.06)	(0.10)	(0.10)	(0.09)	(0.09)
large	-0.43***	-0.44***	-0.10	-0.10	-0.73***	-0.74***
	(0.10)	(0.10)	(0.20)	(0.20)	(0.13)	(0.13)
very large	-1.17***	-1.18***	-1.58***	-1.57***	-1.10***	-1.10***
	(0.16)	(0.16)	(0.39)	(0.39)	(0.18)	(0.18)
exporter	0.33***	0.34***	0.25*	0.26*	0.35***	0.35***
	(0.09)	(0.09)	(0.13)	(0.13)	(0.12)	(0.12)
Credit	0.37***	0.37***	0.42***	0.42***	0.24***	0.24***
	(0.06)	(0.06)	(0.09)	(0.09)	(0.09)	(0.09)
Share	0.23**	0.23**	0.30**	0.31**	0.07	0.07
	(0.08)	(0.08)	(0.14)	(0.14)	(0.11)	(0.11)
Foreign ownership	0.76***	0.76***	0.44***	0.45***	0.78***	0.78***
	(0.08)	(0.08)	(0.14)	(0.14)	(0.13)	(0.13)
Firm age	0.24***	0.24***	0.17***	0.17***	0.30***	0.30***
	(0.03)	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Constant	0.58	0.41	0.90	0.58	1.01*	0.89*
	(0.30)	(0.29)	(0.58)	(0.57)	(0.44)	(0.43)
Number of obs.	5975	5975	2389	2389	2618	2618
R^2	0.47	0.47	0.64	0.64	0.37	0.37

OLS estimation with robust standard errors in parenthesis. The dependent variable is the logarithm of Total Factor Productivity as in Cui, Lapan and Moschini (2012). Columns 1 and 2 presents the results for the whole sample, columns 3 and 4 results for firm without a generator; columns 5 and 6 for firms with a generator. All regressions include country dummy. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.7 – Determinants of generator ownership as in FS.

	(1) Generator ownership	Elasticity	(2) Generator ownership	Elasticity
Number of outages	0.33***	0.13		
	(0.02)			
Small*Numb. Of outages	0.01	0		
	(0.01)			
Large*Numb of outages	-0.05***	-0.02		
	(0.02)			
Very large*Numb. Of outages	-0.17***	-0.07		
	(0.03)			
Hours of outages			0.20***	0.08
			(0.01)	
Small*Hours of outages			0.01	0
			(0.01)	
Large*Hours of outages			-0.04***	-0.02
			(0.01)	
Very large*Hours of outages			-0.13***	-0.05
			(0.02)	
Firm age	0.04**	0.02	0.02	0
	(0.02)		(0.02)	
Number of employees	0.39***	0.15	0.38***	0.15
	(0.03)		(0.03)	
Exporter	0.12**	0.5	0.12**	0.5
	(0.05)		(0.05)	
Constant	-2.60***		-2.23***	
	(0.12)		(0.11)	
Number of observation	7099		6838	

Probit estimation with robust standard errors clustered at the industry-country level in parenthesis. Base country: Angola; base industry: food and beverage processing; base size category: medium (20-100 employees). All regressions include country and industry dummy. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level. Elasticity calculated at base level of dummies (country=Angola and sector=Food and Beverage) and mean level of continuous variables.

Table 1.8 – Summary statistics, sensitive and non-sensitive electricity industry, all costs.

	Electricity sensitive	Non electricity sensitive
Whole sample		
Number of firms	4416	8894
Share reporting electricity as biggest obstacle	18.83%	14.28%
Average number of outage	124	165
Share of firm owning a generator	46.59%	49.36%
Average cost of electricity (share of total cost)	4.50%	3.75%
Very Large		
Number of firms	224	254
Share reporting electricity as biggest obstacle	21.87%	16.06%
Average number of outage	111	167
Share of firm owning a generator	65.00%	66.36%
Average cost of electricity (share of total cost)	3.09%	4.06%
Large		
Number of firms	429	522
Share reporting electricity as biggest obstacle	16.82%	16.15%
Average number of outage	96	154
Share of firm owning a generator	70.00%	70.57%
Average cost of electricity (share of total cost)	3.08%	2.22%
Medium		
Number of firms	1310	2169
Share reporting electricity as biggest obstacle	17.90%	14.07%
Average number of outage	112	147
Share of firm owning a generator	55.49%	59.05%
Average cost of electricity (share of total cost)	4.06%	3.34%
Small		
Number of firms	2440	5920
Share reporting electricity as biggest obstacle	19.46%	14.05%
Average number of outage	138	174
Share of firm owning a generator	35.77%	42.53%
Average cost of electricity (share of total cost)	5.30%	4.27%

Summary statistics of firm sub-samples, electricity reliance calculated considering all production costs.

Table 1.09 – Summary statistics, sensitive and non-sensitive electricity industry, only labour and electricity costs

	Electricity sensitive	Non electricity sensitive
Whole sample		
Number of firms	6308	7002
Share reporting electricity as biggest obstacle	17.09%	14.65%
Average number of outage	122	178
Share of firm owning a generator	43.26%	53.28%
Average cost of electricity (share of total cost)	17.32%	16.56%
Very Large		
Number of firms	243	235
Share reporting electricity as biggest obstacle	18%	20%
Average number of outage	111	166
Share of firm owning a generator	57%	76%
Average cost of electricity (share of total cost)	18.53%	15.84%
Large		
Number of firms	387	564
Share reporting electricity as biggest obstacle	18%	15%
Average number of outage	94	152
Share of firm owning a generator	72%	69%
Average cost of electricity (share of total cost)	15.01%	16.07%
Medium		
Number of firms	1548	1931
Share reporting electricity as biggest obstacle	17%	14%
Average number of outage	107	155
Share of firm owning a generator	56%	58%
Average cost of electricity (share of total cost)	15.24%	15.63%
Small		
Number of firms	4115	4245
Share reporting electricity as biggest obstacle	17%	14%
Average number of outage	132	194
Share of firm owning a generator	34%	47%
Average cost of electricity (share of total cost)	18.31%	17.11%

Summary statistics of firm sub-sample, electricity reliance calculated considering only electricity and labour costs.

Table 1.10 – Determinants of generator ownership as in ADS.

	(1) Generator	(2) Generator	(3) Generator	(4) Generator
Number of outages	0.29***	0.29***	0.29***	0.27***
	(0.02)	(0.02)	(0.02)	(0.02)
Electricity intensive	-0.04	-0.05		
	(0.03)	(0.14)		
Outages*Elec. Int.		0.00		
		(0.03)		
Electricity intense (E&L)			-0.12***	-0.27**
			(0.03)	(0.14)
Outages*Elec. Int. (E&L)				0.04
				(0.03)
Finance constrained	-0.28***	-0.28***	-0.27***	-0.27***
	(0.04)	(0.04)	(0.04)	(0.04)
Access to credit	0.18***	0.18***	0.18***	0.18***
	(0.04)	(0.04)	(0.04)	(0.04)
Firm age	0.16***	0.16***	0.16***	0.16***
	(0.02)	(0.02)	(0.02)	(0.02)
Capital city	0.11***	0.11***	0.10**	0.10**
	(0.03)	(0.03)	(0.03)	(0.03)
Exporter	0.29***	0.29***	0.28***	0.28***
	(0.05)	(0.05)	(0.05)	(0.05)
Foreign ownership (dummy)	0.32***	0.32***	0.31***	0.31***
	(0.04)	(0.04)	(0.04)	(0.04)
Constant	-1.64***	-1.64***	-1.58***	-1.51***
	(0.09)	(0.10)	(0.09)	(0.11)
Number of observation	7018	7018	7018	7018

Probit estimation with robust standard errors clustered at the industry-country level in parenthesis. Columns 1 and 2 presents the results for the electricity intensity measures calculate on labour and electricity costs only; columns 3 and 4 for the one that includes all production costs. Base country: Angola; base industry: food and beverage processing. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.11 – Determinants of generator ownership, sample split by electricity intensity.

	All costs		Electricity & Labour	
	Non Intensive Generator	Intensive Generator	Non Intensive Generator	Intensive Generator
Number of outages	0.29***	0.29***	0.27***	0.31***
	(0.02)	(0.03)	(0.02)	(0.02)
Finance constrained	-0.24***	-0.32***	-0.26***	-0.29***
	(0.05)	(0.06)	(0.05)	(0.05)
Access to credit	0.23***	0.12**	0.13**	0.22***
	(0.05)	(0.06)	(0.05)	(0.06)
Firm age	0.15***	0.18***	0.15***	0.16***
	(0.02)	(0.03)	(0.03)	(0.03)
Capital city	-0.01	0.27***	0.14**	0.06
	(0.04)	(0.05)	(0.05)	(0.04)
Exporter	0.28***	0.29***	0.19**	0.40***
	(0.07)	(0.08)	(0.07)	(0.08)
Foreign ownership (dummy)	0.31***	0.35***	0.35***	0.27***
	(0.06)	(0.06)	(0.06)	(0.06)
Constant	-1.58***	-1.79***	-1.50***	-1.79***
	(0.11)	(0.14)	(0.12)	(0.13)
Number of obs.	4118	2900	3483	3535

Probit estimation with robust standard errors clustered at the industry-country level in parenthesis Columns 1 and 2 presents the results for the electricity intensity measures calculate on labour and electricity costs only; columns 3 and 4 for the one that includes all production costs. Base country: Angola; base industry: food and beverage processing. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 1.12 – Firm size distribution across electricity sensitiveness (all production costs) and infrastructure quality

	Countries below outage mean	Countries above outages mean	Above - Below difference
Firms with <5 employees			
Electricity sensitive	4.81%	3.68%	-1.13%
Non electricity sensitive	12.78%	18.97%	6.19%
Sensitive-non sensitive	-7.97%	-15.29%	-7.32%
Firm with 5-10 employees			
Electricity sensitive	8.43%	5.49%	-2.94%
Non electricity sensitive	18.53%	22.50%	3.97%
Sensitive-non sensitive	-10.10%	-17.01%	-6.91%
Firm with 10-20 employees			
Electricity sensitive	7.37%	3.27%	-4.10%
Non electricity sensitive	13.15%	17.40%	4.25%
Sensitive-non sensitive	-5.78%	-14.13%	-8.35%

Firm size distribution, electricity-reliance (all costs excluding capital depreciation) and infrastructure quality.

Table 1.13 – Firm size distribution across electricity sensitiveness (only labour and electricity costs) and infrastructure quality

	Countries below outage mean	Countries above outages mean	Above - Below difference
Firms with <5 employees			
Electricity sensitive	10.81%	4.03%	-6.78%
Non electricity sensitive	6.77%	18.61%	11.84%
Sensitive-non sensitive	4.04%	-14.58%	
Firm with 5-10 employees			
Electricity sensitive	14.59%	5.90%	-8.69%
Non electricity sensitive	12.38%	22.09%	9.71%
Sensitive-non sensitive	2.21%	-16.19%	
Firm with 10-20 employees			
Electricity sensitive	10.20%	4.18%	-6.02%
Non electricity sensitive	10.32%	16.49%	6.17%
Sensitive-non sensitive	-0.12%	-12.31%	

Firm size distribution, electricity-reliance (only labour and electricity) and infrastructure quality.

Table 1.14 – Determinants of size distribution as in ADS.

	(1) Share of micro firms	(2) Share of micro firms	(3) Share of very small firms	(4) Share of very small firms	(5) Share of small firms	(6) Share of small firms
Average number of outage (/100)	0.19***	0.22***	0.43***	0.43***	0.21***	0.40***
	(0.06)	(0.06)	(0.07)	(0.07)	(0.08)	(0.15)
Interaction term	0.02		-0.18**		-0.04	
	(0.07)		(0.09)		(0.10)	
Interaction term (E&L)		-0.13		-0.18**		-0.28**
		(0.15)		(0.09)		(0.14)
Constant	0.47***	0.46***	0.48***	0.47***	0.47***	0.46***
	(0.11)	(0.10)	(0.06)	(0.06)	(0.07)	(0.07)
Number of observation	484	484	484	484	484	484

Tobit estimation with robust standard errors clustered at the industry-country level in parenthesis. The dependent variable is the share of firm with less than 20 employees in columns 1 and 2; with less than 10 employees in columns 3 and 4; with less than 5 employees in columns 5 and 6. Base country: Angola; base industry: food and beverage processing. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level

Chapter 2

Hydrological modelling of 8 African continental basins with Geospatial Stream Flow Model (GeoSFM).

2.1 Introduction.

Freshwater – apart from its direct use for drinking and sanitation – plays a vital role in agriculture, fishing (which might be generally grouped in "food security") and power generation in the whole SSA. With 62% of the continent's surface composed of international river basins, the trans-boundary nature of African water resources is paramount. In fact, amongst the 11 biggest river basins, 5 are shared between at least 8 countries. This particular characteristic has important implications for the sustainable management of water resources and for the sustainable development of economic activity related to them (de Wit and Stankiewicz 2006, Goulden et al. 2009, Scheumann and Alker 2009).

The continent is also characterized by many different climate regimes. The areas experiencing a mean level of precipitation lower than 400 mm per year ("dry regime") or between 400 and 1000 mm ("intermediate regime") jointly cover 66% of Africa, and especially the latter might be worryingly exposed to change in water availability due to climate change. Intermediate regimes – which feature highly seasonal rainfall – include the southern part of the continent, the majority of East Africa and the whole stripe connecting Senegal to Eritrea. These areas host a disproportionally high amount of the available fresh water resources available (for example, the largest include the Senegal, the Upper Volta, the Niger, the Upper Nile, the Orange, the Limpopo and Lake Chad) and further these areas tend to be densely populated. It is therefore not surprising that some of those rivers, like the Orange or Limpopo in the southern part, are already in a situation of "water scarcity" (less than 1000 m^3 available per person (UNDP 2006), while other, like the Nile or the Volta in the eastern and western part respectively, are approaching water scarcity level (Kumera *et al.* 2008 and Goulden et al. 2009).

Furthermore, there have been increasing episodes of drought[39] or floods around SSA countries with heavy human and economic costs connected[40]. It is of primary importance to be able to forecast and prepare for these situations, as all available evidence point towards an increase in their frequency. The Famine Early Warning System (FEWS), created after the famine of the 1980s by the US Agency for International Development, makes extensive use of hydrological modelling to evaluate the likelihood of extreme events which are the main causes of famine. Specifically, they employ the Geospatial Stream Flow Model (GeoSFM), developed by the US Geological Survey (USGS), which is particularly suited for areas, such as most of SSA, for which little geological and hydrological data is available.

The GeoSFM models has also recently been employed to explore how changing patterns of weather and river-flow behaviour impact on socio-economic activity. A study from Cole *et al.* (2014) analyses the potential effects of different climate scenarios on the increased hydro-dependency of many African countries, using GeoSFM to forecast future river-flow of basins in which hydropower plants are located. The analysis concludes that most of the planned investments appear to be located in regions which should experience less traumatic changes over time, although some concerns still remains for the viability of projects in particular areas. But to the extent in which the hydro-dependence of some African countries might prove problematic, it also offers us a chance to investigate a strong assumption made in the analysis of the first chapter.

---

[39] As there are no agreed upon quantifiable definition for either drought or flood the terms are going to be used throughout the text in the general meanings of prolonged periods of particularly low/high rainfall and river flow with the potential to cause human and economic damage.

[40] At the moment of writing, both Ethiopia and South Africa are experiencing heavy and persistent drought considered to be the worst in respectively 50 and 30 years, see http://www.aljazeera.com/indepth/opinion/2016/01/ethiopia-drought-happen-160121084103587.html and http://www.aljazeera.com/news/2016/01/south-africa-drought-160117111204356.html .

The estimation of the effect of power outages on firm sales through OLS regressions relies on the assumption that the first is exogenous to the second, an assumption which has been increasingly challenged in the most recent literature. Concerns that the exogeneity assumption could be violated are due to different reasons, ranging from measurement error in outage incidence to the existence of policies that affect simultaneously the outage level and firm performance or the simultaneity between economic activity and energy demand. We could though exploit the African hydro-dependency to resolve this problem, as variations in the discharge of river serving hydropower plants should lead to variation in potential electricity production without influencing electricity demand, and therefore leading to more outages, while also remaining exogenous to firm performance. Therefore, through the river flow modelling of African basins in GeoSFM we are able to construct a city-level instrument to control for possible endogeneity in the outages-sales relationship in Chapter 3. Moreover, in Chapter 4 we will use the hydrological measures we are going to developed in this chapter to explore the link between hydro-power production and city- and nation-wide economic activity during the period 2001-2013 with the use of night light data from the DMSP-OLS. Finally, we will also evaluate the capacity of GeoSFM to predict floods and droughts by comparing the model estimates with the available historical data. As this is the most widely used model for the majority of SSA a throughout assessment of its goodness of fit is paramount.

The structure of the chapter is as follow: Section 2 presents a review of the literature; Section 3 presents GeoSFM, the hydrological model used in our analysis; Section 4 presents the results; Section 5 provides a comparison between the simulated and the observed river discharge from a selection of stations; Section 6 concludes.

2.2 Hydrology, water availability and climate change in Africa.

We begin with a review of the most recent literature on the hydrology of the African continent, its water availability and the consequences of possible climate change for water access. Then we focus on some of the issues related to the hydrological modelling of African basins and the particular challenges that it poses.

Climate change has the potential to alter water availability through two main channels. First, changes in mean temperatures and, second, changes in mean precipitation. The former will increase evaporative demand, while the latter will have a direct effect on river flow. Climate models developed by the International Panel on Climate Change (IPCC, Core Writing Team *et al.* 2007) predict an increase in mean temperature between now and 2080 of 3.3°C (+1.8°C as a minimum and +4.7°C as maximum) for West Africa, of 3.2°C (+1.8°C/+4.3°C) in East Africa and of 3.4° (+1.9°C/4.8°C) in Southern Africa during all seasons[41]. With regard to change in rainfall patterns, it is often deemed more useful to present the minimum and maximum changes than the average ones given that seldom models agree over their directions. The annual precipitation in West Africa is predicted to vary between -9% and +13% according to the model selected, and similarly the range goes from -3% to +25% for East Africa and from -12% to +6% for Southern Africa (Core Writing Team *et al.* 2007, Kumera *et al.* 2008). The biggest challenge for the management of water resources is likely to be the increase in extreme events (both droughts and floods), which is already taking place and has caused significant socio-economic disruptions and human suffering in the last 15 years (UNDP 2006). The fluctuation of extreme events is connected to different climate mechanisms in different parts of the continent (e.g. El

[41] Results diverge too much for the Sahelian area to present any consistent values.

Niño for East Africa and the drying of the Sahel for West Africa) and there is need for further research connecting these with global climate change.

Moreover, even ignoring climate change processes, African rainfall and river flow patterns independently show a high degree of variability. A study of Hulme *et al.* (2001) analyses the historical change in climate and rainfall across the continent between 1900 and present day trying to shed some light on the expected trend for the next 100 years. What they find is that the long term trends in variation of temperature and rainfall vary significantly both seasonally and spatially. While it is possible to say that overall the temperature has increased over the last hundred years and it is expected to keep increasing for the next century, rainfall changes are harder to interpret and therefore harder to predict[42]. Regional and local climate characteristics which are hard to account for in models - or that are not yet fully understood conceptually – are among the main explanations.

The connection between rainfall precipitation and river flow varies equally widely across the continent, and this is not particularly surprising when the very diverse nature of its hydro-geology is added to the aforementioned regional climate characteristics. A comprehensive study from Conway et al. (2008) shows how rainfall is the main determinant of river flow variability in West Africa – in which it accounts for 60-70% of the latter – and in Central Africa, that shows a slightly weaker but still predominant link (50% of variability accounted for). Southern and Eastern Africa on the contrary show a much weaker relationship, maybe with the exclusion of the Blue Nile which is less influenced by the complex physiographical features of the Rift

[42] Even so, the contribution of anthropogenic climate change remains undetermined.

Valley. The higher variability of rainfall versus lower variability of river flows in the area is likely to be partially due to the peculiar nature of the geology of the Rift Valley.[43]

It is also worth noting that the main determinant of changes in water availability over the next two decades is likely to remain human activity and not climate change. The continuing process of urbanization, which alters the land cover of substantial areas, is considered to be one of the most relevant climate determinant in Africa and this trend is primarily due to a growing demographic pressure, which in itself will reduce absolute per capita water availability (Hulme et al. 2001). Moreover, the trans-boundary nature of many African basins has already led to over-exploitation of water resources (e.g. Lake Victoria in Uganda, see Pearce 2006) in situations in which low institutional cooperation between riparian states is intertwined with their competing needs for water availability. All of these socio-economic processes are strongly linked to the development of energy networks, and hydro-power development in particular can play a very big role in increasing either international cooperation or water related conflict (Goulden *et al.* 2009, Kumar *et al.* 2011).

We now discuss some of the main issues with hydrological analysis of African basins. The starting point has to be the well acknowledged scarcity of gauge stations across Africa: of the more than 9000 gauge station data available through the Global Runoff Data Centre (GRDC, one of the main providers of publicly accessible hydrological data), only 13% are situated in Africa[44]. This implies both a lack of research relative to many relevant African basins and that

[43] The problem of low conversion rate of rainfall to run-off and its determinant for Eastern and Southern Africa has already been explored by Schulze *et al.* (2001).
[44] A gauge station is a location used by hydrologist to monitor the water level surface elevation and volumetric discharge of a terrestrial water body.

most of the efforts for developing sustainable practices of water management have to rely on hydrological modelling. Many types of models are available to analyse hydrological behaviours, reproducing the cycle as precisely as possible given initial information. The reliance of most of these models on the availability of a wide set of other physical information about the basin makes them though hard to apply to ungauged rivers and in data poor regions, so that uncertainties remain on which is the best approach to follow when information is not readily available (see Kapangaziwiri *et al.* 2009 for a review with special focus on South Africa). Lack of data makes calibration and validation hard in the majority of situations, and when attempted possible discrepancies are not easily explained (see for example Munzimi 2008 also using GeoSFM and Tshimanga *et al.* 2011 and Tshimanga *et al.* 2012 with regard to the Congo Basin).

The research which is more strongly connected with the methodology and areas of this chapter is that by Asante *et al* (2008a), the authors of the GeoSFM package for the FEWS.[45] In their paper they analyse the performance of GeoSFM estimates for Lake Chad, Nile, Zambezi, Congo and Niger when compared to observed streamflow for the period 1998-2005. The authors find that the uncalibrated estimates of absolute streamflow can differ significantly from those observed. Conversely, the performance in terms of anomalies prediction is much higher, as the model was developed to forecast situations of extreme events such as floods or droughts.[46]

[45] Famine Early Warning System, funded in the second half of the 1980s by USAID in response to heavy famines in West and East Africa, aim to provide "objective evidence based analysis to help government decision makers" (FEWS website) and is active in 36 countries in Africa, Asia and the Caribbean.
[46] Anomalies are defined as difference between daily observations and long term mean scaled by the standard deviation.

2.3 The GeoSFM package.

All the river flow estimates presented in the chapter have been obtained using the GeoSFM utilities for ArcView 3.x, a program widely used for the processing of geographical information. These utilities have been developed by USGS for the Earth Resource Observation and Science (EROS) as part of the FEWS. The following explanation is meant to be a summary for the reader, so to help understand the procedure used to obtain the estimates. For a complete treatment of the utilities included in the package refer to Artan G. *et al.* (2008) and Asante *et al.* (2008b), on which most of the first section is based.

2.3.1 Pre-processing modules.

The pre-processing modules included in the package are used to recreate a virtual representation of the geographical zone under analysis, in our case the majority of the African continent, first delimiting the basin, then characterizing its topography and hydrological behaviour. The next session will briefly outline the functioning of the three pre-processing routines, "Terrain Analysis", "Basin Characterization" and "Unit Hydrographic Response"

2.3.1.1 Terrain Analysis.

The first routine, "Terrain Analysis", requires two datasets as inputs: a Digital Elevation Model (DEM)[47] and a shapefile[48] containing information on river networks, which are combined to

[47] Digital Elevation Models are three dimensional representations of geographical surfaces based on elevation data.
[48] Shapefile are geospatial vector data format for geographic information system.

define the spatial extent of the watershed under analysis. In our case, both of these datasets come from Hydro1k, a geographic database developed at the USGS EROS centre derived from the USGS' 30 arc-second DEM of the world, GTOPO30. This database contains hydrologically correct DEMs for every continent, therefore reducing the need of pre-processing to fill spurious pits[49]. The shapefile in the database contains all African watersheds, obtained using information on stream networks and flow directions. Each of them has been classified with the Pfafstetter methodology, which assigns them a unique ID based on the topology of land surface. The method is hierarchal, with watersheds firstly delineated depending on their location within the overall drainage system and secondly divided into three different types. The first, "basin", contains the headwater of the river that defines the watershed and does not receive drainage from any other watershed; the second, "interbasin", in contrast, receives drainage from upstream watersheds; finally there are "internal basins", which neither receive nor contribute to the flow of any other watershed or waterbody. The method is applied recursively as long as four tributaries can be identified for each watershed. Continental watersheds are labelled with level 1 Pfafstetter code (see Figure 2.1). Level 2 codes classify sub-watersheds of level 1 basins and interbasins. Level 3 identifies sub-watersheds of level 2 and so on.[50] The final result is a series of 6-digit codes which uniquely identify the 7,131 African watersheds smaller than 4000 km^2 (Furnans and Olivera 2001, Verdin 1997).

[Figure 2.1 about here]

[49] The program also includes a utility to fill them in case the DEMs used are not hydrologically correct.
[50] The method also permits us to identify quickly any watershed situated upstream or downstream of the one of interest.

Once these two grids are uploaded into the program, the routine determines in which direction water that falls on a grid-cell will flow, comparing its elevation with that of the 8 neighbouring grid-cells and exploiting the basic physics principle that water flows in the direction of the steepest descent (Jenson and Dominque 1988). Every cell is therefore assigned one of the eight compass directions, and a series of grids required for the following steps are subsequently created. These include, on top of the initial one describing flow direction, a grid with the number of upstream cells for any given location; one with the nearest sink; one that groups cells which reside within the same watershed; one identifying the most downstream cell for each river; one identifying every sub-basin which forms the watershed; one which measures the highest change in elevation in the neighbourhood of any given cell and finally one determining the sub-watershed directly downstream of each river reach.

2.3.1.2 Basin Characterization.

The hydrological model of GeoSFM does not solely rely on elevation to determine streamflow behaviour, but also uses a series of information pertaining to land cover and soil characteristics, which play important roles in defining a watershed rate of runoff generation and overland flow transport. As for the previous routine, a series of datasets are required as inputs. The first dataset concerns land cover data, obtained through USGS Global Land Cover Characterization (GLCC) database[51] (Loveland *et al.* 2000) and is used to create an impervious area grid accounting for the existence of water-bodies and wetlands in the sub-basin. The second set of data are those that include soil characteristics, which are created from the Digital Soil Map of the world produced by the United Nations Food and Agriculture Organization in collaboration with the

---

[51] GLCC data were derived from 1-Km AVHRR data and are available as both Interrupted Goode Homolosine and Lambert Equal-Area Azimuthal projection. The second has been used to obtain the results.

United Nations Educational Scientific and Cultural Organization. They are (a) a grid containing information on soil texture, characterized using data from Zobler (1986), (b) a grid relative to soil top and bottom depth, as well as the mixture of salt, silt and clay in the composition of 106 soil types, characterized using data from the Global Data Set of Soil Particle Size Properties developed by National Aeronautics and Space Administration (NASA, Webb *et al* 1993), (c) a grid that describes the soil hydraulic conductivity and (d) a grid relative to the soil water holding capacity, both based on the seven Zobler texture classes. Finally, land cover and soil data are combined to define the last grid needed, that relative to the soil conservation service runoff curve numbers, and are used in determining the amount of incident precipitation becoming surface runoff[52]. All the above information is aggregated, and all cells of the basin grid delineated in the previous step are assigned the appropriate value for each class of soil characteristics. The combination allows the program to determine a long series of other basin characteristics, ranging from the days of interflow residence time to the area of the basin which is covered by water. All of these newly obtained data are written to two different text files, one of which will be used by the soil moisture accounting routine and one by the river flow transport routine.

2.3.1.3 Unit Hydrograph response.

While GeoSFM requires information about precipitation for each watershed, its distribution within the catchment area is left implicit, with a unit hydrograph developed to evaluate the catchment response to a uniformly distributed rain event.[53] The land cover grid from USGS

[52] Assigned based on land cover type and soil hydraulic class of the U.S. Department of Agriculture.
[53] A unit hydrograph is the theoretical response of a catchment in terms of runoff volume and timing to a unit input of rainfall.

GLCC which was used to determine the impervious area in the previous step is the last grid that needs to be created for the pre-processing routines. The unit hydrograph is based on a uniform land velocity for each watershed, determined by its average slope and the predominant land cover type as from equation:

$$VELOCITY = \frac{1}{MANNINGN} * R_H^{\frac{2}{3}} * \sqrt{HILLSLOPE}$$

where MANNINGN is the Manning roughness value for the predominant land cover, R is the hydraulic radius and HILLSLOPE is the elevation change divided by the length from each cell to the outlet (average) . Once the distance along the flow path from each catchment cell to the outlet is calculated, the final distribution of discharge is given by the probability density function (PDF) of the watershed travel times. The unit hydrograph is then the discretization of the PDF over a routing interval, thus it is the probability mass function of flow travel times.[54]

2.3.2 Processing modules.

The processing routines included in the program – "Weather data processing routines", "Soil moisture accounting module" and "River transport module" – are those which yield the final estimation of each catchment's streamflow.

[54] See Asante et al. (2008) for a more in depth coverage of the passages involved.

2.3.2.1 Weather data processing routines.

Weather data, and precipitation data in particular, are the main inputs of any hydrological model. It is worth remembering that GeoSFM was developed for parts of the world without good gauge stations coverage, widely recognized as the best source of precipitation data. Hence, the routines are created to work mainly with satellite-derived rainfall estimates (RFE) and, equally, with satellite-derived estimate of evapotranspiration, which is the second weather input required by the program[55]. Rainfall estimates come from the Climate Prediction Centre of the National Oceanic and Atmospheric Administration (NOAA), and are calculated using a methodology developed by Xie and Arkin (1997)[56]. Their technique exploits the fact that rain is formed when moisture in the atmosphere reaches the point of condensation, 235 Kelvin degree or below. Therefore, by measuring with satellites the temperature of the clouds it is possible to determine where rain events are occurring, and both gauge measurement at experimental sites and from the Global Telecommunications System of the World Meteorological Organization are used to double check the rainfall estimates. The evapotranspiration data, on the other hand, come from two different sources: monthly data for the period 1983-2000 come from the Climate Research Unit of the University of East Anglia, daily data for the period 2001-2010 are produced by the FEWS group at USGS EROS[57]. Both the data-sets have been obtained through the solution of the Penman-Monteith equation, whose

---

[55] Evapotranspiration is a vital part of the water cycle, including both evaporation – that is, movement of water from the soil or waterbody to the air – and transpiration – that is, the movement of water in a plant which terminates with vapour leaving its leaves.

[56] In fact, the methodology was updated during the period of the study, so that estimates from the period 1983-2000 were obtained with the original methodology RFE 1.0 and those for the period 2001-2010 with the updated RFE 2.0. The two methodologies differ mostly on the ground of the availability after 2000 of technologically more advanced satellites and are consistent with one another.

[57] Data from the CRU are calculated on a monthly frequency, while the GeoSFM routine requires them on a daily frequency, format in which they do not exist for the period. The monthly mean has therefore being spread across the month. Also, they have a slightly higher resolution, 50km x 50km, than the one from USGS EROS, which are based on a 100km x 100km grid.

methodology takes into account many different environmental variables - from wind speed to fluxes of long and short wave radiation – to obtain a final value for evapotranspiration. As a final step before running the hydrological model itself, the routine in GeoSFM computes the mean areal precipitation (Xie and Arkin, 1997) and evapotranspiration (Verdin and Klavier 2002) values for each watershed over the decided time interval.

[Figure 2.2 and 2.3 about here]

2.3.2.2 Soil Moisture Accounting Module.

GeoSFM is a continuous hydrological model and hence contains a routine for computing daily runoff and soil moisture, thus separating the rain between atmospheric releases, surface runoff and subsurface flow. Two options are provided: a Linear and a Nonlinear Soil Moisture Accounting. Due to the dimension of the area under analysis, the less computational intensive linear routine has been preferred to the non-linear routine. To begin with, the module evaluates the maximum storage capacity of each sub-basin, obtained multiplying its soil water holding capacity by its soil depth. Then the initial value of soil moisture is computed, assuming no storage in the ground water reservoir at time 0. Consequently, at the beginning of every period the fraction of incident rainfall that becomes directly surface runoff - proportional to the impervious share of the sub-basin - is estimated using:

$$PARATIO = MIN \begin{cases} Max\left[\left(\frac{GWSTORE - SOILDEPTH}{HILLSLOPE - HILLLENGTH}\right), 0\right] \\ 0 \\ 1 - IMPERVIOUS\ RATIO \end{cases}$$

where GWSTORE is storage in groundwater reservoir, HILLSLOPE is as above and HILLLENGTH is the average length between each cell and the catchment outlet. This determines the amount of the watershed surface area that becomes saturated by sub-surface storage, becoming part of the partial contributing area[58]. Excess precipitation, defined as the fraction of rainfall which falls on permanently impervious area or on partial contributing area, is then calculated using:

$$EXCESSRAIN = MAX \left\{ \begin{array}{c} STORE_{t-1} + RAIN - STMAX \\ - \\ RAIN * [PARATIO + IMPERVIOUS\ RATIO] \end{array} \right.$$

where $STORE_{t-1}$ is the quantity of water stored in the subsurface reservoir in the previous period, STMAX is the maximum storage capacity of the subsurface reservoir and PARATIO is the expression defined above. The remaining amount is allowed to either enter the soil or contribute to the subsurface reservoir. The following step is the conversion of potential to actual evapotranspiration, which depends on the availability of moisture from rainfall, runoff sources or soil moisture. The routine then moves on the computation of water losses outside the river boundary, of the amount of water which contributes to the groundwater reservoir generating the base flow and, finally, of the share percolating to regional groundwater systems. Once the base-flow contribution to surface runoff is taken into account, a linear reservoir formulation is

[58] It is from this step onwards the two procedure for soil moisture accounting differ. The one used assumes a linear relationships between the amount of water in sub-surface storage exceeding the soil depth and the partial contributing area. The alternative procedure relaxes the assumption and represents more completely the various sub-surfaces processes. For a sub-set of the basins we tried using the non-linear routine and obtained highly comparable results. See Asante et al. (2008b) for a complete treatment of this issue.

used to calculate the amount of water contributing to the network of rivers and lakes as interflow following:

$$INTERFLOW = \left(\frac{RESIDUAL\ STORAGE}{INTERFLOW\ LAG}\right) * \exp\left(-\frac{1}{INTERFLOW\ LAG}\right)$$

where RESIDUAL STORAGE represents what is left in the groundwater reservoir after the base flow is generated and INTERFLOW LAG is the residence time required for the exchange to happen. The residual moisture in both soil layer and groundwater reservoirs are finally computed, resulting in a representation of moisture fluxes and storage in the period under consideration.

2.3.2.3 River Transport Module.

The final step is the simulation of the horizontal movement of the runoff generated within each catchment from the catchment outlets to the basin outlets. Even in this case GeoSFM offers different options: two linear routines, pure lag and diffusion analog, and a non-linear one, that is called the Muskingum Cunge. Again, due to the dimension of the area under consideration the less computational intensive approach has been preferred (pure lag routing), while the other two have been used as robustness checks for a subset of rivers. The time saving characteristic of the pure lag routing routine is that it does not take account of any attenuation or deformation of the input, so that flow remains unaltered at the discharge point. That is, inflow to the upstream end of each river is moved with a time delay to its downstream end without any change to its magnitude. A mathematical expression for this routine can be given by:

$$D_t = I_{t-t'}$$

where D_t is discharge at time t, t' is the travel time between the input and discharge location and $I_{t-t'}$ is the input at time (t-t'). Intuitively, surplus runoff from excess precipitation, interflow and base-flow is first aggregated over each sub-basin, then converted into flow units (cubic metres per second). Finally, it is translated to the end of the reach[59], taking into account each sub-basin's particular response and allowing for translational losses. The process is repeated, transferring each river discharge to the next downstream reach until it reaches the outlet and saving the discharge from every river reach.

2.4 Simulation results.

We now present the results of the stream-flow simulation for level 1 basins for the period 2001-2010. Given the hydrological diversity and the geographical distance amongst the various basins, the results are presented individually, with maps referring to the average value for each sub-basin across the decade. To further analyse the possibility of common trends in streamflow behaviour across different sub-basins, we perform a principal component analysis (PCA) to individuate hydrological groups.

PCA is a data compression technique which focuses on the eigenvectors of the correlation matrices, allowing for a reduction of dimension in the data without leading to a great loss of

[59] A reach is the length of the stream between any two points in it.

detail. Intuitively, the focus is on the underlying structure of the data: for a given set of N variables which might be correlated, PCA transformation yield at most N-1 uncorrelated variables, which can be seen as grouping together those amongst the original which accounted for similar characteristics. This is due to the nature of the procedure: the first principal component (PC) represents the maximum amount of variance in the data and all the following express the highest unexplained residual variance. As each PC is further uncorrelated with all the others, by concentrating the analysis only on the first significant components, a great deal of simplification of the data is achievable (Westra *et. al.* 2007). That is, instead of having to consider a great number of variables, some of which can play an almost insignificant role in the analysis at hand, one can concentrate on a (possibly) much smaller number which surely play a relevant role.

Various methods have been proposed in the literature for determining the number of significant PC to be retained in the analysis. Due to its computational ease, we decided to use Horn's parallel analysis (Horn 1965), which builds on the observation of Kaiser (Kaiser 1960) that only PC with eigenvalues greater than 1 must be considered meaningful. Shortly, Horn's insight is that, in finite samples, eigenvalues are very likely to be overestimated due to least-square bias and sampling errors, and must hence be adjusted before taking the decision of which PC to retain.

The final technical point which remains to be covered is that of component rotation, that is the application of linear transformation to the outcome of PCA analysis, as the technique allows for infinite alternative solution satisfying the original equation (Richman 1986). Unrotated results presents a series of characteristics which might affect certain kind of analysis: first, the topography of unrotated PC is primarily affected by the shape of the area under analysis (Buell 1979); second, unrotated solutions are more exposed to large sampling errors, especially when neighbouring eigenvalues are close together (North *et al.* 1982); finally, there are many cases in which the rotated solutions are more readily interpretable from a meteorological point of view. Taking these points into account, we decided to apply an oblimin rotation, one of the more commonly used in the literature.[60]

With particular referral to hydrological applications, when the variables considered are time-series of spatially distributed information the eigenvectors of the correlation matrix represent their optimal mode of spatio-temporal variability (Singh and Singh 1996). Given the possibility to represent these vectors in space, this technique has been receiving growing interest since the 1980s, when climate scientist were using it to identify trend in geo-climatic data[61]. In our case the series under analysis are the daily streamflow of all the sub-basins of a continental watershed. For each continental basin we report the result of the Horn's test and present a map with the groupings obtained over the loadings of the first PC, both before and after the rotation.

2.4.1 Lake Chad.

[Figure 2.4 to 2.9 about here]

---

[60] See Richman (1981) for a theoretical discussion of oblimin rotation and Richman and Lamb (1985) for an application to the US.

[61] There are a great number of studies that use this technique in climatology. Some of the best general coverage of the early literature is given by O'Leanic and Levezey (1988), more recent application are covered in Westra *et al.* (2007) and Ssengane *et al.* (2012). For some recent economic application see Barrios, Bertinelli and Strobl (2010).

Figure 2.4 shows the spatial distribution of average stream-flow over the 2001-2010 period for all the sub-basins of Lake Chad, which varies from 0 to 3,246 cubic meters per second. Figure 2.5 presents instead rivers discharges for the average month aggregated over the whole basin. Note that stream-flow starts to rise after April to reach its peak in September. Figure 2.6 shows that the unimodal behaviour peak is consistent all across the time period under consideration, with all maximum flows taking place between August and September and the maximum stream-flow of the decade in 2006. Figure 2.7 represents the result of the Horn's test: non-correction for finite sample bias would lead to retain the first 50 components as opposed to the 39 retained after correcting for the bias. Figure 2.8 shows the spatial grouping over the first PC loading, accounting for 41% of variation in the data, with blue areas representing negative loadings and red positive. Finally, Figure 2.9 shows instead the spatial grouping over the rotated component, highlighting how most of the variation in the component is accounted by only three sub-basins.

2.4.2 Nile.

[Figure 2.10 to 2.15 about here]

As in the previous case, Figure 2.10 shows the spatial distribution of mean stream-flows for all Nile's sub-basins, averaging between 0 and 5,708 cubic meters per second. Taking into account their geographical proximity, it is not surprising that the stream-flow pattern (also for the average month in Figure 2.11) appears similar to that of Lake Chad, with the rises in streamflow starting in May and reaching their peak in August. Figure 2.12 shows the behaviour for all the different years taken into account, which is again consistently unimodal across time with all peaks taking place between August and September and the highest rivers discharge of the decade in 2007. Moving to the PCA, Figure 2.13 reports the result of the Horn's test, leading to reduce the number of retained components from 66 to 42 after adjustments. Loading of the first PC (28% of data variability) is shown in Figure 2.14, with blue areas representing negative loadings and red positive, while oblimin rotation (Figure 2.15) does not lead to a much clearer picture in this case.

2.4.3 Interbasin 3.

[Figure 2.16 to 2.21 about here]

In the case of Interbasin 3, the mean river discharge of its sub-basins varies between 0 and 879 cubic meters per second (Figure 2.16). As Figure 2.17 shows, averaging over an inter-basin system gives a less clear picture than in the case of a basin, as hydro-climatic zones might differ consistently (in this case, the inter-basin spans from Egypt to Mozambique). This is confirmed also by the discharge across the ten years (Figure 2.18). Therefore, for the next interbasins we shall omit the discharge analysis over time. In this case, correcting for finite sample bias reduces the number of relevant components from 30 to 21 (Figure 2.19). Figure 2.20 presents the first PC loadings (50% variation), which are in this case positive all across the area, while Figure

2.21 shows how most of the variation is accounted for by a few sub-basins quite distant from one another.

2.4.4 Zambezi.

[Figure 2.22 to 2.27 about here]

Zambesi's sub-basins exhibit an average stream-flow ranging from 0 to 5,251 cubic meters per second (Figure 2.22). The behaviour of the stream-flow appears to be unimodal with peak in February (Figure 2.23), as confirmed by Figure 2.24 which shows how all peaks happens between February and March. The highest discharge of the decade has been experienced in 2008. Horn's test leads even in this case to a reduction of the number of retained components from 13 to 9, as in Figure 2.25. The first component loading (70% of data variability, Figure 2.26) are positive all across the basin even in this case, while Figure 2.27 highlights how the loading of the rotated component decrease quite consistently on the north-to-south axis.

2.4.5 Interbasin 5.

[Figure 2.28 to 2.33 about here]

Figure 2.28 shows interbasin 5' mean discharge, varying from 0 to 4,837 cubic meters per second. Correcting for finite sample bias reduces the components from 48 to 35 (Figure 2.31), with the blue areas of Figure 2.32 representing the negative loadings of the first PC (35% of variation in the data). Figure 2.33, presenting the rotated components, is not particularly informative in this case.

2.4.6 Congo.

[Figure 2.34 to 2.39 about here]

The average Congo's simulated stream-flow ranges from 0 to 21,131 cubic meters per second (Figure 2.34), with the peak being reached in February as shown by Figure 2.35. While not as clear as in other cases, the behaviour of the river seems to be unimodal[62], with peak reached between November and March as seen in Figure 2.36, with the highest discharge in 2004. Horn's test (Figure 2.37) leads to discard components from 23 to 33 due to their irrelevance after bias' correction. Figure 2.38 presents the loading of the first PC (50% variation), negative over the blue areas and leading to the division of the basin in 4 different quite distinct areas. Figure 2.39 presents the result of oblimin rotation.

---

[62] A river characterized by unimodal streamflow presents a single peak per year, as opposed to bimodal rivers which characterized by more than one.

2.4.7 Interbasin 7.

[Figure 2.40 to 2.45 about here]

Figure 2.40 illustrates the average discharge for the sub-basins of interbasin 7, which varies between 0 and 1,279 cubic meters per second. In this case, the darker red areas of Figure 2.44 (48% of data variation in the first component) are the only one with positive loading, and it can be seen from Figure 2.45 how all of the variation is due to just one sub-basin.

2.4.8 Niger.

[Figure 2.46 to 2.51 about here]

For Niger the average discharge varies from 0 to 7,695 cubic meters per second (Figure 2.46). As can be noticed in Figure 2.47, the basin behaviour seems unimodal with the highest discharge for the year experienced in September. Figure 2.48 confirms this picture, showing that maximum flow is always reached between August and September. The maximum discharge in the period 2001-2010 was in 2010. The result of Horn's test are presented in Figure 2.49, with the number of retained coefficient decreasing from 62 to 39 after bias correction. Figure 2.50 shows the loading for the first PC (42% of data variation), and even in this case an areas of positive (red) and ones of negative (blue) loading are delineated in the map, while the rotated components (Figure 2.51) do not present a clear pattern.

2.4.9 Interbasin 9.

[Figure 2.52 to 2.57 about here]

Lastly, Figure 2.52 presents the mean stream-flow for interbasin 9, ranging from 0 to 3,557 cubic meters per second. Figure 2.56 presents the loadings of the first retained PC (67% of data variation), positive in the whole basin and apparently connected to a small number of rivers in the inner part of the basin (Figure 2.57).

2.5 Comparison with GRDC data.

This section presents a comparison between flow discharges simulated through GeoSFM and the available historical data. As previously mentioned, gauge stations are the most generally used tools to gather hydrological information and the ideal base from which to start any hydrological modelling attempt. Unfortunately, the majority of Sub-Saharan African basins are still ungauged; moreover some of the gauge stations installed have been lost due to poor maintenance, war and general unrest or a combination of the two. In comparison with many other areas of the world there is remarkably little historical data available for African basins, especially over long periods of time.

Amongst the publicly available information, the biggest datasets is that of the GRDC of the German Federal Institute for Hydrology, an international data archive with global coverage and data up to 200 hundred years old, containing river discharge data for more than 9000 gauge stations globally with an average record period of 41 years. To exemplify the initial point about scarcity of gauge stations in the continent, of the 9009 included in the GRDC datasets, 1165 are located in Africa, slightly less than 13%.

[Figure 2.58 about here]

The data for 440 GRDC (Figure 2.58) have been matched to the corresponding basins in the model shapefile and cross referenced using the ALCOM/WWF classification (Verheust and Johnson 1998) given that the Pfafstetter coding used by GeoSFM carries no connection to topographical meaning. This permits the indentification of the river relative to the data but it does not guarantee a perfect geographical association to the exact point in which the gauge station is located. The first assessment of the model's goodness of fit will be given by the linear association between its estimates and the realized stream-flows.

[Table 2.1 about here]

Table 2.1 presents the number of gauge stations in each level 1 basin and the overall correlation between simulated and observed discharge. As it can be noted, the majority of stations show a significant correlation with the GeoSGM estimates, even though the proportion changes across the different basins. At a station level, correlation coefficients vary widely, ranging from -0.38

to 0.99. 52 stations (12%) show a negative correlation[63]. If we look at the geographical distribution amongst Western, Eastern, Central and Southern Africa, it becomes apparent that the lowest correlation are those for rivers located in the southern area of the continent (see Figure 2.59). On the other hand, there is evidence of good correlation in Central Africa (see Figure 2.60) and in Western Africa (Figure 2.61). The only coverage in the eastern part of the continent is offered by the few stations in Ethiopia, with a mean correlation of 0.64 (Figure 2.62).

[Figure 2.59 to 2.62 about here]

The reasons for bad fit of estimates are numerous. With regards to the correlation with pre-1998 observed discharges the unavailability of daily evapotranspiration data and the subsequent use of daily data generated from monthly data is the best candidate, especially for basins in which intra-monthly variability is fairly high. Rainfall distribution with high seasonality might also be over- or under-stimated with satellite derived products, and scarcity of rain gauge makes ex-post adjustment particularly hard with regard to the African continent. Another possibility is that the low resolution of the 1 km elevation grid cannot replicate with enough accuracy the actual topography, while also the shape of the boundaries of some of the medium-small river basins has elsewhere been questioned (Munzimi 2008, Werth et al 2009).

[63] Excluding Interbasin 5 the number drops to 20.

Following the findings by Asante et al. (2008a) we now focus on the analysis of the anomalies[64], comparing them with those from GRDC station. Figure 2.63 shows both density and kernel density of the significant correlations. Overall, only 14 stations shows an insignificant correlation with the estimated anomalies, and for the significant ones the magnitude varies widely, ranging from 0.7 across Lake Chad to 0.36 in the Nile Basin. With regard to Lake Chad, Niger and interbasin 5, more than half the stations have a correlation higher than 0.5 (see Figures 2.64, 2.65 and 2.66 respectively), while the percentage lowers to 41% for Congo and 33% for interbasin 3. Interbasin 3 also hosts the higher proportion of negative coefficients (66%), which are higher than 10% in the only other case of interbasin 9 (35%, see Figures 2.67 and 2.68 respectively).

[Figure 2.63 to 2.68 about here]

One of the drawbacks of correlation as measure of association is that it assumes that the joint distribution of the two variables is elliptical, i.e. symmetric around the mean value. This is though not necessarily the case, as the GeoSFM model might be better suited to forecast droughts in certain geographical areas and floods in others. An alternative way to assess the goodness of fit of the GeoSFM model is then to relax the assumptions requiring the joint distribution to be symmetrical and hence the dependence measure to be linear. This can be done exploiting the properties of Copula functions. Copulas function have been recently receiving increasing attention in hydrology and in climate science more in general, with application

[64] As previously mentioned, anomalies are defined as difference between observations and long term mean expressed in term of standard deviation.

ranging from field significance to discharge-duration-frequency analysis (see Renard and Lang 2006 for a review of case studies).

More precisely, given two random variables X_1 and X_2, this method allows to describe their joint cumulative distribution function $F_{1,2}(x,x)$ by means of their marginal distribution functions $F_1(x)$, $F_2(x)$ and a function (the Copula) describing their dependence. Moreover, as long as the $F_j(x)$ are continuous, the Copula always exists and it is unique. Moreover, there is no requirement for the two variables to follow the same distribution, as the technique divides the estimation of the marginal distribution of each variable from that of their dependence. Copulas are then particularly interesting because they allow to express the dependency between non-normally distributed components (as the extreme events in hydrology, which are known to be non-normal) and to model more openly the relationship amongst the tails of a multivariate distribution, i.e. to have a stronger dependence for lower (drought) or higher (flood) values than for the rest.[65] Two main families of Copula are normally used in hydrology: elliptical (mostly Gaussian and Student t), which can be extended to arbitrary dimension but require radial symmetry, and Archimedean which allows to model upper and lower tail behaviour but can be applied only to bivariate cases.[66]

In this study, we shall use both families to assess different characteristics of GeoSFM model. Firstly, we will apply elliptical copulas to evaluate how well the model reproduces the regional

[65] For a general statistical treatment see for example Mikosch (2006) or Fredricks and Nelsen (2007), while for more hydrology specific one Genest and Favre (2007) and Schölzel and Friedrichs (2008).
[66] The three Archimedean Copulas more commonly used are: Gumbel, for cases in which upper tail behaviour is of primary interest; Clayton, for cases in which lower tail behaviour is the main objective of the study; Frank, for intermediate situations.

dependence amongst sub-basin within the same continental watershed. This is possible because these functions can be extended to more than two dimensions, so that comparing the dependence amongst the gauge stations in a basin to that of the corresponding estimates allows for an assessment of the ability of the model to represent diverse regional hydrological systems. Secondly, Archimedean copulas, which allows for stronger dependence in one or both of the tails than in the central part, will be used as an alternative to the anomaly transformation to see if the model does indeed represents more accurately the incidence of floods and droughts than the average streamflow behaviour.[67]

[Table 2.2 to 2.8 about here]

Tables from 2.2 to 2.8 present the difference between the regional dependence in the 6 sub-basins for which we have data. First, we have modelled both GRDC and GeoSFM data for each continental basins through Gaussian Copulas, obtaining a measure for the rank correlation amongst the different sub-basins in the historical and simulated data; secondly we have calculated the difference amongst the two simply subtracting the second (GeoSFM) from the first (GRDC). Hence, values above zero imply that modelled flows of each sub-basin within the continental one exhibit a lower dependence than that of the historical data. As can be noted by the prevalence of negative over positive signs in all tables, the model tends to overestimate the dependence across streamflow behaviour in all basins. This is not particularly surprising, as the GeoSFM model averages the values of rainfall and evapotranspiration over big geographical

[67] This can be considered a true alternative as Copulas are invariant to positive monotonic transformation, such as the transformation of streamflow into anomalies.

areas, so that by feeding the same weather inputs to different rivers located within the same sub-basins, the model automatically increases the dependence amongst the streamflow. Overall, 25% of the estimated coefficients deviates more than 0.5, either positively or negatively, from the historical one, with the worse situation to be found in the Zambezi and best in interbasin 3.[68]

[Table 2.9 about here]

Table 2.9 presents instead a summary of the Kendall Tau, a non-parametric measure of dependence which can be applied to ranked data, obtained through the modelling of Archimedean Copulas. The number of stations with a long enough record to perform the analysis varies significantly, from a minimum of 2 for the Nile to a maximum of 64 for the Interbasin 5. Overall, 121 out of the 182 total stations show a significant fit for the Copula modelling[69], of which only 4 with a negative dependence (i.e. 3%, a lower proportion than the 12% obtained through "simple" correlation).

Looking at the average Kendall's Tau value in each basin, it can be noted how the modelling through Archimedean copulas does generally lead to a weaker dependence between simulated and historical data than that which will be obtained through simple correlation. Moreover, in the majority of cases (78.5%) the best performing copula[70] appears to be the Gumbel (stronger

---

[68] The results are highly comparable if a Student Copula is used instead of a Gaussian, the amount of coefficient deviating more than 0.5, for example, is in this case 27%.
[69] The significance of Copulas has been determined via the implementation of the Anderson-Darling test in R through the Copula package developed by Hofert *et al.* (2015). Application of the package are presented in Yan (2007), Kojadinovic and Yan (2010) and Hofert and Maechler (2011).
[70] The comparison amongst different possible Copulas has been perforemd through the test based on

right-tail dependence), followed by the Frank (19.8%) while Clayton copulas are selected as the best-fit for only 2 stations across all the continental basins. The copula analysis seems then to confirm what previously observed: the GeoSFM model has an acceptable fit for SSA and fares better in assessing the incidence of flood than in assessing average streamflow as exemplified by the best fit of Gumbel Copulas.

To conclude the comparison we perform a series of regressions aimed at verifying with different techniques the results of the previous analysis. To start, we estimate a panel data with fixed effect having as dependent variable either the daily absolute streamflow or the daily streamflow anomaly from each GRDC station and as explanatory the correspondent value from GeoSFM, the monthly mean and maximum value of rainfall and evapotranspiration (the two weather inputs used by the model) and year dummies[71]. Table 2.10 presents the results for both absolute streamflow and daily anomalies.

[Table 2.10 and 2.11 about here]

As it can be seen, while in the case of absolute streamflow GeoSFM estimates are insignificant in that of anomalies they are positive and significant at the 1% level, further adding weight to the claim that GeoSFM is better suited to estimate deviation from the mean that the mean itself. Table 2.11 shows the coefficients for the GeoSFM estimates disaggregated at the basin level so to understand if there are differences in the fit across the continent. It is possible to notice how

minimization of mean squared error proposed by Barbe *et al.* (1996).

[71] Both models are estimated with robust standard errors.

the lowest point estimates are to be found in the case of the Nile, while the second lowest in the case of Interbasin 5, which accounts for the majority of observation in the sample. Interbasin 3, the Zambezi and Interbasin 9 present instead the highest estimates in the sample.[72]

To further explore if the GeoSFM model does indeed perform better in forecast of extreme events than in that of general streamflow, we also perform quantile regressions[73]. Table 2.12 presents the results for absolute flow, Table 2.13 for anomalies.

[Table 2.12 and 2.13 about here]

[Figure 2.69 to 2.77 about here]

In both cases, GeoSFM estimates are positive and significant at 1% level, but while the point estimates increase in quintile for absolute flow, they decrease for anomalies. This can be partially explained by the fact that absolute flow is naturally lower bounded at zero, while anomalies, normalized by the standard deviation, can vary freely from negative to positive values. Figure 2.72 presents the graph for the anomalies coefficients when all basins are aggregated together, while Figures 2.69 to 2.77 present the results of Table 2.13 graphically. As can be seen from these figures, the same trends can be individuated across all basins with the exclusion of the Nile, which results are though based on only two stations in the same sub-

[72] This is also consistent with the results of the previous analysis in terms of correlations.
[73] The model also include dummies and is estimated with robust standard error.

basins. The GeoSFM model always perform better in the first 2 quantiles, with point estimates consistently higher than the OLS ones, and then loses power.

Table 2.14 presents the results for a quantile panel data regression with time fixed effect and individual fixed effects (at the station level) as in Powell (2014). Again, it can be noted that while the coefficients for absolute streamflow are significant only in the 3 highest quantile, those for anomalies are always positive and significant with decreasing magnitude in quintiles.

[Table 2.14 about here]

We also try to determinate what are the drivers of the differences between the flow forecasted by the model and that of historical data. To do so, we first create a dummy variable equal to one if the difference between the GeoSFM anomaly and the historical anomaly is greater than one standard deviation[74]. This threshold implies 17.6% of overall available data, with the highest concentration to be found in Lake Chad and the Nile and the lowest in Interbasin 3 and 5. We then perform a panel probit regression using as explanatory variables latitude and longitude of the gauge stations, the mean elevation in the basin from Hydro1k, the predominant land cover type from Loveland *et al.* 2000 and the average soil depth from Webb *et al.* 1993. In a successive specification we further include daily rainfall and evapotranspiration anomalies as robustness check, while both model include year dummies. Results are reported in Table 2.15.

[74] Recall that anomalies are already express in term of standard deviation from the long run mean.

[Table 2.15 about here]

As can be seen from Table 2.15, elevation above sea level does not seem to play any significant role in explaining deviation of the forecasted anomalies from the historical one as the coefficient is significant but of irrelevant size when weather variable are excluded and become insignificant when included. Similarly, no role seems to play either latitude or longitude, while increases in the daily anomalies for evapotranspiration and rainfall diminishes the likelihood of a substantially different forecast from the model. Looking at the predominant land cover, the worse fit for the model is to be found in urban landscape (benchmark) as all other land cover types are significant and with negative sign. The best performance are then those for basins predominantly characterized by shrub-land, for those sparsely vegetated or those covered by evergreen forest. A few soil types seem also to influence the likelihood of relevant discrepancies between modelled and actual anomalies, we refer to Webb *et al.* 1993 for the corresponding soil types as the discussion of the different properties of diverse soil composition is outside the scope of the chapter.

Lastly, we want to perform a final check of the ability of GeoSFM to forecast actual streamflow. This will take the form of a series of probit models where the dependent variable is a dummy equal to 1 if both the historical daily anomaly and the results of the model are above or below two given thresholds, which are 1 and 0.5 s.d. (positive, negative and in absolute value). The results of Table 2.16 refer to daily anomalies, while in Table 2.17 they refer to the average of the daily anomalies over a month. The explanatory variables always include the estimated anomalies (only coefficients reported) and year dummies, column 2 also includes weather variables (daily values of evapotranspiration and rainfall in Table 2.16 and their monthly mean in Table 2.17), in column 3 we add latitude and longitude and in column 4 soil depth and land cover types. Standard errors are clustered at the gauge station level.

[Table 2.16 and 2.17 about here]

As it emerges from both tables, GeoSFM anomalies are a positive and significant predictors of the historical anomalies. The model has a better fit for forecasting below-average flow than above-average as the point estimates for negative anomalies are always higher than those for positive anomalies, regardless of the other coefficients included in the regression. Positive anomalies above 1 s.d. represents the only case in which GeoSFM seems not to perform well, especially when we consider the average monthly value. These situation represents though less than 1% of both daily and monthly anomalies, so that overall we can still conclude that the model furnishes a good approximation of flow behaviour. To further appreciate the spatial differences across the African continent, Tables 2.18 and 2.19 report the shares of cases, for daily and monthly anomalies respectively, in which the forecast of GeoSFM falls in the same range of the historical values for each of the 8 continental basins. As can be seen from the tables, the share is never lower than 58%, with the highest correspondence found for Bain 3 and the lowest for Lake Chad (if we exclude the Nile for which only two gauge stations are available).

[Table 2.18 and 2.19 about here]

The previous analyses have demonstrated that GeoSFM is indeed a good predictor of actual river flow, especially in the form of deviation from the historical mean value, or anomaly. In the last regressions we will turn, anticipating the analyses of the following chapters, to how well does our anomaly measure serve the purpose of an instrument for power outages. Unfortunately, no longitudinal information is available regarding the frequency of outages throughout SSA. However, as our identification strategies relies on variation of water available for hydropower affecting actual hydroelectricity production we can rely on this relationship to assess if the instrument is fit for purpose. Moreover, as the exclusion restriction implies that the only channel through which variation of water available influences firms' operation is through outages, there should not be any significant association with industrial electricity demand. As data regarding the latter are unavailable for the SSA region, we will look instead at the relationship of our hydrological variable with the electricity consumption of industry and of all other sectors of the economy as best available proxys.

Information on actual hydroelectricity production, industrial and sectoral electricity consumption (all in GWh) is available from the statistics division of the International Energy Agency for 18 of the countries included in the analyses of the following chapters.[75] The anomaly measures employed in the following regressions are the same used for the national analyses of Chapter 4, in which the average yearly anomaly of all hydropower-serving basins of each country are scaled by their hydro-plant contribution to the state generation portfolio,

[75] These countries are Angola, Cameroon, Congo, Democratic Republic of Congo, Ethiopia, Gabon, Ghana, Ivory Coast, Kenya, Mozambique, Namibia, Nigeria, South Africa, Sudan, Tanzania, Togo, Zambia and Zimbabwe.

either in term of installed or of operational capacity, and then aggregated together[76]. All the following regressions are run as panel OLS with fixed effects and include year dummies.

[Table 2.20 and 2.21 about here]

As it can be seen from the first two columns of Table 2.20, regardless of the weight used the yearly mean anomaly is always positively and significantly associated with actual hydroelectric production, exactly as expected. On the other hand, no meaningful relationship can be individuated with regard to industrial electricity consumption. In Table 2.21 we show that the association with all other sectoral electricity consumption is also insignificant. Although the ideal measure would be industrial demand, for a significant relationship between the latter and the anomaly to exist without it reflecting on consumption when supply increases, a similar effect must interest the demand some of the other sector for which access to electricity is prioritised, and this should reflect on consumption. As we have just shown, this is not the case, so that we are reassured that the exclusion restriction is not violated.

2.6. Conclusion.

The chapter presents the results of a hydrological modelling exercise aimed at obtaining the estimates of the river discharge for 9 of the 10 African continental basins in the period 2001-2013. The literature suggests that a difference exists amongst the various basins from many points of view, ranging from water availability to rainfall-to-runoff conversion rates.

[76] See section 2 of Chapter 3 for a complete description of the instrument-building procedure.

Furthermore, many diverse pictures emerge from the perspective of the possible effects of climate change on water resource availability, with the main agreement being on the general increase of both temperature and incidence of extreme events. The scarcity of available physical data for most basins of the continent makes hydrological modelling particularly difficult, increasing the sources of uncertainties in the results of the models which underpin sustainable water management policies. The basic functioning of the hydrological model GeoSFM, used to obtain the estimates, has been presented in the third section of the chapter. The model has been chosen because of its ability to yield estimates in case of little observed data, as it depends solely on remotely sensed information; further, it can easily handle basin of large dimensions and it is fully integrated into the GIS system which makes its execution particularly tractable.

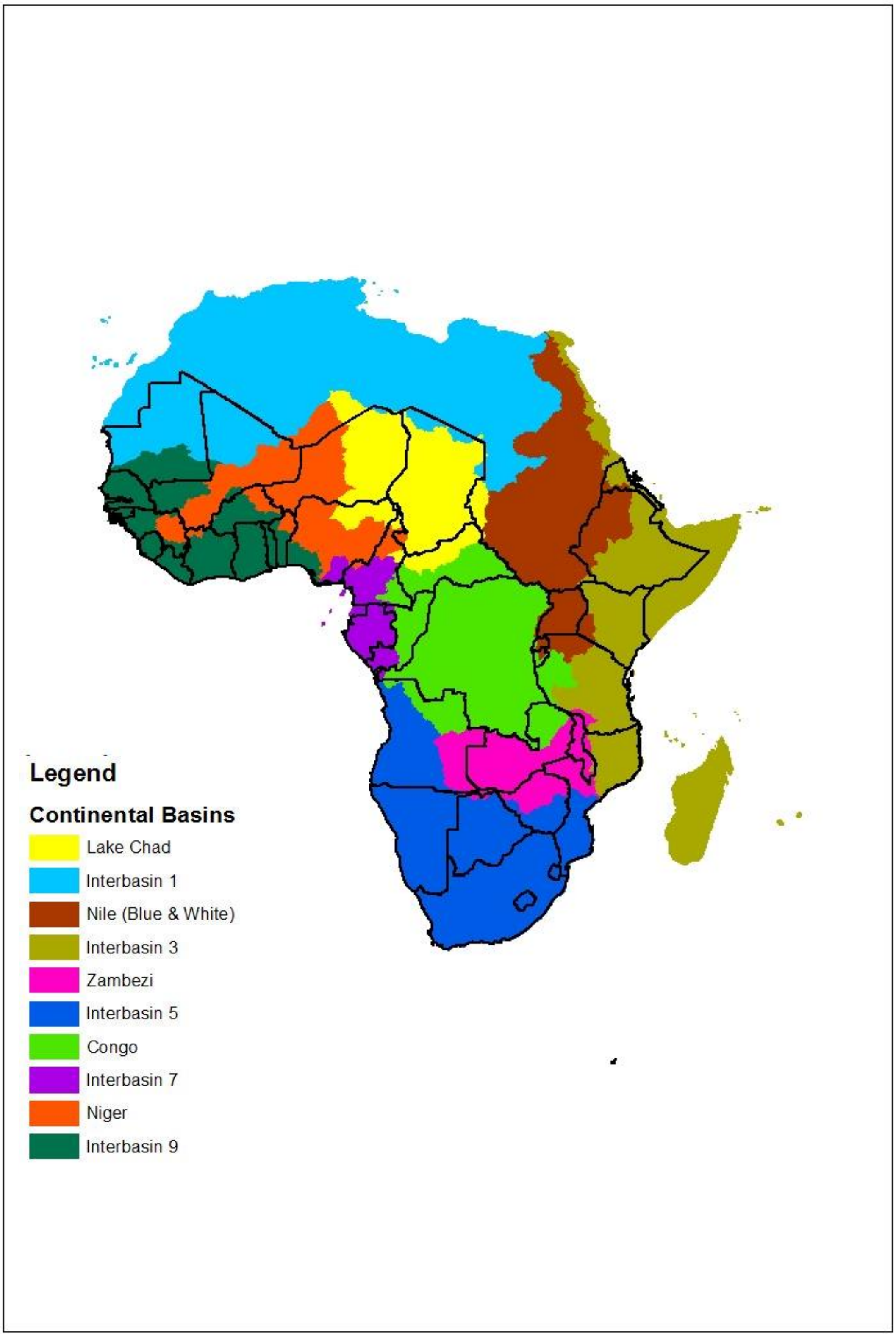
The fourth part of the chapter introduces the estimates, constituting the basis for the construction of a measure for the availability of water for hydropower production at the city-level, which will be used in Chapter 3 as an instrument for power outages and in Chapter 4 to investigate the link between hydroelectricity production and city- and country-wide economic activity. We present the spatial distribution of average stream-flow across each basin and the behaviour of river discharge across the period under consideration, while also presenting the grouping relative to both original and oblimin rotated principal components obtained through a PCA analysis of the streamflow series.

The final part of the chapter covers the comparison between some of the river discharges simulated by the model and their historical counterparts, available through GRDC, the biggest public dataset for streamflow. The overall correlation across the different basins is of 0.35, but many differences exists across them, with most of the divergences being situated in the southern part of the continent. Comparing estimates and observed data under the point of view of anomalies, we confirmed with regard to GRDC data what Asante *et al.* (2008a) did with regard to RivDis data - with a lesser performance of the model for rivers located in interbasins 3 and 9. We then moved to a more rigorous assessment of the better performance of the model with regard to droughts and floods through the use of copula analysis, which allows us to compare estimated and historical streamflow across their whole distribution. The exercise wholly confirms the previously obtained results.

We finally performed a series of checks in the form of fixed effect and quantile regression analysis of the relation between observed and estimated streamflow, also including an analysis of what might drive the discrepancies between the two. These regressions again support the verdict that the model seems to be able to represent the African flow behaviour fairly well as in no cases the correspondence between GeoSFM forecasts and the historical values falls below 58%.Although it would be of interest for further research to investigate more deeply the reason for the discrepancies individuated at the end of section 5, this falls outside the scope of the current study.

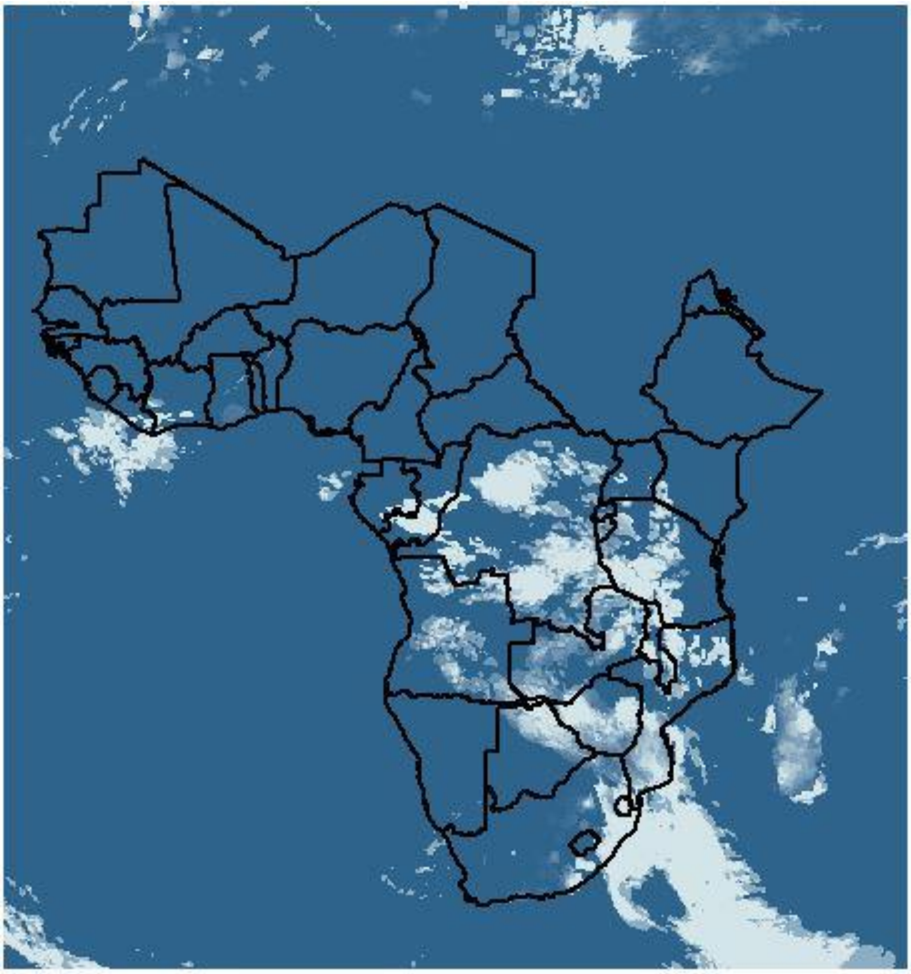
In the final part of the fourth section we introduce the measure of water availability utilised in the following chapters to show that the hydrological modelling exercise achieved its scope: the measure constructed is significantly and positively associated with actual hydropower production but not with industrial energy consumption, in that meeting the exclusion restriction of a valid instrument.

Figure 2.1 – Basins and countries in the study.



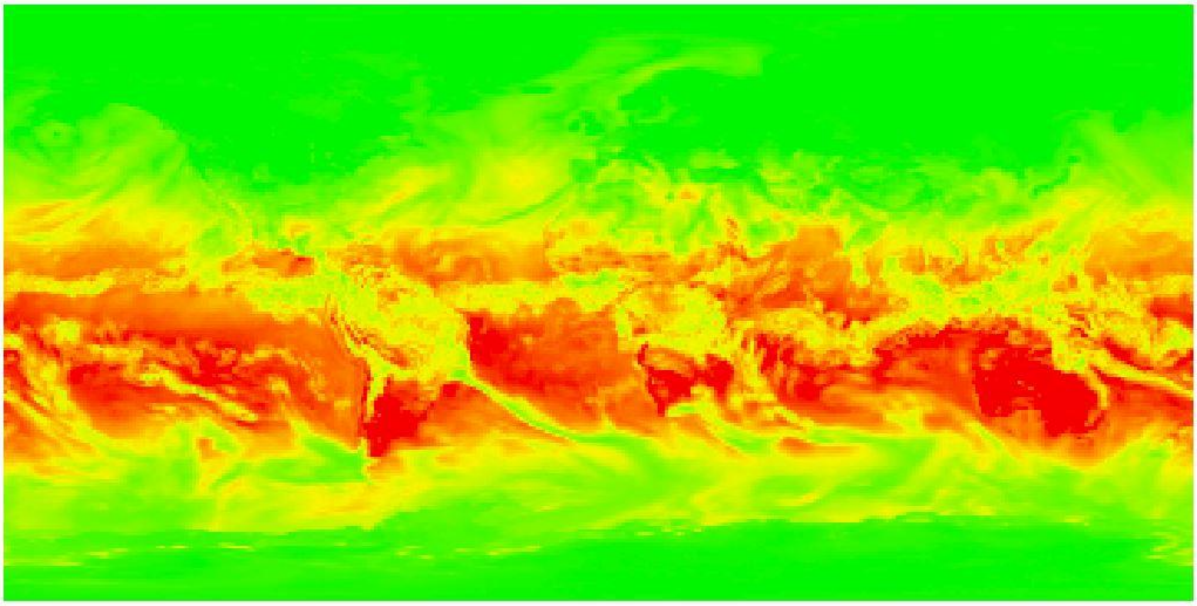
The ten continental basins with political borders of countries in the study.

Figure 2.2 – Rainfall data.



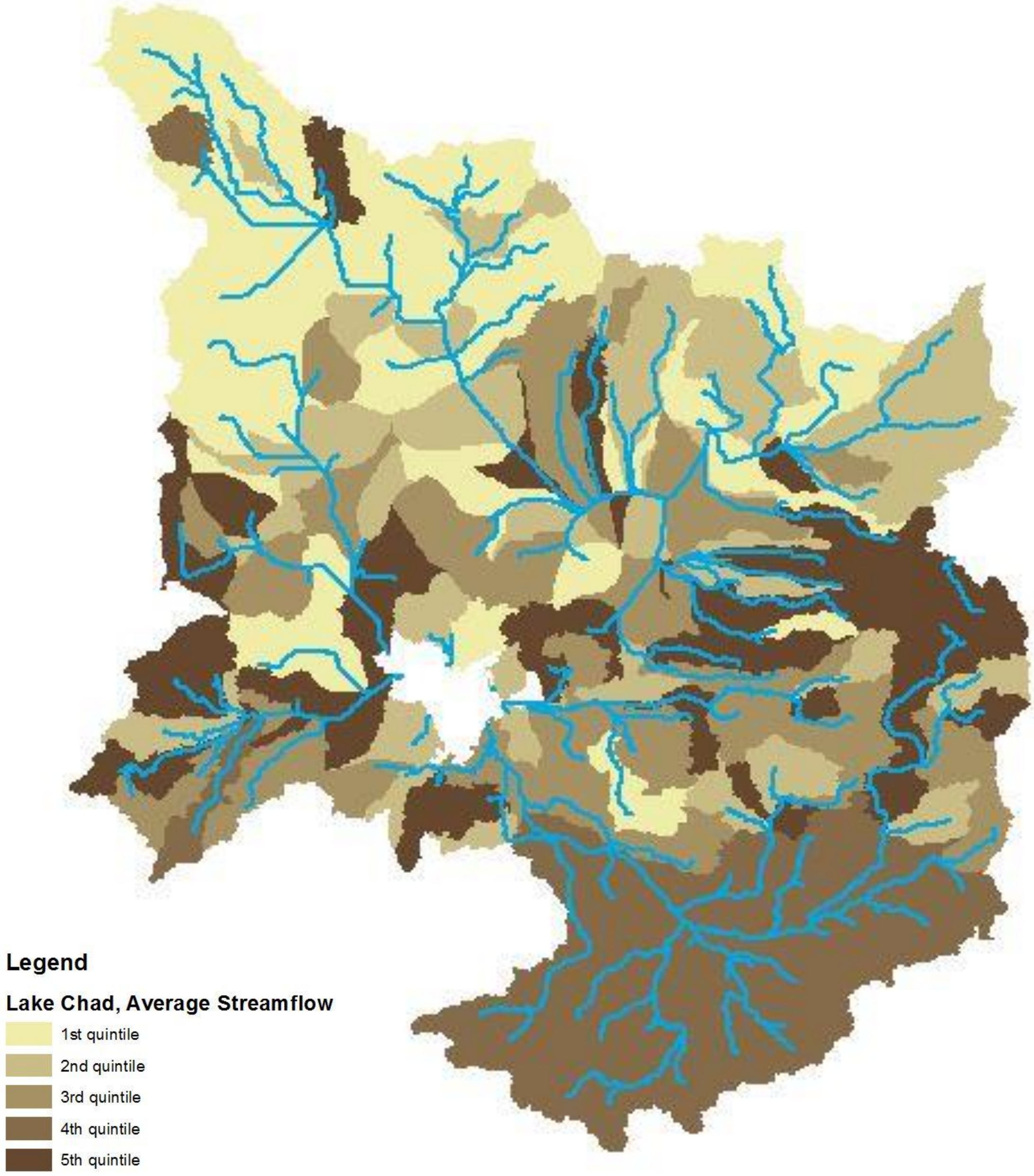
Example of a rainfall grid with superimposed countries' border, 1st January 2001.

Figure 2.3 – Evapotranspiration data.



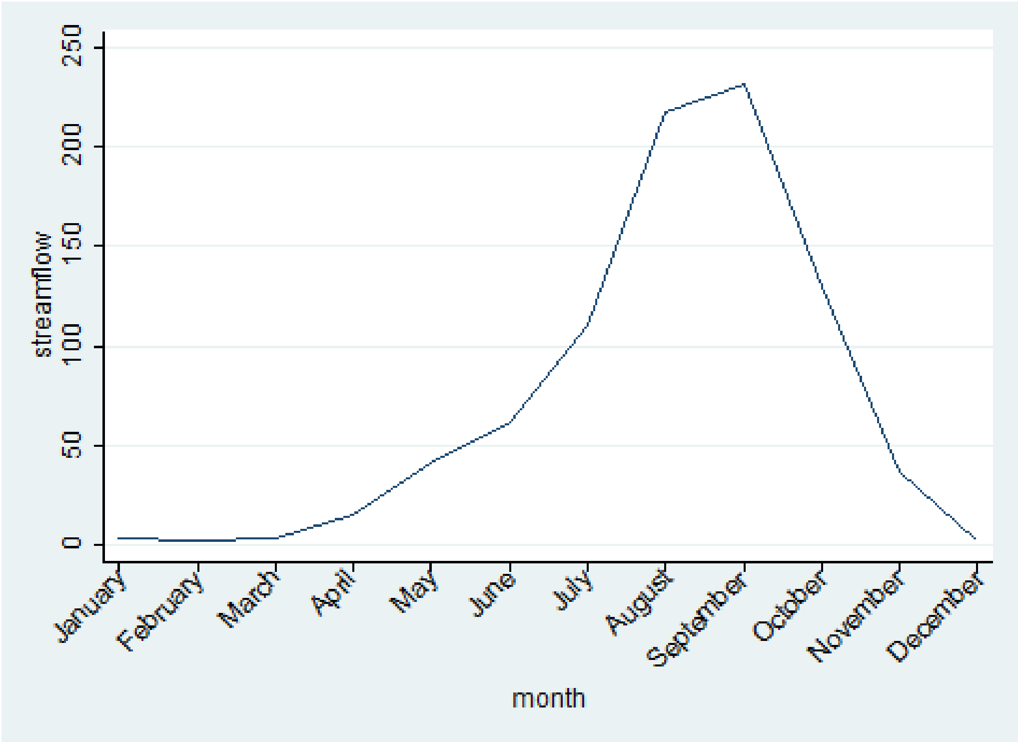
Example of an evapotranspiration grid, 1st January 2001.

Figure 2.4 – Lake Chad.



