Ecole Doctorale des Sciences Economiques, Juridiques, Politiques et de gestion
Centre d'Etudes et de Recherche sur le Développement International (CERDI)

Université Clermont Auvergne, CNRS, IRD, CERDI, F-63000 Clermont-Ferrand, France

EVALUATION DES IMPACTS ECONOMIQUES DE L'ACCES A L'ENERGIE PAR TELEDETECTION ET ENQUETES DE TERRAIN

Thèse présentée et soutenue publiquement le 10/10/2025

pour l'obtention du titre de Docteur en Sciences Economiques

par

Vincent Nossek

sous la direction de Vianney Dequiedt

Membres du Jury

Michael Goujon	Professeur, Université Clermont Auvergne, CERDI	Président
Flore Gubert	Directrice de recherche, IRD, LEDa	Rapporteur
Ahmed Tritah	Professeur, Université de Poitiers, LéP	Rapporteur
Stéphane Straub	Economiste en chef pour les infrastructures, Banque Mondiale	Suffragant
Vianney Dequiedt	Professeur, Université Clermont Auvergne, CERDI	Directeur de thèse

L’université Clermont Auvergne n’entend donner aucune approbation ni improbation aux opinions émises dans cette thèse. Ces opinions doivent être considérées comme propres à leur auteur.

Remerciements

Je tiens à exprimer ma profonde gratitude à l'équipe du programme Transition énergétique et numérique vers le développement durable de la Ferdi, en particulier au Professeur Jean-Claude Berthelemy et à Christophe Angely, pour leur accompagnement précieux tout au long de cette thèse.

Je remercie également mon directeur de thèse, Vianney Dequiedt, pour son encadrement rigoureux, sa disponibilité et son écoute attentive, ainsi que le Professeur Patrick Guillaumont, Président de la Ferdi, pour son soutien constant.

Je souhaite adresser mes sincères remerciements aux membres de mon jury, Professeure Flore Gubert, Professeur Ahmed Tritah, Professeur Stéphane Straub et le président du jury Professeur Michael Goujon, pour avoir pris le temps de lire et d'évaluer ce travail.

Ma reconnaissance va aussi à l'ensemble de mes collègues et amis de la Ferdi, et tout particulièrement à Olivier Santoni, pour son aide précieuse et son appui indéfectible tout au long de mon parcours doctoral. Je pense également à Sosso Feindouno, Matthieu Boussichas, Alou Dama, Edouard Mien, Yannick Bouterige, Florian Léon, Joël Cariolle, Audrey-Anne de Ubeda et Émilie Fabreguettes, pour leur bienveillance et leurs échanges stimulants.

Je remercie aussi chaleureusement les anciens doctorants et amis du Cerdi, notamment Pierre Mandon, Laurène Petitfour, Sékou Keita, Charlotte Dupuy et Camille Laville, pour leur soutien et leurs conseils. Une pensée particulière va également à mes camarades de thèse : Pierre Beaucoral, Paul Vernus, Anouck Daubrée, Habibou Kassoum, Julie Bompas, Melchior Clerc et Rachid Pafadnam, Mariz Kalliny avec qui j'ai partagé des moments précieux.

Enfin, j'exprime toute ma reconnaissance aux enseignants et personnels du Cerdi, en particulier au Professeur Simone Bertoli, directeur du Cerdi, à Samuel Guerineau, directeur de l'École d'Économie, ainsi qu'à Johan Guiot et Chantal Brige, pour leur aide constante et leur disponibilité tout au long de ce parcours.

Résumé

Cette thèse de doctorat analyse les impacts économiques et sociaux de l'électrification sur le continent africain, en mettant l'accent sur les infrastructures décentralisées et centralisées. Elle propose également de nouveaux outils analytiques pour mieux mesurer et évaluer l'accès à l'électricité, en particulier dans les contextes marqués par une faible disponibilité des données. Le travail repose sur trois essais, chacun portant sur une dimension clé de l'électrification : les mini-réseaux, les plateformes énergétiques multiservices, et les grands projets d'infrastructure électrique.

Le premier essai évalue l'impact des mini-réseaux sur les communautés rurales à l'aide de méthodes quasi-expérimentales. Les mini-réseaux améliorent significativement l'accès à l'électricité et l'activité économique dans les zones hors réseau, mais les résultats varient en raison de fréquentes pannes techniques, soulignant l'importance d'une conception robuste, d'un bon entretien et d'une viabilité financière.

Le deuxième essai présente les résultats d'un essai randomisé contrôlé sur les plateformes multiservices « Café Lumière » à Madagascar. L'adoption des services et les effets économiques sont positifs, mais les impacts sociaux sont limités, sauf pour l'amélioration des services de santé. Ces plateformes favorisent l'usage productif de l'énergie, mais leur effet à plus large échelle dépend de services publics complémentaires.

Le troisième essai utilise l'apprentissage automatique et la télédétection pour estimer l'accès à l'électricité en Afrique à haute résolution spatiale. Un modèle de forêts aléatoires, basé sur des enquêtes géolocalisées et des données spatiales, permet d'évaluer des projets d'électrification financés par la Chine et la Banque mondiale. Les résultats montrent des effets positifs et croissants dans les zones traitées, soulignant l'intérêt des données géospatiales pour le suivi des infrastructures.

Ces trois études montrent que les projets peuvent générer des impacts hétérogènes. La thèse plaide pour des stratégies d'électrification contextuelles, combinant mini-réseaux, extensions du réseau national et plateformes multiservices. Elle souligne également l'importance d'une transparence accrue des données, en particulier la géolocalisation des projets, pour permettre une évaluation rigoureuse de leurs effets. Enfin, la thèse ouvre plusieurs pistes de recherche : amélioration des modèles prédictifs via des données de plus haute résolution, intégration de l'électrification aux objectifs de développement plus larges, et analyses comparatives de porteurs de projets. En mobilisant des expériences de terrain, des méthodes quasi-expérimentales et des outils d'intelligence artificielle, ce travail apporte de nouveaux éléments pour concevoir des politiques énergétiques plus efficaces dans les pays en développement.

Mots clés : Électrification, Mini-réseaux, Accès à l'électricité, Économie du développement, Afrique, Infrastructures, Energie renouvelable, Essai randomisé contrôlé (RCT), Apprentissage automatique, Télédétection, Évaluation d'impact, Évaluation des politiques.

Classification JEL: O13, O18, Q40, Q41, Q48, C21, C23, C33, C55, I38

Summary

This PhD thesis examines the economic and social impacts of electrification in Africa, focusing on both decentralized and centralized energy infrastructures. It also develops new tools to better measure and evaluate electricity access, particularly in data-scarce environments. This thesis is structured around three essays that each address a key dimension of electrification: mini-grids, multiservice energy platforms, and large-scale grid infrastructure projects.

The first essay uses quasi-experimental methods to assess the effects of mini-grids in rural communities. Results show increased electricity access and stimulate economic activity in off-grid areas, though results vary greatly due to frequent technical failures. About half of the projects faced reliability issues, highlighting the need for robust design, proper maintenance, and financial sustainability.

The second essay presents findings from a randomized controlled trial of “Café Lumière” multiservice platforms in Madagascar. The intervention led to high adoption and some economic benefits, but shows limited effects on education, health and safety, aside from improvements in healthcare services. These platforms show potential for productive energy use and further economic development but broader impacts rely on complementary public services.

The third essay applies machine learning and remote sensing to estimate electricity access across Africa at high spatial resolution. A Random Forest model, trained on geo-located survey and spatial data, is used to evaluate electrification projects by China and the World Bank. A difference-in-differences analysis reveals significant, increasing impacts on electricity access rates in treated areas, showing the value of combining ML to geospatial data for infrastructure outcomes evaluation.

Together, these essays show that electric infrastructures can yield meaningful benefits, but the scale and speed of these benefits depend on the implementation model, reliability of service, and type of investment. The thesis highlights the need for context-specific electrification strategies that combine multiple energy solutions. It also highlights the importance of improved data transparency, particularly geolocated project information, to support rigorous impact evaluation.

Finally, the research points to future research directions: refining machine learning models with higher-resolution data; integrating electrification with broader development goals; and conducting comparative studies on financing models. By combining field experiments, quasi-experimental methods, and machine learning, this thesis contributes new evidence and tools to inform more effective energy policies in developing countries.

Key words: Electrification, Mini-grids, Energy access, Development economics, Africa, Infrastructure, Renewable energy, Randomized controlled trial (RCT), Machine learning, Remote sensing, Impact evaluation, Policy evaluation.

JEL classification: O13, O18, Q40, Q41, Q48, C21, C23, C33, C55, I38

List of Acronyms

ADB: Asian Development Bank

AfDB: African Development Bank

ANCOVA: Analysis of Covariance

ATT: Average Treatment Effect on the Treated

BRI: Belt and Road Initiative

DiD: Difference-in-Differences

DNN: Deep Neural Networks

ESMAP: Energy Sector Management Assistance Program

IFIs: International Financial Institutions

kNN: k Nearest Neighbors

MDB: Multilateral Development Bank

ML: Machine Learning

NTL: Nighttime Lights

PPP: Public-Private Partnership

RCT: Randomized Controlled Trial

RF: Random Forest

SDG: Sustainable Development Goal

SHS: Solar Home System

SVM: Support Vector Machine

Content

Table des matières

List of tables

Chapter 1: Reaching SDG 7: Shedding a light on the causal effect of mini-grids on rural electrification.

Chapter 2: Experimental impact study on a micro-grid project in rural Madagascar: The Café Lumière project.

Chapter 3: Estimating Electricity Access Impact of Chinese Energy Projects in Africa Using Machine Learning and Remote Sensing Data.

List of figures

Chapter 1: Chapter 1: Reaching SDG 7: Shedding a light on the causal effect of mini-grids on rural electrification.

Chapter 2: Experimental impact study on a micro-grid project in rural Madagascar: The Café Lumière project.

Chapter 3: Estimating Electricity Access Impact of Chinese Energy Projects in Africa Using Machine Learning and Remote Sensing Data.

General introduction

1. Introduction to the Topic and Motivation

1.1 The role of electricity access in economic development

Since the Industrial Revolution, energy, and particularly electricity, has been at the core of economic transformation, enabling industrialization, modernization of agriculture, expansion of services, and improvements in health and education. Despite these well-documented benefits, access to electricity remains unevenly distributed, particularly in low- and middle-income countries, where large portions of the population still lack reliable and affordable power.

Electricity has been seen as a catalyst for economic growth and poverty reduction throughout all the current of developed countries. The relationship between electricity access and economic development operates through multiple channels. At the macroeconomic level, electricity availability is strongly correlated with higher GDP per capita, increased industrial output, and structural transformation (Koščak Kolin et al., 2021). Countries with widespread electricity access tend to have more developed manufacturing and service sectors, contributing to employment generation and economic diversification (Deichmann et al., 2019).

If we focus more at the microeconomic level, access to electricity can improve productivity, facilitate entrepreneurship and business development and foster human capital development. Increased productivity is enabled by mechanization in agriculture, automation in industry, and expansion of digital services, all of which contribute to higher labor productivity and income growth (Fried & Lagakos, 2023). Entrepreneurship and Business Development can be improved as Small and Medium-sized Enterprises (SMEs) benefit from electricity for essential functions such as lighting, refrigeration, and machinery operation. Empirical studies have shown that electrified households and communities experience increased business activity, job creation, and higher incomes (Carabajal et al., 2024; Kooijman-van Dijk, 2012; Moore et al., 2020; Riva et al., 2018). Human capital development can be fostered by extended study hours for students, improves school environments through direct benefits such as lighting, computers, or internet connectivity, but also through indirect benefits such as more attractivity as electrified schools could to attract better teachers and resources. In terms of health and well-being, electrification of healthcare facilities enables 24/24 hours activity, refrigeration of vaccines, sterilization of medical equipment, and the operation of life-saving devices[1].

Persistent Electricity Access Gaps in Developing Countries

Despite its importance, electricity access remains a major challenge in many parts of the world, particularly in Sub-Saharan Africa and South Asia (IEA, 2025). According to the

[1] https://www.who.int/news-room/fact-sheets/detail/electricity-in-health-care-facilities ; https://www.seforall.org/publications/state-of-the-market-report-for-healthcare-facility-electrification

International Energy Agency (IEA), in 2023, approximately 750 million people worldwide lacked access to electricity, with about 600 million of them residing in sub-Saharan Africa. This indicates that sub-Saharan Africa accounts for roughly 80% of the global population without electricity access. The rural-urban divide is stark: while urban electrification rates have improved significantly, rural areas still lag behind due to high infrastructure costs, lower population densities, and logistical challenges[2].

Rural electrification faces key barriers to achieving universal electricity access including: high infrastructure costs, financial constraints, unreliable and intermittent supply, unreliable supply, affordability and willingness to pay issues and data and monitoring challenges. High infrastructure costs are due to the remote and sparsely nature of the population in rural areas. Financial constraints come from Governments and utility companies from developing countries struggling with financial sustainability due to low electricity tariffs, high distribution losses, and insufficient investment. Unreliable or intermittent supply are quite often the norm because of aging production and distribution infrastructures, supply shortages, and poor maintenance. Affordability and willingness to pay issues is also quite common for many low-income households, which cannot afford the upfront connection fees or monthly electricity costs, limiting effective demand for electrification. Finally, data and monitoring challenges are due to the fact that traditional surveys used to track electricity access are infrequent, costly, and often fail to capture granular changes, particularly in rural and peri-urban areas.

1.2 The global policy context

The Relevance of this Research to Sustainable Development Goal 7

Recognizing the transformative power of energy in economic and social progress, the United Nations included universal energy access as a key pillar of its Sustainable Development Goals (SDGs), under SDG 7: Ensure access to affordable, reliable, sustainable, and modern energy for all[3]. This goal reflects a global commitment to ending energy poverty and ensuring that individuals, households, and businesses worldwide can access electricity that is not only available but also reliable and sustainable.

Despite considerable progress in expanding energy access over the past two decades, global efforts remain off track to meet the SDG 7 target by 2030. While countries in South and East Asia have made remarkable progress in electrification, Africa remains the region with the largest access gap, where rural populations continue to rely on traditional biomass, kerosene, or diesel generators for basic energy needs[4].

The challenge is not just about extending electricity infrastructure, but also about ensuring that energy systems are affordable, reliable, and environmentally sustainable. Many of the world's poorest households cannot afford the cost of grid connections, and even where electricity is available, supply is often unreliable, with frequent power outages disrupting

[2] https://www.seforall.org/data-stories/electrification-urbanrural-divide
[3] https://sdgs.un.org/goals/goal7
[4] https://www.worldbank.org/en/news/feature/2024/09/19/five-ways-the-world-bank-will-achieve-mission-300

businesses or healthcare services as discussed previously. Furthermore, as the world moves towards decarbonization, ensuring that electrification efforts rely on renewable energy sources rather than fossil fuels is a growing concern. SDG 7 is therefore closely linked to climate policy, aligning with global efforts such as the Paris Agreement and national commitments to reduce greenhouse gas emissions through clean energy transitions.

The Role of International Development Actors in Financing Energy Infrastructure

Achieving universal electricity access requires not only strong national policies but also significant financial resources. Electrification projects, whether through grid expansion or decentralized systems, require large-scale, long-term investments that often exceed the financial capacity of governments in developing countries. As a result, international development actors, including multilateral institutions, bilateral donors, private investors, and philanthropic organizations, play a crucial role in financing energy infrastructure (Nicolas et al., 2019).

Multilateral Development Banks (MDBs) and International Financial Institutions (IFIs), such as the World Bank, African Development Bank (AfDB), and Asian Development Bank (ADB), have historically been key financiers of electrification projects, providing concessional loans and grants to support grid expansion and rural electrification programs. The World Bank's Energy Sector Management Assistance Program (ESMAP), for example, has financed numerous off-grid and mini-grid initiatives, particularly in Sub-Saharan Africa and South Asia, where the electrification gap remains most pronounced. These institutions not only provide funding but also offer technical assistance, policy guidance, and risk mitigation instruments to attract additional private sector investment. The same kind of support has been made by most of the bilateral aid agencies from countries, which have actively supported energy access programs in low-income countries.

China has also become a major financier of energy infrastructure projects in developing countries, particularly in Africa, through its Belt and Road Initiative (BRI). Chinese energy companies and banks, including the China Development Bank (CDB) and Export-Import Bank of China (Exim Bank), have funded numerous large-scale power generation and transmission projects, often in partnership with national governments (IEA, 2024c). Although many of these projects have historically been fossil-fuel-based, raising concerns about their environmental impact, China has recently signaled a shift toward green energy financing, with increasing investments in hydropower, solar, and wind projects in Africa and other developing regions (IEA, 2024a; UNDP, 2024).

The private sector has also emerged as a critical player in the electrification landscape, particularly in off-grid and renewable energy solutions. Many decentralized electrification projects rely on Public-Private Partnerships (PPPs) or private investments, especially in the Solar Home System (SHS) and mini-grid sectors. Large energy companies and impact investors are increasingly looking at clean energy markets in Africa and Asia as high-growth opportunities, particularly as technological advancements lower the costs of solar, battery storage, and digital payment systems.

Despite the growing role of international development actors, financing gaps remain a major constraint to achieving universal electricity access. The International Energy Agency (IEA) estimates that reaching SDG 7 by 2030 would require an annual investment of approximately $30 billion for off-grid and mini-grid solutions alone, yet current investment levels remain far below this target (IEA, 2024b).

1.3 The challenge of electrification in developing regions

Grid Expansion vs. Off-Grid Solutions: Trade-offs in Electrification Strategies

The challenge of achieving universal electricity access in developing regions is not only a matter of infrastructure deployment but also of economic feasibility, geographic constraints, and policy choices. Broadly, electrification strategies can be now classified into two main approaches: grid extension and off-grid solutions, each presenting unique trade-offs in terms of cost, scalability, and long-term sustainability (ESMAP, 2024).

For many countries, grid extension has long been considered the default approach to electrification, replicating the known model from current developed countries. Large-scale power plants, whether hydroelectric, fossil-fuel based, or increasingly large renewable plants (wind and solar), are connected to transmission and distribution networks that reach homes, businesses, and industries. Extending the grid to urban and peri-urban areas is often the most cost-effective way to provide stable and high-capacity electricity, supporting economic growth, industrialization, and the expansion of essential services to a large number of people. A well-developed national grid facilitates access to diverse energy sources and ensures that electricity can be distributed efficiently across regions with varying demand levels. However, this approach has significant limitations, particularly in rural and remote areas. In many developing countries, extending the grid to sparsely populated regions would be prohibitively expensive. The costs of infrastructure, maintenance, and energy distribution tend to increase exponentially as transmission lines stretch over long distances to serve communities with low and often unpredictable electricity consumption. Many rural electrification projects face low-cost recovery rates, as households struggle to afford connection fees and monthly tariffs, making private sector investments in grid expansion less attractive. Furthermore, even in areas where the grid has been extended, reliability issues persist due to aging infrastructure, insufficient power generation, and poor maintenance, leading to frequent blackouts and voltage instability (Andersen & Dalgaard, 2013; Chen et al., 2023; Ghodeswar et al., 2025).