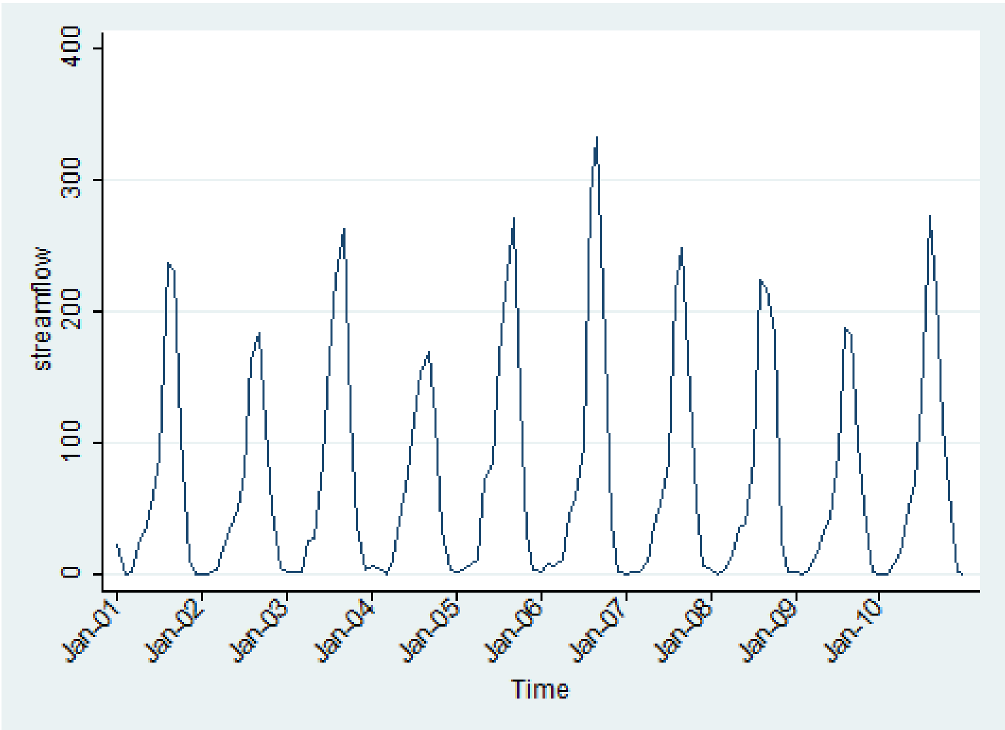
Lake Chad, river network and average streamflow by quintile.

Figure 2.5 – Monthly average streamflow, Lake Chad.



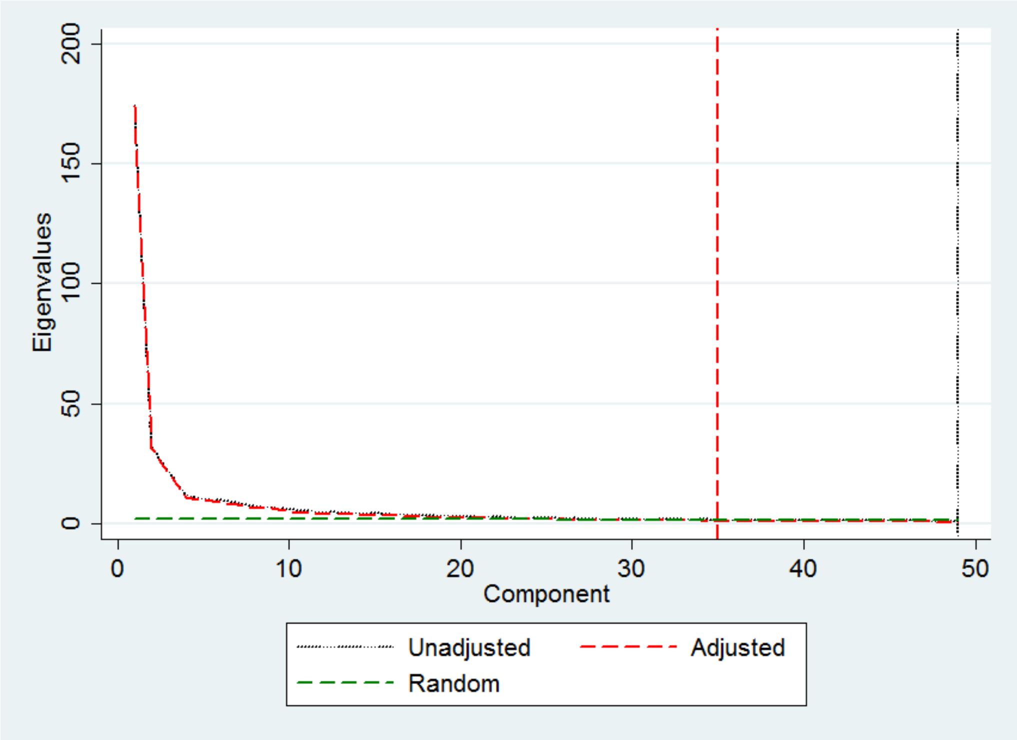
Lake Chad, simulated monthly average stream-flow.

Figure 2.6 – Monthly streamflow 2001-2010, Lake Chad.



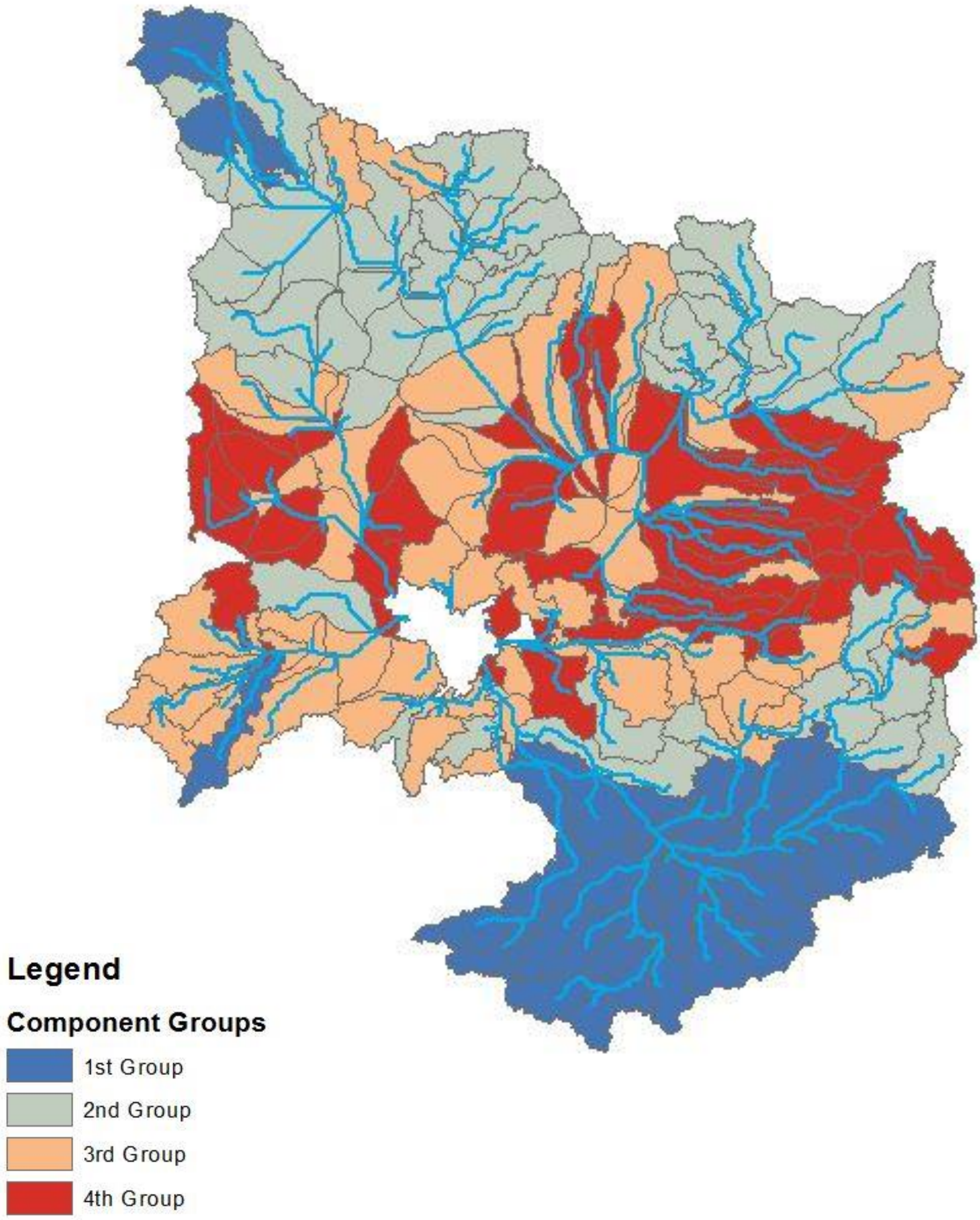
Lake Chad, simulated monthly average stream-flow, 2001-2010.

Figure 2.7 – Horn's test, Lake Chad.



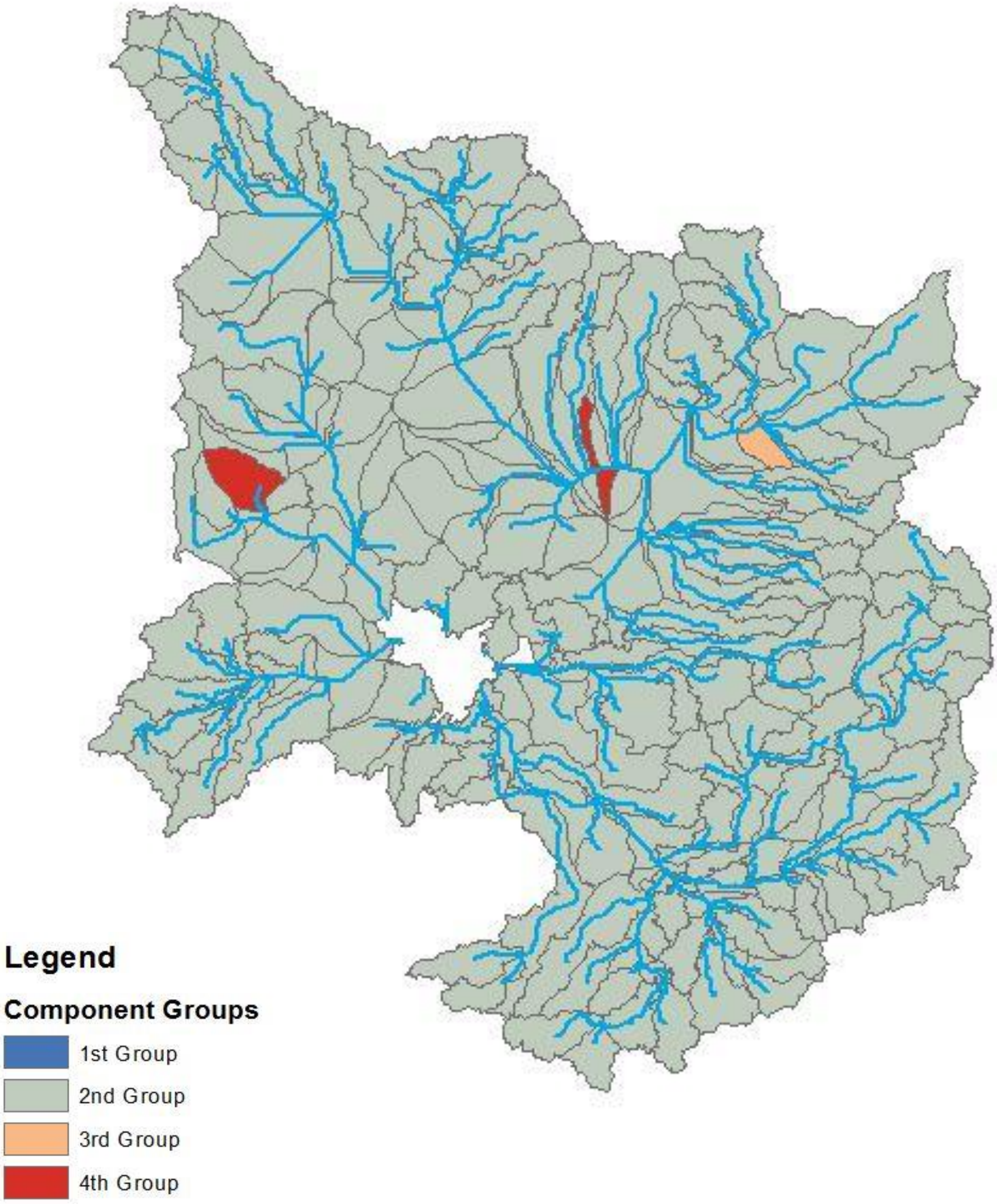
Lake Chad, result of the Horn's test.

Figure 2.8 – First principal component, Lake Chad.



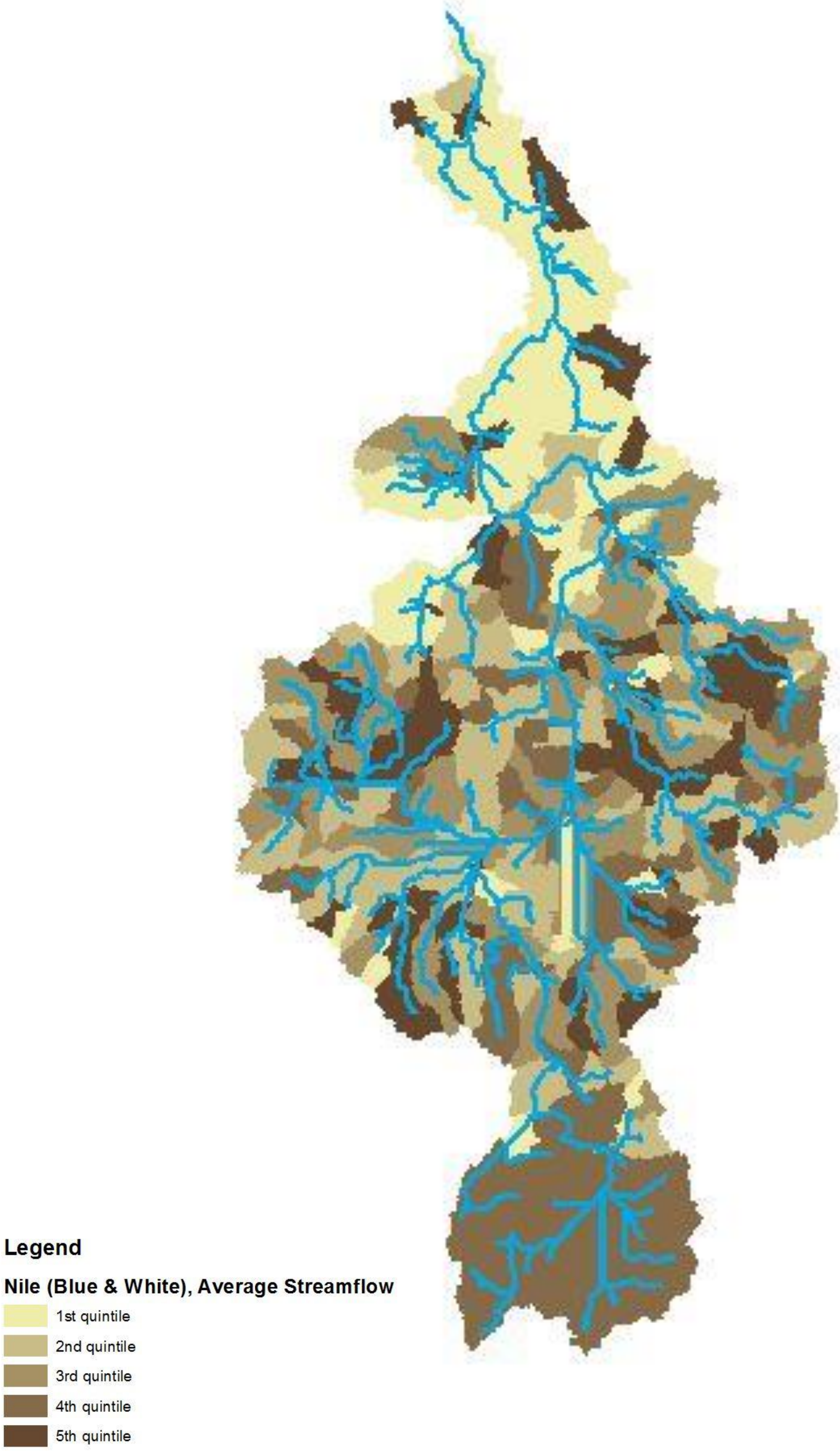
Lake Chad, grouping over the loadings of the first principal component.

Figure 2.9 – First principal component, oblimin rotation, Lake Chad.



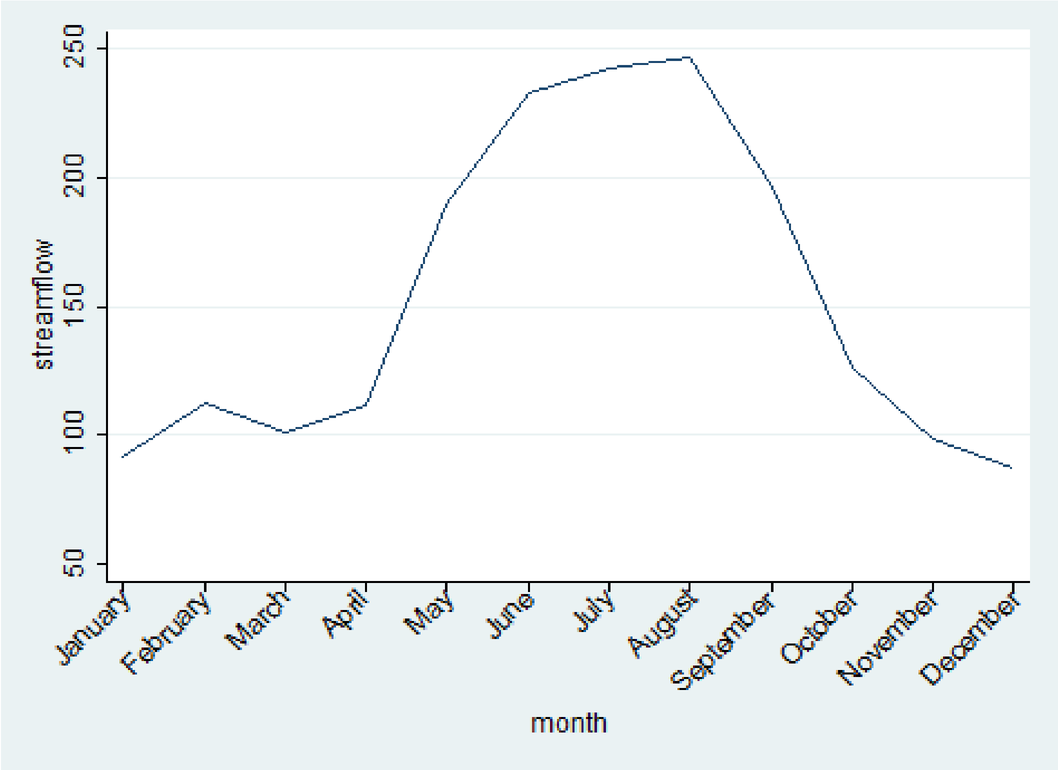
Lake Chad, grouping over the loadings of the first principal component after oblimin rotation.

Figure 2.10 – Nile.



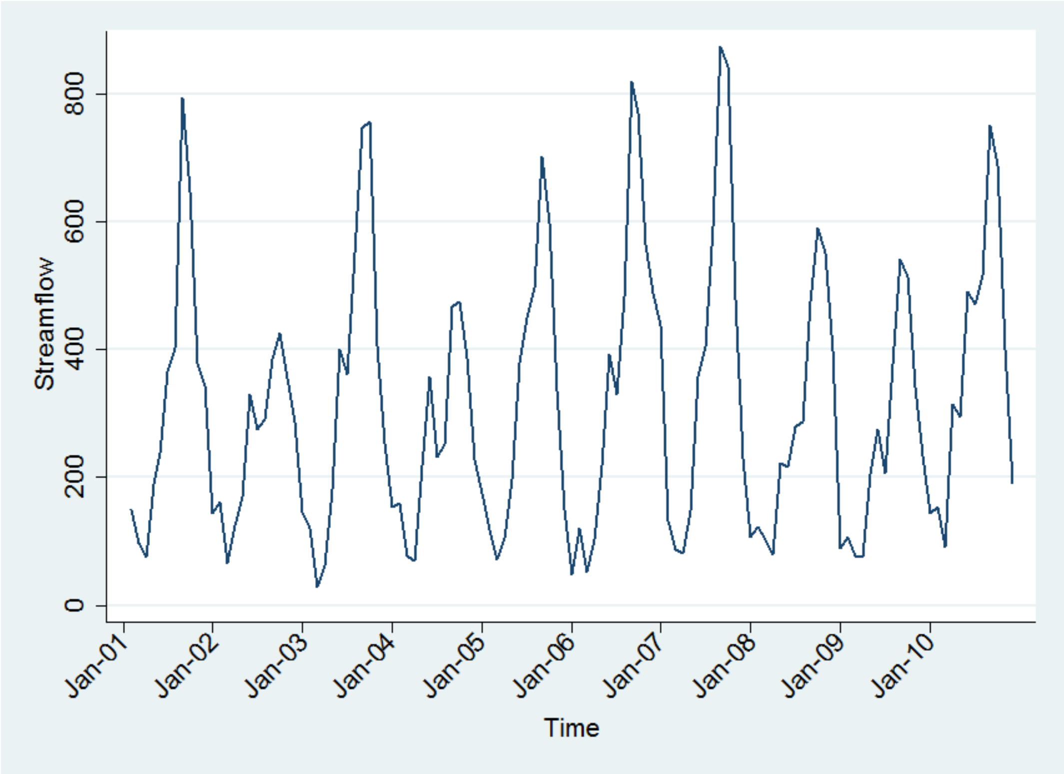
Nile, river network and average streamflow by quintile.

Figure 2.11 – Monthly average streamflow, Nile.



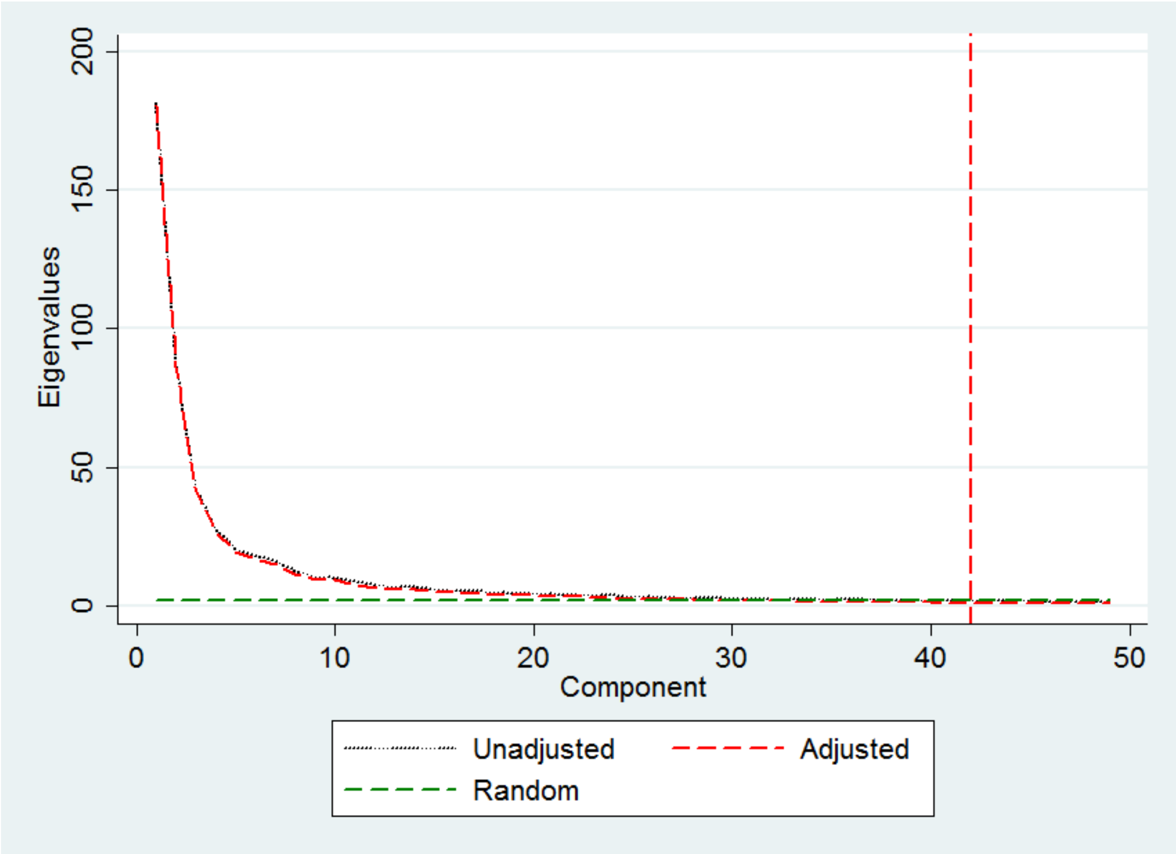
Nile, simulated monthly average stream-flow.

Figure 2.12 – Monthly streamflow 2001-2010, Nile.



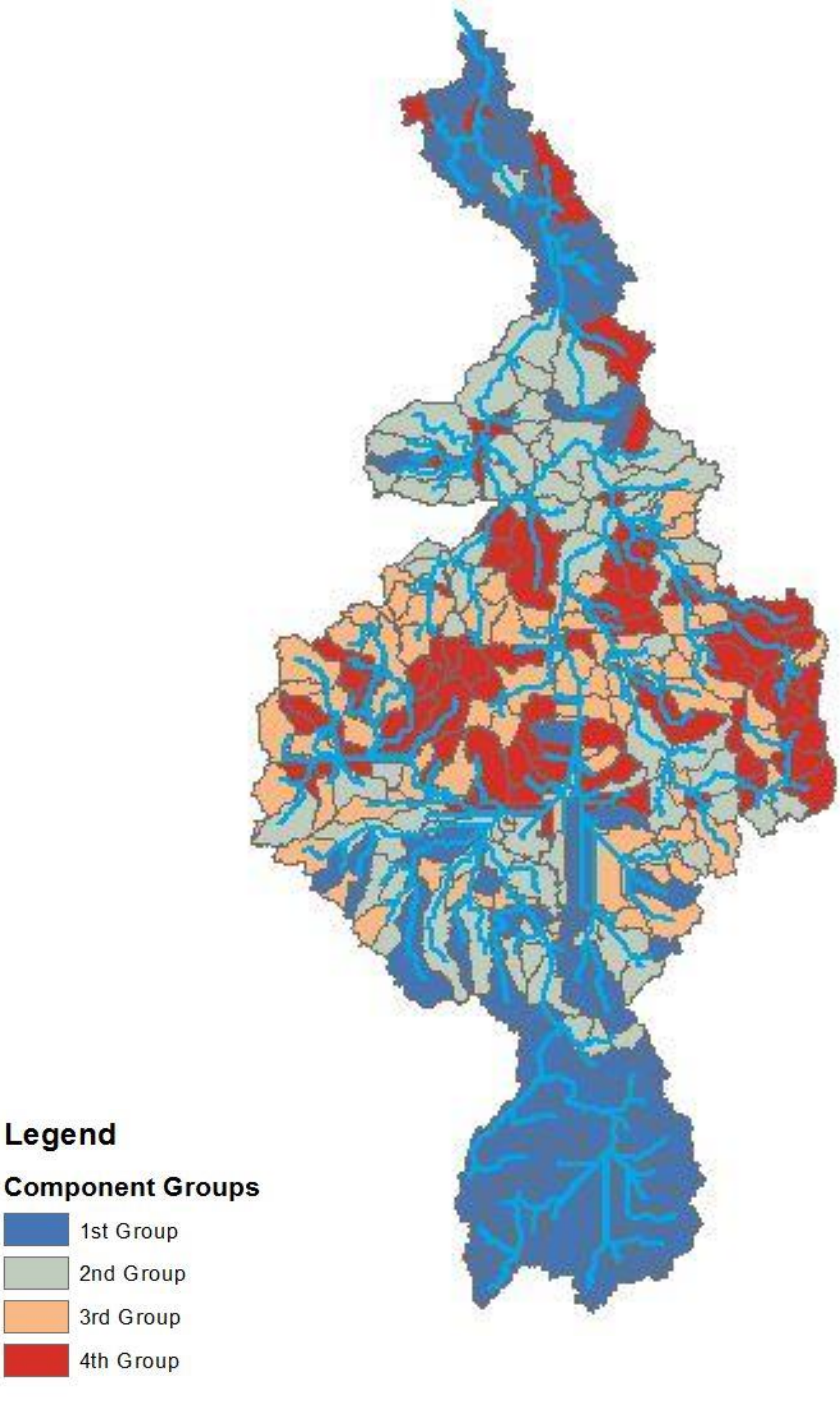
Nile, simulated monthly average stream-flow, 2001-2010.

Figure 2.13 – Horn's Test, Nile.



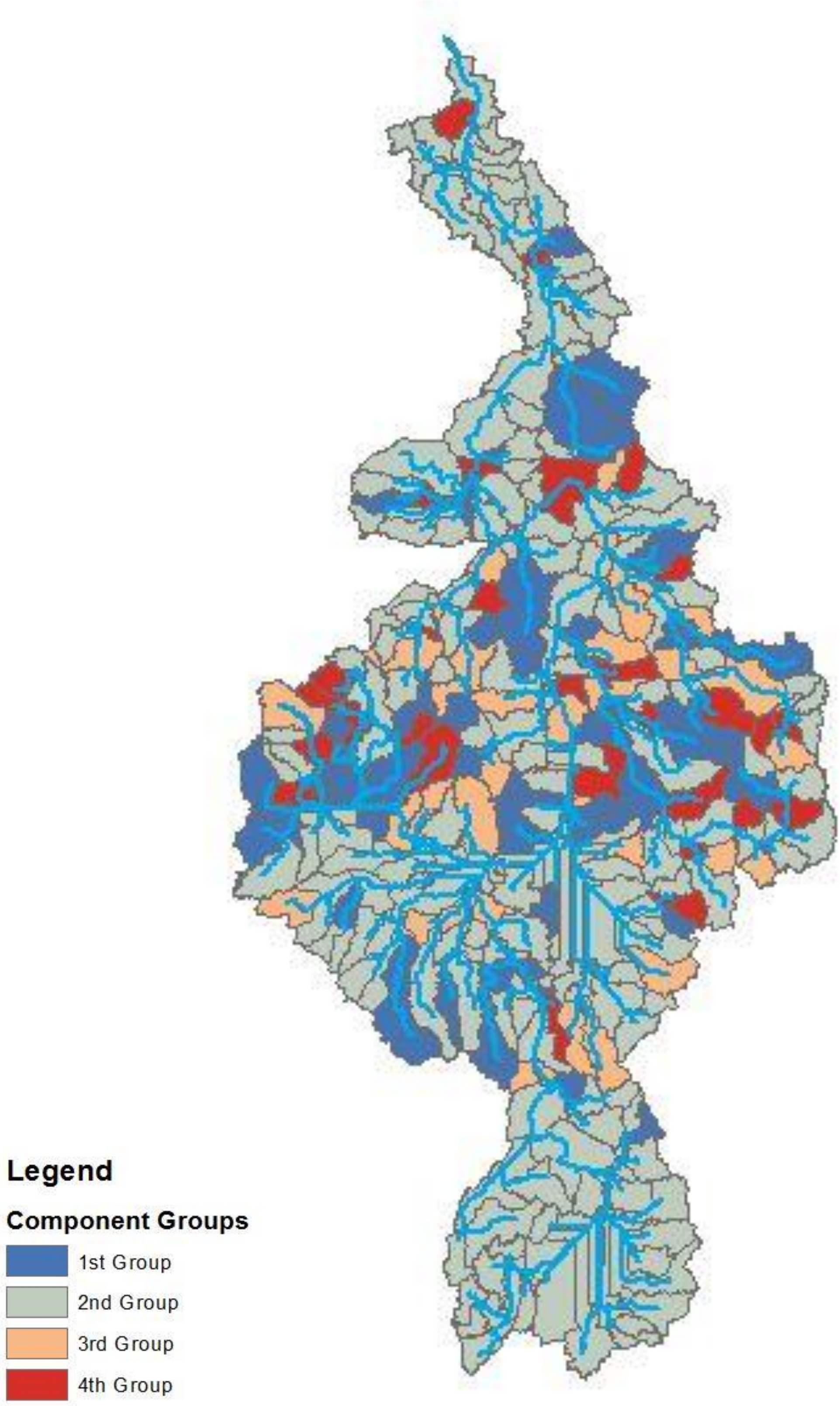
Nile, result of the Horn's test.

Figure 2.14 – First principal component, Nile.



Nile, grouping over the loadings of the first principal component.

Figure 2.15 – First principal component, oblimin rotation, Nile.

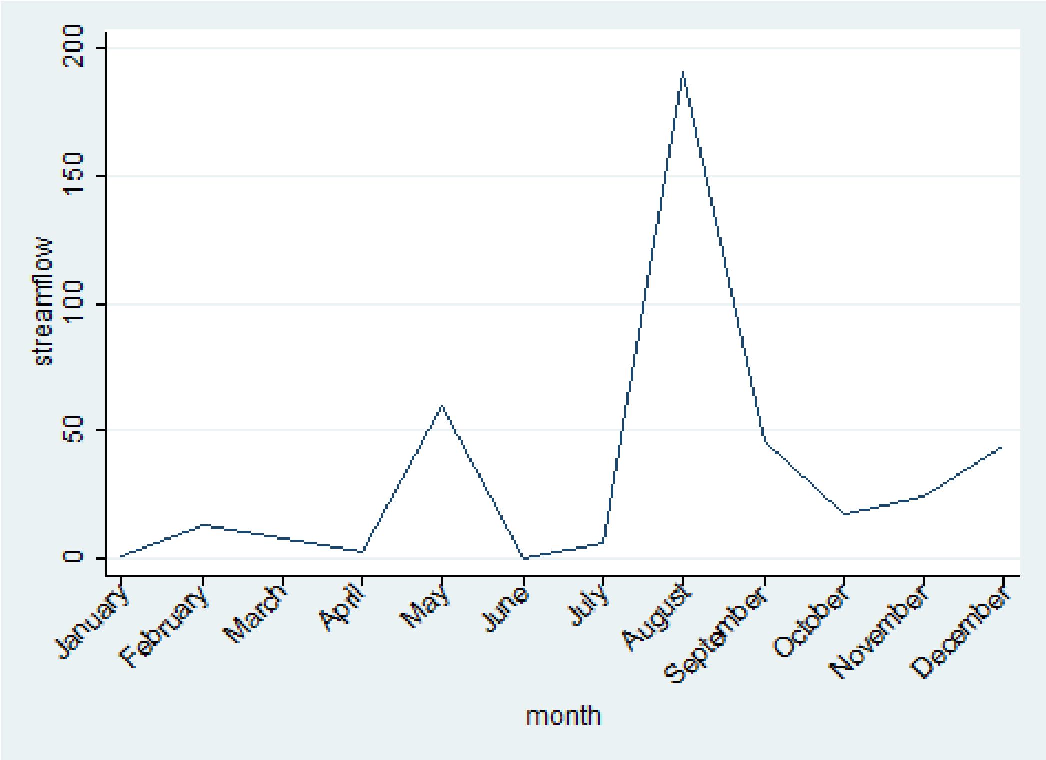


Nile, grouping over the loadings of the first principal component after oblimin rotation.

Figure 2.16 – Interbasin 3.

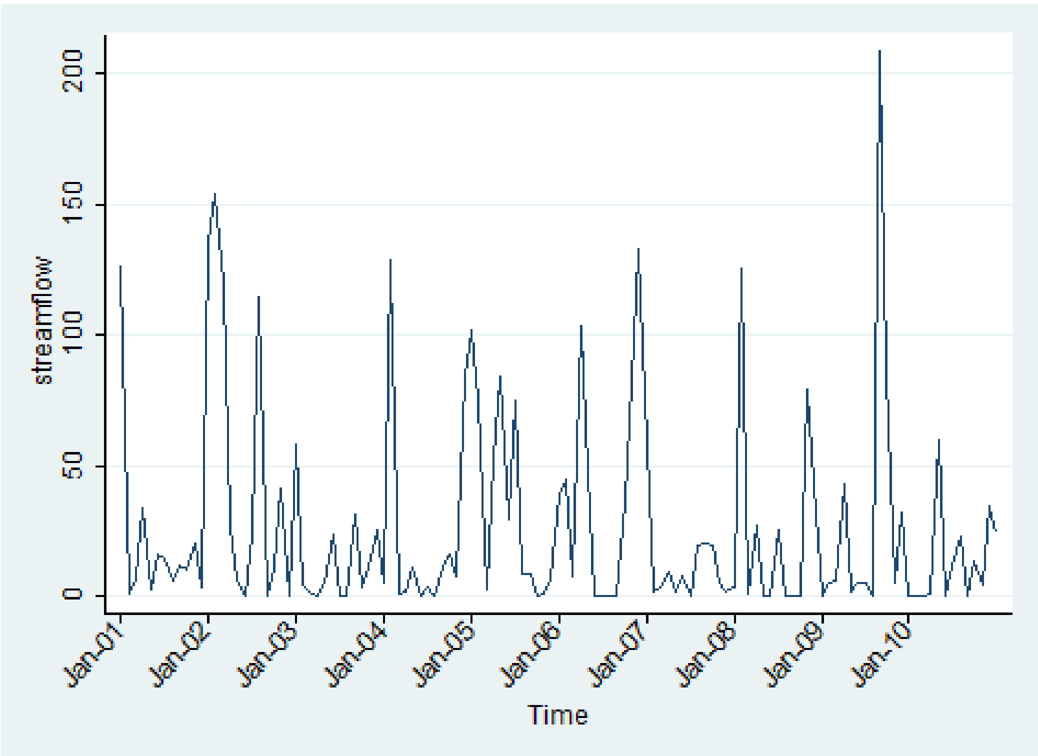
Interbasin 3, river network and average streamflow by quintile.

Figurate 2.17 – Monthly average streamflow, Interbasin 3.



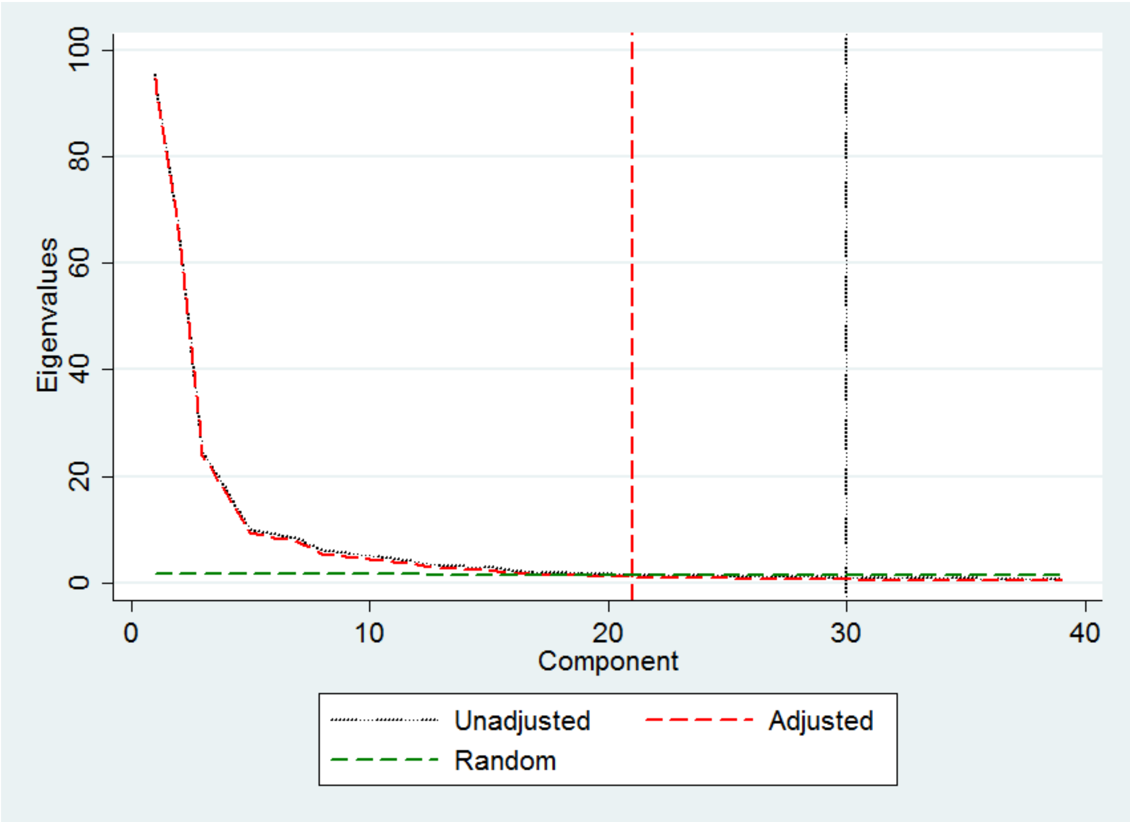
Interbasin 3 simulated monthly average stream-flow.

Figure 2.18 – Monthly streamflow 2001-2010, Interbasin 3.



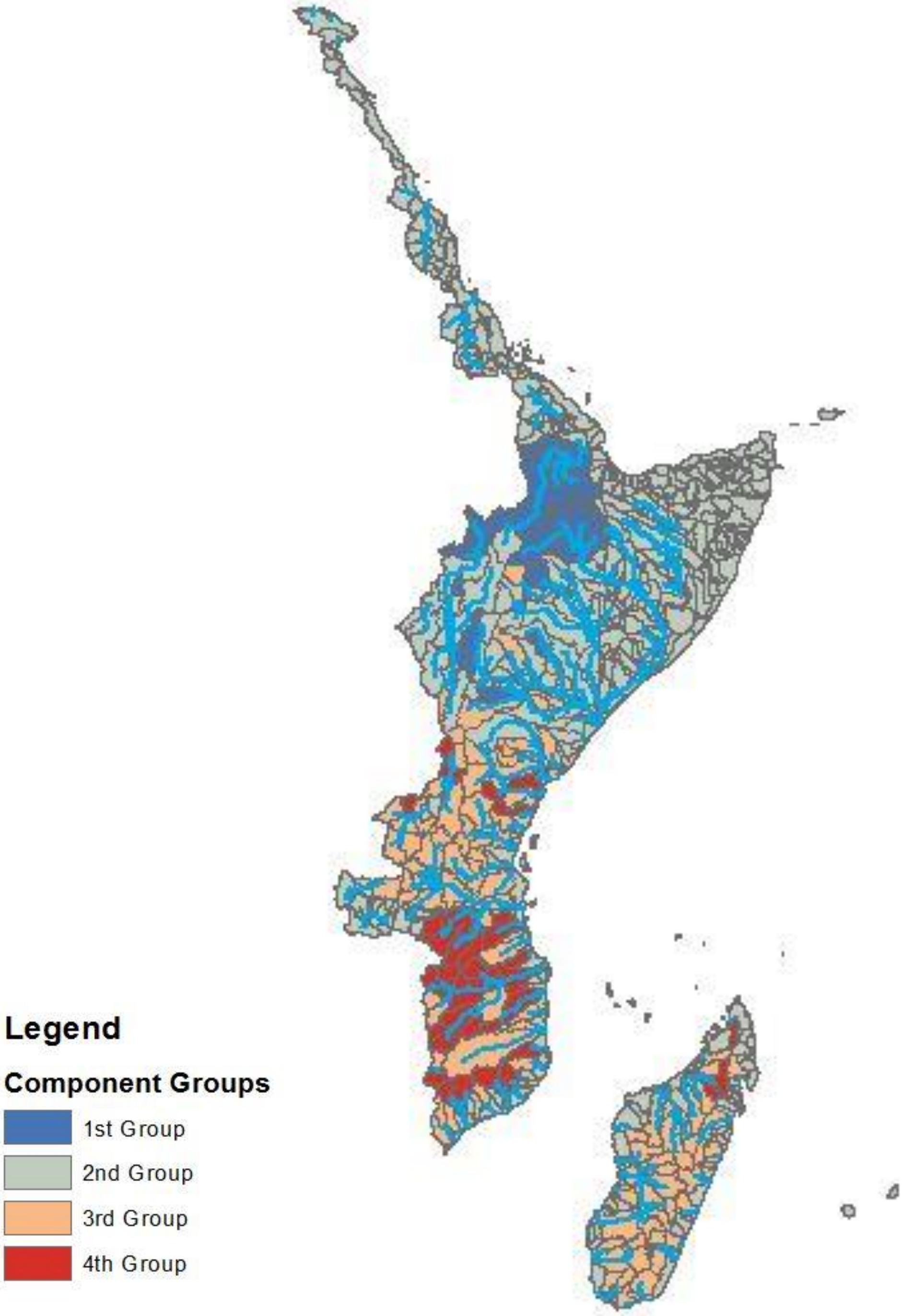
Interbasin 3 simulated monthly average stream-flow, 2001-2010.

Figure 2.19 – Horn's Test, Interbasin 3.



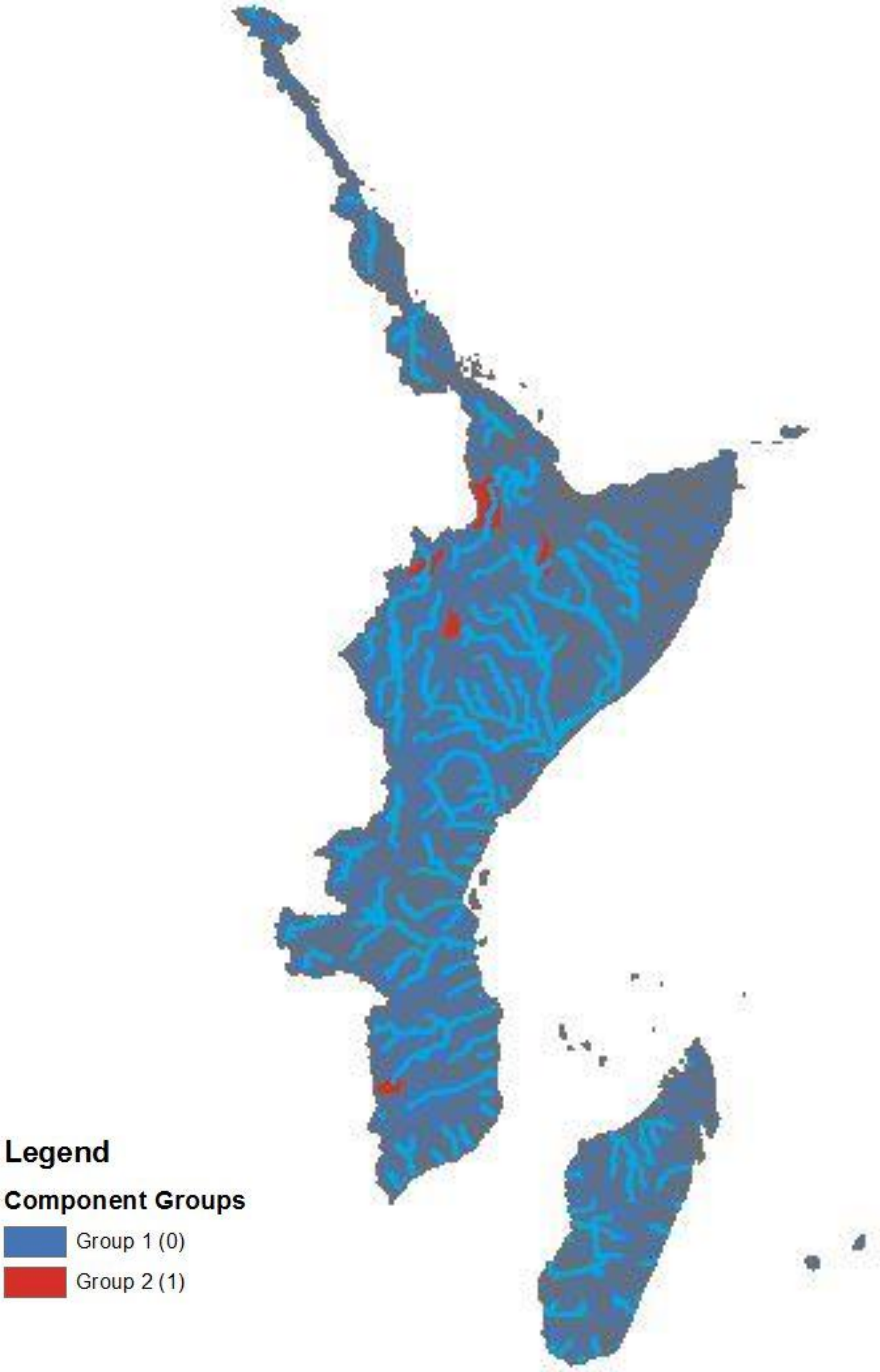
Interbasin 3, result of the Horn's test.

Figure 2.20 – First principal component, Interbasin 3.



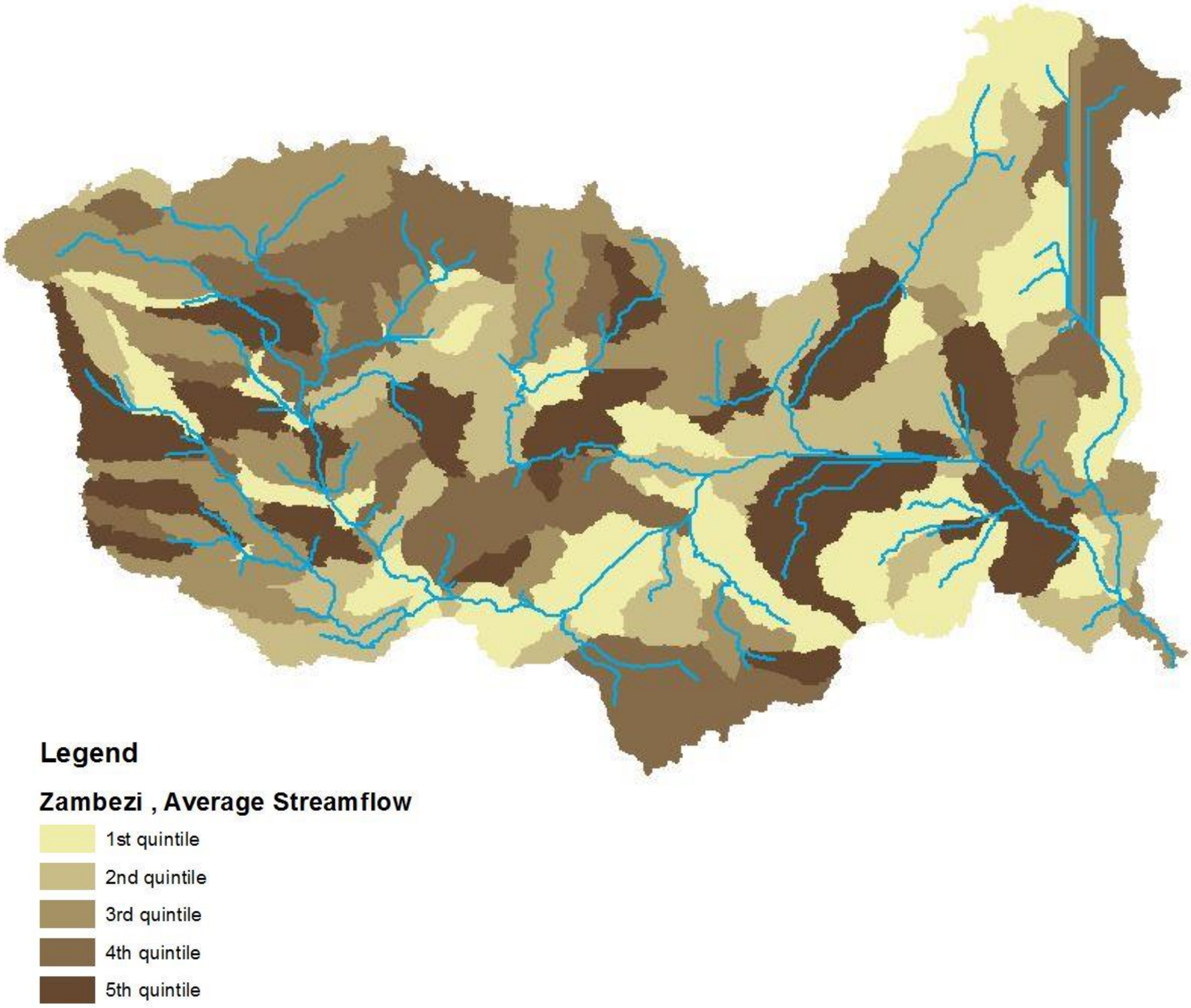
Interbasin 3, grouping over the loadings of the first principal component.

Figure 2.21 – First principal component, oblimin rotation, Interbasin 3



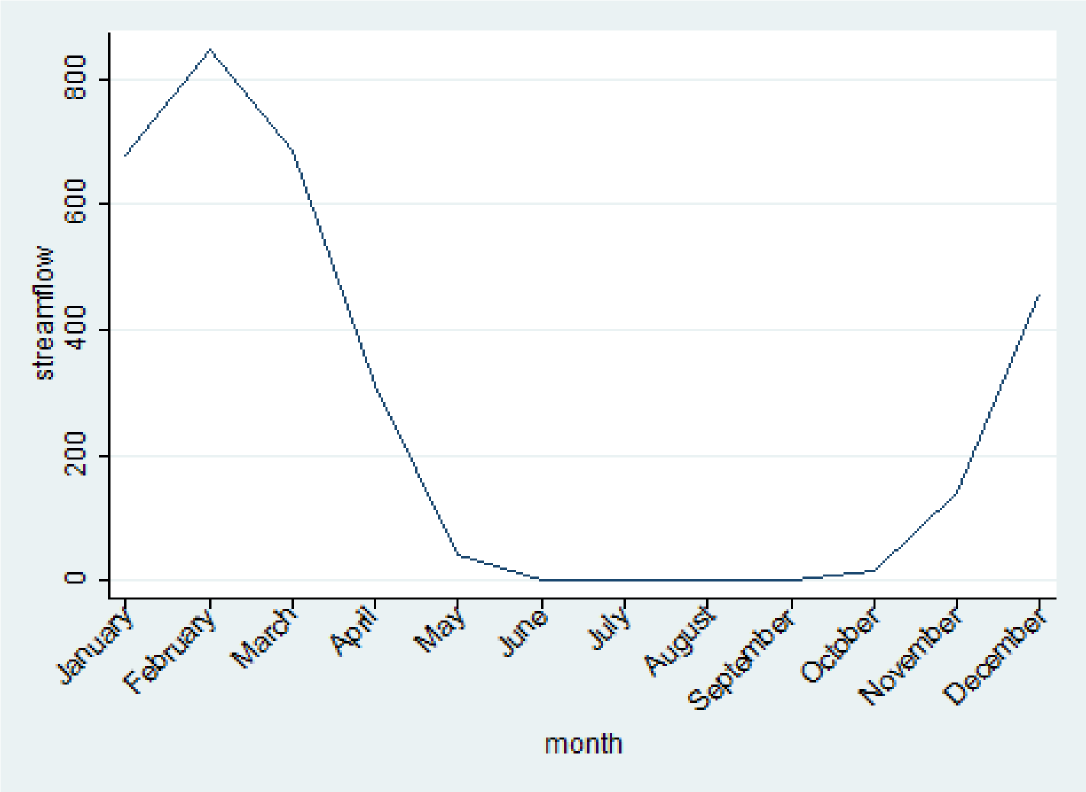
Interbasin 3, grouping over the loadings of the first principal component after oblimin rotation.

Figure 2.22 – Zambezi.



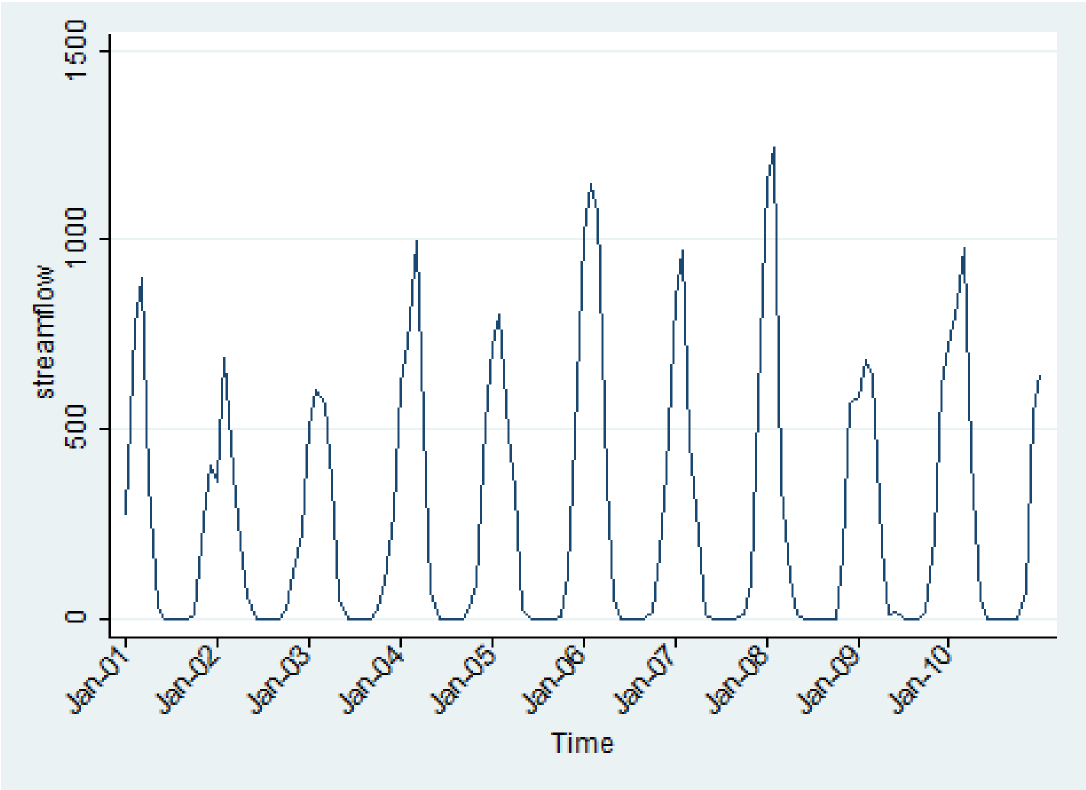
Zambezi, river network and average streamflow by quintile.

Figure 2.23 – Monthly average streamflow, Zambezi.



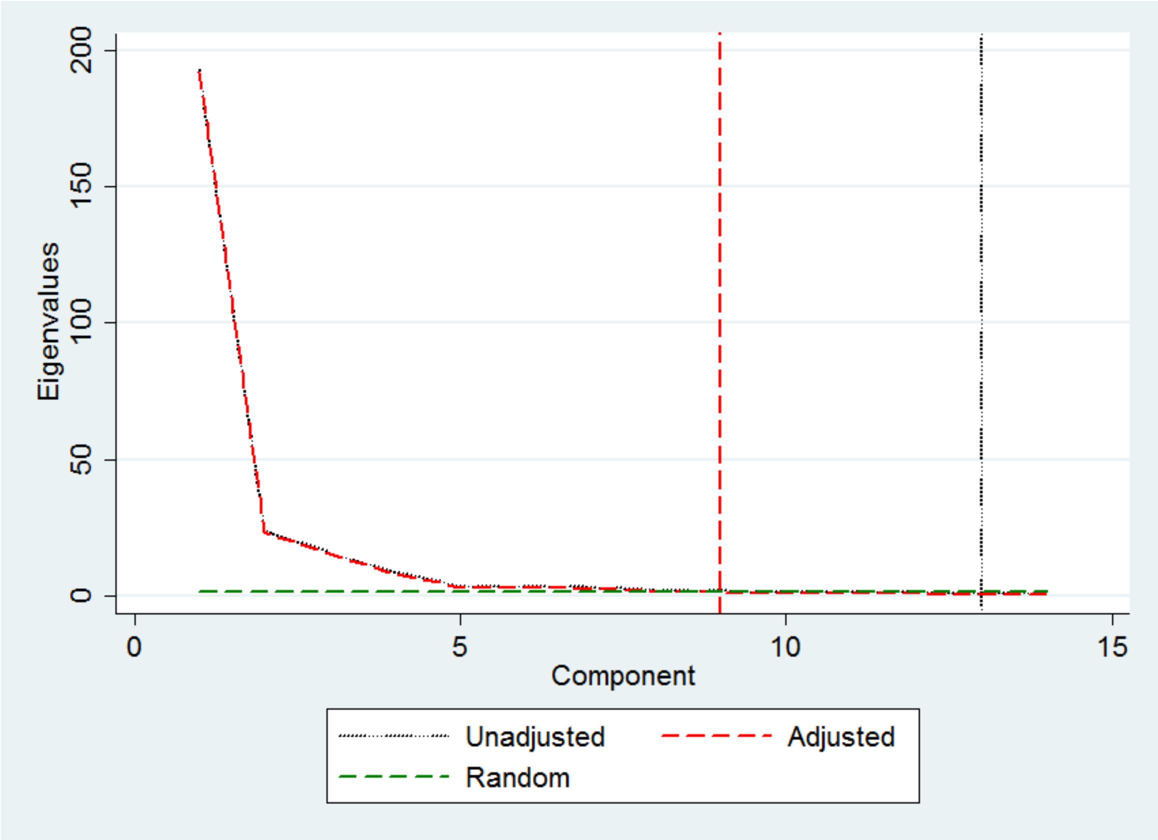
Zambezi, simulated monthly average stream-flow.

Figure 2.24 –Monthly streamflow 2001-2010, Zambezi.



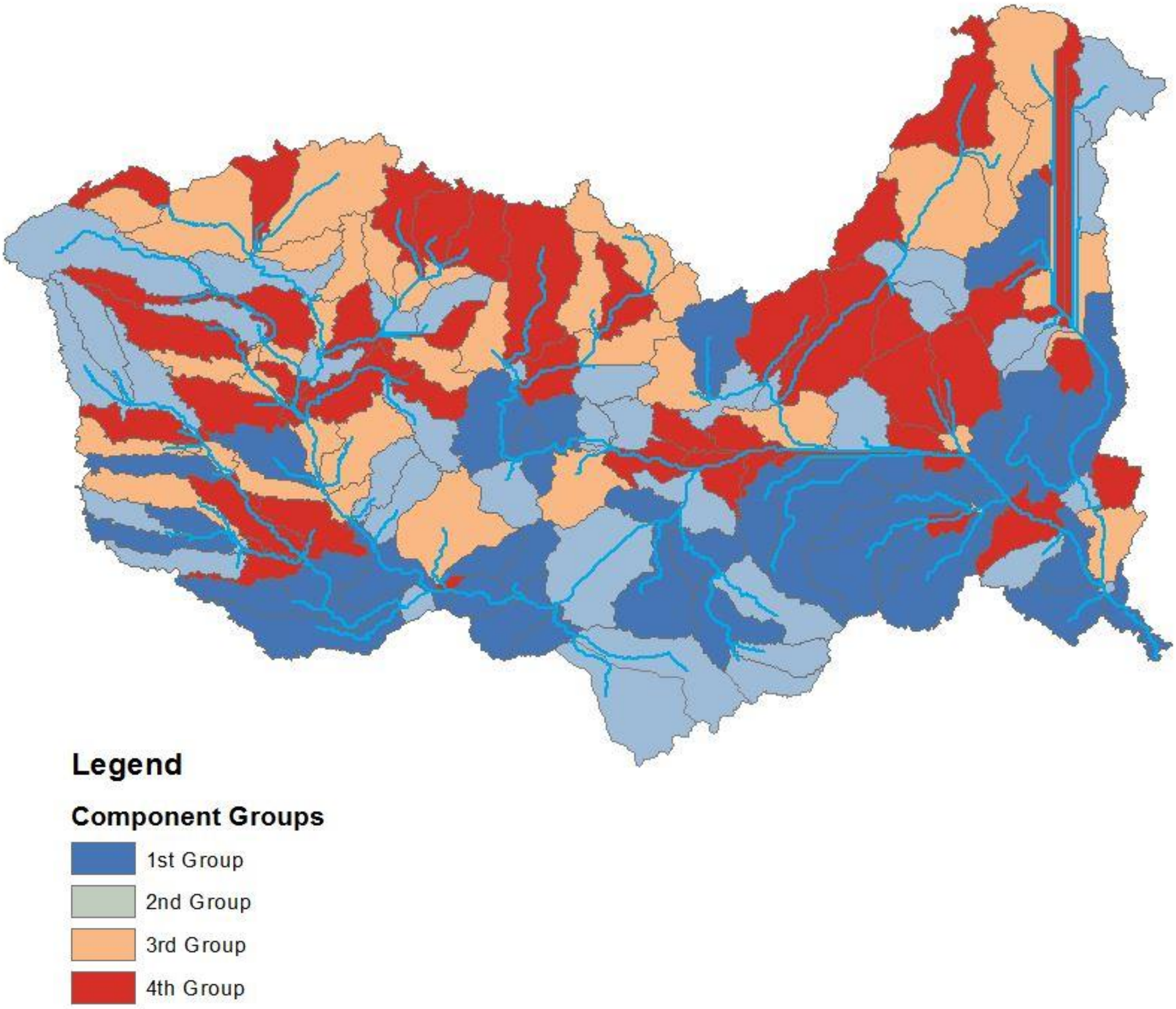
Zambezi, simulated monthly average stream-flow, 2001-2010.

Figure 2.25 – Horn's Test, Zambezi.



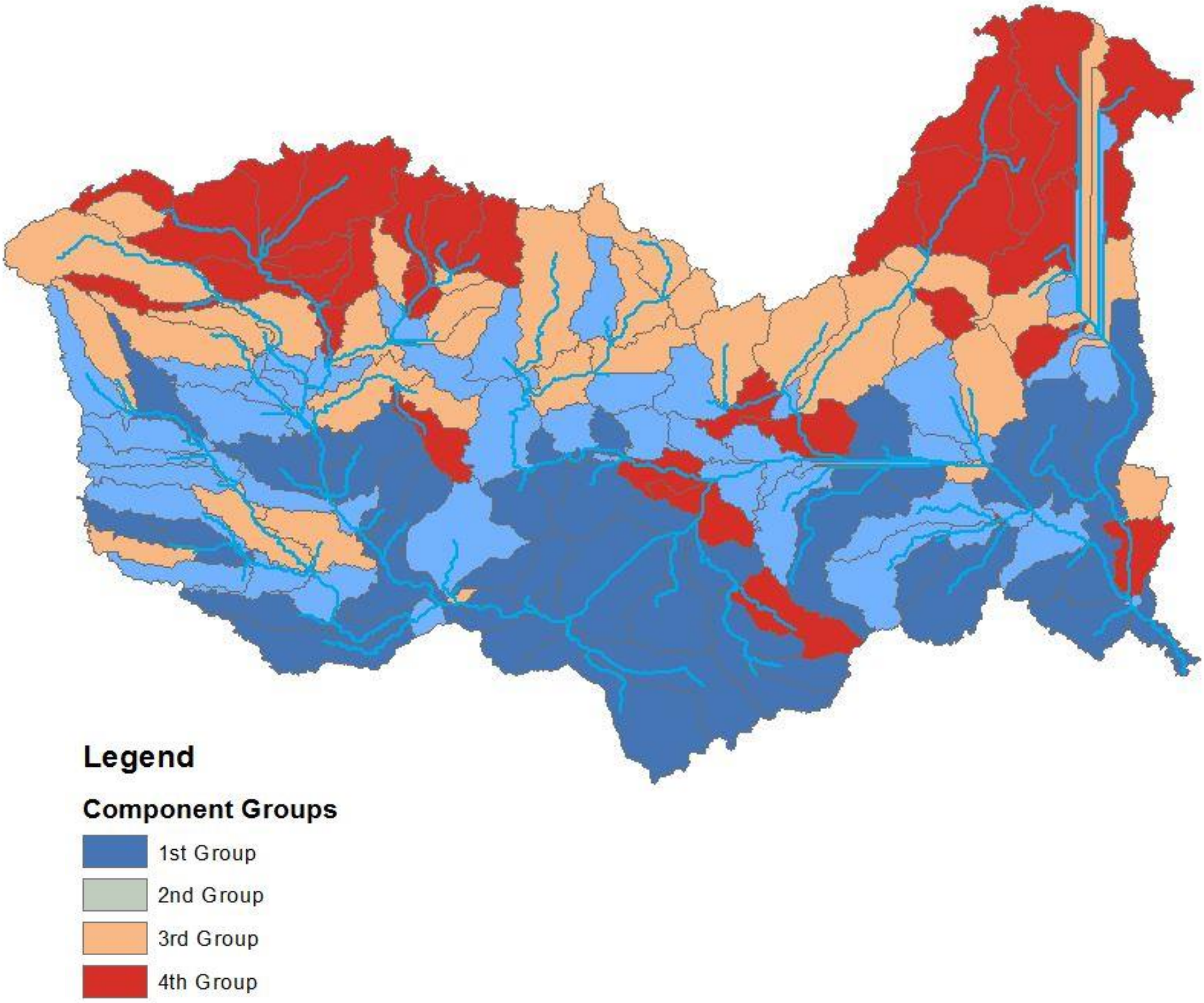
Zambezi, result of the Horn's test.

Figure 2.26 – First principal component, Zambezi.



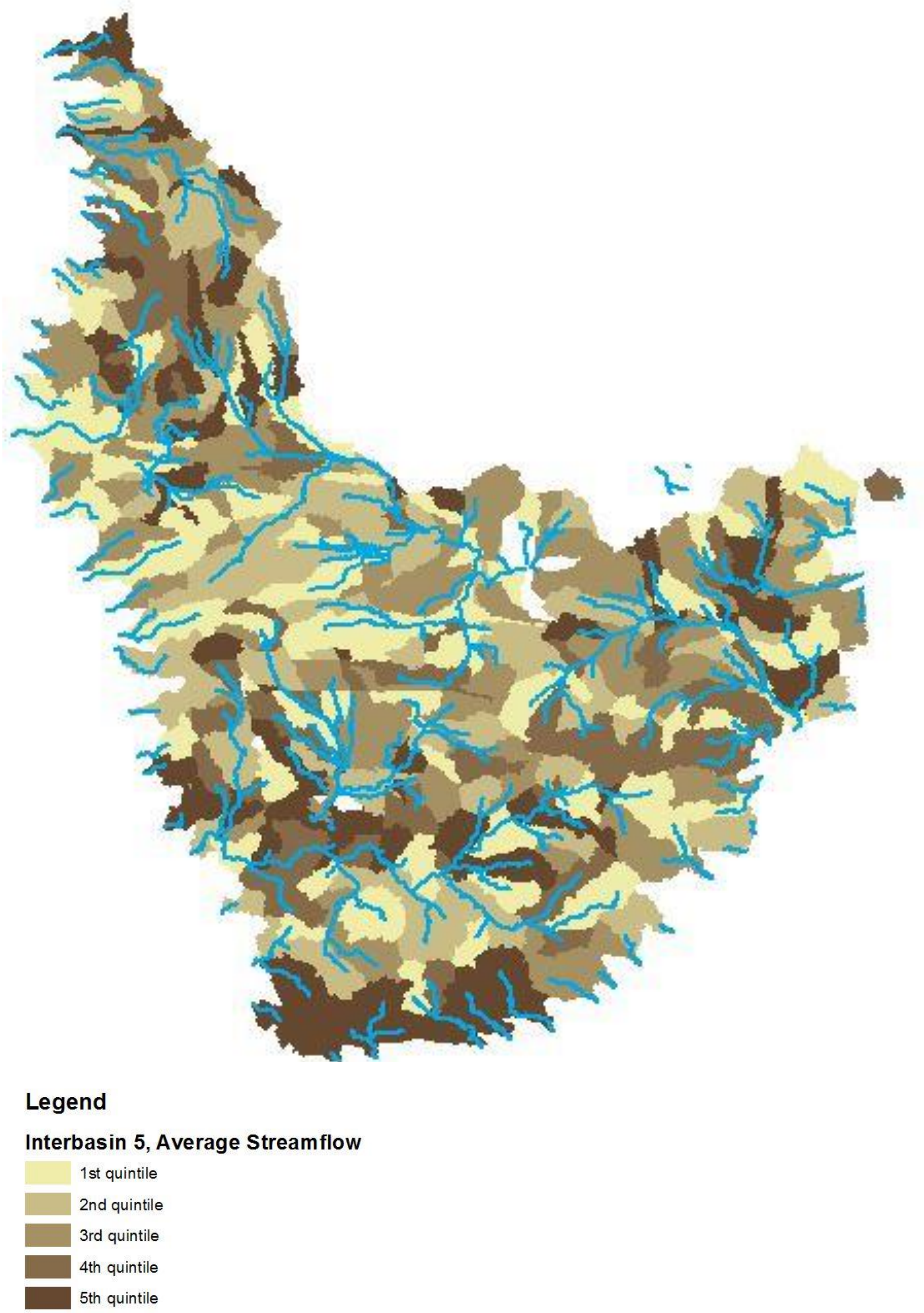
Zambezi, grouping over the loadings of the first principal component.

Figure 2.27 – First principal component, oblimin rotation, Zambezi.



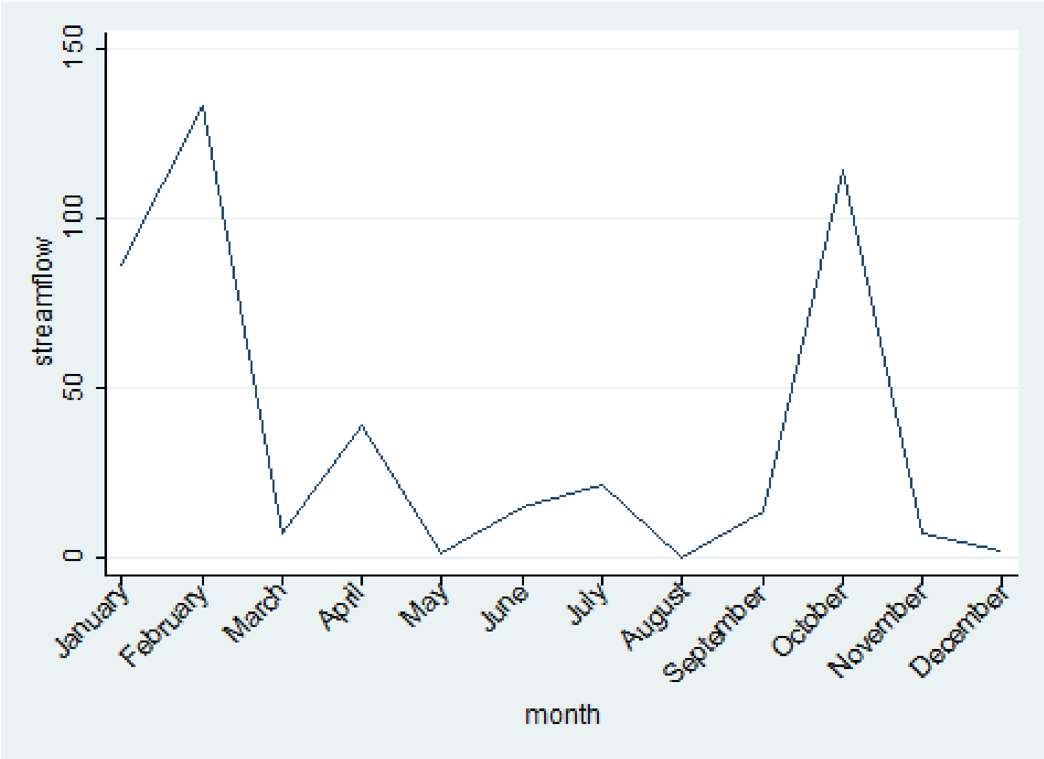
Zambezi, grouping over the loadings of the first principal component after oblimin rotation

Figure 2.28 – Interbasin 5.



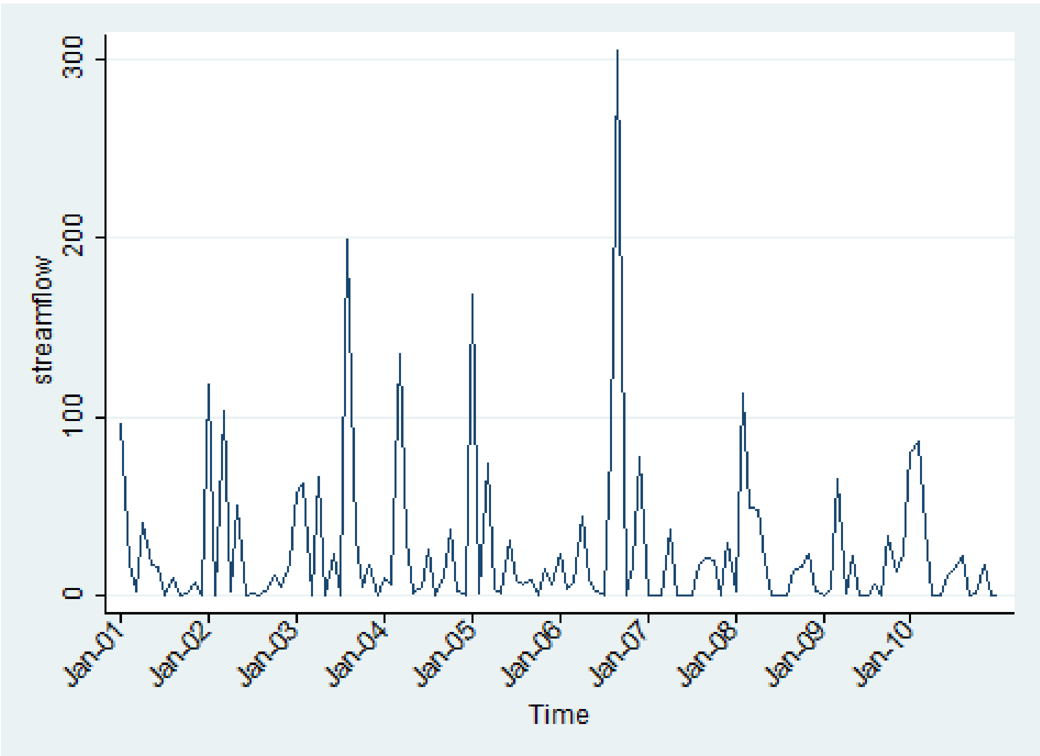
Interbasin 5, river network and average streamflow by quintile.

Figure 2.29 – Monthly average streamflow, Interbasin 5.



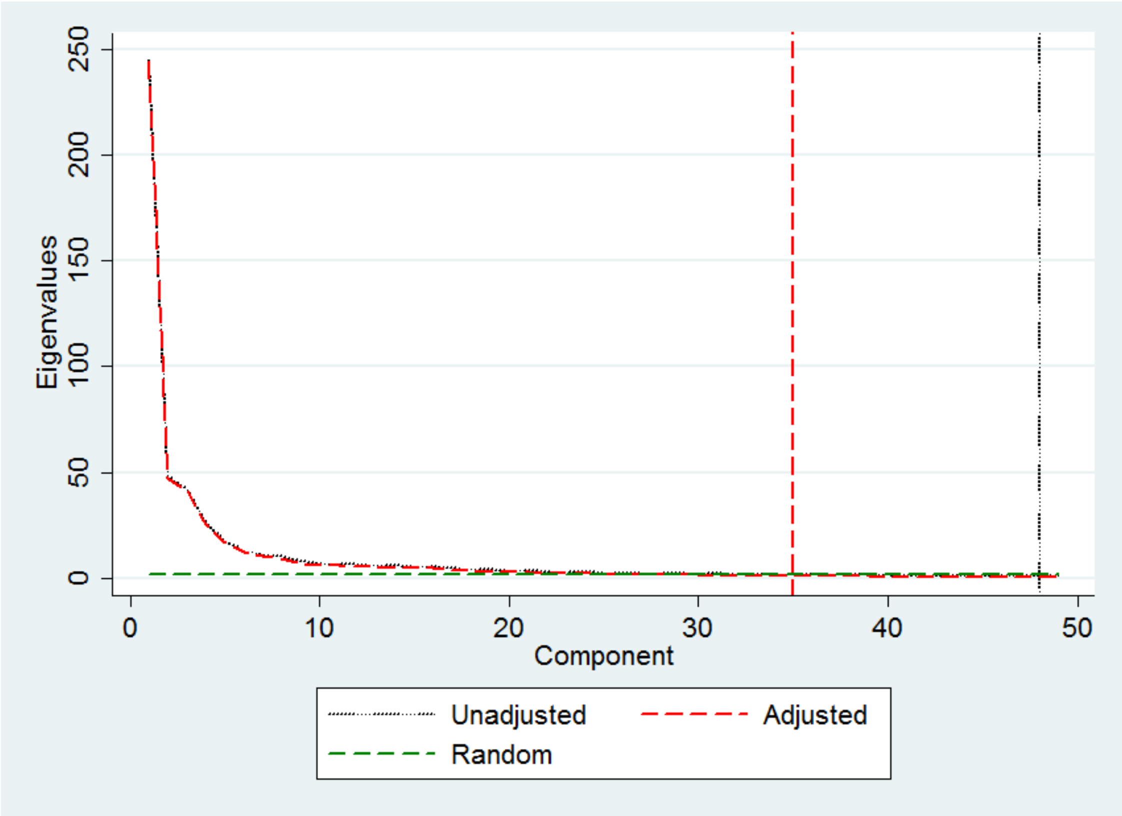
Interbasin 5, simulated monthly average stream-flow.

Figure 2.30 – Monthly streamflow 2001-2010, Interbasin 5.



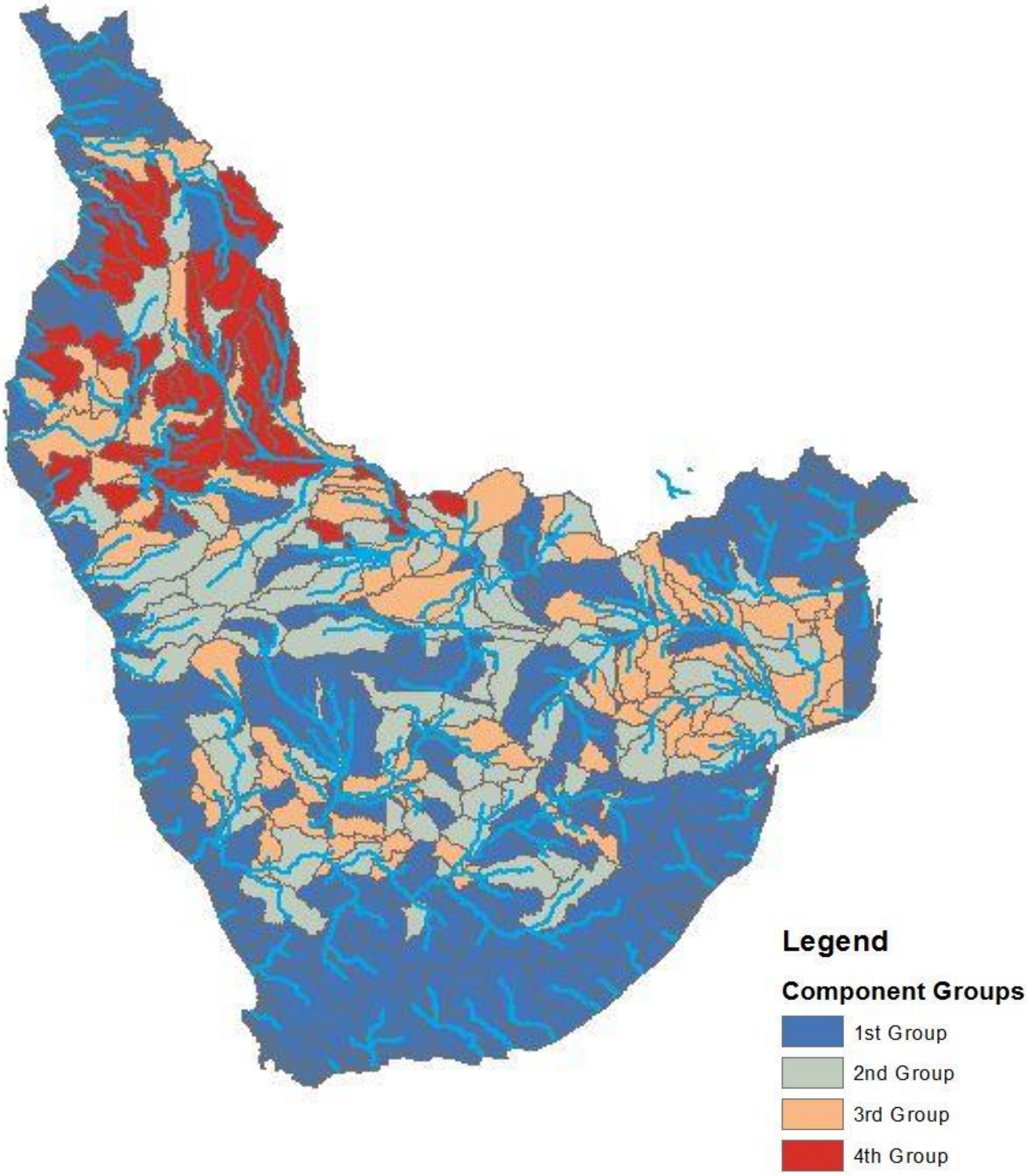
Interbasin 5 simulated monthly average stream-flow, 2001-2010.

Figure 2.31 – Horn's Test, Interbasin 5.



Interbasin 5, result of the Horn's test.

Figure 2.32 – First principal component, Interbasin 5.



Interbasin 5, grouping over the loadings of the first principal component.

Figure 2.33 – First principal component, oblimin rotation, Interbasin 5.

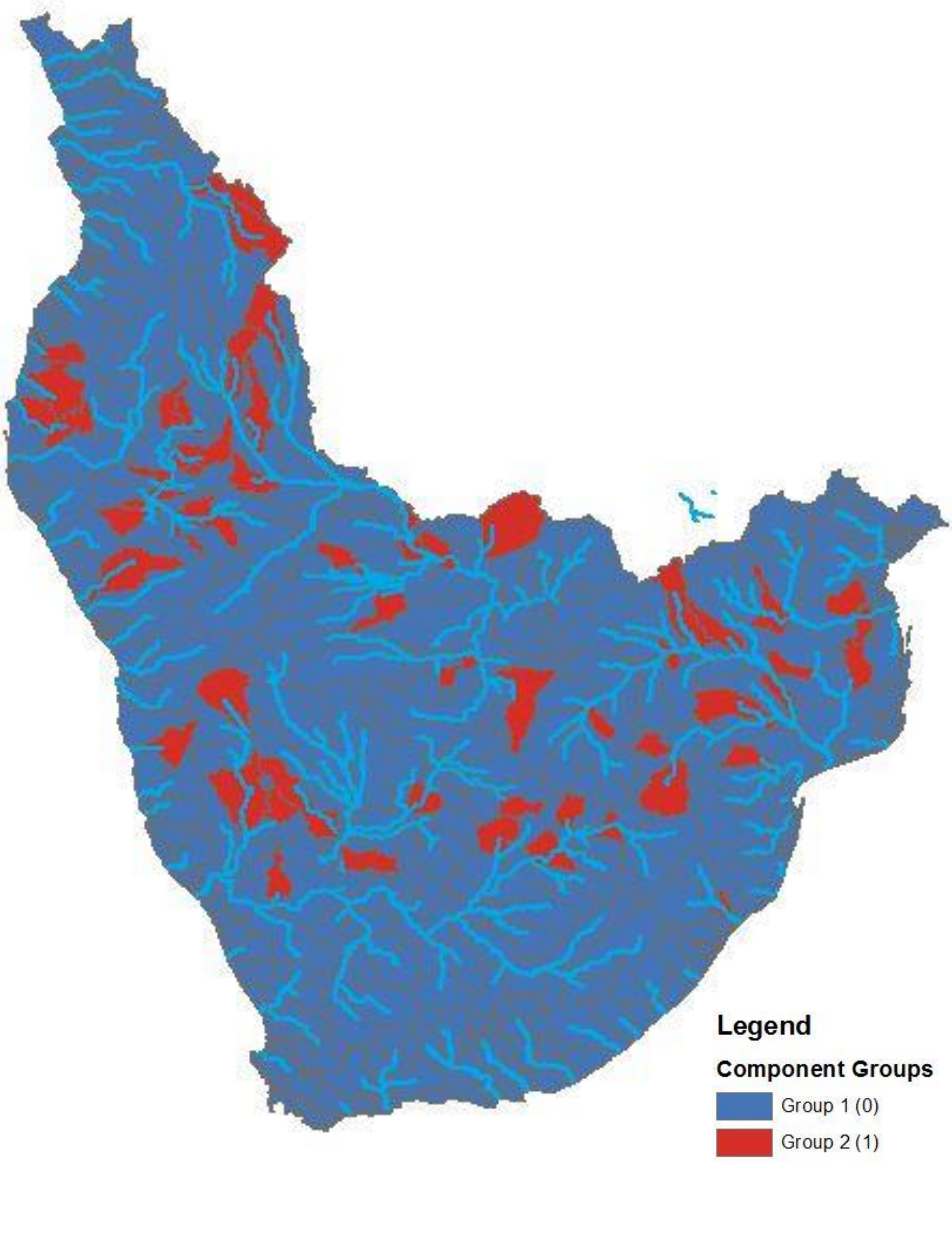
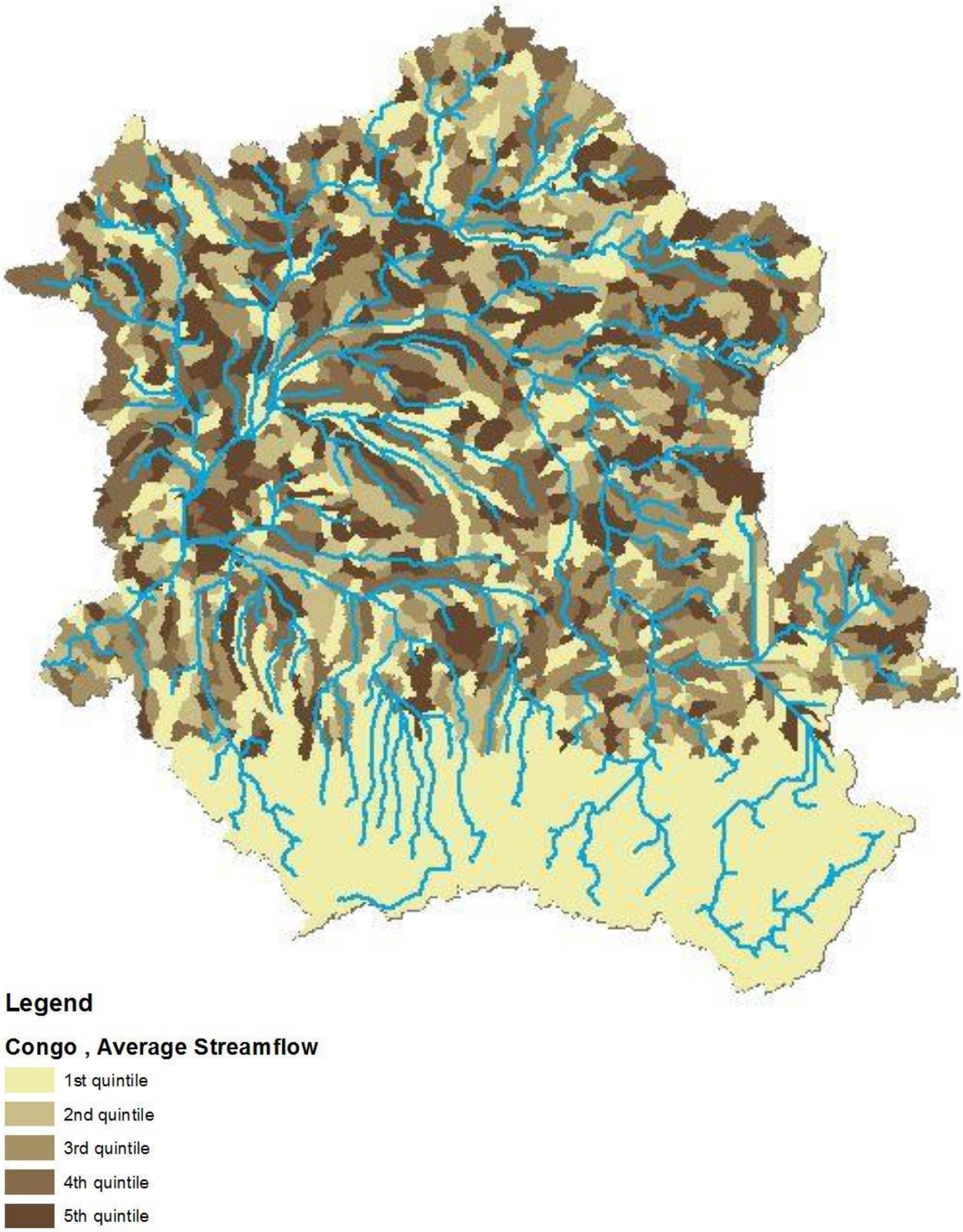


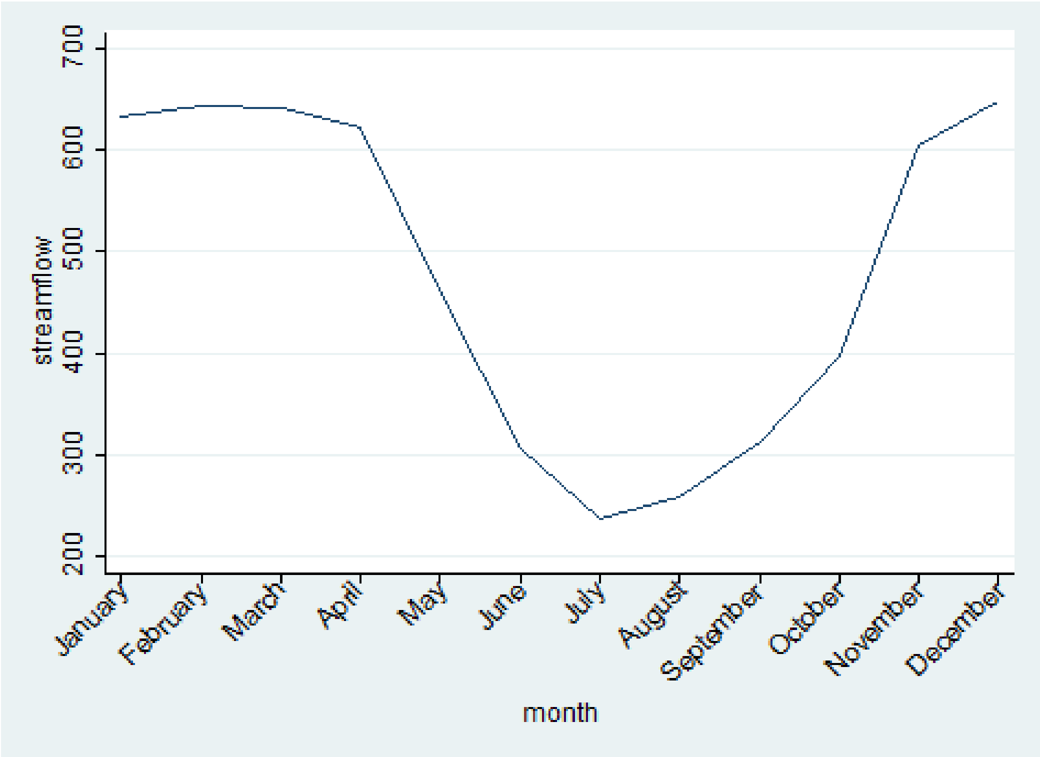
Figure 33. Interbasin 5, grouping over the loadings of the first principal component after oblimin rotation.

Figure 2.34 – Congo.



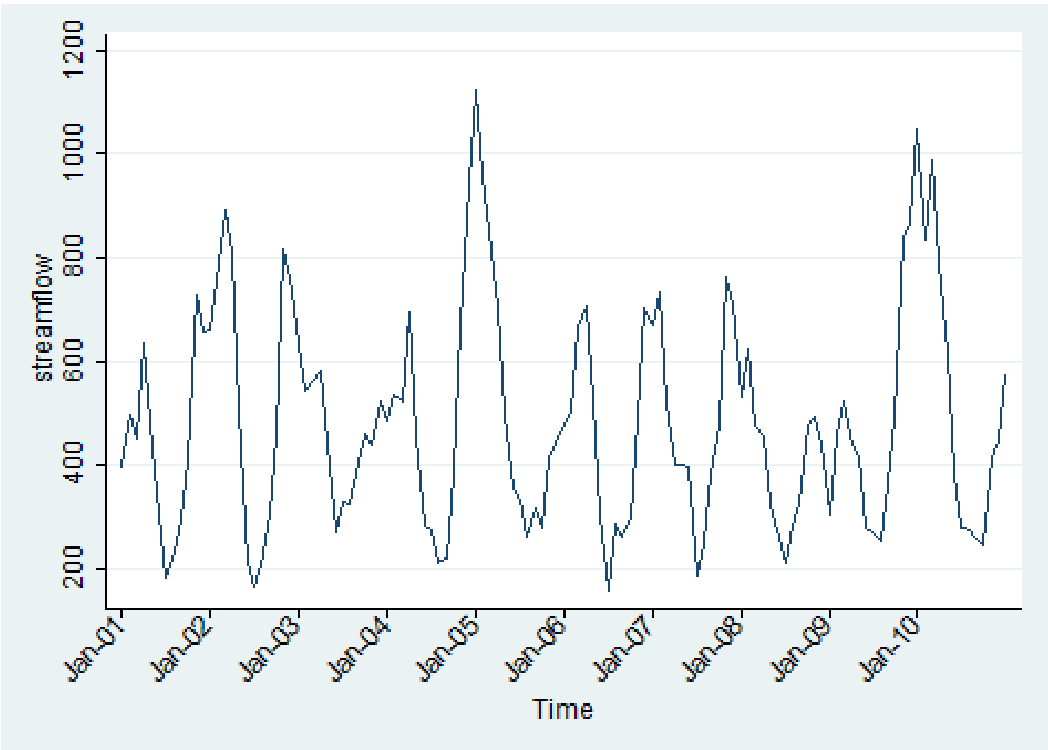
Congo, river network and average streamflow by quintile.

Figure 2.35 – Monthly average streamflow, Congo.



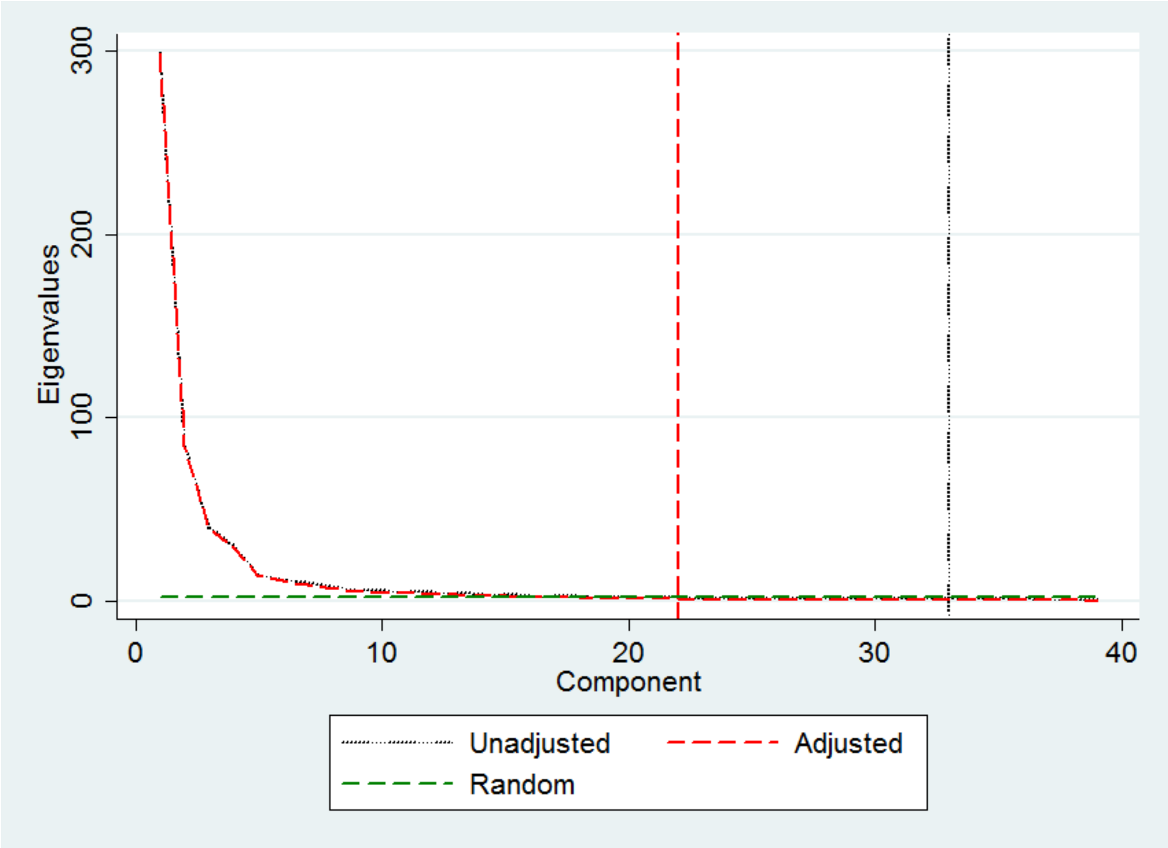
Congo, simulated monthly average stream-flow.

Figure 2.36 –Monthly streamflow 2001-2010, Congo.



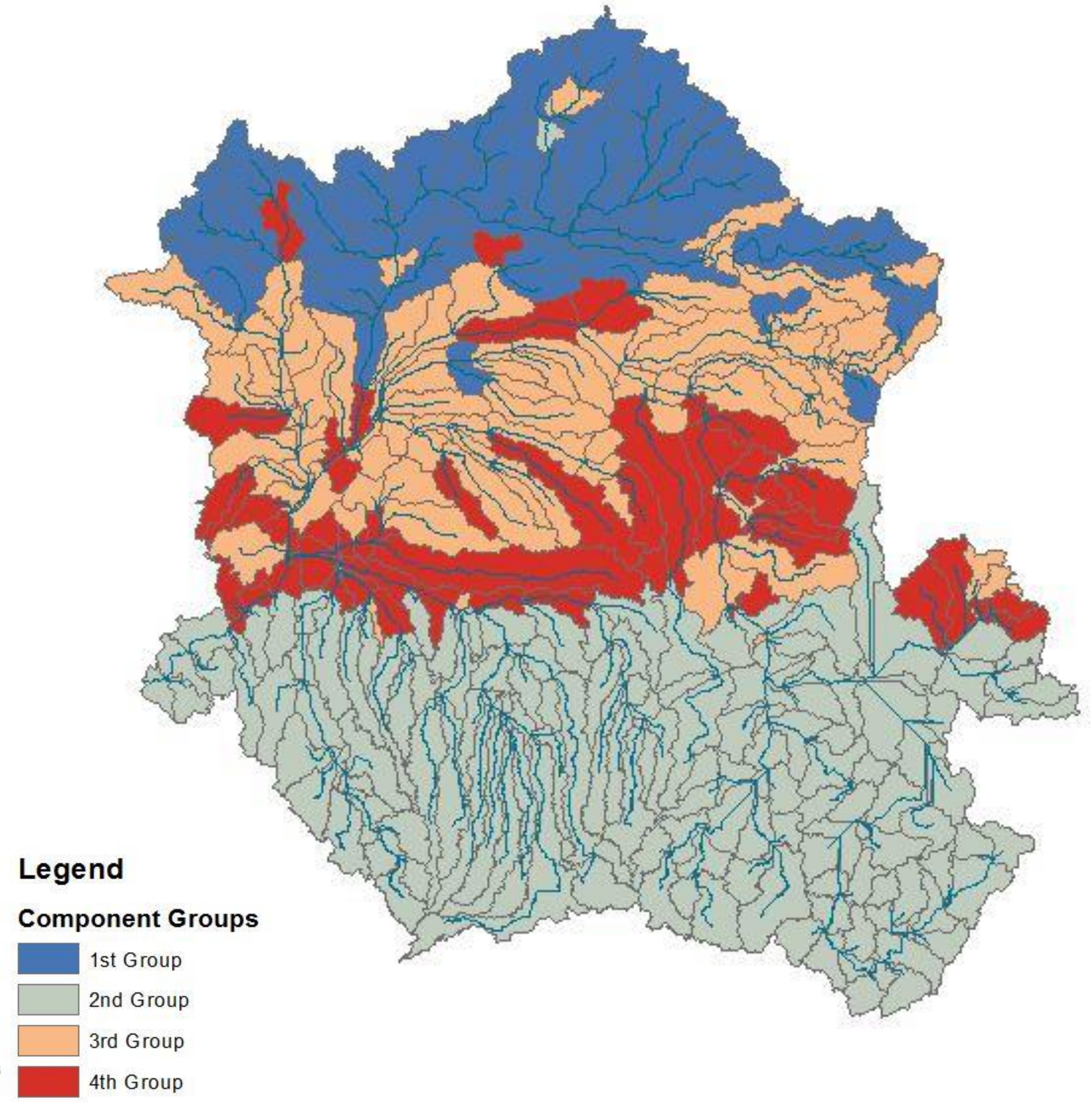
Congo, simulated monthly average stream-flow, 2001-2010.

Figure 2.37 – Horn's Test, Congo.



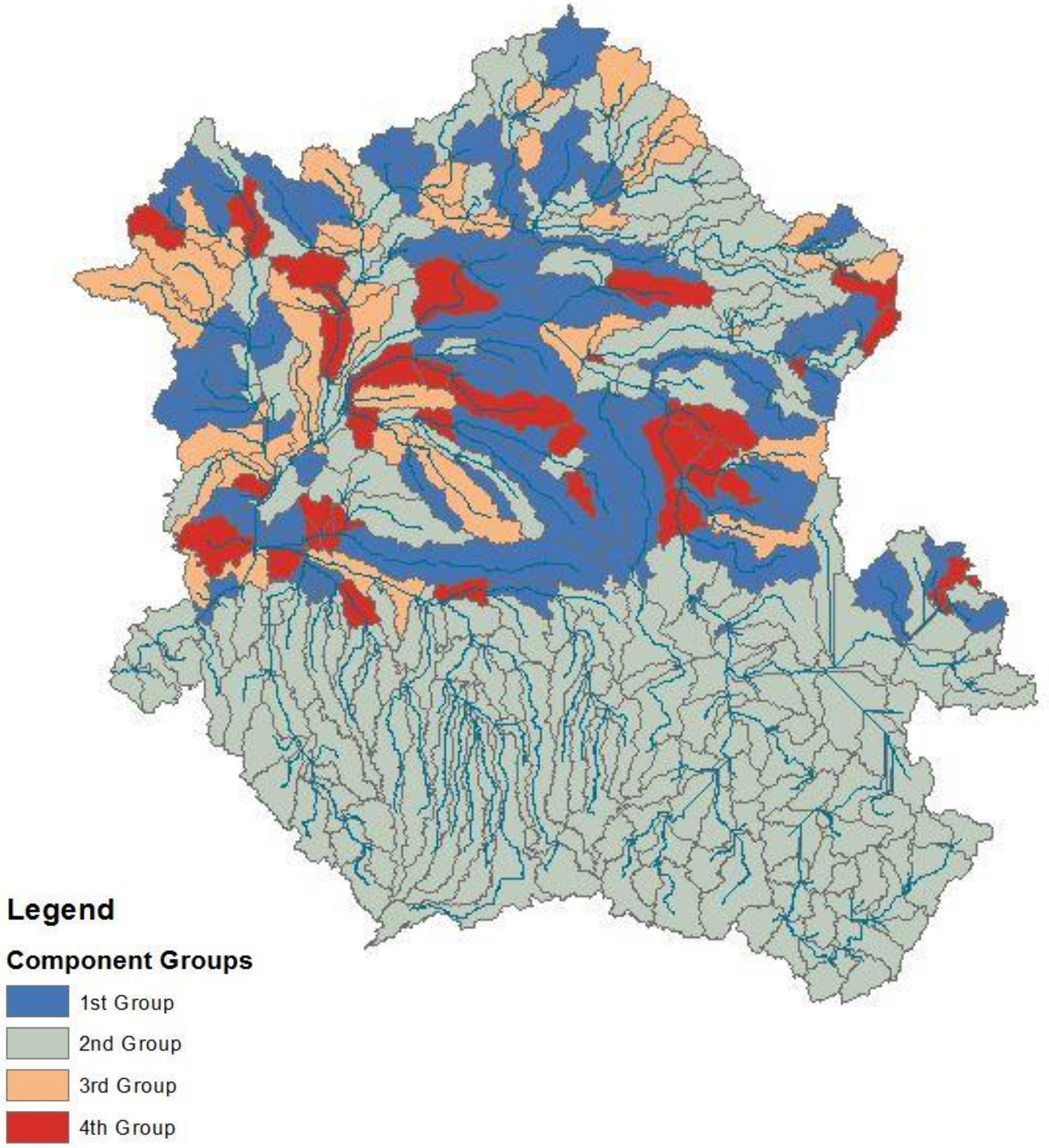
Congo, result of the Horn's test.

Figure 2.38 – First principal component, Congo.



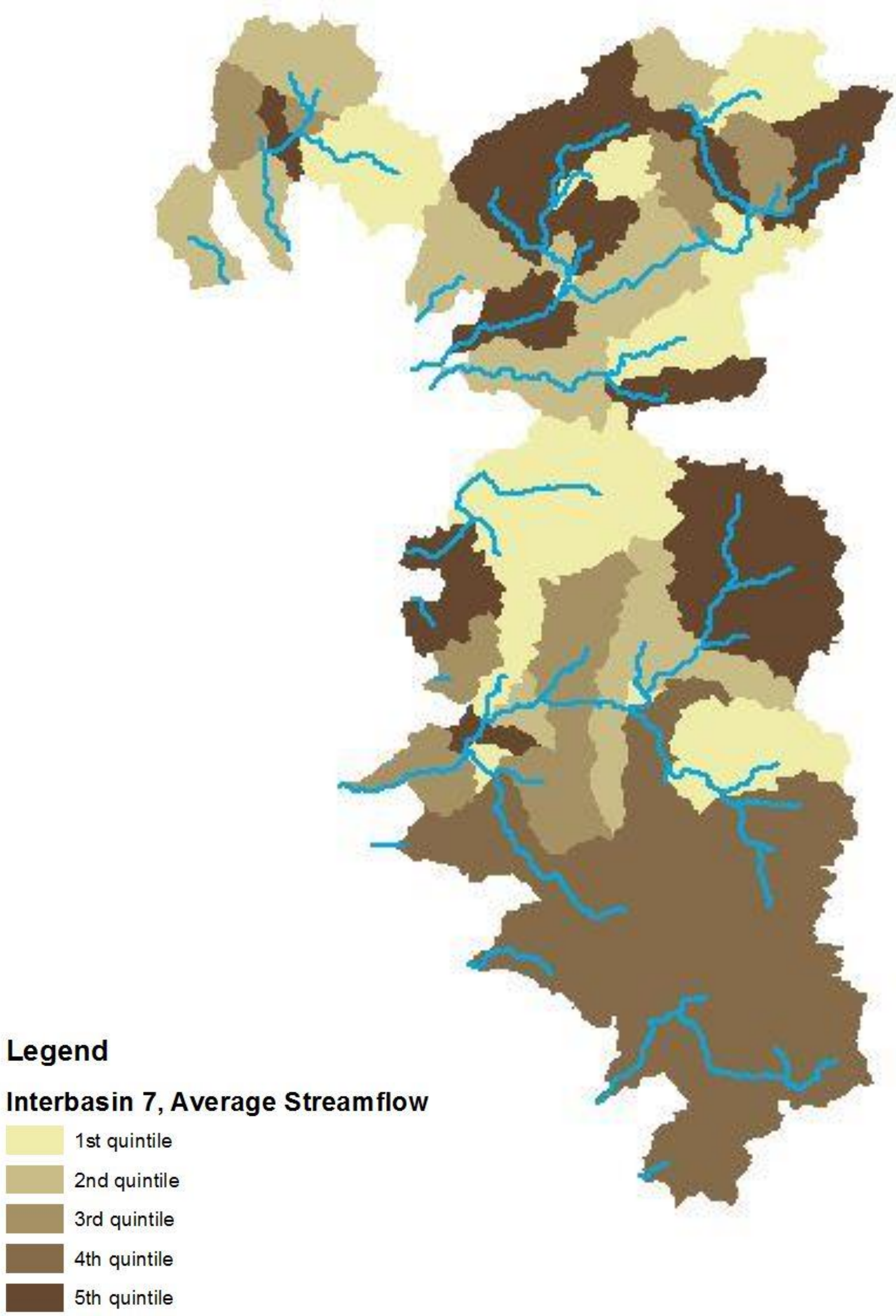
Congo, grouping over the loadings of the first principal component.

Figure 2.39 – First principal component, oblimin rotation, Congo.



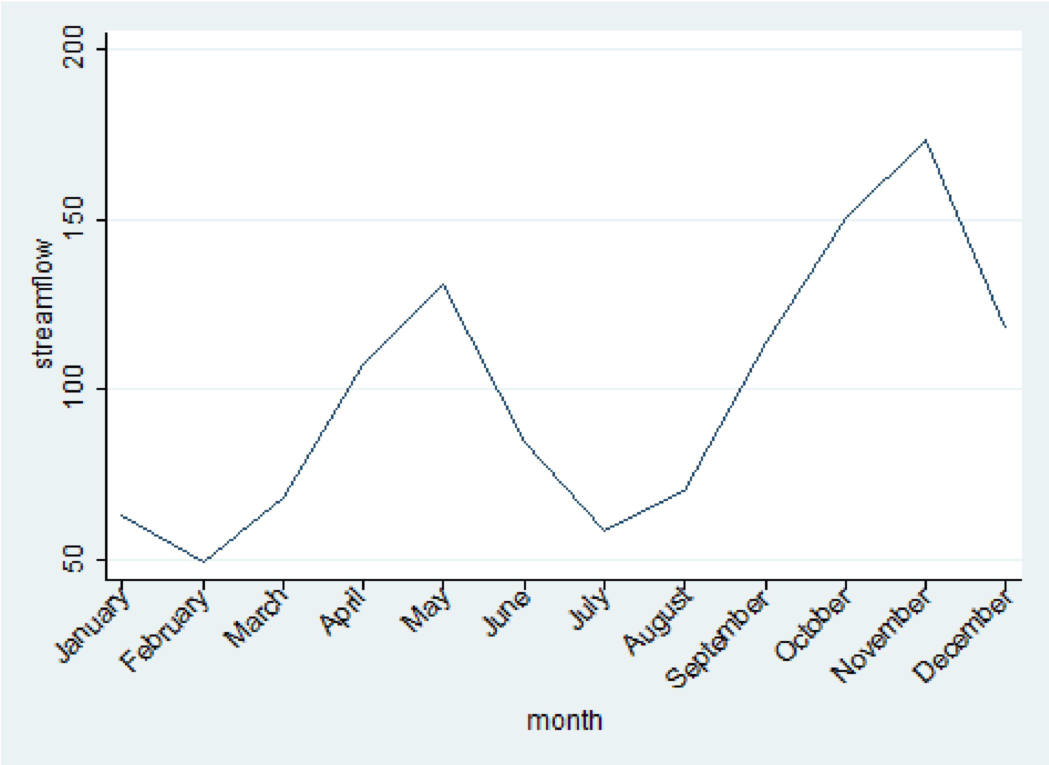
Congo, grouping over the loadings of the first principal component after oblimin rotation.

Figure 2.40 – Interbasin 7.



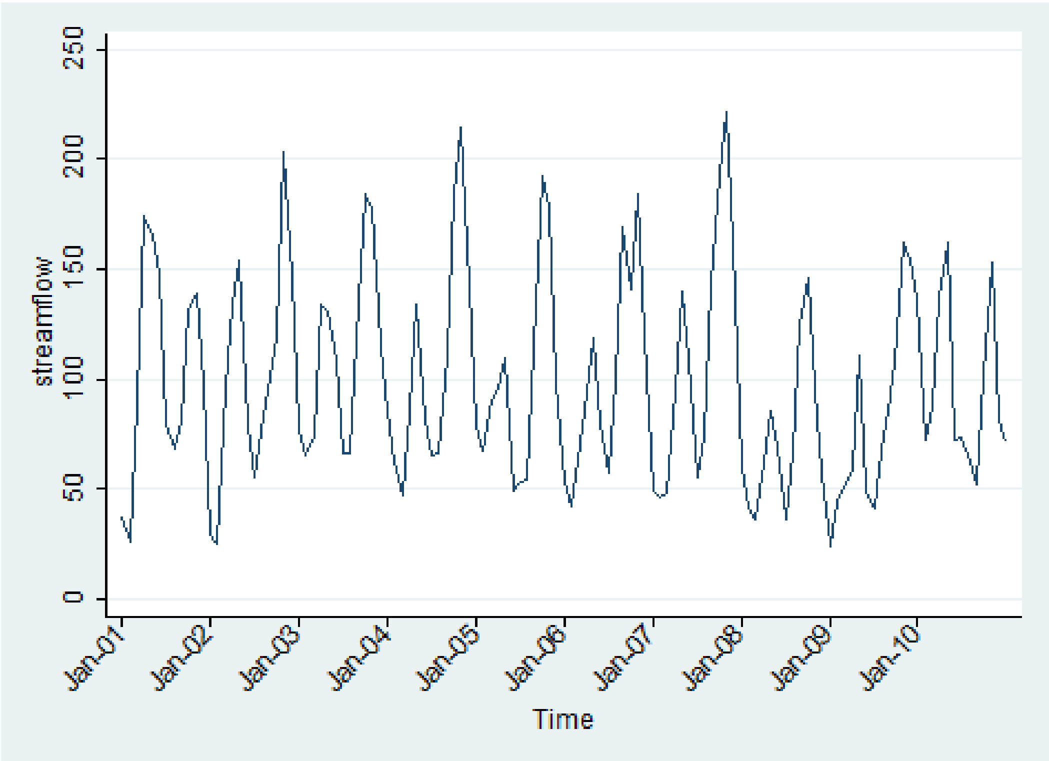
Interbasin 7, river network and average streamflow by quintile.

Figure 2.41 – Monthly average streamflow, Interbasin 7.



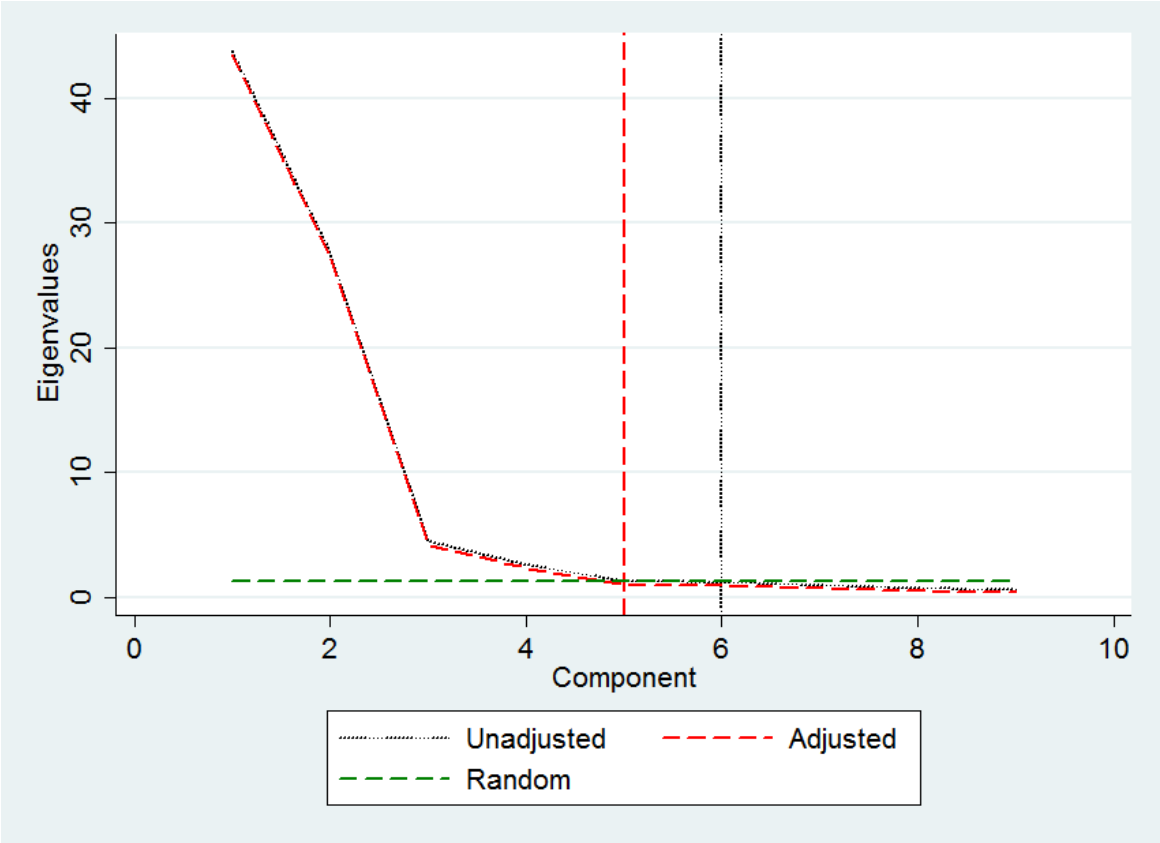
Interbasin 7 simulated monthly average stream-flow.

Figure 2.42 – Monthly streamflow 2001-2010, Interbasin 7.



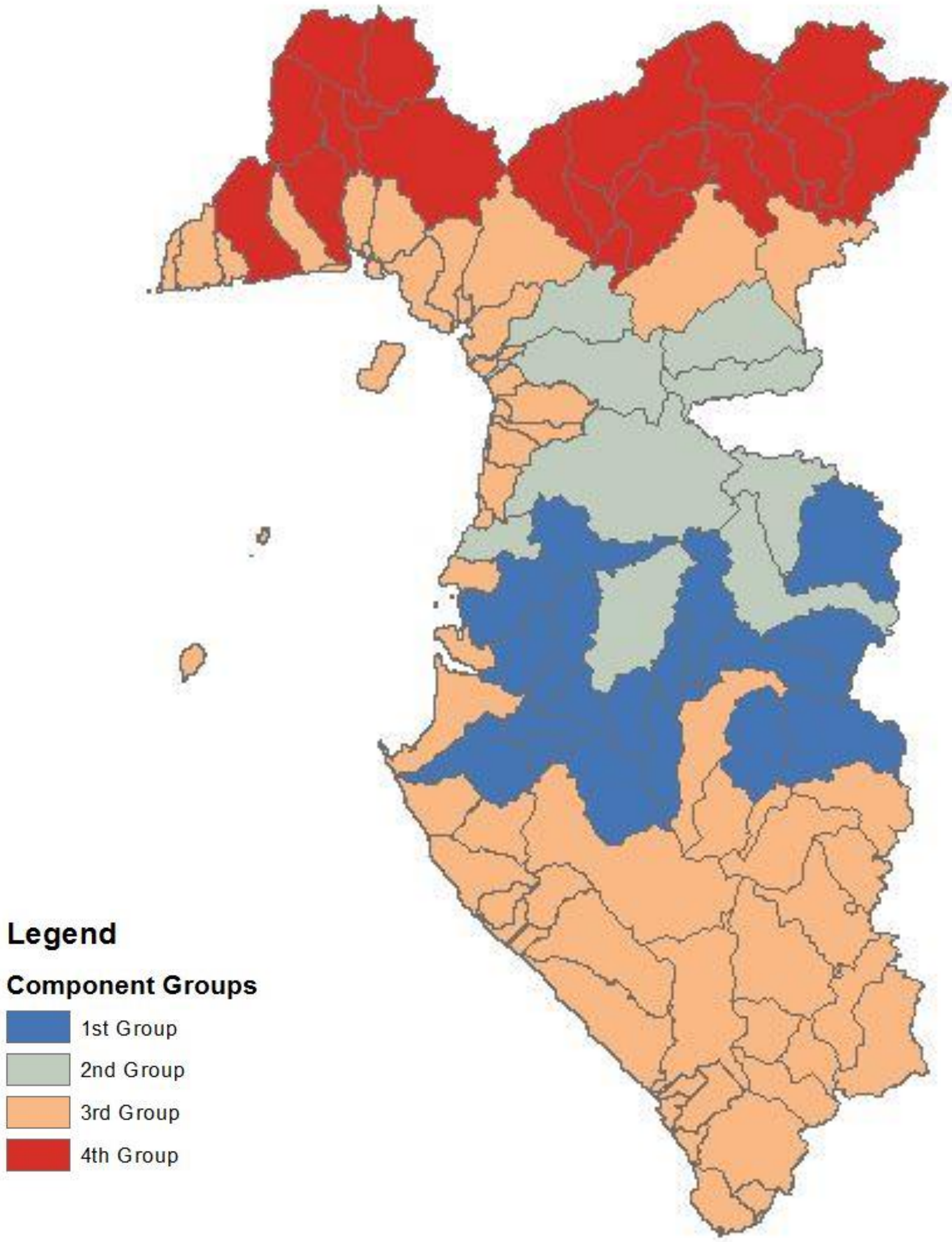
Interbasin 7 simulated monthly average stream-flow, 2001-2010.

Figure 2.43 - Horn's Test, Interbasin 7.



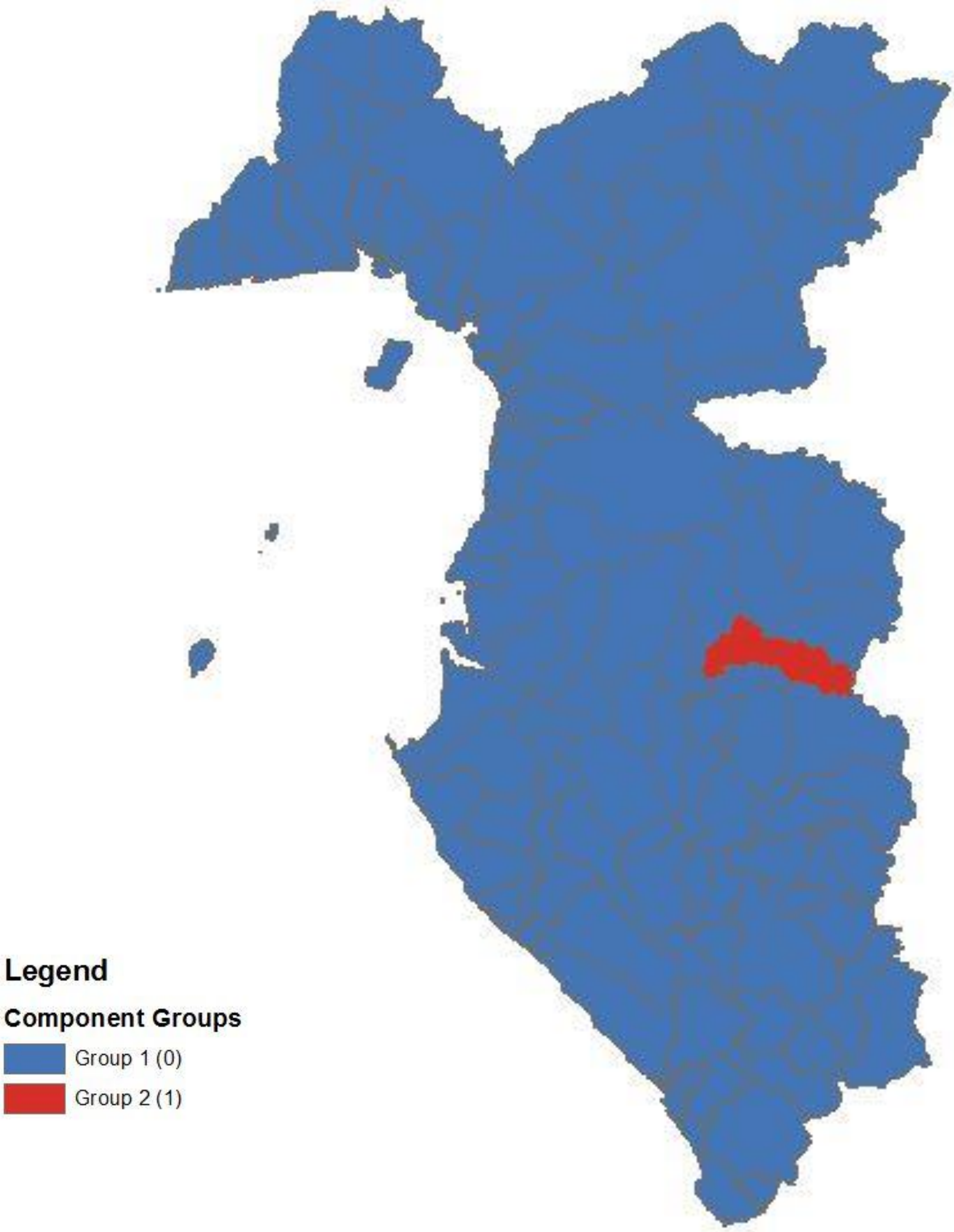
Interbasin 7, result of the Horn's test.

Figure 2.44 – First principal component, Interbasin 7.



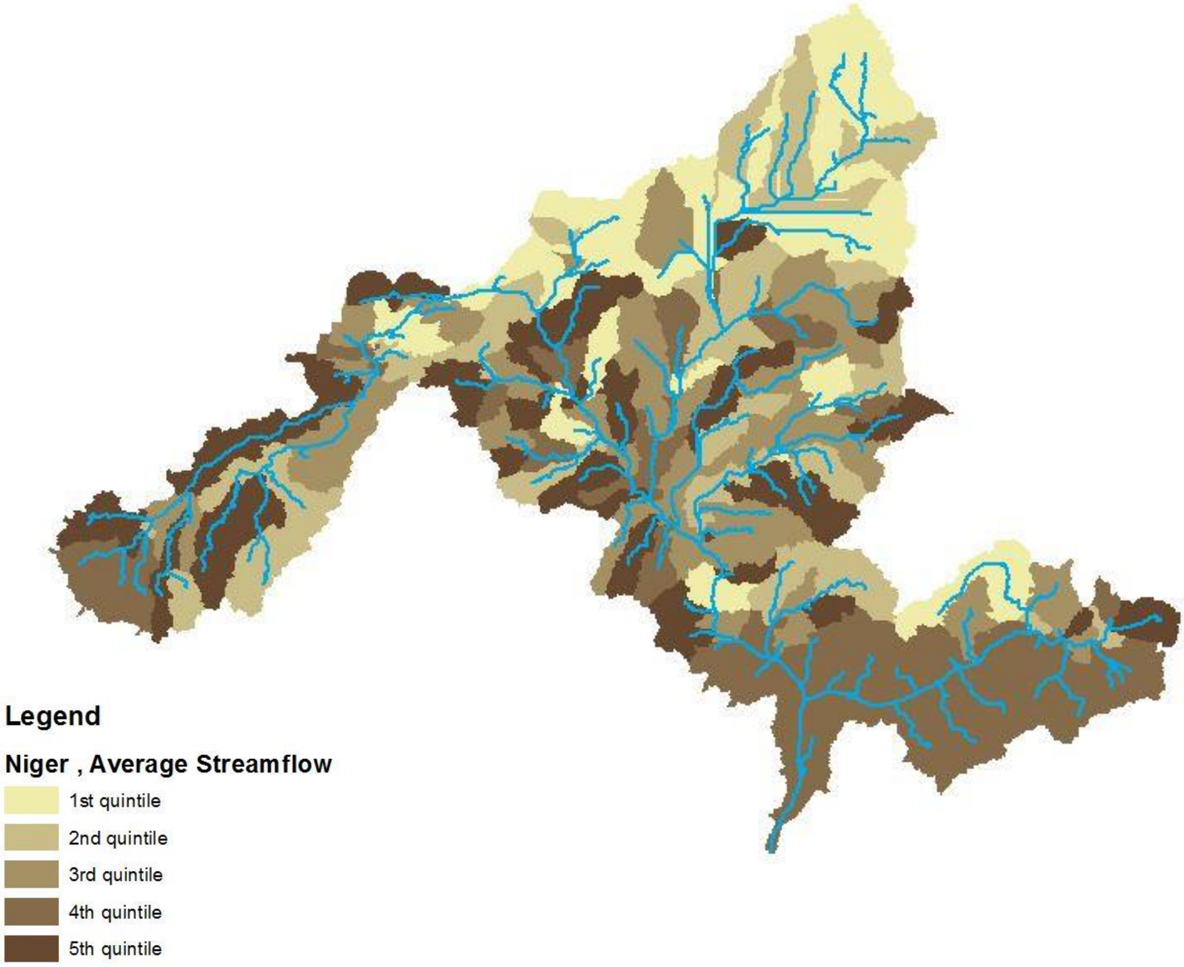
Interbasin 7, grouping over the loadings of the first principal component.

2.45 – First principal component, oblimin rotation, Interbasin 7.



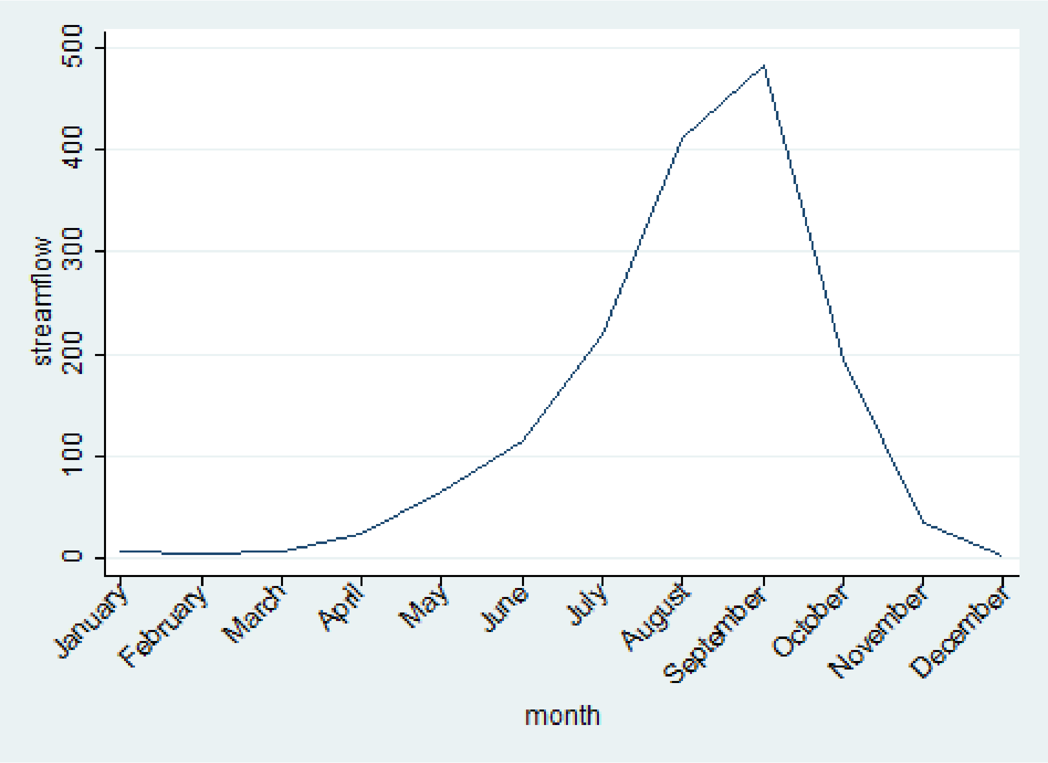
Interbasin 7, grouping over the loadings of the first principal component after oblimin rotation

Figure 2.46 – Niger.



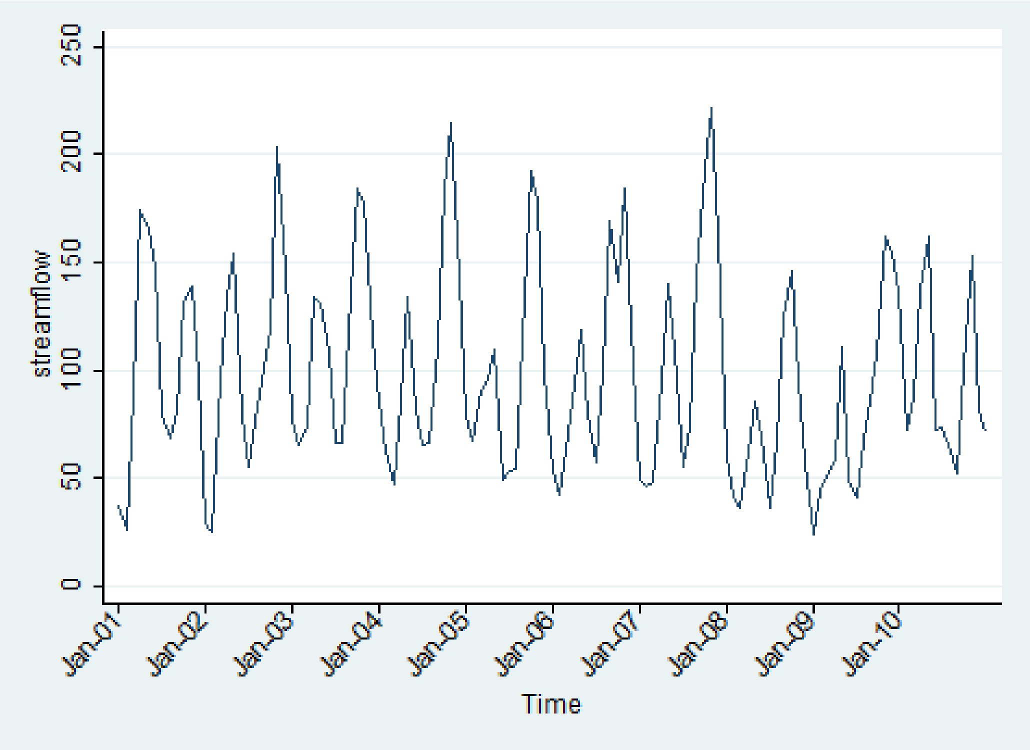
Niger, river network and average streamflow by quintile.

Figure 2.47 – Monthly average stremflow, Niger.



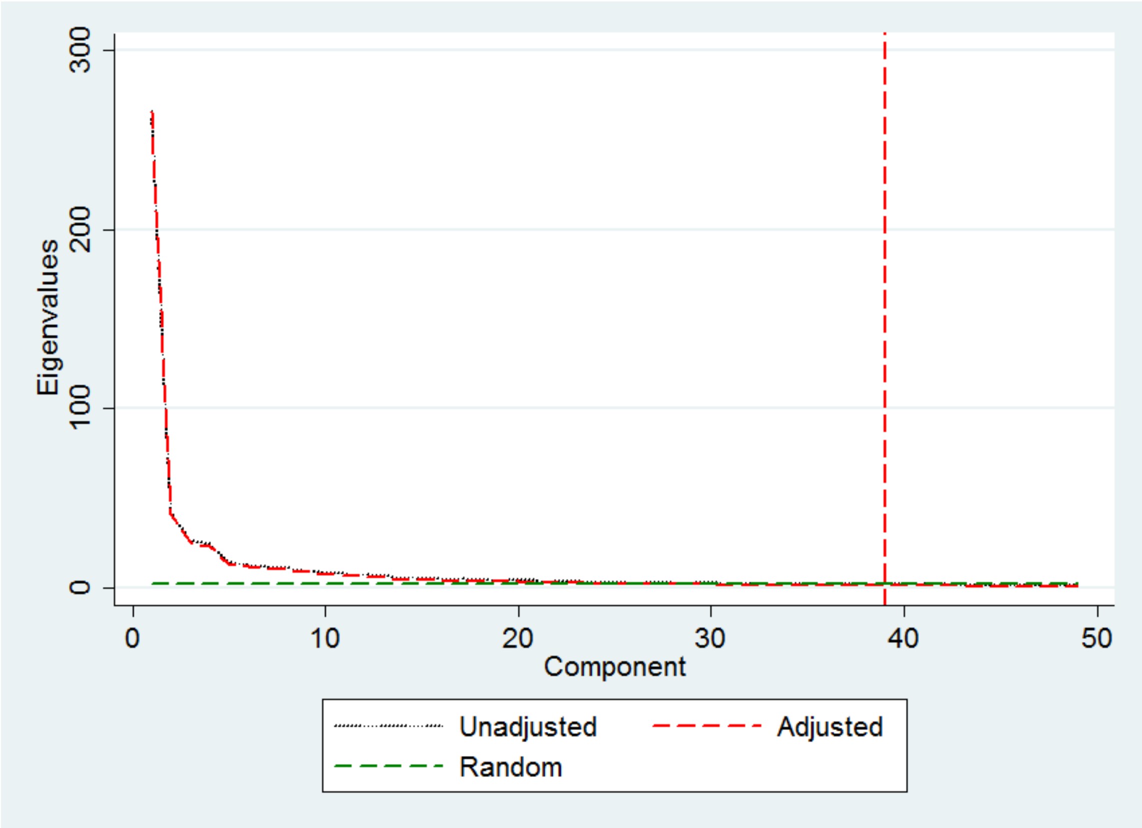
Niger, simulated monthly average stream-flow.

Figure 2.48 – Monthly streamflow 2001-2010, Niger.



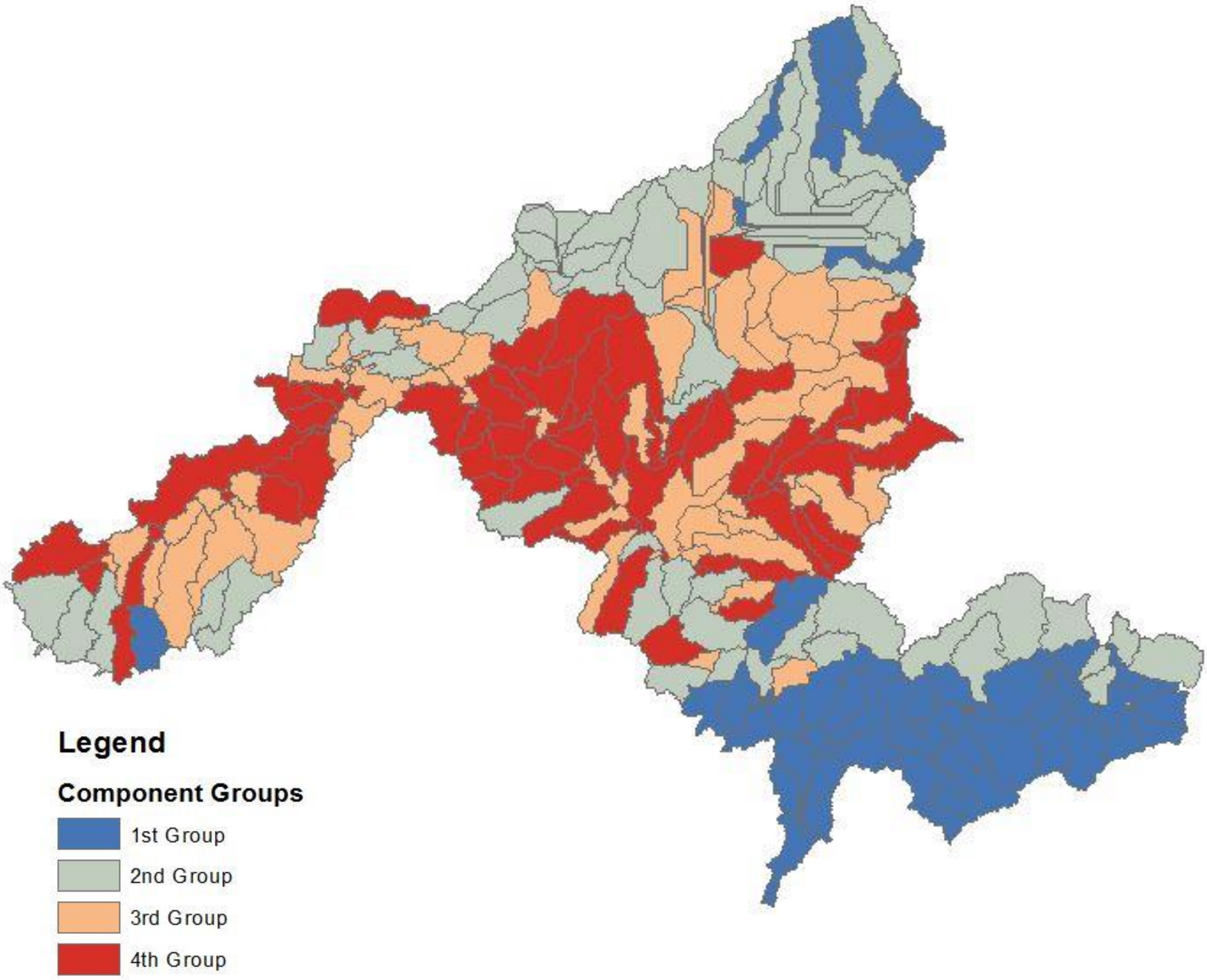
Niger, simulated monthly average stream-flow, 2001-2010.

Figure 2.49 – Horn's Test, Niger.



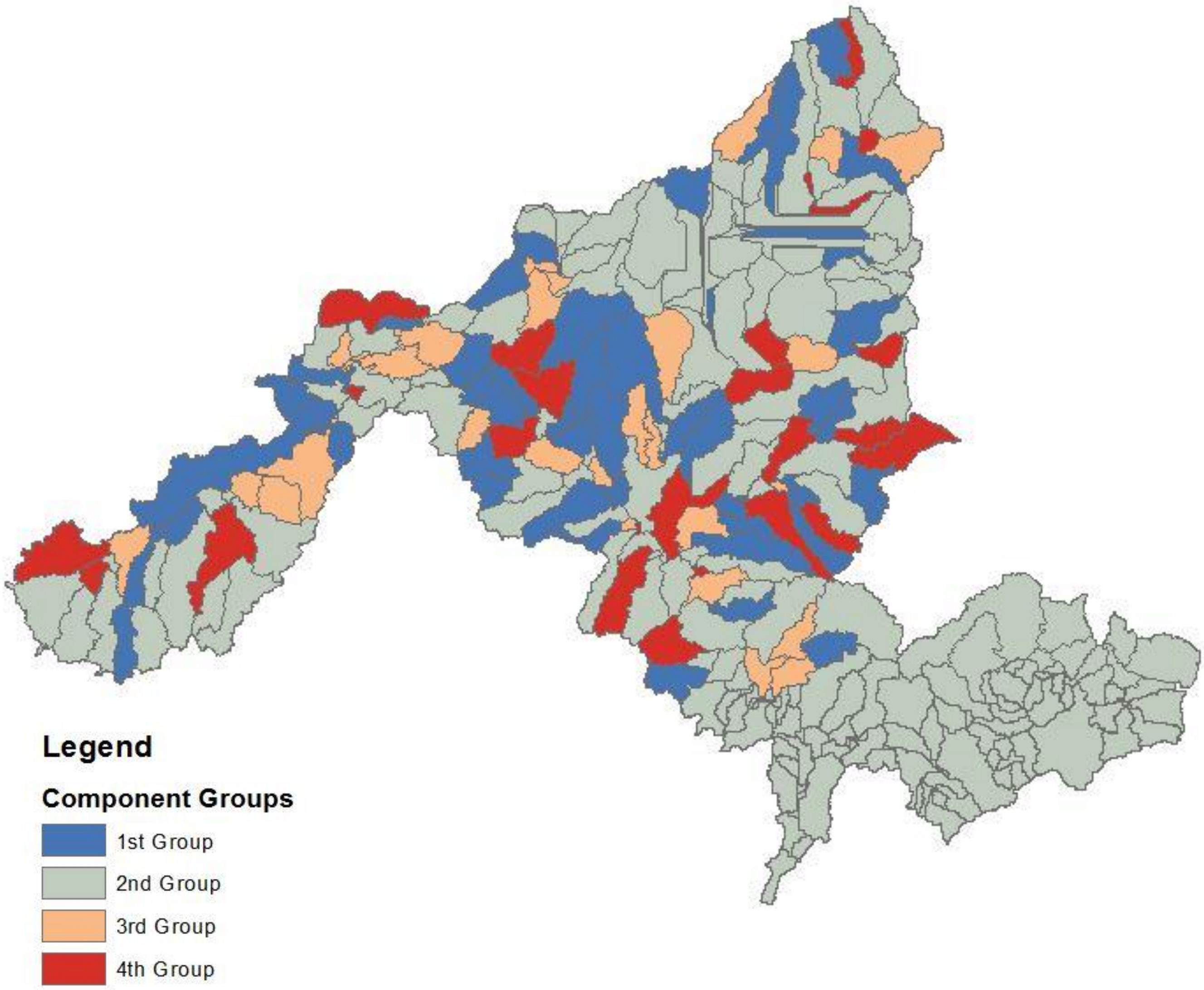
Niger, result of the Horn's test.

Figure 2.50 – First principal component, Niger.



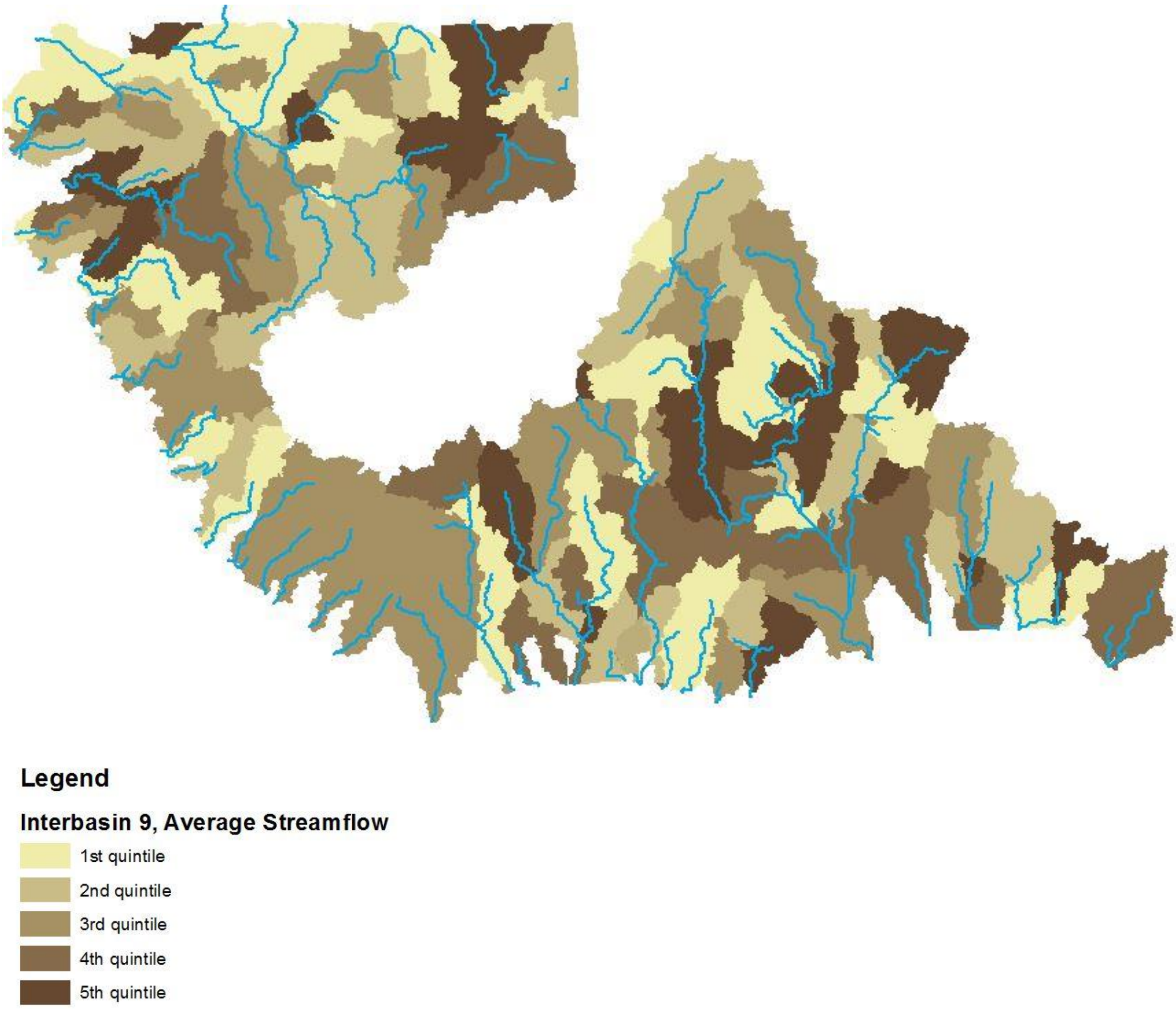
Niger, grouping over the loadings of the first principal component.

Figure 2.51 – First principal component, oblimin rotation, Niger.



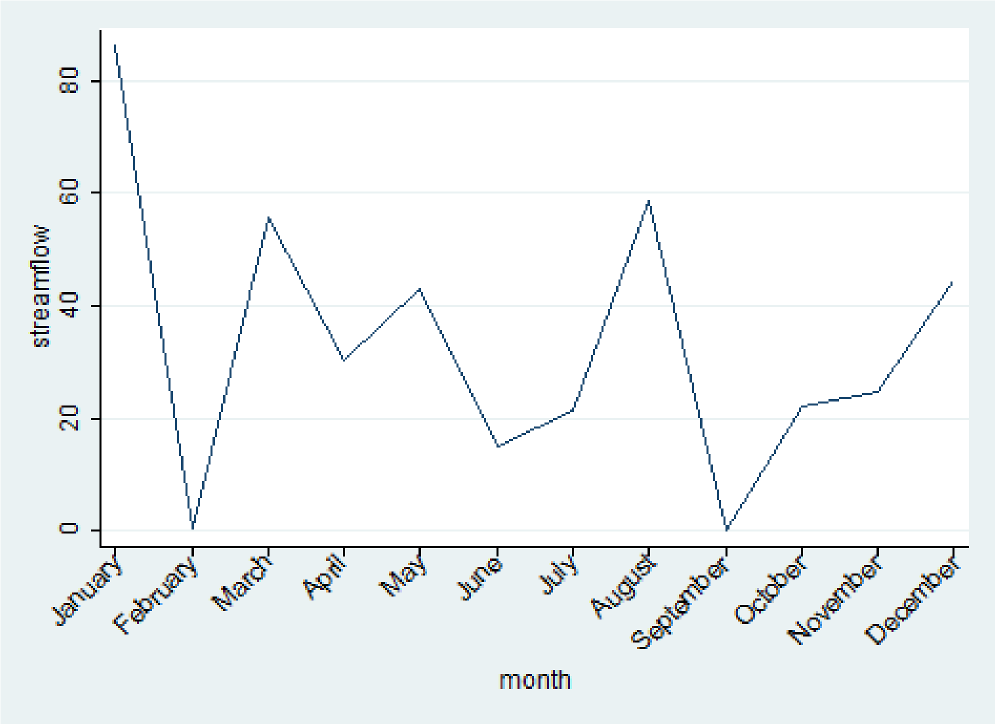
Niger, grouping over the loadings of the first principal component after oblimin rotation.

Figure 2.52 – Interbasin 9.



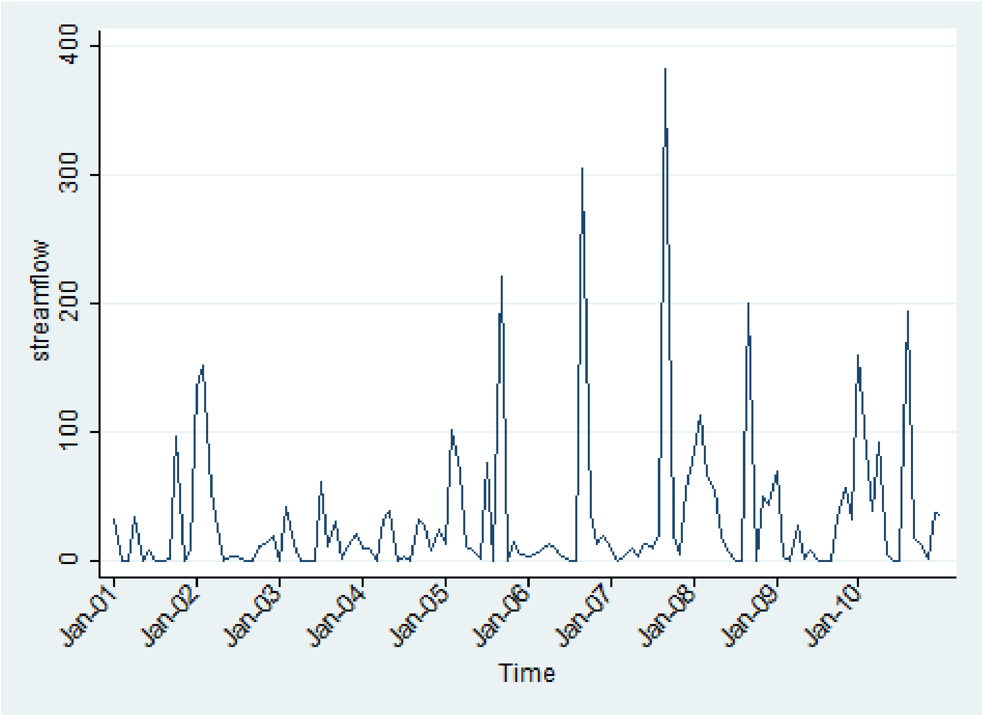
Interbasin 9, river network and average streamflow by quintile.

Figure 2.53 – Monthly average streamflow, Interbasin 9.



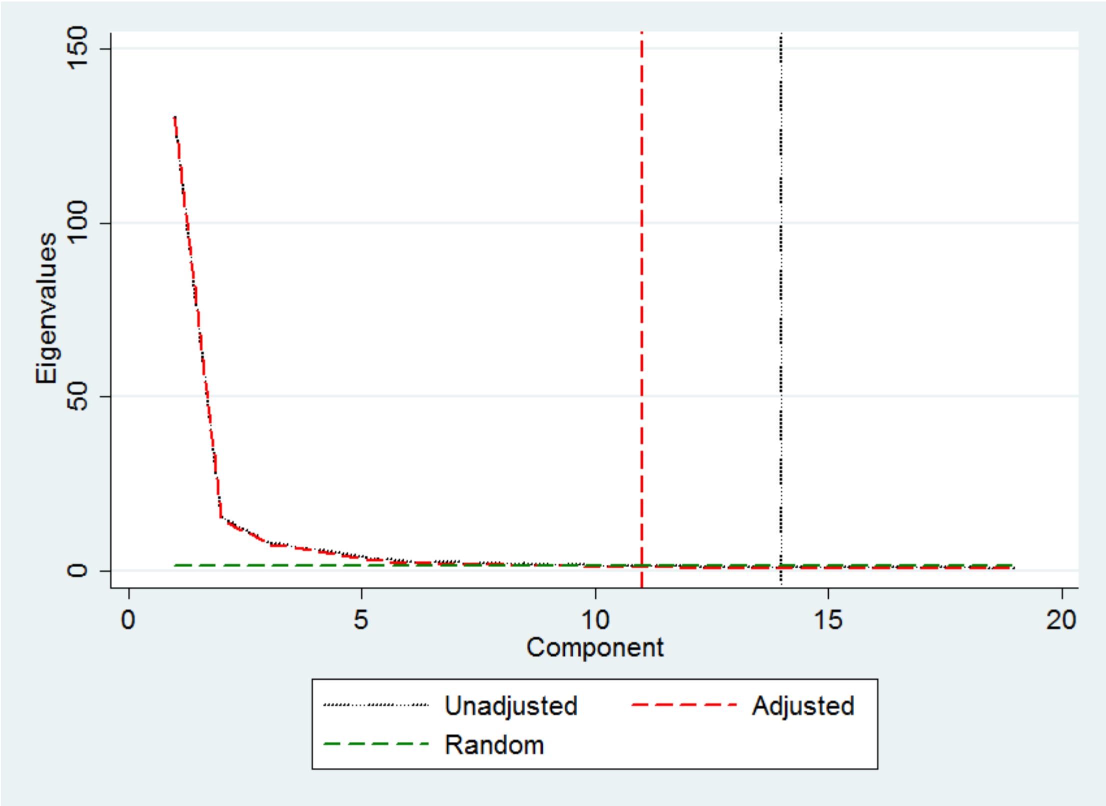
Interbasin 9 simulated monthly average stream-flow in 2005.

Figure 2.54 – Monthly streamflow 2001-2010, Interbasin 9.