Faced with these challenges, off-grid solutions, including mini-grids, micro-grids, and stand-alone solar home systems, have gained prominence as an alternative pathway to electrification and are now an integral part of the electrification strategy of the World Bank through its Mission 300 initiative[5]. Technological advancement in renewable energy, such as solar solutions, have enabled a much more scalable approach to electricity access. Rather than waiting for the grid to reach them, many communities are now accessing electricity through decentralized systems that rely on local energy generation, often from renewable sources. Mini-grids, which serve multiple households and businesses within a localized network, offer

[5] https://www.worldbank.org/en/programs/energizing-africa/overview

a scalable solution that can be tailored to the specific energy needs of a village or rural town. Powered by solar, small hydro, or hybrid systems, mini-grids can provide reliable electricity without the need for extensive transmission infrastructure. For even more remote or dispersed populations, solar home systems (SHS) have become a game-changer, offering individual households an affordable and accessible way to generate their own electricity (ESMAP, 2024). These systems, often sold through pay-as-you-go models, allow families to power basic appliances such as lights, radios, and mobile phones, improving quality of life while reducing dependence on polluting and expensive alternatives like kerosene lamps.

However, decentralized solutions also come with trade-offs. While mini-grids and stand-alone solar systems can be deployed more quickly and at lower upfront costs, they often provide limited production capacity, restricting productive electricity use. Many mini-grids operate independently from national grids, which means that as demand grows, they may struggle to scale up effectively. There is also the issue of long-term sustainability, off-grid projects often rely on subsidies, donor funding, or innovative financing models to remain viable, and the ability to cover maintenance and operational costs over time depends on proper sizing and remains uncertain (Duthie et al., 2024). Integrating these systems into national grids in the future poses another challenge, as many mini-grids have been designed as stand-alone networks without considering how they might connect to larger power infrastructures.

Policymakers and investors must therefore navigate a complex technical landscape, balancing the economic, geographic, and technological realities of electrification. In some regions, extending the grid will remain the most viable solution, particularly where industrial or commercial electricity demand is high enough to justify investment and insure sufficient demand and profitability. In others, decentralized systems will provide the fastest and most cost-effective means of reaching rural populations. Increasingly, hybrid approaches are emerging, where mini-grids are integrated into national grids over time, ensuring a flexible and adaptive approach to electrification, or also as a palliative for unreliable grid systems acting as a form of back up (ESMAP, 2024). This shift reflects a growing recognition that achieving universal energy access requires not a one-size-fits-all model, but a strategic combination of grid and off-grid solutions, tailored to the unique needs and constraints of each region.

The Importance of Reliable, Granular Data to Assess Electricity Access and Infrastructure Effectiveness

One of the most significant barriers to effective electrification planning in developing regions is the lack of reliable, high-resolution data on electricity access and infrastructure performance. Many governments and energy planners rely on national-level statistics and household surveys to track electrification progress, but these sources have several limitations such as low frequency, costs and potentially a lack of spatial granularity. The cost and logistical challenges of conducting large-scale surveys limit their spatial and temporal coverage, making it difficult to capture short-term variations or recent electrification efforts. Also, the national-level electrification rate masks significant disparities between urban and rural areas[6].

Remote sensing, and Machine Learning (ML) approaches, can help alleviate some of those constraints. Satellite-based nighttime lights (NTL) data have for instance emerged as a powerful alternative for tracking electricity access, providing near real-time, high-resolution insights into electrification patterns. Also, machine learning models can integrate multi-source geospatial data (i.e., population density, land use, road networks, etc.) to predict electricity access at fine spatial scales. This approach enhances impact evaluation by allowing policymakers to assess whether electrification projects lead to actual improvements in access and economic activity.

Governments and development organizations need granular, timely, and high-quality data to prioritize investments and evaluate project effectiveness. Using remote sensing and machine learning can help address data gaps and provide continuous monitoring of electricity access trends. By integrating remote sensing and geospatial analytics into energy planning, policymakers can make evidence-based decisions that maximize the impact of electrification initiatives, ensuring that investments lead to tangible development outcomes.

2. Research Questions and Objectives of the Thesis

This thesis aims to contribute directly to the SDG 7 objective by exploring the impact of different electrification pathways, including mini-grids, national grid expansion, and decentralized energy solutions, on energy access, economic development, and sustainability. By leveraging remote sensing, it also addresses one of the critical data challenges in tracking progress toward SDG 7, offering new methodologies to generate high-resolution, real-time estimates of electricity access. This thesis also engages with financing challenges in two folds. Firstly, by assessing the impact, or lack thereof, of different types of energy projects on electricity access. Secondly on focusing on Chinese-funded and World Bank-financed projects and their cost-effectiveness.

2.1 Key gaps in the literature on the impact of electricity infrastructure

Despite the well-documented benefits of electrification such as in the systematic review by Moore et al. (2020), significant knowledge gaps remain in understanding how different types of electricity infrastructure impact access, economic development, and social well-being. However, rigorous impact evaluations of electrification projects remain scarce, particularly for decentralized energy solutions such as mini-grids and multi-service energy platforms. Most of the literature using a proper impact evaluation approach have focused on projects located in Asia and on SHS solutions with regards to decentralized electrification, leaving mini-grids and Africa mostly out of the scope of the literature on electrification impacts (Berthelemy & Millien, 2018). The literature focusing on the African continent have usually mainly covered grid solutions in countries such as Kenya and South Africa (Dinkelman, 2011; Lee et al., 2016,

[6] https://blogs.worldbank.org/en/opendata/access-universal-and-sustainable-electricity-meeting-challenge

2020b, 2020a). Moreover, tracking electricity access progress at a granular spatial scale, as needed for impact evaluation of decentralized electrification, is often constrained by data limitations, making it difficult to assess where energy investments are most effective.

This thesis aims to fill these gaps by investigating the impact of mini-grids, multi-service energy platforms, and large-scale energy infrastructure projects using a combination of causal inference methods, experimental evidence, and remote sensing-based data analytics in countries where research have been scarce. It seeks to answer three fundamental research questions: How do mini-grids influence electricity access over time? How effective are multi-service energy platforms in improving energy availability and socio-economic outcomes? Can remote sensing and machine learning provide a more accurate and scalable way to track electrification and assess the impact of infrastructure investments? These questions are addressed in three distinct but interconnected essays, each contributing to the broader understanding of electricity access and its development implications.

Evaluating the Causal Effects of Mini-Grids on Electricity Access over Time

Mini-grids have been widely promoted as a cost-effective alternative to grid extension, particularly in remote rural areas where traditional grid infrastructure is financially unfeasible. However, the evidence on their actual impact remains limited, partly due to challenges in measuring access at the local level and isolating causal effects from broader socio-economic factors. Many existing studies rely on descriptive analyses or case studies as documented in Berthelemy & Millien (2018), which, while informative, fail to establish robust counterfactuals to assess what would have happened in the absence of mini-grid interventions. Most studies show moderate impacts in the short-term (Aklin et al., 2017; Duthie et al., 2024; Kirubi et al., 2009; Peters et al., 2019). Those moderate impacts might however be explained partially by high failure rates, as documented in the literature outside of the African continent (Berthelemy & Maurel, 2021; Duran & Sahinyazan, 2021).

This thesis contributes to the literature by employing a quasi-experimental approach to measure the causal impact and dynamic of mini-grids on electricity access, leveraging satellite-based NTL data as a proxy for electricity access. The use of NTL is based on the extensive literature studying link between electricity access or production and NTL measures (Elvidge et al., 2020; Falchetta et al., 2019; Min et al., 2024). By comparing areas treated by mini-grid projects to control areas that remain unelectrified, it provides causal estimates of how mini-grid deployment influences electricity access over time. In doing so, it offers insights into the dynamic of the effects, and reliability of mini-grid solutions, helping policymakers determine whether decentralized electrification can serve as a long-term substitute or merely a provisional measure before eventual grid integration.

Assessing the Effectiveness of Multi-Service Energy Platforms Using Experimental Evidence

While mini-grids primarily provide electricity, more integrated and multi-functional energy platforms, which offer additional services such as refrigeration, charging stations or rental of electrical equipment, have emerged as a promising approach to expand energy access while simultaneously addressing other development needs. Unlike traditional electrification efforts that focus solely on energy provision, these platforms aim to create synergies between electricity access, productive uses and socio-economic impacts. This approach fits into the Multi-Tier Framework (MTF) approach developed by the Energy Sector Management Assistance Program (ESMAP) from the World-Bank (Bhatia & Angelou, 2015), which is particularly relevant in a rural context where electricity needs and potential to pay are more limited. The MTF proposes a scale on the level of electricity access based on the uses a household can make with electricity, reflecting a more qualitative assessment rather than a binary outcome[7].

Despite their potential, empirical evidence on the effectiveness of multi-service energy platforms remains scarce, and little is known about whether they generate meaningful economic benefits or improve overall well-being (Duran & Sahinyazan, 2021; Duthie et al., 2024). This thesis addresses this gap by conducting a randomized controlled trial (RCT), based on two survey waves, on a multi-service energy platform initiative implemented in Madagascar. By randomly assigning villages to treatment and control groups, this study isolates the direct effects of access to multi-service energy platforms on universal access to basic electric services, improved social services and improved economic development. The findings from this study will contribute to a growing debate on how to design electrification initiatives that move beyond mere access to electricity and instead aims at fostering a broader and sustainable development and what type of impacts we might expect from such energy solutions platforms.

Leveraging Remote Sensing and Machine Learning to Generate High-Resolution Electricity Access Data and Evaluate Large-Scale Energy Projects

One of the biggest challenges in evaluating the impact of electrification projects is the lack of reliable, granular and frequently updated data on electricity access. Traditional household surveys, while valuable, are conducted infrequently and at different locations, making it difficult to track changes over time and assess the effectiveness of different electrification strategies. As a result, policymakers and researchers often rely on national-level statistics that fail to capture localized disparities in electricity access.

To address this problem, this thesis explores how remote sensing combined with machine learning can be used to generate high-resolution electricity access data, providing a scalable and cost-effective alternative to traditional surveys. Based on the latest advances on the application of machine learning to remote sensing data (Dhorne et al., 2021; Ratledge et al., 2022; Yeh et al., 2020), this chapter builds a machine learning model trained on geo-referenced household survey data, nighttime lights, and other geospatial indicators to predict electricity access across Africa at a fine spatial scale. This approach enables the creation of annual electricity access maps, which can be used to identify underserved areas, monitor electrification progress, and evaluate the impact of energy infrastructure projects. Beyond this

[7] https://www.esmap.org/mtf_multi-tier_framework_for_energy_access

data generation exercise, this study applies the predicted electricity access data to assess the impact of large-scale electrification investments, particularly those financed by China and the World Bank. By employing difference-in-differences and matching, it estimates the causal effect of energy infrastructure projects on electricity access rates and examines whether different funding models (e.g., Chinese vs. multilateral financing) yield different outcomes. This provides empirical evidence on the effectiveness of large-scale energy investments, helping policymakers and development agencies optimize their strategies for achieving Sustainable Development Goal 7.

We can summarize the thesis around these three main research questions:

- What is the impact of mini-grids on rural electrification and economic development, and how can nighttime lights data improve impact evaluation?
- How effective are multi-service energy platforms in improving electricity access and economic activity in remote areas?
- Can machine learning and remote sensing provide reliable, high-resolution data for tracking electricity access and evaluating large-scale infrastructure projects?

2.2 The central role of remote sensing in electricity access evaluation

As seen previously a major challenge in the study and implementation of electrification policies is the lack of granular and frequently updated data on electricity access. Traditional data sources, such as household surveys and national statistics, are often conducted infrequently and at low spatial resolution, making it difficult to track progress, evaluate interventions, and identify underserved areas.

To overcome these limitations, satellite imagery and geospatial datasets have emerged as powerful tools for tracking electrification progress (He et al., 2012; Min et al., 2024), especially in regions where ground-based surveys are scarce or costly. By leveraging remotely sensed indicators, such as NTL, land cover, road networks, or population density, it is possible to generate spatially continuous, high-frequency estimates of electricity access. These methods provide a crucial complement to traditional surveys, allowing for near real-time monitoring and impact assessment of electrification projects across vast and remote areas.

This thesis explores how the use of NTL, in particular when combined with other remotely sensed data can be used as a proxy for electricity access, providing a scalable and cost-effective approach to measuring electrification at a granular level. To address in part some limitations of using only remotely sensed data, the third chapter develops a machine learning model trained on geo-referenced survey data used as ground truth data for calibration. By integrating multiple data sources, this approach enables the generation of high-resolution electricity access maps, offering a more accurate and dynamic tool for tracking electrification efforts and evaluating the impact of energy infrastructure investments. This innovative use of remote sensing and machine learning represents a significant methodological advancement in energy access research.

3. Contribution of the Thesis

3.1 Empirical Contributions

This thesis proposes empirical contributions by employing rigorous impact evaluation methods to assess the effects of electric infrastructure projects. One of the key challenges in electrification research is establishing causal relationships between energy infrastructure investments and socio-economic outcomes. To address this, the thesis utilizes a combination of quasi-experimental and experimental approaches, respectively difference-in-differences (DiD) combined with matching techniques and RCTs. By applying these rigorous experimental frameworks, it provides credible and policy-relevant evidence on the effectiveness of mini-grids and multi-service energy platforms in improving electricity access and fostering economic development and how they might affect economic activities, energy consumption patterns, and household well-being.

Beyond traditional evaluation techniques, this research also advances the use of remote sensing and machine learning to generate high-frequency, high-resolution electricity access data across Africa. While the first chapter shows how combining remote sensing sources can be used to build quasi-experimental samples to conduct causal inference, the third chapter goes beyond this approach by combining remote-sensing with machine learning to infer ground-truth data. Satellite-based nighttime lights (NTL) data, combined with geospatial indicators, with or without machine learning, offer a cost-effective alternative to household surveys, allowing for more continuous and scalable tracking of electrification progress. By integrating these datasets into predictive machine learning models, this thesis not only refines electricity access measurement but also demonstrates how data-driven approaches can enhance infrastructure planning and impact evaluation at a continental scale.

3.2 Methodological Innovations

This thesis introduces several methodological innovations in the study of electricity access and impact evaluation.

First, it advances the use of satellite-based NTL data as a proxy for electrification combined with land-use data to measure causal effects at the village level, assessing its reliability and limitations in measuring electricity access, particularly in rural settings and applied to decentralized electrification solutions. While NTL has been widely used in economic research, this study provides a more refined and localized application, while also integrating project characteristics to study effects heterogeneity (Chapter 1).

Second, this thesis applies a clustered RCT to evaluate the impact of multi-service energy platforms on electricity access and socio-economic outcomes in rural Madagascar. Unlike traditional household-level randomization, this study adopts a clustered design, where entire communities or villages are randomly assigned to treatment and control groups. This approach is particularly well-suited for decentralized electrification projects, as electricity access can generate spillover effects beyond individual households, influencing local businesses, social interactions, and community-wide development.

Third, this research develops a machine learning model that combines remote sensing, geospatial, and survey-based data to predict electricity access at a high spatial resolution (Chapter 3). This approach moves beyond traditional binary electrification metrics by providing continuous, data-driven estimations of energy access, capturing both on-grid and off-grid solutions. The integration of multiple data sources improves granularity and temporal consistency, offering a scalable alternative to infrequent household surveys.

Finally, this thesis demonstrates applications of remotely sensed data to enhance infrastructure impact evaluations, for electrification projects implemented at different scales where survey-based approaches and RCTs design are either unavailable or too costly to implement. By applying quasi-experimental methods to satellite-derived electricity access metrics, this research provides a novel framework for assessing the effects of grid expansion and decentralized electrification projects.

3.3 Policy Relevance

This thesis has important policy implications, particularly in the context of Sustainable Development Goal 7 (SDG 7), which aims to achieve universal access to affordable, reliable, sustainable, and modern energy. By introducing new methodologies for tracking electricity access using remote sensing and machine learning, this research provides policymakers with scalable and cost-effective tools to monitor electrification progress in real-time. Traditional survey-based tracking methods are often too costly and impractical to implement, while satellite-based approaches allow for continuous, high-resolution assessments that can better inform energy planning and investment strategies.

Beyond measurement, this research generates actionable insights for policymakers by evaluating the effectiveness of different electrification approaches, including mini-grids, multi-service energy platforms, and large-scale grid infrastructure investments. The findings provide evidence-based guidance on what policymakers can expect in terms of impacts and potentially quantify realistic expectations by deploying different electrification solutions. Additionally, by analyzing Chinese-financed and World Bank-funded energy infrastructure projects, this thesis sheds light on how different financing models influence electricity access outcomes, offering valuable insights for governments, donors, and multilateral institutions engaged in energy sector investments.

Ultimately, this research contributes to the development of data-driven, evidence-based electrification policies, ensuring that investments in electricity infrastructure are well targeted and efficient in delivering expected impacts. By leveraging the power of remote sensing, causal inference, and experimental methods, this thesis helps bridge the gap between research and policymaking, providing decision-makers with the tools and evidence needed to accelerate progress toward universal energy access.

4. Overview of the Three Essays and Outline

- **Chapter 1: Causal Effects of Mini-Grids in Burkina Faso and Madagascar**

The first essay investigates the impact of mini-grids on rural electrification in Burkina Faso and Madagascar, using a quasi-experimental approach. Mini-grids are increasingly seen as a viable alternative to grid expansion, particularly in remote areas where traditional electrification is costly and logistically challenging. Based on a novel dataset of geolocated mini-grids, this study leverages satellite-based nighttime lights (NTL) data as a proxy for electricity access and, combined with other remote sensing sources and matching, applies a difference-in-differences (DiD) methodology to estimate the causal effects of mini-grid deployment on electrification levels and broader development indicators. By comparing mini-grid sites to similar but unelectrified control areas, the study provides empirical evidence on whether mini-grids effectively increase electricity access, support local economic activities, and improve household well-being. The findings contribute to the debate on whether decentralized electrification solutions can serve as a sustainable and scalable model for energy access in developing countries.