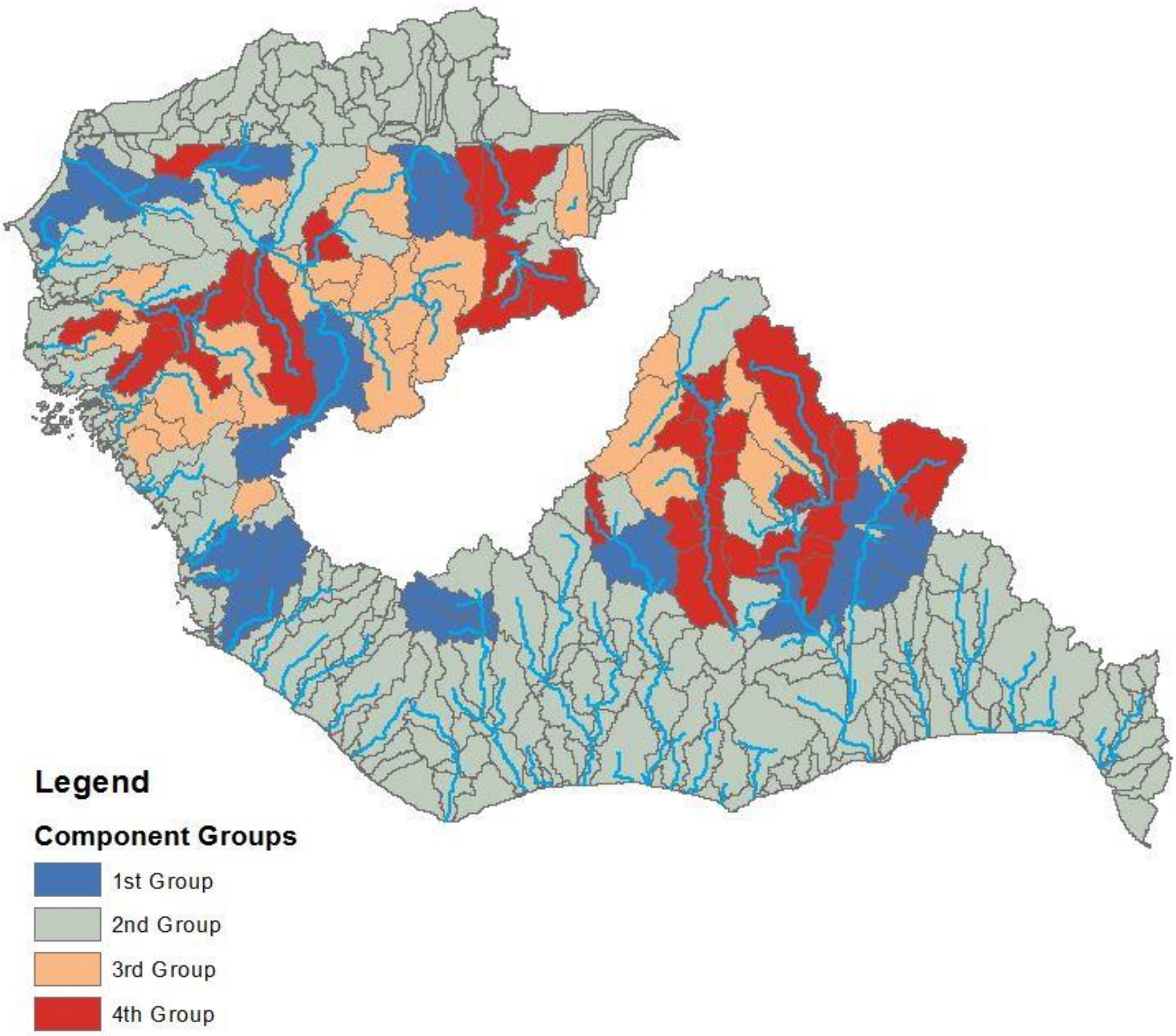
Interbasin 9 simulated monthly average stream-flow, 2001-2010.

Figure 2.55 – Horn's Test, Interbasin 9.



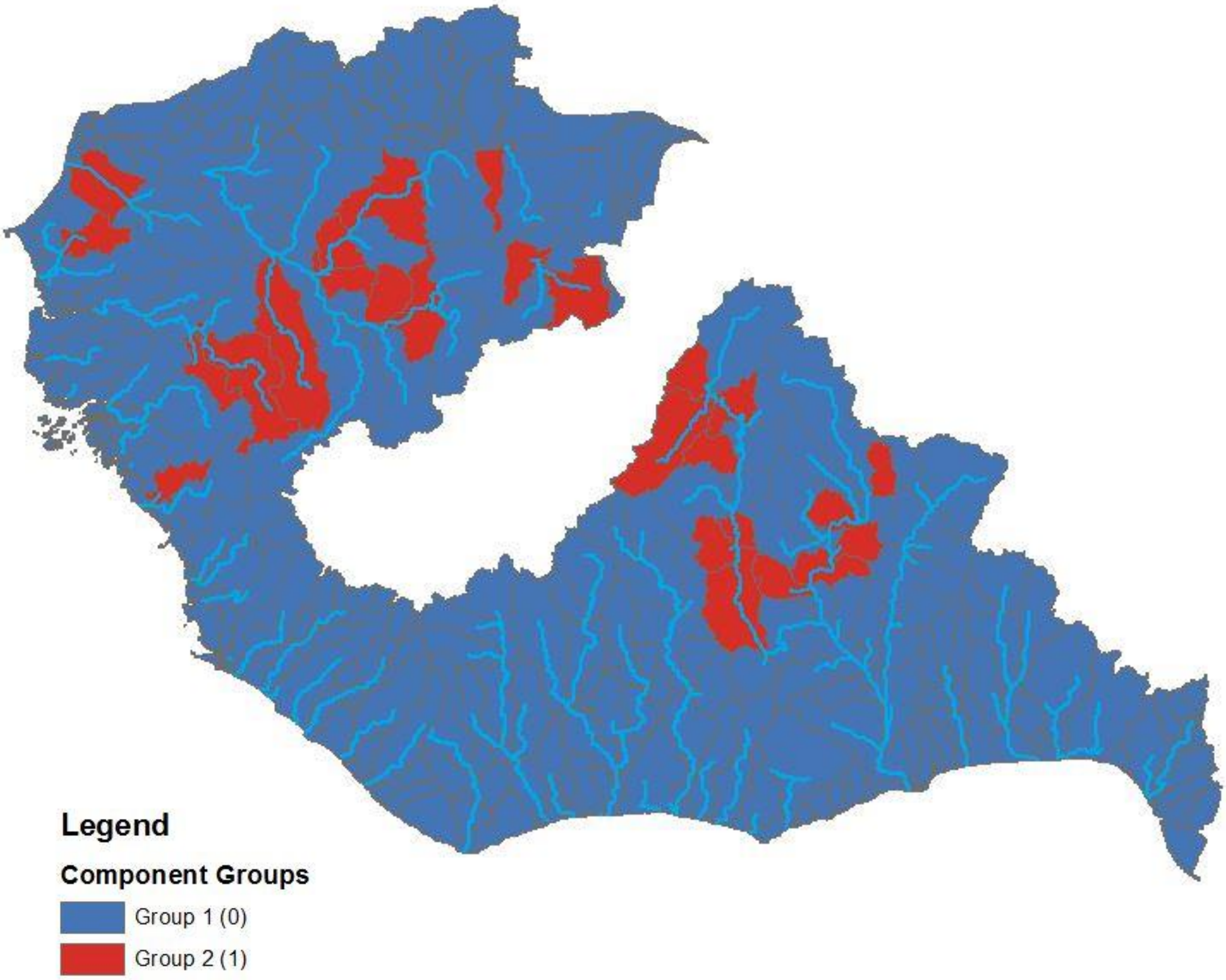
Interbasin 9, result of the Horn's test.

Figure 2.56 – First principal component, Interbasin 9.



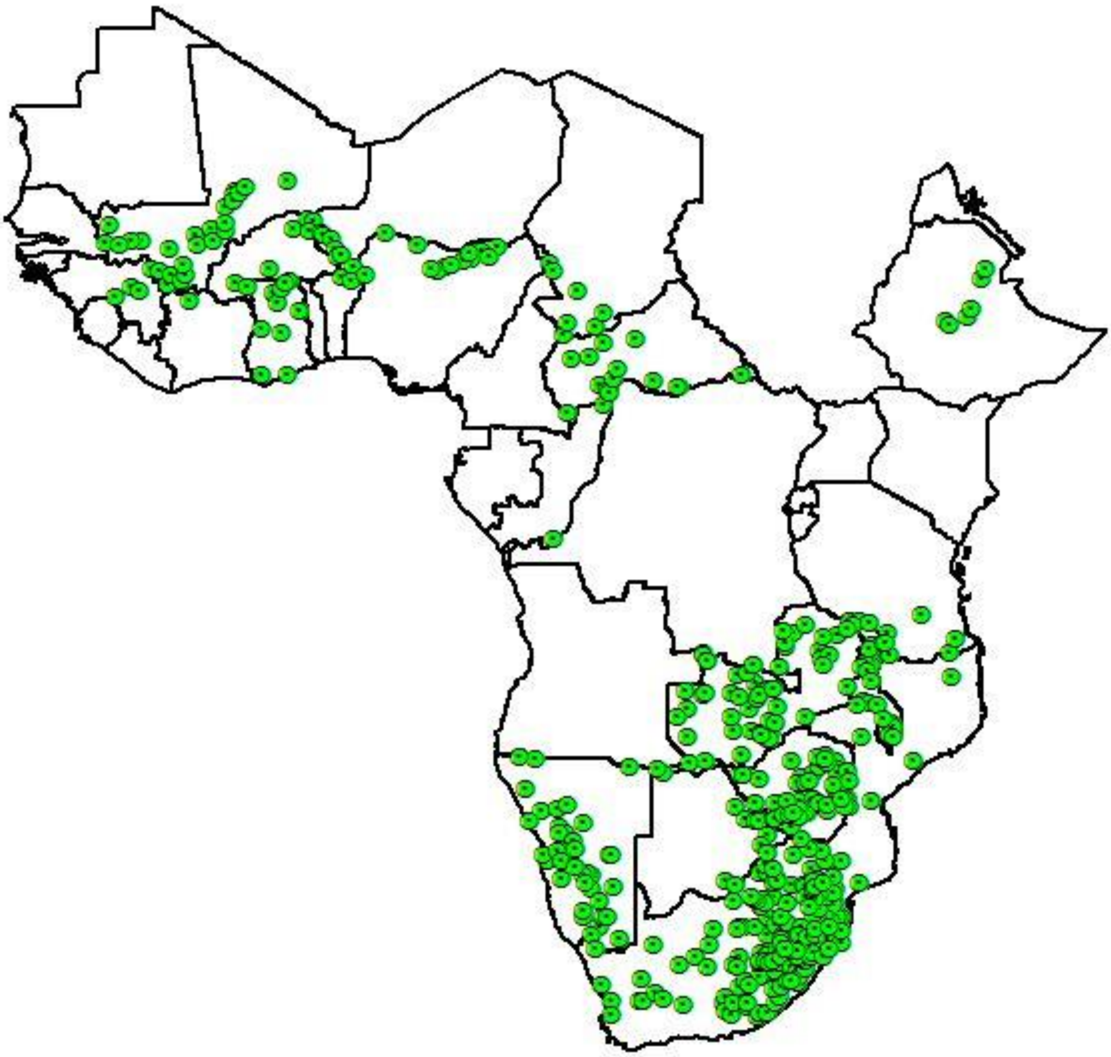
Interbasin 9, grouping over the loadings of the first principal component

Figure 2.57 – First principal component, oblimin rotation, Interbasin 9.



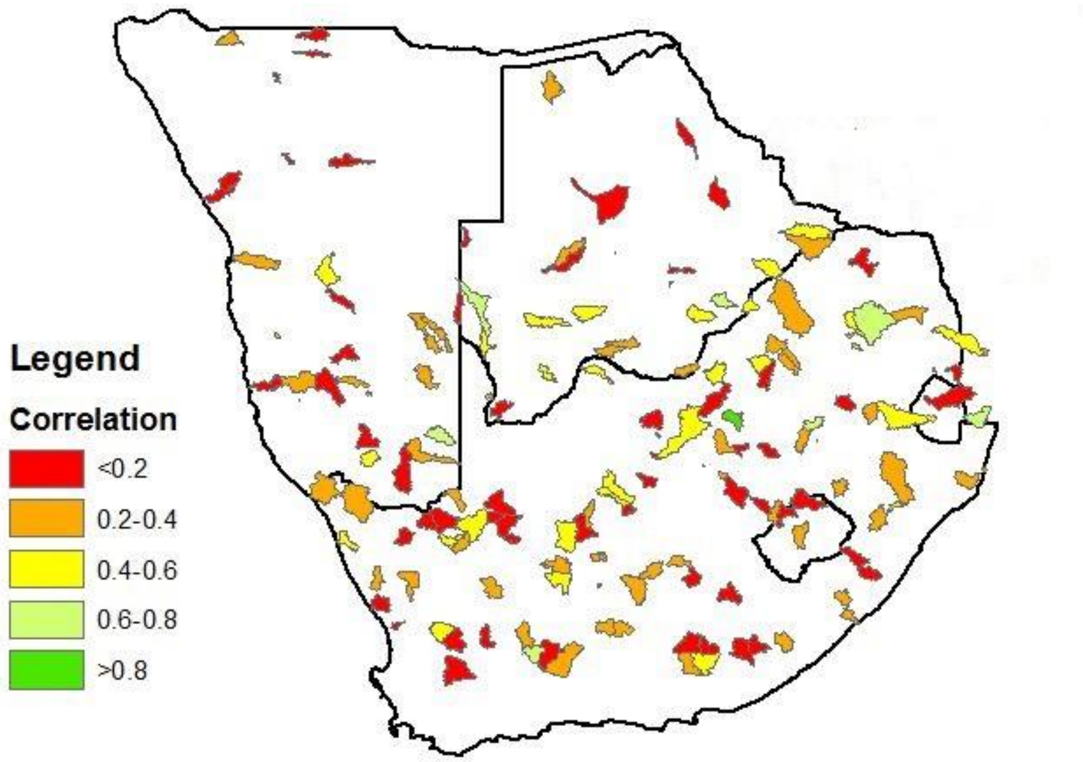
Interbasin 9, grouping over the loadings of the first principal component after oblimin rotation

Figure 2.58 – GRDC gauge stations.



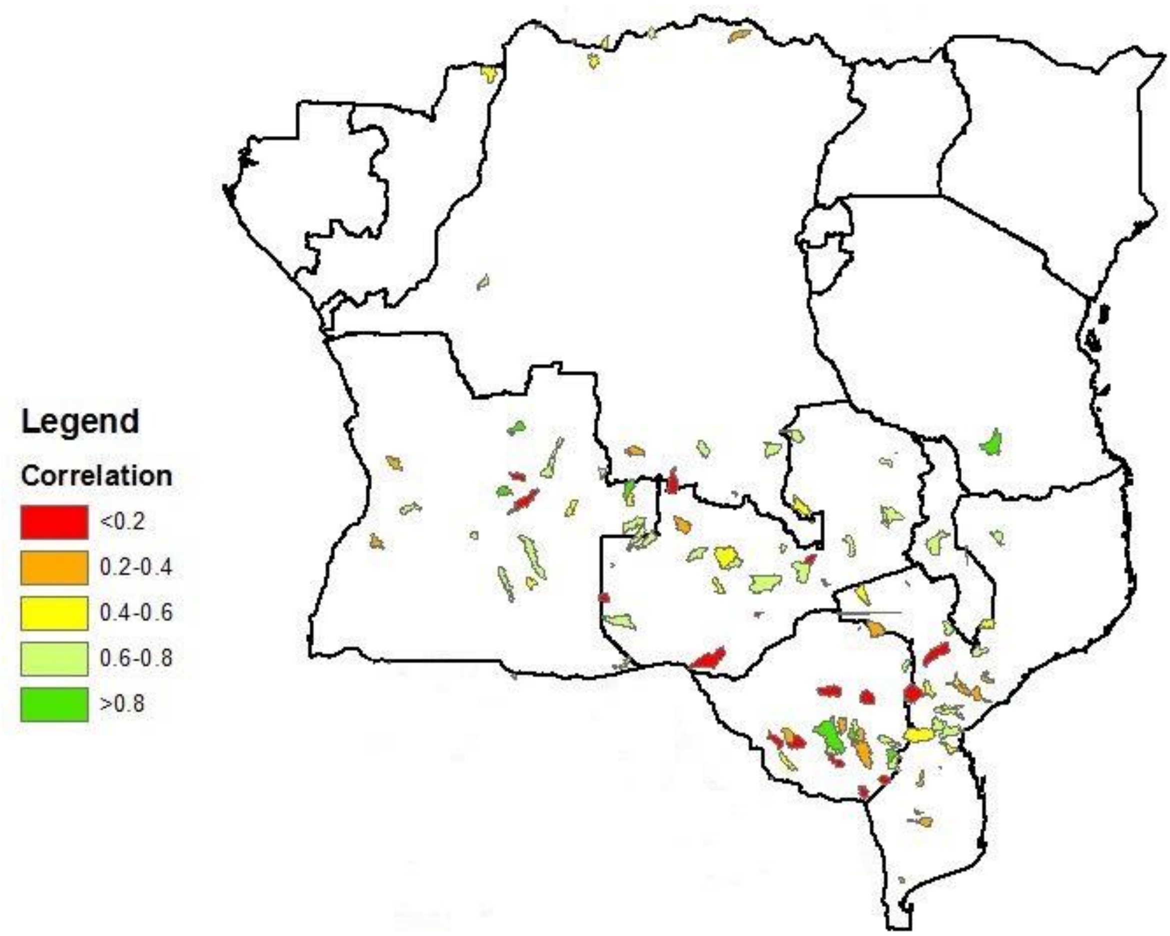
GRDC gauge station location

Figure 2.59 – GRDC-GeoSFM correlation, Southern Africa.



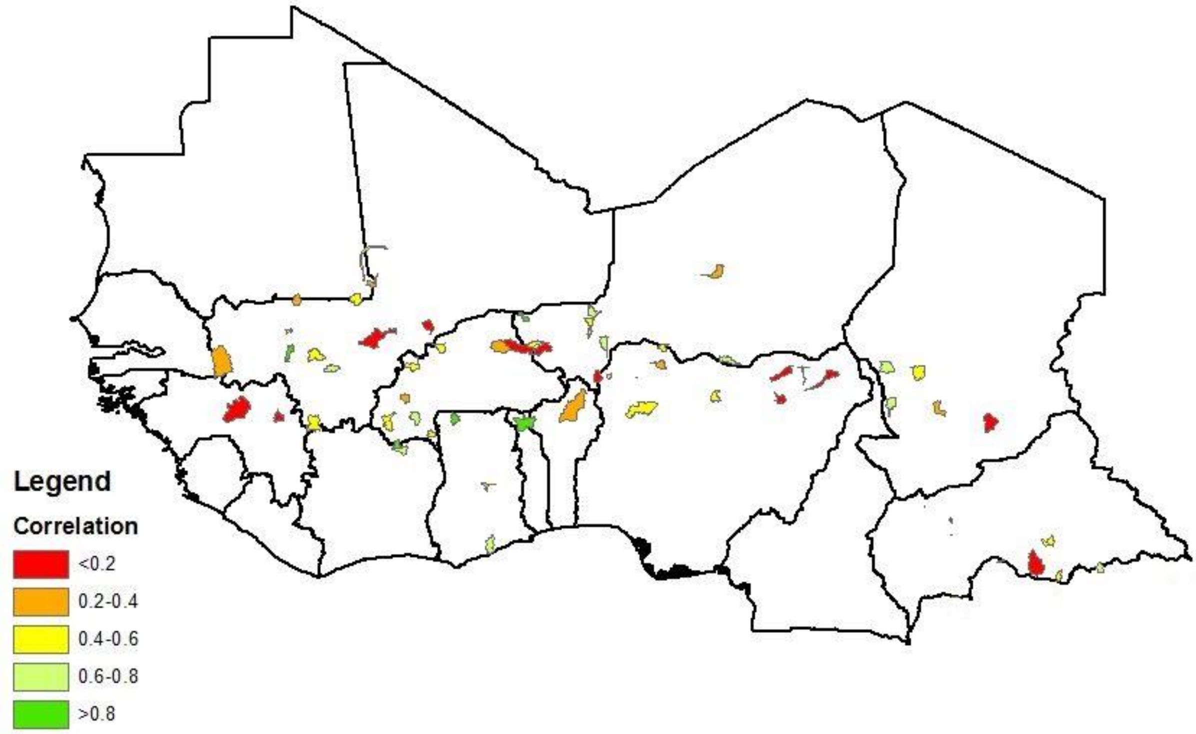
Correlation between model estimates and historical data from GRDC database in Southern Africa.

Figure 2.60 – GRDC-GeoSFM correlation, Central Africa.



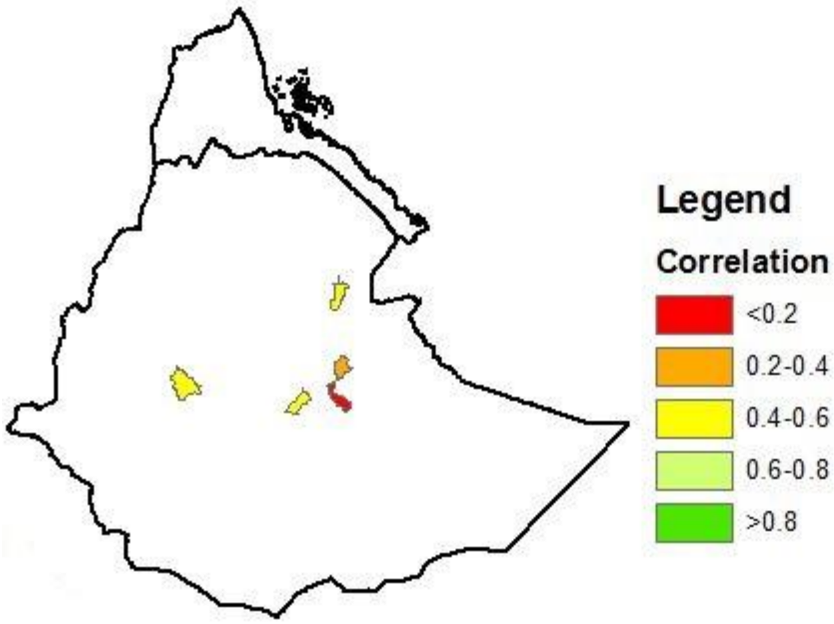
Correlation between model estimates and historical data from GRDC database in Central Africa.

Figure 2.61 – GRDC-GeoSFM correlation, Western Africa.



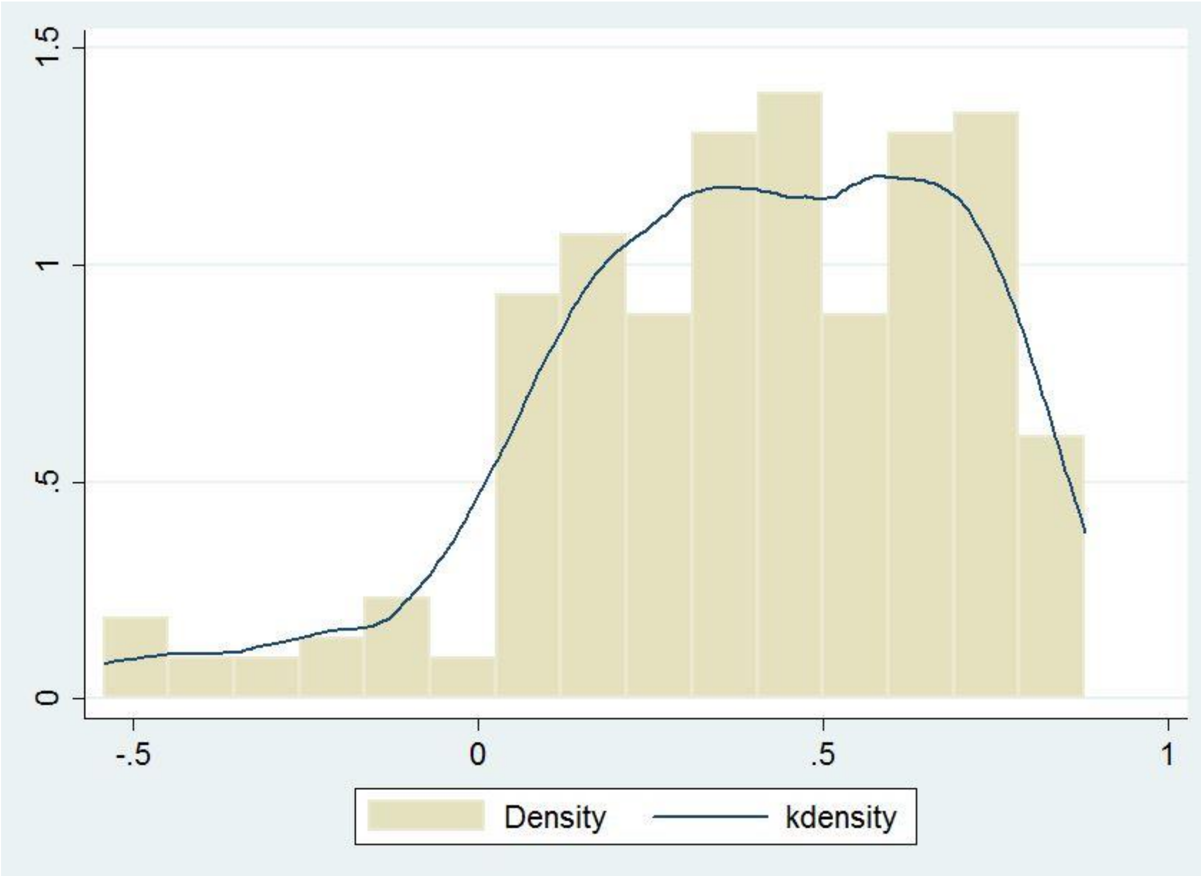
Correlation between model estimates and historical data from GRDC database in Western Africa.

Figure 2.62 – GRDC-GeoSFM correlation, Eastern Africa.



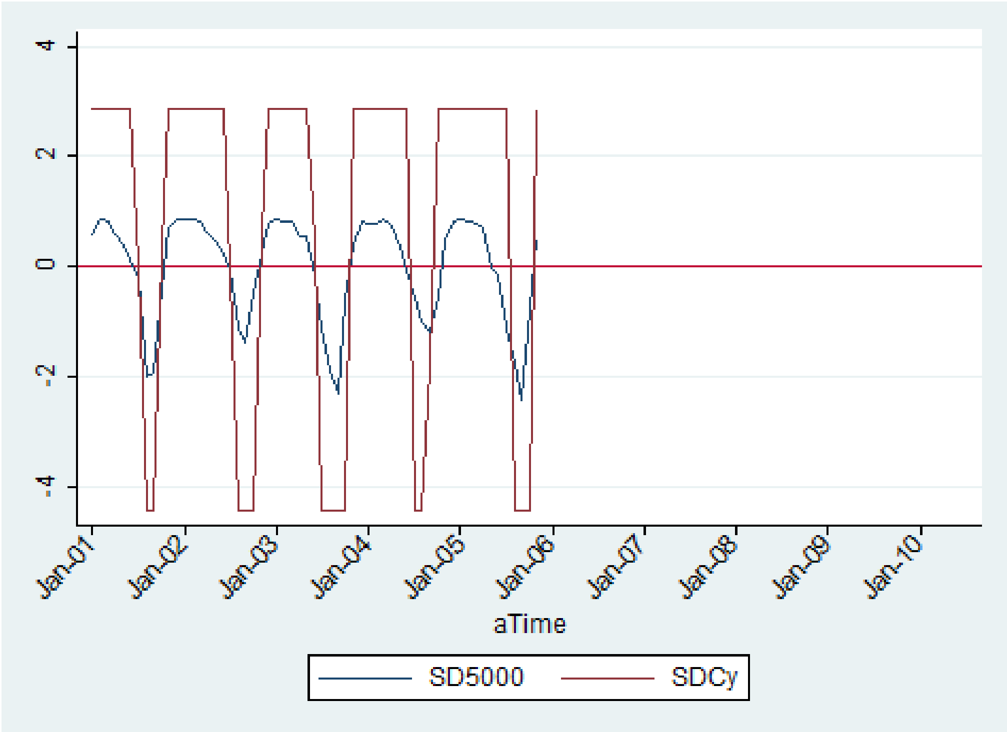
Correlation between model estimates and historical data from GRDC database in Eastern Africa.

Figure 2.63 – Density and Kernel density of significant correlation.



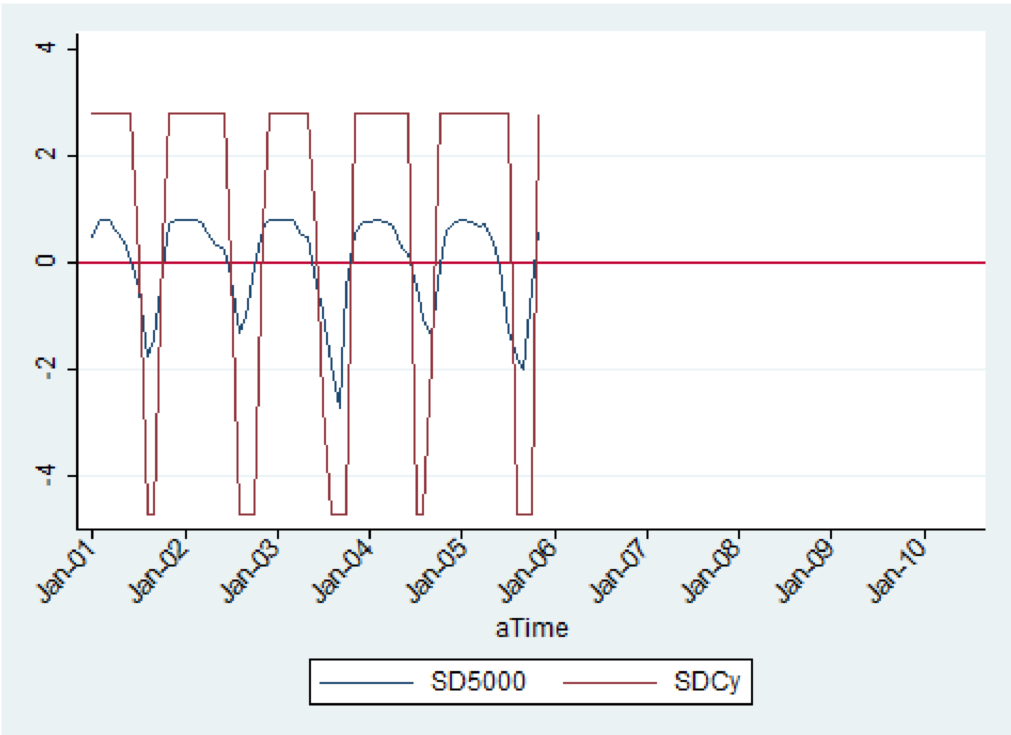
Density (histograms) and Kernel density (blue line) of significant correlation between GRDC stations and GeoSFM estimates.

Figure 2.64 – Historical and simulated anomalies, Lake Chad.



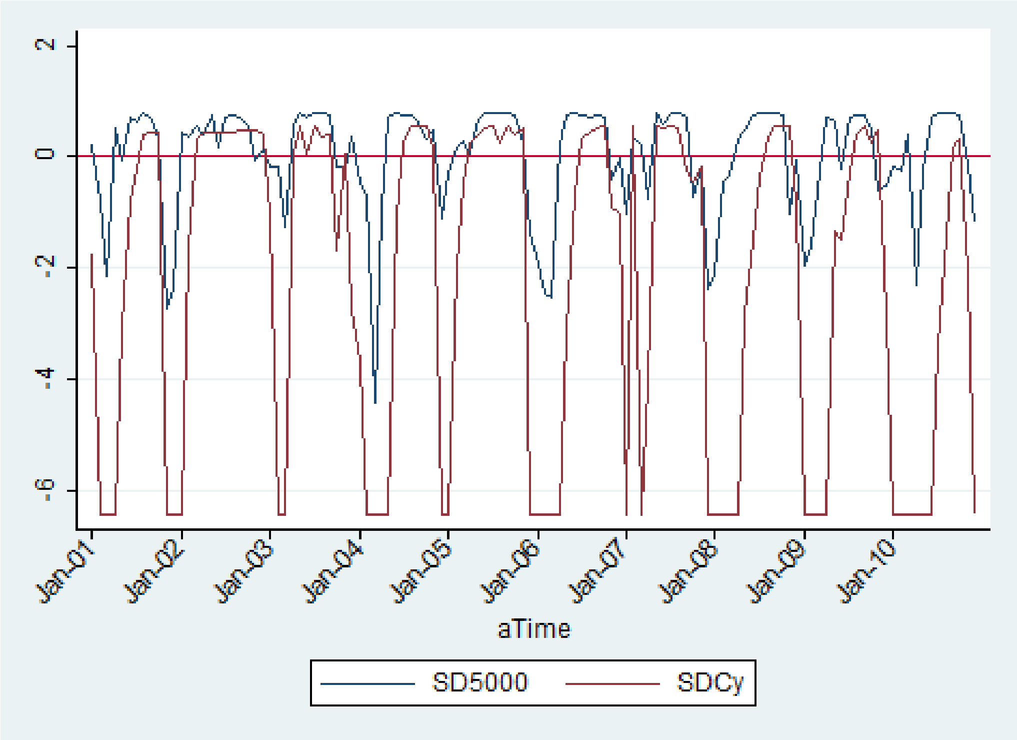
GRDC station in Lake Chad, blue line simulated anomalies, red line observed data.

Figure 2.65 – Historical and simulated anomalies, Niger.



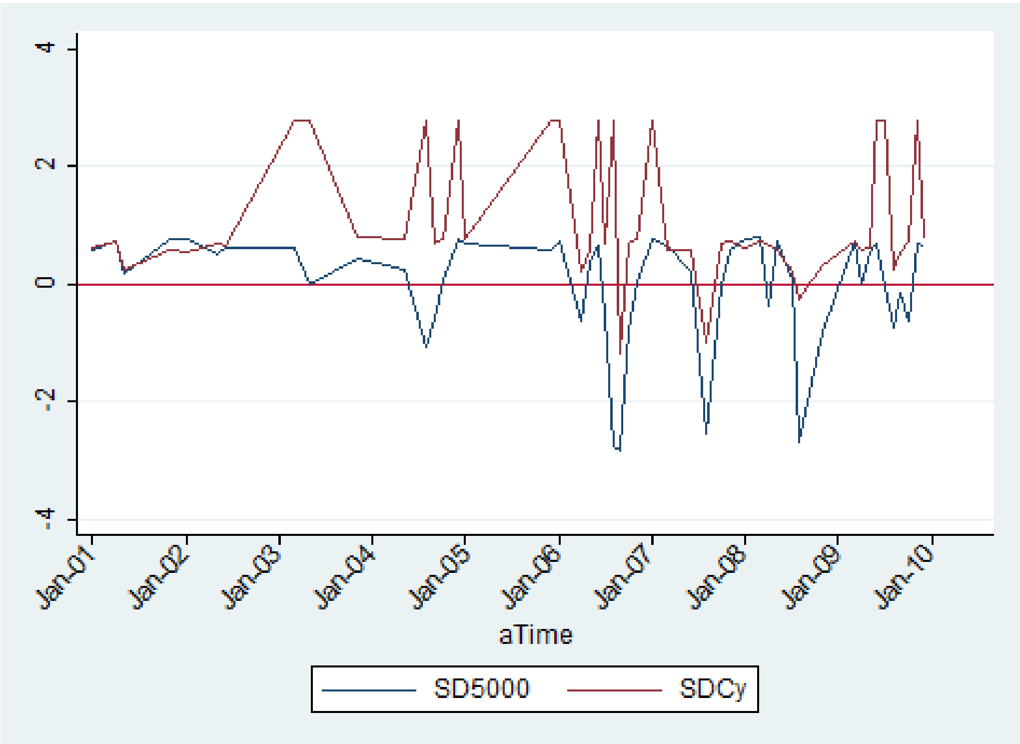
GRDC station in Niger, blue line simulated anomalies, red line observed data.

Figure 2.66 – Historical and simulated anomalies, Interbasin 5.



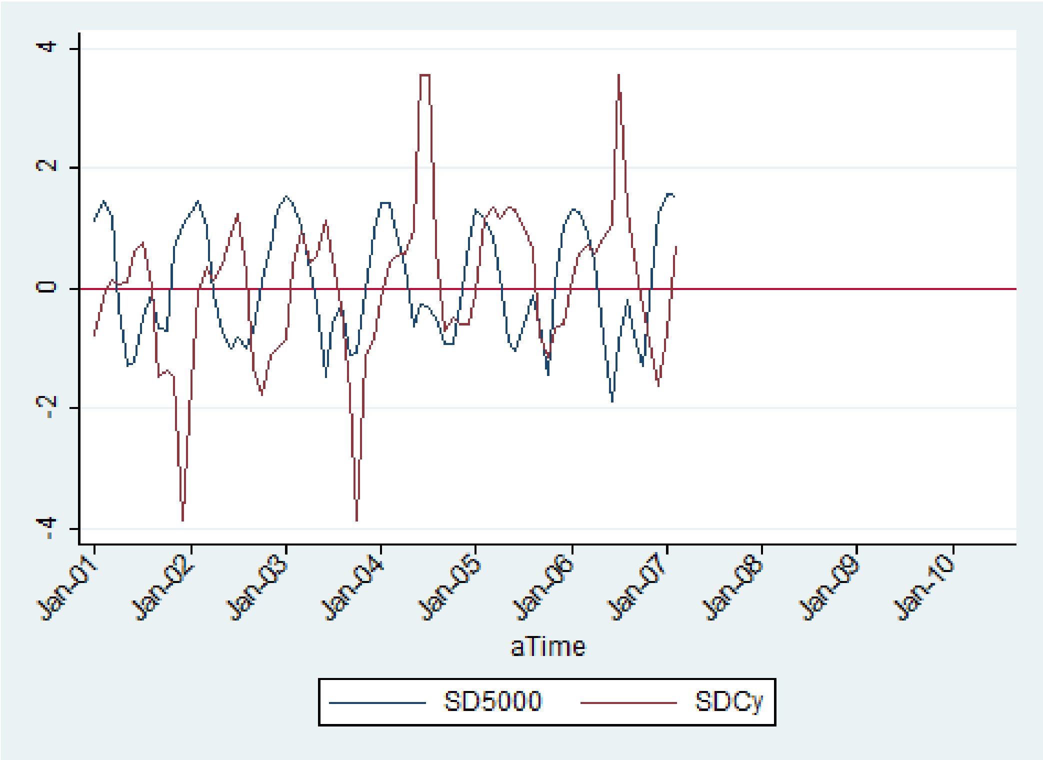
GRDC station in Interbasin 5, blue line simulated anomalies, red line observed data.

Figure 2.67 – Historical and simulated anomalies, Interbasin 3.



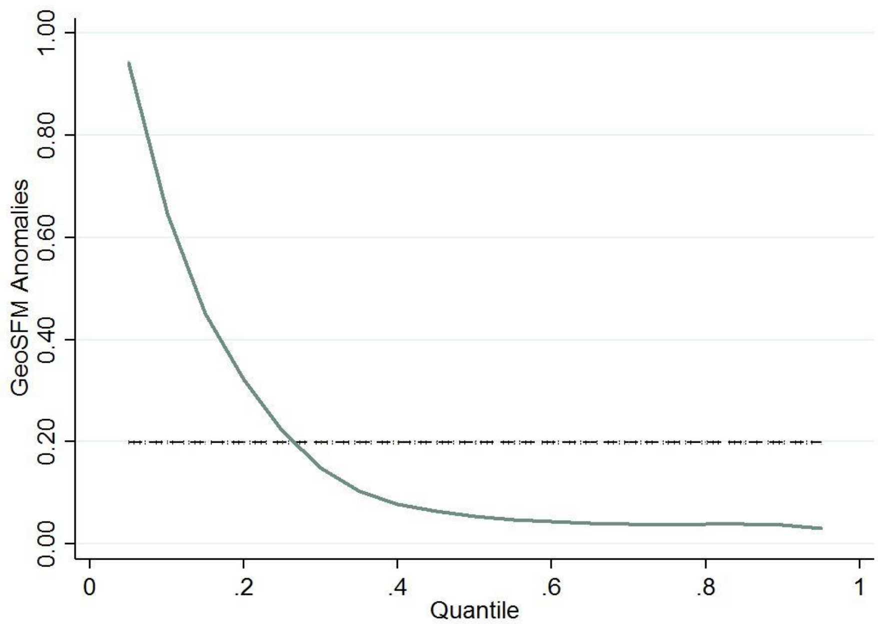
GRDC station in Interbasin 3, blue line simulated anomalies, red line observed data.

Figure 2.68 – Historical and simulated anomalies, Interbasin 9.



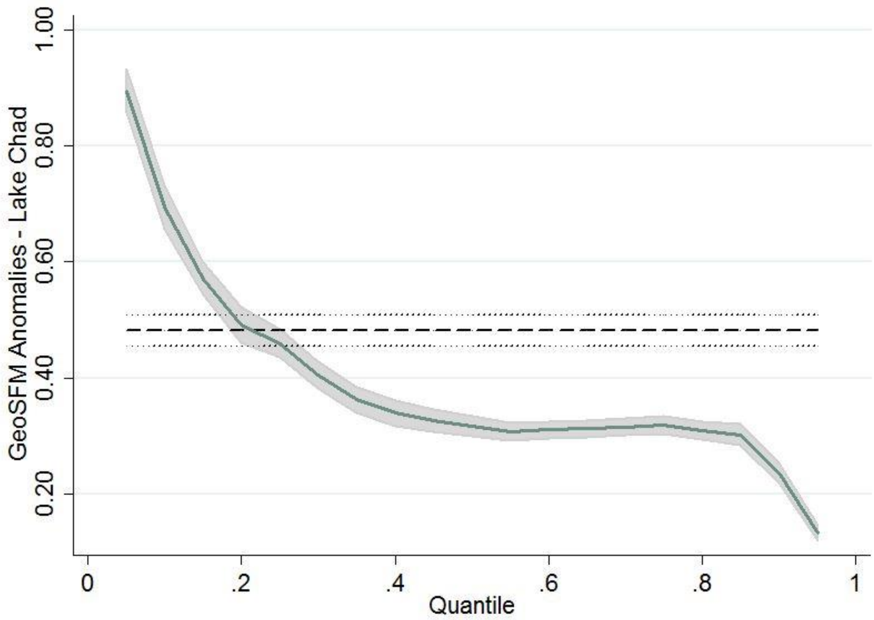
GRDC station in Interbasin 9, blue line simulated anomalies, red line observed data.

Figure 2.69 – Quantile regression and OLS coefficients, all basins.



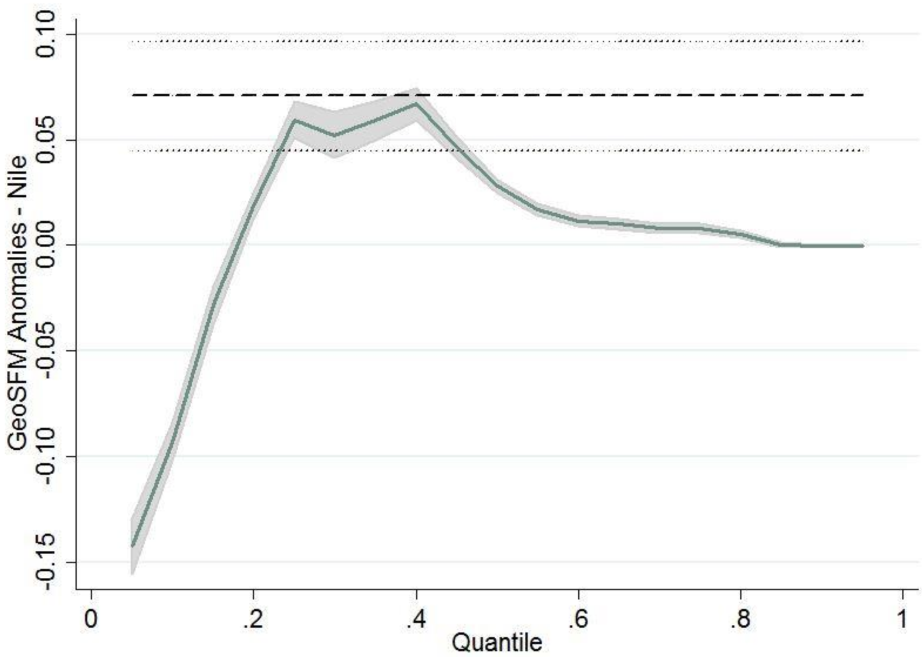
Graphed quantile regression coefficients for GeoSFM anomalies (blue line) including the equivalent OLS coefficient (dotted line) aggregated over all basins. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.70 – Quantile regression and OLS coefficients, Lake Chad.



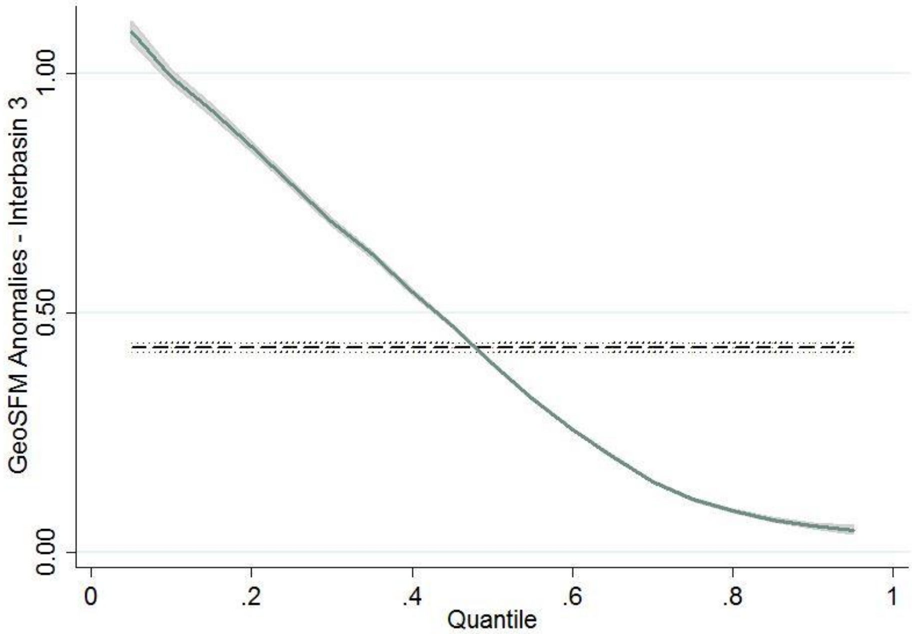
Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for Lake Chad. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.71 – Quantile regression and OLS coefficients, Nile.



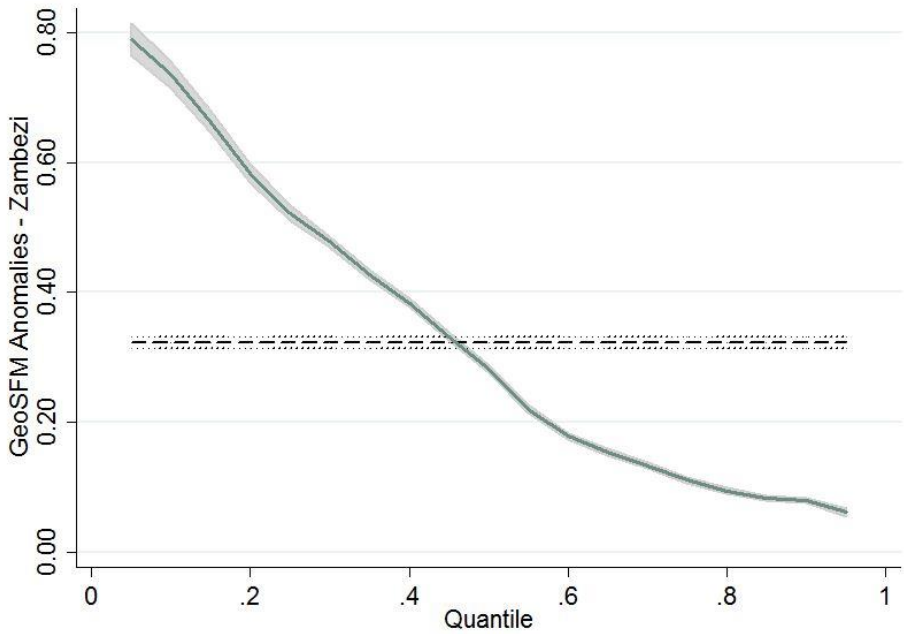
Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for the Nile. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.72 – Quantile regression and OLS coefficients, Interbasin 3.



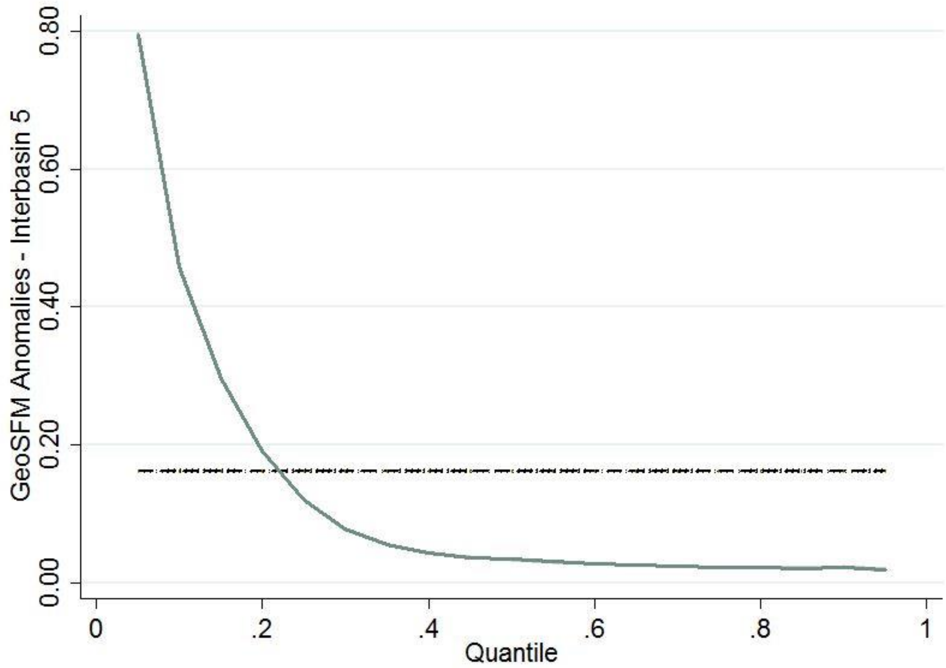
Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for Interbasin 3. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.73 – Quantile regression and OLS coefficients, Zambezi.



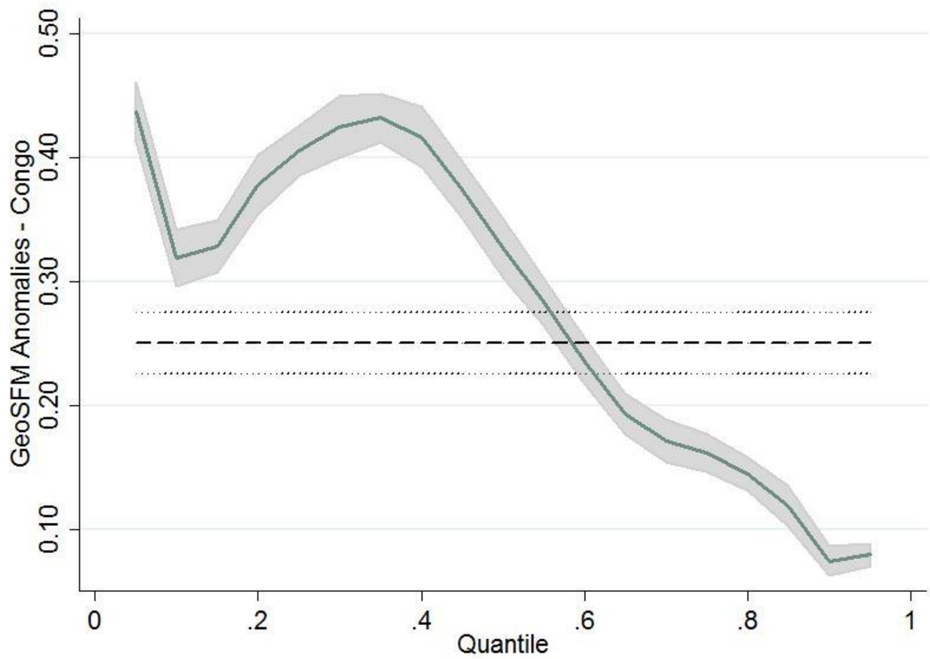
Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for the Zambezi. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.74 – Quantile regression and OLS coefficients, Interbasin 5.



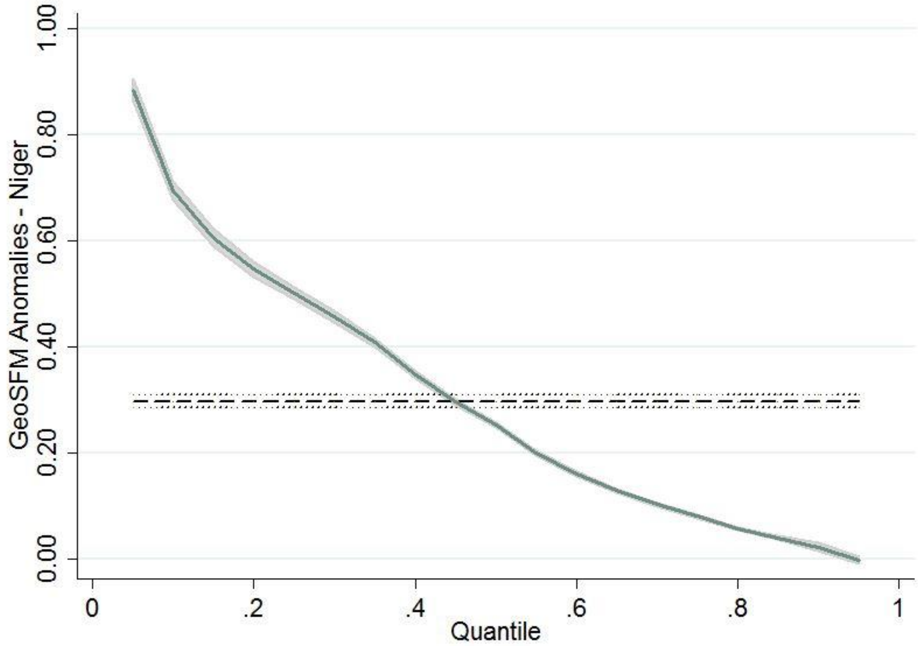
Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for Interbasin 5. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.75 – Quantile regression and OLS coefficients, Congo.



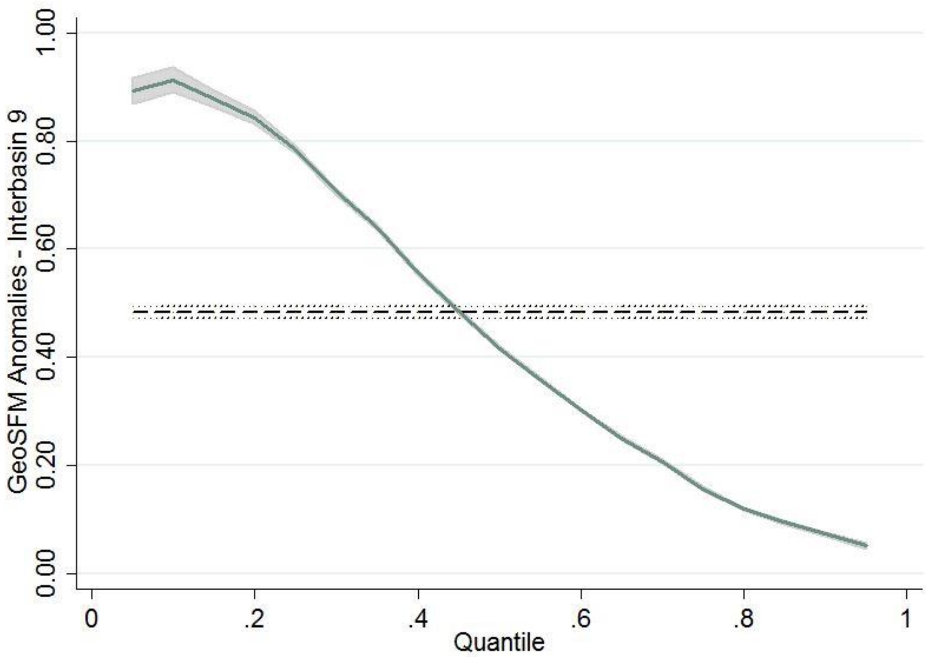
Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for the Congo. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.76 – Quantile regression and OLS coefficients, Niger.



Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for the Niger. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Figure 2.77 – Quantile regression and OLS coefficients, Interbasin 9.



Graphed quantile regression coefficients and confidence interval for GeoSFM anomalies (blue line and grey area) including the equivalent OLS coefficient (dashed line and area between dotted lines) for Interasin 9. Regression includes daily rainfall and evapotranspiration anomalies and their maximum yearly value.

Tables 2.1 – GRDC-GeoSFM correlations by basin

Basin	GRDC corr.	GRDC stations num.	Significant corr. (5%)
Lake Chad	0.55	8	3/8 (37%)
Nile	-	1	-
Basin 3	0.64	13	11/13 (83%)
Zambezi	0.6	40	28/40 (70%)
Basin 5	0.49	239	163/239 (68%)
Congo	0.67	21	14/21 (65%)
Basin 7	-	-	-
Niger	0.54	46	23/46 (50%)
Basin 9	0.75	18	13/18 (69%)

Basins correlation between simulated and observed river discharge for GRDC dataset. The first column reports the average of significant correlations, the second the overall number of gauge stations per basin, the third the percentage of significant correlations.

Table 2.2 – GRDC-GeoSFM dependence comparison, Lake Chad.

Station ID	1837106	1837107	1837200	1837401	1837451	1837560
1837106	0.00	0.02	-0.21	-0.65	-0.33	-0.81
1837107	0.02	0.00	-0.30	-0.59	-0.11	-0.90
1837200	-0.21	-0.30	0.00	0.21	-0.17	0.26
1837401	-0.65	-0.59	0.21	0.00	-0.59	-0.19
1837451	-0.33	-0.11	-0.17	-0.59	0.00	-0.76
1837560	-0.81	-0.90	0.26	-0.19	-0.76	0.00

Comparison of GRDC-GeoSFM regional dependence for Lake Chad obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.3 - GRDC-GeoSFM dependence comparison, Interbasin 3.

Station ID	1577100	1577102	1577101	1577601
1577100	0.00	-0.18	0.45	0.07
1577102	-0.18	0.00	0.05	0.06
1577101	0.45	0.05	0.00	-0.08
1577601	0.07	0.06	-0.08	0.00

Comparison of GRDC-GeoSFM regional dependence for Interbasin 3 obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.4 - GRDC-GeoSFM dependence comparison, Zambezi.

Station ID	1591235	1591231	1591401	1591404	1591410	1591460	1591820
1591235	0.00	-0.05	-0.96	0.08	-0.78	-0.30	-1.10
1591231	-0.05	0.00	-1.02	0.05	-0.77	-0.11	-1.18
1591401	-0.96	-1.02	0.00	-0.82	-1.17	-1.27	-0.64
1591404	0.08	0.05	-0.82	0.00	-0.71	-0.11	-1.26
1591410	-0.78	-0.77	-1.17	-0.71	0.00	-0.69	-0.42
1591460	-0.30	-0.11	-1.27	-0.11	-0.69	0.00	-1.04
1591820	-1.10	-1.18	-0.64	-1.26	-0.42	-1.04	0.00

Comparison of GRDC-GeoSFM regional dependence for the Zambezi obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.5 – GRDC-GeoSFM dependence comparison, Interbasin 5.

Station ID	1159100	1159130	1160881	1196370	1196560	1196600	1159300	1155302	1159650	1159670	1160765	1160820
1159100	0.00	0.08	0.26	0.37	0.35	0.12	0.55	0.46	0.38	0.23	0.31	0.13
1159130	0.08	0.00	-0.45	-0.19	-0.01	-0.05	-0.37	-0.45	-0.51	-0.31	-0.42	-0.05
1160881	0.26	-0.45	0.00	-0.12	0.12	0.10	-0.61	-0.40	-0.14	0.10	0.09	0.32
1196370	0.37	-0.19	-0.12	0.00	0.26	0.25	-0.12	-0.14	-0.01	0.21	-0.03	0.34
1196560	0.35	-0.01	0.12	0.26	0.00	-0.23	0.19	0.23	0.21	0.21	0.07	-0.07
1196600	0.12	-0.05	0.10	0.25	-0.23	0.00	-0.04	-0.02	0.14	0.20	0.14	-0.06
1159300	0.55	-0.37	-0.61	-0.12	0.19	-0.04	0.00	-0.12	-0.24	-0.06	0.08	0.14
1155302	0.46	-0.45	-0.40	-0.14	0.23	-0.02	-0.12	0.00	-0.14	-0.06	0.04	0.12
1159650	0.38	-0.51	-0.14	-0.01	0.21	0.14	-0.24	-0.14	0.00	0.19	0.10	0.29
1159670	0.23	-0.31	0.10	0.21	0.21	0.20	-0.06	-0.06	0.19	0.00	0.22	0.29
1160765	0.31	-0.42	0.09	-0.03	0.07	0.14	0.08	0.04	0.10	0.22	0.00	0.27
1160820	0.13	-0.05	0.32	0.34	-0.07	-0.06	0.14	0.12	0.29	0.29	0.27	0.00

Comparison of GRDC-GeoSFM regional dependence for Interbasin 5 obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.6 – GRDC-GeoSFM dependence comparison, Congo.

Station ID	1593780	1593400	1593600	1593740	1593201	1593770
1593780	0.00	-0.61	-0.60	-0.55	-0.52	-0.58
1593400	-0.61	0.00	-0.45	-0.33	-0.32	-0.35
1593600	-0.60	-0.45	0.00	-0.45	-0.54	-0.48
1593740	-0.55	-0.33	-0.45	0.00	-0.28	-0.29
1593201	-0.52	-0.32	-0.54	-0.28	0.00	-0.29
1593770	-0.58	-0.35	-0.48	-0.29	-0.29	0.00

Comparison of GRDC-GeoSFM regional dependence for the Congo obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.7 – GRDC-GeoSFM dependence comparison, Niger.

Station ID	1234150	1531050	1531420	1134100	1531430	1531450	1531550	1531600	1531800	1837401	1837451
1234150	0.00	-0.52	0.00	-0.48	-0.26	-0.21	-0.57	-0.28	-0.21	-0.47	-0.68
1531050	-0.52	0.00	0.06	0.04	0.06	0.35	-0.99	0.14	0.09	-0.10	-0.13
1531420	0.00	0.06	0.00	0.08	0.15	0.09	-0.66	0.16	-0.35	-0.12	-0.02
1134100	-0.48	0.04	0.08	0.00	-0.10	0.32	-0.96	-0.02	0.21	-0.09	-0.07
1531430	-0.26	0.06	0.15	-0.10	0.00	0.21	-0.94	-0.14	0.48	0.03	-0.07
1531450	-0.21	0.35	0.09	0.32	0.21	0.00	-0.69	0.26	0.38	0.28	0.28
1531550	-0.57	-0.99	-0.66	-0.96	-0.94	-0.69	0.00	-0.60	-0.97	-0.98	-1.04
1531600	-0.28	0.14	0.16	-0.02	-0.14	0.26	-0.60	0.00	0.36	0.01	0.09
1531800	-0.21	0.09	-0.35	0.21	0.48	0.38	-0.97	0.36	0.00	0.05	0.24
1837401	-0.47	-0.10	-0.12	-0.09	0.03	0.28	-0.98	0.01	0.05	0.00	-0.03
1837451	-0.68	-0.13	-0.02	-0.07	-0.07	0.28	-1.04	0.09	0.24	-0.03	0.00

Comparison of GRDC-GeoSFM regional dependence for the Niger obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.8 – GRDC-GeoSFM dependence comparison, Interbasin 9.

Station ID	1531450	1531550	1531600	1531800	1732100	1234150	1531050	1531420	1531430
1531450	0.00	-0.17	-0.35	0.38	-0.20	-0.68	-0.21	-0.09	-0.08
1531550	-0.17	0.00	-0.15	0.32	-0.28	-0.78	-0.20	-0.28	-0.25
1531600	-0.35	-0.15	0.00	0.18	-0.32	-0.72	-0.28	-0.17	-0.18
1531800	0.38	0.32	0.18	0.00	0.46	-0.18	0.38	0.26	0.36
1732100	-0.20	-0.28	-0.32	0.46	0.00	-0.43	-0.19	-0.06	-0.09
1234150	-0.68	-0.78	-0.72	-0.18	-0.43	0.00	-0.66	-0.36	-0.48
1531050	-0.21	-0.20	-0.28	0.38	-0.19	-0.66	0.00	0.06	-0.07
1531420	-0.09	-0.28	-0.17	0.26	-0.06	-0.36	0.06	0.00	-0.20
1531430	-0.08	-0.25	-0.18	0.36	-0.09	-0.48	-0.07	-0.20	0.00

Comparison of GRDC-GeoSFM regional dependence for Interbasin 9 obtained by subtracting the rank correlation between estimated and historical flow. Positive values indicate that the model under-estimate the dependence between the flow behaviour of sub-basins within the same continental basin, negative value the opposite.

Table 2.9 – Archimedean Copulas, summary.

Basin	Kendal Tau	GRDC stations num.	Significant (5%)
Lake Chad	0.42	8	4/8 (50%)
Interbasin 3	0.41	13	11/13 (85%)
Zambezi	0.17	45	42/45 (93%)
Interbasin 5	0.3	64	44/64 (68%)
Congo	0.39	13	3/13 (23%)
Niger	0.43	23	12/23 (52%)
Interbasin 9	0.48	15	5/15 (33%)

Summary of the values for Archimedean Copulas in different basins. The first column indicates the average Kendal Tau amongst the significant copulas, the second the number of stations for which copulas have been modelled and the third the percentage of copulas significant at 5%.

Table 2.10 – Fixed effect regression of historical and simulated streamflow and anomalies, all basins.

	(1) GRDC streamflow	(2) GRDC Anomalies
GeoSFM	0.07	0.20***
	(0.06)	(0.01)
Evap Max	-0.22	0.01**
	(0.21)	(0.00)
Evap Mean	-1.19	-0.00***
	(1.01)	(0.00)
Rain Max	-0.03	-0.02***
	(0.23)	(0.01)
Rain Mean	37.78	-0.00***
	(33.08)	(0.00)
Constant	217.31***	0.16***
	(13.45)	(0.03)
Year dummies	Yes	Yes
Num. of obs.	1065883	1065764

Column 1 presents results for fixed effect regression having as dependent variable the absolute value of streamflow, column 2 those for fixed effect regression having as dependent variables streamflow anomalies. Both include year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.11 – Fixed effect regression of historical and simulated streamflow and anomalies, by basin.

	(1) GRDC streamflow	(2) GRDC Anomalies	Num. of obs.
Lake Chad	0.11	0.54**	6872
Nile	1.17	0.07*	6194
Interbasin 3	0.67***	0.41***	36275
Zambezi	0.12***	0.32***	60023
Interbasin 5	0.05*	0.16***	899257
Congo	-29.69	0.25	6188
Niger	0.17	0.29***	27552
Interbasin 9	0.29***	0.47***	24254

Column 1 presents results for fixed effect regression having as dependent variable the absolute value of streamflow, column 2 those for fixed effect regression having as dependent variables streamflow anomalies. Both include year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.12 – Quantile regression of historical and simulated streamflow, by basin.

	(1) GRDC streamflow	(2) GRDC streamflow	(3) GRDC streamflow	(4) GRDC streamflow	Num. of obs.
Lake Chad	0.11***	0.12***	0.12***	0.11	6872
Nile	1.60***	0.59	1.17***	1.17	6194
Interbasin 3	0.76***	0.49***	0.48***	0.67***	36275
Zambezi	0.21***	0.22***	0.22***	0.12***	60023
Interbasin 5	0.07***	0.06***	0.06***	0.05*	899257
Congo	142.18***	203.77***	193.08***	-29.69	6188
Niger	0.12***	0.14***	0.16***	0.17	27552
Interbasin 9	0.22***	0.22***	0.22***	0.29***	24254

Absolute Streamflow, quantile regressions. Column 1 presents results relative to the first quintile, columns 2 for the second, column 3 for the third and column 4 for the fourth. Reported coefficients are for the simulated absolute flow, all regressions include mean and maximum yearly values for evapotranspiration and rainfall and year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.13 – Quantile regression of historical and simulated anomalies, by basin.

	(1) GRDC Anomalies	(2) GRDC Anomalies	(3) GRDC Anomalies	(4) GRDC Anomalies	Num. of obs.
Lake Chad	0.37***	0.45***	0.48***	0.54**	6872
Nile	0.06***	0.06***	0.07***	0.07*	6194
Interbasin 3	0.51***	0.45***	0.43***	0.41***	36275
Zambezi	0.30***	0.31***	0.32***	0.32***	60023
Interbasin 5	0.19***	0.18***	0.16***	0.16***	899257
Congo	0.47***	0.25***	0.25***	0.25	6188
Niger	0.37***	0.31***	0.30***	0.29***	27552
Interbasin 9	0.54***	0.49***	0.48***	0.47***	24254

Anomalies, quantile regressions. Column 1 presents results relative to the first quintile, columns 2 for the second, column 3 for the third and column 4 for the fourth. Reported coefficients are for the simulated flow anomalies, all regressions include mean and maximum yearly values for evapotranspiration and rainfall and year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.14 – Quantile panel regression with fixed effects of historical and simulated streamflow and anomalies, all basins.

Quantile	0.1	0.2	0.3	0.4	0.5
Streamflow	-4.6	0	0	0	0
Anomaly	0.6***	0.4***	0.2***	0.1***	0.1***
	0.6	**0.7**	**0.8**	**0.9**	**Num. of obs.**
Streamflow	0	0.1***	0.1***	0.2***	1065883
Anomaly	0.1***	0.1***	0.1***	0.1***	1065883

Quantile panel regression with fixed effects. The dependent variable is either absolute historical streamflow (Streamflow) from GRDC or correspondent anomalies (Anomaly), explanatory variable is either simulated streamflow or simulated anomalies. All regressions include year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.15 – Panel probit regression of GRDC-GeoSFM divergences.

	(1) Anomaly difference	(2) Anomaly difference		(1) Anomaly difference	(2) Anomaly difference
Evapotranspiration Anomaly		-0.01***	**XK**	0.29	0.29
		(0.00)		(0.15)	(0.15)
Rainfall Anomaly		-0.18***	**GC/SO/VP**	1.39***	1.41***
		(0.00)		(0.27)	(0.27)
Elevation	-0.00**	-0.00**	**AP**	0.31*	0.30*
	(0.00)	(0.00)		(0.14)	(0.14)
Latitude	0.00***	0.00***	**SG/WD/WE**	-0.03	-0.03
	(0.00)	(0.00)		(0.19)	(0.19)
Longitude	0.00***	0.00***	**BC**	-0.19	-0.20
	(0.00)	(0.00)		(0.21)	(0.21)
Urban and Built-Up Land	0.00	0.00	**AG/BE/BG/DG/FX/QL/RC/SM/ZM**	0.35**	0.34**
	(.)	(.)		(0.14)	(0.14)
Dryland Cropland and Pasture	-0.72***	-0.75***	**BV**	-0.27	-0.28
	(0.23)	(0.23)		(0.26)	(0.26)
Cropland/Grassland Mosaic	-0.69***	-0.71***	**RE**	0.33	0.32
	(0.23)	(0.23)		(0.19)	(0.19)
Cropland/Woodland Mosaic	-0.69***	-0.72***	**AH/GD/LK/WH/WS/XH**	-0.10	-0.11
	(0.24)	(0.24)		(0.16)	(0.16)
Grassland	-0.80***	-0.82***	**AO**	-0.06	-0.07
	(0.23)	(0.23)		(0.17)	(0.17)
Shrubland	-1.11***	-1.13***	**LC/LV**	-0.04	-0.05
	(0.24)	(0.24)		(0.14)	(0.14)
Savanna	-0.74***	-0.76***	**CL/QC/QF**	0.01	-0.00
	(0.22)	(0.23)		(0.13)	(0.13)
Deciduous Broadleaf Forest	-0.63**	-0.65**	**NE/NH**	-0.19	-0.18
	(0.25)	(0.25)		(0.20)	(0.20)
Evergreen Broadleaf Forest	-0.86***	-0.88***	**VC**	-0.02	-0.03
	(0.24)	(0.24)		(0.14)	(0.14)
Water Bodies	-0.52**	-0.53**	**AF/BH/DD/FH/GE/GM/HG/HH/HL/JD/LF/LO/LP/OD/OE/OX/PG/PO/PP/TH/TM/TV/ZO**	0.03	0.02
	(0.23)	(0.23)		(0.13)	(0.13)
Barren or Sparcely Vegetated	-1.08***	-1.11***	**FO**	0.18	0.17
	(0.24)	(0.24)		(0.14)	(0.14)
PF/WX	0.00	0.00	**YY**	0.04	0.02
	(.)	(.)		(0.24)	(0.24)
I	0.15	0.14	**Constant**	0.36	0.38
	(0.15)	(0.15)		(0.26)	(0.26)
			Number of obs.	1065764	1065764

Panel probit regression. The dependent variable is a dummy equal to one if the difference between historical and simulated anomalies is greater than one s.d. and zero otherwise, both regressions include year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.16 – Probit regression of GRDC-GeoSFM daily anomaly correspondence for different values of s.d.

	(1)	(2)	(3)	(4)
Above 1 s.d. (absolute value)	0.71***	0.64***	0.65***	0.67***
Above 1 positive s.d.	0.99***	0.02	0.26***	0.18***
Below 1 negative s.d.	0.77***	0.72***	0.72***	0.74***
Above 0.5 s.d. (absolute value)	0.22***	0.06***	0.05***	0.10***
Above 0.5 positive s.d.	0.42***	0.32***	0.31***	0.44***
Below 0.5 negative s.d.	0.73***	0.68***	0.68***	0.70***

Probit regressions. The dependent variable is a dummy equal to one if the historical daily anomaly correspond to the reported value in term of s.d. , reported values are for a dummy equal to one if the simulated anomaly correspond to the same value. All regression include year dummies, column 2 also include daily values of evapotranspiration and rainfall anomalies, column 3 further includes latitude and longitude, column 4 soil depth and land cover. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.17 – Probit regression of GRDC-GeoSFM monthly anomaly correspondence for different values of s.d.

	(1)	(2)	(3)	(4)
Above 1 s.d. (absolute value)	0.82***	0.75***	0.74***	0.77***
Above 1 positive s.d.	0.88**	-0.08	0.15	0.09
Below 1 negative s.d.	0.87***	0.82***	0.81***	0.84***
Above 0.5 s.d. (absolute value)	0.32***	0.17***	0.15**	0.19***
Above 0.5 positive s.d.	0.53***	0.41***	0.38***	0.51***
Below 0.5 negative s.d.	0.81***	0.77***	0.76***	0.79***

Probit regressions. The dependent variable is a dummy equal to one if the historical average monthly anomaly correspond to the reported value in term of s.d. , reported values are for a dummy equal to one if the simulated anomaly correspond to the same value. All regression include year dummies, column 2 also include monthly average values of evapotranspiration and rainfall anomalies, column 3 further includes latitude and longitude, column 4 soil depth and land cover. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 2.18 – GRDC-GeoSFM daily anomalies correspondence, shares.

	> 1 s.d.	< -1 s.d	> 0.5 s.d	< -0.5 s.d.
Overall	0.99	0.86	0.70	0.80
Lake Chad	0.91	0.74	0.64	0.67
Nile	1.00	0.75	0.59	0.68
Basin 3	0.99	0.87	0.66	0.83
Zambezi	0.97	0.79	0.58	0.72
Basin 5	1.00	0.87	0.72	0.81
Congo	0.83	0.77	0.74	0.71
Niger	0.92	0.80	0.62	0.73
Basin 9	0.95	0.84	0.65	0.79

Share of time in which both the historical and simulated daily streamflow anomalies correspond to the reported level.

Table 2.19 – GRDC-GeoSFM monthly anomalies correspondence, shares.

	> 1 s.d.	< -1 s.d	> 0.5 s.d	< -0.5 s.d.
Overall	0.99	0.88	0.76	0.80
Lake Chad	0.92	0.72	0.65	0.69
Nile	1.00	0.68	0.64	0.64
Basin 3	1.00	0.88	0.67	0.84
Zambezi	0.97	0.79	0.59	0.73
Basin 5	1.00	0.89	0.78	0.81
Congo	0.84	0.75	0.75	0.71
Niger	0.92	0.81	0.64	0.73
Basin 9	0.96	0.85	0.66	0.79

Share of time in which both the historical and simulated monthly streamflow anomalies correspond to the reported level.

Table 2.20– Yearly mean anomaly, hydroelectric production and industrial electricity consumption.

	(1)	(2)	(3)	(4)
	Hydropower prod.		Industrial electricity cons.	
	Installed	Operational	Installed	Operational
Yearly Mean Anomaly	1415.33**	1181.00**	289.52	144.59
	(647.17)	(565.53)	(907.83)	(792.64)
Constant	3456.74***	3460.71***	7426.95***	7427.12***
	(246.95)	(247.22)	(346.41)	(346.50)
Number of obs.	234	234	234	234

Fixed effect regressions. The dependent variable is yearly hydroelectric production in GWh in columns 1 and 2, Industrial electricity consumption in GWh in columns 3 and 4; the dependent variable is the country-wide yearly mean anomaly, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4. All regressions include year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level

Table 2.21– Yearly mean anomaly, other sectoral electricity consumption.

	(1)	(2)	(3)	(4)	(5)	(6)
	Residential electricity cons.		Service electricity cons.		Final electricity cons.	
	Installed	Operational	Installed	Operational	Installed	Operational
Yearly Mean Anomaly	-28.75	165.72	-494.11	-402.79	-327.69	-169.50
	(582.06)	(507.98)	(633.04)	(552.72)	(1778.10)	(1552.30)
Constant	2908.26***	2909.43***	1545.32***	1544.00***	12751.44***	12751.21***
	(222.10)	(222.06)	(241.55)	(241.62)	(678.48)	(678.58)
Number of obs.	234	234	234	234	234	234

Fixed effect regressions. The dependent variable is residential electricity conumption in GWh in columns 1 and 2, service electricity consumption in GWh in columns 3 and 4 and final electricity consumption in GWh in columns 5 and 6; the dependent variable is the country-wide yearly mean anomaly, weighted by installed capacity in columns 1, 3 and 5 and by operational capacity in columns 2, 4 and 6. All regressions include year dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level

Chapter 3

Quality of Electricity Provision and Firm's Revenue: a Solution for the Endogeneity Problem.

3.1 Introduction.

In the first chapter we estimated the firm's revenue elasticity to power outages, finding that the frequency of power outages is negatively and significantly related to it, with a greater effect for firms without generators compared to those firms that have access to backup capacity. However, as we noted in Chapter 1, there are a number of possible sources of endogeneity that might be affecting these results. For example, more productive firms lead to a more profitable tax base, and the subsequent increase in government revenue might lead to improvements in infrastructure spending; or new investments in infrastructure might target zones in which there was previously little incentive to generate new economic activity or conversely where the quality is already high to favour the emergence of economies of scale, hence directly affecting firms' productivity. [77]

These endogeneity concerns will in turn imply that the estimates of the first chapter are biased, and hence in this chapter we take an instrumental variable approach. The latter has been identified in the variation of water available for hydropower production, which is one of the main sources of electricity for the majority of SSA countries. Everything else equal, a lower flow in basins serving hydropower plants will entail a lower capacity to generate hydro-electricity, and if this does not influence demand, the resulting increase in the gap between electricity supplied and demanded should translate in a higher incidence of outages. [78]

[77] For a review of the relevant literature see Chapter 1.
[78] A similar instrument is recently been applied to Indian manufacturing firms (Allcott et al. 2014)

The remainder of the chapter is organized as follow: section 2 explains the procedure used to transform the river flow modelling of chapter 2 into the final instruments, section 3 presents the methodology, section 4 introduces the main results, section 5 discusses a series of robustness checks while section 6 concludes.

3.2 Instrument Construction and Summary Statistics.

3.2.1 Power Plants.

The data about African power plants come from the World Electric Power Plant (WEPP) database from PLATTS, a global provider of energy and commodity information. The database contains information for more than 90,000 plants across the globe, ranging from generation technology and turbine type to installed and operative capacity. With regard to SSA, the WEPP dataset reports information for approximately 1,352 power plants for a total of 3,857 actual generating units. Table 3.1 below reports some of the most relevant summary statistics for each country in our sample.

[Table 3.1 about here]

By looking at the number of plants it is possible to notice the very uneven distribution of electricity infrastructure in the continent, with Nigeria and South Africa accounting for 20% of all power plants. This unevenness is strengthened further if we focus on the actual megawatt (MW) of installed capacity: South Africa alone accounts for 51.58% of the capacity of SSA. Given that Nigeria accounts for 16.61%, this leaves just 31.81% for the remaining 36 countries in the sample. The mean plant dimension is 74.53 MW at a continental level, but of course there is great diversity across countries, from a minimum of 1.32 MW in Burundi to a maximum of 417.13 MW in the Republic of South Africa (both averages).

While the majority of power plants in the continent are oil fuelled (57%), relevant differences are also present in the composition of each country generation portfolio. For the relative majority of countries in the sample (18) the main generation technology, by MW of installed capacity, is hydropower, which is in any case the second also by number of plants (26.4%). Of the 20 remaining countries, 15 have a prevalence of oil in the installed capacity, 3 of coal (Botswana, South Africa, Zimbabwe) and 2 of gas (Cote d'Ivoire, Nigeria). However, hydropower still represents an important source of electricity for many of these countries, as for more than 10 a quarter of the installed capacity consists of hydro, while only 7 are completely without any hydropower production[79]. As the whole identification procedure relies on water availability for hydro generation, these countries will be excluded from the following analysis.

The first step in the instrument creation is to geo-locate the power plants, so to be able to determine which rivers feed into hydro-plants. The WEPP database contains the latitude and longitude coordinates for only 47% of the plants in the study, so that we had to find a way to obtain information about the missing ones. Attempt to contact the relevant ministries and utilities to directly obtain these information were made but, as we did not receive any reply, we had to resort on geo-locating the remainder with Google Earth, using the name of the dam or of the power plant, and when these could not be found the name of the villages/cities in which

[79] These countries are Botswana, Chad, Eritrea, Guinea Bissau, Mauritania, Niger and Senegal.

the plants are located. The procedure increased the share of plants for which we possess coordinates to 80%.

The mean dimension of an un-located plant is 19.01 MW, 50% of them has less than 1 MW and 89% less than 10 MW and of those above 10 MW, 60% are located in countries which are in the top 5 for installed capacity. The distribution of generation technologies across the un-located plant does also closely resemble the overall one, as 54% of them are oil fuelled and 35% hydro fuelled. Considering only the un-located hydro plants, their mean dimension is 14.65 MW, with more than 60% of them smaller than 1 MW and more than 95% smaller than 20 MW (respectively the generally accepted upper boundary for mini and small hydro). Taken together, these facts reassure us that the absence of these un-located plants should not significantly affect our results.

3.2.2 Power Plants & Anomalies.

The next step is to combine information on power plants with the river flow modelling from chapter 2. This means taking all geo-located hydro power plants (264 out of 357) and combining the locations with estimates of the river flow into each plant as predicted by the GeoSFM model. Table 3.2 below summarizes the outcome of this matching process.

[Table 3.2 about here]

The first observation is that we were unable to obtain a complete match as 3.4% (or 9 plants) did not appear to be located on a river. Even in this case though the unmatched hydro plants predominantly belong to the category of mini and small hydro, as only one of them has an installed capacity higher than 20 MW. Again, while it would be undoubtedly desirable to achieve a perfect match, the absence of these plants should not significantly affect the results of the analysis.

Taking the matched hydro plants, the next step is to see how frequently the water available for electricity generation varies and how much it differs from the long run mean on which the plant dimension calculations are normally based upon. As all the information about outages which will be used in the regression analysis are only available as yearly averages, the following summary will also be presented in that form despite the continuous nature of the GeoSFM model.

At a continental level, for the period 2001-2013 the mean number of negative shocks per year was 192.67 with an average magnitude of 0.57 s.d., while the mean number of positive shocks was 95.15 with an average magnitude of 1.11 s.d.. However, these figures hide differences across years and countries. With regard to time variation, the highest value of the positive anomaly[80] index (corresponding to the highest flow at power plants locations) is for 2006, while the highest value for the negative anomaly index (lowest flow) was for 2013. As these are the frequencies of shocks at points where power stations are located, averaged over the whole continent, we do not necessarily expect them to correspond to years of general droughts or floods.[81] Looking instead at country variations, the lowest average of negative shocks in the period is to be found in the Democratic Republic of Congo (DRC), while the highest is that of

[80] Anomalies are defined as the difference between a river daily streamflow and its long term mean expressed in term of standard deviation, see Chapter.2 for further discussion.
[81] For an assessment of the performance of GeoSFM modelling see Chapter 2.

Namibia; on the other hand, the highest mean number of positive shock is that of Gabon, while the lowest average is that of Swaziland. Concerning the mean size of negative shocks, the strongest are those of Gabon (-0.87 long term s.d. in flow) and smallest those of Swaziland (-0.38 long term s.d.); the highest average size for positive shocks is that of Malawi (1.29 long term s.d.), the lowest that of Gabon (0.8 long term s.d.).

Any variation of flow below or above the long run mean will count as a negative or positive shock despite its size, and we know from the analysis in Chapter 2 that many rivers in the continent have a strong seasonal behaviour, which could influence the measure of shocks over a year. Since we only have the average number and hours of outages in a year, to try and take this into account we further divide our measures between strong and weak anomalies, using as a cut-off point the average negative and positive shock size for the basin across the period. This should allow us to at least partially differentiate between the incidences of different types of shocks: anomalies with a magnitude higher (in absolute value) than the average negative anomaly in the basin are defined as "strong negative anomalies"; those with a magnitude between the average and 0 as "weak negative anomalies". Conversely, anomalies with a magnitude between 0 and the average positive anomaly in the basin are defined as "weak positive anomalies" and those with a magnitude higher than the average as "strong positive anomalies".[82]

For the whole of SSA, the average level of strong negative anomaly is 109.44 per year, with the lowest incidence in Malawi and the highest in Guinea and the highest mean value again in

[82] As a robustness check we have also performed the analysis using the median value instead of the average.

Gabon while the lowest in Swaziland[83]; for weak negative anomalies the continental average is 83.24, the lowest incidence is found in Mali (the weakest magnitude in Lesotho) while the highest frequency is that of Namibia (and the strongest magnitude in Malawi). The mean incidence of strong positive anomalies is 39.42 at continental level (55.73 for weak positive anomalies), at a country level the highest incidence is that of Gabon (which also has the highest incidence of weak ones) and the lowest that of Swaziland (Namibia for weak anomalies) while with regard to magnitude the strongest is that of Swaziland (Tanzania for weak ones) and the weakest that of Gabon (Sierra Leone for weak anomalies).

It is worth remembering that the regression analysis will be performed in a cross country setting given the very low availability of panel data for African firms. If we consider only the relevant years, the above picture is only slightly changed. The highest incidence of weak negative shock is still in Namibia while the lowest is now in Togo, while the lowest magnitude is that of Guinea and the strongest is that of Malawi. With regard to strong negative shocks, the highest frequency is that of Togo while the lowest that of Malawi, while the weakest magnitude is again found in Swaziland and the strongest in Gabon. Considering instead weak positive anomalies, the highest and lowest frequencies are again that of Gabon and Namibia respectively (while the strongest and weakest magnitude become that of Togo and Gabon). The highest incidence of strong positive anomalies is that of Rwanda while the lowest becomes that of Namibia, the strongest magnitude that of Rwanda and the weakest that of Gabon.

[83] Both the figures for lowest incidence and lowest magnitude excludes Namibia as all its negative anomalies qualify as weak.

3.2.3 Power Plants and cities.

The final step in the creation of our instrument consists of matching the power plants with the productive centres they serve. In this case, it is the 109 cities included in the sample. This is done by selecting all the closest power plants to each city, but we need to determine a criterion for the selection. As we can see from Table 3.3, the density of power plants around cities varies considerably in the sample, so that Swaziland has all power plants within one hundred kilometres radius from the capital while Sudan has none. While electricity is immediately available for consumption at any point of the grid once it has been generated, given the figures reported in the first chapter about the average condition of transmission and distribution lines across the continent there is little doubt that the probability of a power outage increases with the distance from the power plant, especially in a context in which imports and export of electricity seldom account for much electricity consumption in any given country.

[Table 3.3 about here]

In the analysis that follows we apply four different radiuses (50, 100, 200 and 300 km) to enable us to pick up enough variation in power plants proximity for different countries in the sample. As the density of power plants around any production centre (city) is clearly going to be related to the area of the country, we will use exactly this criterion to determine which radius is going to be applied to which country. We then use the smallest radius for countries which fall in the smallest quartile for area, the second smaller to those in the second quartile and so on.[84] The list of countries by radius dimension is reported in Table 3.4. With regard to the last part of the

[84] To ensure that the results are not driven by the particular radius picked, we also performed the analysis dropping the biggest and smallest radius and using only the central ones.

analysis of Chapter 4, in which we investigate the relation between hydropower production and national economic activity, this step is skipped, and we consider all power plants in a country weighting them as of next paragraph.

[Table 3.4 about here]

Once the production centres have been connected to the plants, the final step is to try and account for the different importance of different plants, as shocks to (hydro) plants with a bigger generation capacity will be more relevant than shocks to plants with a smaller capacity. We then scale the anomaly variables by plant dimension before aggregating them at the city level (or country level for the version used at the end of Chapter 4). As we possess information about both installed and operative capacity, we use them alternatively to see how the results change, but given that we cannot determine if the reported operative capacity in the WEPP database corresponds to the period for which we have outage information we consider the installed capacity as the benchmark for the analysis.

As more or stronger shocks hit the basins on which hydro-plants are located, there will be *ceteris paribus* less electricity generated, which should in itself increase the number and duration of outages. We have already shown at the end of Chapter 2 that our instrument is not related to the electricity consumed by any sector of the economy, as the exclusion restriction implies the absence of reverse causality between the level of output and the level of water available for hydro-production. Generally, the existence of a structural relationship between this two variables seems to be highly unlikely, if not maybe for producers of hydro-turbines or of electric generators. As both of these activities fall under the industrial category "Machinery and Equipment” which represents 1.47% of the overall sample, and at most 3.74% in any given country, this objection does not seem to be relevant in our case.

3.3 Methodology.

Our estimation strategy is to instrument both outage measures used in Chapter 1 with some combination of the information about the water available for hydro-generation using the following:

$$Y_i = \alpha + \beta_1 X_i + \beta_{ji} Z_i + \varepsilon_i$$

where Y_i is the logarithm of total sales, X_i is the instrumented version of one of our two measures of outages (log-numbers or log-hours) and Z_i is an array of usual control variables (size dummies, exporter status, age, structure of ownership and access to credit). As in the first chapter, country dummies are included in all specifications, while we consider both specifications with industry dummies and with a simple manufacturing dummy, as including only the latter allows us to increase the sample size with the addition of a small number of countries. Finally, we cluster standard errors at the city level.

The main challenges to the identification procedure is the need to explain firm-level variation in outages using city-level variation only (as the absence of more firm data prevent us from using a panel framework) and to account for the high variability of water availability throughout the year using a single average measure, which we are forced to employ given the structure of the questionnaires. We are ideally proxying for hydroelectricity production, and we have shown in Chapter 2 that the instrument is indeed correlated with actual hydropower production at the national level for the sub-sample of countries for which data is available. Although it would also be preferable to further assess the performance of our instruments at a finer level (i.e. for some specific hydro-plant), these data are not available. We cannot but further stress the necessity of improved data collection for SSA, ranging from firms to energy information.

It is hard to decide ex-ante which forms of the instruments are to be preferred. Neither the average value of the anomalies throughout the year, nor the frequencies or the magnitudes of different kinds of shocks alone will fully capture the effect on hydro generation: frequent but weak negative shocks can be more problematic for electricity generation than stronger and rare ones and the same logic stands for excessive flow which might damage hydro plants. We have hence opted for presenting two different sets of results: the first uses the simplest possible version of the instrument, that is the average yearly value of anomalies; the second uses instead the interactions between frequency and magnitude of strong positive and negative anomalies and that of weak negative anomalies[85], which can be seen as relative indexes of the hydro-shocks, as greater absolute values imply more relevant obstacle to hydro-generation. [86] To show that the estimates are robust to different combinations of hydro variables we will also present a series of alternative first stage specifications as robustness checks.

[85] We have decided to exclude the index for weak positive anomalies as excess flow of that magnitude is unlikely to cause any issue to hydro generation. A specification including is presented amongst the robustness checks.
[86] This is because a higher absolute value is connected with a higher frequency, a stronger magnitude or both.

3.4 Results.

We start the discussion of the results by presenting in Table 3.5 estimates for an OLS regression on the final sample. The model estimated corresponds to that of Chapter 1 but all countries without any installed hydro-power capacity have been excluded, implying a decrease of roughly 1,600 observations.

[Table 3.5 about here]

As can be noted from Table 3.5, this exclusion leads to a loss of significance for the coefficients on the (log) hours of outages but does not alter the results for (log) numbers: a negative effect of power outages on firms' sales can already be noted in the overall sample, the coefficient on outages is still greater for firms without back-up capacity, while this time no significant association between the two variables can be found for firms with a generator. Table 3.6 and Table 3.7 present then estimates for the first stage of 2 stages least square (2SLS) regressions in which our outage measures have been instrumented either with the average yearly value of the anomalies (Table 3.6) or with the indexes for weak and strong negative anomalies and for strong positive anomalies (Table 3.7). In both cases the instruments have been constructed by using all 4 radiuses and weighted by the installed capacity. Columns 1 and 2 of both tables refer to the regressions run on the whole sample, columns 3 and 4 to those run only on the firms without generators while columns 5 and 6 to those run only on firm with available backup.

[Table 3.6 and 3.7 about here]

As it is possible to notice from Table 3.6, the yearly average value of the anomalies is highly statistically significant and with the expected sign, but only in the regressions run on the firms without generators. On the other hand, from Table 3.7 we can see that when we include the three anomaly indexes some of the instruments are always significant, regardless of the outage measure used and of the sample on which we run the regression. As we expected, the yearly average anomaly takes a negative sign (the more water available for hydro production, the less outages); on the other hand, the strong negative anomaly index should be negative (as it is upper bounded to the average negative shock) and the positive should be positive (as greater values correspond to stronger excess flows). Despite this, the fitted values for both outages measures have the expected signs and are strongly correlated with the observed ones regardless of the instrument combination (0.68 with regard to log-number and 0.61 with regard to log-hours), suggesting that we are indeed picking up what we intended to and that the problems with signs might be related to some collinearity between our instruments.

Tables 3.6 and 3.7 also report a series of test statistics. In the first row the test statistic of the Wooldridge Score Test, which can be used to assess the presence of endogeneity between regressor and dependent variable when the s.e. are clustered, is reported; the second row reports the Stock- Yogo F-Statistic, a minimum eigenvalue statistic elaborated from the work of Cragg and Donald (1993) and of Shea (1997) by Stock, Wright and Yogo (2002) and Stock and Yogo (2005) which aims at assessing the strength of the instrument. The authors propose a definition of weak instrument connected to either the acceptable bias of the instrumental variable estimate relative to the OLS one (which is equal to its absolute bias in a case with a single endogenous regressor such as ours) or to the bias in the size of the confidence interval. Critical values for assessing such biases are also provided, although they are calculated for i.i.d. error terms, which is not the case of our model as error terms are clustered at the city level. As critical values for non i.i.d. cases have never been developed, we present them as they furnish at least a crude comparison. When the yearly average anomaly is used, a minimum eigenvalue of 20.29 (third column of Table 3.6) implies a maximum bias of less than 10% (see Table 3.8 for the critical values in the case of one endogenous regressor and one exogenous instrument) and would be acceptable also for a situation in which i.i.d. errors could be assumed. On the other hand, considering the case in which the anomaly indexes are used, a minimum eigenvalue of 4.39 (third column of Table 3.7) implies an absolute bias of the IV estimate of at least 25% (see Table 3.9), so that we cannot exclude that the combination of these instruments is weak.

There are two ways to take this into account in the following analysis. The first is to use tests which have the correct size even in presence of weak instruments, such as the Anderson-Rubin statistic (Anderson and Rubin 1949) and the conditional likelihood ratio (CLR, Moreira 2003, Andrews, Moreira and Stock 2007), both of which jointly test the structural parameters and the over-identification restriction. These are in turn the fourth and the fifth row of Table 3.6 and Table 3.7, while the third is a test of under-identification. The second way is to move from a 2SLS setting to a Limited-Information Maximum Likelihood (LIML) one, which has been shown by the authors (Stock, Wright and Yogo 2002, Stock and Yogo 2005) to perform better in cases in which more than one instrument is used (such as our). In Table 3.9 we report the critical values for one endogenous regressor and three instruments in case of i.i.d. errors for both 2SLS and LIML. The behaviour of the critical values, although relative to an i.i.d. case, gives at least a rough indication of why the second method is to be preferred.

[Table 3.8 and 3.9 about here]

We can now consider the results of the second stages of the 2SLS regressions. In Table 3.10 those where the instrument is the average yearly anomaly are presented, while Table 3.11 we use instead the three anomaly indexes. As all coefficients other than those on the outages measures are almost identical to the OLS case regardless of the instrument choice, the main differences between instrumented and non-instrumented estimates are twofold: first, both outage measures are now insignificant when the whole sample is considered; second, both coefficients are now significant for firms without generator and of much greater magnitude than in the OLS case. At the same time though, results for the endogeneity, under-identification and over-identification test for the regression with a single instrument (first two columns of Table 3.6 referring to those of Table 3.10), while only the result for the endogeneity test for the case with the three disaggregated hydrological measure (first two columns of Table 3.7 referring to those of Table 3.11), suggest that the OLS estimates are more efficient than the 2SLS ones. That is, we do not find direct confirmation that outages are endogenous to productivity in the whole sample, so that we can conclude that they have an overall negative effect, at least when measured in numbers. On the other hand, when we look at all test results for firms without generator, these lead us to prefer 2SLS estimates over their OLS equivalent, confirming that there are indeed some endogeneity issues in this relationship. Furthermore, comparing Table 3.10 with Table 3.11 it seems that the presence of weak instruments in the second table biases the coefficients for firms without generator downwards, as both coefficients on log-numbers and log-hours are higher when the average yearly anomaly is used than in the specification with the three indexes. This is further confirmed by Table 3.12, in which we estimate the same model of Table 3.11 using LIML, which is proven to be more robust to the presence of weak instruments.

[Tables 3.10 to 3.12 about here]

As it can be seen, the LIML regression of Table 3.12 also points toward a downward bias in the 2SLS estimates when all indexes are used, as the coefficients for both measures of outages remain significant at the same confidence level and are slightly bigger in magnitude, while again no other coefficients exhibit any relevant change.

It appears then that once the possible endogeneity between the quality of electricity service and firms revenue is taken into account, the effect on firms which do not have access to back-up generation is much stronger than OLS estimates would suggest, while, as we do not find confirmation in the data that this relationship is endogenous for the whole sample, a much weaker effect is still perceivable on the overall sample. Using again the example of the first chapter, a reduction in the average hours of outage for the average firm without a generator in the sample to the level of the average South African firm (corresponding to a reduction of 79.3%) will entail an increase of revenue of 77% as opposed to 3.2% (corresponding to roughly 16 mln $ at 2005 PPP). The coefficient on the number of outages allows us instead to compare the effect between the overall sample (for which OLS estimates are preferred) and only for firms without generator (for which 2SLS estimates are preferred): in this case, a reduction of 70% in their number (roughly the difference between an average firm and its South African counterpart) entails an increase in revenue of 5.6% in the overall sample but of almost 96% for a firm without generator.

3.5 Robustness Checks.

3.5.1 Instrument form

The first robustness check regards the actual choice of the instrument combination. As previously noted, a few different options were taken into account when deciding which combination was going to become our baseline, and we eventually opted for the simple yearly average and for the indexes for weak and strong negative anomalies and that for strong positive anomalies. To show that the results are not driven by this choice we present in Table 3.13 a summary of the different considered combinations, all weighted by the installed capacity and constructed using the four different radiuses. We include both the second stage coefficients of the outage measures from 2SLS regressions and the relevant first stage tests. The first row combination includes the yearly average anomaly, as in our main specification, together with its standard deviation; the second the magnitude of average positive and negative anomalies, their frequency and both indexes (without differentiating between weak and strong); the third and fourth rows correspond to the second but they consider only the weak and strong anomalies respectively; the fifth includes the positive and negative indexes alone without differentiating between weak and strong; the sixth and seventh correspond to the fifth but considering only the weak or the strong anomalies respectively while the eighth corresponds to our other baseline specification but also includes the index for weak positive anomalies.

[Table 3.13 about here]

As it can be seen, the estimates for the number of outages never changes sign or magnitude considerably and remain significant in all but one case (when the strong positive and negative indexes are used together), although once it is significant only at 10%. The coefficients on the hours of outage remain significant in all but two cases (the same combination for which the number of outage is also insignificant plus the one in which we instrument with the frequency, magnitude and indexes of strong positive and negative anomalies), and never considerably deviates from the value of the baseline regression. In the same way, in all cases in which estimates are significant we pass all relevant tests, except once it which the endogeneity of log hours of outage is rejected.

[Tables 3.14 and 3.15 about here]

Successively, Table 3.14 and Table 3.15 present both the first and second stages regressions obtained by substituting the industry dummies with a simple manufacturing dummy, which allows us to expand the analysis to Benin, the Central African Republic, Guinea, Malawi, Rwanda, Swaziland and Togo. Even in this case it can be noted that the results do not vary substantially from the baseline specifications, possibly also because the majority of the firms which enter the analysis possess a generator. Similarly, switching the weight from the installed capacity to the operational capacity does not alter significantly any of the estimates, neither those obtained by instrumenting with the average yearly anomaly (Table 3.16) nor those obtained with the three indexes (Table 3.17). In neither case using LIML instead of 2SLS brings any relevant change to the estimates, which are only slightly bigger in magnitude (for these and all the following robustness checks we will refer to the results of LIML regressions which are reported in the appendix A3).

[Tables 3.16 and 3.17 about here]

Switching the number of radiuses used to connect the cities to the power plants from four to two (i.e. dropping the two extremes of 50 and 300 km using instead 100 and 200 km radiuses respectively) has a more noticeable effect, as now the coefficients on the (log) hours of outage is significant only at 10% for both instrument combinations (yearly average anomaly on its own in Table 3.18 and the three indexes in Table 3.19). This is coupled with the failure of the Kleibergen-Paap under-identification test for these regressions, as both combinations of instruments become insignificant in the first stage. Again, it is worth noticing that the results remain consistent when LIML is used instead of 2SLS, with the only difference being that the spurious coefficients on the hours of outage for firms without generator becomes now insignificant. The number of observations is also lower than in other regressions as there are now cities for which no hydro-power plants are present within the used radiuses.

[Tables 3.18 and 3.19 about here]

The next robustness checks use instead only one of the two intermediate radiuses. Table 3.20 and Table 3.21 present the results when all cities are connected to power plants within a 100 km (instrumenting with the yearly average anomaly and the three indexes respectively), in Table 3.22 and Table 3.23 the 200 km radius is used instead. In all these cases the number of observations included in the regressions is different from the main specifications, due to either cities not having any hydro-power plants in the selected radius (100 km) or now having at least one while previously none (200 km). The significantly smaller number of observation (3777 instead of 5048) included when we use the 100 km radius only is probably behind the loss of significance of the estimates of Table 3.20, in which the yearly average anomaly is used as instrument. When the three indexes are used instead (Table 3.21) the log-number of outages remains significant for firms without generator and furthermore becomes now significant at 10% also in the overall sample, although the endogeneity test is passed only in the regression for firms without generator, while log-hours are never significant.

As the number of cities associated with at least one hydro-power plant grows, the results obtained by using only the 200 km radius are already closer to the baseline specifications. Regardless of the combination of instruments used, the estimates for the number of outages always remain significant for firms without any available backup, while the estimates for log-hours are insignificant when instrumented with the yearly average anomaly (Table 3.22) and significant, although only at 10%, when instrumented with the three indexes (Table 3.23).

[Tables 3.20 to 3.23 about here]

The following robustness check entails instead the inclusion amongst the explanatory variables of the average level of sales in the different cities, to verify that the results are not driven by some unobservable effect at the city level correlated with our instruments (Table 3.24 and Table 3.25, yearly average anomaly and three indexes respectively). The newly included variables are always positive and significant, indicating that there might indeed be some further city level effect which is relevant for firm revenues. At the same time the sign and significance of the effect of outages does not change, although there is a decrease in their magnitude. A difference between the instrument combinations is perceivable: in the estimates obtained with the yearly average anomaly, the endogeneity test confirms that 2SLS has to be preferred over OLS, while the opposite stands true for the estimates obtained with the anomaly indexes, probably due to their weak relationship with the instrumented variable.

[Tables 3.24 and 3.25 about here]

Table 3.26 and Table 3.27 present the second-last robustness check, which consists in the inclusion of the long term values of our instruments amongst the explanatory variables (those used as instruments refer instead to the year for which firms data are available) as further way to verify the exclusion restrictions. In these cases, none of the newly included variables are even close to significance, while little changes interest the main estimates. The coefficients on log-number of outages remain significant regardless of the combination of instruments used in the regressions on firms without generators and become now significant at 5% also in the overall sample; the estimate for the log-hours of outage remains significant when the yearly average anomaly is used as instrument but becomes insignificant when the three indexes are used.

[Tables 3.26 and 3.27 about here]

Finally, Table 3.28 presents the results for the case in which we use as cut-off point between weak and strong anomalies the median anomaly in the basin instead of the average one. As it can be seen, also in this case the estimates are widely consistent with all those previously presented, and they remain so if the estimation procedure is changed to LIML.

[Table 3.28 about here]

3.5.2 Sample Selection Bias.

Given the high frequency of black-outs across SSA and the existence of some evidence that firms rarely achieve complete back-up (Oseni and Pollitt 2013), the finding of an insignificant effect of outages on firms owning a generator might depend on the fact that we are ignoring how firms self-select in this category, that is a sample selection bias is present. In this section of the robustness checks we will then try to control for this possible bias applying the model developed by Heckman (Heckman 1976, Heckman 1979, Heckman *et al.* 1999). The procedure consist of two steps: first, estimating a probit model to obtain the probability of being selected in the sample (the selection equation), to then include the inverse Mills ratio[87] from this regression amongst the explanatory variable in the equation of interest (the outcome equation) to control for selection bias.

There are two different identification criteria for sample selection issues in the Heckman model: one can either rely only on the nonlinearity of the inverse Mills ratio or include some instruments in the selection equation which should affect the probability of selection but not the outcome of interest. The latter identification strategy is usually preferred, as the nonlinearity

[87] The Inverse Mills Ratio is the ratio between the probability density function and the cumulative density function of function.

of the inverse Mills ratio crucially depends on the normality of the link function in the initial probit, although the inclusion of non-theoretically justified variables as sources of identification is also debated (Leung and Yu 2000, Sartori 2003).

In our specific case, it has proven hard to individuate a variable which influence generator ownership but not firms' sales, as the probit models for ownership presented in chapter 1 include the same regressors of the sales analysis. Three candidates which might satisfy the exclusion restriction have been individuated: the mean value of the anomaly indexes used as an instrument in the third chapter over the 5 years preceding the WBES[88]; dummies for firms finding their expansion constrained by the quality of the electricity supply or by their access to finance and these last two variables as averages in the country instead that firm-level values. As we have seen in the robustness check of the third chapter, long-term hydrological measures are not a significant predictor of sales, and their past values should influence the choice of generator ownership. The other two variables are obtained from a question in the WBES in which firms are asked to decide the most relevant constraint for expansion from a series of different choices. As they are given the choice to answer that the question does not apply in their case, and only 0.7% of firms in the sample responds in that way, there should not be any direct connection between the level output and which particular obstacle firms' individuate as the most relevant. To see if this is not the case, we will also run the models using the average share at the country level instead that the firms' response.

We will now present the results, all of which have been obtained through Maximum Likelihood estimation with either robust standard errors (for the case with no exclusion restriction and

[88] 5 years is as far back as we can go with the hydrological data for the older firms' data included in the analysis.

when firms' responses were included), standard error clustered at the city level (for mean value of the three lagged indexes) or at the country level (for the two specifications using country level shares).

[Table 3.29 about here]

As it can be seen from the table, the coefficients for the selection equation are fairly similar to those obtained from the probit in the first chapter, allowing for the fact that countries without any installed hydropower are now excluded from the analysis, so that the sample mirrors that of the previous sections. The overall number of observations is that for the probit selection equation, while the uncensored observations are those on which the outcome OLS equation is run. The Rho coefficients represent the correlations between the errors of the selection equation and those of the outcome equation, if they are significantly different from 0 than selection bias is present in the model, and this is the case in Table 3.29. Although it is peculiar that the correlation between the errors is negative in the case of the (log) numbers of outage and positive in the case of (log) hours, this might depend on the fact that no exclusion restriction is applied to the model. At any rate, it would seem that once selection bias is accounted for, the coefficient for number of outages becomes significant also for firms without generator, while that for hours remains insignificant.

To see if the results are driven by the absence of an exclusion restriction we now turn on the different specifications of the model. First, in Table 3.30 we use as selection variables the average lagged anomaly indexes. We would expect both negative indexes to be negative (the higher their values, the nearer to the historical mean the streamflow was, hence the lower the incidence of outages) while the strong positive index should be positive (as it is proxying for floods).

[Table 3.30 about here]

As it can be seen from the table, only the strong negative anomaly index is significant in the selection equation, and with the opposite sign from what expected, while all other variables maintain the same sign and significance of the previous specification. Also in this case both Rho coefficients are highly statistically significant, but contrary to the previous one they are now both positive. This leads both measure of outages to be insignificant in the outcome equation, while the only relevant change amongst the covariates is the now significant exporter dummy in the regression on the number of outages.

[Table 3.31 and 3.32 about here]

The next two selection models will instead use for the exclusion restriction the firm-level answer to which obstacle was more relevant in constraining their expansion. Specifically, these will be two dummy variables, either taking value 1 if the answer was "quality of the electricity supply" (Table 3.31) or value 1 if the answer was "access to finance" (Table 3.32)[89]. Given the analysis in the first chapter, we would expect firms which find the quality of electricity supply the main obstacle for expansion to have a higher chance of acquiring a generator, while those

[89] The slightly smaller number of observations included in these two models is due to missing observations in the question of which was the major obstacle (2%).

which are constrained by finance to have a lower likelihood. As we can see from Tables 3.31 and 3.32, both variables have the expected sign and significance, while all other covariates are widely comparable to previous specifications. Once again the Rho coefficients are highly statistically significant, confirming the presence of selection bias. However, they take the opposite signs in the two models: positive in Table 3.31 and negative in Table 3.32. Once again this lead to diverging results in the outcome equation, as both outage coefficients are insignificant when being constrained by electricity is used in the probit model, while they are both significant if being constrained by finance is used.

[Table 3.33 and Table 3.34]

To see if the previous results might be due to a violation of the exclusion restriction, the final two models use the share of firms in the country being constrained by electricity (Table 3.33) or by finance (Table 3.34) instead that the firm-level values. We would still expect these variables to take the same values as before, and indeed we can see in both tables that they do, as do all other covariates. Again, the significance of the Rho coefficients points toward the presence of selection bias, but in both cases there are difference from the previous versions in which the firm level value were used. In both Tables 3.33 and 3.34 we can see that to a negative correlation between the errors of the outcome and the selection equation correspond a negative and significant effect of the number of outages. This was insignificant in Table 3.31 when the firm level value was used, associated with a positive correlation between the errors, exactly as it is now for hours of outage in Table 3.34 (contrary to what was happening with the firm level value in Table 3.32).

We have then see that, although we can conclude that issues of selection bias are present when analysing the effect of outages on firm with generators, the direction of the bias depends to a great extent on which exclusion restriction are used to identify the model. All models give significant Rho values, corresponding to a correlation between the error terms of the selection and outcome equations, but the coefficients for outages are significant in the latter only when this correlation is negative. This happens in 4 out 6 cases with regard to the number of outages and only in 1 out of 6 cases with regard to the number of outages, so that further work on the variable to be used for the exclusion restriction is required before being able to convincingly identify the direction of the bias.

3.6 Conclusions.

As already noted, while there are surely many other elements which constrain the industrial development of many African states, the poor quality of the electricity infrastructure is without doubt one of the key issues that policy makers will have to tackle to spur significant structural transformation processes. As infrastructure upgrades are always expensive and many African economies have been historically financially constrained, it is paramount to be able to assess which are the expectable gains. This chapter focuses on one of the most relevant gains which will accrue from a more stable supply of electricity, namely the increase in firm revenue.

Building on the previous two chapters, we are able to resolve the endogeneity issue which biased our initial estimates thanks to an instrument built to account for the variation in water available for hydro-power production. Once we switch from an OLS to a 2SLS framework, the true relevance of the drag represented by frequent power outages on the production of firms without a generator manifests itself. If the hours of outage of the average firm without a generator in the sample (562 hours) could be reduced to that of the average South African firm in the sample (118 hours), this would entail an increase in sales of 77%, or roughly 16 million 2005 international dollars. As we find evidence of an endogenous relationship between outages and firm productivity only for firms without access to back-up generation capacity, the results of the OLS estimates for the whole sample remain valid, indicating that, although of a much lower magnitude, electricity shortages are a perceivable issue also for the overall economy. No significant effect of outages could be individuated for the sub-sample of firms with access to in-house generation in either OLS or 2SLS regressions. Although from the modelling performed in the robustness check part there is evidence that this might depend on sample selection issue, no clear indication of the direction of the bias has emerged. As this might depend on the variable used to satisfy the exclusion restriction of the selection equation, not easy to find in our case, we intend to better address the issue in future research.

The main policy suggestion which stems from the chapter is then that African states who have enough available funds should continue investing in upgrading their energy infrastructure, as this will lead to a general increase in profitability for all firms in their economy, and a much more significant one for those which cannot access back-up generation, constituting 45.7% of the sample.

Appendix A3.

This appendix includes the LIML regressions equivalent to those reported in Tables 3.15, 3.17, 3.19, 3.21, 3.23, 3.25, 3.27 and 3.28.

[Table A3.1 to A3.8 about here]

Table 3.1 – WEPP information by country.

	Number of power plants	Installed Capacity (MW)	Main generation technology	Share of HP	Mean plant dimension
Angola	67	1,319.41	Hydropower	66.84%	19.69
Benin	17	104.194	Oil	0.48%	6.13
Botswana	9	244.238	Coal	0%	27.15
Burkina Faso	46	275.061	Oil	13.06%	5.98
Burundi	32	42.283	Hydropower	82.93%	1.32
Cameroon	27	1,032.86	Hydropower	71.71%	38.25
Central African Republic	18	42.558	Oil	44.03%	2.36
Chad	38	187.562	Oil	0%	4.94
Congo	11	196.408	Hydropower	50.41%	17.86
Cote d'Ivoire	14	1,378.45	Gas	47.66%	98.46
DRC	90	2,631.66	Hydropower	97.66%	29.24
Eritrea	7	167.346	Oil	0%	23.91
Ethiopia	60	932.373	Hydropower	74.20%	15.54
Gabon	43	401.579	Hydropower	42.42%	9.34
Ghana	29	2,473.66	Hydropower	48.62%	85.3
Guinea	21	929.52	Hydropower	70.59%	44.26
Guinea Bissau	3	33.115	Oil	0%	11.04
Kenya	65	1,917.90	Hydropower	52.62%	29.51
Lesotho	14	79.546	Hydropower	97.69%	5.69
Liberia	24	528.407	Oil	32.15%	22.02
Malawi	18	337.161	Hydropower	84.44%	18.73
Mali	28	602.304	Oil	43.11%	21.51
Mauritania	27	236.793	Oil	0%	8.77
Mozambique	27	2,439.33	Hydropower	89.52%	90.35
Namibia	12	450.76	Hydropower	57.68%	75.13
Niger	13	135.9	Oil	0%	10.45
Nigeria	134	16,791.68	Gas	31.49%	125.31
Rwanda	25	67.906	Oil	41.59%	2.72
Senegal	54	787.066	Oil	0%	14.58
Sierra Leone	22	207.934	Oil	26.93%	9.45
South Africa	125	52,141.71	Coal	4.35%	417.13
Sudan	80	1,961.10	Oil	33.74%	49.03
Swaziland	16	132.093	Hydropower	48.82%	8.26
Tanzania	92	1,447.53	Hydropower	44.10%	15.73
Togo	8	251.64	Oil	26.75%	31.46
Uganda	30	1,094.18	Hydropower	77.89%	36.47
Zambia	32	1,906.61	Hydropower	92.04%	59.58
Zimbabwe	20	4,862.55	Coal	30.46%	243.13

Selected summary statistics for Africa power plants from the PLATT WEPP database. The first column presents the number of power plants, the second the total installed capacity in MW, the third the main generation technology in the country by MW of installed capacity, the third the share of MW of installed capacity corresponding to hydro power, the fourth the mean plant dimension in term of MW.

Table 3.2 – Hydropower plants in WEPP.

	WEPP	WEPP, geo-located	Joined with GeoSFM
Angola	28	25	24
Benin	1	1	1
Burkina	4	3	3
Burundi	28	21	21
Cameroon	7	6	5
Central African Republic	3	3	3
Congo	3	3	3
Cote d'Ivoire	5	5	5
DRC	49	37	37
Ethiopia	14	9	9
Gabon	6	5	2
Ghana	2	2	2
Guinea	8	6	6
Kenya	23	17	17
Lesotho	8	4	4
Liberia	2	2	2
Malawi	7	5	5
Mali	5	2	2
Mozambique	6	6	6
Namibia	2	1	1
Nigeria	13	6	6
Rwanda	15	9	9
Sierra Leone	2	2	2
South Africa	26	20	19
Sudan	5	5	5
Swaziland	8	6	6
Tanzania	41	30	28
Togo	2	1	1
Uganda	10	6	5
Zambia	13	11	11
Zimbabwe	11	5	5
Total	357	264	255

The first columns present the number of hydro power plants present in the WEPP PLATT database, the second the number of hydro power plants for which coordinates have been obtained, the third the number of power plants which coordinates correspond a basin in the GeoSFM model.

Table 3.3 – Power plants and radiuses.

	50 km	100 km	200 km	300 km	Country size, Km^2
Angola	0	6	12.58	24.98	1,246,700
Benin	0	8	9	9	114,763
Burkina	1.66	5.97	18.57	29.82	274,200
Burundi	**7.24**	17.96	22	22	27,384
Cameroon	1.97	5.66	10.77	19.23	475,442
Central African Republic	1	3	5.52	7	622,984
Congo	3.33	4	5.67	7	342,000
Cote d'Ivoire	4.04	4.83	6.11	12.07	322,463
DRC	2.31	2.96	5.03	8.32	2,345,409
Ethiopia	3.08	4.89	11.15	17.18	1,104,300
Gabon	3	4.41	11.41	18.77	267,667
Ghana	4.57	6.79	12.56	20.56	238,535
Guinea	2.48	5.52	12.52	14.04	245,836
Kenya	3.32	11.71	22.78	29.96	581,309
Lesotho	**2**	2	8	8	30,355
Liberia	**3.84**	6.42	8.99	14.21	111,369
Malawi	4.05	5.79	7.29	9.68	118,484
Mali	1.65	1.65	5.97	11.03	1,240,192
Mozambique	2.24	2.85	4.47	5.18	801,590
Namibia	2.3	2.3	2.18	3.73	825,615
Nigeria	0	0	36.53	45.65	923,768
Rwanda	**3.11**	9.32	13	13	1,221,037
Sierra Leone	**3**	3.56	7.56	10	26,338
South Africa	7.76	14.28	21.86	28.7	71,740
Sudan	0	0	11	16	1,886,068
Swaziland	**5.69**	13	13	13	17,364
Tanzania	4.47	4.89	11.16	16.36	947,303
Togo	**5**	5	6	6	56,785
Uganda	4.7	7.64	9.79	15.33	241,038
Zambia	4.54	5.55	7.35	10.5	752,618
Zimbabwe	1	1.25	1.75	5.5	390,757

The first column present the average number of power plants present in a 50km radius from the city, the second in a 100 km radius, the third in a 200km radius, the fourth in a 300 km radius while the fifth presents the country dimension in squared kilometres.

Table 3.4 – Country quartiles and radiuses.

1st Quartile: 50 km	2nd Quartile: 100 km	3rd Quartile: 200 km	4th Quartile: 300 km
Swaziland	Benin	Congo	Namibia
Rwanda	Malawi	Zimbabwe	Nigeria
Burundi	Ghana	Cameroon	Tanzania
Lesotho	Uganda	Kenya	Ethiopia
Togo	Guinea	CAR	RSA
Sierra	Gabon	Zambia	Mali
Liberia	Burkina	Mozambique	Angola
	Cote		Sudan
			DRC

Countries by dimension quartile with associated radius.

Table 3.5 – OLS baseline estimation results.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.08**		-0.12**		-0.07	
	(0.03)		(0.05)		(0.06)	
Hours of PO		-0.03		-0.03		-0.03
		(0.02)		(0.03)		(0.03)
small	-1.35***	-1.35***	-1.37***	-1.37***	-1.20***	-1.20***
	(0.07)	(0.07)	(0.10)	(0.10)	(0.10)	(0.10)
large	1.07***	1.06***	1.31***	1.30***	0.74***	0.73***
	(0.12)	(0.12)	(0.19)	(0.19)	(0.16)	(0.16)
very large	1.79***	1.79***	1.20***	1.19***	2.04***	2.04***
	(0.18)	(0.18)	(0.31)	(0.31)	(0.23)	(0.23)
exporter	0.66***	0.67***	0.56***	0.57***	0.60***	0.61***
	(0.10)	(0.10)	(0.14)	(0.14)	(0.14)	(0.14)
Credit	0.54***	0.54***	0.56***	0.56***	0.40***	0.41***
	(0.07)	(0.07)	(0.10)	(0.10)	(0.11)	(0.11)
Share	0.37***	0.37***	0.43***	0.44***	0.15	0.14
	(0.10)	(0.10)	(0.16)	(0.16)	(0.13)	(0.13)
Foreign ownership	0.97***	0.97***	0.76***	0.76***	0.93***	0.93***
	(0.10)	(0.10)	(0.17)	(0.17)	(0.14)	(0.14)
Firm age	0.38***	0.39***	0.27***	0.27***	0.46***	0.46***
	(0.04)	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)
Constant	13.81***	13.67***	14.57***	14.30***	13.68***	13.56***
	(0.32)	(0.31)	(0.62)	(0.59)	(0.47)	(0.43)
Number of obs.	5048	5048	2116	2116	2407	2407

OLS estimation with robust standard errors in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.6 – First stage regression, baseline specification, single instrument.

	(1) All Number	(2) All Hours	(3) No generator Number	(4) No generator Hours	(5) Generator Number	(6) Generator Hours
Yearly Mean Anomaly	-0.75	-0.33	-2.34***	-3.31***	0.05	0.55
Wooldridge Score Test	0.09	0.11	9.82***	10.38***	0.46	0.45
F-Statistic (Stock - Yogo)	2.3	0.15	20.29	13.96	0	0.34
Kleibergen-Paap Wald statistic	2.36	0.16	21.16***	14.56***	0	0.37
Anderson-Rubin Wald Chi²	0.12	0.12	12.28***	12.28***	0.47	0.47

First stage regression for the baseline specification. The dependent variables are either the log number of outage per year or the log hour of outage per year. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.7 – First stage regression, baseline specification, multiple instruments.

	(1) All Number	(2) All Hours	(3) No generator Number	(4) No generator Hours	(5) Generator Number	(6) Generator Hours
Strong Negative Anomaly Index	0.11**	0.15**	0	-0.02	0.16***	0.21***
Strong Positive Anomaly Index	0	0.02	-0.06**	-0.09*	0.02	0.06
Weak Negative Anomaly Index	-0.14**	-0.17	-0.28***	-0.45***	-0.04	-0.04
Wooldridge Score Test	0.11	0.7	10.18***	11.88***	0.14	0.01
F-Statistic (Stock - Yogo)	5.55	2.39	4.39	3.49	3.74	2.84
Kleibergen-Paap Wald statistic	17.07***	7.35*	16.58***	10.93**	11.68***	8.85**
Anderson-Rubin Wald Chi²	6.25*	6.25*	16.58***	16.58***	3.28	3.28
Conditional Likelihood Ratio	5.26**	5.72**	15.57***	15.83***	1.15	2.07

First stage regression for the baseline specification. The dependent variables are either the log number of outage per year or the log hour of outage per year. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.8 – Stock and Yogo 2005 critical values, single instrument.

	2SLS
10% bias	16.38
15% bias	8.96
20% bias	6.66
25% bias	5.53

Critical value for one endogenous regressor and one exogenous instrument from Stock and Yogo 2005.

Table 3.9 – Stock and Yogo 2005 critical values, three instruments.

	2SLS	LIML
10% maximal size	22.3	6.46
15% maximal size	12.83	4.36
20% maximal size	9.54	3.69
25% maximal size	7.8	3.32

Critical value for one endogenous regressor and three exogenous instrument from Stock and Yogo 2005.

Table 3.10 – Second stage regression, baseline specification, single instrument.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.60		-1.37***		-33.99	
	(1.54)		(0.42)		(520.67)	
Hours of PO		-1.35		-0.97***		-2.84
		(4.78)		(0.35)		(5.66)
small	-1.35***	-1.37***	-1.33***	-1.35***	-0.67	-1.18***
	(0.15)	(0.16)	(0.22)	(0.21)	(8.04)	(0.19)
large	1.09***	0.83	1.40***	1.32***	1.92	0.17
	(0.21)	(0.82)	(0.24)	(0.26)	(18.04)	(1.30)
very large	1.82***	1.88***	1.39*	1.46*	1.64	1.82**
	(0.30)	(0.45)	(0.55)	(0.64)	(7.25)	(0.65)
exporter	0.61**	0.66***	0.44**	0.52**	-2.23	0.70
	(0.23)	(0.19)	(0.14)	(0.19)	(43.66)	(0.39)
Credit	0.53***	0.59**	0.57***	0.61***	-0.77	0.56
	(0.09)	(0.20)	(0.12)	(0.13)	(17.91)	(0.40)
Share	0.35**	0.28	0.35*	0.50**	1.51	-0.13
	(0.13)	(0.35)	(0.15)	(0.18)	(21.18)	(0.73)
Foreig ownership	0.94***	0.76	0.55*	0.38	0.59	0.75
	(0.17)	(0.74)	(0.26)	(0.30)	(5.75)	(0.42)
Firm age	0.37***	0.47	0.27***	0.36***	-1.88	0.47**
	(0.06)	(0.32)	(0.06)	(0.07)	(36.30)	(0.15)
Constant	15.85**	20.88	19.25***	18.87***	149.06	28.92
	(6.02)	(25.99)	(1.64)	(1.73)	(2079.96)	(31.30)
Number of obs.	5048	5048	2116	2116	2407	2407

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level

Table 3.11 – Second stage regression, baseline specification, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.33		-1.33***		0.63	
	(0.79)		(0.48)		(1.93)	
Hours of PO		-0.40		-0.86**		-0.08
		(0.47)		(0.38)		(1.12)
small	-1.35***	-1.36***	-1.33***	-1.35***	-1.21***	-1.20***
	(0.15)	(0.15)	(0.22)	(0.21)	(0.19)	(0.19)
large	1.08***	0.99***	1.40***	1.32***	0.72***	0.72*
	(0.18)	(0.18)	(0.23)	(0.25)	(0.21)	(0.28)
very large	1.81***	1.82***	1.39*	1.42*	2.05***	2.03***
	(0.28)	(0.28)	(0.54)	(0.63)	(0.21)	(0.22)
exporter	0.64***	0.67***	0.45**	0.52**	0.66*	0.61**
	(0.16)	(0.15)	(0.14)	(0.18)	(0.29)	(0.22)
Credit	0.54***	0.56***	0.57***	0.60***	0.43**	0.41***
	(0.09)	(0.09)	(0.12)	(0.12)	(0.16)	(0.11)
Share	0.36**	0.35**	0.36*	0.49**	0.12	0.14
	(0.11)	(0.13)	(0.15)	(0.17)	(0.22)	(0.13)
Foreig ownership	0.96***	0.91***	0.55*	0.43	0.93***	0.92***
	(0.14)	(0.16)	(0.25)	(0.28)	(0.15)	(0.16)
Firm age	0.38***	0.41***	0.27***	0.35***	0.51**	0.46***
	(0.06)	(0.05)	(0.06)	(0.07)	(0.19)	(0.08)
Constant	14.82***	15.68***	19.08***	18.30***	10.90	13.85*
	(2.87)	(2.47)	(1.87)	(1.88)	(7.62)	(6.12)
Number of obs.	5048	5048	2116	2116	2407	2407

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.12 – LIML regression, baseline specification, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.43		-1.35***		1.02	
	(1.09)		(0.50)		(3.09)	
Hours of PO		-0.64		-0.87**		-0.30
		(0.84)		(0.39)		(5.52)
small	-1.35***	-1.36***	-1.33***	-1.35***	-1.21***	-1.20***
	(0.15)	(0.15)	(0.22)	(0.21)	(0.20)	(0.18)
large	1.08***	0.95***	1.40***	1.32***	0.70***	0.68
	(0.19)	(0.21)	(0.23)	(0.25)	(0.23)	(1.10)
very large	1.81***	1.83***	1.39**	1.43**	2.05***	2.02***
	(0.28)	(0.28)	(0.54)	(0.63)	(0.22)	(0.46)
exporter	0.63***	0.66***	0.44***	0.52***	0.69	0.62**
	(0.18)	(0.15)	(0.15)	(0.18)	(0.36)	(0.28)
Credit	0.53***	0.56***	0.57***	0.60***	0.44**	0.42
	(0.10)	(0.10)	(0.12)	(0.12)	(0.19)	(0.28)
Share	0.36***	0.33**	0.35**	0.50***	0.10	0.12
	(0.11)	(0.13)	(0.15)	(0.17)	(0.27)	(0.47)
Foreign ownership	0.95***	0.87***	0.55**	0.42	0.94***	0.91**
	(0.15)	(0.19)	(0.25)	(0.28)	(0.17)	(0.39)
Firm age	0.38***	0.43***	0.27***	0.35***	0.53**	0.46***
	(0.06)	(0.06)	(0.06)	(0.07)	(0.26)	(0.08)
Constant	15.20***	17.00***	19.17***	18.39***	9.34	15.01
	(4.02)	(4.43)	(1.93)	(1.94)	(12.26)	(30.11)
Number of obs.	5048	5048	2116	2116	2407	2407

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.13 – Second stage results and first stage tests for different instrument combinations.

Instrument Combination	Outage Measure	Coefficient	W S T	S-Y	K-P W	A-R C	CLR
Anomaly - S.D.	Number	-1.36***	9.34***	9.98	20.83***	12.04***	12.04***
	Hours	-0.98***	10.22***	6.37	13.3***	12.04***	12.03***
Positive - Negative	Number	-1.43***	14.97***	8.61	54.03***	20.86***	17.17***
	Hours	-1.05***	14.79***	4.39	27.52***	20.86***	18.46***
Weak Positive - Negative	Number	-1.28***	8.01***	4.91	30.78***	24.57***	19.16***
	Hours	-0.43*	2.71	4.14	26.02***	24.57***	6.83**
Strong Positive - Negative	Number	-1.58***	5.14**	4.96	31.1***	43.29***	24.38***
	Hours	-0.47	0.85	0.73	4.6	9.44	1.96
Index (Positive - Negative)	Number	-1.55***	11.87***	5.26	10.99***	15.88***	15.07***
	Hours	-0.95***	12.37***	8.03	16.76***	15.88***	13.76***
Weak Index (Positive - Negative)	Number	-1.24*	4.94**	3.25	6.78**	7.4**	7.4**
	Hours	-0.69*	5.34**	3.18	6.65**	7.4**	7.14**
Strong Index (Postive -Negative)	Number	0.07	0.01	0.38	0.8	0.01	0
	Hours	0.08	0.08	0.21	0.45	0.01	0
Strong - Weak Index (Positive - Negative)	Number	-1.38***	12.99***	3.12	13.04**	18.91***	17.81***
	Hours	-0.76**	7.69***	2.79	11.65**	18.91***	15.32***

Summary of the different considered combinations of instruments. The first column show the combination to which the others refer, the second indicates which outage measures is considered, the third shows the value and significance of coefficients for firms without a generator, the fourth presents the Wooldridge Score Test, the fifth for the Stock-Yogo F-test, the sixth for the Kleibergen-Paap Wald Test, the seventh for the Anderson-Rubin Wald Chi² Test and the eighth for the Conditional Likelihood Ratio Test.

Table 3.14 – 2SLS regression, manufacturing dummy, single instrument.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.53		-1.29***		-30.39	
	(1.34)		(0.41)		(448.51)	
Hours of PO		-1.24		-0.87***		-1.94
		(4.27)		(0.34)		(3.53)
small	-1.45***	-1.45***	-1.44***	-1.47***	-0.74	-1.20***
	(0.14)	(0.14)	(0.21)	(0.21)	(8.06)	(0.23)
large	1.18***	0.96	1.42***	1.31***	2.26	0.51
	(0.20)	(0.65)	(0.24)	(0.25)	(20.95)	(0.72)
very large	1.82***	1.89***	1.30**	1.29*	1.63	1.86***
	(0.29)	(0.47)	(0.49)	(0.56)	(6.18)	(0.40)
exporter	0.58**	0.58*	0.45***	0.47**	-2.33	0.61*
	(0.20)	(0.25)	(0.13)	(0.17)	(42.98)	(0.26)
Credit	0.58***	0.65**	0.67***	0.73***	-0.66	0.47*
	(0.09)	(0.24)	(0.11)	(0.12)	(15.58)	(0.21)
Share	0.41**	0.32	0.44**	0.52**	0.48	-0.01
	(0.14)	(0.42)	(0.16)	(0.18)	(4.49)	(0.51)
Foreig ownership	1.01***	0.88*	0.59*	0.44	2.90	0.93***
	(0.13)	(0.44)	(0.24)	(0.26)	(27.78)	(0.22)
Firm age	0.36***	0.43	0.25***	0.33***	-1.52	0.45***
	(0.05)	(0.23)	(0.05)	(0.07)	(29.59)	(0.12)
Manuf dummy	-0.18	-0.09	-0.01	0.02	-0.73	-0.12
	(0.09)	(0.31)	(0.17)	(0.16)	(7.91)	(0.23)
Constant	15.82**	20.68	18.87***	18.44***	141.81	24.94
	(5.44)	(24.18)	(1.57)	(1.68)	(1894.17)	(20.99)
Number of obs.	5752	5752	2277	2277	2698	2698
Yearly Mean Anomaly	-0.83*	-0.35	-2.29***	-3.4***	0.05	0.76
Wooldridge Score Test	0.09	0.11	7.24***	7.77***	0.44	0.42
F-Statistic (Stock - Yogo)	2.84	0.17	22.04	18.18	0	0.61
Kleibergen-Paap Wald statistic	2.8	0.17	22.67***	18.7***	0	0.62
Anderson-Rubin Wald Chi²	0.13	0.13	8.56***	8.56***	0.44	0.44

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test and the Anderson-Rubin Wald Chi² Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.15 – 2SLS regression, manufacturing dummy, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.09		-1.23***		1.03	
	(0.74)		(0.42)		(1.82)	
Hours of PO		-0.29		-0.82**		0.27
		(0.52)		(0.34)		(1.10)
small	-1.45***	-1.45***	-1.45***	-1.47***	-1.31***	-1.30***
	(0.14)	(0.14)	(0.21)	(0.20)	(0.17)	(0.18)
large	1.15***	1.10***	1.41***	1.31***	0.79***	0.88***
	(0.18)	(0.17)	(0.23)	(0.25)	(0.21)	(0.25)
very large	1.79***	1.81***	1.29**	1.28*	2.00***	2.01***
	(0.27)	(0.27)	(0.49)	(0.56)	(0.21)	(0.23)
exporter	0.63***	0.62***	0.45***	0.47**	0.67*	0.56**
	(0.15)	(0.13)	(0.13)	(0.17)	(0.28)	(0.20)
Credit	0.59***	0.60***	0.67***	0.73***	0.44**	0.40***
	(0.09)	(0.09)	(0.11)	(0.12)	(0.15)	(0.11)
Share	0.43***	0.41***	0.45**	0.52**	0.20	0.24
	(0.12)	(0.12)	(0.16)	(0.18)	(0.18)	(0.13)
Foreign ownership	1.00***	0.98***	0.60**	0.46	0.95***	1.02***
	(0.13)	(0.15)	(0.23)	(0.24)	(0.19)	(0.14)
Firm age	0.37***	0.39***	0.25***	0.32***	0.53**	0.47***
	(0.05)	(0.04)	(0.05)	(0.07)	(0.17)	(0.07)
Manuf dummy	-0.18	-0.16	-0.02	0.02	-0.18	-0.21
	(0.09)	(0.10)	(0.16)	(0.15)	(0.12)	(0.12)
Constant	14.03***	15.31***	18.65***	18.16***	9.17	11.90
	(2.80)	(2.81)	(1.62)	(1.75)	(7.51)	(6.40)
Number of obs.	5752	5752	2277	2277	2698	2698
Strong Negative Anomaly Index	0.12***	0.14**	0.02	-0.02	0.16***	0.22***
Strong Positive Anomaly Index	-0.01	0.02	-0.06**	-0.09**	0.02	0.06
Weak Negative Anomaly Index	-0.16**	-0.15	-0.3***	-0.47***	-0.05	-0.02
Wooldridge Score Test	0.01	0.26	10.35***	14.17***	0.4	0.08
F-Statistic (Stock - Yogo)	7.13	1.86	6.12	4.2	4.41	3.2
Kleibergen-Paap Wald statistic	21.78***	5.76	18.92***	12.97***	13.59***	9.86**
Anderson-Rubin Wald Chi²	5.12	5.12	20.79***	20.79***	1.99	1.99
Conditional Likelihood Ratio	3.7*	4.48*	20.19***	20.36***	0.05	0.45

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.16 – 2SLS regression, operational capacity, single instrument.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.44		-1.32***		33.69	
	(1.46)		(0.43)		(462.37)	
Hours of PO		-0.95		-0.94***		-5.85
		(3.92)		(0.36)		(19.51)
small	-1.35***	-1.36***	-1.33***	-1.35***	-1.72	-1.17***
	(0.15)	(0.15)	(0.22)	(0.21)	(7.43)	(0.34)
large	1.08***	0.90	1.40***	1.32***	-0.43	-0.43
	(0.20)	(0.67)	(0.24)	(0.26)	(16.47)	(4.11)
very large	1.81***	1.85***	1.39*	1.45*	2.44	1.59
	(0.29)	(0.39)	(0.55)	(0.63)	(5.74)	(1.79)
exporter	0.63**	0.66***	0.45**	0.52**	3.42	0.81
	(0.22)	(0.17)	(0.14)	(0.18)	(38.62)	(0.91)
Credit	0.53***	0.57***	0.57***	0.61***	1.57	0.72
	(0.09)	(0.17)	(0.12)	(0.13)	(16.27)	(1.19)
Share	0.36**	0.31	0.36*	0.50**	-1.21	-0.43
	(0.12)	(0.27)	(0.15)	(0.18)	(18.52)	(2.14)
Foreign ownership	0.95***	0.83	0.55*	0.40	1.27	0.55
	(0.16)	(0.61)	(0.25)	(0.29)	(5.07)	(1.26)
Firm age	0.38***	0.45	0.27***	0.36***	2.79	0.47
	(0.06)	(0.26)	(0.06)	(0.07)	(31.69)	(0.27)
Constant	15.24**	18.67	19.05***	18.70***	-121.07	45.35
	(5.59)	(21.21)	(1.61)	(1.73)	(1844.15)	(106.95)
Number of obs.	5048	5048	2116	2116	2407	2407
Yearly Main Anomaly	-0.73*	-0.34	-2.17***	-3.06***	-0.05	0.28
Wooldridge Score Test	0.05	0.07	16.73***	9.5***	0.5	0.49
F-Statistic (Stock - Yogo)	3.07	0.17	14.97	9.19	0.01	0.09
Kleibergen-Paap Wald statistic	3.15*	0.17	15.61***	9.59***	0.01	0.1
Anderson-Rubin Wald Chi²	0.08	0.08	11.05***	11.05***	0.51	0.51

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test and the Anderson-Rubin Wald Chi² Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.17 – 2SLS regression, operational capacity, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.63		-1.53***		2.01	
	(0.90)		(0.41)		(2.73)	
Hours of PO		-0.47		-1.01**		0.87
		(0.55)		(0.35)		(1.63)
small	-1.35***	-1.36***	-1.32***	-1.35***	-1.23***	-1.20***
	(0.15)	(0.15)	(0.23)	(0.21)	(0.21)	(0.22)
large	1.09***	0.98***	1.42***	1.32***	0.67**	0.91**
	(0.19)	(0.19)	(0.24)	(0.26)	(0.24)	(0.34)
very large	1.82***	1.82***	1.42**	1.47*	2.06***	2.11***
	(0.28)	(0.28)	(0.54)	(0.64)	(0.25)	(0.29)
exporter	0.61***	0.67***	0.43**	0.52**	0.77*	0.58*
	(0.17)	(0.15)	(0.15)	(0.19)	(0.37)	(0.26)
Credit	0.53***	0.56***	0.57***	0.61***	0.47*	0.36*
	(0.10)	(0.09)	(0.12)	(0.13)	(0.19)	(0.16)
Share	0.35**	0.34**	0.34*	0.50**	0.06	0.23
	(0.11)	(0.13)	(0.15)	(0.18)	(0.29)	(0.17)
Foreig ownership	0.94***	0.90***	0.52*	0.37	0.95***	0.98***
	(0.14)	(0.16)	(0.25)	(0.29)	(0.21)	(0.19)
Firm age	0.37***	0.41***	0.27***	0.37***	0.60*	0.46***
	(0.06)	(0.05)	(0.06)	(0.07)	(0.25)	(0.08)
Constant	15.97***	16.03***	19.85***	19.07***	5.38	8.63
	(3.20)	(2.76)	(1.61)	(1.78)	(10.66)	(8.66)
Number of obs.	5048	5048	2116	2116	2407	2407
Strong Negative Anomaly Index	0.05	0.09	-0.05	-0.08	0.09**	0.14**
Strong Positive Anomaly Index	-0.04*	-0.01	-0.09***	-0.11**	-0.01	0.02
Weak Negative Anomaly Index	-0.22***	-0.27**	-0.36***	-0.52***	-0.11	-0.15
Wooldridge Score Test	0.38	0.75	16.15***	16.15***	0.6	0.29
F-Statistic (Stock - Yogo)	7.1	2.64	8.8	5.8	4.29	1.69
Kleibergen-Paap Wald statistic	21.82***	8.11**	27.57***	18.17****	13.38***	5.26
Anderson-Rubin Wald Chi²	3.57	3.57	21.26***	21.26***	1.28	1.28
Conditional Likelihood Ratio	1.71	1.7	15.65***	16.31***	0.04	0.18

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.18 – 2SLS regression, 100 and 200 km radius, single instrument.

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	No generator	No generator	Generator	Generator
	Total Sale	Total Sale	Total Sale	Total Sale	Total Sale	Total Sale
Number of PO	-5.83		-1.75**		3.01	
	(5.92)		(0.60)		(5.67)	
Hours of PO		4.66		-1.81*		1.49
		(7.99)		(1.09)		(2.23)
small	-1.33***	-1.27***	-1.32***	-1.30***	-1.22***	-1.18***
	(0.19)	(0.35)	(0.23)	(0.23)	(0.24)	(0.26)
large	1.34**	1.85	1.46***	1.45***	0.63*	1.04*
	(0.45)	(1.43)	(0.25)	(0.33)	(0.31)	(0.49)
very large	2.09***	1.30	1.33*	1.74*	2.04***	2.14***
	(0.54)	(1.12)	(0.57)	(0.88)	(0.30)	(0.38)
exporter	0.09	0.93	0.38*	0.40	0.97	0.69*
	(0.73)	(0.51)	(0.17)	(0.30)	(0.56)	(0.28)
Credit	0.48	0.35	0.58***	0.70***	0.51*	0.33
	(0.27)	(0.49)	(0.13)	(0.17)	(0.25)	(0.24)
Share	0.23	0.64	0.37*	0.58*	0.00	0.28
	(0.30)	(0.60)	(0.15)	(0.24)	(0.39)	(0.30)
Foreign ownership	0.72*	1.67	0.44	0.01	0.94***	1.03***
	(0.31)	(1.21)	(0.24)	(0.53)	(0.26)	(0.23)
Firm age	0.27	0.09	0.27***	0.45**	0.69	0.47***
	(0.19)	(0.54)	(0.06)	(0.14)	(0.45)	(0.10)
Constant	36.49	-12.05	20.63***	22.93***	1.35	5.29
	(23.51)	(43.77)	(2.32)	(5.34)	(22.90)	(12.25)
Number of obs.	4945	4945	2065	2065	2367	2367
Yearly Main Anomaly	-0.34	0.42	-1.77***	-1.72	0.54	1.09
Wooldridge Score Test	2.74	2.86*	9.59***	10.07***	0.66	0.64
F-Statistic (Stock - Yogo)	0.82	0.45	8.52	2.58	1.15	2.51
Kleibergen-Paap Wald statistic	0.84	0.46	8.91***	2.7	1.19	2.61
Anderson-Rubin Wald Chi²	2.86*	2.86*	11.48***	11.48***	0.63	0.63

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test and the Anderson-Rubin Wald Chi² Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.19 – 2SLS regression, 100 and 200 km radius, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.93		-1.26**		-0.17	
	(0.72)		(0.55)		(1.24)	
Hours of PO		-0.32		-0.73*		-0.18
		(0.57)		(0.42)		(0.89)
small	-1.34***	-1.35***	-1.33***	-1.34***	-1.19***	-1.19***
	(0.15)	(0.15)	(0.22)	(0.21)	(0.19)	(0.19)
large	1.10***	1.00***	1.42***	1.37***	0.74***	0.70**
	(0.20)	(0.19)	(0.24)	(0.25)	(0.22)	(0.27)
very large	1.79***	1.77***	1.25*	1.32	2.01***	1.99***
	(0.29)	(0.30)	(0.58)	(0.68)	(0.22)	(0.22)
exporter	0.64***	0.73***	0.43**	0.50**	0.70**	0.72***
	(0.16)	(0.14)	(0.15)	(0.16)	(0.22)	(0.19)
Credit	0.55***	0.58***	0.58***	0.63***	0.41***	0.43***
	(0.10)	(0.10)	(0.12)	(0.12)	(0.13)	(0.13)
Share	0.34**	0.34*	0.39**	0.49**	0.14	0.11
	(0.13)	(0.14)	(0.15)	(0.17)	(0.16)	(0.18)
Foreign ownership	0.94***	0.94***	0.53*	0.46	0.95***	0.94***
	(0.15)	(0.19)	(0.26)	(0.31)	(0.13)	(0.15)
Firm age	0.36***	0.40***	0.27***	0.35***	0.45**	0.46***
	(0.05)	(0.06)	(0.06)	(0.07)	(0.14)	(0.09)
Constant	17.16***	15.25***	18.81***	17.71***	14.08**	14.40**
	(3.15)	(3.39)	(2.12)	(2.09)	(5.27)	(5.17)
Number of obs.	4945	4945	2065	2065	2367	2367
Strong Negative Anomaly Index	0.13***	0.15*	0.04	0.03	0.16***	0.19**
Strong Positive Anomaly Index	0.03	0.06**	0	0.03	0.04**	0.06**
Weak Negative Anomaly Index	-0.08	-0.04	-0.16*	-0.14	0	-0.02
Wooldridge Score Test	1.27	0.23	7.74***	3.87*	0.01	0.03
F-Statistic (Stock - Yogo)	4.15	3.1	3.22	1.47	4.06	4.14
Kleibergen-Paap Wald statistic	12.78***	9.56**	10.11**	4.6	12.69***	12.92***
Anderson-Rubin Wald Chi²	10.9**	10.9**	10**	10**	6.85*	6.85*
Conditional Likelihood Ratio	6.92**	4.37**	8.81***	8.07**	1.75	1.1

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.20 – 2SLS regression, 100 km radius, single instrument.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-2.83		-1.22		-3.74	
	(4.29)		(0.76)		(25.43)	
Hours of PO		-3.04		-0.67		-0.64
		(11.84)		(0.43)		(3.12)
small	-1.46***	-1.39*	-1.49***	-1.49***	-1.12	-1.29***
	(0.17)	(0.62)	(0.25)	(0.23)	(1.76)	(0.36)
large	1.46**	0.78	1.49***	1.38***	1.31	0.81*
	(0.47)	(1.62)	(0.26)	(0.25)	(2.93)	(0.46)
very large	2.18***	2.17*	1.75***	1.83**	2.26*	2.02***
	(0.39)	(0.86)	(0.51)	(0.62)	(1.18)	(0.45)
exporter	0.32	0.67*	0.42**	0.53**	0.30	0.65*
	(0.52)	(0.33)	(0.20)	(0.19)	(2.02)	(0.34)
Credit	0.49**	0.78	0.63***	0.70***	0.13	0.37
	(0.17)	(1.05)	(0.12)	(0.13)	(1.26)	(0.27)
Share	0.55**	0.59	0.51***	0.59**	0.72	0.30
	(0.21)	(0.64)	(0.16)	(0.19)	(3.08)	(0.20)
Foreig ownership	0.81**	0.51	0.57**	0.52*	0.96***	0.96***
	(0.25)	(1.79)	(0.27)	(0.29)	(0.28)	(0.20)
Firm age	0.31***	0.53	0.28***	0.36***	0.18	0.37***
	(0.09)	(0.77)	(0.06)	(0.08)	(1.29)	(0.07)
Constant	28.06	34.40	20.44***	19.29***	32.21	19.90
	(18.00)	(70.89)	(3.69)	(3.04)	(109.51)	(18.59)
Number of obs.	3777	3777	1722	1722	1616	1616
Yearly Mean Anomaly	-0.22	-0.2	-0.89*	-1.62**	0.09	0.51
Wooldridge Score Test	0.9	0.89	1.96	2.16	0.83	0.04
F-Statistic (Stock - Yogo)	0.32	0.08	3.17	4.65	0.05	0.34
Kleibergen-Paap Wald statistic	0.33	0.08	3.34*	4.91**	0.05	0.55
Anderson-Rubin Wald Chi²	0.94	0.94	2.57	2.57	0.05	0.05

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test and the Anderson-Rubin Wald Chi² Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.21 – 2SLS regression, 100 km radius, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.97*		-1.00**		-0.01	
	(0.53)		(0.34)		(0.95)	
Hours of PO		-0.08		-0.25		-0.04
		(0.29)		(0.23)		(0.39)
small	-1.51***	-1.54***	-1.49***	-1.51***	-1.36***	-1.36***
	(0.14)	(0.13)	(0.24)	(0.23)	(0.16)	(0.15)
large	1.27***	1.16***	1.47***	1.37***	0.89***	0.88***
	(0.17)	(0.16)	(0.23)	(0.22)	(0.22)	(0.19)
very large	2.07***	2.02***	1.73***	1.73***	2.10***	2.09***
	(0.23)	(0.21)	(0.49)	(0.49)	(0.21)	(0.20)
exporter	0.53***	0.64***	0.45**	0.57***	0.60***	0.60***
	(0.13)	(0.14)	(0.16)	(0.14)	(0.22)	(0.20)
Credit	0.51***	0.52***	0.63***	0.65***	0.31**	0.32**
	(0.11)	(0.10)	(0.11)	(0.11)	(0.13)	(0.13)
Share	0.48***	0.44***	0.52**	0.57***	0.27	0.27**
	(0.09)	(0.10)	(0.16)	(0.17)	(0.17)	(0.12)
Foreig ownership	0.93***	0.98***	0.61*	0.69**	0.98***	0.98***
	(0.16)	(0.18)	(0.27)	(0.28)	(0.14)	(0.15)
Firm age	0.33***	0.34***	0.28***	0.31***	0.37***	0.37***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.09)	(0.06)
Constant	20.27***	16.70***	19.56***	16.96***	16.14***	16.36***
	(2.58)	(2.17)	(2.56)	(2.40)	(4.35)	(2.66)
Number of obs.	3777	3777	1722	1722	1616	1616
Strong Negative Anomaly Index	0.16***	0.17***	0.17***	0.12*	0.08*	0.18***
Strong Positive Anomaly Index	0.04*	0.09***	0.02	0.06	0.05**	0.09***
Weak Negative Anomaly Index	-0.07	-0.11	-0.13*	-0.16	0.02	-0.12
Wooldridge Score Test	2.26	0.02	3.63*	0.62	0.01	0.01
F-Statistic (Stock - Yogo)	7.56	7.56	10.28	6.04	2.08	6.7
Kleibergen-Paap Wald statistic	23.48***	23.49***	32.63***	16.88***	6.6*	21.25***
Anderson-Rubin Wald Chi²	19.25***	19.25***	16.88***	19.18***	12.93***	12.93***
Conditional Likelihood Ratio	14.89***	0.34	4.03**	0.44	6.52**	0.4

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.22 – 2SLS regression, 200 km radius, single instrument.

	(1)	(2)	(3)	(4)	(5)	(6)
	All	All	No generator	No generator	Generator	Generator
	Total Sale	Total Sale	Total Sale	Total Sale	Total Sale	Total Sale
Number of PO	-8.44		-1.78**		2.97	
	(15.91)		(0.72)		(5.83)	
Hours of PO		-6.09		-1.50		3.51
		(14.69)		(0.92)		(7.39)
small	-1.45***	-1.46***	-1.32***	-1.31***	-1.16***	-1.16**
	(0.40)	(0.40)	(0.23)	(0.23)	(0.24)	(0.37)
large	1.70	0.43	1.44***	1.46***	0.56	1.31
	(1.21)	(1.65)	(0.26)	(0.31)	(0.44)	(1.31)
very large	2.37*	2.58	1.29**	1.66*	1.88***	2.05***
	(1.28)	(2.14)	(0.59)	(0.85)	(0.30)	(0.52)
exporter	-0.32	0.53	0.32*	0.41	1.05	0.70*
	(1.96)	(0.60)	(0.17)	(0.25)	(0.74)	(0.42)
Credit	0.42	0.60*	0.53***	0.57***	0.48*	0.28
	(0.40)	(0.36)	(0.13)	(0.15)	(0.26)	(0.40)
Share	-0.03	-0.10	0.37**	0.55**	0.14	0.56
	(0.91)	(1.48)	(0.16)	(0.21)	(0.25)	(0.95)
Foreig ownership	0.52	-0.00	0.40	0.11	0.99***	1.21*
	(0.77)	(2.36)	(0.24)	(0.44)	(0.29)	(0.66)
Firm age	0.12	0.66	0.26***	0.40***	0.69	0.47**
	(0.54)	(0.71)	(0.06)	(0.11)	(0.49)	(0.16)
Constant	47.14	47.18	20.72***	21.44***	1.30	-6.25
	(63.73)	(81.52)	(2.76)	(4.54)	(23.71)	(41.30)
Number of obs.	5234	5234	2081	2081	2440	2440
Yearly Mean Anomaly	-0.23	-0.32	-1.62**	-1.92*	0.53	0.45
Wooldridge Score Test	1.74	1.72	7.17***	7.29***	0.6	0.56
F-Statistic (Stock - Yogo)	0.23	0.19	6.02	2.85	1.09	0.45
Kleibergen-Paap Wald statistic	0.23	0.2	6.29**	2.98*	1.13	0.47
Anderson-Rubin Wald Chi²	1.79	1.79	8.45***	8.45***	0.57	0.45

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.23 – 2SLS regression, 200 km radius, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.94		-1.24**		-0.32	
	(0.67)		(0.51)		(1.08)	
Hours of PO		-0.68		-0.71*		-0.46
		(0.56)		(0.39)		(0.83)
small	-1.35***	-1.35***	-1.34***	-1.34***	-1.15***	-1.15***
	(0.15)	(0.14)	(0.22)	(0.21)	(0.19)	(0.18)
large	1.15***	1.01***	1.40***	1.37***	0.77***	0.68***
	(0.20)	(0.19)	(0.25)	(0.26)	(0.23)	(0.25)
very large	1.80***	1.82***	1.20**	1.31	1.96***	1.94***
	(0.30)	(0.31)	(0.60)	(0.70)	(0.24)	(0.22)
exporter	0.58***	0.67***	0.38**	0.45**	0.66***	0.70***
	(0.16)	(0.14)	(0.15)	(0.17)	(0.23)	(0.18)
Credit	0.51***	0.53***	0.54***	0.57***	0.37***	0.39***
	(0.09)	(0.09)	(0.12)	(0.11)	(0.12)	(0.11)
Share	0.36***	0.36**	0.41***	0.52**	0.15	0.10
	(0.13)	(0.15)	(0.15)	(0.17)	(0.15)	(0.19)
Foreig ownership	0.89***	0.83***	0.49*	0.42	0.90***	0.87***
	(0.15)	(0.18)	(0.26)	(0.29)	(0.14)	(0.18)
Firm age	0.35***	0.41***	0.26***	0.33***	0.43***	0.45***
	(0.06)	(0.06)	(0.06)	(0.07)	(0.13)	(0.09)
Constant	17.16***	17.18***	18.70***	17.57***	14.58**	15.83**
	(2.99)	(3.39)	(2.00)	(1.99)	(4.69)	(4.96)
Number of obs.	5234	5234	2081	2081	2440	2440
Strong Negative Anomaly Index	0.14***	0.16*	0.05	-0.01	0.19***	0.24**
Strong Positive Anomaly Index	0.04*	0.05*	0.01	0.03	0.04***	0.06**
Weak Negative Anomaly Index	-0.07	-0.08	-0.14*	-0.16	-0.01	-0.08
Wooldridge Score Test	1.5	1.41	8.56***	3.83*	0.04	0.26
F-Statistic (Stock - Yogo)	5.58	3.01	10.34**	1.57	7.67	5.09
Kleibergen-Paap Wald statistic	17.16***	9.26**	10.91**	4.93	23.94***	15.9***
Anderson-Rubin Wald Chi²	9.86**	9.86**	3.29	3.48**	6.86*	6.86*
Conditional Likelihood Ratio	6.97**	7.02**	9.71***	8.3***	2.11	2.69

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value the Wooldridge Score Test, for the Stock-Yogo F-test, the Kleibergen-Paap Wald Test and the Anderson-Rubin Wald Chi² Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.24 – 2SLS regression, average city sale, single instrument.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	0.71		-0.83**		-39.41	
	(1.26)		(0.40)		(563.53)	
Hours of PO		1.53		-0.53*		-3.64
		(4.19)		(0.27)		(6.30)
small	-1.29***	-1.26***	-1.32***	-1.33***	-0.51	-1.06***
	(0.16)	(0.19)	(0.21)	(0.20)	(8.00)	(0.22)
large	1.07***	1.36	1.37***	1.32***	2.15	0.07
	(0.16)	(0.75)	(0.22)	(0.23)	(19.19)	(1.45)
very large	1.75***	1.68**	1.26*	1.27*	1.60	1.80*
	(0.33)	(0.53)	(0.57)	(0.62)	(8.18)	(0.75)
exporter	0.72***	0.66**	0.48***	0.53***	-2.70	0.70
	(0.21)	(0.21)	(0.13)	(0.14)	(47.07)	(0.45)
Credit	0.49***	0.43*	0.54***	0.56***	-1.01	0.51
	(0.08)	(0.17)	(0.11)	(0.11)	(18.86)	(0.49)
Share	0.40**	0.48	0.35**	0.44**	1.78	-0.13
	(0.14)	(0.37)	(0.14)	(0.15)	(22.50)	(0.85)
Foreig ownership	0.98***	1.18	0.64**	0.57*	0.49	0.63
	(0.16)	(0.67)	(0.24)	(0.25)	(5.89)	(0.47)
Firm age	0.36***	0.25	0.26***	0.31***	-2.30	0.40*
	(0.05)	(0.28)	(0.05)	(0.06)	(38.91)	(0.18)
City Average Sale	0.79***	0.77***	0.44***	0.51***	0.49	0.86***
	(0.13)	(0.12)	(0.11)	(0.10)	(5.42)	(0.23)
Constant	-0.61	-5.89	11.14***	9.56***	163.59	20.78
	(6.44)	(23.77)	(2.82)	(2.34)	(2326.01)	(35.35)
Number of obs.	5048	5048	2116	2116	2407	2407
Yearly Main Anomaly	-0.69	-0.32	-2.07***	-3.28***	0.05	0.54
Wooldridge Score Test	0.53	0.46	3.73*	4.06**	5.36**	5.14**
F-Statistic (Stock - Yogo)	2.25	0.14	17.7	11.51	0.01	0.36
Kleibergen-Paap Wald statistic	2.31	0.15	18.48***	12.01***	0	0.37
Anderson-Rubin Wald Chi²	0.47	0.47	5.03**	5.03**	5.42**	5.42**

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value the Wooldridge Score Test, for the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.25– 2SLS regression, average city sale, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.42		-0.75**		-0.75	
	(0.38)		(0.36)		(0.68)	
Hours of PO		-0.31		-0.41*		-0.73
		(0.27)		(0.24)		(0.61)
small	-1.28***	-1.29***	-1.32***	-1.33***	-1.07***	-1.08***
	(0.15)	(0.15)	(0.21)	(0.20)	(0.18)	(0.17)
large	1.12***	1.05***	1.36***	1.32***	0.82***	0.65**
	(0.17)	(0.16)	(0.22)	(0.22)	(0.21)	(0.23)
very large	1.81***	1.80***	1.24*	1.23*	2.07***	2.03***
	(0.29)	(0.29)	(0.58)	(0.62)	(0.23)	(0.21)
exporter	0.60***	0.64***	0.49***	0.53***	0.51*	0.60**
	(0.15)	(0.15)	(0.12)	(0.13)	(0.21)	(0.21)
Credit	0.47***	0.49***	0.54***	0.55***	0.29**	0.36**
	(0.09)	(0.09)	(0.11)	(0.11)	(0.11)	(0.12)
Share	0.37***	0.36***	0.36**	0.43**	0.25*	0.15
	(0.09)	(0.10)	(0.13)	(0.15)	(0.11)	(0.14)
Foreig ownership	0.93***	0.90***	0.66**	0.62*	0.85***	0.81***
	(0.13)	(0.14)	(0.24)	(0.25)	(0.11)	(0.14)
Firm age	0.34***	0.37***	0.26***	0.30***	0.34***	0.40***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.09)	(0.08)
City Average Sale	0.73***	0.75***	0.45***	0.52***	0.84***	0.85***
	(0.11)	(0.10)	(0.11)	(0.09)	(0.11)	(0.11)
Constant	4.73	4.49*	10.69***	8.86***	4.17	5.06
	(2.56)	(2.19)	(2.60)	(2.13)	(3.29)	(3.59)
Number of obs.	5048	5048	2116	2116	2407	2407
Strong Negative Anomaly Index	0.12**	0.15**	0.01	-0.02	0.16***	0.21**
Strong Positive Anomaly Index	0	0.02	-0.05**	-0.09*	0.02	0.06
Weak Negative Anomaly Index	-0.12*	-0.17	-0.24***	-0.44***	-0.03	-0.04
Wooldridge Score Test	0.84	1.01	2.71	2.41	1.03	1.78
F-Statistic (Stock - Yogo)	4.62	2.35	3.04	2.96	3.91	2.87
Kleibergen-Paap Wald statistic	14.21***	7.23*	9.53**	9.28**	12.21***	8.95**
Anderson-Rubin Wald Chi²	1.35	1.35	4.86	4.86	2.55	2.55
Conditional Likelihood Ratio	0.94	1.12	3.95*	2.93	0.56	0.99

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.26 – 2SLS, long term anomaly, single instrument.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-1.60		-1.31***		-1.40	
	(2.82)		(0.50)		(3.58)	
Hours of PO		18.03		-0.85**		-0.73
		(396.10)		(0.37)		(2.24)
small	-1.33***	-1.17	-1.33***	-1.36***	-1.18***	-1.20***
	(0.16)	(4.11)	(0.22)	(0.21)	(0.19)	(0.17)
large	1.13***	4.25	1.40***	1.32***	0.77**	0.58
	(0.25)	(70.23)	(0.24)	(0.25)	(0.26)	(0.53)
very large	1.86***	0.72	1.39*	1.43*	2.00***	1.96***
	(0.31)	(23.54)	(0.55)	(0.63)	(0.29)	(0.31)
exporter	0.52	0.74	0.44**	0.52**	0.52	0.65**
	(0.31)	(2.18)	(0.14)	(0.17)	(0.31)	(0.24)
Credit	0.52***	-0.07	0.56***	0.60***	0.36	0.44**
	(0.11)	(13.66)	(0.12)	(0.12)	(0.20)	(0.15)
Share	0.34*	1.38	0.35*	0.48**	0.22	0.09
	(0.14)	(21.89)	(0.15)	(0.17)	(0.26)	(0.23)
Foreig ownership	0.86***	4.01	0.56*	0.45	0.87***	0.85***
	(0.22)	(66.93)	(0.25)	(0.29)	(0.14)	(0.21)
Firm age	0.37***	-0.94	0.27***	0.34***	0.39	0.48***
	(0.06)	(29.16)	(0.06)	(0.07)	(0.26)	(0.09)
Long term MA	2.41E+07	-2.06E+08	-3.42E+06	-1.01E+07	3.57E+07	2.77E+07
	(4.04E+07)	(4.62E+09)	(9.82E+06)	(1.13E+07)	(4.71E+07)	(3.38E+07)
Constant	20.11	-87.56	18.99***	18.14***	19.42	17.72
	(11.66)	(2221.29)	(2.03)	(1.97)	(14.94)	(12.68)
Number of obs.	5048	5048	2116	2116	2407	2407
Yearly Main Anomaly	-0.4	0.04	-2.15***	-3.32***	0.64	1.22
Wooldridge Score Test	0.05	0.08	10.21***	10.36***	0.47	0.22
F-Statistic (Stock - Yogo)	2.3	0.15	20.28	13.95	0	0.36
Kleibergen-Paap Wald statistic	2.36	0.16	21.16***	14.56***	0	0.37
Anderson-Rubin Wald Chi²	0.21	0.21	10.59***	10.59***	0.11	0.11

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.27 – 2SLS, long term anomaly, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-1.31**		-1.26***		-3.41	
	(0.65)		(0.47)		(2.22)	
Hours of PO		-1.09		-0.76		-2.39
		(0.85)		(0.49)		(2.14)
small	-1.33***	-1.35***	-1.32***	-1.34***	-1.11***	-1.15***
	(0.15)	(0.15)	(0.22)	(0.21)	(0.21)	(0.18)
large	1.14***	0.89***	1.39***	1.30***	0.90**	0.30
	(0.20)	(0.25)	(0.23)	(0.24)	(0.35)	(0.58)
very large	1.87***	1.88***	1.38*	1.40*	2.04***	1.90***
	(0.27)	(0.29)	(0.54)	(0.61)	(0.39)	(0.37)
exporter	0.54***	0.67***	0.45**	0.53**	0.33	0.71*
	(0.15)	(0.16)	(0.14)	(0.17)	(0.27)	(0.28)
Credit	0.51***	0.56***	0.57***	0.60***	0.23	0.47*
	(0.10)	(0.11)	(0.12)	(0.12)	(0.20)	(0.24)
Share	0.33**	0.30	0.35*	0.47**	0.32	-0.05
	(0.12)	(0.17)	(0.15)	(0.17)	(0.24)	(0.39)
Foreig ownership	0.88***	0.77***	0.56*	0.46	0.83***	0.71*
	(0.13)	(0.20)	(0.25)	(0.31)	(0.19)	(0.30)
Firm age	0.37***	0.46***	0.28***	0.35***	0.21	0.45***
	(0.05)	(0.09)	(0.06)	(0.07)	(0.19)	(0.13)
Long term Weak N.A. Index	1.75	2.00	0.82	1.30	2.92	2.37
	(1.21)	(1.43)	(0.63)	(0.89)	(2.21)	(2.07)
Long term Strong N.A. Index	-0.27	-0.33	-0.16	-0.28	-0.26	-0.07
	(0.27)	(0.34)	(0.20)	(0.30)	(0.41)	(0.37)
Long term Strong P.A. Index	0.13	0.15	0.06	0.08	0.23	0.25
	(0.09)	(0.11)	(0.04)	(0.05)	(0.18)	(0.21)
Constant	19.71***	20.49***	19.20***	18.35***	29.47**	28.58*
	(3.07)	(5.00)	(1.89)	(2.33)	(10.19)	(12.95)
Number of obs.	5048	5048	2116	2116	2407	2407
Strong Negative Anomaly Index	0.08	0.2	-0.26	-0.6*	0.16	0.36
Strong Positive Anomaly Index	0	0.04	-0.09**	-0.15**	0.03	0.08
Weak Negative Anomaly Index	-0.21	-0.1	-0.74**	-1.4**	-0.06	0.19
Wooldridge Score Test	5.74**	2.73	5.76**	4.16**	2.95*	2.33
F-Statistic (Stock - Yogo)	4.01	1.52	3.97	2.71	2.66	1.3
Kleibergen-Paap Wald statistic	12.34***	4.67	12.44***	8.49**	8.29*	4.06
Anderson-Rubin Wald Chi²	6.53*	6.53*	10.72***	10.72**	4.51	4.51
Conditional Likelihood Ratio	6.39**	6.46**	10.55***	8.3***	3.55*	3.93*

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.28 – 2SLS, median as cutting point, multiple instruments.