- **Chapter 2: Experimental Evidence from Multi-Service Energy Platforms in Madagascar**

The second essay evaluates the effectiveness of multi-service energy platforms in improving rural energy access and fostering local development. Unlike conventional electrification projects that focus solely on providing electricity, these platforms integrate additional services, aiming to create a more productive and development-oriented energy ecosystem. To assess their impact, this study employs a clustered randomized controlled trial (RCT), randomly assigning villages to treatment and control groups. The analysis examines the effects of access to multi-service energy platforms on household electricity consumption, business creation, agricultural productivity, and social welfare. By using a rigorous experimental framework, this chapter provides causal evidence on how decentralized energy solutions can extend beyond basic electrification to drive broader economic and social improvements.

- **Chapter 3: Machine Learning and Remote Sensing for Electricity Access and Infrastructure Evaluation**

The third essay focuses on the measurement challenge in electricity access research, addressing the limitations of traditional survey-based approaches. It develops a machine learning model trained on geo-referenced survey data, remote sensing indicators, and other geospatial features to predict electricity access at a high spatial resolution across Africa. This method allows for the generation of continuous, yearly electrification estimates, offering an alternative to costly and infrequent household surveys. Beyond improving measurement, this study applies the predicted electricity access data to evaluate energy infrastructure projects financed by different actors, namely Chinese investments and World Bank-funded electrification programs. Using quasi-experimental methods such as difference-in-differences, the study assesses the effectiveness of these infrastructure projects in expanding electricity access. By integrating remote sensing, machine learning, and impact evaluation techniques, this chapter advances new data-driven methodologies for monitoring electrification progress and evaluating energy sector investments.

- **Outline**

This thesis is structured into five main parts, beginning with a general introduction that outlines the motivation, research questions, and key contributions of the study. Following this, the three core essays each explore a distinct but interconnected aspect of electricity access and infrastructure evaluation through chapters 1, 2 and 3. The first essay employs a quasi-experimental approach to assess the impact of mini-grids on rural electrification and socio-economic outcomes. The second essay takes a randomized controlled trial (RCT) approach to evaluate the effectiveness of multi-service energy platforms in improving energy access and fostering economic development. The third essay shifts focus to methodological advancements, developing a machine learning model trained on remote sensing and survey data to generate high-resolution electricity access estimates and apply them in the evaluation of large-scale energy projects. Finally, the thesis concludes with a synthesis of the findings, discussing their broader implications for energy policy, electrification strategies, and future research directions in development economics.

References

Aklin, M., Bayer, P., Harish, S. P., & Urpelainen, J. (2017). Does basic energy access generate socioeconomic benefits? A field experiment with off-grid solar power in India. *Science Advances*, *3*(5), e1602153. https://doi.org/10.1126/sciadv.1602153

Andersen, T. B., & Dalgaard, C.-J. (2013). Power outages and economic growth in Africa. *Energy Economics*, *38*, 19-23. https://doi.org/10.1016/j.eneco.2013.02.016

Berthelemy, J.-C., & Maurel, M. (2021). A new approach for evaluation of the economic impact of decentralized. *Ferdi*, *WP284*. https://ferdi.fr/publications/a-new-approach-for-evaluation-of-the-economic-impact-of-decentralized-electrification-projects

Berthelemy, J.-C., & Millien, A. (2018). *Impact of Decentralized Electrification Projects on Sustainable Development : A Meta-Analysis*. https://halshs.archives-ouvertes.fr/halshs-01965653

Bhatia, M., & Angelou, N. (2015). Beyond Connections : Energy Access Redefined. *ESMAP Technical Report*, *008/15*. https://openknowledge.worldbank.org/entities/publication/a896ab51-e042-5b7d-8ffd-59d36461059e

Carabajal, A. T., Orsot, A., Moudio, M. P. E., Haggai, T., Okonkwo, C. J., III, G. T. J., & Selby, N. S. (2024). Social and Economic Impact Analysis of Solar Mini-Grids in Rural Africa : A Cohort Study from Kenya and Nigeria. *Environmental Research: Infrastructure and Sustainability*, *4*(2), 025005. https://doi.org/10.1088/2634-4505/ad4ffb

Chen, H., Jin, L., Wang, M., Guo, L., & Wu, J. (2023). How will power outages affect the national economic growth : Evidence from 152 countries. *Energy Economics*, *126*, 107055. https://doi.org/10.1016/j.eneco.2023.107055

Deichmann, U., Reuter, A., Vollmer, S., & Zhang, F. (2019). The relationship between energy intensity and economic growth : New evidence from a multi-country multi-sectorial dataset. *World Development*, *124*, 104664. https://doi.org/10.1016/j.worlddev.2019.104664

Dhorne, M., Nicolas, C., Arderne, C., & Besnard, J. (2021). *Tracking Advances in Access to Electricity Using Satellite-Based Data and Machine Learning to Complement Surveys* [Brief]. World Bank. https://openknowledge.worldbank.org/handle/10986/35473

Dinkelman, T. (2011). The Effects of Rural Electrification on Employment : New Evidence from South Africa. *American Economic Review*, *101*(7), 3078-3108. https://doi.org/10.1257/aer.101.7.3078

Duran, A. S., & Sahinyazan, F. G. (2021). An analysis of renewable mini-grid projects for rural electrification. *Socio-Economic Planning Sciences*, *77*, 100999. https://doi.org/10.1016/j.seps.2020.100999

Duthie, M., Ankel-Peters, J., Mphasa, C., & Bhat, R. (2024). The elusive quest for sustainable mini-grid electrification : New evidence from Indonesia. *Energy for Sustainable Development*, *80*, 101454. https://doi.org/10.1016/j.esd.2024.101454

Elvidge, C. D., Ghosh, T., Hsu, F.-C., Zhizhin, M., & Bazilian, M. (2020). The Dimming of Lights in China during the COVID-19 Pandemic. *Remote Sensing*, *12*(17), Article 17. https://doi.org/10.3390/rs12172851

ESMAP. (2024). *Off-Grid Solar Market Trends Report 2024 : Outlook*. Energy Sector Management Assistance Program (ESMAP), World Bank Group. https://www.esmap.org/Off-Grid_Solar_Market_Trends_Report_2024

Falchetta, G., Pachauri, S., Parkinson, S., & Byers, E. (2019). A high-resolution gridded dataset to assess electrification in sub-Saharan Africa. *Scientific Data*, *6*(1), 110. https://doi.org/10.1038/s41597-019-0122-6

Fried, S., & Lagakos, D. (2023). Electricity and Firm Productivity : A General-Equilibrium Approach. *American Economic Journal: Macroeconomics*, *15*(4), 67-103. https://doi.org/10.1257/mac.20210248

Ghodeswar, A., Bhandari, M., & Hedman, B. (2025). Quantifying the economic costs of power outages owing to extreme events : A systematic review. *Renewable and Sustainable Energy Reviews*, *207*, 114984. https://doi.org/10.1016/j.rser.2024.114984

He, C., Ma, Q., Li, T., Yang, Y., & Liu, Z. (2012). Spatiotemporal dynamics of electric power consumption in Chinese Mainland from 1995 to 2008 modeled using DMSP/OLS stable nighttime lights data. *Journal of Geographical Sciences*, *22*(1), 125-136. https://doi.org/10.1007/s11442-012-0916-3

IEA. (2024a). *Clean Energy Market Monitor – March 2024*. IEA. https://www.iea.org/reports/clean-energy-market-monitor-march-2024

IEA. (2024b). *Tracking SDG7 : The Energy Progress Report, 2024 – Analysis*. IEA, IRENA, UNSD, World Bank, WHO. https://www.iea.org/reports/tracking-sdg7-the-energy-progress-report-2024

IEA. (2024c, décembre 18). *China's evolving footprint in global energy development finance – Analysis*. IEA. https://www.iea.org/commentaries/china-s-evolving-footprint-in-global-energy-development-finance

IEA. (2025). *Electricity 2025*. IEA. https://www.iea.org/reports/electricity-2025

Kirubi, C., Jacobson, A., Kammen, D. M., & Mills, A. (2009). Community-Based Electric Micro-Grids Can Contribute to Rural Development : Evidence from Kenya. *World Development*, *37*(7), 1208-1221. https://doi.org/10.1016/j.worlddev.2008.11.005

Kooijman-van Dijk, A. L. (2012). The role of energy in creating opportunities for income generation in the Indian Himalayas. *Energy Policy*, *41*, 529-536. https://doi.org/10.1016/j.enpol.2011.11.013

Koščak Kolin, S., Karasalihović Sedlar, D., & Kurevija, T. (2021). Relationship between electricity and economic growth for long-term periods : New possibilities for energy prediction. *Energy*, *228*, 120539. https://doi.org/10.1016/j.energy.2021.120539

Lee, K., Miguel, E., & Wolfram, C. (2016). Appliance Ownership and Aspirations among Electric Grid and Home Solar Households in Rural Kenya. *American Economic Review*, *106*(5), 89-94. https://doi.org/10.1257/aer.p20161097

Lee, K., Miguel, E., & Wolfram, C. (2020a). Does Household Electrification Supercharge Economic Development? *The Journal of Economic Perspectives*, *34*(1), 122-144.

Lee, K., Miguel, E., & Wolfram, C. (2020b). Experimental Evidence on the Economics of Rural Electrification. *Journal of Political Economy*, *128*(4), 1523-1565. https://doi.org/10.1086/705417

Min, B., O'Keeffe, Z. P., Abidoye, B., Gaba, K. M., Monroe, T., Stewart, B. P., Baugh, K., & Sánchez-Andrade Nuño, B. (2024). Lost in the dark : A survey of energy poverty from space. *Joule*. https://doi.org/10.1016/j.joule.2024.05.001

Moore, N., Glandon, D., Tripney, J., Kozakiewicz, T., Shisler, S., Eyres, J., Zalfou, R., Leon, M., Kurkijan, V., Snilstveit, B., & Perdana, A. (2020). *Systematic Review on the Impact of Access to Electricity on*

Household Welfare [ADB Independent Evaluation]. ADB. https://www.adb.org/documents/systematic-review-impact-access-electricity-household-welfare

Nicolas, C., Samson, B., & Rozenberg, J. (2019). Meeting the Sustainable Development Goal for Electricity Access—Using a Multi-Scenario Approach to Understand the Cost Drivers of Power Infrastructure in Sub-Saharan Africa. *World Bank*. https://openknowledge.worldbank.org/server/api/core/bitstreams/7af293a4-243b-51a8-b62e-59c427e3f804/content

Peters, J., Sievert, M., & Toman, M. A. (2019). Rural electrification through mini-grids : Challenges ahead. *Energy Policy*, *132*, 27-31. https://doi.org/10.1016/j.enpol.2019.05.016

Ratledge, N., Cadamuro, G., de la Cuesta, B., Stigler, M., & Burke, M. (2022). Using machine learning to assess the livelihood impact of electricity access. *Nature*, *611*(7936), Article 7936. https://doi.org/10.1038/s41586-022-05322-8

Riva, F., Ahlborg, H., Hartvigsson, E., Pachauri, S., & Colombo, E. (2018). Electricity access and rural development : Review of complex socio-economic dynamics and causal diagrams for more appropriate energy modelling. *Energy for Sustainable Development*, *43*, 203-223. https://doi.org/10.1016/j.esd.2018.02.003

UNDP. (2024). *Leveraging China's International Development Cooperation to Drive the Global Transition Towards Renewable Energy*. United Nations Development Program. https://www.undp.org/china/publications/leveraging-chinas-international-development-cooperation-drive-global-transition-towards-renewable-energy

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., & Burke, M. (2020). Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. *Nature Communications*, *11*(1), Article 1. https://doi.org/10.1038/s41467-020-16185-w

Chapter 1:
Reaching SDG 7, Shedding a light on the causal effect of mini-grids on rural electrification.

Vincent Nossek[a*] and Jean-Claude Berthelemy[b]

[a] *Université Clermont Auvergne, CNRS, IRD, CERDI and FERDI*

[b] *Université Paris 1 Panthéon Sorbonne and FERDI*

* Corresponding author contact: vincent.nossek@uca.fr

Published June 2025 in *The Journal of Development Studies* as:
Nossek, V., & Berthelemy, J.-C. (2025). Reaching SDG 7 : Shedding a Light on the Causal Effect of Mini-Grids on Rural Electrification. *The Journal of Development Studies*, 1-26. https://doi.org/10.1080/00220388.2025.2510636

Reaching SDG 7: Shedding a light on the causal effect of mini-grids on rural electrification.

Abstract

Mini-grids are increasingly becoming a popular solution to electrify rural and remote locations to reach SDG 7. However, the impact of such solutions has been sparsely studied, in particular in Africa. Our study examines the impact of 144 mini-grids deployed in Burkina Faso and Madagascar launched between 2015 and 2020. We mobilize geo-spatial and remote sensing data to perform a quasi-experimental study. We utilize matching and a difference-in-differences method with panel data to estimate the impact of mini-grids on the Nighttime Light (NTL), which is considered a good proxy for electrification. Our results show that mini-grids have a causal and increasing positive impact on the NTL which is increasingly positive over the first three years after installation. We then study the heterogeneity of results among different countries and project characteristics, such as production capacity and technical solutions employed, including renewable technologies and grid connectivity. Overall mini-grids seem to be viable solutions to improve electricity access in a rural context, in particular solutions based on solar energy.

Keywords: electricity access; renewable energy; mini-grids; rural electrification; impact evaluation; nighttime lights.

1. Introduction

Access to electricity is widely recognized as a cornerstone of socioeconomic development, with profound implications for improving living standards and boosting economic productivity. The centrality of electricity to development has reflected in Sustainable Development Goal (SDG) 7, which aims to "ensure access to affordable, reliable, sustainable, and modern energy for all" by 2030[1]. However, achieving this goal remains a significant challenge in many parts of the world, particularly in sub-Saharan Africa. Countries such as Burkina Faso and Madagascar exemplify the struggle to expand electricity access where, as of 2022, only about respectively 3.4% and 11% of rural population had access to electricity[2].

The barriers to achieving universal electrification in these regions are numerous and complex. In rural areas, populations are often dispersed across vast and difficult-to-reach terrains, making grid expansion both technically and economically challenging. Furthermore, the existing power grids in these countries are frequently unreliable, suffering from frequent outages and maintenance issues that further exacerbate the energy access gap. Financial constraints compound these problems, as both the limited capacity of consumers to pay for electricity and the chronic underinvestment in energy infrastructure by governments and private entities create significant hurdles to expanding and modernizing the electricity sector.

In light of these challenges, decentralized energy solutions, stand-alone solutions such as solar home systems (SHS) as well as mini-grids, have emerged as a promising alternative to conventional grid-based electrification strategies. Mini-grids, which are small-scale distributed electricity systems, often based on renewable energy, offer a high degree of flexibility in both load capacity and technology. This adaptability has led to a growing interest in mini-grid projects across rural Africa, with significant support from international development agencies and financial institutions, including the World Bank[3].

Despite the increasing reliance on mini-grids, there remains a significant gap in the literature regarding their long-term impact on electricity access and broader development outcomes. Much of the existing research on rural electrification has focused on the impacts of national grid extensions and of SHS, with relatively little attention paid to the specific contributions of mini-grids. Studies on grid extensions, such as those by Moradi and Schmidt (2022) in Burkina Faso, have documented substantial benefits in terms of economic development and human capital improvement. Similarly, research on SHS, which constitutes approximately 70% of the literature on decentralized electrification according to a meta-analysis by Berthelemy and Millien (2018), has shown positive effects on household income, employment, and overall quality of life, while Duthie et al. (Duthie et al., 2024) tend to show more mixed results in

---

1 https://sdgs.un.org/goals/goal7

2 https://data.worldbank.org/indicator/EG.ELC.ACCS.RU.ZS?locations=MG-BF

3 https://www.worldbank.org/en/topic/energy/publication/mini-grids-for-half-a-billion-people ; https://blogs.worldbank.org/energy/putting-africa-path-universal-electricity-access#:~:text=At%20this%20slow%20pace%2C%20263,residents%2C%20have%20access%20to%20electricity

particular due to low financial sustainability. However, these studies do not fully address the unique dynamics and potential of mini-grids, particularly in contexts where the grid extension is not feasible and where SHS may not provide sufficient power load for productive uses.

The high failure rate of mini-grid projects is another critical issue that has been brought to light in the literature. Berthelemy and Maurel (2024) and Duran and Sahinyazan (2021), report that up to 50% of mini-grid projects fail within the first five years, often due to technical breakdowns, inadequate maintenance, or poor financial viability. The literature suggests that these failures are not just isolated incidents but reflect broader challenges in the design, implementation, and management of mini-grid projects. For instance, Akinyele et al. (2018) and Ikejemba et al. (2017) emphasize the importance of proper organizational structures and community engagement in ensuring the long-term success of mini-grid projects. The failure to address these critical factors can lead to a rapid decline in the reliability and operability of mini-grids, ultimately undermining their potential to contribute sustainably to rural electrification and development. Debunking positive impacts that would be only short-lived requires to monitor over time these impacts.

Given the high stakes and the significant knowledge gaps identified in the literature, there is a pressing need for more systematic and rigorous evaluations of mini-grid projects, particularly those that can assess their long-term sustainability and impact on electricity access, conditioning impact on socio-economic development. Traditional evaluation methods, which rely heavily on field surveys, are costly, time-consuming, and logistically challenging in remote areas. They are also hardly implementable on a multi-periods scale, preventing the tracking of impacts over time. This has led to calls for alternative approaches that can provide more efficient and scalable assessments. While Nighttime Lights (NTL) data have been used as a general proxy to measure the impact of aid projects on development at different spatial scales (Bitzer and Gören, 2024; Khomba and Trew, 2022), it has also been used more directly as a proxy of local electricity access. NTL data have been successfully correlated with administrative data on electricity access in rural India (Dugoua et al., 2018) and also in rural villages of Mali and Senegal (Min et al., 2013). More recently, NTL data have been used effectively in various contexts to evaluate the effectiveness of policy initiatives aiming at developing rural electrification, in particular the evaluation of an electrification reform in rural India (Chindarkar and Goyal, 2023). Berthelemy and Maurel (2024) have also used NTL to assess impacts of mini-grids on electricity access in a quasi-experimental setting, with a sample of 48 mini-grids mainly located in Asian countries.