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.47		-1.45**		0.30	
	(1.11)		(0.56)		(2.36)	
Hours of PO		-0.57		-0.88**		-0.47
		(0.58)		(0.37)		(1.25)
small	-1.35***	-1.36***	-1.32***	-1.35***	-1.20***	-1.19***
	(0.15)	(0.15)	(0.23)	(0.21)	(0.19)	(0.18)
large	1.09***	0.96***	1.41***	1.32***	0.73***	0.64***
	(0.19)	(0.20)	(0.24)	(0.25)	(0.22)	(0.32)
very large	1.81***	1.83***	1.41***	1.43**	2.04***	2.00***
	(0.28)	(0.28)	(0.54)	(0.63)	(0.21)	(0.23)
exporter	0.62***	0.67***	0.43***	0.52***	0.63**	0.62***
	(0.18)	(0.15)	(0.15)	(0.18)	(0.30)	(0.21)
Credit	0.53***	0.56***	0.57***	0.60***	0.42**	0.43***
	(0.10)	(0.09)	(0.12)	(0.12)	(0.17)	(0.12)
Share	0.36***	0.33**	0.35**	0.50***	0.13	0.10
	(0.11)	(0.13)	(0.15)	(0.17)	(0.23)	(0.15)
Foreig ownership	0.95***	0.88***	0.53**	0.42	0.93***	0.90***
	(0.15)	(0.17)	(0.25)	(0.28)	(0.14)	(0.16)
Firm age	0.38***	0.42***	0.27***	0.35***	0.48**	0.46***
	(0.06)	(0.05)	(0.06)	(0.07)	(0.21)	(0.09)
Constant	15.36***	16.62***	19.53***	18.43***	12.21	15.96*
	(4.12)	(3.08)	(2.17)	(1.84)	(9.35)	(6.88)
Number of obs.	5048	5048	2116	2116	2407	2407
Strong Negative Anomaly Index	0.1*	0.15*	-0.03	-0.06	0.15**	0.21**
Strong Positive Anomaly Index	0.01	0.04	-0.06*	-0.09*	0.03	0.08*
Weak Negative Anomaly Index	-0.09	-0.11	-0.25***	-0.42***	0.03	0.03
Wooldridge Score Test	0.12	0.95	9.38***	10.45***	0.03	0.13
F-Statistic (Stock - Yogo)	1.99	1.95	2.99	3.56	2.77	2.99
Kleibergen-Paap Wald statistic	6.11	6	9.4**	11.16**	8.62**	9.33**
Anderson-Rubin Wald Chi²	8.69**	8.69**	15.67***	15.67***	5.97	5.97
Conditional Likelihood Ratio	7.96**	8.4**	15.39***	15.38***	3.51*	4.84**

2SLS estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. The table also reports the value for the Wooldridge Score Test, the Stock-Yogo F-test, the Kleibergen-Paap Wald Test, the Anderson-Rubin Wald Chi² Test and the Conditional Likelihood Ratio Test. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.29 – Heckman selection model for generator ownership, inverse Mills ratio as exclusion restriction.

	Selection Equation Generator Ownership		Outcome Equation Total Sales	
Number of PO	0.15***		-0.19***	
	(0.05)		(0.07)	
Hours of PO		0.05**		0.03
		(0.02)		(0.04)
small	-0.52***	-0.50***	-0.79***	-1.69***
	(0.07)	(0.07)	(0.25)	(0.18)
large	0.30***	0.29***	0.53**	1.02***
	(0.08)	(0.08)	(0.21)	(0.22)
very large	0.17	0.18	1.93***	2.17***
	(0.23)	(0.22)	(0.30)	(0.27)
exporter	0.26***	0.23**	0.41	0.84***
	(0.07)	(0.07)	(0.22)	(0.24)
Credit	0.08	0.09	0.30**	0.51***
	(0.06)	(0.06)	(0.13)	(0.14)
Share	0.11	0.09	0.05	0.27
	(0.10)	(0.09)	(0.15)	(0.20)
Foreign ownership	0.37***	0.36***	0.60***	1.31***
	(0.09)	(0.09)	(0.20)	(0.14)
Firm age	0.07**	0.09**	0.37***	0.54***
	(0.03)	(0.04)	(0.08)	(0.09)
Constant	0.64*	0.86***	14.85***	12.48***
	(0.28)	(0.26)	(0.92)	(0.95)
Number of obs.	4523	4523		
Censored	2116	2116		
Uncensored	2407	2407		
Rho	-0.7***	0.79***		

Heckman selection model with robust standard error. The dependent variables are a dummy equal to one if the firm owns a generator in the selection equation and the logarithm of total sale expressed in PPP 2005$ in the outcome equation, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.30 – Heckman selection model for generator ownership, lagged hydrological indexes as exclusion restriction.

	Selection Equation Generator Ownership		Outcome Equation Total Sales	
Number of PO	0.13***		0.08	
	(0.04)		(0.09)	
Hours of PO		0.05**		0.03
		(0.02)		(0.04)
small	-0.51***	-0.50***	-1.70***	-1.70***
	(0.07)	(0.07)	(0.18)	(0.18)
large	0.27***	0.28***	1.01***	1.02***
	(0.08)	(0.08)	(0.22)	(0.22)
very large	0.18	0.18	2.16***	2.17***
	(0.22)	(0.21)	(0.28)	(0.28)
exporter	0.25***	0.24***	0.84***	0.84***
	(0.07)	(0.07)	(0.25)	(0.24)
Credit	0.09	0.08	0.51***	0.50***
	(0.06)	(0.06)	(0.13)	(0.14)
Share	0.08	0.09	0.27	0.26
	(0.09)	(0.09)	(0.20)	(0.20)
Foreign ownership	0.36***	0.36***	1.30***	1.30***
	(0.09)	(0.09)	(0.14)	(0.14)
Firm age	0.10**	0.09**	0.55***	0.54***
	(0.03)	(0.03)	(0.09)	(0.09)
Weak neg. anomaly Index, -5 year mean	0.13	0.13		
	(0.14)	(0.15)		
Strong neg. anomaly Index, -5 year mean	0.11**	0.12**		
	(0.04)	(0.05)		
Strong pos. anomaly index, -5 year mean	0.07	0.07		
	(0.05)	(0.05)		
Constant	0.77**	1.06***	12.30***	12.47***
	(0.33)	(0.28)	(1.05)	(0.95)
Number of obs.	4523	4523		
Censored	2116	2116		
Uncensored	2407	2407		
Rho	0.79***	0.79***		

Heckman selection model with robust standard error. The dependent variables are a dummy equal to one if the firm owns a generator in the selection equation and the logarithm of total sale expressed in PPP 2005$ in the outcome equation, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.31 – Heckman selection model for generator ownership, electricity as main obstacle (firm-level) as exclusion restriction.

	Selection Equation Generator Ownership		Outcome Equation Total Sales	
Number of PO	0.13***		0.08	
	(0.05)		(0.11)	
Hours of PO		0.04**		0.02
		(0.02)		(0.04)
small	-0.51***	-0.50***	-1.67***	-1.67***
	(0.06)	(0.06)	(0.22)	(0.22)
large	0.28***	0.29***	1.03***	1.04***
	(0.11)	(0.11)	(0.28)	(0.29)
very large	0.22	0.22	2.19***	2.20***
	(0.26)	(0.26)	(0.27)	(0.27)
exporter	0.24***	0.23***	0.85**	0.84**
	(0.07)	(0.07)	(0.28)	(0.27)
Credit	0.09	0.08	0.50**	0.50**
	(0.07)	(0.07)	(0.19)	(0.19)
Share	0.10	0.10	0.27	0.26
	(0.15)	(0.14)	(0.28)	(0.28)
Foreign ownership	0.37***	0.37***	1.32***	1.32***
	(0.09)	(0.09)	(0.16)	(0.16)
Firm age	0.10**	0.09**	0.54***	0.54***
	(0.04)	(0.04)	(0.09)	(0.09)
Constrained by elec., firm level	0.16**	0.16**		
	(0.07)	(0.06)		
Constant	0.63**	0.90***	12.36***	12.53***
	(0.23)	(0.18)	(0.59)	(0.39)
Number of obs.	4443	4443		
Censored	2086	2086		
Uncensored	2357	2357		
Rho	0.78***	0.78***		

Heckman selection model with robust standard error. The dependent variables are a dummy equal to one if the firm owns a generator in the selection equation and the logarithm of total sale expressed in PPP 2005$ in the outcome equation, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.32 – Heckman selection model for generator ownership, finance as main obstacle (firm-level) as exclusion restriction.

	Selection Equation Generator Ownership		Outcome Equation Total Sales	
Number of PO	0.15**		-0.17***	
	(0.06)		(0.05)	
Hours of PO		0.05**		-0.08*
		(0.02)		(0.04)
small	-0.52***	-0.51***	-0.76**	-0.77**
	(0.06)	(0.06)	(0.34)	(0.34)
large	0.29***	0.31***	0.55**	0.53*
	(0.10)	(0.11)	(0.27)	(0.28)
very large	0.21	0.23	1.89***	1.88***
	(0.26)	(0.27)	(0.35)	(0.35)
exporter	0.27***	0.26***	0.41*	0.43*
	(0.06)	(0.06)	(0.23)	(0.24)
Credit	0.07	0.07	0.31**	0.32**
	(0.06)	(0.06)	(0.13)	(0.13)
Share	0.10	0.10	0.02	0.02
	(0.15)	(0.14)	(0.15)	(0.15)
Foreign ownership	0.36***	0.36***	0.62**	0.62**
	(0.09)	(0.08)	(0.25)	(0.25)
Firm age	0.07**	0.07**	0.36***	0.37***
	(0.03)	(0.03)	(0.07)	(0.08)
Constrained by finance., firm-level	-0.18**	-0.18**		
	(0.07)	(0.08)		
Constant	0.69**	0.97***	14.79***	14.51***
	(0.22)	(0.18)	(0.30)	(0.33)
Number of obs.	4443	4443		
Censored	2086	2086		
Uncensored	2357	2357		
Rho	-0.7***	0.7***		

Heckman selection model with robust standard error. The dependent variables are a dummy equal to one if the firm owns a generator in the selection equation and the logarithm of total sale expressed in PPP 2005$ in the outcome equation, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.33 – Heckman selection model for generator ownership, electricity as main obstacle (country-level) as exclusion restriction.

	Selection Equation Generator Ownership		Outcome Equation Total Sales	
Number of PO	0.15**		-0.19***	
	(0.06)		(0.05)	
Hours of PO		0.05**		0.03
		(0.02)		(0.04)
small	-0.52***	-0.50***	-0.79*	-1.69***
	(0.06)	(0.06)	(0.34)	(0.22)
large	0.30**	0.29**	0.53*	1.02***
	(0.10)	(0.11)	(0.26)	(0.28)
very large	0.17	0.18	1.93***	2.17***
	(0.24)	(0.24)	(0.34)	(0.26)
exporter	0.26***	0.23**	0.41	0.84**
	(0.06)	(0.07)	(0.23)	(0.27)
Credit	0.08	0.09	0.30*	0.51**
	(0.06)	(0.07)	(0.13)	(0.19)
Share	0.11	0.09	0.05	0.27
	(0.15)	(0.15)	(0.15)	(0.28)
Foreign ownership	0.37***	0.36***	0.60*	1.31***
	(0.09)	(0.09)	(0.24)	(0.16)
Firm age	0.07*	0.09*	0.37***	0.54***
	(0.03)	(0.04)	(0.08)	(0.09)
Constrained by elec., country share	18.37***	16.40***		
	(1.35)	(1.07)		
Constant	-0.27	0.05	14.85***	12.48***
	(0.23)	(0.16)	(0.29)	(0.39)
Number of obs.	4523	4523		
Censored	2116	2116		
Uncensored	2407	2407		
Rho	-0.7***	0.79***		

Heckman selection model with robust standard error. The dependent variables are a dummy equal to one if the firm owns a generator in the selection equation and the logarithm of total sale expressed in PPP 2005$ in the outcome equation, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 3.33 – Heckman selection model for generator ownership, finance as main obstacle (country-level) as exclusion restriction.

	Selection Equation Generator Ownership		Outcome Equation Total Sales	
Number of PO	0.15**		-0.19***	
	(0.06)		(0.05)	
Hours of PO		0.05**		0.03
		(0.02)		(0.04)
small	-0.52***	-0.50***	-0.79*	-1.69***
	(0.06)	(0.06)	(0.34)	(0.22)
large	0.30**	0.29**	0.53*	1.02***
	(0.10)	(0.11)	(0.26)	(0.28)
very large	0.17	0.18	1.93***	2.17***
	(0.24)	(0.24)	(0.34)	(0.26)
exporter	0.26***	0.23**	0.41	0.84**
	(0.06)	(0.07)	(0.23)	(0.27)
Credit	0.08	0.09	0.30*	0.51**
	(0.06)	(0.07)	(0.13)	(0.19)
Share	0.11	0.09	0.05	0.27
	(0.15)	(0.15)	(0.15)	(0.28)
Foreign ownership	0.37***	0.36***	0.60*	1.31***
	(0.09)	(0.09)	(0.24)	(0.16)
Firm age	0.07**	0.09**	0.37***	0.54***
	(0.03)	(0.04)	(0.08)	(0.09)
Constrained by finance, country share	-11.37***	-10.15***		
	(0.84)	(0.66)		
Constant	0.64**	0.86***	14.85***	12.48***
	(0.22)	(0.18)	(0.29)	(0.39)
Number of obs.	4523	4523		
Censored	2116	2116		
Uncensored	2407	2407		
Rho	-0.7***	0.79***		

Heckman selection model with robust standard error. The dependent variables are a dummy equal to one if the firm owns a generator in the selection equation and the logarithm of total sale expressed in PPP 2005$ in the outcome equation, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country and industry dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.1 – LIML regression, manufacturing dummy, multiple instruments (equivalent to 3.15).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.09		-1.26**		1.36	
	(0.97)		(0.44)		(2.44)	
Hours of PO		-0.60		-0.83*		0.79
		(1.22)		(0.35)		(3.07)
small	-1.45***	-1.45***	-1.45***	-1.47***	-1.32***	-1.33***
	(0.14)	(0.13)	(0.21)	(0.20)	(0.17)	(0.22)
large	1.15***	1.05***	1.41***	1.31***	0.77***	0.97
	(0.18)	(0.23)	(0.23)	(0.25)	(0.23)	(0.53)
very large	1.79***	1.83***	1.30**	1.28*	2.00***	2.04***
	(0.28)	(0.29)	(0.49)	(0.56)	(0.22)	(0.32)
exporter	0.63***	0.61***	0.45***	0.47**	0.70*	0.55*
	(0.16)	(0.15)	(0.13)	(0.17)	(0.33)	(0.23)
Credit	0.59***	0.62***	0.67***	0.73***	0.45**	0.38*
	(0.09)	(0.10)	(0.11)	(0.12)	(0.17)	(0.16)
Share	0.43***	0.38**	0.44**	0.52**	0.19	0.29
	(0.12)	(0.14)	(0.16)	(0.18)	(0.19)	(0.31)
Foreign ownership	1.00***	0.94***	0.59**	0.45	0.93***	1.04***
	(0.13)	(0.18)	(0.23)	(0.25)	(0.22)	(0.20)
Firm age	0.37***	0.40***	0.25***	0.32***	0.55**	0.47***
	(0.06)	(0.06)	(0.05)	(0.07)	(0.20)	(0.08)
Manuf dummy	-0.18	-0.14	-0.01	0.02	-0.17	-0.23
	(0.10)	(0.13)	(0.16)	(0.15)	(0.13)	(0.17)
Constant	14.05***	17.07*	18.77***	18.24***	7.75	8.82
	(3.74)	(6.78)	(1.69)	(1.80)	(10.12)	(18.00)
Number of obs.	5752	5752	2277	2277	2698	2698

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.2 – LIML regression, operational capacity, multiple instruments (equivalent to Table 3.17).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.96		-1.58***		2.56	
	(1.47)		(0.43)		(3.43)	
Hours of PO		-1.27		-1.09**		2.72
		(2.28)		(0.40)		(6.05)
small	-1.34***	-1.37***	-1.32***	-1.34***	-1.24***	-1.21***
	(0.15)	(0.15)	(0.23)	(0.21)	(0.22)	(0.32)
large	1.11***	0.84*	1.42***	1.32***	0.65*	1.28
	(0.21)	(0.42)	(0.24)	(0.27)	(0.26)	(1.27)
very large	1.84***	1.87***	1.43**	1.49*	2.07***	2.25***
	(0.28)	(0.33)	(0.54)	(0.65)	(0.27)	(0.66)
exporter	0.57**	0.66***	0.42**	0.51*	0.82	0.51
	(0.21)	(0.17)	(0.15)	(0.20)	(0.43)	(0.45)
Credit	0.52***	0.58***	0.57***	0.61***	0.49*	0.26
	(0.11)	(0.14)	(0.13)	(0.13)	(0.22)	(0.42)
Share	0.34***	0.29	0.34*	0.51**	0.04	0.41
	(0.10)	(0.20)	(0.16)	(0.19)	(0.33)	(0.67)
Foreign ownership	0.93***	0.78*	0.51*	0.34	0.95***	1.10*
	(0.15)	(0.37)	(0.25)	(0.30)	(0.24)	(0.49)
Firm age	0.37***	0.47**	0.27***	0.37***	0.64*	0.46***
	(0.06)	(0.15)	(0.06)	(0.08)	(0.30)	(0.13)
Constant	17.30**	20.42	20.02***	19.45***	3.17	-1.48
	(5.42)	(12.19)	(1.69)	(2.00)	(13.51)	(32.90)
Number of obs.	5048	5048	2116	2116	2407	2407

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.3 – LIML regression, 100 and 200 km radius, multiple instruments (equivalent to Table 3.19).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-1.33		-1.34**		-0.27	
	(1.13)		(0.62)		(2.51)	
Hours of PO		-0.94		-1.06		-1.84
		(2.08)		(0.81)		(14.06)
small	-1.34***	-1.36***	-1.33***	-1.32***	-1.19***	-1.20***
	(0.16)	(0.15)	(0.22)	(0.22)	(0.20)	(0.18)
large	1.12***	0.90*	1.43***	1.39***	0.75**	0.36
	(0.22)	(0.36)	(0.24)	(0.29)	(0.24)	(2.89)
very large	1.82***	1.83***	1.26*	1.45	2.01***	1.85
	(0.30)	(0.38)	(0.58)	(0.76)	(0.22)	(1.25)
exporter	0.59**	0.71***	0.42**	0.47*	0.69*	0.75*
	(0.19)	(0.18)	(0.15)	(0.21)	(0.29)	(0.32)
Credit	0.54***	0.60***	0.58***	0.65***	0.41**	0.53
	(0.11)	(0.14)	(0.12)	(0.13)	(0.14)	(0.87)
Share	0.33*	0.30	0.38**	0.52*	0.14	-0.05
	(0.13)	(0.20)	(0.15)	(0.20)	(0.19)	(1.39)
Foreign ownership	0.92***	0.85*	0.52	0.32	0.95***	0.86
	(0.16)	(0.38)	(0.27)	(0.45)	(0.12)	(0.75)
Firm age	0.36***	0.44**	0.27***	0.38***	0.44	0.45**
	(0.06)	(0.14)	(0.06)	(0.11)	(0.23)	(0.14)
Constant	18.74***	18.64	19.12***	19.29***	14.50	23.48
	(4.73)	(11.58)	(2.38)	(3.93)	(10.35)	(77.03)
Number of obs.	4945	4945	2065	2065	2367	2367

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.4 – LIML regression, 100 km radius, multiple instruments (equivalent to Table 3.21).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-1.35		-1.19**		56.07	
	(0.78)		(0.42)		(60850.23)	
Hours of PO		-0.21		-0.67		0.16
		(1.02)		(0.68)		(17.26)
small	-1.50***	-1.53***	-1.49***	-1.49***	-5.08	-1.38
	(0.14)	(0.14)	(0.24)	(0.23)	(4029.12)	(1.93)
large	1.31***	1.15***	1.49***	1.38***	-5.50	0.91
	(0.18)	(0.18)	(0.24)	(0.26)	(6926.86)	(2.18)
very large	2.10***	2.03***	1.74***	1.83**	-0.26	2.12
	(0.25)	(0.23)	(0.51)	(0.65)	(2556.89)	(2.19)
exporter	0.49**	0.64***	0.42*	0.53**	5.15	0.59
	(0.15)	(0.14)	(0.17)	(0.19)	(4937.65)	(1.45)
Credit	0.50***	0.53***	0.63***	0.70***	3.05	0.30
	(0.12)	(0.13)	(0.12)	(0.13)	(2972.56)	(1.38)
Share	0.49***	0.45***	0.51**	0.59**	-6.41	0.26
	(0.10)	(0.12)	(0.16)	(0.20)	(7249.82)	(0.77)
Foreign ownership	0.90***	0.96***	0.58*	0.52	1.33	0.99
	(0.16)	(0.27)	(0.27)	(0.42)	(379.67)	(0.70)
Firm age	0.32***	0.35***	0.28***	0.36**	3.29	0.37
	(0.05)	(0.07)	(0.06)	(0.11)	(3165.37)	(0.21)
Constant	21.85***	17.45**	20.32***	19.32***	-225.54	15.12
	(3.52)	(6.31)	(2.76)	(4.26)	(262251.67)	(102.43)
Number of obs.	3777	3777	1722	1722	1616	1616

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.5 – LIML regression, 200 km radius, multiple instruments (equivalent to Table 3.23).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-1.17		-1.33**		-0.47	
	(0.89)		(0.58)		(1.77)	
Hours of PO		-1.11		-1.01		-1.23
		(1.11)		(0.73)		(2.65)
small	-1.35***	-1.36***	-1.34***	-1.33***	-1.15***	-1.15***
	(0.15)	(0.14)	(0.22)	(0.22)	(0.19)	(0.17)
large	1.17***	0.97***	1.41***	1.41***	0.78**	0.56
	(0.21)	(0.22)	(0.25)	(0.29)	(0.25)	(0.50)
very large	1.82***	1.88***	1.22*	1.45	1.97***	1.92***
	(0.30)	(0.35)	(0.60)	(0.78)	(0.25)	(0.25)
exporter	0.55**	0.66***	0.37*	0.44*	0.65*	0.70***
	(0.17)	(0.16)	(0.15)	(0.20)	(0.28)	(0.19)
Credit	0.51***	0.54***	0.54***	0.57***	0.37**	0.42**
	(0.10)	(0.10)	(0.12)	(0.12)	(0.13)	(0.15)
Share	0.35*	0.32	0.40**	0.53**	0.15	0.01
	(0.14)	(0.18)	(0.15)	(0.19)	(0.15)	(0.38)
Foreign ownership	0.88***	0.76**	0.48	0.30	0.90***	0.81**
	(0.15)	(0.25)	(0.26)	(0.40)	(0.15)	(0.31)
Firm age	0.34***	0.43***	0.26***	0.35***	0.41*	0.44***
	(0.06)	(0.08)	(0.06)	(0.09)	(0.18)	(0.11)
Constant	18.12***	19.55**	19.04***	19.05***	15.19*	20.13
	(3.85)	(6.39)	(2.26)	(3.60)	(7.45)	(15.05)
Number of obs.	5234	5234	2081	2081	2440	2440

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.6 – LIML regression, average city sale, multiple instruments (equivalent to Table. 3.25).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.43		-0.77**		-0.81	
	(0.39)		(0.36)		(0.74)	
Hours of PO		-0.32		-0.42*		-0.77
		(0.28)		(0.25)		(0.65)
small	-1.28***	-1.29***	-1.32***	-1.33***	-1.07***	-1.08***
	(0.15)	(0.15)	(0.21)	(0.20)	(0.18)	(0.17)
large	1.12***	1.04***	1.37***	1.32***	0.83***	0.65**
	(0.17)	(0.16)	(0.22)	(0.22)	(0.22)	(0.24)
very large	1.81***	1.81***	1.25*	1.24*	2.07***	2.02***
	(0.29)	(0.29)	(0.58)	(0.62)	(0.23)	(0.21)
exporter	0.60***	0.64***	0.48***	0.53***	0.51*	0.60**
	(0.15)	(0.15)	(0.12)	(0.13)	(0.21)	(0.21)
Credit	0.47***	0.49***	0.54***	0.55***	0.29**	0.36**
	(0.09)	(0.09)	(0.11)	(0.11)	(0.11)	(0.12)
Share	0.37***	0.36***	0.36**	0.43**	0.25*	0.15
	(0.09)	(0.10)	(0.13)	(0.15)	(0.11)	(0.14)
Foreign ownership	0.93***	0.90***	0.65**	0.61*	0.85***	0.81***
	(0.13)	(0.14)	(0.24)	(0.25)	(0.11)	(0.14)
Firm age	0.34***	0.37***	0.26***	0.30***	0.34***	0.40***
	(0.05)	(0.05)	(0.05)	(0.06)	(0.09)	(0.08)
City Average Sale	0.73***	0.75***	0.45***	0.52***	0.84***	0.85***
	(0.11)	(0.10)	(0.11)	(0.09)	(0.11)	(0.11)
Constant	4.76	4.54*	10.77***	8.93***	4.41	5.25
	(2.59)	(2.23)	(2.65)	(2.18)	(3.52)	(3.78)
Number of obs.	5048	5048	2116	2116	2407	2407

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.7 – LIML, long term anomaly, multiple instruments (equivalent to Table 3.27).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-1.40*		-1.27***		-4.78	
	(0.73)		(0.48)		(3.72)	
Hours of PO		-1.20		-0.79		-2.98
		(1.00)		(0.52)		(2.96)
small	-1.33***	-1.35***	-1.32***	-1.34***	-1.08***	-1.15***
	(0.15)	(0.15)	(0.22)	(0.21)	(0.24)	(0.20)
large	1.14***	0.87**	1.39***	1.30***	0.96*	0.19
	(0.20)	(0.27)	(0.23)	(0.25)	(0.44)	(0.74)
very large	1.87***	1.88***	1.38*	1.41*	2.03***	1.86***
	(0.27)	(0.30)	(0.54)	(0.61)	(0.49)	(0.45)
exporter	0.53***	0.67***	0.45**	0.53**	0.22	0.73*
	(0.16)	(0.17)	(0.14)	(0.17)	(0.39)	(0.33)
Credit	0.50***	0.56***	0.57***	0.60***	0.17	0.49
	(0.11)	(0.12)	(0.12)	(0.12)	(0.27)	(0.29)
Share	0.32**	0.29	0.35*	0.48**	0.39	-0.10
	(0.12)	(0.18)	(0.15)	(0.17)	(0.33)	(0.48)
Foreign ownership	0.88***	0.76***	0.56*	0.45	0.80**	0.66
	(0.13)	(0.22)	(0.25)	(0.31)	(0.26)	(0.38)
Firm age	0.37***	0.47***	0.28***	0.36***	0.12	0.45**
	(0.05)	(0.10)	(0.06)	(0.08)	(0.29)	(0.15)
Long term Weak N.A. Index	-0.27	-0.34	-0.16	-0.29	-0.29	-0.04
	(0.27)	(0.34)	(0.20)	(0.30)	(0.48)	(0.42)
Long term Strong N.A. Index	0.13	0.16	0.06	0.08	0.28	0.28
	(0.09)	(0.12)	(0.04)	(0.05)	(0.22)	(0.25)
Long term Strong P.A. Index	1.78	2.07	0.82	1.32	3.49	2.58
	(1.22)	(1.48)	(0.63)	(0.90)	(2.69)	(2.33)
Constant	20.08***	21.10***	19.26***	18.49***	35.48*	32.03
	(3.38)	(5.81)	(1.92)	(2.47)	(16.65)	(17.83)
Number of obs.	5048	5048	2116	2116	2407	2407

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A3.8 – LIML, median as cutting point, multiple instruments (equivalent to Table 3.28).

	(1) All Total Sale	(2) All Total Sale	(3) No generator Total Sale	(4) No generator Total Sale	(5) Generator Total Sale	(6) Generator Total Sale
Number of PO	-0.76		-1.45**		0.93	
	(1.93)		(0.56)		(6.82)	
Hours of PO		-0.90		-0.88**		-1.62
		(1.05)		(0.37)		(4.82)
small	-1.35***	-1.36***	-1.32***	-1.35***	-1.21***	-1.19***
	(0.15)	(0.15)	(0.23)	(0.21)	(0.20)	(0.17)
large	1.10***	0.91***	1.41***	1.32***	0.71**	0.41
	(0.22)	(0.25)	(0.24)	(0.25)	(0.32)	(1.01)
very large	1.83***	1.85***	1.41**	1.43*	2.05***	1.92***
	(0.29)	(0.29)	(0.54)	(0.63)	(0.23)	(0.45)
exporter	0.59**	0.66***	0.43**	0.52**	0.69	0.66*
	(0.24)	(0.16)	(0.15)	(0.18)	(0.61)	(0.28)
Credit	0.53***	0.57***	0.57***	0.61***	0.44	0.49
	(0.11)	(0.11)	(0.12)	(0.12)	(0.30)	(0.28)
Share	0.35***	0.31**	0.35**	0.50**	0.11	-0.01
	(0.10)	(0.15)	(0.15)	(0.17)	(0.40)	(0.46)
Foreign ownership	0.94***	0.83***	0.53*	0.42	0.94***	0.82**
	(0.17)	(0.22)	(0.25)	(0.28)	(0.18)	(0.36)
Firm age	0.37***	0.44***	0.27***	0.35***	0.53	0.47***
	(0.07)	(0.07)	(0.06)	(0.07)	(0.51)	(0.11)
Constant	16.48*	18.42**	19.54***	18.44***	9.66	22.26
	(7.30)	(5.60)	(2.17)	(1.85)	(27.17)	(26.39)
Number of obs.	5048	5048	2116	2116	2407	2407

LIML estimation with standard errors clustered at the city level in parenthesis. The dependent variables is the logarithm of total sale expressed in PPP 2005$, the explanatory variables of main interest are the logarithm of number of power outages per year and the logarithm of the hours of power outages per year. All regressions include country dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Chapter 4

Water availability for electricity generation and night lights.

4.1 Introduction.

To avoid problems related to endogeneity, the analysis presented in the previous chapter exploited the high reliance of many Sub-Saharan African (SSA) countries on hydropower to construct a valid instrument for assessing how detrimental power outages are to firms. The hydrological measures of water available for hydroelectricity generation developed in Chapter 2 and Chapter 3 can though also be used as a starting point to explore other interesting empirical questions related to the effect of relying so heavily on a single technology to generate electricity. One such question is if yearly variations in hydro-power generation impact on the general economic activity of African countries, and to this analysis we turn in this Chapter.

The question is motivated by the fact that over the last two decades there have been several reports of rain failures or droughts leading to a diminished hydropower production in different SSA countries, with some projects failing to meet expected returns on investments (Karekezi et al. 2009, Cole et al 2014, UNECA 2016a). For countries in which hydropower represents more than 60% of installed capacity, this is likely to also result in diminished economic activity. To investigate this issue we have though to abandon the firm data used in Chapter 1 and 3, as they are not collected in a timely enough manner to allow for longitudinal exploration. The most widely available longitudinal measure of general economic activity is GDP, but as our hydrological variables have been developed at the city level we would ideally look for more disaggregated information, which are though hard to come by for the African continent.

As it has been recently demonstrated, one of the best candidates for a sub-national measure of economic activity in developing countries is the night light data collected by the Defense Meteorological Satellite Program – Operational Linescan System (DMSP OLS) and processed by the National Oceanic and Atmospheric Administration (NOAA). This data has been collected since the 1970s by satellites circling around every location of the earth multiple time per days and take the form of world-wide pictures in which each pixel, representing an area equal to 30 arc-second (slightly less than a square kilometre at the equator), is assigned a value ranging from 1 to 63 depending on the intensity and frequency of the light emitted in its area.

DSMP OLS data has been receiving a fair amount of attention from social scientists since it became available in digital format in the 1990s. The reason for this attention is easy to grasp: as most human activities require electricity, and most usage of electricity after the sun sets happen in an environment in which lighting is required, a measure of the intensity of night lights can give information about many interesting aspects of human behaviour. To give a few examples, over the last few years the DMSP OLS data has been used for applications as diverse as measuring urban extension, mapping nocturnal squid-fishing, connecting anthropogenic emissions to sea-life breeding behaviour or estimating sub-national level GDP. It is the definition of the data which makes them so suitable for diverse uses, as each researcher can focus on a specific area of the globe and aggregate the information at the necessary level, be that a borough, a city, a region, a state or a continent.

In our specific case, the data is first aggregated at the city-level so to cover the same sample for which city-level instruments were developed in Chapter 3. As the urban masks available for SSA cities were constructed in the year 2000, we also explore how accounting for the urbanisation process who took place over the last 15 years influences the results. Furthermore, using the hydrological variables constructed at the country-level, the resolution of the DSMP-OLS data allows us to explore the relationship between hydroelectric generation and economic activity also at the national level. The remainder of the chapter is organized as follow: section 2 reviews the relevant literature; section 3 introduces the data used in the study; section 4 presents the methodology; section 5 discusses the results and section 6 concludes.

4.2 Literature Review.

The use of satellite data has become much more common in the social sciences over the last couple of decades, so much so that the Center for International Earth Science Information Network (CIESIN), together with the Socioeconomic Data and Applications Center (SEDAC) of the US National Aeronautics and Space Administration (NASA) published in 2002 a guide called "Social Science Applications of Remote Sensing" (de Sherbinin *et al.* 2002), meant to be a useful starting point for someone taking its first steps into the field. This section of the chapter reviews the most relevant literature related to the application in social sciences of night light imagery. The majority of these papers have investigated questions related to urbanization processes, estimation of the earth's population, energy use and, more recently, estimation of economic activity at sub-national levels or some combination of the above[90].

The first application of night light satellite imagery is that of Tobler (1969), who used a portion of a picture taken by the Apollo VI on Texas to add further empirical weight to a series of claim made by investigators on the possibility of estimating the radius of settlements and their built up area with knowledge of its number of inhabitant. In the paper, the author proves that also the reverse is true by deriving within a 2% accuracy the population of Dallas from the picture,

[90] The papers reviewed in this section were selected from those found by searching for different combinations of the terms "DMSP OLS", "Night Light", "Population", "GDP", "Africa" and "Infrastructure" in EBSCO host and the bibliographies therein.

concluding that further studies need to be directed towards the development of the concept of build-up areas.

Following further developments, the first author to find a practical application of an earlier version of DMSP OLS data was Welch in 1980, who recognized the potential of satellite data for understanding the quality of urban planning trends in developing countries and of energy demands in developed ones (Welch 1980). Specifically, the DMSP OLS data was found to be highly correlated with the known distribution of urban centres in the US and a good starting point to monitor and evaluate differences in urban population and energy utilization at subnational scales. However, the author concludes that further technological advancements in satellites' sensors are required before this data can be fully evaluated, and similarly the complementary data needed must be gathered with a higher precision.

As technology advances, with the development of sensors capable of discerning more sources of lights (Foster 1983) and the first complete satellite picture of the whole earth being composed (Sullivan 1989), more complex applications of DMSP OLS became available, such as that of Imhoff *et al.* in 1997 (Imhoff *et al.* 1997). The authors look at the conflict between the preservation of the best agricultural soils and the continuing process of urbanization in the US, first creating spatial masks for all urban areas by experimenting with different levels of light threshold in the DMSP, then joining those with FAO Digital Soil Map of the World dataset described in Chapter 2. Through their analysis, they were able to determine that only 3.14% of the US soil which can be classified as agriculturally productive had been urbanized, but focusing on cities with higher than average urban expansion and high agricultural productivity they also noticed how urbanization was often taking place on the most productive lands or on the second best when the best were preserved.

Also in 1997, Elvidge *et al.* similarly analyses the potential of DMSP OLS for updating urbanization data at a lower processing power cost than with other datasets (Elvdige *et al.* 1997). Furthermore, they show how with the recent development in technologies the analyses first proposed by Tobler and Welch were now easier to conduct, demonstrating how there exists a significant linear relationship between area lit, population, energy consumption and GDP for 21 countries using a composite of pictures taken between 1994 and 1995. However, especially with regard to population the said relationship was stronger in developed than in developing countries, as advances in sensor technology were still needed to fully appreciate light emissions from rural areas and to properly record light intensity. In any case, the night light dataset was praised for being a further tool which could be utilized in various types of analysis, amongst which rural electrification and frequency of power outages.

The first attempt to use the DMSP OLS dataset at a global scale is that of Doll *et al.* (2000), who stress how relevant the knowledge of the spatial distribution of global population is to understand dynamics such as anthropogenic climate change. By combining the same composite DMSP OLS image of the previous study with a series of other geo-referenced data, the authors show how the significant regional or country relationship found between area lit, population, energy consumption and GDP hold also at a global level. They further shows how area lit also relates to CO_2 emissions, although they also stress how these, like the previous relationships, should be analysed separately for countries at different development stages.

Two studies of Sutton and other authors from the early 2000s analyse other possible applications of the DMSP OLS night light imagery to population questions (Sutton 2001, Sutton *et al.* 2003). Sutton 2001 shows how, combining the relationship between extension of urban areas and population empirically proved by Tobler with knowledge of the division between urban and rural areas in each country, 1,383 geo-located city masks could be used to estimate the global population for the year 2000 with a margin of error of 7%. As suggested in the previous literature, the author further uses a series of different light intensity thresholds to try and differentiate the urban expansion of cities in different countries. Sutton *et al.* 2003 combines DMSP OLS with two other geo-located datasets (LandScan and Gridded Population of the World) to obtain a first approximation of a method to calculate a temporally averaged density of population, relevant for many policy uses amongst which decision of where to focus population censuses in countries with scarce resources.

This author looked at two other questions in the same period: firstly in 2002, along with Costanza (Sutton and Costanza 2002), he combined the DMSP OLS data with a dataset on land cover and GDP to obtain a spatially explicit ecologically-augmented measure of GDP, accounting for non-marketed ecosystem services, which positively correlates with other sustainability index and environmental impact measures; alone in 2003 (Sutton 2003) he tries instead to develop a city-specific aggregate measure of per capita land consumption, also in this case experimenting on how to best set the threshold of light luminosity to define the urban extent, depending on the interest for commuter zones contiguous to city which tend to emit fewer lights.

In the following years, three different papers looked more in depth at how to use night light data as a starting point for the estimation of sub-national GDP. Doll *et al.* (2005) analysed how the relationship between night lights and economic activity changed moving from the national to different sub-national levels for the US and 11 European countries, finding that it is still correlated with gross regional product but that the relationship varied across different regions. While incorporating land use in the analyses improved the estimates, for some regions it was still impossible to identify any significant linear relationship between the two (Doll *et al.* 2005).

In the same year, Ebener *et al.* (2005) made a similar attempt using GDP figures representative for the regions of 26 countries across all 5 continents in order to obtain an approximation of the distribution of health indicators which normally correlate with income per capita (Ebener *et al.* 2005). The authors found that, with access to income per capita figures at sub-national level, a country specific model relating economic activity to night light could be developed for many different countries, but, given the impossibility of identifying any general regional model, other data were still needed as a starting point to fully exploit DMSP OLS.

Finally, Sutton *et al.* (2007) compared the methodology developed by Ebener *et al.* (2005) to one which also took into account differences in population density across regions through an application to China, India, Turkey and the US (Sutton *et al.* 2007). The authors conclude that the estimation of GDP through night lights could be helpful in countries in which statistical offices have little capacity but can add little to the measures produced by those with more substantial budgets and advanced techniques. They also note how the increasing use of satellite data in the social sciences and in society in general was critically leading to a rise in the ethical, political and legal challenges associated with the increase in spatio-demographic information.

Elvidge *et al.* in 2009 investigates instead if the DMSP OLS data could also be used for creating a poverty map, which had been used often since the early 2000s to target program intervention but are normally based on rounds of surveys which do not have complete coverage and regular roll out (Elvidge *et al.* 2009). They found that a simple index constructed by dividing the average value of light intensity by the number of people had a very strong correlation with that constructed by the World Development Index (WDI) based on the 2$ poverty line. The final estimate of global poverty obtained through this method was 2.2 billion, lower but comparable to the 2.6 billion calculated by the WDI, with the underestimation probably due to the difficulties in accounting for settlements with very low emissions in rural areas.

A paper from the following year, Doll and Pachauri (2010), focuses exactly on rural areas, specifically trying to derive an estimate of the rural population without access to electricity in the developing world. The authors rely on a combination of the DMSP OLS, gridded population of the world data and the residual areas from urban masks to estimate the share of rural population without electricity access in different countries. They then compare the figures with those from the International Energy Agency, finding that the former generally overestimate the latter. The two main reasons for the overestimation are found in the very low density of population of rural unlit areas, which is often missed by dataset assuming homogeneous population density across administrative boundaries, and in the energy use of rural areas being undetectable by satellites, either because not dense enough or because mostly happening indoors.

Another two papers from the end of the 2000s use night light data as a mean to improve on the calculation of GDP at sub-national level, either to assess the contribution of informal economies in different states (Gosh *et al.* 2009), or to further try to spatially distribute the contribution of agriculture (Gosh *et al.* 2010). The motivation of the first paper lies in the fact that many economic activities, especially in countries at lower level of income, take place in informal settings which will not be reflected in official GDP estimates but which might influence energy consumption and light emitted. By analysing the differences between official figures and those obtained through country-calibrated models based on night lights, one can have a further approximation of the contribution of informal activities. After developing a model based on the US and "blindly applying" it to Mexico, a country in which the relevance of informal activities is well established, the analysis of the results suggests to the authors that official estimates of their contribution might seriously underestimate their relevance. While recognizing that the methodology needs to be further developed, the authors stress how a refined model developed through the use of night time data might be useful to address many different socio-economic variables (Gosh *et al.* 2009).

In the second paper, the authors expand the previous sub-regional analysis to include also China and India and apply it globally at a national scale. The authors also develop a method to spatially distribute the contribution of agriculture to GDP, relating it to population density so to obtain a global map for economic activity for the year 2006 similar to that developed by Doll *et al.* for the population (Doll *et al.* 2000, Gosh *et al.* 2010). Despite recognizing that there are limited ways to verify the procedure, apart from comparing its results to the available sub-national figures for the US, Mexico, China and India, the initial attempt is deemed promising and seems to point towards a general underestimation of GDP in almost all countries of the world. Given the frequency with which satellite data becomes available, the method offers a further way to analyse the trends in the global distribution of wealth.

Ma *et al.* 2011 provide instead a comparative analysis of how the relationship between night time emissions and different socio-economic variables varies across cities in the same country using data from 200 prefectural level Chinese municipalities (Ma *et al.* 2011). Specifically, the authors look at which functional forms amongst linear, power law and exponential best fit the relationship between the average night-light emission of each city and its population, GDP, built-up area and electric consumption. As urbanization is a complex phenomenon, encompassing a nexus of dynamics between demographic pressure, land cover and energy flow, the authors are not surprised that different functional forms best fit different relationships, with night light-population better described by a linear form and night light-GDP/electric consumption better described by exponential forms.

Two papers from the beginning of the 2010s, Xi and Nordhaus 2011 and Henderson *et al.* 2012, reassess the contribution that night light can give in the adjustment of GDP figures at a national and sub-national level in light of the longer data series now available and of a more explicit modelling of the relationship between the measurement errors in the two data generation processes (Xi and Nordhaus 2011, Henderson *et al.* 2012). These authors confirm the previous finding that the procedure adds very little to GDP estimates from developed countries. However, the analysis for countries whose statistical offices receives the lowest grades in either indexes of the Penn World Table or of the World Bank points to significant contribution to be had by augmenting the figure with nigh light data, with the estimates of Henderson *et al.* suggesting that real GDP growth for the 30 countries with the lowest grade might differ by up to 3.2% per year.

Following these last improvements in methodology, a series of recent papers have been finding different empirical applications for night light data across the globe, and we will review those referring specifically to the African continent. To start with, two works from the same authors, Michaloupoulos and Papaioannou 2013 and 2014, combine anthropological data on the different political and economic traits of various African ethnic groups before colonisation with contemporary sub-national development as proxied by night light data. In the first work, the authors explore how pre-colonial African institutions and ethnicity have shaped regional development. Their results show a strong positive relationships between pre-colonial political centralisation (i.e. the existence of an early state hierarchically organised) and modern regional development, robust to the inclusion of different control, to diverse estimation strategies and to different level of analysis (i.e. original ethnic homeland or pixel level), while other pre-colonial characteristics seem to be insignificant.

In their second work (Michaloupolous and Papaioannou 2014), the authors exploit the same data to answer the opposite question, namely how contemporary national institutions have influenced the development of regions originally belonging to the same pre-colonial homeland partitioned in different states when the African border were drawn. The analysis confirms their previous findings, as the economic performance of areas who belonged to the same ethnic homeland, again as proxied by night light, does not systematically differ across borders regardless of differences in national institutions, and also in this case the findings are robust to a different specifications and estimation strategies. The results are further qualified by a more in depth analysis of the ethnic groups for which national institutions, contrary to the overall pattern, significantly shape economic performances. The reason for this heterogeneous effect is individuated in the closeness of both the partitioned area of the pre-colonial state to their respective capital cities.

Finally, Mveyange 2015 explores the extent to which income inequality at a regional level can be proxied by regional inequalities in night lights, given that the latter have been proved to be valid proxies for both income and wealth, with an application to the whole of Africa. Combining spatially explicit income data available for 32 of the 54 African countries with their respective DMSP-OLS, the author first derives inequality indexes for both at a regional level, to subsequently establish in a panel setting a significant and robust relationship between the two. This allows him to extend the analysis to the whole continent and to show the existence of heterogeneous inequality trends across different regions. Although no general trend exists between 1992 and 2003, a decrease in inequality can be observed between 2004 and 2012, although the effect of the financial crisis of 2008-2010 is noticeable. Moreover, these trends appear to be mostly dominated by between-regions inequalities, as within-region inequality across all countries analysed has been fairly low in the period under consideration.

To conclude, over the last 35 years DMSP-OLS data has been used to investigate a growing number of socio-economic relationships. Many of the first studies have been concerned with the urbanisation process in more developed economies, as the data lent themselves to the investigation of the expansion of urban areas over time in countries were night light emission was intense enough to be picked up with the technology of the date. The necessity to account for differences amongst the economic activities taking place in diverse countries and cities, so to properly assess how much they were expanding, was one of the main results of this first wave of analyses. Once the link between night lights and extension of cities was firmly established and the improvement in satellite technology allowed to detect the more feeble lights characterising remote areas of developing countries, it was only natural to move onto the analysis of population patterns at national and global scales, as cities represent the most densely populated areas on the planet and now rural population could be accounted for.

The data proved to be a good starting point also for this type of analyses, so that the successive step was to connect night -light with measures of national and sub-national economic activity, as urban conglomerates are generally the most productive areas of any country. Various analyses confirmed that DMSP-OLS can indeed be used as a proxy for both national and sub-national GDP, and that especially for poorer countries, in which statistical offices are not so-well developed and informal activities account for a relevant share of the economy, the use of night light data might lead to significant improvements over official GDP estimates.

At this point, a series of empirical investigations of the effects of different socio-economic variables on sub-national development became possible for countries in which regional economic data was previously available. It is to this strand of the literature that this chapter contributes. To the best of our knowledge, this is the first work which analyses the city-wide effect of energy infrastructure in Africa using night light data with the aim of assessing their contribution to local development. Following various examples in the literature, we also try to account for the process of urbanisation which affected SSA cities by experimenting with different level of thresholding, some common across all cities in the continent and some allowed to vary from city to city, so to explore how this might influence the results.

4.3 Data and Summary Statistics.

The two sources of the data used in the chapter are the DMSP OLS for the night time light values and the GeoSFM simulation for the water available for hydropower generation. As the latter has been extensively discussed in Chapter 2 and Chapter 3, we refer to those for the general presentation. The hydro-related variables used in the city level analysis are the same used for the analysis in Chapter 3, so that the only difference regards those used in the national level analysis, as the final step in their construction, joining each city to the nearest power plants[91], has been skipped and the anomalies of all hydro-power plants in the country have been used.

Beginning in the 1970s, satellites from the US Air Force DMSP have orbited around the planet 14 times per day registering the emission of light from earth. The OLS sensor mounted on the satellites is able to pick up both very low intensity visible and near-infrared (VNIR) light sources and thermal infrared, a range covering from fires and moonlit clouds to city lights. The data are first processed by the satellites, which smooth the finer 0.5 km resolution picture originally taken over a 5x5 area, and then further processed by researchers at NOAA, who clean them from a series of spurious light emissions which might be due to forest fires, auroral activities or other sources, so that man-made light remains the main source of emissions picked up. The presence of cloud cover is also taken into account, as heavy clouds might restrain the ability of the sensor to pick up VNIR emissions and light cloud cover might instead diffuse light creating a further source of spurious emissions. The final dataset are then constructed as an average over all nights of all orbits of each satellite, and can be accessed either in their raw format (i.e. as taken by the satellites, including light by fires, auroral activities etc.) or in their processed version, called “stable lights”, which is the one used in this study (an example for the year 2010 is presented in Figure 4.1).

[Figure 4.1 about here]

[91] Described in Capter 3, Section 2.3.

These data are represented in a grid in which each pixel, corresponding to 30 arc-second or slightly less than a square kilometres at the equator, is given a light-intensity value as a six-bit digital number, which is an integer ranging from 0 (or no lights) to 63. A couple of points have to be noted: first, the data cover from 180° W to 180° E longitude and 75° N to 65° S, hence excluding the whole of the Antarctic Circle and a sensible part of Artic Circle, although only approximately 10,000 people live in these areas. Second, the digital number describing each pixel is not perfectly proportional to the quantity of light emitted, partially because of sensor saturation, i.e. the incapacity of the satellite sensor to distinguish differences in intensity in the very bright cores of urban centres which are all coded with the maximum value of 63, and partially because the scaling factor which the sensor uses to convert light into digital numbers varies for unexplained reasons. This night light emission are then a reflection of both indoor and outdoor lighting, which are connected to a plethora of different human activities and are likely to be influenced by socio-cultural characteristics varying across the world, so that cross country analysis has to account for these.

[Figure 4.2 about here]

The analysis in the chapter focuses on the subset of the 29 SSA countries which have at least some hydropower capacity installed (the same on which the analysis in Chapter 3 is focused, see Figure 4.2) during the period 2001-2013. To give an overview of how data vary across different countries, Table 4.1 presents the distribution of the digital numbers both for the whole set of countries included in the studies and for 6 selected countries, while Figure 4.3 to Figure 4.8 presents their respective DMSP OLS pictures for the year 2010.

[Table 4.1 about here]

[Figure 4.3 to 4.8 about here]

As can be seen from Table 4.1 and from Figure 4.2 to Figure 4.8, the vast majority of pixels in all countries in the study are unlit (97.58%), with the lowest fraction of unlit pixels at a country level being the one reported for the Republic of South Africa (RSA, 86.23%), followed by that of Nigeria (90.85%, not reported), while the highest fractions are those of DRC and Mali (both 99.75%, not reported). Looking instead at the most luminous pixels, those with a value of 63, they represent only 0.02% of all the pixels included in the study, with the highest share at a country level being that of RSA (0.2%), followed by that of Ghana (0.07%) and that of Nigeria (0.06%), while five countries (Burundi, Central African Republic, Guinea, Liberia and Sierra Leone) do not have a single pixel with a value of 63 across the whole period. This implies that some of the problems which other studies (for examples Xi and Nordhaus 2011 or Ma *et al.* 2011) have reported with sensor saturation will be almost irrelevant in our application as the core of urban areas are not bright enough for top-coding to be a problem.

Table 4.1 also reports the sum of nigh light intensity averaged over the country dimension and its average yearly growth rate in the period 2001-2013, both considering and excluding unlit cells. In the case of the average value including unlit cells, the highest is that of RSA (1.48), followed by Nigeria (1.14) and Ghana (0.68); if we exclude unlit cells, the highest value becomes that of the Republic of Congo (13.67), followed by DRC (12.97) and Gabon (12.54). The growth rate also varies significantly across the continent, with Gabon representing the only country having experienced a negative average growth rate over the period when unlit cells are included (-1.34%), while Liberia has experienced the most dramatic average increase in light intensity (+21.52%). Nigeria represents instead the only country in which the average growth rate excluding unlit cells is negative, although only just (-0.79%), while the fastest one is that of Togo (+6.59%).

As the original hydro-related variables were developed for a series of cities across the countries included in the study, the main interest is investigating the hydropower-night light intensity relationship at that level. SEDAC, a data centre part of NASA's Earth Observing System Data and Information System hosted by the Center for International Earth Science Information Network at Columbia University, provides a raster representation of urban areas at 30 arc-second developed through a combination of population counts in the year 2000 and night time light. Specifically, where city buffers were not already provided by the national statistical agencies, all contiguous lighted cells for which the total combined population was greater than 5,000 inhabitants were considered as urban areas. Joining these rasters with the centroids of the administrative boundaries of the Gridded Population of the World, also developed by SEDAC, has permitted to identify the urban extent of 83 of the 133 cities included in Chapter 3.[92] Table 4.2 presents then the same statistics of Table 4.1 but only for the cities of the selected countries.

[Table 4.2 about here]

As it can be seen from Table 4.2, once the analysis is moved to the city level the relevance of unlit cells decreases significantly, as their average share across all considered cities is now

[92] No cities could be individuated following the described procedure in Liberia and Togo, while cities other than the capital could be individuated for the Central African Republic and Guinea.

2.41%, while the majority of lit cells, both overall and in most urban areas of considered countries, are in the 21-62 range. The highest average value of all considered urban pixels is that reported for Angola (42.33), followed by that of RSA (38.05) and that of Ghana (35.77), while the lowest are those of Burundi (14.76), Sierra Leone (14.71) and Central African Republic (8.8). The lowest growth rate of light intensity in the sample is that of Zimbabwean cities (+0.06%), while the fastest is that of Guinean ones (+62.01%); if we focus only on capital cities, the growth rate of Harare, Zimbabwe, remains the slowest one (+0.15%) while that of Kigali, Rwanda, becomes the fastest (+6.56%).

As the urban masks used for the above analysis were developed with reference to population and night time data of the year 2000 and most countries in SSA have been experiencing high urbanization rates over the last 16 years, we have decided to try and account for the likely expansion of the urban boundaries. Two different procedures have been used. In the first case, the lowest average value of light intensity amongst all the original boundaries of the capital cities for each year has been used as benchmark, and all the pixels contiguous to the urban boundaries having that value have been added to the boundaries determined in the previous year (that is, a single value has been used to update the boundaries of all cities of all countries every single year). We also repeated the same exercise with half of the average values and with a quarter of it as alternative cut-off points for the expansion of the boundaries.

The actual procedure consists of, after transforming the original DMSP-OLS picture into an array of points, each representing a pixel of the original, first selecting all points falling within the original city boundaries, then selecting only the points with night-light values higher than the benchmark, then merging the two together and reshaping the points as polygons. If any of the points/pixels contiguous to the original boundaries were equal to or higher than the benchmark, they will now be part of the expanded city polygons, while all the spurious polygons corresponding to other cluster of lights scattered through the countries were dropped.

There are some suggestions in the literature that each country (Sutton *et al.* 2001), if not city (Ma *et al.* 2011), should be analysed on its own. The reasons for this lie both with different cultural uses of light at different levels of standard of living (e.g. how long after sunset would a child keep on reading a book, presence of garden lights to deter burglars and so on) and with a different association between night light and economic activity at different stages of industrialization (e.g. relevance of chemical sector and other heavy industries whose machinery is never shut off), so that an unique relationship between night light and economic activity is unlikely to exist. This implies that the previous procedure, by using a unique threshold for expanding the boundaries of all cities across the continent, is likely to either over- or underestimate the urbanization rate in many cases. To account for this, for the second procedure we have then restricted the analysis to only the capital of each country in the study, and for each year we have updated each capital's boundaries using the specific mean value (or a half or a quarter of it) of the pixels contained in its original boundaries, so to have a single threshold for each city every year.

[Table 4.3 about here]

[Figure 4.9 to Figure 4.14 about here]

Table 4.3 presents a comparison of the results of the two procedures across all the capitals in the study and those of selected countries. The first thing to notice is how the average values of lights when the capital boundaries are expanded using the mean value in both procedures is always very close to those of the original boundaries: the average across all original boundaries is 30.0622 (see Table 4.2), that across boundaries expanded with the first procedure 30.587, that across boundaries expanded with the second procedure 29.7007. Secondly, we can notice how changes in the average values obtained using different threshold (i.e. half or a quarter of the mean instead of the mean) vary depending on the procedure used, and specifically how using city specific values leads to smaller changes than using a single value across the whole continent. Similar differences can also be noted amongst the growth rates of light intensity.

This is made clearer by Figures from 4.9 to 4.14. Figure 4.9 shows Accra, the Ghanaian capital, in 2010 when the boundaries have been expanded using a single mean value common for the whole continent, while in Figure 4.10 the Accra-specific mean has been used; Figure 4.11 and Figure 4.12 present the same for half the mean (common and specific respectively) and Figure 4.13 and Figure 4.14 for a quarter of the mean (again common and specific respectively). The boundaries of the city expanded using the common mean (Figure 4.9) are slightly larger than that obtained using half of the Accra-specific mean (Figure 4.12), and those obtained by using half of the common mean (Figure 4.11) are slightly larger than those obtained by using a quarter of the Accra-specific mean (Figure 4.13). This confirms what it has been previously noted, most recently by Ma *et al.* 2011, i.e. that relevant differences in the analysis will arise depending on the choice of using or not country or city specific values when updating urban boundaries.

4.4 Methodology.

Our aim is to estimate the impact of the quality of energy infrastructure on sub-national development, using water availability for hydro-generation as measure of the former and night lights captured from space as proxy for the former. We use the following specification:

$$Y_{it} = \alpha + \beta_1 X_{it} + \varepsilon_{it}$$

where Y_{it} is the yearly average night light intensity in the city or country i (depending on the setting) in the year t, furthermore, for the country level regressions we will consider both an average calculated using unlit cells and one without them given their relevant share in all countries in the sample (the minimum value, that of RSA, is 86.23%); X_{it} will alternatively be the yearly mean anomaly or the same set of indexes for weak/strong positive/negative anomalies used in Chapter 3 for the city or country i in the year t. The models are firstly estimated in level as pooled OLS regression, and then, to correct for possible sources of endogeneity, both as first-difference in a pooled OLS setting and in a panel OLS setting with fixed effect

Alternative specifications explored include a restriction of the sample to different subsets of countries in which hydropower represents at least a given share - either 30%, 40%, 50% - of installed capacity, listed in Table 4.4; the substitution of year fixed effects with country- or city-specific linear time trends; finally, quantile regression will be performed instead of panel regression to explore if the effect of variations in water available for hydro-power changes with light intensity.

[Table 4.4 about here]

4.5 Results.

4.5.1 City level.

The first results to be presented are those relative to the 83 cities for which we were able to determine the urban boundaries, so that both the average night light intensity and the hydrological variables refer to the city-level. Tables 4.5, 4.6 and 4.7 present the benchmark results for the specification using the yearly mean anomaly, first in level (Table 4.5), then in first difference (Table 4.6) and last with fixed effect (Table 4.7). As can be noted from all tables, the yearly mean anomaly is always insignificant, regardless of the weight, the radius or the estimation strategy used.

[Table 4.5 to 4.7 about here]

Tables 4.8, 4.9 and 4.10 present instead the results for the specifications using the same indexes developed in Chapter 3, which further disaggregate water available amongst strong/weak and positive/negative shocks. As a higher value of the indexes for the negative shock implies more water available for hydro-electricity production (as these are upper bounded to 0, the situation in which the streamflow is equal to the long-term mean), we expect both of them to take positive signs, while we expect the strong positive shock index to take a negative sign because in flood situation damages can be incurred by the hydro-turbines. We would instead expect the weak positive shock index to be insignificant, as having a slightly higher flow than the long term mean should not have an influence on the generation of hydro-electricity.

These expectations are though not met by the models' results, as across all three estimation strategies and all different weights only twice we have some significant coefficients, and then only at 10%: that on the weak positive anomalies index, which we expected to be insignificant, in the level form regression (Table 4.8) and that on the strong negative anomalies index in the first difference regression (Table 4.9), although only for the version of the instrument calculated using only two radiuses. None of the hydrological variables are ever significant in the fixed effect regression (Table 4.10).

[Table 4.8 to Table 4.10 about here]

All the above regressions do not account in any way for the process of urbanization which interested the continent during the period under consideration. In Table 4.11 to Table 4.16 we then replicate the exercise after having adjusted the data so to account at least crudely for this process following the first procedure outlined in the third section of the chapter, that relying on the lowest average light intensity amongst the continent capitals. Again as before, no effect is to be found when the yearly average anomaly is used as regressor, regardless of weight, radius, estimation strategy or definition of urban growth, as all coefficients in Tables 4.11 to 4.13 are always insignificant. Tables 4.14 to 4.16 also closely resemble those where the urbanization process is ignored, as again the only significant variables are the weak positive anomaly index and the strong negative anomaly index, and also in this case only at 10%.

[Table 4.11 to 4.16 about here]

Finally, Tables 4.17 to 4.22 present the results for the models in which the boundaries of the capital cities have been expanded using city-specific values. Once again, the results closely resemble those run on the original sample. In this case, the only difference lies in the fact that now the only disaggregated hydrological variable which remains significant is the weak positive anomaly index in the first difference regression of Table 4.22, exactly the one we expected to be insignificant

[Table 4.17 to 4.22 about here]

From our baseline specifications it appears then that we cannot find a stable direct effect of the availability of water for hydropower generation on the general economic activity in SSA, as the first of our measures of water availability, the yearly mean anomaly, is insignificant across all specifications, while only two of our four disaggregated indexes are sometimes significant, one of which we expected to be irrelevant and in any case often only at 10%, and further show an erratic behaviour across the different models.

4.5.2 Alternative specifications.

Three main alternative specifications have been tried to explore possible differences from the baseline estimates: the first entails the restriction of the sample to cities in countries which have at least a minimum share of installed generation capacity taking the form of hydro-power plants; the second consists of switching the year dummies with country specific time trends and the third in using quantile regressions instead of pooled or panel OLS. Furthermore, all alternative specifications have also been tried on the urbanization-adjusted datasets, but, as for the benchmarks, in all cases the results are extremely similar to those which ignore the process of urbanization so that they will be reported only when significant divergences exist.

The first set of alternative results to be presented are those regarding the restriction of the sample to only the cities in countries in which hydropower represents at least a given share of the installed capacity (30%, 40% and 50%).[93] Tables 4.24 to 4.26 present the results for the specifications in which the simple yearly mean anomaly was used (in level, first difference and fixed effects respectively), Tables 4.27 to 4.28 present those in which the disaggregated hydrological measures have been used (again in level, first difference and fixed effects respectively).

[Table 4.24 to 4.28 about here]

As can be seen from Tables 4.23 and 4.24, all the coefficients on the yearly mean anomaly appear to be insignificant in both the level and the first difference regression regardless of the weight and radius used or of the threshold applied. On the other hand, both measures using four radiuses have the right sign and are statistically significant, although only at 10%, in the fixed effect regressions when we consider only countries in which hydropower represents at least

[93] See Table 4.4 in the methodology section for a list of the countries included in each sample.

50% of the installed capacity (Table 4.23). The results relative to the specifications using the disaggregated forms of the hydrological variables closely resemble those of the baseline specifications, as almost none of the indexes is ever significant, excluding again the strong negative anomaly index in the first difference regression on the sample of cities for which hydropower represents at least 30% of the installed capacity when only two radiuses are used, and also in this case only at 10% (Table 4.27).

This is the specification for which the expansion of the urban boundaries plays a significant role, at least when they are updated using the capital-specific values of the second procedure. Table 4.29 and Table 4.30 present the results for the case in which we estimate the model with panel fixed effects, the explanatory variable is the yearly mean anomaly and urban boundaries have been updated using the mean light intensity in the core-area of the capital or half the mean light intensity respectively. As it appears from both tables, once the city boundaries are updated there is a significant association, although often only at 10%, between an higher value of yearly mean anomaly and an higher intensity of night-light, and this could already be perceived from the restriction of the sample to countries in which hydropower represent at least 30% of installed capacity. Furthermore, as we would expect, the higher the relevance of hydropower in the energy portfolio of the country, the stronger the effect which water availability for electricity generation has on light intensity. When the boundaries are updated with a quarter of the mean light intensity the relationship turns insignificant again, and similarly no significance could be found for the regressions in which the indexes were used.

[Table 4.29 and 4.30 about here]

These results indicate that a relationship between hydroelectricity production and night-light intensity, and hence economic activity, might exist, although we are able to individuate it only for the capital cities of countries in which hydropower represents a significant share in the generation portfolio. In addition, it is necessary to also consider how the urbanisation process which interested the SSA region over the last decade is accounted for, as the estimates are significant only in the case in which the urban boundaries of the capital cities are updated with city specific values.

Next we focus on the specifications in which the time dummies are substituted by city-specific linear time trends. Tables 4.31 to Table 4.33 present the results for the regressions using the yearly average anomaly as explanatory variable (in level, first difference and fixed effect respectively), while Tables 4.34 to Table 4.36 those in which the disaggregated forms of the variables are used. Again, the results are not particularly instructive, as there is no significant relationship between the yearly average anomalies and night lights across all specifications, while in the case of the more disaggregated indexes we have both the strong positive anomaly index and the weak negative anomaly index significant in the level regression (Tables 4.34), the first with the wrong sign and the second when expected insignificant, while in the first difference regression the strong negative anomaly index turns positive and significant (Table 4.35).

[Table 4.31 to 4.36 about here]

The final set of results for the city-level analysis is the one in which we switch the estimation method from OLS to quantile regression, as we are interested to see if the effect of a variation in the water available for hydropower generation changes with night-light intensity. Table 4.37 and Table 4.38 present the results for the case in which the yearly average anomaly is used as regressor in growth and level form respectively, while Table 4.39 and Table 4.40 present the equivalent results for the disaggregated forms of the variable. As it can be seen from Table 4.37, a growth in the yearly mean anomaly is almost always significantly and positively related to a growth in night light regardless of the weight or radius used, with the coefficients decreasing in size as the quantile increases, suggesting that the influence of variation in water availability is more relevant for dimmer than for brighter lights (consistently with the load-base nature of hydropower). A significant, although less stable, relationship can also be found in the level form regression (Table 4.38), although in this case the coefficients' size grows bigger with the quantile. Table 4.39 and Table 4.40 on the other hand seem to show a very significant relationship between all disaggregated form of the hydrological variables and night lights, but a more careful examination reveals sign switching from quantile to quantile, so that most variables take the wrong sign in at least one quantile of each specification. This does not seem to be restricted to any particular form of the dependent variable (growth/level) or weight used for the construction of the hydrological variables (installed/operational), so that it becomes hard to give any satisfactory explanation.

[Table 4.37 to Table 4.40 about here]

4.5.3 National Level.

We now move away from the city level analysis to focus on the national level. In all the following regressions the night light intensity has been averaged over the entirety of each country, while for the construction of the hydrological variables all hydropower plants have been considered, so that each power plant weight corresponds to its contribution to the national generation portfolio. The first results to be presented are those relative to the regressions having as dependent variable the average night light intensity, either in level (Table 4.41), in first difference (Table 4.42) or with fixed effects (Table 4.33). The average has been calculated including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4.

[Table 4.41 to 4.43 about here]

As it can be seen from the above tables, changes in the yearly average surplus or deficit of water available for hydropower generation are never significantly related to changes in night light intensity, regardless of the weight used for the calculation of the hydrological variables (installed or operative capacity), of the exclusion or inclusion of unlit cells, or the estimation strategy. In Table 4.44 and Table 4.46 the disaggregated forms of the hydrological variables are used instead.

[Table 4.44 to 4.46 about here]

The picture emerging from these regressions resemble those of the city level regression in which the disaggregated indexes were used. In the level regression, no index is ever significant when the dependent variable includes unlit cells, while if unlit cells are excluded there is, contrary to what expected, a negative and significant relationship between night light intensity and the weak negative anomaly, although only at 10% if the operational capacity is used as weight (columns 3 and columns 4 of Table 4.44). All variables are instead insignificant when the equivalent regression in first difference is run (Table 4.45), while when it is estimated with fixed effects the only significant result is a positive association between the values of the strong negative anomaly index and that of night lights when unlit cells are excluded, although only at 10% when the installed capacity is used as a weight (Table 4.46, columns 3 and 4), this time in line with the original hypothesis.

It is worth noting that we have also tried to include, alternatively, only the indexes for the weak anomalies or the indexes for the strong anomalies (and for this latter case we also tried each index at a time). In all those cases, the estimates are very closely in line with those reported, and as we do not feel that these models add significantly to the analysis we have decided not to report the results. Similarly, all the alternative specifications which were considered for the city-level analysis have also been applied to the country-level one, but they give a limited contribution to the understanding of the relationship between hydropower production and night light intensity, so that we leave their discussion to the Appendix A.4.

4.6 Conclusions.

In the last two decades, the increase in quality and frequency of satellite-based data collection efforts has allowed the exploration of a series of questions relevant for social scientists and policy makers which could not previously be tackled. Specifically, observation of night-light emitted by human settlements from space has led to an increased understanding of the geographical distribution of human population and of the drivers and effects of an ever increasing urbanization trend. More recently, the focus in the literature has been on establishing a link between night light and GDP at both national and sub-national level, and various analyses have proven that this data is indeed a good proxy for economic activity, especially in developing countries. More nuanced empirical investigations have swiftly followed.

The chapter presents the results of the initial attempt of connecting the intensity of night-light from SSA cities and countries with the water available for hydro-electricity production, to the best of our knowledge the first time that this data has been used to analyse the city-wide economic effect of infrastructure in Africa. Two different line of inquiries have been attempted, carrying out the analysis for the period 2001-2013 both at the city level for 83 of the 113 cities which composed the sample of Chapter 3 and at the national level for 29 countries. Furthermore, as suggested by the literature, we have developed a method to update the boundaries of the cities included in the study in order to account for the changes brought about by the high urbanization rates which characterised the continent in the last two decades.

Overall, the results presented in the chapter point towards an irrelevant contribution of hydropower production to the city- and country-wide economic activity as measured by night lights. Some signs of the existence of the relationship under investigation have emerged, but they are not conclusive and interest only the city-level analysis, where it has proven important which procedure to update the city boundaries is chosen and which countries enter the analysis. Specifically, the only significant results are those for the capital cities of countries in which hydropower represents a significant share of installed capacity (>30%) if the urbanization trends over the period 2001-2013 are allowed to vary from city to city. In this case, fixed effect estimates using the yearly average anomaly are significant and with the expected sign (see Appendix B4 for a summary of the results). On the other hand, estimates in which the more disaggregated hydro measures have been used have proven less stable than the latter also in this case. Similar points can be made for the national level analysis, in which signs of the same relationship between the yearly average anomaly and night light intensity could be found only in some alternative specifications, and only when unlit cells were excluded, discussed in Appendix A.4.

Future improvements on this analysis will then have to continue the work started in the chapter by further expanding the sample of cities for which a specific urbanisation trend is constructed from only the capitals to all cities, to see if this influence the results, both overall and for countries in which hydropower represents a significant share of installed generation capacity. Moreover, when better and updated household and industrial connection data, or data relative to the energy infrastructure in general, will become available, new procedures of scaling the night-light intensity by usage/capacity of the network instead than by dimension of the country could also be attempted.

Appendix A.4.

In this appendix we present the same set of alternative specifications of the country level analysis which were considered for the city level one. Tables A4.1 to A4.6 present the results relative to the first alternative, the restriction of the sample to countries in which hydropower represents a significant share of installed capacity, when the simple yearly mean anomaly is used. Three different threshold levels have been used: in columns 1 and 2 of each table countries with less than 30% of installed capacity consisting in hydropower have been excluded from the analysis[94], in columns 3 and 4 the threshold is increased to 40% and in columns 5 and 6 it is further increased to 50%.

[Table A4.1 to A4.6 about here]

As the tables makes clear, the restriction of the sample does not lead to any noticeable change from the regression using the entire one, as the yearly mean anomaly remains statistically insignificant regardless of the weight used, the inclusion or exclusion of unlit cells and the estimation strategy. Almost exactly the same can be said of all regressions run on the restricted samples using instead the more disaggregated forms of the hydrological variables, shown in Tables 4.7 to 4.12. In these cases, the only significant variables are in regression where the dependent variable excludes unlit cells, namely the weak negative anomaly index in the level

[94] See Table 4.4 in the methodology section for a list of the countries included in each sample.

regression (with the wrong sign, Table A4.10) and the strong positive anomaly index in the fixed effect regression, this time with the right sign (Table A4.12).

[Table A4.7 to A4.12 about here]

The next set of results to be introduced is the one in which the time dummies have been replaced with country-specific (linear) time trends. As usual, the first results presented are those relative to the use of the simple yearly average anomaly, in level form in Table A4.13, in first difference in Table A4.14 and with fixed effects in Table A4.15. The main difference for this set of results appears in columns Tables A4.14 and A4.15: the yearly mean anomaly is now positively related to night light intensity regardless of the weight used, but only when unlit cells are included in first differences and only when they are excluded in with fixed effects.

[Table A4.13 to A4.15 about here]

Once again the results relative to the use of the more disaggregated hydro-variables (Table A4.16 to Table A4.18) present an erratic behaviour: all variables are insignificant in the level form regression (Table A4.16), only the weak negative anomaly index is significant in first difference (as in the other cases with the wrong sign, Table A4.17) and only when the dependent variable includes unlit cells, while only the strong negative anomaly index is significant with fixed effects, this time with the expected sign (Table A4.18).

[Table A4.16 to A4.18 about here]

The final set of results for the country-level analysis is the one in which we switch the estimation method from panel OLS to quantile regression, as we are interested to see if the effect of a variation in the water available for hydropower generation changes with night-light intensity. Table A4.19 to Table A4.22 above present the results for the specification using the simple yearly mean anomaly. This is the set of results using a single hydrological variable for which the inclusion or exclusion of unlit cells make the biggest difference: as it can be seen from Table A4.19 and A4.20, as long as they are included the yearly average mean is significant and takes the opposite sign from what expected in all but the highest quantile of light intensity (columns 4 and 8 in both tables); on the other hand, when unlit cells are excluded (Table A4.21 and Table A4.22) the yearly average anomaly is almost always significant and takes the wrong sign in only one case (column 1 of Table A4.21).

[Table A4.19 to 4.26 about here]

As in all the previous cases, the specifications using the more disaggregated hydro-measures (Table A4.23 to Table A4.26 above) are the least consistent, with sign switching from quantile to quantile and consequently m most variables taking the wrong sign in at least one quantile of each specification. This does not seem to be restricted to any particular form of the dependent variable (growth/level) or weight used for the construction of the hydrological variables (installed/operational), so that it becomes hard to give any satisfactory explanation.

Appendix B.4.

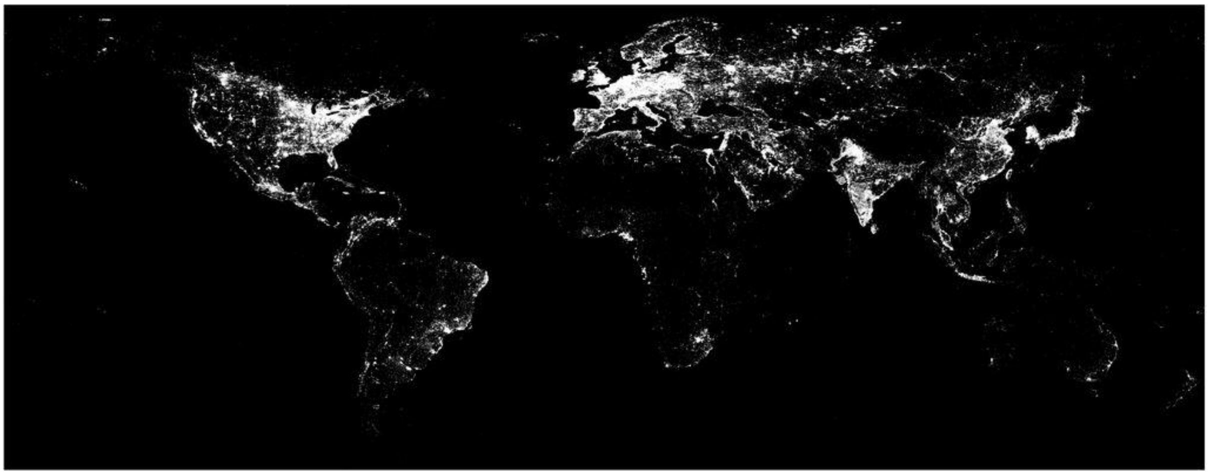
Although the overall results of Chapter 4 point towards the insignificance of hydroelectric production as a direct determinant of economic activity across SSA, the fixed effect regressions in the sample of capital cities of states in which hydropower accounts for a significant share of installed capacity yielded some stable and significant results. In these regressions the masks used to individuate the boundaries of the city were expanded through the procedure described in the end of section 4.3, so to account for the strong urbanisation rates characterising the African continent since the year 2000, year when the original urban masks were created.

Tables B4.1 and B4.2 report the coefficients for the effect of an increase of one standard deviation in the streamflow of the rivers serving hydropower stations connected to the capital cities on their average night light intensity, expressed in term of the digital numbers used to measure night light intensity. To better contextualise the results, the tables also report the average night-light intensity across the capitals in the sample and the average value of the measure for water availability. In Table B4.1 the urban masks have been expanded each year by including all contiguous pixels with a night light intensity at least equal to the mean value of the previous year, in Table B4.2 by including all pixels in which the intensity was at least half of this value.

[Table B4.1 and B4.2 about here]

As can be expected, the higher the relevance of hydropower in the installed capacity of the country, the higher is the coefficient, and similarly higher are all coefficients in which the mean anomaly is scaled by the operational capacity. For countries in which hydropower accounts for at least 30% or 40% of the installed capacity, when the urban masks are expanded with the mean night light intensity an increase of one s.d. leads to an increase in intensity of around 10% of the average city value. Although this effect might seem high, it must be noted that the average anomaly values, however scaled, are much lower than one full s.d,, so that an increase of this magnitude is unlikely to ever take place. These effects are of a lower size and significance when the boundaries are expanded with half of the mean value of the previous year mask, as in this case the areas of capital cities are bigger and include neighbourhoods with a lower average light intensity.

Figure 4.1 – DMSP OLS, 2010.



DMSP OLS composite for the year 2010.

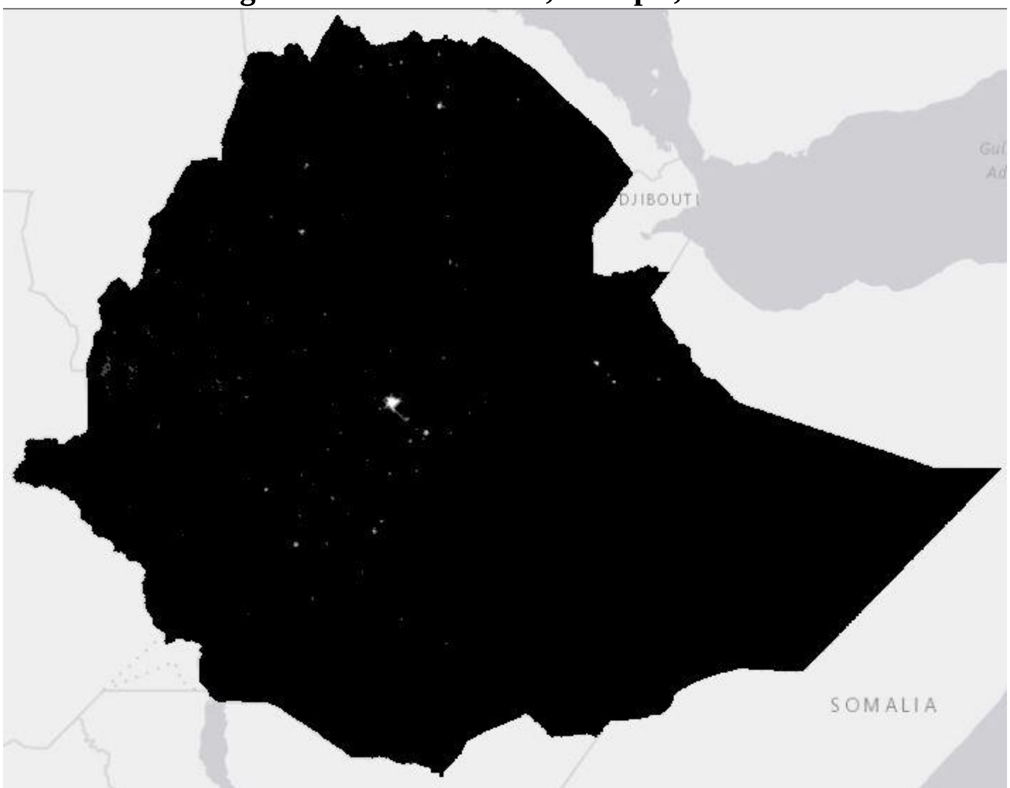
Figure 4.2 – DMSP OLS, countries in the sample, 2010.

DMSP OLS composite for the countries in the sample, year 2010.

Figure 4.3 – DMSP OLS, Angola, 2010.

DMSP OLS composite for Angola, year 2010.

Figure 4.4 – DMSP OLS, Ethiopia, 2010.



DMSP OLS composite for Ethiopia, year 2010.

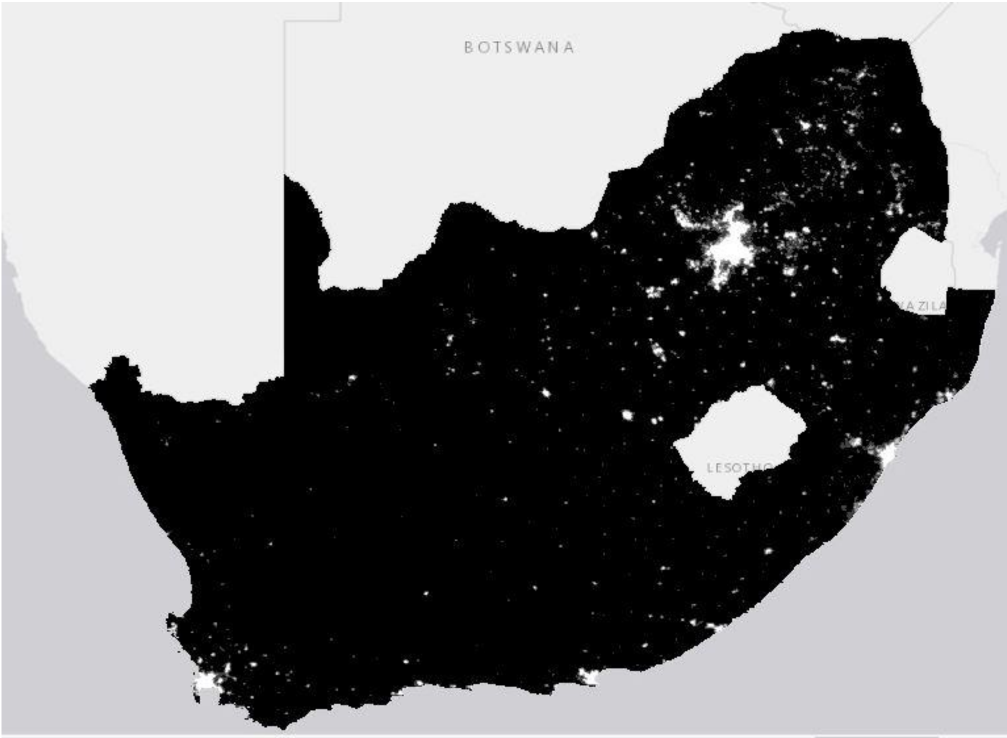
Figure 4.5 – DMSP OLS, Ghana, 2010.

DMSP OLS composite for Ghana, year 2010.

Figure. 4.6- DMSP OLS, Kenya, 2010.

DMSP OLS composite for Kenya, year 2010.

Figure 4.7 – DMSP OLS, South Africa, 2010.

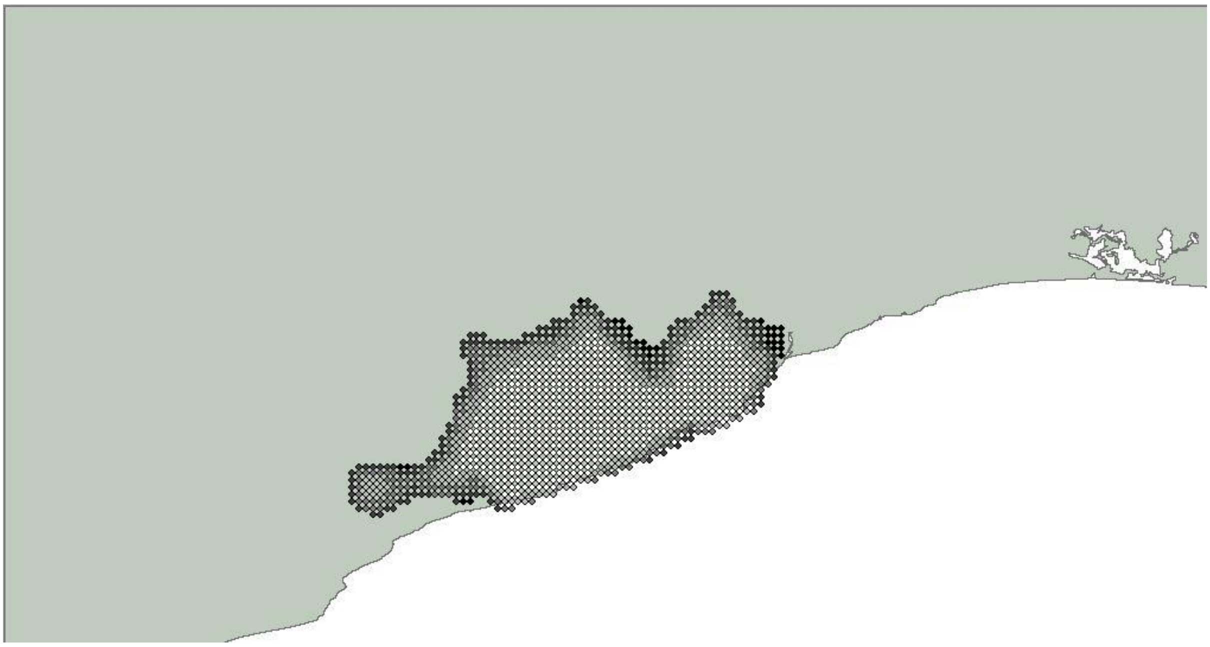


DMSP OLS composite for the Republic of South Africa, year 2010.

Figure 4.8 – DMSP OLS, Tanzania, 2010.

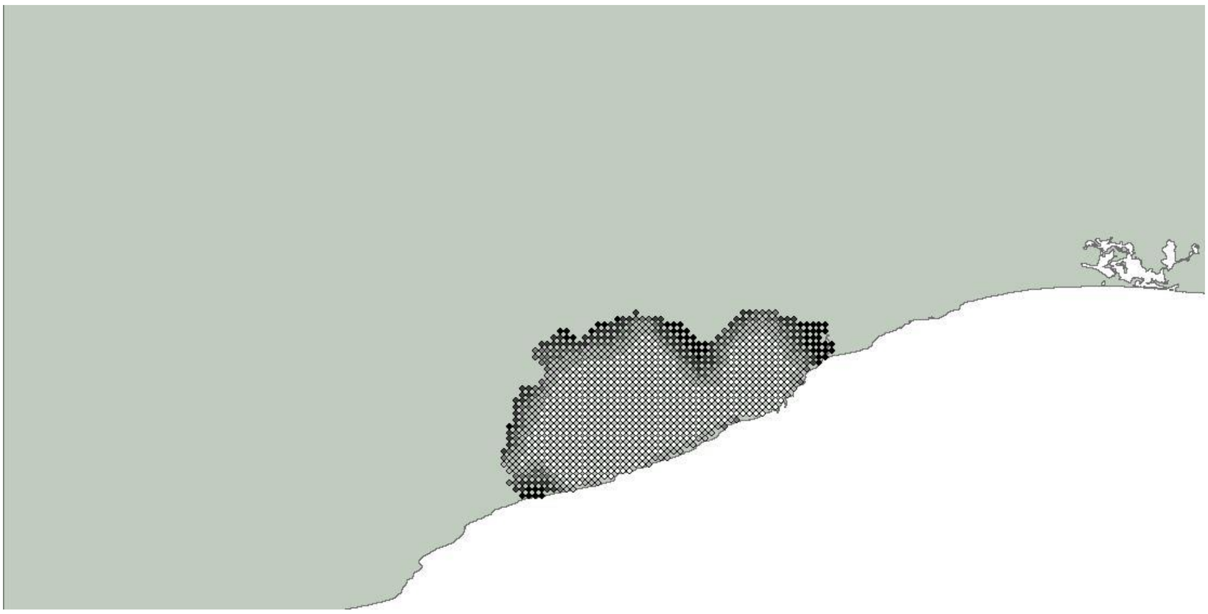
DMSP OLS composite for Tanzania, year 2010.

Figure 4.9 – Accra, continental benchmark, mean value, 2010.



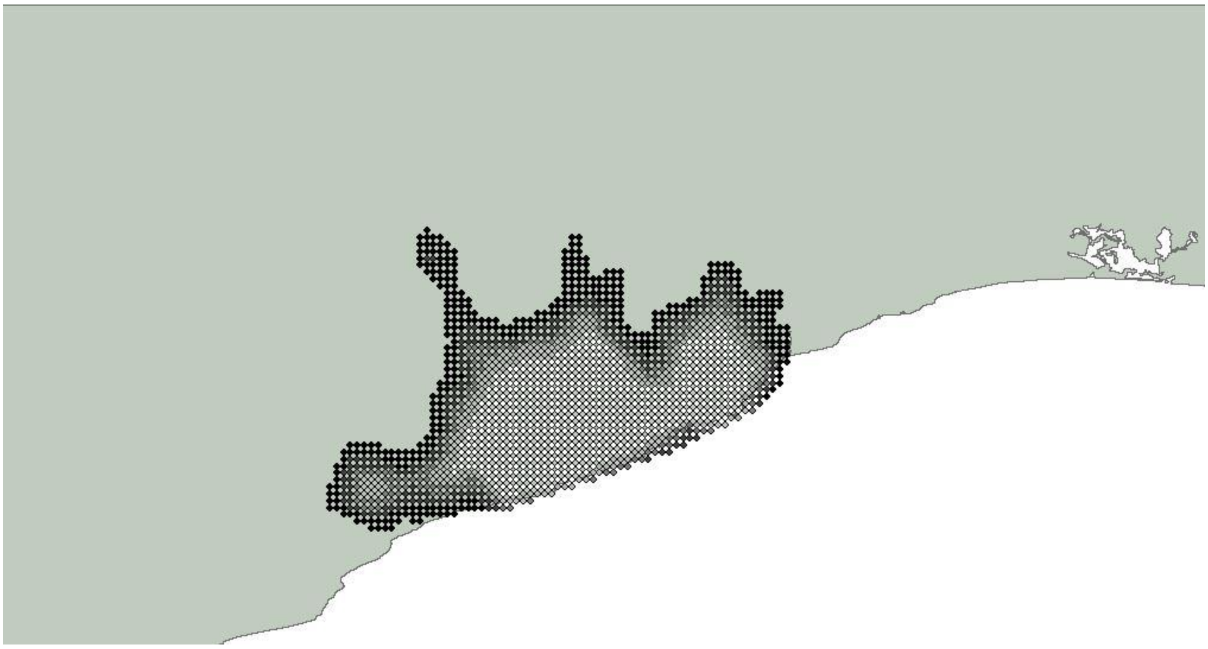
Dimension and luminosity of Accra obtained by updating city boundaries with a mean value common to the whole continent, mean value, year 2010.

Figure 4.10 – Accra, city specific benchmark, mean value, 2010.



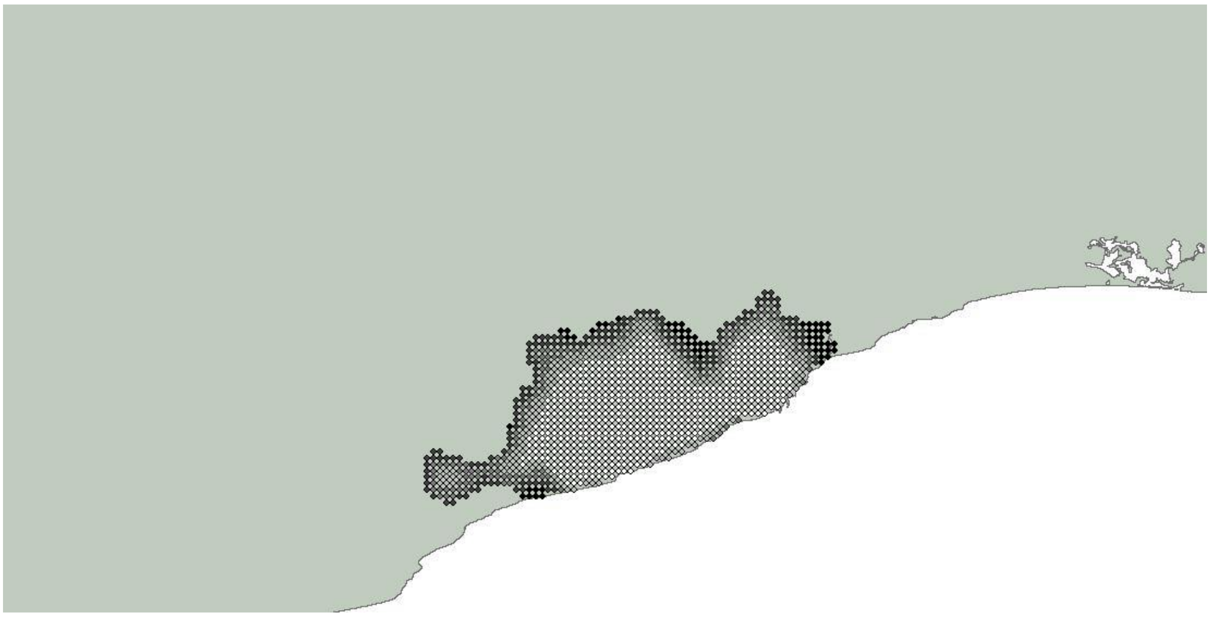
Dimension and luminosity of Accra obtained by updating city boundaries with a mean value specific to each capital, mean value, year 2010.

Figure 4.11 – Accra, continental benchmark, half the mean value, 2010.



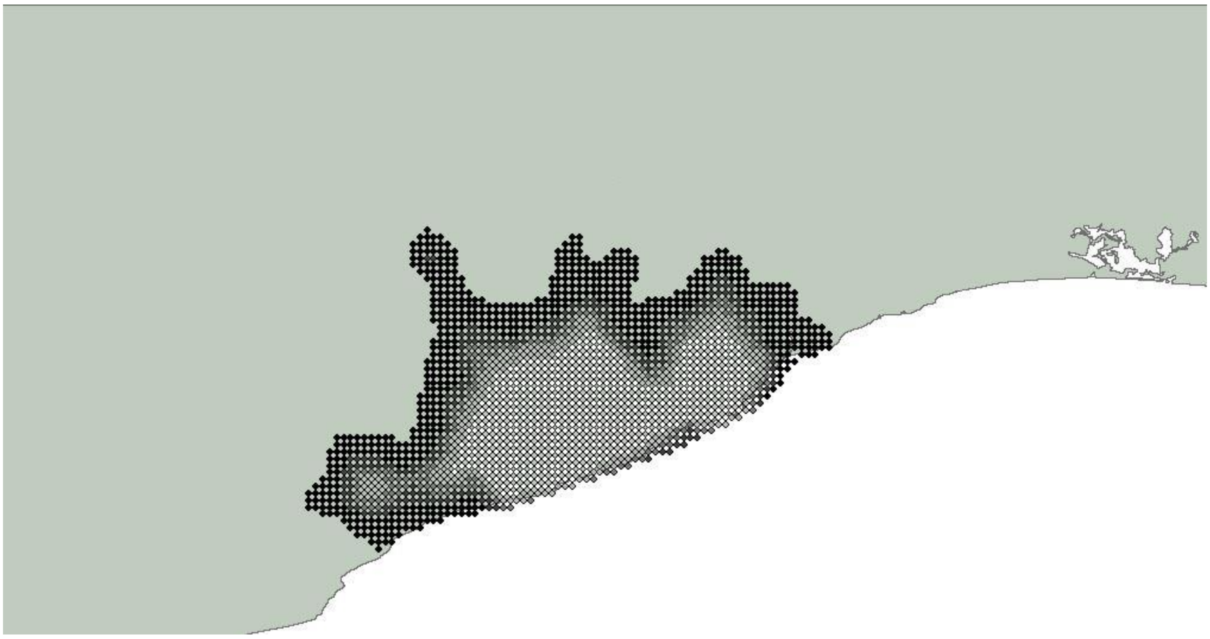
Dimension and luminosity of Accra obtained by updating city boundaries with a mean value common to the whole continent, half the mean value, year 2010.

Figure 4.12 – Accra, city specific benchmark, half the mean value, 2010.



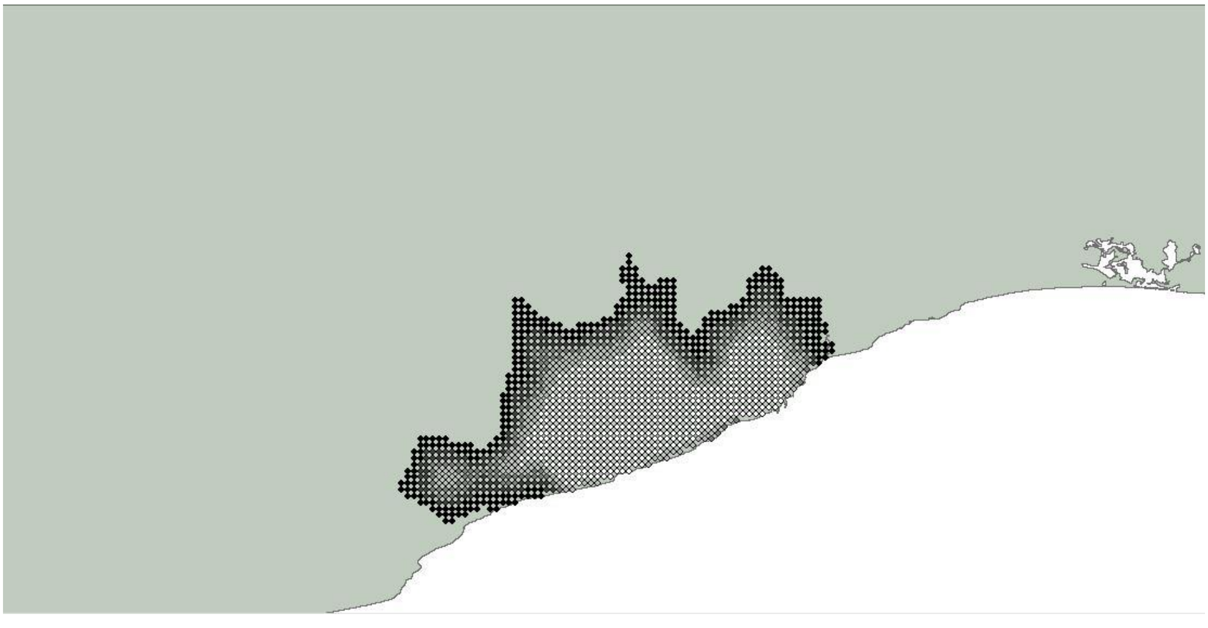
Dimension and luminosity of Accra obtained by updating city boundaries with a mean value specific to each capital, half the mean value, year 2010.

Figure 4.13 – Accra, continental benchmark, quarter of the mean value, 2010.



Dimension and luminosity of Accra obtained by updating city boundaries with a mean value common to the whole continent, quarter of the mean value, year 2010.

Figure 4.14 – Accra, city specific benchmark, quarter of the mean value, 2010.



Dimension and luminosity of Accra obtained by updating city boundaries with a mean value specific to each capital, a quarter of the mean value, year 2010.

Table 4.1 – Summary statistics, country sample.

	SSA	Angola	Ethiopia	Ghana	Kenya	RSA	Tanzania
0	97.58%	99.18%	99.28%	92.65%	98.08%	86.23%	99.17%
1-5	1.18%	0.27%	0.32%	3.74%	0.81%	6.72%	0.32%
6-10	0.70%	0.29%	0.28%	2.36%	0.77%	3.73%	0.34%
11-20	0.27%	0.10%	0.07%	0.62%	0.21%	1.62%	0.10%
21-62	0.25%	0.10%	0.05%	0.57%	0.13%	1.51%	0.06%
63	0.02%	0.02%	0.00%	0.07%	0.00%	0.20%	0.00%
Average	0.2608	0.0954	0.0646	0.6815	0.1723	1.4823	0.0792
Average Growth	7.20%	13.85%	9.52%	5.38%	6.74%	2.76%	6.66%
Average (no 0)	10.5700	11.4808	8.5338	9.2631	8.7962	10.7685	9.3608
Average Growth (no 0)	3.13%	5.50%	3.50%	4.42%	2.82%	1.91%	2.82%

Summary statistics and night light distribution for the full country-sample (SSA) and for selected countries.

Table 4.2 – Summary statistics, city sample.

	SSA	Angola	Ethiopia	Ghana	Kenya	RSA	Tanzania
0	2.41%	0.36%	7.64%	0.38%	1.56%	0.08%	4.47%
1-5	4.86%	4.17%	8.19%	1.72%	4.81%	2.23%	8.12%
6-10	16.56%	9.66%	22.22%	11.32%	25.84%	9.60%	22.25%
11-20	20.63%	11.29%	21.23%	18.86%	30.59%	17.95%	24.28%
21-62	46.52%	46.13%	39.79%	58.64%	36.99%	54.39%	37.73%
63	9.02%	28.39%	0.94%	9.07%	0.22%	15.76%	3.14%
Average	30.4054	42.3306	22.4019	35.7673	21.9318	38.0467	22.5120
Average Growth	5.49%	13.44%	6.79%	4.37%	5.10%	1.08%	4.14%
Average (capital)	30.0622	45.5484	28.6991	44.6935	23.9188	30.2293	32.1226
Average Growth (capital)	3.66%	6.07%	4.53%	2.83%	3.91%	0.91%	3.33%

Summary statistics and night light distribution for the full city-sample (SSA), and for the cities of selected countries.

Table 4.3 – Summary statistics, urbanization procedure.

	SSA		Angola		Ethiopia		Ghana		Kenya		RSA		Tanzania	
	Capital sp.	Lowest	Capital sp.	Lowest	Capital sp.	Lowest	Capital sp.	Lowest	Capital sp.	Lowest	Capital sp.	Lowest	Capital sp.	Lowest
Average (Mean)	29.7007	30.5870	45.5271	42.2562	28.6682	30.5817	44.8292	42.7686	23.8997	24.8176	30.3650	30.8768	32.4748	33.4775
Average Growth (Mean)	3.74%	4.14%	6.16%	4.21%	4.70%	5.39%	2.85%	1.71%	3.99%	4.57%	0.89%	1.77%	3.34%	3.39%
Average (Mean/2)	29.0396	26.5292	43.6617	34.9837	28.5323	27.0678	43.8529	34.8838	23.6566	22.5824	29.8491	24.4143	32.2134	29.7113
Average Growth (Mean/2)	3.41%	3.10%	5.51%	3.07%	4.59%	3.14%	2.33%	1.70%	3.74%	3.27%	0.75%	1.39%	3.08%	1.53%
Average (Mean/4)	25.0494	17.7081	38.4191	24.3567	24.7411	18.3615	39.4228	22.1276	17.9616	13.8965	24.7574	13.5878	28.9124	19.9193
Average Growth (Mean/4)	3.66%	2.10%	6.10%	1.65%	0.19%	0.71%	0.30%	2.85%	3.32%	2.46%	0.24%	0.15%	-1.02%	1.68%

Average value of night light and of night light yearly growth for the full sample (SSA) and for the capital of selected countries. Columns under the heading “Capital sp.” present the data for the second procedure (each capital’s boundary is updated using its specific mean of light intensity); columns under the heading “Lowest” present the data for first procedure (each city’s boundary has been updated using the lowest mean value of light intensity amongst all capitals).

Table 4.4 – List of countries in the full and in restricted samples.

Full Sample	HP > 30%	HP > 40%	HP > 50%
Angola	Angola	Angola	Angola
Benin	Burundi	Burundi	Burundi
Burkina	Cameroon	Cameroon	Cameroon
Burundi	Congo	Congo	DRC
Cameroon	DRC	DRC	Ethiopia
Congo	Ethiopia	Ethiopia	Kenya
DRC	Gabon	Gabon	Malawi
Ethiopia	Ghana	Ghana	Mozambique
Gabon	Ivory	Kenya	Namibia
Ghana	Kenya	Malawi	Sudan
Ivory	Malawi	Mali	Uganda
Kenya	Mali	Mozambique	Zambia
Malawi	Mozambique	Namibia	
Mali	Namibia	Rwanda	
Mozambique	Nigeria	Sudan	
Namibia	Rwanda	Tanzania	
Nigeria	Sudan	Uganda	
RSA	Tanzania	Zambia	
Rwanda	Uganda		
Sierra	Zambia		
Sudan	Zimbabwe		
Tanzania			
Uganda			
Zambia			
Zimbabwe			

Countries in the full-sample and in the sample restricted to countries in which hydropower represents at least a given share; shares in the heading.

Table 4.5 – Level regression, city-level baseline specification, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Night Lights			
Mean Anomaly (Inst, 4 rad)	2.97			
	(2.21)			
Mean Anomaly (Op, 4 rad)		2.70		
		(1.73)		
Mean Anomaly (Inst, 2 rad)			2.13	
			(1.90)	
Mean Anomaly (Op, 2 rad)				2.53
				(2.40)
Constant	19.42***	19.41***	19.29***	19.26***
	(1.02)	(1.02)	(1.02)	(1.02)
Num of obs.	1079	1079	1092	1092
R^2	0.13	0.13	0.13	0.13

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4, constructed using 4 radiuses changing with country dimension in columns 1 and 2 and with 2 radiuses in columns 3 and 4. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.6 – First difference regression, city-level baseline specification, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Δ Night Lights			
Δ Mean Anomaly (Inst, 4 rad)	0.33			
	(0.29)			
Δ Mean Anomaly (Op, 4 rad)		0.33		
		(0.21)		
Δ Mean Anomaly (Inst, 2 rad)			0.04	
			(0.32)	
Δ Mean Anomaly (Op, 2 rad)				0.05
				(0.32)
Constant	1.03***	1.02***	1.04***	1.05***
	(0.17)	(0.17)	(0.17)	(0.17)
Num of obs.	996	996	1008	1008
R^2	0.67	0.67	0.67	0.67

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference in average night-late intensity, the explanatory variable is the first difference in the yearly mean anomaly, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4, constructed using 4 radiuses changing with country dimension in columns 1 and 2 and with 2 radiuses in columns 3 and 4. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.7 – Fixed effect panel regression, city-level baseline specification, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Night Lights			
Mean Anomaly (Inst, 4 rad)	0.72			
	(0.83)			
Mean Anomaly (Op, 4 rad)		0.34		
		(0.70)		
Mean Anomaly (Inst, 2 rad)			0.66	
			(0.92)	
Mean Anomaly (Op, 2 rad)				0.57
				(0.92)
Constant	19.50***	19.51***	19.34***	19.34***
	(0.41)	(0.41)	(0.41)	(0.41)
Num of obs.	1079	1079	1092	1092

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4, constructed using 4 radiuses changing with country dimension in columns 1 and 2 and with only two radiuses in columns 3 and 4. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.8 – Level regression, city-level baseline specification, strong/weak positive/negative anomalies.

	Night Lights			
	4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	-0.21	-0.30	-0.05	-0.34
	(0.82)	(0.81)	(0.81)	(0.86)
St. Pos. Anomaly Index	-0.16	-0.08	-0.02	-0.03
	(0.20)	(0.22)	(0.24)	(0.25)
Wk. Neg. Anomaly Index	1.01	1.14	2.33	1.96
	(1.85)	(1.93)	(2.30)	(2.34)
Wk. Pos. Anomaly Index	-2.85*	-2.49	-0.67	-0.61
	(1.63)	(1.68)	(2.08)	(2.21)
Constant	23.19***	22.48***	21.42***	20.77***
	(1.40)	(1.35)	(1.41)	(1.32)
Num of obs.	1079	1079	1092	1092
R^2	0.23	0.20	0.17	0.16

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variables are the weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4, constructed using 4 radiuses changing with country dimension in columns 1 and 2 and with only two radiuses in columns 3 and 4. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.9 – First difference regression, city-level baseline specification, strong/weak positive/negative anomalies.

	Δ Night Lights			
	4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative
Δ St. Neg. Anomaly Index	0.14	0.12	0.22*	0.20
	(0.13)	(0.13)	(0.13)	(0.13)
Δ St. Pos. Anomaly Index	-0.03	-0.03	-0.02	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)
Δ Wk. Neg. Anomaly Index	-0.18	-0.15	-0.19	-0.08
	(0.28)	(0.29)	(0.27)	(0.28)
Δ Wk. Pos. Anomaly Index	-0.13	-0.18	0.06	0.01
	(0.13)	(0.13)	(0.14)	(0.14)
Constant	1.06***	1.06***	1.05***	1.06***
	(0.17)	(0.18)	(0.17)	(0.18)
Num of obs.	996	996	1008	1008
R^2	0.68	0.68	0.67	0.67

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4, constructed using 4 radiuses changing with country dimension in columns 1 and 2 and with only two radiuses in columns 3 and 4. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.10 – Fixed effect panel regression, city-level baseline specification, strong/weak positive/negative anomalies.

	Night Lights			
	4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	0.26	0.38	0.07	0.12
	(0.42)	(0.41)	(0.44)	(0.44)
St. Pos. Anomaly Index	0.02	0.03	-0.00	-0.01
	(0.07)	(0.07)	(0.07)	(0.07)
Wk. Neg. Anomaly Index	0.61	0.42	0.91	1.18
	(0.79)	(0.81)	(0.83)	(0.86)
Wk. Pos. Anomaly Index	-0.49	-0.22	-0.50	-0.24
	(0.43)	(0.44)	(0.44)	(0.46)
Constant	20.68***	20.36***	20.45***	20.42***
	(0.62)	(0.60)	(0.60)	(0.58)
Num of obs.	1079	1079	1092	1092

Note for Table 4.10 Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable are the weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 and 3 and by operational capacity in columns 2 and 4, constructed using 4 radiuses changing with country dimension in columns 1 and 2 and with only two radiuses in columns 3 and 4. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.11 – Level regression, city-level sample expanded through the first urbanization procedure, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
Mean Anomaly (Inst, 4 rad)	2.38				1.58				0.60			
	(1.69)				(1.27)				(0.86)			
Mean Anomaly (Op, 4 rad)		2.16				1.18				0.18		
		(1.31)				(1.02)				(0.75)		
Mean Anomaly (Inst, 2 rad)			1.95				1.57				0.65	
			(1.50)				(1.15)				(0.83)	
Mean Anomaly (Op, 2 rad)				2.24				1.63				0.68
				(1.83)				(1.37)				(0.94)
Constant	19.91***	19.90***	19.76***	19.73***	18.61***	18.62***	18.47***	18.45***	14.41***	14.42***	14.30***	14.30***
	(1.02)	(1.02)	(1.02)	(1.02)	(0.89)	(0.89)	(0.89)	(0.89)	(0.60)	(0.60)	(0.60)	(0.60)
Num of obs.	1079	1079	1092	1092	1079	1079	1092	1092	1066	1066	1079	1079
R^2	0.22	0.22	0.22	0.22	0.20	0.20	0.19	0.19	0.17	0.17	0.17	0.17

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.12 – First difference regression, city-level sample expanded through the first urbanization procedure, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Δ Night Lights, Mean				Δ Night Lights, Mean/2				Δ Night Lights, Mean/4			
Δ Mean Anomaly (Inst, 4 rad)	-0.03				-0.19				-0.26			
	(0.37)				(0.31)				(0.25)			
Δ Mean Anomaly (Op, 4 rad)		-0.13				-0.25				-0.28		
		(0.26)				(0.22)				(0.18)		
Δ Mean Anomaly (Inst, 2 rad)			0.34				0.18				-0.07	
			(0.36)				(0.30)				(0.31)	
Δ Mean Anomaly (Op, 2 rad)				0.46				0.29				-0.02
				(0.38)				(0.31)				(0.31)
Constant	1.06***	1.06***	1.08***	1.09***	0.92***	0.91***	0.94***	0.95***	0.55**	0.55**	0.57**	0.57**
	(0.17)	(0.17)	(0.17)	(0.17)	(0.16)	(0.16)	(0.15)	(0.15)	(0.17)	(0.17)	(0.17)	(0.17)
Num of obs.	996	996	1008	1008	996	996	1008	1008	984	984	996	996
R^2	0.72	0.72	0.71	0.71	0.73	0.73	0.73	0.73	0.71	0.71	0.71	0.71

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is first difference of average night-late intensity, the explanatory variable is the first difference of the yearly mean anomaly, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.13 – Fixed effect panel regression, city-level sample expanded through the first urbanization procedure, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
Mean Anomaly (Inst, 4 rad)	0.43				0.12				-0.30			
	(0.68)				(0.52)				(0.38)			
Mean Anomaly (Op, 4 rad)		0.08				-0.27				-0.49		
		(0.57)				(0.44)				(0.32)		
Mean Anomaly (Inst, 2 rad)			0.65				0.50				-0.18	
			(0.76)				(0.58)				(0.42)	
Mean Anomaly (Op, 2 rad)				0.54				0.25				-0.35
				(0.76)				(0.58)				(0.42)
Constant	19.97***	19.98***	19.80***	19.80***	18.66***	18.68***	18.50***	18.51***	14.44***	14.45***	14.33***	14.34***
	(0.34)	(0.34)	(0.34)	(0.34)	(0.26)	(0.26)	(0.26)	(0.26)	(0.19)	(0.19)	(0.19)	(0.19)
Num of obs.	1079	1079	1092	1092	1079	1079	1092	1092	1066	1066	1079	1079

Panel estimation with city fixed effect. The dependent variables is the average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.14 – Level regression, city-level sample expanded through the first urbanization procedure, strong/weak positive/negative anomalies.

	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	-0.34	-0.44	-0.15	-0.46	-0.26	-0.36	-0.06	-0.33	0.06	0.00	0.20	0.07
	(0.86)	(0.86)	(0.85)	(0.90)	(0.74)	(0.74)	(0.74)	(0.78)	(0.49)	(0.48)	(0.48)	(0.49)
St. Pos. Anomaly Index	-0.17	-0.10	-0.03	-0.04	-0.15	-0.10	-0.02	-0.04	-0.11	-0.09	-0.02	-0.02
	(0.21)	(0.22)	(0.24)	(0.26)	(0.18)	(0.19)	(0.21)	(0.22)	(0.13)	(0.13)	(0.14)	(0.15)
Wk. Neg. Anomaly Index	0.52	0.70	1.85	1.54	0.23	0.31	1.44	1.18	0.17	0.14	0.87	0.73
	(1.91)	(2.01)	(2.39)	(2.46)	(1.68)	(1.76)	(2.12)	(2.18)	(1.18)	(1.21)	(1.41)	(1.43)
Wk. Pos. Anomaly Index	-3.11*	-2.72	-0.90	-0.81	-2.72*	-2.43	-0.71	-0.62	-1.20	-1.11	-0.07	-0.04
	(1.70)	(1.76)	(2.19)	(2.33)	(1.48)	(1.53)	(1.93)	(2.07)	(0.99)	(1.01)	(1.18)	(1.27)
Constant	23.40***	22.69***	21.68***	21.04***	21.53***	20.95***	20.05***	19.50***	16.14***	15.86***	15.22***	14.94***
	(1.35)	(1.31)	(1.37)	(1.29)	(1.12)	(1.10)	(1.17)	(1.10)	(0.75)	(0.75)	(0.78)	(0.74)
Num of obs.	1079	1079	1092	1092	1079	1079	1092	1092	1066	1066	1079	1079
R^2	0.31	0.29	0.25	0.24	0.29	0.27	0.23	0.22	0.24	0.23	0.20	0.19

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variables are the weak/strong positive/negative anomalies, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.15 – First difference regression, city-level sample expanded through the first urbanization procedure, strong/weak positive/negative anomalies.

	Δ Night Lights, Mean				Δ Night Lights, Mean/2				Δ Night Lights, Mean/4			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
Δ St. Neg. Anomaly Index	0.14	0.11	0.25*	0.21	0.12	0.10	0.18	0.14	0.13	0.12	0.16*	0.16*
	(0.14)	(0.14)	(0.15)	(0.15)	(0.12)	(0.12)	(0.12)	(0.12)	(0.09)	(0.09)	(0.09)	(0.09)
Δ St. Pos. Anomaly Index	-0.01	-0.02	0.00	0.01	-0.02	-0.03	-0.01	-0.01	-0.02	-0.03	-0.02	-0.02
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Δ Wk. Neg. Anomaly Index	-0.22	-0.19	-0.26	-0.14	-0.17	-0.15	-0.16	-0.05	-0.33*	-0.33	-0.25	-0.21
	(0.29)	(0.30)	(0.30)	(0.32)	(0.24)	(0.25)	(0.25)	(0.26)	(0.19)	(0.20)	(0.20)	(0.21)
Δ Wk. Pos. Anomaly Index	-0.20	-0.26*	-0.01	-0.08	-0.16	-0.20	-0.03	-0.08	-0.12	-0.13	-0.03	-0.03
	(0.14)	(0.14)	(0.14)	(0.15)	(0.12)	(0.12)	(0.12)	(0.12)	(0.10)	(0.10)	(0.11)	(0.11)
Constant	1.08***	1.09***	1.07***	1.09***	0.94***	0.95***	0.94***	0.95***	0.56**	0.56**	0.56**	0.56**
	(0.18)	(0.18)	(0.18)	(0.18)	(0.16)	(0.16)	(0.16)	(0.16)	(0.18)	(0.18)	(0.18)	(0.18)
Num of obs.	996	996	1008	1008	996	996	1008	1008	984	984	996	996
R^2	0.72	0.72	0.71	0.71	0.73	0.73	0.73	0.73	0.71	0.71	0.71	0.71

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.16 – Fixed effect panel regression, growth form in the city-level sample expanded through the first urbanization procedure, strong/weak positive/negative anomalies.

	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	0.07	0.09	0.01	-0.05	0.14	0.16	0.11	0.07	0.14	0.14	-0.03	-0.05
	(0.34)	(0.34)	(0.36)	(0.36)	(0.26)	(0.26)	(0.28)	(0.28)	(0.19)	(0.19)	(0.20)	(0.20)
St. Pos. Anomaly Index	0.02	0.03	0.00	-0.00	0.00	-0.00	-0.00	-0.02	0.01	0.00	-0.00	-0.01
	(0.05)	(0.05)	(0.06)	(0.06)	(0.04)	(0.04)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Wk. Neg. Anomaly Index	0.47	0.33	0.90	1.18*	-0.21	-0.36	0.28	0.35	-0.61	-0.70	-0.24	-0.22
	(0.65)	(0.66)	(0.68)	(0.71)	(0.50)	(0.51)	(0.52)	(0.54)	(0.36)	(0.37)	(0.38)	(0.39)
Wk. Pos. Anomaly Index	-0.49	-0.30	-0.47	-0.30	-0.43	-0.32	-0.39	-0.33	-0.14	-0.08	-0.19	-0.14
	(0.35)	(0.36)	(0.36)	(0.38)	(0.27)	(0.28)	(0.28)	(0.29)	(0.19)	(0.20)	(0.20)	(0.21)
Constant	20.80***	20.49***	20.78***	20.73***	19.09***	18.89***	19.15***	19.10***	14.27***	14.15***	14.31***	14.29***
	(0.51)	(0.49)	(0.49)	(0.48)	(0.39)	(0.38)	(0.38)	(0.37)	(0.28)	(0.28)	(0.27)	(0.27)
Num of obs.	1079	1079	1092	1092	1079	1079	1092	1092	1066	1066	1079	1079

Panel estimation with city fixed effect. The dependent variables is the average night-late intensity, the explanatory variable are the weak/strong positive/negative anomalies, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.17 – Level regression, capital sample expanded through the second urbanization procedure, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
Mean Anomaly (Inst, 4 rad)	4.85				4.37				3.13			
	(4.62)				(4.07)				(3.33)			
Mean Anomaly (Op, 4 rad)		1.94				1.73				1.20		
		(3.04)				(2.67)				(2.12)		
Mean Anomaly (Inst, 2 rad)			5.10				4.58				3.19	
			(4.78)				(4.20)				(3.47)	
Mean Anomaly (Op, 2 rad)				6.76				5.88				4.31
				(5.03)				(4.46)				(3.68)
Constant	26.16***	26.15***	26.17***	26.16***	26.08***	26.08***	26.09***	26.08***	24.29***	24.29***	24.30***	24.29***
	(1.30)	(1.30)	(1.30)	(1.30)	(1.30)	(1.30)	(1.30)	(1.30)	(1.40)	(1.39)	(1.40)	(1.40)
Num of obs.	325	325	325	325	325	325	325	325	325	325	325	325
R^2	0.32	0.32	0.32	0.33	0.31	0.30	0.31	0.31	0.17	0.16	0.16	0.17

Pooled OLS regression with s.e clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All capitals' boundaries were expanded using their specific yearly average night light intensity in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.18 – First difference regression, capital sample expanded through the second urbanization procedure, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Δ Night Lights, Mean				Δ Night Lights, Mean/2				Δ Night Lights, Mean/4			
Δ Mean Anomaly (Inst, 4 rad)	-0.00				-0.05				-0.32			
	(0.48)				(0.43)				(0.39)			
Δ Mean Anomaly (Op, 4 rad)		-0.11				-0.14				-0.22		
		(0.25)				(0.22)				(0.20)		
Δ Mean Anomaly (Inst, 2 rad)			0.12				0.13				-0.32	
			(0.66)				(0.59)				(0.50)	
Δ Mean Anomaly (Op, 2 rad)				0.25				0.22				-0.22
				(0.61)				(0.54)				(0.47)
Constant	1.55***	1.54***	1.55***	1.56***	1.44***	1.44***	1.45***	1.45***	0.45	0.45	0.45	0.45
	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)	(0.25)	(0.26)	(0.25)	(0.25)	(0.25)	(0.25)
Num of obs.	300	300	300	300	300	300	300	300	300	300	300	300
R^2	0.76	0.76	0.76	0.76	0.77	0.77	0.77	0.77	0.74	0.74	0.74	0.74

Pooled OLS regression with s.e clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variable is the first difference of the yearly mean anomaly, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All capitals' boundaries were expanded using their specific yearly average night light intensity in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.19 – Fixed effect panel regression, level form in the capital sample expanded through the second urbanization procedure, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
Mean Anomaly (Inst, 4 rad)	1.25				1.02				0.19			
	(1.29)				(1.10)				(0.93)			
Mean Anomaly (Op, 4 rad)		0.07				-0.04				-0.22		
		(0.85)				(0.72)				(0.61)		
Mean Anomaly (Inst, 2 rad)			1.82				1.56				0.40	
			(1.54)				(1.32)				(1.11)	
Mean Anomaly (Op, 2 rad)				2.60				2.05				0.76
				(1.48)				(1.27)				(1.07)
Constant	26.18***	26.18***	26.18***	26.17***	26.10***	26.11***	26.10***	26.10***	24.31***	24.32***	24.31***	24.31***
	(0.63)	(0.64)	(0.63)	(0.63)	(0.54)	(0.54)	(0.54)	(0.54)	(0.46)	(0.46)	(0.46)	(0.46)
Num of obs.	325	325	325	325	325	325	325	325	325	325	325	325

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All capitals' boundaries were expanded using their specific yearly average night light intensity in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.20 – Level regression, capital sample expanded through the second urbanization procedure, strong/weak positive/negative anomalies.

	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	-0.94	-0.93	0.10	-0.04	-0.91	-0.91	0.17	0.03	-1.04	-1.03	0.40	0.23
	(0.59)	(0.67)	(0.65)	(0.76)	(0.56)	(0.63)	(0.65)	(0.74)	(0.71)	(0.74)	(0.81)	(0.86)
St. Pos. Anomaly Index	0.12	0.23	0.15	0.26	0.16	0.25	0.18	0.27	-0.03	0.07	0.00	0.09
	(0.26)	(0.27)	(0.30)	(0.33)	(0.26)	(0.27)	(0.30)	(0.34)	(0.29)	(0.29)	(0.34)	(0.38)
Wk. Neg. Anomaly Index	0.06	0.73	2.09	2.20	-0.21	0.41	1.88	1.94	-1.09	-0.53	1.87	1.79
	(1.66)	(1.59)	(1.99)	(1.90)	(1.63)	(1.54)	(2.04)	(1.92)	(1.75)	(1.62)	(2.65)	(2.49)
Wk. Pos. Anomaly Index	-2.34	-1.43	0.46	0.51	-2.44	-1.55	0.42	0.47	-2.96	-2.09	1.02	0.99
	(1.94)	(1.82)	(2.39)	(2.20)	(2.00)	(1.86)	(2.53)	(2.32)	(2.13)	(1.95)	(3.21)	(2.97)
Constant	26.63***	25.90***	26.85***	26.37***	26.38***	25.65***	26.69***	26.22***	24.83***	24.17***	25.13***	24.63***
	(1.87)	(1.82)	(1.80)	(1.67)	(1.78)	(1.72)	(1.74)	(1.61)	(1.81)	(1.77)	(1.75)	(1.63)
Num of obs.	325	325	325	325	325	325	325	325	325	325	325	325
R^2	0.33	0.33	0.33	0.33	0.31	0.32	0.31	0.31	0.18	0.18	0.18	0.17

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variables are the weak/strong positive/negative anomalies, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All capitals' boundaries were expanded using their specific yearly average night light intensity in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.21 – First difference regression, capital sample expanded through the second urbanization procedure, strong/weak positive/negative anomalies.

	Δ Night Lights, Mean				Δ Night Lights, Mean/2				Δ Night Lights, Mean/4			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
Δ St. Neg. Anomaly Index	-0.05	-0.07	0.06	0.12	-0.10	-0.15	0.04	0.06	-0.16	-0.20	0.05	0.07
	(0.33)	(0.35)	(0.30)	(0.32)	(0.29)	(0.31)	(0.26)	(0.27)	(0.27)	(0.28)	(0.27)	(0.27)
Δ St. Pos. Anomaly Index	-0.01	-0.01	-0.03	-0.02	-0.01	-0.00	-0.02	-0.01	-0.04	-0.03	-0.04	-0.04
	(0.04)	(0.04)	(0.05)	(0.05)	(0.04)	(0.03)	(0.04)	(0.04)	(0.03)	(0.03)	(0.03)	(0.03)
Δ Wk. Neg. Anomaly Index	0.06	0.10	-0.55	-0.43	0.17	0.25	-0.39	-0.24	0.21	0.28	-0.12	-0.00
	(0.62)	(0.64)	(0.54)	(0.56)	(0.55)	(0.56)	(0.50)	(0.50)	(0.52)	(0.54)	(0.46)	(0.47)
Δ Wk. Pos. Anomaly Index	-0.58**	-0.49*	-0.57**	-0.49*	-0.53*	-0.45	-0.54**	-0.46*	-0.44*	-0.37*	-0.38*	-0.33
	(0.28)	(0.28)	(0.27)	(0.27)	(0.28)	(0.27)	(0.26)	(0.27)	(0.22)	(0.21)	(0.23)	(0.22)
Constant	1.50***	1.52***	1.50***	1.52***	1.41***	1.42***	1.41***	1.42***	0.42	0.43	0.44	0.45
	(0.24)	(0.25)	(0.25)	(0.26)	(0.25)	(0.25)	(0.25)	(0.26)	(0.25)	(0.25)	(0.25)	(0.25)
Num of obs.	300	300	300	300	300	300	300	300	300	300	300	300
R^2	0.76	0.76	0.76	0.76	0.78	0.78	0.78	0.78	0.75	0.74	0.74	0.74

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All capitals' boundaries were expanded using their specific yearly average night light intensity in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.22 – Fixed effect panel regression, level form in the capital sample expanded through the second urbanization procedure, strong/weak positive/negative anomalies.

	Night Lights, Mean				Night Lights, Mean/2				Night Lights, Mean/4			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	0.31	0.21	0.75	0.97	0.14	0.02	0.52	0.63	-0.12	-0.17	0.21	0.34
	(0.74)	(0.72)	(0.83)	(0.80)	(0.63)	(0.62)	(0.71)	(0.69)	(0.53)	(0.52)	(0.60)	(0.58)
St. Pos. Anomaly Index	0.07	0.13	0.06	0.10	0.09	0.13	0.06	0.09	0.04	0.07	0.02	0.04
	(0.12)	(0.11)	(0.12)	(0.12)	(0.10)	(0.10)	(0.10)	(0.10)	(0.09)	(0.08)	(0.09)	(0.09)
Wk. Neg. Anomaly Index	1.00	1.06	0.79	1.31	0.95	1.04	0.81	1.30	0.45	0.48	0.29	0.58
	(1.35)	(1.37)	(1.36)	(1.34)	(1.16)	(1.17)	(1.16)	(1.15)	(0.98)	(0.99)	(0.98)	(0.97)
Wk. Pos. Anomaly Index	-0.95	-0.55	-0.16	-0.17	-0.91	-0.54	-0.25	-0.25	-0.73	-0.49	-0.23	-0.25
	(0.78)	(0.75)	(0.78)	(0.75)	(0.67)	(0.64)	(0.67)	(0.64)	(0.57)	(0.54)	(0.56)	(0.54)
Constant	28.01***	27.39***	27.71***	28.23***	27.57***	27.00***	27.38***	27.78***	25.00***	24.64***	24.95***	25.27***
	(0.96)	(0.94)	(1.02)	(0.98)	(0.82)	(0.80)	(0.87)	(0.84)	(0.70)	(0.68)	(0.74)	(0.71)
Num of obs.	325	325	325	325	325	325	325	325	325	325	325	325

Panel estimation with city fixed effect. The dependent variables is the average night-late intensity, the explanatory variable are the weak/strong positive/negative anomalies, weighted by installed capacity in odd columns and by operational capacity in even columns, constructed using 4 radiuses changing with country dimension in columns 1, 2, 5, 6, 9, 10 and with only two radiuses in columns 3, 4, 7, 8, 11, 12. All cities' boundaries were expanded using the lowest average night light intensity amongst all capitals in columns 1 to 4, half of that value in columns 5 to 8 and a quarter in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.23 – Level regression, city-level restricted sample, yearly mean anomaly.

	30%HP Night Lights				40% HP Night Lights				50% HP Night Lights			
Mean Anomaly (Inst, 4 rad)	2.93				3.81				4.79			
	(2.46)				(2.72)				(3.19)			
Mean Anomaly (Op, 4 rad)		3.31				4.11				5.13		
		(2.64)				(2.88)				(3.29)		
Mean Anomaly (Inst, 2 rad)			2.26				3.05				3.58	
			(1.98)				(2.29)				(2.54)	
Mean Anomaly (Op, 2 rad)				2.67				3.37				3.97
				(2.48)				(2.79)				(3.07)
Constant	18.39***	18.36***	18.41***	18.37***	18.04***	18.02***	18.06***	18.03***	16.95***	16.91***	17.02***	16.98***
	(1.01)	(1.01)	(1.01)	(1.01)	(1.20)	(1.20)	(1.20)	(1.20)	(1.31)	(1.31)	(1.31)	(1.30)
Num of obs.	975	975	975	975	780	780	780	780	559	559	559	559
R^2	0.14	0.14	0.14	0.14	0.13	0.14	0.13	0.13	0.14	0.14	0.14	0.14

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.24 – First difference regression, city-level restricted sample, yearly mean anomaly.

	30%HP Δ Night Lights				40% HP Δ Night Lights				50% HP Δ Night Lights			
Δ Mean Anomaly (Inst, 4 rad)	-0.21				-0.16				0.01			
	(0.31)				(0.34)				(0.36)			
Δ Mean Anomaly (Op, 4 rad)		-0.23				-0.22				0.02		
		(0.32)				(0.34)				(0.36)		
Δ Mean Anomaly (Inst, 2 rad)			0.04				0.07				0.05	
			(0.32)				(0.36)				(0.37)	
Δ Mean Anomaly (Op, 2 rad)				0.16				0.19				0.18
				(0.32)				(0.35)				(0.36)
Constant	1.22***	1.21***	1.23***	1.24***	1.22***	1.22***	1.24***	1.25***	1.11**	1.11**	1.12**	1.13**
	(0.26)	(0.26)	(0.26)	(0.26)	(0.31)	(0.31)	(0.31)	(0.31)	(0.36)	(0.36)	(0.36)	(0.36)
Num of obs.	912	912	912	912	732	732	732	732	528	528	528	528
R^2	0.63	0.63	0.63	0.63	0.61	0.61	0.61	0.61	0.58	0.58	0.58	0.58

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variable is the first difference of the yearly mean anomaly, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.25 – Fixed effect panel regression, level form in the city-level restricted sample, yearly mean anomaly.

	30%HP Night Lights				40% HP Night Lights				50% HP Night Lights			
Mean Anomaly (Inst, 4 rad)	0.94				1.53				1.96*			
	(0.88)				(1.00)				(1.17)			
Mean Anomaly (Op, 4 rad)		1.05				1.54				2.00*		
		(0.88)				(0.99)				(1.15)		
Mean Anomaly (Inst, 2 rad)			0.68				1.18				1.46	
			(0.96)				(1.10)				(1.24)	
Mean Anomaly (Op, 2 rad)				0.58				0.97				1.24
				(0.96)				(1.08)				(1.22)
Constant	18.46***	18.45***	18.46***	18.46***	18.13***	18.13***	18.14***	18.15***	17.10***	17.09***	17.13***	17.14***
	(0.45)	(0.45)	(0.45)	(0.45)	(0.52)	(0.53)	(0.53)	(0.53)	(0.66)	(0.67)	(0.67)	(0.67)
Num of obs.	975	975	975	975	780	780	780	780	559	559	559	559

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.26 – Level regression, city-level restricted sample, strong/weak positive/negative anomalies.

	Night Lights, 30% HP				Night Lights, 40% HP				Night Lights, 50% HP			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	-0.39	-0.46	-0.20	-0.48	-0.25	-0.19	-0.07	-0.27	-0.12	-0.01	0.68	0.50
	(0.83)	(0.82)	(0.82)	(0.86)	(0.88)	(0.87)	(0.88)	(0.92)	(0.95)	(0.95)	(0.92)	(0.97)
St. Pos. Anomaly Index	-0.14	-0.11	0.01	-0.01	-0.09	-0.08	0.14	0.12	-0.08	-0.10	-0.02	-0.06
	(0.21)	(0.21)	(0.24)	(0.25)	(0.22)	(0.23)	(0.26)	(0.26)	(0.25)	(0.24)	(0.25)	(0.25)
Wk. Neg. Anomaly Index	0.54	0.43	1.94	1.60	1.00	1.14	2.75	2.54	1.38	1.34	3.58	3.26
	(1.87)	(1.94)	(2.32)	(2.35)	(1.95)	(2.03)	(2.42)	(2.44)	(2.30)	(2.30)	(2.67)	(2.71)
Wk. Pos. Anomaly Index	-2.66	-2.43	-0.55	-0.43	-2.73	-2.34	-0.22	0.08	-2.63	-2.26	0.21	0.68
	(1.63)	(1.68)	(2.08)	(2.21)	(1.74)	(1.78)	(2.18)	(2.28)	(1.96)	(2.02)	(2.40)	(2.54)
Constant	21.59***	20.97***	20.13***	19.48***	22.23***	21.89***	20.06***	19.39***	22.37***	22.02***	21.72***	20.59***
	(1.46)	(1.36)	(1.43)	(1.30)	(1.82)	(1.73)	(1.81)	(1.69)	(2.30)	(2.20)	(2.24)	(2.13)
Num of obs.	975	975	975	975	780	780	780	780	559	559	559	559
R^2	0.20	0.18	0.16	0.15	0.21	0.20	0.16	0.15	0.22	0.21	0.22	0.19

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.27 – First difference regression, city-level restricted sample, strong/weak positive/negative anomalies.

	Δ Night Lights, 30% HP				Δ Night Lights, 40% HP				Δ Night Lights, 50% HP			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
Δ St. Neg. Anomaly Index (Inst)	0.16	0.15	0.23*	0.22*	0.12	0.10	0.18	0.17	0.16	0.13	0.21	0.21
	(0.13)	(0.13)	(0.13)	(0.13)	(0.14)	(0.14)	(0.15)	(0.14)	(0.15)	(0.15)	(0.15)	(0.16)
Δ St. Pos. Anomaly Index (Inst)	-0.03	-0.03	-0.01	-0.01	-0.02	-0.03	-0.01	-0.01	-0.02	-0.02	-0.01	-0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)
Δ Wk. Neg. Anomaly Index (Inst)	-0.16	-0.12	-0.17	-0.06	-0.01	0.02	-0.04	0.05	-0.04	0.02	-0.07	0.02
	(0.29)	(0.30)	(0.28)	(0.29)	(0.30)	(0.31)	(0.29)	(0.31)	(0.33)	(0.35)	(0.31)	(0.33)
Δ Wk. Pos. Anomaly Index (Inst)	-0.14	-0.19	0.06	0.01	-0.14	-0.21	0.04	-0.01	-0.09	-0.16	0.03	-0.02
	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.14)	(0.15)	(0.15)	(0.16)	(0.16)	(0.16)	(0.16)
Constant	1.25***	1.26***	1.24***	1.25***	1.25***	1.26***	1.24***	1.25***	1.11**	1.12**	1.11**	1.13**
	(0.26)	(0.26)	(0.27)	(0.27)	(0.31)	(0.32)	(0.32)	(0.32)	(0.37)	(0.37)	(0.37)	(0.37)
Num of obs.	912	912	912	912	732	732	732	732	528	528	528	528
R^2	0.63	0.64	0.63	0.63	0.61	0.61	0.61	0.61	0.58	0.58	0.58	0.58

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.28 – Fixed effect panel regression, city-level restricted sample, strong/weak positive/negative anomalies.

	Night Lights, 30% HP				Night Lights, 40% HP				Night Lights, 50% HP			
	4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses		4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index (Inst)	0.26	0.42	0.06	0.12	0.24	0.42	-0.01	0.06	0.26	0.44	0.12	0.20
	(0.43)	(0.43)	(0.45)	(0.46)	(0.47)	(0.47)	(0.50)	(0.50)	(0.54)	(0.54)	(0.55)	(0.55)
St. Pos. Anomaly Index (Inst)	0.02	0.02	-0.00	-0.01	0.07	0.07	0.04	0.02	0.06	0.06	0.04	0.02
	(0.07)	(0.07)	(0.07)	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)	(0.09)	(0.09)	(0.09)	(0.09)
Wk. Neg. Anomaly Index (Inst)	0.66	0.57	0.90	1.17	0.87	0.71	1.10	1.30	1.17	0.99	1.29	1.51
	(0.82)	(0.83)	(0.86)	(0.89)	(0.90)	(0.91)	(0.94)	(0.97)	(1.05)	(1.07)	(1.04)	(1.09)
Wk. Pos. Anomaly Index (Inst)	-0.53	-0.32	-0.50	-0.23	-0.40	-0.18	-0.42	-0.14	-0.53	-0.30	-0.37	-0.08
	(0.44)	(0.45)	(0.45)	(0.47)	(0.48)	(0.49)	(0.50)	(0.52)	(0.58)	(0.59)	(0.56)	(0.59)
Constant	19.86***	19.70***	19.68***	19.65***	19.51***	19.35***	19.36***	19.36***	19.31***	19.10***	19.01***	18.98***
	(0.69)	(0.68)	(0.67)	(0.65)	(0.86)	(0.86)	(0.84)	(0.83)	(1.14)	(1.14)	(1.13)	(1.11)
Num of obs.	975	975	975	975	780	780	780	780	559	559	559	559

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.29 – Fixed effect panel regression, level form in the capital sample expanded through the second urbanization procedure, capital specific mean, yearly mean anomaly.

	30%HP, Mean Night Lights				40% HP, Mean Night Lights				50% HP, Mean Night Lights			
Mean Anomaly (Inst, 4 rad)	2.40*				2.90*				3.67**			
	(1.43)				(1.52)				(1.79)			
Mean Anomaly (Op, 4 rad)		2.82**				3.20**				4.04**		
		(1.39)				(1.46)				(1.72)		
Mean Anomaly (Inst, 2 rad)			1.96				2.61				3.54*	
			(1.60)				(1.73)				(2.05)	
Mean Anomaly (Op, 2 rad)				2.75*				3.23**				4.26**
				(1.53)				(1.61)				(1.91)
Constant	26.36***	26.36***	26.36***	26.35***	26.92***	26.93***	26.91***	26.92***	25.93***	25.95***	25.96***	25.99***
	(0.71)	(0.71)	(0.71)	(0.71)	(0.76)	(0.76)	(0.76)	(0.76)	(1.00)	(0.99)	(1.00)	(0.99)
Number of obs.	273	273	273	273	234	234	234	234	156	156	156	156

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All capitals' boundaries were expanded using their specific yearly average night light intensity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.30 – Fixed effect panel regression, level form in the capital sample expanded through the second urbanization procedure, capital specific mean, yearly mean anomaly.

	30%HP, Mean/2 Night Lights				40% HP, Mean/2 Night Lights				50% HP, Mean/2 Night Lights			
Mean Anomaly (Inst, 4 rad)	2.03*				2.37*				2.89*			
	(1.23)				(1.30)				(1.52)			
Mean Anomaly (Op, 4 rad)		2.27*				2.60**				3.17**		
		(1.20)				(1.24)				(1.46)		
Mean Anomaly (Inst, 2 rad)			1.67				2.13				2.83	
			(1.37)				(1.47)				(1.74)	
Mean Anomaly (Op, 2 rad)				2.17				2.61*				3.39**
				(1.32)				(1.37)				(1.62)
Constant	26.30***	26.30***	26.30***	26.29***	26.85***	26.85***	26.84***	26.85***	25.87***	25.88***	25.89***	25.91***
	(0.61)	(0.61)	(0.61)	(0.61)	(0.65)	(0.65)	(0.65)	(0.65)	(0.85)	(0.84)	(0.85)	(0.85)
Number of obs.	273	273	273	273	234	234	234	234	156	156	156	156

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only cities of countries in which hydropower represents at least 30% of installed capacity are included in columns 1 to 4, the threshold is increased to 40% in columns 5 to 8 and to 50% in columns 9 to 12. All capitals' boundaries were expanded using half of their specific yearly average night light intensity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.31 –Level regression, city-level sample, city specific linear time-trend, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Night Lights			
Mean Anomaly (Inst, 4 rad)	-0.60			
	(1.56)			
Mean Anomaly (Op, 4 rad)		-0.19		
		(1.19)		
Mean Anomaly (Inst, 2 rad)			-1.26	
			(1.50)	
Mean Anomaly (Op, 2 rad)				-0.90
				(1.74)
Constant	36.49***	36.49***	36.49***	36.49***
	(0.52)	(0.52)	(0.52)	(0.53)
Num of obs.	1079	1079	1092	1092
R^2	0.77	0.77	0.77	0.77

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include city specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.32 – First difference regression, city-level sample, city specific linear time-trend, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Δ Night Lights			
Δ Mean Anomaly (Inst, 4 rad)	1.63			
	(0.98)			
Δ Mean Anomaly (Op, 4 rad)		0.69		
		(0.69)		
Δ Mean Anomaly (Inst, 2 rad)			2.08	
			(1.47)	
Δ Mean Anomaly (Op, 2 rad)				1.82
				(1.37)
Constant	4.49***	4.49***	4.49***	4.49***
	(0.00)	(0.00)	(0.00)	(0.00)
Num of obs.	996	996	1008	1008
R^2	0.06	0.06	0.07	0.06

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variable is the first difference of the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include city specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.33 – Fixed effect panel regression, level form in the city-level sample, city specific linear time-trend, yearly mean anomaly.

	Night Lights			
Mean Anomaly (Inst, 4 rad)	0.49			
	(0.87)			
Mean Anomaly (Op, 4 rad)		0.23		
		(0.74)		
Mean Anomaly (Inst, 2 rad)			0.47	
			(0.96)	
Mean Anomaly (Op, 2 rad)				0.23
				(0.97)
Constant	36.99***	36.99***	36.99***	36.99***
	(1.08)	(1.08)	(1.07)	(1.07)
Num of obs.	1079	1079	1092	1092

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variable is the average growth in the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include city specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.34 – Level regression, city-level sample, city specific linear time-trend, strong/weak positive/negative anomalies.

	Night Lights			
	4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	-0.21	-0.09	-0.44	-0.45
	(0.62)	(0.60)	(0.72)	(0.72)
St. Pos. Anomaly Index	-0.17**	-0.12	-0.19**	-0.15**
	(0.08)	(0.08)	(0.07)	(0.07)
Wk. Neg. Anomaly Index	2.57**	2.50	2.80**	3.44**
	(1.24)	(1.28)	(1.29)	(1.49)
Wk. Pos. Anomaly Index	-1.64***	-1.51***	-1.58***	-1.44***
	(0.50)	(0.45)	(0.58)	(0.51)
Constant	39.12***	38.04***	39.10***	37.65***
	(0.98)	(0.65)	(1.10)	(0.63)
Num of obs.	1079	1079	1092	1092
R^2	0.77	0.77	0.78	0.78

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the average night-late intensity, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include city specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.35 – First difference regression, city-level sample, city specific linear time-trend, strong/weak positive/negative anomalies.

	Δ Night Lights			
	4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative
Δ St. Neg. Anomaly Index	-0.13	-0.03	-0.13	-0.03
	(0.29)	(0.30)	(0.28)	(0.32)
Δ St. Pos. Anomaly Index	0.13***	0.12***	0.12***	0.12***
	(0.04)	(0.04)	(0.04)	(0.04)
Δ Wk. Neg. Anomaly Index	0.11	-0.21	0.10	-0.22
	(0.64)	(0.66)	(0.70)	(0.78)
Δ Wk. Pos. Anomaly Index	-0.24	-0.18	-0.24	-0.24
	(0.26)	(0.28)	(0.19)	(0.21)
Constant	4.46***	4.48***	4.46***	4.47***
	(0.01)	(0.01)	(0.01)	(0.01)
Num of obs.	996	996	1008	1008
R^2	0.07	0.06	0.06	0.06

Pooled OLS regression with s.e. clustered at the city level. The dependent variables is the first difference of average night-late intensity, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include city specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.36 – Fixed effect panel regression, level form in the city-level sample, city specific linear time-trend, strong/weak positive/negative anomalies.

	Night Lights			
	4 Radiuses		2 Radiuses	
	Installed	Operative	Installed	Operative
St. Neg. Anomaly Index	0.09	0.22	0.20	0.27
	(0.45)	(0.45)	(0.47)	(0.48)
St. Pos. Anomaly Index	0.01	0.01	-0.02	-0.02
	(0.07)	(0.07)	(0.07)	(0.07)
Wk. Neg. Anomaly Index	1.07	0.74	1.15	0.82
	(0.85)	(0.86)	(0.87)	(0.91)
Wk. Pos. Anomaly Index	-0.62	-0.45	-0.53	-0.35
	(0.46)	(0.47)	(0.46)	(0.49)
Constant	38.09***	37.54***	38.39***	37.52***
	(1.16)	(1.10)	(1.20)	(1.10)
Num of obs.	1079	1079	1092	1092

Panel estimation with city fixed effect. The dependent variables is the yearly average night-late intensity, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include city specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.37 – Quantile regression, growth form in the city-level sample, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Δ Night Lights				Δ Night Lights				Δ Night Lights				Δ Night Lights			
Δ Mean Anomaly (Inst, 4 rad)	0.0001054***	0.0000728***	0.0000499**	4.62e-06***												
	(0.00)	(0.00)	(0.00)	(0.00)												
Δ Mean Anomaly (Op, 4 rad)					0.0003098***	0.0003205***	0.0000476	0.0000967***								
					(0.00)	(0.00)	(0.00)	(0.00)								
Δ Mean Anomaly (Inst, 2 rad)									0.000088***	0.0000747***	0.0000748***	0.0000185***				
									(0.00)	(0.00)	(0.00)	(0.00)				
Δ Mean Anomaly (Op, 2 rad)													0.000088***	0.0000747***	0.0000748***	0.0000185***
													(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.04***	0.06***	0.23***	0.49***	0.01***	0.04***	0.21***	0.48***	0.01***	0.04***	0.21***	0.47***	0.01***	0.04***	0.21***	0.48***
	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)
Num of obs.	996	996	996	996	996	996	996	996	1008	1008	1008	1008	1008	1008	1008	1008

Quantile regression. The dependent variables is the yearly growth of average night-late intensity, the explanatory variable is the average growth in the yearly mean anomaly, weighted by installed capacity or by operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.38 – Quantile regression, level form in the city-level sample, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Night Lights				Night Lights				Night Lights				Night Lights			
Mean Anomaly (Inst, 4 rad)	0.05***	0.30***	0.51***	0.59***												
	(0.00)	(0.07)	(0.15)	(0.00)												
Mean Anomaly (Op, 4 rad)					-0.36***	-0.04	0.23*	0.57***								
					(0.00)	(0.09)	(0.12)	(0.00)								
Mean Anomaly (Inst, 2 rad)									0.19***	0.10	0.04	0.95***				
									(0.00)	(0.17)	(0.12)	(0.00)				
Mean Anomaly (Op, 2 rad)													0.12***	0.06	0.04	1.01***
													(0.00)	(0.19)	(0.11)	(0.00)
Constant	11.34***	35.65***	51.63***	53.15***	11.32***	35.65***	51.66***	53.17***	11.34***	35.83***	51.71***	53.20***	11.34***	35.84***	51.71***	53.29***
	(0.00)	(0.09)	(0.19)	(0.00)	(0.00)	(0.14)	(0.19)	(0.00)	(0.00)	(0.21)	(0.15)	(0.00)	(0.00)	(0.23)	(0.14)	(0.00)
Num of obs.	1079	1079	1079	1079	1079	1079	1079	1079	1092	1092	1092	1092	1092	1092	1092	1092

Quantile regression. The dependent variables is the yearly growth of average night-late intensity, the explanatory variable is the average growth in the yearly mean anomaly, weighted by installed capacity or by operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.39 – Quantile regression, growth form in the city-level sample, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	4 Radiuses, Installed, Δ Night Lights				4 Radiuses, Operative, Δ Night Lights				2 Radiuses, Installed, Δ Night Lights				2 Radiuses, Operative, Δ Night Lights			
Δ St. Neg. Anomaly Index	0.0009765***	-0.01***	0.04***	0.06***	-0.03***	-0.03***	-0.06***	-0.0026578***	0.0009676***	-0.01***	0.07***	0.13***	-0.03***	-0.02***	-0.03***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Δ St. Pos. Anomaly Index	0.0001458***	0.0001565***	0.0000927***	0.0000267***	0.0001795***	0.0001648***	0.0001051***	00000137***	0.0001445***	0.0001715***	0.0001097***	0.0000128***	0.0001892***	0.0001857***	0.0001114***	-0.0000111***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Δ Wk. Neg. Anomaly Index	0.01***	0.03***	0.01***	0.01***	0.02***	0.03***	0.01***	0.02***	0.05***	0.05***	0.04***	0.01***	0.04***	0.04***	0.03***	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Δ Wk. Pos. Anomaly Index	0.0000306***	-0.0032327*	-0.0033366**	-0.0010554***	-0.0041237***	-0.0036167***	-0.0049292***	0.0022142***	0.0042515***	-0.01***	0.0007604	0.01***	-0.0031516***	-0.01***	-0.0012441	0.01***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.00***	0.01	0.18***	0.44***	-0.00***	0.00	0.18***	0.43***	0.01***	0.01***	0.18***	0.44***	0.14***	0.17***	0.35***	0.62***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num of obs.	996	996	996	996	996	996	996	996	1008	1008	1008	1008	1008	1008	1008	1008

Quantile regression. The dependent variables is the yearly growth of average night-late intensity, the explanatory variables are the growth rate of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.40 – Quantile regression, level form in the city-level sample, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	4 Radiuses, Installed, Night Lights				4 Radiuses, Operative, Night Lights				2 Radiuses, Installed, Night Lights				2 Radiuses, Operative, Night Lights			
St. Neg. Anomaly Index	0.29***	0.23***	0.28***	0.14***	0.16***	0.26***	0.51***	0.33***	0.16***	0.25***	0.13***	0.40***	0.05***	0.17***	0.60***	0.43***
	(0.00)	(0.04)	(0.03)	(0.00)	(0.00)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)	(0.01)	(0.00)	(0.00)	(0.02)	(0.04)	(0.00)
St. Pos. Anomaly Index	0.02***	0.02**	0.03***	).0017467*	0.02***	0.02***	0.01*	-0.01***	0.01***	-0.01***	-0.00	-0.01***	0.01***	-0.01***	-0.01	-0.02***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
Wk. Neg. Anomaly Index	0.51***	0.60***	0.14***	-0.23***	0.75***	0.68***	-0.16***	-0.34***	0.90***	0.43***	0.44***	-0.10***	1.32***	0.91***	0.73***	0.04***
	(0.00)	(0.08)	(0.05)	(0.00)	(0.00)	(0.03)	(0.03)	(0.00)	(0.00)	(0.03)	(0.03)	(0.00)	(0.00)	(0.04)	(0.07)	(0.00)
Wk. Pos. Anomaly Index	0.11***	-0.25***	-0.18***	-0.28***	0.21***	-0.03*	-0.16***	-0.09***	0.04***	-0.32***	-0.15***	-0.02***	0.08***	-0.03	-0.18***	0.08***
	(0.00)	(0.05)	(0.03)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)	(0.00)	(0.02)	(0.01)	(0.00)	(0.00)	(0.02)	(0.04)	(0.00)
Constant	11.88***	36.41***	51.88***	53.47***	11.54***	36.20***	51.78***	53.22***	12.03***	36.46***	52.13***	53.68***	11.50***	36.05***	52.30***	53.43***
	(0.00)	(0.12)	(0.07)	(0.00)	(0.00)	(0.04)	(0.05)	(0.00)	(0.00)	(0.04)	(0.04)	(0.00)	(0.00)	(0.05)	(0.09)	(0.00)
Num of obs.	1079	1079	1079	1079	1079	1079	1079	1079	1092	1092	1092	1092	1092	1092	1092	1092

Quantile regression. The dependent variables is the yearly average night-late intensity, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.41 – Level regression, country-level baseline specification, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Night Lights		Night Lights (0 censored)	
Mean Anomaly (Inst)	-0.05		0.88	
	(0.03)		(0.61)	
Mean Anomaly (Op)		-0.04		0.82
		(0.03)		(0.53)
Constant	0.24**	0.24**	10.05***	10.05***
	(0.07)	(0.07)	(0.36)	(0.36)
Num. of obs.	377	377	377	377
R^2	0.03	0.03	0.41	0.41

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.42 – First difference regression, country-level baseline specification, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Δ Night Lights		Δ Night Lights (0 censored)	
Δ Mean Anomaly (Inst)	-0.02		-0.35	
	(0.02)		(0.31)	
Δ Mean Anomaly (Op)		-0.01		-0.22
		(0.01)		(0.27)
Constant	0.01	0.01*	-0.08	-0.07
	(0.00)	(0.00)	(0.17)	(0.17)
Num. of obs.	348	348	348	348
R^2	0.38	0.38	0.68	0.68

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variable is the first difference of the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.43 – Fixed effect panel regression, level form of country-level baseline specification, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Night Lights		Night Lights (0 censored)	
Mean Anomaly (Inst)	-0.04		0.63	
	(0.03)		(0.42)	
Mean Anomaly (Op)		-0.03		0.58
		(0.03)		(0.37)
Constant	0.24***	0.24***	10.05***	10.05***
	(0.01)	(0.01)	(0.17)	(0.17)
Num. of obs.	377	377	377	377

Panel estimation with country fixed effect. The dependent variables is the yearly average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.44 – Level regression, country-level baseline specification, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)
	Night Lights		Night Lights (0 censored)	
St. Neg. Anomaly Index (Inst)	0.04		-0.29	
	(0.04)		(0.28)	
St. Pos. Anomaly Index (Inst)	-0.00		-0.09	
	(0.01)		(0.06)	
Wk. Neg. Anomaly Index (Inst)	0.07		-1.02**	
	(0.08)		(0.49)	
Wk. Pos. Anomaly Index (Inst)	-0.07		-0.09	
	(0.06)		(0.46)	
St. Neg. Anomaly Index (Op)		0.04		-0.19
		(0.03)		(0.28)
St. Pos. Anomaly Index (Op)		0.00		-0.08
		(0.01)		(0.06)
Wk. Neg. Anomaly Index (Op)		0.07		-0.88*
		(0.07)		(0.48)
Wk. Pos. Anomaly Index (Op)		-0.06		-0.12
		(0.04)		(0.37)
Constant	0.42*	0.41*	9.33***	9.53***
	(0.16)	(0.16)	(0.57)	(0.62)
Num. of obs.	377	377	377	377
R^2	0.14	0.14	0.45	0.44

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.45 – First difference regression, country-level baseline specification, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)
	Δ Night Lights		Δ Night Lights (0 censored)	
Δ St. Neg. Anomaly Index (Inst)	0.00		-0.25	
	(0.01)		(0.21)	
Δ St. Pos. Anomaly Index (Inst)	-0.00		-0.04	
	(0.00)		(0.02)	
Δ Wk. Neg. Anomaly Index (Inst)	-0.00		0.05	
	(0.01)		(0.28)	
Δ Wk. Pos. Anomaly Index (Inst)	-0.01		-0.10	
	(0.01)		(0.16)	
Δ St. Neg. Anomaly Index (Op)		0.01		-0.19
		(0.01)		(0.21)
Δ St. Pos. Anomaly Index (Op)		-0.00		-0.04
		(0.00)		(0.02)
Δ Wk. Neg. Anomaly Index (Op)		-0.02		0.06
		(0.01)		(0.27)
Δ Wk. Pos. Anomaly Index (Op)		-0.01		-0.11
		(0.01)		(0.14)
Constant	0.01*	0.01	-0.10	-0.09
	(0.00)	(0.00)	(0.17)	(0.17)
Num. of obs.	348	348	348	348
R^2	0.38	0.38	0.69	0.69

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table 4.46 – Fixed effect panel regression, level form of baseline specification, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)
	Δ Night Lights		Δ Night Lights (0 censored)	
St. Neg. Anomaly Index (Inst)	-0.03		0.45*	
	(0.02)		(0.25)	
St. Pos. Anomaly Index (Inst)	-0.00		0.03	
	(0.00)		(0.03)	
Wk. Neg. Anomaly Index (Inst)	0.01		-0.08	
	(0.03)		(0.41)	
Wk. Pos. Anomaly Index (Inst)	-0.01		0.17	
	(0.02)		(0.22)	
St. Neg. Anomaly Index (Op)		-0.02		0.50**
		(0.02)		(0.23)
St. Pos. Anomaly Index (Op)		-0.00		0.03
		(0.00)		(0.03)
Wk. Neg. Anomaly Index (Op)		0.01		-0.06
		(0.03)		(0.39)
Wk. Pos. Anomaly Index (Op)		-0.02		0.13
		(0.02)		(0.20)
Constant	0.23***	0.23***	10.34***	10.47***
	(0.02)	(0.02)	(0.29)	(0.29)
Num. of obs.	377	377	377	377

Panel estimation with country fixed effect. The dependent variables is the yearly average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.1 – Level regression, country-level restricted sample, unlit cells included, yearly mean anomaly.