In this paper we focus on a new sample of 144 mini-grids projects implemented since 2015 located in Burkina Faso and Madagascar. We put the emphasis on the dynamics of progresses in electricity access, tracking changes in electrification levels over several years after project commissioning, using NTL from the Visible Infrared Imaging Radiometer Suite (VIIRS) data, as it is the most precise source for measuring electrification in rural and remote areas (Elvidge et al., 2013; Falchetta et al., 2019).

This paper contributes to the literature by documenting the impact of mini-grids using remote sensing data, and in particular the effect dynamics across time. This is also the first paper to document effects with a large sample of mini-grids located in African countries. We contribute to the literature on the potential high failure rates of mini-grids in the African context while also providing heterogeneity analysis around technologies used.

The remainder of the paper is structured as follows. Section 2 presents data and the methodology used to build the sample and the empirical strategy to estimate the impact of mini-grids on NTL. Results are presented in section 3 and followed by discussion and robustness checks in section 4. Finally, section 5 provides conclusions.

2. Data and methods

2.1. Data and sources

Our study covers Madagascar and Burkina Faso for the period 2013-2022. Our evaluation period is defined by the availability of the NTL data from the VIIRS source which begin in 2013. We evaluate mini-grid projects implemented in both countries from 2015 to 2020.

i. Club-ER mini-grid projects data

Data on mini-grid projects are provided by Club-ER, an association regrouping African rural electrification agencies, from 32 countries. Given the variations in the nature of information provided by the members on their respective mini-grids projects, data must be carefully double-checked and we decided at this stage to retain only Madagascar and Burkina Faso as they provided the most complete information on locations and dates of implementation, which are the two main information needed for our study. Location information have been double-checked by cross referencing region, department, county and village name information with latitude and longitude provided. Each mini-grid electrifies only one location and cluster projects (i.e., multiple mini-grids near to each other's) are registered as their specific location of implementation. Since we do not possess information at the household level, we only consider treatment status, and therefore effects, at the village level. With regards to the starting date of the mini-grids they are not always defined at a monthly scale in our database. We therefore decided to retain only the years and define T 0 as the year when the mini-grids are put into service. Our sample consists of 144 mini-grid projects, with 75 located in Burkina Faso[4] and 69 located in Madagascar. We evaluate only projects that were

[4] The dataset provided by Club ER contains 76 mini-grid projects after 2015 but we removed one project due to its specific location. This project is located in a village next to an industrial mine, Youga Industrial Mine, and was clearly electrified before the mini-grid implementation.

launched since 2015, as the NTL VIIRS data series only begin in 2013 and we need at least two periods before treatment to apply our difference in differences approach[5].

As shown in Table 1, most of the mini-grids (65%) were launched after 2018, which will limit highly our conclusion beyond 3 years after treatment. Burkina Faso and Madagascar share the same dynamics in project completion with both having most projects launched in the latest years.

Table 1: Projects by country and year of implementation.

Start Year	Country		Total
	Burkina Faso	Madagascar	
2015	1	5	6
2016	2	8	10
2017	24	4	28
2018	3	3	6
2019	30	22	52
2020	15	27	42
Total	75	69	144

Table 2 provides information on projects characteristics. With regards to mini-grid's capacity, representing power generation capacity of the installation, the information is mostly lacking for Burkina Faso. However, it can be noted that from the few projects with available data, a few are hydroelectric projects with a power output over 1 MW, which bias Burkina Faso's average well above the capacity's measured for Madagascar. Although mini-grids located in Burkina-Faso may have higher production capacity we cannot draw a clear conclusion from our sample due to the high number of missing data.

[5] Only one project in Madagascar and two in Burkina Faso we launched in 2013 and 2014. No projects were found before 2013.

Table 2: Projects characteristics.

Characteristic	Overall, N = 144[1]	Burkina Faso, N = 75[1]	Madagascar, N = 69[1]
Technology			
Diesel	2 (1.4%)	1 (1.3%)	1 (1.4%)
Hybrid Solar PV / Diesel	32 (22%)	17 (23%)	15 (22%)
Hydro	26 (18%)	6 (8.0%)	20 (29%)
Réseau	27 (19%)	27 (36%)	0 (0%)
Solar PV	57 (40%)	24 (32%)	33 (48%)
Capacity (in MW)			
Mean (SD)	0.28 (0.98)	1.02 (1.89)	0.06 (0.09)
[Min-Max]	[0.00, 7.80]	[0.01, 7.80]	[0.00, 0.66]
Unknown	*61*	*56*	*5*
Peak capacity for solar solutions (in MW)			
Mean (SD)	0.07 (0.16)	0.15 (0.30)	0.05 (0.10)
[Min-Max]	[0.00, 1.10]	[0.01, 1.10]	[0.00, 0.66]
Unknown	*84*	*63*	*21*

[1] n (%)

With regards to the technology used to provide power to the mini-grids, 62% of the projects rely on solar PV, either solely or partially like diesel hybrid solutions. However, Burkina Faso and Madagascar both have specificities in terms of technology used. Around 36% of mini-grids set-up in Burkina Faso rely solely on a connection to the national grid without a local production unit. These mini-grids without production units, called COOPELs (Cooperative for electricity), are set up, built and managed by local communities. Madagascar on the other hand cannot rely on its existing grid network, but can use its wide array of rivers running across the country to power hydroelectric production units, which represents 29% of their mini-grid projects.

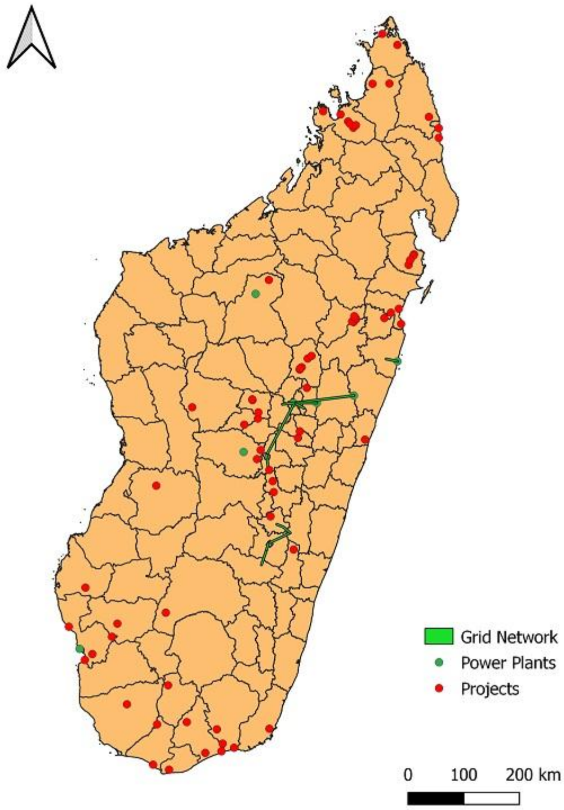
ii. Development projects related to the energy sector and grid infrastructures

To ensure the validity of our approach we aim to consider as much as possible other energy related projects and infrastructures that could impact the control group we aim to build.

To this end we mobilize data provided by AidData, as it is the main provider of geo-localized development projects data. From the AidData databases on the World Bank and China we were able to recover 10 projects related to the energy sector located in Burkina Faso. We delete all potential control villages located within a 3 km radius of those 10 projects. We follow a slightly more conservative approach than Ratledge et al. (Ratledge et al., 2022) that considered villages within a 2 km distance of electric infrastructures as treated.

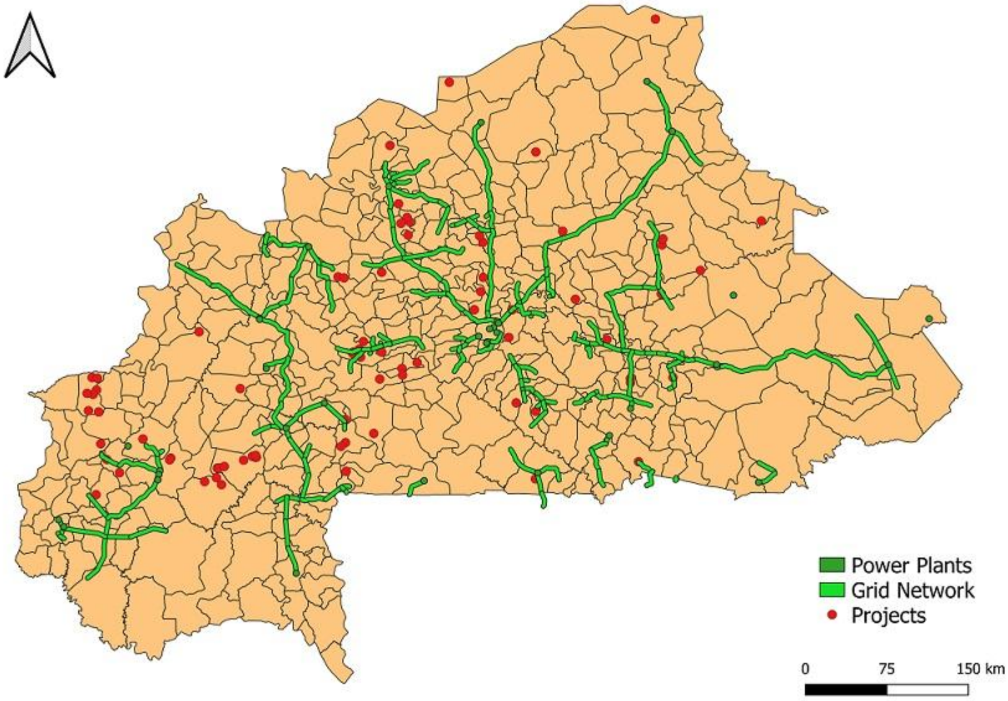
We also identify already, or up-coming, electrified villages with the data provided by the World Bank platform Energydata.info[6]. This platform provides information on the localization of power production units, as well as the distribution network for existing grid and planned grid extensions. According to this source, Burkina Faso has 24 existing production units and 12 planned, while Madagascar has 5 existing production stations and 2 planned. The production unit's localization as well as the grid network is used to remove potentially already electrified villages. For both countries we exclude all villages within a 3km radius of a national power plant unit or within a 2km radius of a distribution line network. As seen in the Figure 1 and 2, the extent of the grid infrastructure is quite different in Burkina Faso and Madagascar, while Burkina Faso's totalizes 3 895 km of distribution lines, Madagascar counts only 483 km.

Figure 1: Madagascar's electricity infrastructure



[6] https://energydata.info/dataset

Figure 2: Burkina Faso's electricity infrastructures



iii. NTL data

The main variable used for evaluation is the Nighttime Lights (NTL) from 2013 to 2022, provided by the Earth Observation Group, Colorado School of Mines. The NTL provides the radiance values in nanoWatts per centimeter squared per steradian (nW/cm²/sr) at a resolution of approximately 500 meters (15 arc second). NTL data can have slightly negative radiance values in very low lighting area such as in rural Africa due to noise and calibration process. We therefore replace the few negative values by 0, as it is common practice in the literature using raw VIIRS data (Beyer et al., 2022; Falchetta and Noussan, 2019)[7]. NTL is used for evaluation post-implementation of projects but is also used before implementation to avoid potentially already electrified villages. NTL is defined at the village level by taking the average of NTL within the polygon delimiting the village extent. We consider NTL as qualitative assessment of electricity access rather than a direct quantitative measure of electricity consumption. NTL is indeed imperfectly correlated to electric power consumption due to the many uses made which do not emit any light measurable by NTL (Gibson et al., 2021; Min et al., 2024). Potential differences in elasticities between electricity consumption and NTL output are context dependent and could lead to under-estimation of the actual electricity consumption since we only measure a subset of the total electric output. Although it is not perfect proxy for consumption, nor it captures fully the multidimensional expected benefits from a mini-grid, NTL remains however correlated with electricity access and consumption, even at the local level (Berthelemy and Maurel, 2024).

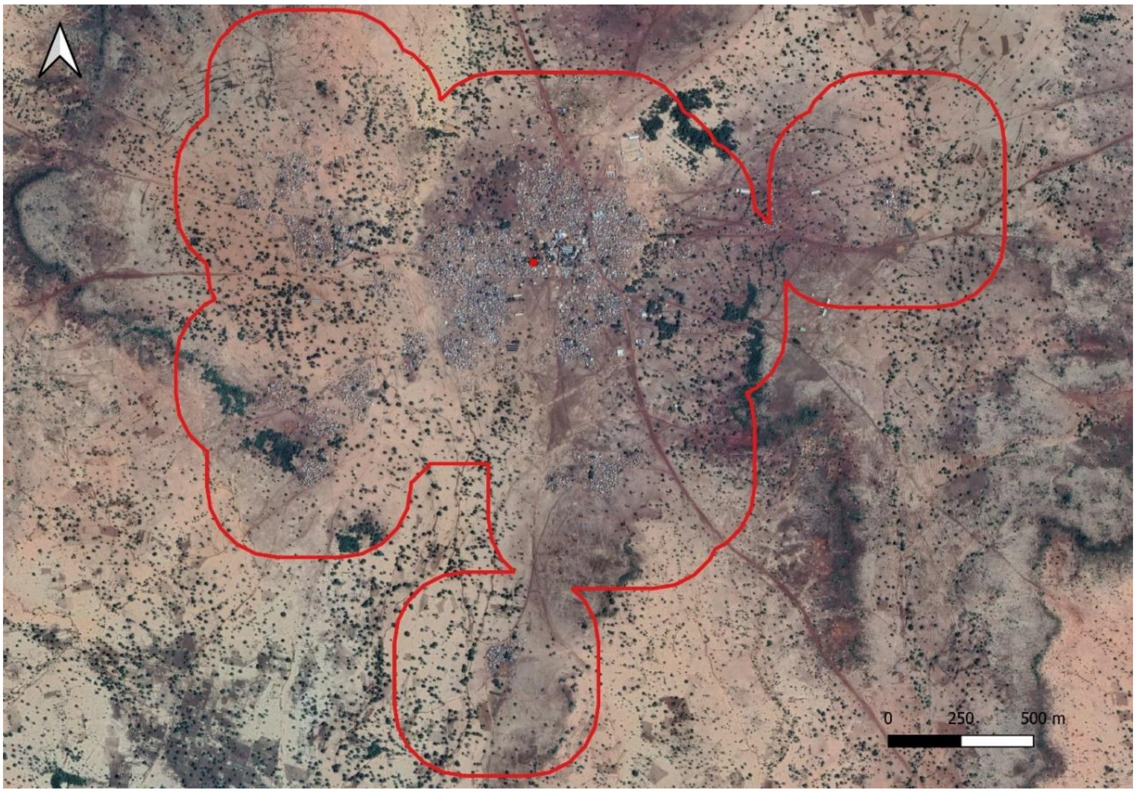
[7] We however provide results in Annex Table A.5 with NTL uncorrected for negative values.

iv. Sample Building

To build our sample of potential control villages we use NTL with a combination of other remote sensing data. First, we make use of the land cover data from the Copernicus Global Land Service database to identify built-up areas with a resolution of 100m x 100m.

We refer to the identified aggregates of built-up areas as villages but our definitions might differ slightly from the villages defined in the national administrative data of each country.[8] Figure 3 illustrates the village of Ziga in Burkina Faso, which received a mini-grid, with its delimitation area drawn in red from the Copernicus built-up layer. This dataset allows us to build a map of all rural villages in which mini-grids have been implemented and identify potential control villages located between a 3km and 50km radius[9] around the treated villages.[10].

Figure 3: Ziga, Yatenga, Burkina Faso, project id: BFAproj4



Note: Point in red represents the coordinates of the projects, the line in red represents the boundary of the treated village area drawn from the Copernicus built-up layer data.

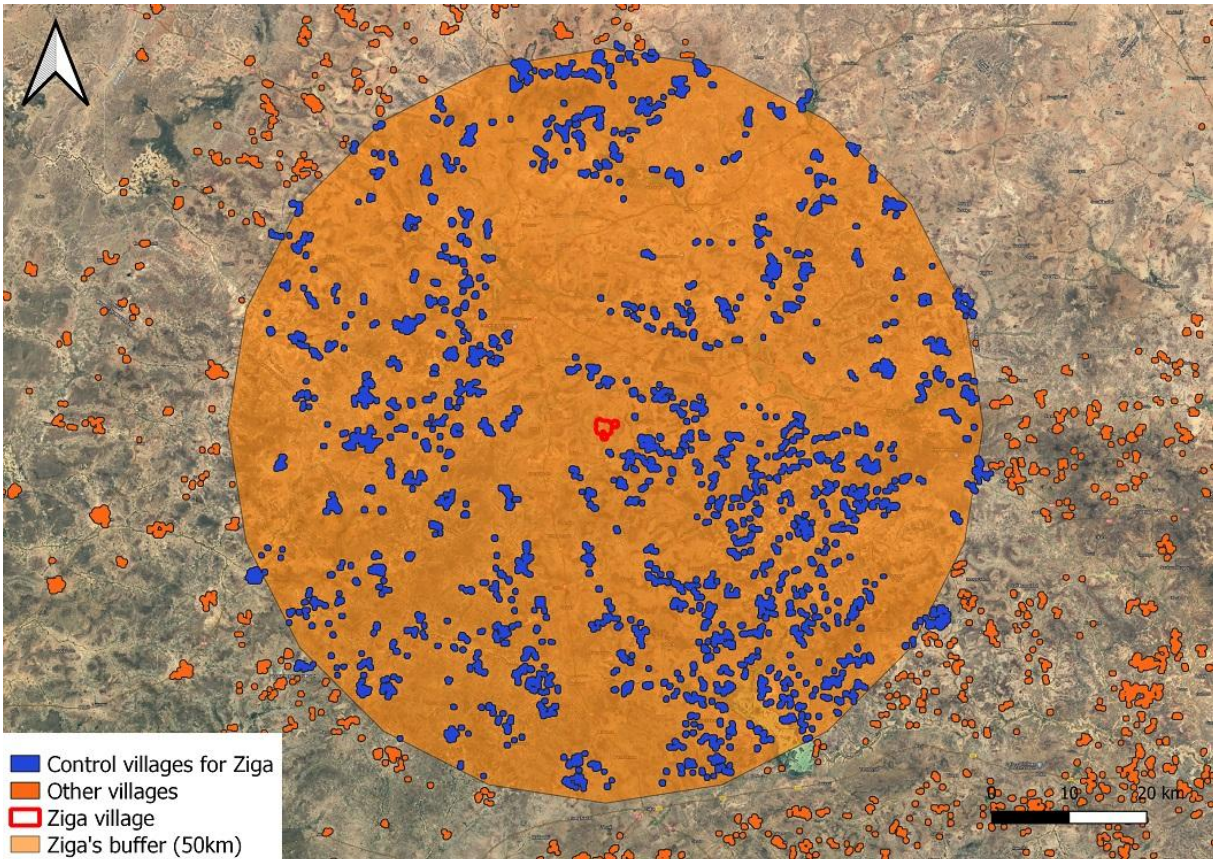
[8] Two adjacent villages could be merged as a single unit in our data. However official geo-localized data on cities and rural villages not being available for Madagascar and Burkina Faso, this is the most precise data to build a map of rural villages, as they reflect true land use rather than administrative boundaries.