	Night Lights					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Mean Anomaly (Inst)	-0.06		-0.03		-0.01	
	(0.03)		(0.02)		(0.02)	
Mean Anomaly (Op)		-0.04		-0.03		-0.00
		(0.02)		(0.02)		(0.02)
Constant	0.21**	0.21**	0.14**	0.14**	0.10***	0.10***
	(0.06)	(0.06)	(0.04)	(0.04)	(0.02)	(0.02)
Num. of obs.	312	312	260	260	169	169
R^2	0.04	0.04	0.08	0.08	0.18	0.18

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity including unlit cells, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.2 – First difference regression, country-level restricted sample, unlit cells included, yearly mean anomaly.

	Δ Night Lights					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Δ Mean Anomaly (Inst)	-0.01		-0.00		-0.01	
	(0.01)		(0.01)		(0.01)	
Δ Mean Anomaly (Op)		0.00		-0.00		-0.01
		(0.01)		(0.01)		(0.01)
Constant	0.01	0.01	0.01*	0.01*	0.01**	0.01**
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num. of obs.	288	288	240	240	156	156
R^2	0.39	0.39	0.42	0.42	0.57	0.57

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of average night-late intensity including unlit cells, the explanatory variable is the first difference in the yearly mean anomaly, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.3 – Fixed effect panel regression, level form in the country-level restricted sample,unlit cells included, yearly mean anomaly.

	Night Lights					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Mean Anomaly (Inst)	-0.03		-0.01		0.00	
	(0.03)		(0.03)		(0.01)	
Mean Anomaly (Op)		-0.02		-0.01		0.01
		(0.03)		(0.02)		(0.01)
Constant	0.21***	0.21***	0.14***	0.14***	0.10***	0.10***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Num. of obs.	312	312	260	260	169	169

Panel estimation with country fixed effect. The dependent variables is the yearly average night late intensity including unlit cells, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.4 – Level regression, country-level restricted sample, unlit cells excluded, yearly mean anomaly.

	Night Lights (0 censored)					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Mean Anomaly (Inst)	0.96		0.84		0.88	
	(0.64)		(0.65)		(0.90)	
Mean Anomaly (Op)		0.89		0.82		0.95
		(0.58)		(0.55)		(0.81)
Constant	10.05***	10.05***	10.10***	10.09***	9.80***	9.80***
	(0.42)	(0.42)	(0.35)	(0.35)	(0.44)	(0.43)
Num. of obs.	312	312	260	260	169	169
R^2	0.38	0.38	0.45	0.45	0.48	0.48

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity excluding unlit cells, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.5 – First difference regression, country-level restricted sample, unlit cells excluded, yearly mean anomaly.

	Δ Night Lights (0 censored)					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Δ Mean Anomaly (Inst)	-0.23		-0.18		0.05	
	(0.31)		(0.32)		(0.34)	
Δ Mean Anomaly (Op)		-0.16		-0.05		0.17
		(0.27)		(0.27)		(0.29)
Constant	-0.04	-0.03	0.16	0.17	-0.01	0.00
	(0.20)	(0.20)	(0.16)	(0.16)	(0.19)	(0.19)
Num. of obs.	288	288	240	240	156	156
R^2	0.70	0.70	0.72	0.72	0.73	0.73

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of the average night-late intensity excluding unlit cells, the explanatory variable is the first difference in the yearly mean anomaly, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.6 – Fixed effect panel regression, level form in the country-level restricted sample, unlit cells excluded, yearly mean anomaly.

	Night Lights (0 censored)					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Mean Anomaly (Inst)	0.72		0.65		0.54	
	(0.44)		(0.46)		(0.51)	
Mean Anomaly (Op)		0.66*		0.63		0.60
		(0.39)		(0.40)		(0.46)
Constant	10.06***	10.06***	10.10***	10.10***	9.81***	9.81***
	(0.19)	(0.19)	(0.20)	(0.20)	(0.24)	(0.24)
Num. of obs.	312	312	260	260	169	169

Panel estimation with country fixed effect. The dependent variables is the yearly average night late intensity including unlit cells, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.7 – Level regression, country-level restricted sample, unlit cells included, strong/weak positive/negative anomalies.

	Night Lights					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
St. Neg. Anomaly Index (Inst)	0.02		0.02		0.03	
	(0.02)		(0.02)		(0.02)	
St. Pos. Anomaly Index (Inst)	0.00		0.00		0.00	
	(0.01)		(0.01)		(0.00)	
Wk. Neg. Anomaly Index (Inst)	0.02		-0.01		-0.02	
	(0.05)		(0.03)		(0.03)	
Wk. Pos. Anomaly Index (Inst)	-0.08		-0.03		0.03	
	(0.06)		(0.04)		(0.02)	
St. Neg. Anomaly Index (Op)		0.02		0.01		0.02
		(0.03)		(0.02)		(0.02)
St. Pos. Anomaly Index (Op)		0.00		0.01		0.00
		(0.01)		(0.01)		(0.00)
Wk. Neg. Anomaly Index (Op)		0.03		-0.01		-0.02
		(0.06)		(0.03)		(0.03)
Wk. Pos. Anomaly Index (Op)		-0.07		-0.03		0.03
		(0.04)		(0.03)		(0.02)
Constant	0.34*	0.33*	0.19*	0.17*	0.08*	0.08*
	(0.14)	(0.16)	(0.07)	(0.06)	(0.04)	(0.03)
Num. of obs.	312	312	260	260	169	169
R^2	0.10	0.10	0.11	0.10	0.39	0.38

Notes for Table A4.7 Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity including unlit cells, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.8 – First difference regression, country-level restricted sample, unlit cells included, strong/weak positive/negative anomalies.

	Δ Night Lights					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Δ St. Neg. Anomaly Index (Inst)	0.00		0.01		0.01	
	(0.01)		(0.01)		(0.01)	
Δ St. Pos. Anomaly Index (Inst)	-0.00		-0.00		-0.00	
	(0.00)		(0.00)		(0.00)	
Δ Wk. Neg. Anomaly Index (Inst)	-0.01		-0.01		-0.01	
	(0.01)		(0.01)		(0.01)	
Δ Wk. Pos. Anomaly Index (Inst)	-0.01		-0.01		-0.00	
	(0.01)		(0.01)		(0.01)	
Δ St. Neg. Anomaly Index (Op)		0.01		0.01		0.01*
		(0.01)		(0.01)		(0.01)
Δ St. Pos. Anomaly Index (Op)		-0.00		0.00		-0.00
		(0.00)		(0.00)		(0.00)
Δ Wk. Neg. Anomaly Index (Op)		-0.01		-0.01		-0.01
		(0.01)		(0.01)		(0.01)
Δ Wk. Pos. Anomaly Index (Op)		-0.01		-0.00		-0.00
		(0.01)		(0.01)		(0.01)
Constant	0.01	0.01	0.01	0.01	0.01*	0.01*
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Num. of obs.	288	288	240	240	156	156
R^2	0.39	0.39	0.42	0.42	0.57	0.58

Notes for Table A4.8 Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of average night-late intensity including unlit cells, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.9 – Fixed effect panel regression, level form in the country-level restricted sample, unlit cells included, strong/weak positive/negative anomalies.

	Night Lights					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
St. Neg. Anomaly Index (Inst)	-0.03		-0.01		-0.00	
	(0.02)		(0.02)		(0.01)	
St. Pos. Anomaly Index (Inst)	-0.00		0.00		-0.00	
	(0.00)		(0.00)		(0.00)	
Wk. Neg. Anomaly Index (Inst)	0.02		0.01		0.01	
	(0.03)		(0.03)		(0.01)	
Wk. Pos. Anomaly Index (Inst)	-0.01		-0.01		-0.01	
	(0.02)		(0.01)		(0.01)	
St. Neg. Anomaly Index (Op)		-0.02		-0.01		0.00
		(0.02)		(0.01)		(0.01)
St. Pos. Anomaly Index (Op)		-0.00		0.00		0.00
		(0.00)		(0.00)		(0.00)
Wk. Neg. Anomaly Index (Op)		0.01		0.01		0.01
		(0.03)		(0.03)		(0.01)
Wk. Pos. Anomaly Index (Op)		-0.01		-0.01		-0.01
		(0.01)		(0.01)		(0.01)
Constant	0.20***	0.21***	0.13***	0.14***	0.13***	0.13***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	(0.01)
Num. of obs.	312	312	260	260	169	169

Notes for Table A4.9 Panel estimation with country fixed effect. The dependent variables is the yearly average night-late intensity including unlit cells, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.10 – Level regression, country-level restricted sample, unlit cells excluded, strong/weak positive/negative anomalies.

	Night Lights (0 censored)					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
St. Neg. Anomaly Index (Inst)	-0.37		-0.36		-0.35	
	(0.27)		(0.31)		(0.28)	
St. Pos. Anomaly Index (Inst)	-0.10		-0.11		-0.03	
	(0.07)		(0.07)		(0.04)	
Wk. Neg. Anomaly Index (Inst)	-1.27**		-1.32**		-1.62***	
	(0.50)		(0.49)		(0.47)	
Wk. Pos. Anomaly Index (Inst)	-0.10		-0.33		0.02	
	(0.49)		(0.44)		(0.41)	
St. Neg. Anomaly Index (Op)		-0.25		-0.20		-0.28
		(0.29)		(0.32)		(0.32)
St. Pos. Anomaly Index (Op)		-0.09		-0.09		-0.04
		(0.07)		(0.06)		(0.04)
Wk. Neg. Anomaly Index (Op)		-1.10**		-1.00*		-1.36**
		(0.53)		(0.51)		(0.54)
Wk. Pos. Anomaly Index (Op)		-0.20		-0.31		-0.09
		(0.40)		(0.39)		(0.44)
Constant	8.89***	9.28***	9.09***	9.55***	7.42***	8.02***
	(0.84)	(0.89)	(0.79)	(0.79)	(0.84)	(0.80)
Num. of obs.	312	312	260	260	169	169
R^2	0.44	0.42	0.51	0.48	0.64	0.59

Notes for Table A4.10 Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity excluding unlit cells, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.11 – First difference regression, country-level restricted sample, unlit cells excluded, strong/weak positive/negative anomalies.

	Δ Night Lights (0 censored)					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
Δ St. Neg. Anomaly Index (Inst)	-0.08		-0.06		0.09	
	(0.21)		(0.21)		(0.20)	
Δ St. Pos. Anomaly Index (Inst)	-0.04		-0.04		-0.02	
	(0.02)		(0.02)		(0.03)	
Δ Wk. Neg. Anomaly Index (Inst)	-0.11		-0.17		-0.16	
	(0.27)		(0.27)		(0.30)	
Δ Wk. Pos. Anomaly Index (Inst)	-0.01		-0.01		0.08	
	(0.15)		(0.15)		(0.17)	
Δ St. Neg. Anomaly Index (Op)		-0.03		-0.00		0.17
		(0.21)		(0.21)		(0.21)
Δ St. Pos. Anomaly Index (Op)		-0.04		-0.03		-0.01
		(0.02)		(0.02)		(0.02)
Δ Wk. Neg. Anomaly Index (Op)		-0.12		-0.18		-0.19
		(0.26)		(0.26)		(0.31)
Δ Wk. Pos. Anomaly Index (Op)		-0.02		-0.01		0.10
		(0.14)		(0.14)		(0.16)
Constant	-0.06	-0.05	0.14	0.14	-0.02	-0.01
	(0.20)	(0.20)	(0.16)	(0.16)	(0.20)	(0.20)
Num. of obs.	288	288	240	240	156	156
R^2	0.70	0.70	0.73	0.73	0.73	0.73

Notes for Table A4.11 Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of average night-late intensity excluding unlit cells, the explanatory variables are the first difference of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.12 – Fixed effect panel regression, growth form in the country-level restricted sample, unlit cells excluded, strong/weak positive/negative anomalies.

	Night Lights (0 censored)					
	(1)	(2)	(3)	(4)	(5)	(6)
	30% HP		40% HP		50% HP	
St. Neg. Anomaly Index (Inst)	0.61**		0.64**		0.63**	
	(0.25)		(0.26)		(0.27)	
St. Pos. Anomaly Index (Inst)	0.04		0.02		0.01	
	(0.04)		(0.04)		(0.04)	
Wk. Neg. Anomaly Index (Inst)	-0.21		-0.42		-0.37	
	(0.43)		(0.44)		(0.46)	
Wk. Pos. Anomaly Index (Inst)	0.21		0.21		0.15	
	(0.23)		(0.23)		(0.25)	
St. Neg. Anomaly Index (Op)		0.64***		0.68***		0.68**
		(0.24)		(0.24)		(0.26)
St. Pos. Anomaly Index (Op)		0.03		0.03		0.02
		(0.03)		(0.03)		(0.04)
Wk. Neg. Anomaly Index (Op)		-0.18		-0.34		-0.28
		(0.41)		(0.41)		(0.47)
Wk. Pos. Anomaly Index (Op)		0.16		0.19		0.12
		(0.21)		(0.21)		(0.24)
Constant	10.47***	10.63***	10.50***	10.67***	10.31***	10.48***
	(0.35)	(0.34)	(0.39)	(0.38)	(0.49)	(0.47)
Num. of obs.	312	312	260	260	169	169

Notes for Table A4.12 Panel estimation with country fixed effect. The dependent variables is the yearly average night-late intensity excluding unlit cells, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. Only countries in which hydropower represents at least 30% of installed capacity are included in columns 1 and 2, the threshold is increased to 40% in columns 3 and 4 and to 50% in columns 5 and 6. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.13 – Level regression, country-level sample, country specific linear time-trend, yearly mean anomaly.

	(1) Night Lights	(2)	(3) Night Lights (0 censored)	(4)
Mean Anomaly (Inst)	-0.00		1.46	
	(0.03)		(0.86)	
Mean Anomaly (Op)		0.00		1.23
		(0.02)		(0.80)
Constant	0.09***	0.09***	11.32***	11.31***
	(0.00)	(0.00)	(0.08)	(0.08)
Num. of obs.	377	377	377	377
R^2	0.93	0.93	0.48	0.48

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include country specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.14 – First difference regression, country-level sample, country specific linear time-trend, yearly mean anomaly.

	(1) Δ Night Lights	(2)	(3) Δ Night Lights (0 censored)	(4)
Δ Mean Anomaly (Inst)	0.04**		0.66	
	(0.02)		(0.50)	
Δ Mean Anomaly (Op)		0.03*		0.43
		(0.02)		(0.46)
Constant	0.01	0.01	0.58	0.58
	(0.01)	(0.01)	(0.41)	(0.41)
Num. of obs.	348	348	348	348
R^2	0.02	0.02	0.01	0.01

Pooled OLS regression with s.e. clustered at the country level. The dependent variables is the first difference of average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variable is the first difference in the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include country specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.15 – Fixed effect panel regression, level form in the country-level sample, country specific linear time-trend, yearly mean anomaly.

	(1)	(2)	(3)	(4)
	Night Lights		Night Lights (0 censored)	
Mean Anomaly (Inst)	0.02		1.66**	
	(0.03)		(0.66)	
Mean Anomaly (Op)		0.02		1.49**
		(0.03)		(0.59)
Constant	-22.04***	-22.04***	-572.91***	-572.85***
	(1.90)	(1.90)	(38.90)	(38.89)
Num. of obs.	377	377	377	377

Panel estimation with country fixed effect. The dependent variables is the yearly average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variable is the yearly mean anomaly, weighted by either installed or operational capacity. All regressions include country specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.16 – Level regression, country-level sample, country specific linear time-trend, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)
	Night Lights		Night Lights (0 censored)	
St. Neg. Anomaly Index (Inst)	0.00		0.62	
	(0.01)		(0.62)	
St. Pos. Anomaly Index (Inst)	-0.00		0.09	
	(0.00)		(0.06)	
Wk. Neg. Anomaly Index (Inst)	0.01		0.88	
	(0.02)		(0.72)	
Wk. Pos. Anomaly Index (Inst)	-0.02*		-0.45	
	(0.01)		(0.45)	
St. Neg. Anomaly Index (Op)		0.00		0.56
		(0.01)		(0.59)
St. Pos. Anomaly Index (Op)		-0.00		0.08
		(0.00)		(0.06)
Wk. Neg. Anomaly Index (Op)		0.02		1.06
		(0.01)		(0.64)
Wk. Pos. Anomaly Index (Op)		-0.02**		-0.43
		(0.01)		(0.42)
Constant	0.14***	0.15***	13.63***	13.99***
	(0.02)	(0.02)	(0.81)	(0.73)
Num. of obs.	377	377	377	377
R^2	0.93	0.93	0.50	0.50

Pooled OLS regression with s.e. clustered at the country level The dependent variables is the average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include country specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.17 – First difference regression, country-level sample, country specific linear time-trend, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)
	Δ Night Lights		Δ Night Lights (0 censored)	
Δ St. Neg. Anomaly Index (Inst)	0.02		-0.14	
	(0.01)		(0.36)	
Δ St. Pos. Anomaly Index (Inst)	0.00		0.06	
	(0.00)		(0.04)	
Δ Wk. Neg. Anomaly Index (Inst)	-0.03**		-0.06	
	(0.01)		(0.50)	
Δ Wk. Pos. Anomaly Index (Inst)	-0.00		-0.30	
	(0.01)		(0.31)	
Δ St. Neg. Anomaly Index (Op)		0.02		-0.14
		(0.01)		(0.40)
Δ St. Pos. Anomaly Index (Op)		0.00		0.04
		(0.00)		(0.04)
Δ Wk. Neg. Anomaly Index (Op)		-0.03**		-0.01
		(0.01)		(0.49)
Δ Wk. Pos. Anomaly Index (Op)		-0.00		-0.30
		(0.01)		(0.28)
Constant	0.01	0.01	0.58	0.57
	(0.01)	(0.01)	(0.40)	(0.41)
Num. of obs.	348	348	348	348
R^2	0.03	0.03	0.02	0.02

Pooled OLS regression with s.e. clustered at the country level The dependent variables is the first difference of average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variables are the first differences of weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include country specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.18 – Fixed effect panel regression, level form in the country-level sample, country specific linear time-trend, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)
	Night Lights		Night Lights (0 censored)	
St. Neg. Anomaly Index (Inst)	0.03		0.67	
	(0.02)		(0.40)	
St. Pos. Anomaly Index (Inst)	0.00		0.09	
	(0.00)		(0.05)	
Wk. Neg. Anomaly Index (Inst)	-0.00		0.86	
	(0.03)		(0.64)	
Wk. Pos. Anomaly Index (Inst)	-0.00		-0.37	
	(0.02)		(0.34)	
St. Neg. Anomaly Index (Op)		0.03		0.72**
		(0.02)		(0.37)
St. Pos. Anomaly Index (Op)		0.00		0.08
		(0.00)		(0.05)
Wk. Neg. Anomaly Index (Op)		0.00		0.90
		(0.03)		(0.61)
Wk. Pos. Anomaly Index (Op)		-0.00		-0.29
		(0.02)		(0.32)
Constant	-21.72***	-21.76***	-559.75***	-560.01***
	(1.90)	(1.89)	(38.17)	(37.94)
Num. of obs.	377	377	377	377

Panel estimation with country fixed effect. The dependent variables is the yearly average night-late intensity, including unlit cells in columns 1 and 2 and excluding them in columns 3 and 4, the explanatory variables are the weak/strong positive/negative anomalies, weighted by either installed or operational capacity. All regressions include country specific linear time-trend. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.19 – Quantile regression, growth form in the country-level sample, unlit cells included, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Δ Night Lights				Δ Night Lights			
Δ Mean Anomaly (Inst)	0.0000121	-0.000113	-0.0003226	0.0003012***				
	(0.00)	(0.00)	(0.00)	(0.00)				
Δ Mean Anomaly (Op)					-0.0000586	-0.0001338	-0.000282	0.000195***
					(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.08***	0.09**	0.12**	0.16***	0.15***	0.18***	0.25***	0.27***
	(0.02)	(0.03)	(0.04)	(0.02)	(0.01)	(0.03)	(0.03)	(0.00)
Num. of obs.	345	345	345	345	345	345	345	345

Quantile regression. The dependent variables is the yearly growth of average night-late intensity including unlit cells, the explanatory variable is the average growth in the yearly mean anomaly, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.20 – Quantile regression, level form in the country-level sample, unlit cells included, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Night Lights				Night Lights			
Mean Anomaly (Inst)	-0.0087185***	-0.0026843	-0.0045535	0.0007554				
	(0.00)	(0.01)	(0.03)	(0.01)				
Mean Anomaly (Op)					-0.0045447	-0.0009267	-0.0015482	0.0029379
					(0.00)	(0.01)	(0.02)	(0.02)
Constant	0.05***	0.07***	0.08***	0.09***	0.05***	0.07***	0.08***	0.09***
	(0.00)	(0.01)	(0.02)	(0.00)	(0.00)	(0.01)	(0.02)	(0.01)
Num. of obs.	377	377	377	377	377	377	377	377

Quantile regression. The dependent variables is the yearly average night-late intensity including unlit cells, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. *** =significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.21 – Quantile regression, growth form in the country-level sample, unlit cells excluded, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Δ Night Lights (0 cens)				Δ Night Lights (0 cens)			
Δ Mean Anomaly (Inst)	-0.00000451***	0.0004862	0.0003879***	0.0002457				
	(0.00)	(0.00)	(0.00)	(0.00)				
Δ Mean Anomaly (Op)					0.000077**	0.0001668	0.0003203*	0.0002789***
					(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.04***	0.00	0.04***	0.09***	0.11***	0.20	0.21***	0.26***
	(0.00)	(0.09)	(0.01)	(0.02)	(0.01)	(0.13)	(0.02)	(0.01)
Num. of obs.	345	345	345	345	345	345	345	345

Quantile regression. The dependent variables is the yearly growth of average night-late intensity including unlit cells, the explanatory variable is the average growth in the yearly mean anomaly, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.22 – Quantile regression, level form in the country-level sample, unlit cells excluded, yearly mean anomaly.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Night Lights (0 cens)				Night Lights (0 cens)			
Mean Anomaly (Inst)	0.76***	0.60**	0.90	0.40				
	(0.00)	(0.25)	(1.57)	(0.55)				
Mean Anomaly (Op)					0.58***	0.36	0.83	0.56***
					(0.00)	(0.46)	(1.19)	(0.07)
Constant	9.73***	11.13***	11.40***	12.49***	9.77***	11.16***	11.38***	12.15***
	(0.00)	(0.19)	(1.10)	(0.46)	(0.00)	(0.40)	(0.92)	(0.06)
Num. of obs.	377	377	377	377	377	377	377	377

Quantile regression. The dependent variables is the yearly average night-late intensity including unlit cells, the explanatory variable is the yearly mean anomaly, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. *** =significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.23 – Quantile regression, growth form in the country-level sample, unlit cells included, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Δ Night Lights				Δ Night Lights			
Δ St. Neg. Anomaly Index (Inst)	-0.01***	-0.04**	-0.07***	-0.13***				
	(0.00)	(0.01)	(0.00)	(0.00)				
Δ St. Pos. Anomaly Index (Inst)	-0.01***	-0.00***	-0.00***	-0.00***				
	(0.00)	(0.00)	(0.00)	(0.00)				
Δ Wk. Neg. Anomaly Index (Inst)	0.00***	0.02**	0.04***	0.10***				
	(0.00)	(0.01)	(0.00)	(0.00)				
Δ Wk. Pos. Anomaly Index (Inst)	0.01***	0.01	-0.01***	0.00***				
	(0.00)	(0.00)	(0.00)	(0.00)				
Δ St. Neg. Anomaly Index (Op)					-0.05***	-0.05*	-0.07**	-0.14***
					(0.00)	(0.03)	(0.03)	(0.00)
Δ St. Pos. Anomaly Index (Op)					0.00***	0.00**	-0.00***	-0.00***
					(0.00)	(0.00)	(0.00)	(0.00)
Δ Wk. Neg. Anomaly Index (Op)					0.02***	0.02	0.02	0.10***
					(0.00)	(0.02)	(0.02)	(0.00)
Δ Wk. Pos. Anomaly Index (Op)					0.01***	0.01	0.01	-0.00***
					(0.00)	(0.01)	(0.01)	(0.00)
Constant	0.00***	0.01	0.01***	0.04***	0.00	0.01	0.02	0.04***
	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.02)	(0.02)	(0.00)
N	332	332	332	332	332	332	332	332

Note for Table A4.23 - Quantile regression. The dependent variables is the yearly growth of average night-late intensity including unlit cells, the explanatory variables are the growth rate of weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.24 – Quantile regression, growth form in the country-level sample, unlit cells included, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Night Lights				Night Lights			
St. Neg. Anomaly Index (Inst)	-0.01***	0.0038247*	-0.00	-0.0011281***				
	(0.00)	(0.00)	(0.00)	(0.00)				
St. Pos. Anomaly Index (Inst)	-0.0012109***	-0.00	-0.00	-0.0002992***				
	(0.00)	(0.00)	(0.00)	(0.00)				
Wk. Neg. Anomaly Index (Inst)	0.01***	0.00	0.01	0.01***				
	(0.00)	(0.00)	(0.01)	(0.00)				
Wk. Pos. Anomaly Index (Inst)	-0.01***	-0.00	-0.01	-0.0019742***				
	(0.00)	(0.00)	(0.00)	(0.00)				
St. Neg. Anomaly Index (Op)					-0.01***	0.00447***	0.00	0.0002432***
					(0.00)	(0.00)	(0.00)	(0.00)
St. Pos. Anomaly Index (Op)					-0.0011549***	0.00	-0.00	-0.000266***
					(0.00)	(0.00)	(0.00)	(0.00)
Wk. Neg. Anomaly Index (Op)					0.01***	0.01***	0.01**	0.01***
					(0.00)	(0.00)	(0.00)	(0.00)
Wk. Pos. Anomaly Index (Op)					-0.01***	-0.00	-0.01**	-0.0039612***
					(0.00)	(0.00)	(0.00)	(0.00)
Constant	0.05***	0.08***	0.08***	0.09***	0.06***	0.09***	0.10***	0.10***
	(0.00)	(0.00)	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)
N	377	377	377	377	377	377	377	377

Note for Table A4.24 - Quantile regression. The dependent variables is the yearly average night-late intensity including unlit cells, the explanatory variables are the weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.25 – Quantile regression, growth form in the country-level sample, unlit cells excluded, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Δ Night Lights (0 cens)				Δ Night Lights (0 cens)			
Δ St. Neg. Anomaly Index (Inst)	0.05***	0.06***	0.03**	0.06***				
	(0.00)	(0.01)	(0.01)	(0.01)				
Δ St. Pos. Anomaly Index (Inst)	-0.0026259***	0.0040657***	0.0037741***	0.0029589***				
	(0.00)	(0.00)	(0.00)	(0.00)				
Δ Wk. Neg. Anomaly Index (Inst)	-0.03***	-0.01**	-0.004075	-0.02***				
	(0.00)	(0.01)	(0.01)	(0.01)				
Δ Wk. Pos. Anomaly Index (Inst)	-0.02***	-0.04***	-0.02***	-0.04***				
	(0.00)	(0.00)	(0.00)	(0.00)				
Δ St. Neg. Anomaly Index (Op)					0.08***	0.09***	0.04***	0.06***
					(0.00)	(0.01)	(0.01)	(0.02)
Δ St. Pos. Anomaly Index (Op)					0.0003848***	0.0003404***	0.0002314***	0.0001975***
					(0.00)	(0.00)	(0.00)	(0.00)
Δ Wk. Neg. Anomaly Index (Op)					-0.04***	-0.04***	-0.02**	-0.02**
					(0.00)	(0.01)	(0.01)	(0.01)
Δ Wk. Pos. Anomaly Index (Op)					-0.03***	-0.04***	-0.03***	-0.04***
					(0.00)	(0.00)	(0.00)	(0.00)
Constant	-0.02***	0.05***	0.06***	0.10***	-0.00	0.04***	0.05***	0.10***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.00)	(0.01)	(0.01)	(0.01)
N	332	332	332	332	332	332	332	332

Note for Table 4.25 - Quantile regression. The dependent variables is the yearly growth of average night-late intensity excluding unlit cells, the explanatory variables are the growth rate of weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 to 4 and by operational c apacity in columns 4 to 8. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table A4.26 – Quantile regression, level form in the country-level sample, unlit cells excluded, strong/weak positive/negative anomalies.

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Night Lights (0 cens)				Night Lights (0 cens)			
St. Neg. Anomaly Index (Inst)	0.21***	0.12***	0.04	0.42***				
	(0.00)	(0.02)	(0.31)	(0.00)				
St. Pos. Anomaly Index (Inst)	0.06***	0.04***	0.07*	0.02***				
	(0.00)	(0.00)	(0.04)	(0.00)				
Wk. Neg. Anomaly Index (Inst)	0.03***	0.34***	-0.12	-0.37***				
	(0.00)	(0.04)	(0.52)	(0.00)				
Wk. Pos. Anomaly Index (Inst)	0.25***	0.29***	-0.02	-0.33***				
	(0.00)	(0.02)	(0.29)	(0.00)				
St. Neg. Anomaly Index (Op)					0.11***	0.06***	0.48	0.49***
					(0.00)	(0.02)	(0.30)	(0.00)
St. Pos. Anomaly Index (Op)					0.04***	0.03***	0.06	0.01***
					(0.00)	(0.00)	(0.04)	(0.00)
Wk. Neg. Anomaly Index (Op)					0.03***	0.39***	-0.12	-0.53***
					(0.00)	(0.03)	(0.50)	(0.00)
Wk. Pos. Anomaly Index (Op)					0.16***	0.14***	-0.07	-0.26***
					(0.00)	(0.02)	(0.27)	(0.00)
Constant	9.76***	11.22***	11.20***	12.90***	9.69***	11.36***	12.61***	12.84***
	(0.00)	(0.04)	(0.60)	(0.00)	(0.00)	(0.03)	(0.60)	(0.00)
N	377	377	377	377	377	377	377	377

Note for Table A4.26 - Quantile regression. The dependent variables is the yearly average night-late intensity excluding unlit cells, the explanatory variables are the weak/strong positive/negative anomalies, weighted by installed capacity in columns 1 to 4 and by operational capacity in columns 4 to 8. All regressions include time dummies. ***=significant at the 1% level; **=significant at the 5% level; *=significant at the 10% level.

Table B4.1 – Summary of key results, capital samples expanded through the second urbanisation procedure, mean value.

	HP > 30%	HP > 40%	HP > 50%
Yearly Mean Anomaly, Installed capacity and 4 radiuses	2.4*	2.9*	3.67**
Yealy Mean Anomaly, Operational capacity and 4 radiuses	2.82**	3.2**	4.04**
Average Anomaly , Installed capacity	0.005	0.006	0.009
Average Anomaly , Operational capacity	0.006	0.007	0.011
Average Light Intensity	29.87	30.56	29.81

Summary of key results. The first two lines report the coefficients for the yearly mean anomaly constructed with 4 radiuses as of Table. 4.29, that is when the capital sample has been expanded through the second urbanisation procedure and the mean value, the third and fourth the average values of the yearly mean anomaly in the different sample, the fifth the average night-light intensity in the capitals of the sample.

Table B4.2 – Summary of key results, capital samples expanded through the second urbanisation procedure, half of the mean value.

	HP > 30%	HP > 40%	HP > 50%
Yearly Mean Anomaly, Installed capacity and 4 radiuses	2.03*	2.37*	2.89*
Yealy Mean Anomaly, Operational capacity and 4 radiuses	2.27*	2.6**	3.17**
Average Anomaly , Installed capacity	0.005	0.006	0.009
Average Anomaly , Operational capacity	0.006	0.007	0.011
Average Light Intensity	29.24	29.9	29.17

Summary of key results. The first two lines report the coefficients for the yearly mean anomaly constructed with 4 radiuses as of Table. 4.29, that is when the capital sample has been expanded through the second urbanisation procedure and half of the mean value, the third and fourth the average values of the yearly mean anomaly in the different sample, the fifth the average night-light intensity in the capitals of the sample.

Conclusions, Limitations and Future Research.

Many SSA economies have recently been going through a period of sustained growth, with the region experiencing an average growth rate of 5% over the period 2008-2015. Together with improved economic performances relative to the last decades of the 20^{th} century, SSA has also experienced improvements in many social indicators, ranging from primary school enrolment to a reduction in the incidence of HIV/AIDS. However, for these positive developments to really lead to a more prosperous future for the inhabitants of the region, all SSA economies need, to differing extents, to continue and intensify the process of industrial diversification. The recent growth performance has been mainly due to a prolonged peak in the demand and price of the natural resources of which the continent is rich. Although their production and export offers a chance to increase government revenues, providing the necessary new funds to finance various projects, primary commodity sectors have not contributed much to the availability of non-vulnerable employment in the region, which increased by only 2.3% since 2001. The agricultural sector, accounting for around 20% of the region's GDP and 65% of its employment, remains largely unmodernised and the share of economic output deriving from manufacturing industries, historically the main contributors to stable employment, has decreased by 1.5% since 1980.

Although the reasons for the lack of industrial diversification are varied, one of the main obstacles to a firm's expansion often quoted in surveys is the quality of the electricity supply. Indeed, the status of the energy infrastructure in SSA has long left much to be desired. Generation capacity in the region has grown from 68 GW to 90 GW since 2000 but is still much below what would be required to cover demand; transmission and distribution losses are more than double the world average; the tariffs in many countries are amongst the highest in the world and, although electricity access has increased from 23% to 32% since 2000, population growth outstripped the electrification rate, so that the absolute number of people lacking access to electricity has grown of 100 million. However, it has to be recognized that the trend of under-investment in energy infrastructure has started to reverse over the last 10 years, with a significant increase in funding by national governments, private investors and overseas development funding agencies, with an ever growing role of China in both of the latter. Importantly, much of the recent additions to capacity generation have taken the form of renewable energy plants: in developing the energy resources of SSA, the growing threat of anthropogenic climate change, of which the energy sector is a main contributor, must be taken into account. The continent, despite its low contribution to green-house gases emissions, is already experiencing significant stresses due to increased extreme weather events, many of which will continue in the future.

Notwithstanding the positive developments in recent years, unreliable power supply remains a major drag on firms' ability to expand their production and successfully compete in national and international markets. Frequent power outages are a common feature of all economies in the SSA region, and their overall cost has been estimated at 2.1% of its GDP and 4.9% of its total sales, with the acquisition of expensive backup generation having become the normal coping strategy for firms which can afford it. Economic studies analysing the cost of power outages in SSA, and in developing countries more in general, are still limited, especially those focusing on its impact on firms, although this literature has been growing over the last few years. This thesis represents a further contribution to this strand of the literature, as its main aim is to provide an endogeneity-free estimate of the elasticity of firm's sales to power outages across SSA.

In the first chapter, we started with the analysis of the most recent rounds of World Bank Enterprise Surveys, covering 13,310 firms located in 38 SSA countries and having been collected between 2006 and 2014, which confirmed the relevance of both the incidence of outages (in average 150.95 per year or 881.05 hours) and of backup generation ownership (46% of firms own one). As it appears from a series of OLS regressions, the frequency and duration of outages have a significant and negative impact on firms' sales across the continent. Moreover, these negative overall effects appeared to be driven mostly by firms without access to backup generation, as point estimates for this group were always of a greater magnitude and more robust to different specifications than those for firms which had access to it. We therefore moved to the analysis of the main determinants of generator ownership, applying two different methodologies, one developed by Foster and Steinbucks (2010) and already applied to the previous rounds of WBES and one from Alby, Dethier and Straub (2011). The results from the first model closely resembled the originals, with frequency of power outages being the second most relevant determinant of generator ownership after firm size, showing how little has changed over the previous 6 years despite the increased investment in electricity infrastructure. From the second model it also emerged the significant role that having access to credit plays in being able to afford backup generation and how the effect of frequent outages on the choice of acquiring backup seems to be the same for both firm in electricity intensive sectors and those outside of them. The model also allowed for the investigation of the effect of frequent outages on the evolution of the industrial structure of SSA economies, confirming its relevance as a constraint on a firm's ability to grow in size, this time slightly more so for firms in electricity intensive sectors.

However, concerns exist about the endogeneity of the firm level relationship between performance and quality of the power supply, particularly due to possible influences of both the quality of energy infrastructure on initial plant location and of the interaction between policies directed to enhance firms' performance and a country outage level. If these concerns are founded, it follows that the previous estimates will be biased. In the second chapter, we then set forth to obtain the basis for a valid instrument affecting the frequency of outages but not directly related to firms' performance, namely the quantity of water available for hydropower production. This choice is driven both by the relevance of hydropower in the generation portfolio of many SSA countries, making it an optimal candidate to influence energy supply, and by the low number of countries without any hydropower installed in our sample, only 7 out of 38. The first step in the construction of the instrument is the modelling of the streamflow of 8 of the 9 African continental basins, which we performed through the Geospatial Streamflow Model, developed by the US Geological Survey for the Famine Early Warning System and particularly suited for data scarce regions such as SSA. After an extensive presentation of the model results, we assessed its performance by comparing them with the historical records for 440 gauge stations from the Global Runoff Data Centre, the biggest public provider of hydrological data. Our analyses showed how the final judgment on the model-fit partially dependeds on how the comparison was performed: by looking at correlation only, hence assuming that the joint distribution of simulated results and historical data is elliptical, or by using instead Copula function analysis, allowing the joint distribution to be non-symmetrical and hence the relationship to be stronger in the lower (drought) or upper (flood) tail. We also performed a series of fixed effect and quantile regressions to add further weight to our previous analyses, and probit regressions to better assess in which cases the model seems to underperform. Taken together, our results suggest that the GeoSFM model does produce better estimates of extreme rather than average values of streamflow, as the Copula which better fitted the data is the one allowing for stronger dependence in the upper tail, while quantile regressions of the simulated anomalies, defined as the difference between daily streamflow values and their long-term average scaled by their standard deviation, often yielded higher and more significant estimates for higher or lower quantiles. Moreover, this last measure, suggested by Asante *et al.* (2008a), the authors of the GeoSFM model, as the best form in which to look at its result, does indeed normally exhibit a stronger relation with the historical data than the level of streamflow. Finally, the analyses also showed how the model tends to overestimate interdependence between rivers in the same basin, probably due to the nature of the weather inputs it uses, and how performances change from basin to basin, with predominant land cover and soil type in an area being the main determinants of the model deviations from historical record. Overall, the results indicate that the GeoSFM estimates are a solid base on which to build our instrument, as the concept of anomaly is well suited to be related with hydropower production, and we have shown at the end of the chapter that yearly average anomalies are indeed a significant predictor of hydroelectricity generation in the continent.

In the third chapter we are then able to combine the economic analysis developed in the first with the hydrological analysis of the second to obtain an endogeneity free estimate of a firm's sales elasticity to power outages. The first step involved the connection of each hydropower plant with its associated basin, while the second the connection of each city for which we have firms data with the nearest surrounding power plants, both of which were performed in ArcGIS, the successor of the software of which GeoSFM is an utility. The procedure allowed us to take into account each hydropower plant's contribution to the generation capacity, and we selected two particular forms of the instrument, a simple yearly average anomaly and a series of more disaggregated indexes taking into account the incidence of negative and positive anomalies of different strengths, which were both used in 2SLS and LIML regressions. We found proof of an endogenous relationship between sales and outages for firms without generator, but not for the overall sample or for firms with access to back-up capacity. By correcting the endogeneity bias, we have shown how the negative effect for firms without access to backup generation is of a higher order of magnitude than the original OLS regressions suggested. We presented tests verifying the robustness of the results, and also performed a series of robustness checks using different specifications of both the first and second stage, which confirmed the validity of the results. In the final part of the chapter we have also investigate the presence of selection bias in the analysis for firms which own a generator, confirmed by the application of Heckman selection models. However, we were not able to determine the direction of the bias, as this was connected to the variable chosen to satisfy the exclusion restriction of the selection equation.

Taken together, our results indicate that the unreliable power supply does indeed represent a significant constraint on the performance of firms, and much more so for those without access to backup generation. By accounting for the endogenous nature of this relationship, we were able to show how substantial the gains from even a partial reduction in incidence of outages would be for the latter: if the average firm without access to backup generation could face the average hours of outage per year which an average South African firm faces, that is 118 hours instead of 562, its sales would increase by 77% or approximately 16 million dollars. As the increased funding of electricity infrastructure from government and private investors will necessarily require some time to achieve even the partial reduction in the incidence of outages to which we just referred, the analysis has relevant policy consequences. By adding further weight to the body of evidence on the damaging economic effects of unstable electricity supplies, it must serve as a reminder of the necessity to maintain, if not step up, the recent effort in improving the energy infrastructure, while also furnishing a mean to quantify part of the advantages from such investments. Any reduction in blackout incidence leads to an almost proportional increase in sales for firms who cannot switch to own generation, almost half or our sample, with various positive consequences, amongst which an increase in tax revenue for SSA governments, vital in a moment in which many are struggling to expand their tax base.

Finally, in the fourth and last chapter we investigated the connection between variations in hydropower generation and the general economic activity in SSA as proxied by night light intensity. Recently, night light data has been receiving growing attention from social scientists, and it has been found to be highly correlated with many other series, such as population distribution, urbanization rate, green-house gas emission and GDP. We modelled this relationship both at the city level and at the national level for all countries included in the third chapter, and we further attempted to account for the urbanization process which is occurring in the majority of SSA cities through two different procedures. The results of the analysis did not validate the existence of a direct link between hydroelectric production and overall economic activity, although we found evidence of a significant impact of variation in hydroelectricity generation on night-light intensity for the capital cities of countries in which hydropower represents a significant share of the installed capacity when we allowed the city boundaries of each to evolve independently from the other.

Limitations. The results of the main research question tackled in the thesis are in line with the limited literature available on the topic, although we find only a partial confirmation for the endogenous relationship between electricity and outages. Some differences also exist with regard to the work of Oseni and Pollitt (2013), who found that, although access to backup capacity substantially reduced the damages occurring from outages, firms owning a generator still incurred some impact from the unreliable power supply as few had achieved complete backup. We could not find confirmation in our data, as coefficients on outages for both OLS and 2SLS regressions were always insignificant, although this might depend on the direction of the sample selection bias which was indicated by the Heckman selection models. This difference might also be explained by the different models (revealed preferences or subjective evaluation) and estimation method (two limit tobit) used by the authors, by the different countries included in their study (12, 4 of which we do not include as they are situated in Northern Africa) and by the different round of WBES used, that of 2007. Up to that year, more specific questions about the dimension of the generator acquired and the quantity of electricity self-generated were included in the WBES, which permitted to calculate the cost of acquiring the generator and of producing a KWh of electricity, hence allowing a more explicit modelling of firms' backup decision. Unfortunately, from the following year these questions were dropped from the surveys, so that the only countries included in the study for which we have these data available are Mozambique and Senegal, not enough to explicitly verify to what extent the application of the same models to the new round of surveys would yield different results.

The results presented are robust to a series of different specifications of both the second and first stage, and we have already discussed the validity of the exclusion restriction in the second and third chapter. However a couple of points about the instrument construction must be raised.

First, As we do not have access to the most recent version of the World Electric Power Plant data, the share of each power plant installed capacity used to weight its overall contribution might not exactly coincide with the real one, and this is even truer for the operational capacity, has each year new damages can take place or restoration projects can be completed. However, given that the period under consideration in the study is that in which infrastructure investment was starting to take off, it is unlikely that each plant share in installed generation capacity used in the study will differ substantially, and the operational capacity has only been used as a robustness check. Second, despite our best effort some hydropower plants could not be geo-located and hence have been excluded from the analysis, and amongst those geo-located, some seemed not to be placed on a river, so that again they were not taken into account. This might have led to further discrepancies between the weights used to evaluate each power plant contribution, but, as discussed in the third chapter, given the average dimension of the un-located plant these discrepancies are unlikely to play a big role in driving the results, and we have shown in the second chapter that our instrument was still a significant predictor of hydroelectric generation.

A series of the study limitations are due to the general lack of data of good quality for the SSA region. It would have been ideal to complement the general analysis presented in the thesis with a key study for a single macro-region or country, as many differences exist amongst African economies and the inclusion of country dummies can account for only so much. This would have also allowed for a more specific evaluation of the performances of the chosen hydrological model, as we could have focused on a single continental basin instead of 8. Equally interesting would have been to look at a longer time frame, moving into panel data instead of limiting ourselves to cross country regressions, hence gaining further information of the interaction between energy infrastructure quality and firms' performance over time and allowing us to control for unobservable firm characteristics . However, none of these points could be tackled with the data available at this moment as there are only a handful of countries for which two, let alone three, rounds of WBES have been carried out. We attempted to work with this data, but the combination of small sample numbers, low percentages of firms which could be matched in the two round of surveys across different countries (5.1% in Ghana, 21% in Kenya or 18% in Tanzania) and low response rates did not allow for any meaningful panel analysis. The high non-response rates were also a relevant issue in the surveys used for the study, especially those to cost-related questions such as fuel or raw material cost, imposing further constraints on the measure which could be used as dependent variable. Two last points must also be raised about the structure of the WBES questionnaires. First, it would be desirable to reintroduce the question about generator capacity in the survey, as this will allow for a more detailed modelling of back-up decisions by firms. Second, there is a need of harmonising questions about capital expenditure and depreciation between the service and manufacturing questionnaires, as they are completely absent in the former and of varying quality depending on the survey's round in the latter. With regard to the possibility of using government run surveys, which might occur more frequently than the one organized by the World Bank, they either did not contain the required information (Ethiopia) or, when they did, could not be accessed despite repeated attempts (Rwanda).

Finally, as the results from the regressions of the fourth chapter in which the urbanisation rate of each capital city was modelled independently from the other showed a more stable relationship between hydropower generation and economic activity then others, we could have expanded this method to the other cities included in the sample. However, time-limitations forced us to focus on that particular sub-sample, and for similar reasons we could not experiment more with the different available versions of night light data.

Future Research. With more data availability many of the limitations of this thesis could be addressed. As more governments fully realise the extent to which electricity supply constrains their economic development and more round of WBES are carried out we expect that firm and energy related data from SSA will become easier to come by. Apart from permitting us to use more refined productivity measures and to move into a panel data framework, these might allow for a more explicit modelling of how energy related issues affect a firm's capital intensity or its level of technology, which is a relevant determinant of its chances to successfully compete in national and international markets. Were more information about backup generation to make it back into the questionnaire, it would also be possible to dedicate more attention to which firms stand to gain the most from accessing it, which could also be useful to give more specific policy suggestions on how to tackle short and medium term deficiencies of the power sector. Similarly, more work can be directed to identify other possible variables to be used for the exclusion restriction of the Heckman selection models presented in the last part of Chapter 3.

Moreover, with more time available further experimentation with the night light data would be possible. First, we could expand our hydrological modelling back to 1992, the first year for which night light data is available, substantially increasing the coverage of the study. Secondly, we could continue with our modelling effort of the urbanization process, expanding the methodology used for the capital cities to all other cities located in countries where hydropower generation is present. Thirdly, we could explore the use of different versions of the night light dataset to obtain some other measures of the incidence of outages as at least one other author has tried to do recently. If a more stable relationship could be found, this would allow for a more general assessment of the contribution of hydropower generation to economic activity in SSA.

Finally, it would be of extreme interest to the author to assess the contribution which off-grid mini, micro and pico hydropower can give to the electrification of rural areas. A necessary condition for achieving higher standards of living for the majority of people in the continent is the fruitful development of the agricultural sector. At present the sector's lack of access to modern inputs, including irrigation which critically depends on electricity, represents a major obstacle on that path. Furthermore, the electrification of rural areas plays a vital role in fostering the growth of agriculture-related industries which have normally given an important contribution to the first stages of industrial development. There is still a lack of studies on the economic effect of rural electrification in SSA, probably due to the absence of readily available data from rural electrification agencies across the continent (a fact confirmed by an interview with the Tanzanian Rural Energy Agency in 2013), without which it is hard to strongly argue for its relevance. With more time and resources at hand, this would be a research agenda surely worth pursuing.

Bibliography.

Abeberese, A.B. (2012) "Electricity Cost and Firm Performance: Evidence from India", *Columbia University Working Paper*

Adenikinju, A.F. (2003) "Electric Infrastructure Failure in Nigeria: A Survey-based Analysis of the Costs and Adjustments Responses", *Energy Policy* 31, pp. 1519-1530

AFREPREN (2004) "The Regulation of the Power Sector in Africa: Attracting Investment and Protecting the Poor" eds. by Marandu, E. and Kayo, D., African Energy Policy Research Network, Nairobi

Agbemabiese, Nkoma and Sokona (2012) "Enabling Innovations in Energy Access: An African Perspective", *Energy Policy* 47, pp. 38-47;

Alam, M.M. (2014) "Coping with Blackouts: Power Outages and Firm Choices", (Order No. 3580619). Available from ProQuest Dissertations & Theses Global. (1539578867). Retrieved from http://search.proquest.com/docview/1539578867?accountid=8630

Alby, P., Dethier, J.J. And Straub, S. (2011) "Let There Be Light! Firms Operating under Electricity Constraints in Developing Countries" *Toulouse School of Economics Working Paper*

Allcott, H., Collard-Wexler, A. and O'Connell,S. (2014) "How do Electricity Shortages Affect Productivity? Evidence from India", *NBER Working Paper* 19977

Andersen, T.B. and Dalgaard, C.J. (2013) "Power Outages and Economic Growth in Africa", *Energy Economics* 38, pp. 19-23

Anderson, T.W. and Rubin, H. (1949) "Estimation of the parameters of a single equation in a complete system of stochastic equations" *Annals of Mathematical Statistics* 20, pp. 46-63;

Andrews, D.W.K., Moreira, M.J. and Stock, J.H. (2007) "Performance of conditional Wald tests in IV regression with weak instrument" *Journal of Econometrics* 139, pp. 116-132;

Apeaning, R.W. and Thollander, P. (2013) "Barriers to and driving forces for industrial energy efficiency improvements in African Industries – a case study of Ghana largest industrial area" *Journal of Cleaner Production* 53, pp. 204-213

Artan, G.A. *et al.* (2008) "User Manual for the Geospatial Stream Flow Model (GeoSFM)", U.S. Geological Survey Open-File Report 2007-1440

Asante, K.O. *et al.* (2008a) "A lineal geospatial streamflow modelling system for data sparse environment", *International Journal of River Basin Management* (6), pp. 233-241

Asante, K.O. *et al.* (2008b) "Technical Manual for the Geospatial Stream Flow Model (GeoSFM)", U.S. Geological Survey Open-File Report 2007-1441

Aschauer , D.A. (1989) "Is Public Expenditure Productive?" *Journal of Monetary Economics* 23, pp. 177-200

Barbe, P., Genest, C., Ghoudi, K. and Remillard B. (1996) "On Kendall's process" *Journal of Multivariate Analysis* 58, pp. 197-229

Barouni, M. and Broeke, S. (2014) "The Returns to Education in Africa: Some New Estimates", *The Journal of Development Studies (50),* pp.1593-1613

Barrios, S., Bertinelli, L. and Strobl, E. (2010) "Trends in rainfall and economic growth in Africa: a neglected cause of the African growth tragedy", *The Review of Economics and Statistics* (92), pp. 350-366

Bartelsman, E.J., Haltwinger, J.C. and Scarpetta, S. (2009) "Cross-Country Differences in Productivity: the Role of Allocation and Selecting" *NBER Working Paper Series* 15490

Brown, Muller and Dobrotkova (2011) “Renewable Energy: Markets and Prospects by Technology”, *International Energy Agency*, Paris

Buell, C.E. (1975) “The topography of empirical orthogonal functions”, *[Preprints of the] Fourth Conference on Probability and Statistics in Atmospheric Sciences*, Tallahassee, FL.

Calderon, C. (2009) “Infrastructure and Growth in Africa”, *World Bank Policy Research Working Paper 4914*

Canning, D. and Perdoni, P. (1999) “Infrastructure and Long Run Economic Growth” *CAE Working Paper* 90-99

Canning, D. and Bennathan, E. (2000) “The Social Rate of Return on Infrastructure Investments”, *World Bank Policy Research Working Paper 2390*

Christensen, L.R. and Greene, W.H. (1976) “Economies of Scale in U.S. Electric Power Generation”. *Journal of Political Economy* 84, pp. 655-676

Cleeve, E.A., Debrah, Y. and Yiheys, Z. (2015) “Human Capital and FDI Inflow: An Assessment of the African Case”, *World Development (74)*, pp. 1-14

Colclough, C., Kingdon, G. and Patrinos, H.A. (2010) “The changing pattern of wage returns to education and its implication” *Development Policy Review (28)*, pp.733-747

Cohen, A.J. and Harcourt, G.C. (2003) “Whatever happened to the Cambridge Capital Theory Controversies?” *Journal of Economic Perspectives* 17, pp. 199-214

Cole, M.A., Elliott, R.J. and Strobl, E. (2014) “Climate Change, Hydro-dependency, and the African Dam Boom”, *World Development* (60), pp. 84-98

Collier, P. (2006) “African Growth: Why a ‘Big Push’?”, *Journal of African Economies (15),* pp.188-211

Conway, D. *et al.* (2009) "Rainfall and Water Resource Availability in Sub-Saharan Africa during the 20th Century", *Journal of Hydrometeorology* (10), pp. 41-59

Cook, P. (2011) "Infrastructure, Rural Electrification and Development" *Energy for Sustainable Development* 15, pp. 304-313

Core Writing Team, Pachauri, R.K. and Reisinger, A. (2007) "Climate Change 2007: Synthesis Report. Contribution of Working Groups I, II and III to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change", *International Panel on Climate Change,* Geneva

Cui, J., Lapan, H. and Moschini, G. (2012) "Are Exporters More Environmentally Friendly than Non Exporters? Theory and Evidence" *Iowa State University Staff General Research Paper* 35548

Cragg, J.G. and Donald, S.G. (1993) "Testing Identifiability and Specification in Instrumental Variable Models" *Econometric Theory* 9, pp. 222-240

Danquah, M. and Ouattara, B. (2014) "Productivity Growth, Human Capital and Distance to Frontier in Sub-Saharan Africa", *Journal of Economic Development (39),* pp.27-48.

Davies, J., Kang, A. and Vincent, J. (2001) "How Important is Improved Water Infrastructure for Microenterprises? Evidence from Uganda" *World Development* 29, pp. 1753-1767

De Sherbinin *et al.* (2002) "Social Science Applications of Remote Sensing" A CIESIN Thematic Guide, Center for International Earth Science Information Network of Columbia University

de Wit, M. and Stankiewicz, J. (2006) "Changes in surface water supply across Africa with predicted climate change", *Science (*31), pp. 1917-1921

Diagne, A. and Diene, B. (2011) "Estimating returns to higher education: A survey of models, methods and empirical evidence", *Journal of African Economies (20),* pp. iii80-iii132

Dinkelman, T. (2011) "The Effects of Rural Electrification on Employment: New Evidence from South Africa", *American Economic Review* 101, pp. 3078-3108

Dinwiddy, C.L. and Teal, F. (1996) "Principle of Cost-Benefit Analysis for Developing Countries" Cambridge University Press, Cambridge, UK

Doll, C.N.H *et al.* (2000) "Night-time Imagery as a Tool for Global Mapping of Socioeconomic Parameters and Greenhouse Gas emissions", *Ambio (29),* pp.57-162

Doll, C.N.H *et al.* (2005) "Mapping regional economic activity from night-time light satellite imagery", *Ecological Economics (57),* pp.75-92

Doll, C.N.H. and Pachauri, S. (2010) "Estimating rural populations without access to electricity in developing countries through night-time light satellite imagery", *Energy Policy (38),* pp.5661-5670.

Duflo, E. and Pande, R. (2007) "Dams", *Quarterly Journal of Economics* 122, pp. 601-646

Ebener, S. *et al.* (2005) "From wealth to health: modelling the distribution of income per capita at the sub-national level using night-time light imagery", *International Journal of Health Geographics (4),* pp.1-17

Eberhard, A. *et al.* (2008) "Underpowered: The State of the Power Sector in Sub-Saharan Africa", *Africa Infrastructure Country Diagnostic –World Bank* ,Washington

Eberhard, A. *et al.* (2011) “Africa’s power infrastructure: investment, integration, efficiency”, *Africa Infrastructure Country Diagnostic –World Bank*, Washington

Eifert, B. and Gelb, A. (2008) “The Cost of Doing Business in Africa: Evidence from Enterprise Survey Data”, *World Development* 36, pp. 1531-1546

Elias, R.J. and Victor, D.G. (2005) “Energy Transitions in Developing Countries: a Review of Concepts and Literature”, *Stanford Working Paper* 40

Elvidge, C.D. *et al.* (1997) “Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption”, *International Journal of Remote Sensing (18),* pp.1373-1379

Elvidge, C.D. *et al.* (2009) “A global poverty map derived from satellite data”, *Computers & Geoscience (35),* pp.1652-1660

Escribano, A. , Guasch, J.L. And Pena, J. (2009) “Assessing the impact of Infrastructure Quality on Firm Productivity in Africa”, *Africa Infrastructure County Diagnostic Working Paper* 9 ;

Esfahani, H.S. And Ramirez, M.T. (2003) “Institutions, Infrastructure and Economic Growth”, *Journal of Development Economics* 70, pp. 443-477

Fisher-Vanden, K., Mansur, E. and Wang, Q.J. (2008) “Costly Blackouts: Measuring Productivity and Environmental Effects of Electricity Shortages” *NBER Working Paper* 1774

Foster, J.L (1983) "Observations Of The Earth Using Nighttime Visible Imagery", *Proc. SPIE 0414*, Optical Engineering for Cold Environments, 187 (September 22, 1983);

Foster, V. and Briceño-Garmendia (2009) “Africa’s Infrastructure: A Time for Transformation”, *Africa Development Forum Series,* World Bank, Washington

Fosu, A., Bates, R. and Hoeffler, A. (2006) "Institutions, Governance and Economic Development in Africa: An Overview" *Journal of African Economies (15),* pp. 1-9

Fredricks, G.A. and Nelsen, R.B. (2007) "On the relationship between Spearman's rho and Kendall's tau for pairs of continuous random variable", *Journal of Statistical Planning and Inference* (137), pp. 2143-2150

Foster, V. and Steinbucks, J. (2009) "Paying the Price for Unreliable Power Supplies. In-House Generation of Electricity by Firms in Africa.", *World Bank Policy Research Working Paper* 4913 ;

Furnans, J. and Olivera, F. (2001) "Watershed Topology – the Pfafstetter System", Proceedings from the 21st ESRI User Conference, ESRI, San Diego, CA

Geginat, C. and Ramalho, R. (2015) "Electricity Connections and Firm performance in 183 Countries", *World Bank Policy Research Working Paper* 7460

Genest, C. and Favre, A.C. (2007) "Everything You Always Wanted to Know about Copula Modeling but Were Afraid to Ask", *Journal of Hydrologic Engeneering* (12), pp. 347-368

Gosh, T. *et al.* (2009) "Estimation of Mexico's Informal Economy and Remittances Using Nighttime Imagey", *Remote Sensing (1),* pp.418-444

Gosh, T. *et al.* (2010) "Shedding Light on the Global Distribution of Economic Activity", *The Open Geography Journal (3),* pp.148-161

Goulden, M. *et al.* (2009) "Adaptation to Climate Change in International River Basin in Africa: a Review", *Hydrological Science* (54), pp. 805-828

Gramlich, E.M. "Infrastructure Investment: A Review Essay", *Journal of Economics Literature* 32, pp. 1176-1196

Gutman, J. *et al.* (2015) "Financing Africa Infrastructure: Can the World Deliver?", *Global Economy and Development,* Brookings Institution, Washington

Harrison, A.E., Lin, J.Y. And Xiu, L.C. (2013) "Explaining Africa's (Dis)advantage. The Curse of Party Monopoly" *World Bank Policy Research Paper* 6316;

Henderson, J.V. *et.al.* (2012) "Measuring Economic Growth from Outer Space", *America Economic Review (102),* pp.994-1028

Heckman, J. (1976) "The common structure of statistical models of truncation, sample selection and limited dependent variables" *Annals of Economic and Social Measurement (5),* pp. 475-492

Heckman, J. (1979) "Sample Selection Bias as a Specification Error" *Econometrica (47),* pp.153-161

Heckman, J., La Londe, R. and Smith, J. (1999) "The economics and econometrics of active labor market programs", in Ashenfelter O. and Card D. (eds) *Handbook of Labor Economics (3A),* pp. 1865-2097

Hickey, S. (2012) "Beyond 'Poverty Reduction through Good Governance': The New Political Economy of Development in Africa", *New Political Economy (17),* pp.683-690

Hirschman, A.O. (1958) "The Strategy of Economic Development", Yale University Press

Hofert, M. and Maechler, M (2011) "Nested Archimedean Copulas Meet R: The Nacopula Package", *Journal of Statistical Software* 39(9), pp. 1-20

Hofert, M., Kojadinovic, I., Maechler, M. and Yan, J. (2015) copula: Multivariate Dependence with Copulas. R package version 0.999-13

Holmstorm, B. and Tirole, J. (1997) "Financial Intermediation, Loanable Funds, and the Real Sector" *Quarterly Journal of Economics* 112, pp. 663-691

Horn, J. (1965) "A rationale and test for the number of factors in factor analysis", *Psychometrika* 30(2), pp. 179-185

Hulme, M. *et al.* (2001) "African climate change: 1900-2100", *Climate Research* (17), pp. 145-168

IEA (2014) "Africa Energy Outlook: A Focus on Energy Prospects in Sub-Saharan Africa", *World Energy Outlook Special Report*, IEA, Paris

IEA (2016) "Boosting the Power Sector in Sub-Saharan Africa: China's Involvement", *Partner Country Series*, IEA, Paris

Iimi, A. (2011) "Effects of Improving Infrastructure Quality on Business Costs: Evidence from Firm Level Data in Eastern Europe and Central Asia", *The Developing Economics* 49, pp. 121-147

ILO (2014) "Global Employment Trend 2014: Risk of a jobless recovery?", ILO, Geneva

IMF (2015) "Regional Economic Outlook: Sub-Saharan Africa, April 2015: Navigating Headwinds", IMF, Washington

IMF (2016) "Regional Economic Outlook: Sub-Saharan Africa, April 2016: Time for a Policy Reset", IMF, Washington

Imhoff, M.L. *et al.* (1997) "Using Nighttime DMSP/OLS Images of City Lights to Estimate the Impact of Urban Land Use on Soil Resources in the United States", *Remote Sensing of Environment (59),* pp.105 117

Jensen, S.K. and Dominque, J.O. (1988) "Extracting Topographic Structure from Digital Elevation Data for Geographic Information System Analysis", *Photogrammetric Engineering and Remote Sensing* (54), pp. 1593-1600

Kaiser, H. (1960) "The Application of Electronic Computers to Factor Analysis" *Educational and Psychological Measurement* 20(1), pp. 141-151

Kapangaziwiri, E. *et al.* (2009) "Towards the development of a consistent uncertainty framework for hydrological predictions in South Africa", in *New Approaches to Hydrological Prediction in Data-sparse Regions,* IAHS Publ.333, pp.84-93

Karekezi, S., Kimani, J., Onguru, O., & Kithyoma, W. (2009) "Large hydropower, renewable energy and adaptation to climate change: Climate change and energy security in east and horn of Africa", AFREPREN/FWD Occasional Paper No. 33

Khan, M.H. (2012) "Governance during Social Transformations: Challenges for Africa", *New Political Economy (17),* pp.667-675

Kojadinovic, I. and Yan, J. (2010) "Modeling Multivariate Distributions with Continuous Margins Using the copula R Package", *Journal of Statistical Software* 34(9), pp. 1-20

Kumar *et al.* (2011) "Hydropower. IPCC Special Report on Renewable Energy Sources and Climate Change Mitigation", *International Panel on Climate Change,* Geneva

Lee, K. S., Anas, A. and Oh, G.T. (1996) "Cost of Infrastructure Deficiencies in Manufacturing in Indonesia, Nigeria, and Thailand", *World Bank Policy Research Paper* 1604

Leung, S.F. and Yu, S. (2000) "Collinearity and Two-Step Estimation of Sample Selection Models: Problems, Origins, and Remedies" *Computational Economics (15),* pp.173-199

Lio, M. and Liu, M.C. (2006) "ICT and agricultural productivity: evidence from cross-country data" *Agricultural Economics* 34, pp. 221-228

Lipscomb, M., Mobarak, A.M. And Barham. T (2013) "Development Effect of Electrification: Evidence from the Topographic Placement of Hydropower Plants in Brazil" *Applied Economics Journal: Applied Economics* 5, pp. 200-231

Loveland, T.R. *et al.* (2000) "Development of a Global Land Cover Characteristics Database and IGBP DISCover from 1-km AVHRR Data", *International Journal of Remote Sensing* (21), pp.1303-1330

Luo, X. (2004) "The Role of Infrastructure Investment Location in China's Western Development" *World Bank Working Paper* 3345

Ma, T. *et al.* (2011) "Quantitative estimation of urbanization dynamics using time series of DMSP/OLS nighttime light data: A comparative case study from China's cities", *Remote Sensing of Environment (124),* pp.99-107

Michalopoulos, S. and Papaioannou, E. (2013) "Pre-colonial ethnic institutions and contemporary African development" *Econometrica (81),* pp.113-152

Michalopoulos, S. and Papaioannou, E. (2014) "National Institutions and Subnational Development in Africa", *The Quarterly Journal of Economics (129),* pp.151-213

Mikosch, T. (2006) "Copulas: Tales and Facts", *Extremes* (9), pp. 3-20

Moreira, M.J. (2003) "A conditional likelihood ratio test for structural model" *Econometrica* 71, pp. 1027-1048;

Moyo, B. (2012a) "Infrastructure Quality and Manufacturing Exports in Africa: a Firm Level Analysis", *South African Journal of Economics* 80, pp. 367-386

Moyo, B. (2012b) "Do Power Cuts Affect Productivity? A Case Study of Nigerian Manufacturing Firms", *International Business & Economics Research Journal* 11, pp. 1163-1174

Muller, Marmion and Beerepot (2011) "Renewable Energy: Markets and Prospects by Region", *International Energy Agency*, Paris

Munzimi, Y. (2008) "Satellite derived rainfall estimates (TRMM Products) used for the hydrological prediction of the Congo River flow: overview and preliminary results", START Report

Mveyange, A. (2015) "Night lights and regional income inequality in Africa", WIDER Working Paper 2015/085.

Nerlove, M. (1963) "Return to Scale in Electricity Supply" In *Measurement in Economics: Studies in Mathematical Economics and Econometrics in Memory of Yehuda Grunfeld,* by Christ, C.F. e*t al.*

Ng, F. and Yeats, A. (1999) "Good Governanc and Trade Policy: Are those the Keys to Africa's Global Integration and Growth?", The World Bank, Policy Research Working Paper Series 2038

Niang, I. *et al.* (2014) "Africa" in *Climate Change 2014: Impacts, Adaptation, and Vulnerability. Part B: Regional Aspects. Contribution of Working Group II to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change,* eds. by Barros, V.R. *et al.*, Cambridge University Press, Cambridge

Noman, A. and Stiglitz, J.E. (2012) "Strategies for African Development" in Noman, A., Botchwey, K., Stein, H., and Stiglitz, J.E. (eds.) *Good Growth and Governance for Africa: Rethinking Development Strategies,* pp. 3-47

North, G.L., Bell, T.L., Cahalan, R.F. and Moeng, F.J. (1982) "Sampling Errors in the Estimate of Empirical Orthogonal Functions", *Monthly Weather Review* (10), pp. 609-706

O'Lenic, E. and Livezey, R.E. (1988) "Practical considerations in the use of rotated principal components analysis (RPCA) in diagnostic studies of upper_air height fields". *Monthly Weather Review* (116), pp. 1682-1689

Oseni, M.O. and Pollitt, M.G. (2013) "The Economic Costs of Unsupplied Electricity: Evidence from Backup Generation among African Firms", *Cambridge Working Paper in Economics* 1351

Page, J. (2012) "Can Africa Industrialize?", *Journal of African Economies* 21, pp. ii86-ii124 ;

Pearce, F. (2006) "Uganda pulls plug on Lake Victoria", *New Scientist* (2538), p.12

Peters, J., Vance, C., Harsdorff, M. (2010) "Grid Extension in Rural Benin: Micro-Manufacturers and the Electrification Trap", *World Development (39),* pp. 773-783

Powell, D. (2014) "Did the Economic Stimulus Payment of 2008 Reduce Labour Supply? Evidence from Quantile Panel Data Estimation", *RAND Corporation Publications Department Working Papers*

Psacharopoulos, G. and Patrinos, H.A. (2004) "Returns to investment in education: A further update", *Education Economics (12),* pp.111-134

Quah, E. (2012) "Cost-Benefit Analysis in Developing Countries: What's Different?" *Economic Growth Centre Report No: 2012/05*

Ram (2008) “Renewable Energy Development in Africa: Challenges, Opportunities and Way Forward”, *African Development Bank*, Tunis

Renard, B. and Lang, M. (2007) “Use of Gaussian Copula for multivariate extreme value analysis: Some case studies in hydrology”, *Advances in Water Resources* (30), pp. 897-912

Reinikka, R and Svensson, J. (2002) “Coping with poor public capital” *Journal of Development Economics* 69, pp. 51-69

Richman, M.B. (1981) “Obliquely Rotated Principal Components: An Improved Meteorological Map Typing Technqiue?”, *Journal of Applied Meteorology* 20, pp. 1145-1159

Richman, M.B. (1986) “Rotation of Principal Components”, *Journal of Climatology* 6, pp. 293-335

Richman, M.B. and Lamb, P.J. (1985) “Climatic pattern analysis of 3- and 7-day summer rainfall in central United States: Some methodological considerations and a regionalization” *Journal of Climate and Applied Meteorology* 24, pp. 1325-1343

Rosenberg, N. (1998) “The Role of Electricity in Industrial Development”, *The Energy Journal* 19, pp. 7-24

Sachs, J.D. *et al* (2004) “Ending Africa’s Poverty Trap”, *Brooking Papers on Economic Activity 1:2004*, pp.117-240

Sartori, A. (2003) “An Estimator for some binary-outcome selection models without exclusion restriction” *Political Analysis (11)*, pp. 111-138

Scheumann, W. and Alker, M. (2009) “Cooperation on Africa’s transboundary aquifers – conceptual ideas”, *Hydrological Science Journal* (54), pp.793-802

Schölzel, C. and Friedrichs, P. (2008) "Multivariate non-normally distributed random variables in climate research – introduction to the copula approach", *Nonlinear Processes in Geophysics* (15), pp. 761-772

Schulze, R. *et al.* (2001) "Present and potential future vulnerability of eastern and southern Africa's hydrology and water resources", *South African Journal of Science* (97), pp. 150-160

Shea, J. (1997) "Instrument Relevance in Multivariate Linear Models: a Simple Measure" *Review of Economics and Statistics* 79, pp.348-352

Singh, K.K. and Singh, S.V. (1996) "Space-time variation and regionalization of seasonal and monthly summer monsoon rainfall of the Sub-Himalayan region and Gangetic plane of India", *Climate Research* 6, pp. 251-262

Sokona, Mulugetta and Gujba (2012) "Widening Energy Access in Africa: Towards Energy Transition", *Energy Policy* 47, pp. 3-10

Söderbom, M. and Teal, F. (2000) "Skills, Investments and Exports from Manufacturing Firms in Africa", *The Journal of Development Studies (37),* pp. 13-43

Ssegane, H., Tollner, E.W., Mohamoud, Y.M., Rasmussen, T.C. and Dowd, J.F. (2012) "Advances in variable selection methods I: Causal selection methods versus stepwise regression and principal component analysis on data of known and unknown functional relationships" *Journal of Hydrology* 438-439, pp. 16-25

Ssozi, J. and Asongu, S.A. (2016) "The Comparative Economics of Catch-Up in Output per Worker, Total Factor Productivity and Technological Gain in Sub-Saharan Africa", *African Development Review (28)*, pp. 215-28

Steinbucks, J. and Foster, V. (2010) "When do firms generate? Evidence on in-house electricity supply in Africa", *Energy Economics* 32, pp. 505-514

Stock, J.H., Wright, J.H. and Yogo, M.(2002) “A survey of weak of instruments and weak identification in generalized method of moments” *Journal of Business and Economic Statistics* 20, pp. 518-529

Stock, J.H. and Yogo, M. (2005) “Testing for weak instruments in linear IV regression” in “*Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg”,* ed. Andrews, D.W.K. and Stock, J.H., Cambridge University Press.

Sullivan, W. T. III (1989) “A 10 km resolution image of the entire night-time Earth based on cloud-free satellite photographs in the 400-1100 nm band”, *International Journal of Remote Sensing, (10),* pp.1-5

Sutton, P.C. *et al.* (2001) “Census from Heaven: an estimate of the global human population using night-time satellite imagery”, *International Journal of remote sensing (22),* pp.3061-3076

Sutton, P.C. and Costanza, R. (2002) “Global estimates of market and non-market values derived from nighttime satellite imagery, land cover, and ecosystem service valuation”, *Ecological Economics (41),* pp.509-527

Sutton, P.C. (2003) “A scale-adjusted measure of “Urban sprawl” using nighttime satellite imagery” *Remote Sensing of the Environment (86),* pp.353-369

Sutton, P.C. *et al.* (2003) “Building and Evaluating Models to Estimate Ambient Population Density”, *Photgrammetric Engineering & Remote Sensing (69),* pp.545-553

Teal, F. (2010) “Higher Education and Economic Development in Africa: a Review of Channels and Interaction”, *Centre of Study of the African Economies Working Paper Series* 2010-25

Tobler, W.R. (1969) "Satellite confirmation of settlement size coefficients", *Area (1),* pp.31-34

Troutman, B.M. and Karlinger M.R. (2003) "Regional flood probabilities", *Water Resources Research* (39), pp.1-15

Tshimanga, R.M. *et al.* (2011) "Initial Calibration of a semi-distributed rainfall runoff model for the Congo River basin", *Physics and Chemistry of the Earth* (36), pp. 761-774

Tshimanga, R.M. *et al.* (2012) "Climate change and impacts on the hydrology of the Congo Basin: The case of the northern sub-basins of the Oubangui and the Sangha Rivers", *Physics and Chemistry of the Earth* (50-52), pp. 72-83

Tybout, J.R. (2000) "Manufacturing Firms in Developing Countries: How Well do They Do, and Why?", *Journal of Economic Literature* 38, pp. 11-44

UNDP (2006) "Beyond scarcity: Power, poverty and the global water crisis", *Human Development Report*, UNDP, New York

UNECA (2011) "Governing development in Africa - the role of the state in economic transformation", UNECA, Addis Ababa

UNECA (2013) "Making the Most of Africa's Commodities: Industrializing for Growth, Jobs and Economic Transformation" UNECA, Addis Ababa

UNECA (2014) "Dynamic Industrial Policy in Africa: Innovative Institutions, Effective Processes and Flexible Mechanisms" UNECA, Addis Ababa

UNECA (2015) "MDG Report 2015: Assessing Progress in Africa toward the Millennium Development Goals", UNECA, Addis Ababa

UNECA (2016a) "Greening Africa's Industrialization" UNECA, Addis Ababa

UNECA (2016b) ”Transformative Industrial Policy for Africa” UNECA, Addis Ababa

Verdin, K.L. (1997) “A System for Topologically Coding Global Drainage Basins and Stream Networks”, Proceedings from the 17th ESRI User Conference, ESRI, San Diego, California

Verdin, J.P. and Klavier, R.W. (2002) “Grid Cell Based Crop Water Accounting for the Famine Early Warning System”, *Hydrologic Processes* 16, pp. 1617-1630

Verheust, L. and Johnson, G. (1998) “The Watershed Database for Sub-equatorial Africa, Structure and User Interface”, *ALCOM Working Paper No.* 16

Webb, R.S. *et al.* (1993) “Specifying Land Surface Characteristics in General Circulation Model: Soil Profile Data Set and Derived Water-holding Capacities”, *Global Biogeochemical Cycles* 7, pp.97-108

Welch, R. (1980) “Monitoring urban population and energy utilization patterns from satellite data”, *Remote Sensing of the Environment (9),* pp.1-9

Werth, S. (2009) “Integration of GRACE mass variation into a global hydrological model”, *Earth Planet Science Letters* (1-2), pp. 166-173

Westra, S., Brown, C., Lall, U. and Sharma, A. (2007) “Modelling multivariable hydrological series: Principal component analysis or independent component analysis” *Water Resource Research* 43, W06429

World Bank (1994) “World development report 1994: Infrastructure for development”, World Bank, Washington.

World Bank (2016) “Global Economic Prospects, June 2016: Divergences and Risk”, World Bank, Washington

World Commission on Dams (2000) “Dams and Development: A New Framework for

Decision-Making", *EarthScan Publications,* London

Xi, C. and Nordhaus, W.D. (2011) "Using luminosity data as a proxy for economic status" *PNAS (108),* pp.8589-8594

Xie, P. and Arkin P.A. (1997) "A 17-year Monthly Analysis Based on Gauge Observations, Satellite Estimates and Numerical Model Outputs", *Bulletin of the American Meteorological Society* 78, pp. 2539-2558

Yan, J. (2007) "Enjoy the Joy of Copulas: With a Package copula", *Journal of Statistical Software* 21(4), pp.1-21

Zobler, L. (1986) "A World Soile File for Global Climate Modelling", *NASA Technical Memorandum* 87802