[9] Although there is no clear proof of spatial spillover effects from mini-grids projects we decided to take a conservative approach and excluded all villages within a 3km radius around a project.

[10] For Burkina Faso we have on average 459 control villages, with a minimum of 89 and a maximum of 718. For Madagascar we have on average 781 control villages, with a minimum of 29 and a maximum of 1451.

For each country, we exclude all villages smaller than the smallest treated village and those larger than twice the size of the largest treated village nationwide. Figure 4 illustrates one cluster composed of a treated village, in red, with its remaining potential controls in blue after deletion of villages within the 3 km radius of Ziga and other potential electric infrastructures.

Figure 4: Village of Ziga, Yatenga, Burkina Faso, with its potential controls in blue



We then remove villages with a NTL value superior to the highest value of treated villages of each country, which could indicate already electrified villages. We also compute the NTL correlation before project implementation between control villages and their treated counterpart and filter out controls with a correlation inferior to the first decile (0.3) of the correlation distribution. This selection step aims to improve the similarity between the NTL dynamic between treated and controls.

Given the high number of potential controls per treated villages, we restrict or control sample to villages with characteristics similar to treated villages to reduce the risk of selection bias. For that purpose, we resort to matching techniques to select the best fitting controls. To conduct our matching, we rely on the following geocoded variables:

- Average of NTL before project implementation
- Village area computed from the Copernicus Global Land Service;

- Altitude and terrain gradient from SRTM 90m of CGIAR-CSI data;
- Population density from the Gridded Population of the World, Version 4 (GPWv4) 2015 of the NASA's SEDAC;
- Photovoltaic potential from energydata.info, is the long-term yearly average of potential photovoltaic electricity production in kWh/kWp, covering the period 1994-2018;
- Distance to major cities (pop. > 50k.) from Travel time to cities and ports in the year 2015 by Nelson et al. (Nelson et al., 2019).

The main matching techniques present in the literature of impact evaluation revolve around Propensity Score Matching (PSM) and Mahalanobis or Euclidean distance techniques. Recent literature demonstrated that PSM tend to perform very poorly for matching precisely a limited number of controls to treated observations and tools relying on Mahalanobis or Euclidean distance should be favored (King and Nielsen, 2019) . These methods are able to produce a matching more comparable to a Randomized Controlled Trial (RCT) with a fully blocked experimental design (i.e. stratified sampling). We therefore compute a ranking from the most to the least similar control village within the vicinity of each treated villages using the Mahalanobis distance matching method. We also use Euclidean Distance (ED) matching as a robustness check. Retaining only one control per treated might be too restrictive and could give less robust coefficients estimation. To avoid this lack of robustness due to a strong dependency on the sample selected we therefore favor the interpretation of the results with 3 controls, as there is less chance that the first 3 best matching controls for one treated village have all potentially received an unseen development project.

We present in table 3 and 4 the descriptive statistics for Project villages and their respective 3 best matched controls.[11]

Table 3 presents descriptive statistics for Burkina Faso and displays that after matching, there is no significant differences in the variables considered between treated and control villages.

Table 4 provides descriptive statistics for Madagascar after matching. Contrary to Burkina Faso, treated village tend to be slightly larger than control village surface of the villages, however the NTL level are not significantly different between villages. We can also note that treated villages located in Madagascar are in average around 50% smaller than one's located in Burkina Faso.

[11] Table A.1 and A.2 in annex present comparative tables with 1, 3, 5 and 10 best matches.

Table 3: Burkina Faso descriptive statistics after matching (3 best matches per control)

Characteristic	Overall, N = 300	Projects, N = 75	Controls, N = 225	p-value[1]
Surface area (km²)				
Mean (SD)	2.82 (1.67)	2.92 (1.71)	2.79 (1.66)	0.6
[Min-Max]	[0.37, 12.23]	[0.37, 8.70]	[0.37, 12.23]	
Altitude (meters)				
Mean (SD)	311 (38)	309 (39)	311 (37)	0.6
[Min-Max]	[217, 540]	[240, 511]	[217, 540]	
Land gradient				
Mean (SD)	1.62 (0.50)	1.61 (0.52)	1.62 (0.49)	>0.9
[Min-Max]	[0.91, 4.56]	[0.91, 3.76]	[0.92, 4.56]	
Population (hab/km²)				
Mean (SD)	69 (60)	70 (60)	69 (60)	>0.9
[Min-Max]	[8, 363]	[8, 362]	[18, 363]	
Distance to cities (minutes)				
Mean (SD)	68 (47)	69 (49)	68 (46)	>0.9
[Min-Max]	[5, 215]	[5, 215]	[9, 205]	
Potential photovoltaic production (kWh/kWp)				
Mean (SD)	4.58 (0.06)	4.58 (0.06)	4.58 (0.06)	>0.9
[Min-Max]	[4.44, 4.74]	[4.46, 4.73]	[4.44, 4.74]	
NTL pre-project				
Mean (SD)	0.087 (0.055)	0.101 (0.097)	0.082 (0.030)	0.10
[Min-Max]	[0.025, 0.787]	[0.033, 0.787]	[0.025, 0.221]	

[1] Welch Two Sample t-test

Table 4: Madagascar descriptive statistics after matching (3 best matches per control)

Characteristic	Overall, N = 276	Projects, N = 69	Controls, N = 207	p-value[1]
Surface area (km²)				
Mean (SD)	1.40 (1.41)	1.82 (1.73)	1.26 (1.26)	0.015
[Min-Max]	[0.35, 9.26]	[0.35, 9.26]	[0.35, 7.22]	
Altitude (meters)				
Mean (SD)	526 (558)	528 (548)	525 (563)	>0.9
[Min-Max]	[4, 1,943]	[6, 1,845]	[4, 1,943]	
Land gradient				
Mean (SD)	6.2 (4.0)	6.6 (4.6)	6.1 (3.7)	0.4
[Min-Max]	[0.8, 19.5]	[0.9, 19.5]	[0.8, 16.8]	
Population (hab/km²)				
Mean (SD)	57 (46)	60 (48)	55 (46)	0.5
[Min-Max]	[6, 300]	[6, 188]	[6, 300]	
Distance to cities (minutes)				
Mean (SD)	298 (291)	294 (291)	300 (291)	0.9
[Min-Max]	[7, 1,157]	[27, 1,051]	[7, 1,157]	
Potential photovoltaic production (kWh/kWp)				
Mean (SD)	4.70 (0.35)	4.69 (0.37)	4.71 (0.34)	0.7
[Min-Max]	[3.82, 5.26]	[3.82, 5.26]	[3.93, 5.23]	
NTL pre-project				
Mean (SD)	0.071 (0.028)	0.076 (0.029)	0.069 (0.028)	0.12
[Min-Max]	[0.011, 0.298]	[0.011, 0.176]	[0.020, 0.298]	

[1] Welch Two Sample t-test

2.2. Methodology

To estimate the causal treatment effect on villages that received a project we rely on the difference-in-differences (DiD) approach. Given we work with panel data, different treatments timing, and a staggered adoption setting we use an event study DiD design, also called Two Way Fixed Effects (TWFE), and estimate the model defined as follows:

$$NTL_{i,t} = \alpha + \sum_{e=-5}^{-2} \beta_e^{lead} D_{i,t}^e + \sum_{e=0}^{3} \beta_e^{lag} D_{i,t}^e + V_i + Y_t + \varepsilon_{i,t} \qquad (1)$$

Where i is a village, t a time period and e the event study time (i.e. elapsed time since treatment). $NTL_{i,t}$ is the night time light level of village i at time t, $D^e_{i,t}$ a dummy variable denoting the treatment status of village i being e periods after or before treatment at time t. $D^e_{i,t}$ takes the value 1 when e >=0, for instance, $D^e_{i,t}$ will take the value 1 for a village (i=1) that was treated in 2016 at the calendar year 2018 ($D^2_{1,2018}$=1). V_i is a village fixed effect, Y_t a year fixed effect and and $\varepsilon_{i,t}$ the error term. β_e^{lag} represents the Average Treatment on the Treated (ATT) at period e after treatment, which is our coefficient of interest reported in the following regression tables. We are able to estimate up to 3 periods after treatment due to the structure of our sample. Indeed, as seen in section 1, most of the projects are quite recent and we do not have enough time elapsed yet to conclude reliably outside this time window[12]. The effects are computed by taking *e-1* (i.e., the year before project is launched) as the base year to compute the effect. We cluster all standard errors at the treatment level, which is the village level in our case.

To estimate equation (1), we cannot rely on regular TWFE estimates as the recent literature demonstrates that a number of problems can arise when doing so. As shown in de Chaisemartin and D'Haultfoeuille (de Chaisemartin and D'Haultfœuille, 2020) this approach can suffer from inconsistent estimation due to very restrictive hypothesis made on the homogeneity of the effects across time and cohorts[13]. As we have no a priori knowledge on the ATT over time, we prefer to not rely on this hypothesis and use the estimator proposed by de Chaisemartin and D'Haultfoeuille (2024) as our main estimator and also use Callaway and Sant'Anna (2021) estimator for robustness checks.

While the regular TWFE estimated by OLS tries to recover directly (i.e. in one step) a weighted average effect from the all sample at once, de Chaisemartin and D'Haultfoeuille, as well as Callaway and Sant'Anna, first estimate the effect by cohort (g) and time period (t), which allows for heterogeneity of the estimated effect and then aggregate those estimates as an event study. This approach computes multiple Average Treatment on the Treated for each cohort and time denominated ATT(g,t), which is defined as follow:

$$ATT(g,t) = \mathbb{E}\left[NTL_t(g) - NTL_t(0)\middle|G_g = 1\right] \quad (2)$$

Where g is the year of a cohort of projects starting the same year, t is the time period and G_g is a dummy variable that takes the value 1 when the cohort g is treated. Once these ATT(g,t) are estimated for each cohort g and time periods t they are aggregated over a common calendar, relative to the elapsed time since treatment noted e ($e = t - g$). This aggregation recovers the estimates corresponding to an event study type. The aggregation to obtain the ATT of an event study (ATT_{es}) use cohort-dependent weights w(g,t) and follows the form:

$$ATT_{es}(e) = \sum_{g \in G} w(g,t) * ATT(g,t) \quad (3)$$

[12] We provide estimates up to 5 periods in annex table A.4.

[13] They found that more than 40% of these published papers reports biased estimates.

This aggregated parameter $ATT_{es}(e)$ is the estimator of β_e^{lag} found in equation (1). The term w(g,t) is the weight based on the sub-sample size of the cohort g at time t in the overall sample at elapsed time e. Intuitively this means that the larger a cohort is, the larger is its weight in the estimated effect at time e. Standard errors and confidence intervals are computed by using a multiplier bootstrap procedure to conduct asymptotically valid inference[14]. This procedure allows to use information from the full sample at each iteration and accounts for dependence across time and between cohorts.

Given the high failure rate on electrification projects documented in the literature, we also estimate the effects on two subsamples of mini-grids based on their likelihood of failure. To that effect we apply the methodology from Berthelemy and Maurel (2024), which aims at computing the probability of success of mini-grid projects. This method is based on computing the NTL trend of a village before implementation of a project and then measuring the NTL deviation from it during the 3 years after project implementation. We compute an NTL trend for each project village independently. If the average deviation is significantly above the trend trajectory, we then consider that the project has succeeded or failed. Although we could expect this method to be somewhat biased in favor of capacity project Berthelemy and Maurel (2024), found no link between power capacity and success rates, even when using DMSP NTL data, which is much less sensitive to low radiance emissions than VIIRS NTL used in this paper. We then estimate the effect of mini-grids on two subsamples of projects defined as either failed or successful. Those latter estimates aim to provide the proper effect of electrification from well-designed, properly managed and maintained mini-grids, as estimates on the all sample have a high probability of containing failed projects and therefore biasing downward the estimated ATT.

[14] Contrary to a regular bootstrap procedure that redraw subsamples, each iteration of a multiplier bootstrap perturbs by a random weight the influence function of the estimate, which estimates dependency of the results on each observation.

3. Results

3.1. Pooled results

Table 5 : Dynamic effects for all projects

	(1)	(2)	(3)	(4)
VARIABLES	MD1	MD3	MD5	MD10
Post treatment average	0.0269***	0.0280***	0.0286***	0.0288***
	(0.00407)	(0.00379)	(0.00377)	(0.00375)
T 0	0.0130***	0.0132***	0.0137***	0.0138***
	(0.00300)	(0.00296)	(0.00296)	(0.00296)
T+1 year	0.0240***	0.0252***	0.0258***	0.0258***
	(0.00427)	(0.00405)	(0.00403)	(0.00402)
T+2 years	0.0322***	0.0338***	0.0346***	0.0350***
	(0.00529)	(0.00490)	(0.00485)	(0.00482)
T+3 years	0.0431***	0.0445***	0.0452***	0.0455***
	(0.00769)	(0.00727)	(0.00722)	(0.00718)
Observations	2188	4492	6796	12548
R-squared	0.55	0.61	0.63	0.66
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: MD#: Mahalanobis distance; # refers to the number of controls per treated retained.

F-test for equality of the effects on MD3: p-value = .00001817

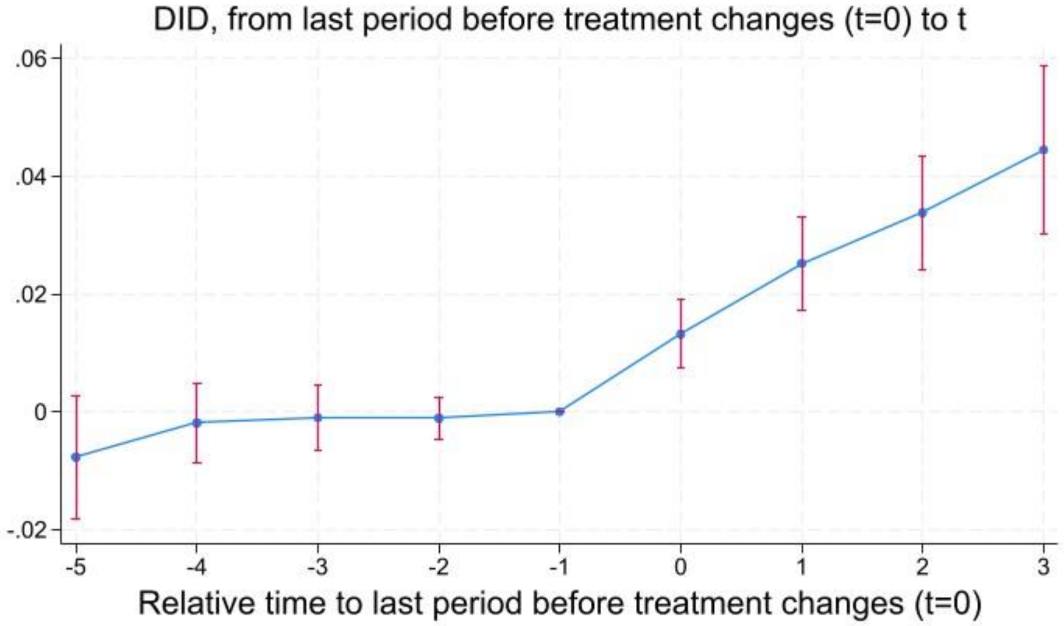
Table 5 shows the results for all the projects, Burkina Faso and Madagascar combined, and the 3 periods after implementation. Figure 5 presents results from up to 3 periods after implementation and all periods before. Although we can compute effects up to six years after implementation, our sample is composed mostly (69%) of projects implemented in 2018 or after, therefore results after 3 years should be taken with caution and might not be very robust due to the very small group of observations[15]. Also results in T 0 year should not really be interpreted as some projects can be launched at the end of a year and therefore their impact can only be fully measured during later periods.

Table 5 compares the results using different control samples, namely control groups composed of either 1, 3, 5 and 10 best matching controls per treated village. The post treatment average coefficients, which provides the average effect over all the periods, are all positive and significant at the 1% level irrespective of the control group used, with an ATT ranging from +0.027 to +0.029 nW/cm^2/sr per year. The post treatment average coefficients remain quite stable with respect to the number of selected controls (i.e., MD1, MD3, MD5 and MD10), with a slight increase in coefficients as the number of controls increases.

If we look into more details at the dynamic effect (i.e., year to year coefficients), we see that all coefficients are positive and significant at the 1% level from the T+1 year to T+3 year, irrespective of the control groups used. If we look at the evolution of the effect over time, we see an increase in NTL impact, between each year after implementation, irrespective of the control groups used. The effect reaches its peak during the third year with an increase of the NTL between 0.043 and 0.046 nW/cm^2/sr compared to the reference year (i.e. T -1). Figure 5 displays the results from the column (2) in table 5 and shows the sharp increase in NTL from T +1 year to T +3 year. Figure 5 also show that no pretends exists prior to implementation of a project and pre-trend tests are available in annex table A.3. In order to confirm the significance of this upward trend we use a joint F-test procedure provided by de Chaisemartin and D'Haultfoeuille, testing for equality of coefficients. Notes under the table 5 provides the F-test results, rejecting the equality hypothesis with a p-value < 1%. Since coefficients from the first 3 years cannot be considered equal, we can state that our results show a clear upward trend of the NTL over time, as displayed in Figure 5.

[15] See Table A4 in Annex for results with 6 periods.

Figure 5: Dynamic ATT on all projects



3.2. Results by country

Table 6: Dynamic effects for Burkina Faso's and Madagascar's projects

	Burkina Faso		Madagascar	
	(1)	(2)	(3)	(4)
VARIABLES	MD3	MD5	MD3	MD5
Post treatment average	0.0295***	0.0305***	0.0257***	0.0255***
	(0.00509)	(0.00505)	(0.00536)	(0.00535)
T 0	0.00865***	0.00905***	0.0191***	0.0190***
	(0.00318)	(0.00320)	(0.00503)	(0.00501)
T+1 year	0.0235***	0.0243***	0.0259***	0.0255***
	(0.00555)	(0.00554)	(0.00586)	(0.00580)
T+2 years	0.0388***	0.0405***	0.0316***	0.0313***
	(0.00633)	(0.00626)	(0.00697)	(0.00694)
T+3 years	0.0513***	0.0525***	0.0268***	0.0269***
	(0.00980)	(0.00971)	(0.00954)	(0.00952)
Observations	2339	3539	2153	3257
R-squared	0.64	0.68	0.65	0.64
Treated units/controls	75/225	75/375	69/207	69/345
Mean of NTL	0.1536	0.1496	0.1154	0.1125

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Notes: MD#: Mahalanobis distance; # refers to the number of controls per treated retained.
F-test for equality of the effects on Burkina Faso MD3: p-value = 3.468e-09;
F-test for equality of the effects on Madagascar MD3: p-value = .14371509.

Table 6 presents the effects for projects located in Burkina Faso's in column (1) and (2), and in Madagascar in column (3) and (4). As seen previously with the all sample, both countries display positive and significant post treatment average effect, although the effect tend to be slightly larger in Burkina Faso. The effect reaches its peak for Burkina Faso in the third year at +0.051 nW/cm^2/sr, while reaching its peak in the second year at +0.032 nW/cm^2/sr for Madgascar. While the NTL increases in Burkina Faso for each period, the increase seems to taper off in the third year for Madagascar’s projects. After three years, NTL level have increased by 51% in Burkina Faso and by 36% in Madagascar compared to pre-average level. While we can conclude to a dynamic growth over time for Burkina Faso projects, we cannot clearly draw similar conclusions for Madagascar's projects, since we are not able to reject equality of the effects between periods.

4. Discussion

The dynamic trend seen in figure 5 can shows an increase of the emitted NTL at the village location, which can be interpreted as an increase in electricity access during the first 3 years. If the measured increase in NTL was the sole effect of public lighting, we would expect to see an increase in the first year followed by a plateau and not a continuous increase over the period as public lighting tends to be quite stable over time. However, the measured increase of the coefficients between periods shows that more NTL is emitted over the years. This increase could be interpreted either as an increase of consumption or as an increase in the number of people using electricity or as some combination of both factors. Unfortunately, the sole NTL data does not allow to discriminate clearly between those driving factors. The fact that the overall magnitude of the effect is similar in both countries tend to show that, although the contexts are different, the expected outcome of the mini-grid projects are quite similar, in particular during the first two years. However, projects located in Burkina Faso sustain an increase of the effect for all periods, while projects in Madagascar seem to plateau after two years. This could indicate a lower performance of projects located in Madagascar over the long term.

4.1 Heterogeneity by likelihood of success

A point regularly raised by field experts and supported by Berthelemy and Maurel (2024) is that the failure rate of mini-grid projects is relatively high. Their estimates show that up to 50% of the projects end up with some types of failure during the first few years, usually linked to a breakdown or lack of maintenance of either the electric production unit or the distribution network. The high number of failing projects can bias downward the overall impact measured by the NTL and by removing potentially failed projects we can provide a new estimate without this underestimation.

To this end we use the methodology developed by Berthelemy and Maurel (2024) to filter out projects for which we cannot detect any significant change in their NTL trend after project implementation. First, we compute the trend on monthly data up until the end of year T-1 and then we measure the deviations from the trend during years T+1, T+2 and T+3 [16]. We then compute the probability with a t-test, that the average deviation from the predicted trend is different from 0. If the probability of the t-test is superior to 0.9 and the average deviation from the trend is positive we assume that the project has succeeded during the first 3 years and failed otherwise. We therefore classify here as failed a project for which we are not able to detect a significant positive change from its NTL trend. We then run the previous difference-in-differences model on the two sub-samples of estimated successful and failed projects. Although this method requires limited information, we must notice it is necessarily limited and lack in nuances compared to on the field surveys, and may therefore miss on the multidimensional aspect of success or failure.

As shown in the table 7 our sample displays results quite similar to experts' findings with a failure rate averaging around 49% across countries. These results are in line with the findings in Berthelemy and Maurel (2024) although the analysis period and geographical extent are very different between the sample used. We can notice a lower failure rate computed for Burkina Faso's mini-grids compared to Madagascar. This difference seems to be driven by a higher failure rate of hydro powered mini-grids classified as failures. We can only speculate on the causes of failures, such as tariffication, economic models or governance models, given we do not have access to on-field data.

Table 7: Descriptive statistics of Success and failures measured by the deviation from the NTL trend

Trend deviation test	Burkina Faso	Madagascar	Total
Failed (p-value > 0.1)	32	39	71
Successful (p-value < 0.1)	43	30	73
Total	75	69	144
Failure rate	*43%*	*57%*	*49%*

Table 8 reports the estimated effects for the two subsamples and, as it could be expected, the failed group shows a lower effect, although it remains significant on the overall post treatment average. The post treatment average effect for the first 3 years for all projects, as reported in table 5 with the complete sample, goes from 0.028 nW/cm^2/sr to 0.043 nW/cm^2/sr for the success group only, representing an increase of 54% of the effect. This means high failure rates could lead to an important underestimation of the potential effect due to reliable

[16] The trend is computed with monthly fixed effects. If the trend computed is not statistically significant at the 10% level, we use the NTL pre-average to replace predicted values from the trend. The year of implementation is excluded from our data because we do not know, within that year, what is the exact month of implementation for most projects (68%).

electrification. Although we could expect to see no significant effect with the failed group, we still measure an overall significant increase in the NTL over the periods. However, the Failed group yields an effect that is about 4 times smaller than the Successful group. One possible explanation to still measure some positive effects on the failed group is the fact that in the failure group some projects might have endured quite small breakdowns which have been repaired over time. Another possible explanation is some form of learning mechanism where communities learned to use the infrastructure. Therefore, few projects were not complete failures per say and worked with some lag, leading to driving the mean of the failed group to a positive and significant effect overall. This is one limit of our method of classification, as it is imperfect and based on remote sensing data which cannot capture the multidimensionality aspect of failures[17].

Table 8: Dynamic effects by success or failure

VARIABLES	(1) Successful	(2) Failed
Post treatment average	0.0434***	0.0110***
	(0.00553)	(0.00338)
T 0	0.0211***	0.00389
	(0.00498)	(0.00287)
T+1 year	0.0447***	0.00396
	(0.00592)	(0.00402)
T+2 years	0.0509***	0.0147***
	(0.00704)	(0.00510)
T+3 years	0.0633***	0.0252***
	(0.0106)	(0.00565)
Observations	2236	2256
R-squared	0.54	0.72
Treated units/controls	73/219	71/213
Mean of NTL	0.1415	0.1289

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
Note: estimations use MD3 control group

[17] This blog provides some preliminary results for micro-grids implemented with RCT design in Madagascar: https://www.entreprenanteafrique.com/en/measuring-the-impact-of-decentralised-electrification-projects-4-4-characterisation-of-the-impacts-of-decentralised-electrification-projects-on-access-to-electricity-using-household-data/

4.2. Heterogeneity by project characteristics

We hereafter investigate how projects characteristics might influence the effect measured. Indeed mini-grids are quite diverse in terms of technical solutions used and those different engineering designs could yield quite different effects.

Table 9 displays the results on different sub-samples based on projects power and grid connection status. Columns 1 displays the average results for the all sample serving as a baseline. Columns 2 and 3 displays the results for Madagascar's projects only, split in two sub-samples, a High-power sub-sample, defined by a production capacity over the median value of 0.036 MW, and a Low-power sub-sample for projects under the median[18]. Both power groups represent quite well the different technologies used, namely solar, hydro and hybrid solar, even though hybrid tend to be more represented in the high-power group. Although there is no clear difference in the effect measured at the post treatment average level, effects measured at the third year decreased and are only significant at the 10% level. This might indicate that smaller installations will not yield expected results of electrification over the long term. Another distinction we can make with the lower power group is the absence of a continuous positive effect, which is confirmed by F-test (see note under table 9) showing that coefficients are not different from each other.

Columns 4 and 5 provide results for projects located in Burkina Faso's and its specificity linked to mini-grids being connected to the grid, either as for power smoothing and paired to a local production unit (column 4) or as the sole provider of power (column 5). Connection of the mini-grid to the grid does not seem to yield very different estimates of the effect and performs well. Column 5 provides results for Burkina Faso's COOPEL, which are mini-grids connected to the national grid without local production unit, which is a sub-sample of column 4. Although the number of these mini-grids type is quite limited, results indicate that the COOPEL perform well compared to the entire sample.

[18] We excluded projects from Burkina Faso as many are missing power production value and could be biased in reporting mainly high-power projects.

Table 9: Dynamic effects by Projects power and grid connection

VARIABLES	(1) ALL sample	(2) High Power	(3) Low Power	(4) Connected to grid	(5) No local Production
Post treatment average	0.0280***	0.0293***	0.0256***	0.0312***	0.0383***
	(0.00379)	(0.00972)	(0.00550)	(0.00602)	(0.00779)
T 0	0.0132***	0.0154**	0.0254***	0.00451	0.0115**
	(0.00296)	(0.00778)	(0.00728)	(0.00349)	(0.00455)
T+1 year	0.0252***	0.0232**	0.0316***	0.0269***	0.0341***
	(0.00405)	(0.00973)	(0.00660)	(0.00641)	(0.00874)
T+2 years	0.0338***	0.0407***	0.0267***	0.0391***	0.0397***
	(0.00490)	(0.0125)	(0.00763)	(0.00839)	(0.0107)
T+3 years	0.0445***	0.0440**	0.0136*	0.0548***	0.0680***
	(0.00727)	(0.0182)	(0.00816)	(0.0123)	(0.0160)
Observations	4492	1040	959	1163	537
R-squared	0.61	0.61	0.74	0.82	0.82
Treated units/controls	144/432	33/99	31/93	37/111	27/81
Mean of NTL	0.1353	0.1217	0.1121	0.1478	0.1525

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: All estimates are using MD3 control group; Columns (2) and (3) use only Madagascar's projects; Columns (4) and (5) use only Burkina Faso's projects.

F-test for equality of the effects for Low Power (3): p-value = .14026467.

Table 10 provides results by renewable technology used as the main source of electricity production. Columns 2 and 3 show the results for mini-grid projects, in both countries, using either solar as the main power source, including hybrid solutions, or hydro. The post treatment average coefficients show significant and positive effects for both technologies, however hydro powered projects have a magnitude of about half of solar powered projects. Hydro powered projects do not show clear significant effect after the first year of implementation. Column 4 provides the results for hydro powered mini-grids located only in Madagascar and confirms those findings, although those hydro mini-grids are composed of high and low power projects. This result suggests that hydro technology might yield lower impacts particularly on the long term, especially since we do not find evidence for a positive dynamic during the first three years.

Table 10: Dynamic effects by Projects by renewable technology

VARIABLES	(1) ALL sample	(2) Solar powered	(3) Hydro powered	(4) Hydro powered Madagascar only
Post treatment average	0.0280***	0.0319***	0.0160***	0.0192***
	(0.00379)	(0.00511)	(0.00548)	(0.00616)
T 0	0.0132***	0.0175***	0.0200***	0.0237***
	(0.00296)	(0.00398)	(0.00704)	(0.00842)
T+1 year	0.0252***	0.0283***	0.0237***	0.0266***
	(0.00405)	(0.00533)	(0.00732)	(0.00904)
T+2 years	0.0338***	0.0410***	0.00579	0.00994
	(0.00490)	(0.00641)	(0.00585)	(0.00649)
T+3 years	0.0445***	0.0466***	0.0142*	0.0155*
	(0.00727)	(0.00979)	(0.00749)	(0.00871)
Observations	4492	2336	696	528
R-squared	0.61	0.64	0.37	0.71
Treated units/controls	144/432	89/267	26/78	20/60
Mean of NTL	0.1353	0.1637	0.1276	0.1070

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: All estimates are using MD3 control group; Columns (1), (2) and (3) use both countries and column (4) use Madagascar's projects only. F-test for equality of the effects for hydro Madagascar (4): p-value = .10941125.

4.3. Robustness

We investigate the effect of changing the matching methodology, using Euclidean distance rather than the Mahalanobis distance. Table A.7 in annex reproduces results from Table 5 using Euclidean distance matching. Results are very similar across control groups and remains significant and positive with similar magnitudes and dynamics to the one found with Mahalanobis distance.

Table A.8 and Figure A.1 in annex reproduces the main results from Table 5 using the Callaway and Sant'Anna estimator as well as regular TWFE estimator rather than de Chaisemartin & D'Haultfœuille. All results remain significant and positive and the magnitude of the effects are very similar. We also confirm the dynamic discussed previously of an increasing effect during the first 3 years the effect.

Overall estimates are robust to different matching method changes, the estimator used and log transformation of the NTL variable.

5. Conclusion

Our findings provide robust evidence that mini-grids have a positive and significant impact on the NTL in both Burkina Faso and Madagascar. Three years after implementation, NTL increased on average by 0.045 nW/cm^2/sr compared to the reference period. Additionally, the NTL levels show a dynamic continuous increase during the initial 3 years after implementation, indicating a rise in electricity access over time, which can be interpreted as an improvement toward the achievement of SDG7. However, this dynamic could vary slightly between countries and we are unable to conclude on the effect over the long term, given most projects are quite recent. Although we find positive effects in both countries, which covers quite different levels of grid development, differences found in the magnitude of the effects pushes for caution with regards to generalization of those results for other countries.

We further examined projects characteristics and their impact on the effect magnitude. Higher power capacity mini-grids yield higher measured effect by the NTL, specifically over the long term, although low power mini-grids still provide positive and significant results. Our findings also suggest solar-powered mini-grids yields a greater and more sustainable effect on the NTL than hydro-powered solutions. Hydro-powered mini-grids showed moderate effects, with significant impacts only during the first year after implementation. Concerning mini-grids that are connected to the national grid, as seen in Burkina Faso, it does not seem to clearly improve nor hinder the measured impact on the NTL.

Our analysis also included a methodology to assess the likelihood of mini-grid project success or failure. We found, based on this new sample, a potential failure rate averaging around 49%, which is consistent with previous findings. These failed projects had a significant impact on the overall measured effects. Excluding potentially failed projects increased the post-treatment average effect from +0.028 $nW/cm^2/sr$ to +0.043 $nW/cm^2/sr$.

Although mini-grids have in their vast majority a positive impact on the electricity access, as measured by the increase in NTL, the results over the long term (i.e. above 3 years) are still unclear due to the novelty of their use as a way to electrify rural villages. These findings are promising for the use of mini-grids as a viable solution for rural electrification with renewable energy sources, contributing to the achievement of SDG 7. Nonetheless, the probable high rates of failure, as reported by experts and gauged by NTL data are concerning and calls for further data collection and research to find the determinants of those failures. Addressing these issues is crucial to realizing the full potential of mini-grids in enhancing the lives of rural communities.

Acknowledgement

This work was supported by the Agence Nationale de la Recherche of the French government through the program "France 2030" (grant number ANR-16-IDEX-0001). We want to thank Olivier Santoni, Geomatician at Université Clermont Auvergne, CNRS, IRD, CERDI.

Declaration of interest statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this article.

References

Akinyele, D., Belikov, J., Levron, Y., 2018. Challenges of Microgrids in Remote Communities: A STEEP Model Application. Energies 11, 432. https://doi.org/10.3390/en11020432

Berthelemy, J.-C., Maurel, M., 2024. A sky view evaluation of the impact of mini-grid projects on progress towards SDG7. J. Dev. Eff. 0, 1–20. https://doi.org/10.1080/19439342.2024.2309376

Berthelemy, J.-C., Maurel, M., 2021. A new approach for evaluation of the economic impact of decentralized. Ferdi WP284.

Berthelemy, J.-C., Millien, A., 2018. Impact of Decentralized Electrification Projects on Sustainable Development: A Meta-Analysis.

Beyer, R., Hu, Y., Yao, J., 2022. Measuring Quarterly Economic Growth from Outer Space. Policy Res. Work. Pap. https://doi.org/10.1596/1813-9450-9893

Bitzer, J., Gören, E., 2024. The Impact of Foreign Aid on Local Development: A Grid Cell Analysis. J. Dev. Stud. 0, 1–35. https://doi.org/10.1080/00220388.2024.2377294

Callaway, B., Sant'Anna, P.H.C., 2021. Difference-in-Differences with multiple time periods. J. Econom., Themed Issue: Treatment Effect 1 225, 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chindarkar, N., Goyal, N., 2023. Did it increase energy consumption? A difference-in-differences evaluation of a rural electrification policy in Gujarat, India using night-time lights data. Energy Policy 183, 113814. https://doi.org/10.1016/j.enpol.2023.113814

de Chaisemartin, C., D'Haultfœuille, X., 2024. Difference-in-Differences Estimators of Intertemporal Treatment Effects. Rev. Econ. Stat. 1–45. https://doi.org/10.1162/rest_a_01414

de Chaisemartin, C., D'Haultfœuille, X., 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. Am. Econ. Rev. 110, 2964–2996. https://doi.org/10.1257/aer.20181169

Dugoua, E., Kennedy, R., Urpelainen, J., 2018. Satellite data for the social sciences: measuring rural electrification with night-time lights. Int. J. Remote Sens. 39, 2690–2701.

Duran, A.S., Sahinyazan, F.G., 2021. An analysis of renewable mini-grid projects for rural electrification. Socioecon. Plann. Sci. 77, 100999. https://doi.org/10.1016/j.seps.2020.100999

Duthie, M., Ankel-Peters, J., Mphasa, C., Bhat, R., 2024. The elusive quest for sustainable mini-grid electrification: New evidence from Indonesia. Energy Sustain. Dev. 80, 101454. https://doi.org/10.1016/j.esd.2024.101454

Elvidge, C., Baugh, K., Zhizhin, M., Hsu, F.-C., 2013. Why VIIRS data are superior to DMSP for mapping nighttime lights. Proc. Asia-Pac. Adv. Netw. 35, 62–69. https://doi.org/10.7125/APAN.35.7

Falchetta, G., Noussan, M., 2019. Interannual Variation in Night-Time Light Radiance Predicts Changes in National Electricity Consumption Conditional on Income-Level and Region. Energies 12, 456. https://doi.org/10.3390/en12030456

Falchetta, G., Pachauri, S., Parkinson, S., Byers, E., 2019. A high-resolution gridded dataset to assess electrification in sub-Saharan Africa. Sci. Data 6, 110. https://doi.org/10.1038/s41597-019-0122-6

Gibson, J., Olivia, S., Boe-Gibson, G., Li, C., 2021. Which night lights data should we use in economics, and where? J. Dev. Econ. 149, 102602. https://doi.org/10.1016/j.jdeveco.2020.102602

Ikejemba, E.C.X., Mpuan, P.B., Schuur, P.C., Van Hillegersberg, J., 2017. The empirical reality & sustainable management failures of renewable energy projects in Sub-Saharan Africa (part 1 of 2). Renew. Energy 102, 234–240. https://doi.org/10.1016/j.renene.2016.10.037

Khomba, D.C., Trew, A., 2022. Aid and Local Growth in Malawi. J. Dev. Stud. 58, 1478–1500. https://doi.org/10.1080/00220388.2022.2032668

King, G., Nielsen, R., 2019. Why Propensity Scores Should Not Be Used for Matching. Polit. Anal. 27, 435–454.

Min, B., Gaba, K.M., Sarr, O.F., Agalassou, A., 2013. Detection of rural electrification in Africa using DMSP-OLS night lights imagery. Int. J. Remote Sens. 34, 8118–8141. https://doi.org/10.1080/01431161.2013.833358

Min, B., O'Keeffe, Z.P., Abidoye, B., Gaba, K.M., Monroe, T., Stewart, B.P., Baugh, K., Sánchez-Andrade Nuño, B., 2024. Lost in the dark: A survey of energy poverty from space. Joule. https://doi.org/10.1016/j.joule.2024.05.001

Moradi, A., Schmidt, M., 2022. Community effects of electrification: evidence from Burkina Faso's grid extension. CSAE Work. Pap.

Nelson, A., Weiss, D.J., van Etten, J., Cattaneo, A., McMenomy, T.S., Koo, J., 2019. A suite of global accessibility indicators. Sci. Data 6, 266. https://doi.org/10.1038/s41597-019-0265-5

Ratledge, N., Cadamuro, G., de la Cuesta, B., Stigler, M., Burke, M., 2022. Using machine learning to assess the livelihood impact of electricity access. Nature 611, 491–495. https://doi.org/10.1038/s41586-022-05322-8

Annexes

Table A.1: Burkina Faso's descriptive Statistics after matching

		Number of control units selected per treated							
		10		5		3		1	
Variables	**Projects, N = 75**[1]	**Controls, N = 749**[1]	***p-value***[2]	**Controls, N = 375**[1]	***p-value***[2]	**Controls, N = 225**[1]	***p-value***[2]	**Controls, N = 75**[1]	***p-value***[2]
Surface area (km²)	2.92 (1.71)	2.57 (1.49)	0.089	2.71 (1.60)	0.3	2.79 (1.66)	0.6	2.88 (1.87)	>0.9
Altitude (meters)	309 (39)	312 (37)	0.5	312 (40)	0.5	311 (37)	0.6	309 (39)	>0.9
Land gradient	1.61 (0.52)	1.57 (0.48)	0.5	1.58 (0.45)	0.6	1.62 (0.49)	>0.9	1.57 (0.37)	0.6
Population (hab/km²)	70 (60)	69 (59)	0.9	69 (60)	>0.9	69 (60)	>0.9	69 (59)	>0.9
Distance to cities (minutes)	69 (49)	71 (46)	0.7	71 (47)	0.7	68 (46)	>0.9	68 (47)	>0.9
Potential photovoltaic production (kWh/kWp)	4.58 (0.06)	4.58 (0.06)	>0.9	4.59 (0.06)	0.9	4.58 (0.06)	>0.9	4.59 (0.06)	>0.9
NTL pre-project	0.101 (0.097)	0.080 (0.031)	0.073	0.081 (0.030)	0.088	0.082 (0.030)	0.10	0.08 (0.03)	0.13

[1] Mean (SD)

[2] Welch Two Sample t-test

Table A.2: Madagascar's descriptive Statistics after matching

		Number of control units selected per treated							
		10		5		3		1	
Variables	**Projects, N = 69**[1]	**Controls, N = 690**[1]	***p-value***[2]	**Controls, N = 345**[1]	***p-value***[2]	**Controls, N = 207**[1]	***p-value***[2]	**Controls, N = 69**[1]	***p-value***[2]
Surface area (km²)	1.82 (1.73)	1.17 (1.25)	0.003	1.21 (1.22)	0.006	1.26 (1.26)	0.015	1.37 (1.29)	0.086
Altitude (meters)	528 (548)	530 (564)	>0.9	527 (564)	>0.9	525 (563)	>0.9	516 (564)	>0.9
Land gradient	6.6 (4.6)	6.0 (3.5)	0.3	6.1 (3.8)	0.3	6.1 (3.7)	0.4	5.9 (3.7)	0.3
Population (hab/km²)	60 (48)	54 (42)	0.3	54 (43)	0.4	55 (46)	0.5	60 (56)	>0.9
Distance to cities (minutes)	294 (291)	302 (290)	0.8	301 (293)	0.8	300 (291)	0.9	303 (303)	0.8
Potential photovoltaic production (kWh/kWp)	4.69 (0.37)	4.71 (0.34)	0.6	4.71 (0.33)	0.7	4.71 (0.34)	0.7	4.71 (0.34)	0.8
NTL pre-project	0.076 (0.029)	0.068 (0.026)	0.029	0.068 (0.027)	0.050	0.069 (0.028)	0.12	0.073 (0.038)	0.7

[1] Mean (SD)

[2] Welch Two Sample t-test

Table A.3: Pre-trend coefficients on pooled results

	(1)	(2)	(3)	(4)
VARIABLES	MD1	MD3	MD5	MD10
Placebo_1	-0.00122	-0.00107	-0.000592	-0.000664
	(0.00180)	(0.00182)	(0.00173)	(0.00168)
Placebo_2	0.000130	-0.000996	-0.00136	-0.00154
	(0.00304)	(0.00287)	(0.00278)	(0.00273)
Placebo_3	-0.00106	-0.00191	-0.00311	-0.00423
	(0.00382)	(0.00348)	(0.00341)	(0.00335)
Placebo_4	-0.00551	-0.00773	-0.00953*	-0.0103**
	(0.00609)	(0.00530)	(0.00514)	(0.00502)
Observations	2282	4586	6890	12642
R-squared	0.55	0.61	0.63	0.66
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

As we can see in Table A.3, there is no consistent and significant pre-trend coefficients, meaning that there is no difference in the NTL trend between the villages that received mini-grid projects and the control villages, either with 1, 3, 5 or 10 controls per treated selected.

Table A.4: All projects all periods

VARIABLES	(1) MD1	(2) MD3	(3) MD5	(4) MD10
Post treatment average	0.0296***	0.0310***	0.0318***	0.0320***
(6 years)	(0.00479)	(0.00443)	(0.00439)	(0.00436)
T 0	0.0130***	0.0132***	0.0137***	0.0138***
	(0.00300)	(0.00296)	(0.00296)	(0.00296)
T+1 year	0.0240***	0.0252***	0.0258***	0.0258***
	(0.00427)	(0.00405)	(0.00403)	(0.00402)
T+2 years	0.0322***	0.0338***	0.0346***	0.0350***
	(0.00529)	(0.00490)	(0.00485)	(0.00482)
T+3 years	0.0431***	0.0445***	0.0452***	0.0455***
	(0.00769)	(0.00727)	(0.00722)	(0.00718)
T+4 years	0.0392***	0.0417***	0.0433***	0.0439***
	(0.0130)	(0.0125)	(0.0124)	(0.0124)
T+5 years	0.0519***	0.0550***	0.0570***	0.0578***
	(0.0161)	(0.0156)	(0.0156)	(0.0155)
Observations	2282	4586	6890	12642
R-squared	0.54	0.60	0.62	0.64
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288

Table A.5: Results with raw NTL data (no correction for negative radiance values)

	(1)	(2)	(3)	(4)
VARIABLES	MD1	MD3	MD5	MD10
Post treatment average	0.0281***	0.0291***	0.0298***	0.0300***
	(0.00416)	(0.00388)	(0.00386)	(0.00384)
T 0	0.0135***	0.0137***	0.0142***	0.0143***
	(0.00309)	(0.00305)	(0.00305)	(0.00305)
T+1 year	0.0246***	0.0259***	0.0264***	0.0264***
	(0.00433)	(0.00412)	(0.00410)	(0.00410)
T+2 years	0.0337***	0.0353***	0.0361***	0.0365***
	(0.00538)	(0.00498)	(0.00494)	(0.00490)
T+3 years	0.0454***	0.0468***	0.0475***	0.0478***
	(0.00777)	(0.00735)	(0.00730)	(0.00726)
Observations	2188	4492	6796	12548
R-squared	0.57	0.64	0.66	0.69
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1378	0.1300	0.1265	0.1235

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: MD#: Mahalanobis distance; # refers to the number of controls per treated retained.

Table A.7: Dynamic effects for all projects using Matching by Euclidean distance (ED).

	(1)	(2)	(3)	(4)
VARIABLES	ED1	ED3	ED5	ED10
Post treatment average	0.0271***	0.0292***	0.0294***	0.0281***
	(0.00402)	(0.00382)	(0.00377)	(0.00430)
T 0	0.0119***	0.0137***	0.0141***	0.0136***
	(0.00304)	(0.00296)	(0.00295)	(0.00308)
T+1 year	0.0242***	0.0264***	0.0265***	0.0246***
	(0.00422)	(0.00406)	(0.00402)	(0.00496)
T+2 years	0.0334***	0.0354***	0.0355***	0.0338***
	(0.00520)	(0.00493)	(0.00486)	(0.00553)
T+3 years	0.0439***	0.0461***	0.0464***	0.0456***
	(0.00766)	(0.00730)	(0.00723)	(0.00742)
Observations	2188	4492	6796	12548
R-squared	0.57	0.62	0.65	0.57
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: ED# : Euclidean distance ; # refers to the number of controls per treated retained.

TableA.8: All projects estimate using Callaway and Sant'Anna estimator for TWFE model

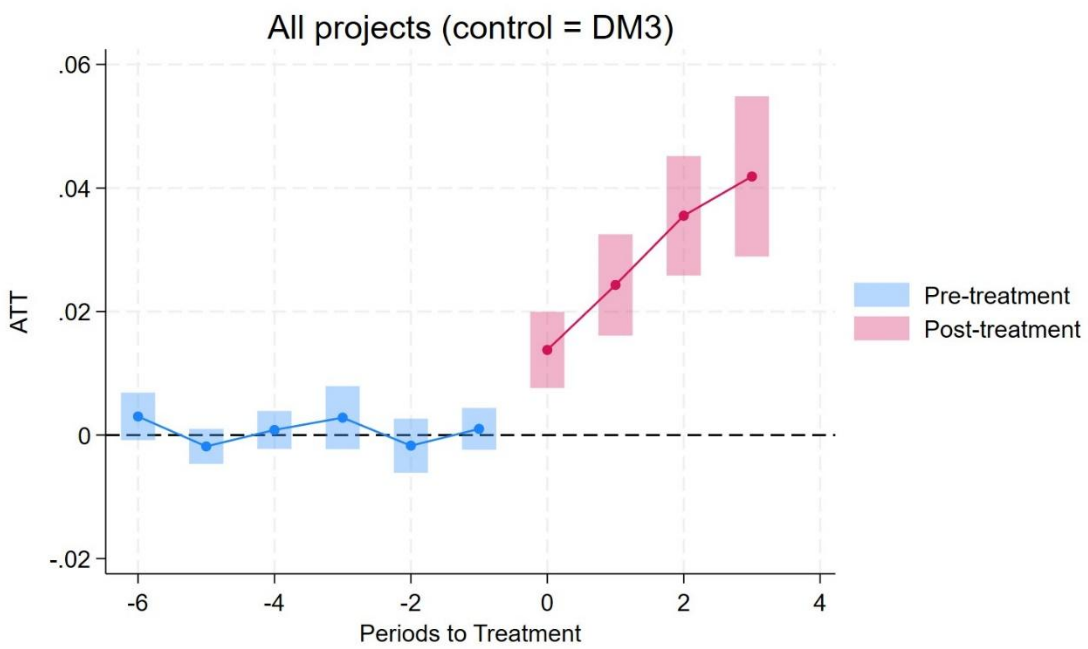
VARIABLES	(1) MD1	(2) MD3	(3) MD5	(4) MD10
Post treatment average	0.0277***	0.0288***	0.0292***	0.0301***
	(0.00418)	(0.00390)	(0.00368)	(0.00374)
T 0	0.0145***	0.0141***	0.0143***	0.0147***
	(0.00321)	(0.00312)	(0.00310)	(0.00308)
T+1 year	0.0229***	0.0239***	0.0242***	0.0253***
	(0.00431)	(0.00416)	(0.00393)	(0.00398)
T+2 years	0.0324***	0.0349***	0.0357***	0.0367***
	(0.00525)	(0.00493)	(0.00453)	(0.00455)
T+3 years	0.0409***	0.0425***	0.0427***	0.0434***
	(0.00718)	(0.00667)	(0.00652)	(0.00667)
Observations	1920	3840	5760	10560
R-squared	0.55	0.61	0.63	0.66
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Notes: MD#: Mahalanobis distance; # refers to the number of controls per treated retained.

Figure A.1: Dynamic effect computed by Callaway and Sant'Anna method.



Note: MD3 control group is used.

Table A.9: Regular TWFE estimator

VARIABLES	(1) MD1	(2) MD3	(3) MD5	(4) MD10
Post treatment average	0.0244***	0.0295***	0.0312***	0.0330***
	(0.00487)	(0.00476)	(0.00479)	(0.00479)
T 0	0.00993	0.0127**	0.0117**	0.0134***
	(0.00628)	(0.00572)	(0.00523)	(0.00498)
T+1 year	0.0160*	0.0236***	0.0232***	0.0230***
	(0.00898)	(0.00758)	(0.00705)	(0.00678)
T+2 years	0.0307***	0.0362***	0.0343***	0.0365***
	(0.00783)	(0.00718)	(0.00681)	(0.00657)
T+3 years	0.0304***	0.0309***	0.0301***	0.0284***
	(0.0110)	(0.0105)	(0.0101)	(0.0103)
Observations	2,648	5,296	7,944	14,556
R-squared	0.853	0.849	0.851	0.852
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table A.10: Regular TWFE estimator with Country Year interaction

VARIABLES	(1) MD1	(2) MD3	(3) MD5	(4) MD10
Post treatment average	0.0241***	0.0292***	0.0307***	0.0326***
	(0.00467)	(0.00462)	(0.00466)	(0.00466)
T 0	0.00993***	0.0127***	0.0117***	0.0135***
	(0.00372)	(0.00398)	(0.00352)	(0.00330)
T+1 year	0.0160***	0.0236***	0.0232***	0.0231***
	(0.00531)	(0.00493)	(0.00456)	(0.00442)
T+2 years	0.0307***	0.0362***	0.0343***	0.0366***
	(0.00669)	(0.00628)	(0.00599)	(0.00580)
T+3 years	0.0304***	0.0309***	0.0301***	0.0284***
	(0.0102)	(0.0100)	(0.00975)	(0.0100)
Observations	2,648	5,296	7,944	14,556
R-squared	0.870	0.868	0.870	0.871
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of NTL	0.1429	0.1353	0.1318	0.1288
Country Year interaction	YES	YES	YES	YES

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A.11: Results with log of NTL

VARIABLES	(1) MD1	(2) MD3	(3) MD5	(4) MD10
Post treatment average	0.0804***	0.0809***	0.0827***	0.0811***
	(0.0229)	(0.0223)	(0.0222)	(0.0220)
T 0	0.0424*	0.0410*	0.0428*	0.0423*
	(0.0223)	(0.0220)	(0.0219)	(0.0218)
T+1 year	0.0700***	0.0701***	0.0715***	0.0693***
	(0.0233)	(0.0229)	(0.0228)	(0.0227)
T+2 years	0.108***	0.111***	0.113***	0.111***
	(0.0256)	(0.0245)	(0.0243)	(0.0241)
T+3 years	0.110***	0.111***	0.113***	0.110***
	(0.0322)	(0.0311)	(0.0309)	(0.0308)
Observations	1920	3840	5760	10560
R-squared	0.75	0.77	0.78	0.79
Treated units/controls	144/144	144/432	144/720	144/1439
Mean of log NTL	-2.2776	-2.3170	-2.3375	-2.3523

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Notes: MD#: Mahalanobis distance; # refers to the number of controls per treated retained. We define log of NTL as log(NTL+0.01) to avoid issues with 0 values.

Chapter 2:
Experimental impact study on a micro-grid project in rural Madagascar: The Café Lumière project

Jean-Claude Berthélemy [a,c], Vincent Nossek [b,c,*], Victor Béguerie [c]

[a] *University Paris 1 Pantheon-Sorbonne ;* [b] *Université Clermont Auvergne, CNRS, IRD, CERDI ;*

[c] *FERDI.*

** Corresponding author contact:* vincent.nossek@uca.fr

ABSTRACT

This paper evaluates the impact of the Café Lumière project, an initiative launched by the NGO Electriciens sans Frontière (ESF) to address the lack of reliable electricity in rural Madagascar. Implemented in six villages across the Vakinankaratra and Itasy highland regions, the project integrates solar-powered micro-grids and energy platforms to provide affordable electricity and basic electricity services for households, businesses, and public services. The study employs a randomized control trial (RCT) design, with data collected through two waves of household surveys in 2017-2018 and 2023, complemented by locality data. The experimental dimension of the paper is based on a treatment implemented at the locality level. A dozen of comparable localities had been initially selected after a common pre-feasibility study as suitable for the project, and the project has been randomly implemented in only half of them. Findings reveal notable improvements in household access to electricity, particularly in the use of modern energy services like lighting and phone charging. The project has also facilitated economic activities, especially secondary income-generating occupations, though the overall impact on household wealth and income remains limited. Social outcomes, including health, education, and public safety, have seen mixed results, with only minor improvements in health. The limited extension of all these observed positive impacts is also compounded by their spatial narrowness. It is possible that a longer observation period would be necessary for the impacts of the project to fully materialize. However, more ambitious complementary initiatives foreshadowed in the project's theory of change but not yet fully implemented may be needed to achieve the broader economic and social transformation initially anticipated.

Keywords: electricity access; renewable energy; mini-grids; rural electrification; Randomized controlled trial (RCT); impact evaluation; Multiservice energy platforms.

1. Introduction

Madagascar's rural population faces significant challenges in accessing reliable electricity (rural electrification rate of 10,9% (2022, World Development Indicator), which limits its social and economic development in many dimensions.

The "Café Lumière" project by the NGO Electriciens sans Frontière (ESF), implemented since 2018, proposes to address these multiple issues through an innovative solution combining a solar-energy powered micro-grid with an energy platform. The power plant feeding the micro-grid provides power available for income generating activities, part of households, collective services and the platform. The platform is essentially an electricity boutique producing basic electricity services with the objective of supplying these services at an affordable cost to the whole village population. Within the overall framework of sustainable development goals (SDGs), this project is designed to simultaneously fulfill the public policy objective of universal access to basic electricity services embedded in the SDG7, develop an economically sustainable business model for rural electrification that supports economic development objectives and improve collective services, thereby promoting social and human development progress. These three specific objectives are actually intertwined because the development of sustained electricity demand is necessary to ensure economic sustainability of the project and to deliver welfare improvements necessary to sustain engagement of all community actors involved In the Café Lumière context, the interaction between social and economic objectives is reinforced by the introduction of a cross-subsidy mechanism, in which the electricity consumption of public services is subsidized by a local tax collected by the operator from private micro-grid customers and redistributed to the municipal authorities.

Evaluation of the impacts of micro-grid solutions is important because it has been poorly studied in the literature, in particular within a Randomized Control Trial (RCT) setting (Berthelemy and Millien, 2018). Most of the literature related to impacts of electricity access are based on grid network extension and usually covering Asian countries(Moore et al., 2020) and have typically relied on instrumentation strategy (Dinkelman, 2011). However this literature is poorly relevant to assess rural electrification through micro-grids as services offered to the population are quite different, as demonstrated by the Multi-Tier Framework (MTF) and the related literature (Nicolas et al., 2019). Micro-grids are now seen as a robust alternative to bring electricity access in rural areas, in particular in Africa, but have been criticized to under-deliver in their expected impacts, in particular for driving economic activity, due to unsustainable business models over the long term, improper sizing, pricing or property ownership model (Duran and Sahinyazan, 2021; Duthie et al., 2024).

Most of the rural electrification impact literature using RCT designs has focused on standalone systems usually leading to quite limited impacts on topics such as health or education (Kudo et al., 2019a, 2019b). One similar study to our paper also based on an RCT design, which covered effects of microgrids installed in India, also found limited impacts, being able to conclude only on significant effects on electricity access rate and kerosene expenditures, but did not find significant effects on households wealth or consumption, local economic development or other broader indicators of socioeconomic development (Aklin et al., 2017). Findings of limited outcomes related to economic and social topic are corroborated by another study focusing on rural electrification using on-grid randomized grid connections (Lee et al., 2020). Our study enriches the existing literature by providing rigorous RCT-based evidence on the impacts of micro-grid projects, the Café Lumière project, in rural Africa, specifically highlighting socio-economic outcomes in the under-researched context of rural Madagascar.

The Café Lumière concept has been initially implemented on a pilot project financed by Agence Française de Développement (AFD) in six villages located in the Vakinankaratra and Itasy regions of Madagascar. This pilot has served as a feasibility test and an evaluation strategy has been put in place from the beginning to study its impacts. This paper reports the conclusions of this impact evaluation, implemented by independent researchers with the collaboration of the project developer (ESF) and the operator of the Café Lumière (Anka Madagascar) for data collection.

The core of the impact evaluation strategy presented in this paper is based on a two-wave household survey implemented in 2017-2018 and in 2023, covering six villages where the pilot has been implemented between 2018 and 2021 and six comparable control villages. These two sub-groups were randomly selected from an initial list of 12 villages identified by a preliminary study as suitable for the Café Lumière project. This random assignment of treatment at the village level is at the core of the RCT design, which has been put forward as the gold-standard design for impact evaluation and in particular in research in development economics aimed at poverty reduction (Banerjee et al., 2019). To facilitate the interpretation and validation of results obtained from this quantitative analysis, the household survey data were complemented by monthly implementation data provided by the project operator, as well as locality-level data collected concurrently with the household surveys.

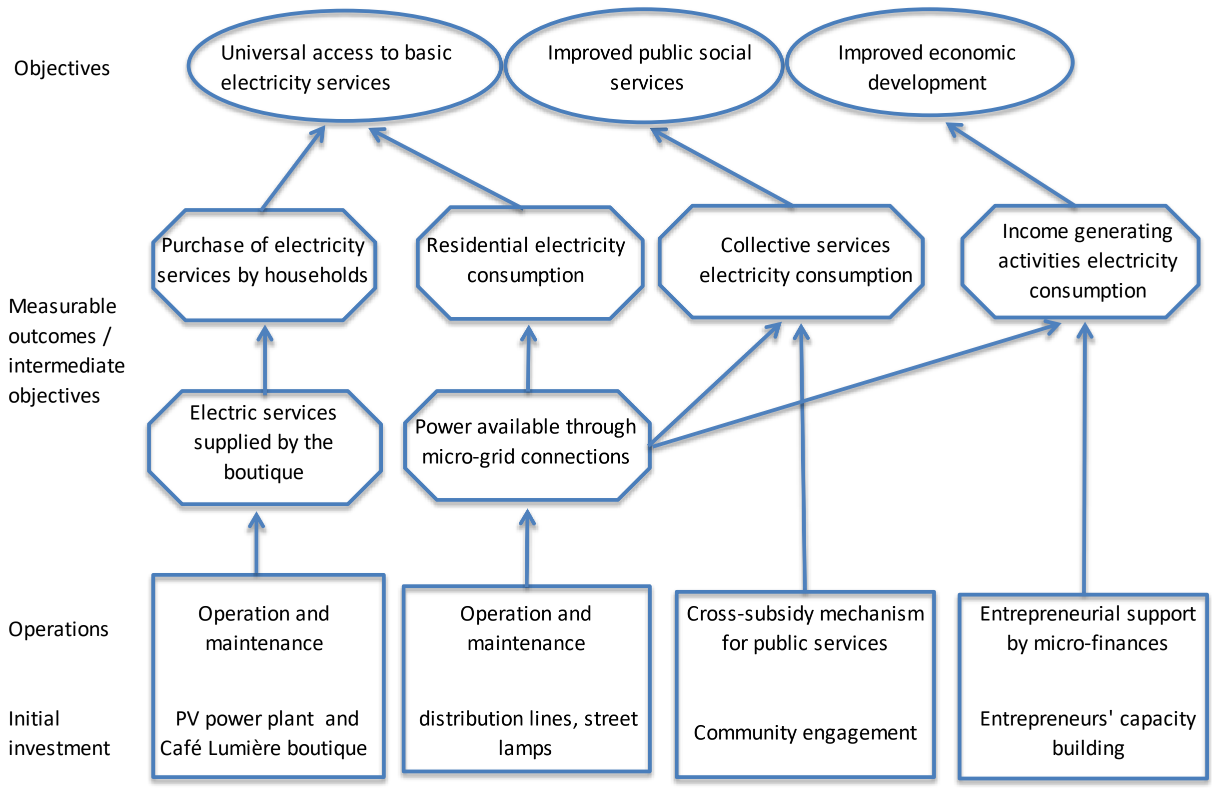
Our findings highlight significant positive impacts of the Café Lumière project on household access to basic electricity services, particularly through increased use of electricity for lighting and mobile phone charging. Although broader socio-economic outcomes such as wealth and income improvements remain modest, the project demonstrates clear benefits for economic activities reliant on electricity, notably by promoting mechanized cereal processing and small-scale entrepreneurship. Furthermore, we document notable improvements in maternal and child healthcare services enabled by the electrification of health centers. This paper contributes to the existing literature by providing rigorous, RCT-based evidence on the effectiveness of integrated micro-grid solutions in rural Africa, filling a crucial gap by examining socio-economic impacts within the previously under-researched context of Madagascar.

The rest of this paper is structured as follows. In the first section we present the Café Lumière project and its theory of change and data related to its implementation phase. We then present data on the surveys and how they relate to the objective from the theory of change as well as the principal econometric method proposed to derive tests of impacts in this statistical context. Results of expected impacts observed on households and/or individuals' levels are reported in the next section. Then, we discuss these results by crosschecking statistical test results with alternative methods and by triangulating conclusions with locality data. The project has produced favorable outcomes, although with mixed results in several dimensions, which we discuss in conclusion.

2. The Café Lumière Project

2.1 Theory of Change

Figure 1: Theory of change of the Café Lumière project



The *Café Lumière* project aims at promoting SDGs through 3 complementary specific objectives: universal access to basic electricity services, improved public social services and improved economic development. Figure 1 describes the theory of change of the project, designing the investments and operations implemented with a view to meet these specific objectives.

The technical solution of the Café Lumière does not differ from the majority of micro-grid projects implemented in developing countries. It consists in the production of electricity by solar panels, backed by batteries and an emergency diesel generator, distributed through a micro-grid. Based on previous experience, the project aims at promoting both social and economic development in villages of implementation.

The social dimension of the project primarily rests on promoting universal access to basic electricity services. Universal access was considered by ESF unattainable by the mere provision of electricity through the micro-grid, given that connection to this grid may entail fixed costs that are too high for the poor. For this reason, an electricity services provision is proposed in a boutique managed by the operator supplying notably recharges of cell phones as well as recharges and renting of solar lamps. Social development also rests on the development of social services, through the connection of health centers, schools and streetlamps. A cross-subsidy mechanism reducing the cost of public services electricity bills has been introduced, initially through direct support by the operator and since 2023 through a local tax paid by the operator to the municipal authorities. The outcome of this financial support to collective services depends of course on the engagement of all community actors.

On the economic dimension, the micro-grid provides households and entrepreneurs, named henceforth very small size enterprises (VSEs) access to more reliable load, on a longer daily schedule and with potentially higher electric intensity than solar home systems, which were already used by some household and VSEs before the start of the project. A few VSEs used also diesel generators, but at a high operational cost. These immediate outcomes contribute to increased productivity of economic activities by enabling local businesses to operate for longer hours and offer a wider range of services. Over time, it is expected that these benefits will foster entrepreneurship and the creation of new businesses, leading to economic diversification. In order to promote this economic transformation, micro-finance interventions were planned to help VSEs to invest in electricity-powered equipment. Capacity building of VSEs, on technical and managerial skills, was also considered as an important factor of success of economic transformation.

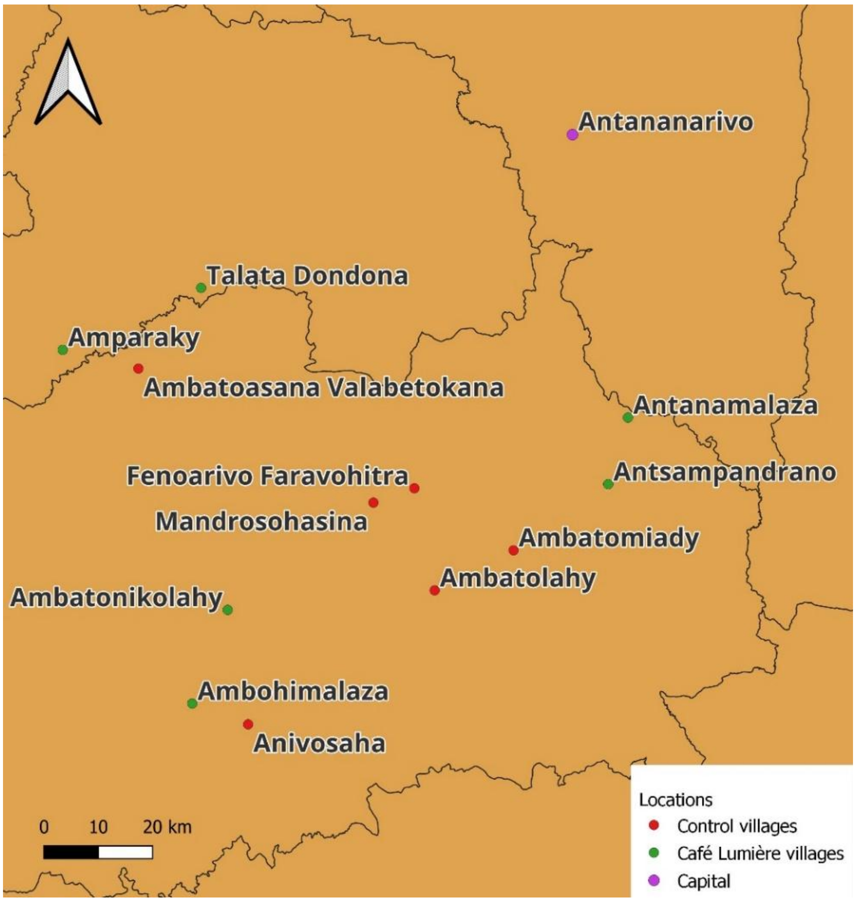
Increased household wealth associated with the expected economic transformation is a key determinant of the economic sustainability of operations of the project and the ability to maintain equipment and invest, contributing to a virtuous cycle of economic growth. This economic growth, in turn, encourages further investment in education and health, creating a sustainable pathway to long-term socio-economic development. The sustainability of the project over time depends also on financial results of the operator, whose activity relies mostly on fixed costs, and therefore on electricity demand.

2.2 Project implementation

The Café Lumière project equipped 6 villages located in the southwest of Antananarivo, more precisely in the regions of Vakinankaratra and Itasy between 2018 and 2021[26]. Figure 2 shows the location of the 12 villages.

The analysis in this section is based on the reporting done by Anka on its operations, available on a monthly basis from 2019 to 2023 in the 6 equipped villages. Such data document the actual implementation of the project. Table 1 lists the principal implementation data used in this section. These data are useful to understand the theory of change, by documenting in a single matrix action and expected outcomes. When it comes to tests the impact of the project, we also mobilize survey data collected in treated and non-treated villages. To illustrate the overall conceptual consistency of our statistical space, we list in the same matrix survey data, which will be analyzed in the following sections, to visualize how such data relate to the expected outcomes.

Figure 2: Map of project and control villages



[26] Ambatonikolahy in 2018; Ambohimalaza in 2019; Antanamalaza, Antsampandrano and Dondona in 2020; Amparaky in 2021. See Table S1 in supplementary material for more details.

Table 1: mapping of data sources

Observations / Objectives	Outcomes reported by the operator	Tests of impacts using households' survey	Tests of impacts using individuals survey	Triangulation using locality survey data
Universal access to basic electricity services		Private household electricity access		
	Phone & lamp recharges by the Café Lumière boutiques	Light sources at home (electricity, solar lamps, kerosene lamps)	Mobile phones possession, recharging and usage	
	Electricity subscriptions by households			
Improved public social services	Electricity subscriptions by collective services			
		Incidence of illnesses at household level		Electrified primary health care center
		Quality of water access and water treatment		
			Birth delivery with electricity	
			Children vaccination	
			School attendance and completion	Electrified primary and secondary schools
	Electricity supplied for street lighting	Incidence of thefts		
Improved economic development	Electric power consumption at the village level	Wealth index	Secondary incomes	
	Electricity subscriptions by VSEs			Number and electrification of services and commerce
	Electricity subscriptions by households	Possession of high-power appliances by households		
		Electricity used by households for economic purposes	Individual activities and their reliance on electricity	
		Mechanized shelling of cereals		
	Capacity building and microfinance		Access to information	

Universal access to basic electricity services:

The demand by households for direct electricity access through the micro-grid has grown rapidly in several villages, notably in the latest equipped villages (Figure 3). But this does not necessarily contribute a lot to the universal access objective, as it concerns only relatively few households (about 60 households per village at the end of 2023, which is about ten percent of the targeted population).

Compared to the planned design of the project, the weight given to the principal action foreseen to meet the universal access objective, namely supplying basic electricity services to household, has declined. While in the first equipped villages there have been noticeable numbers of recharges of cell-phones and solar lamps, this activity has decline rapidly and is absent in the last equipped villages (Figure 4).

It seems that the number of households potentially interested by this service had been overestimated, partly because the same services have been offered by private businesses and partly because the subscriptions of the households to the micro-grid has grown more rapidly than expected. In turn, this may have led the operator to reduce its implication on this activity becoming possibly less profitable than expected.

Figure 3: Micro-grid subscriptions (households)

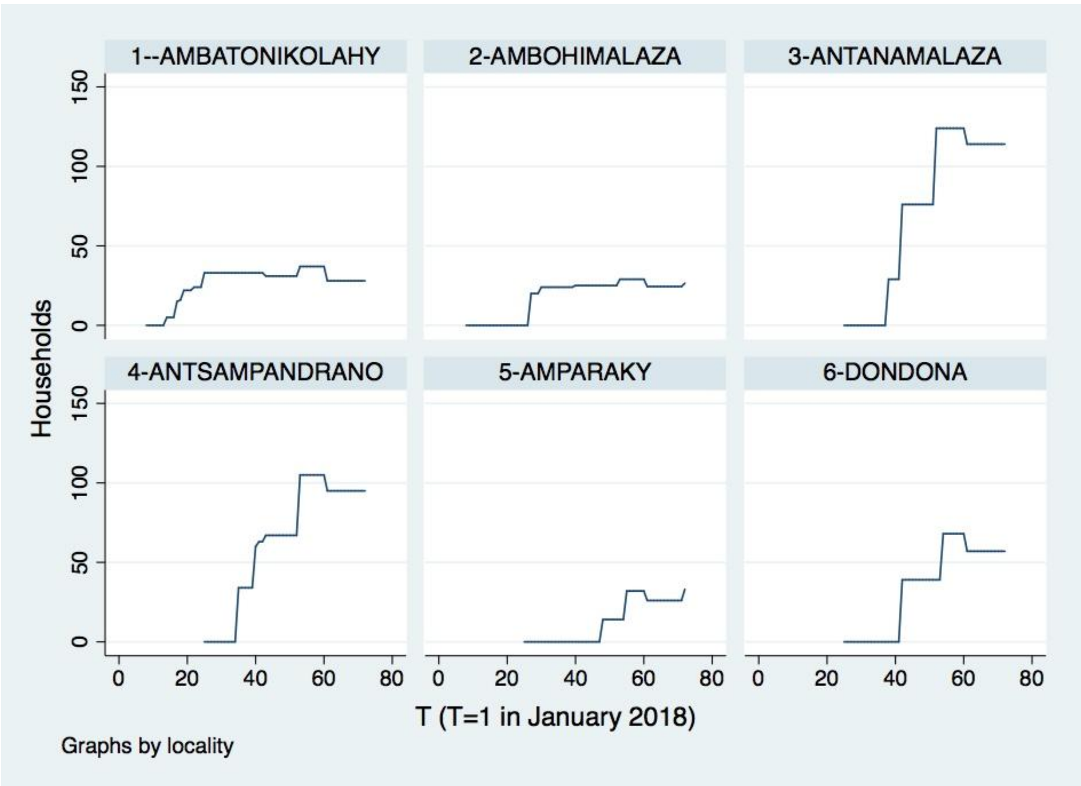
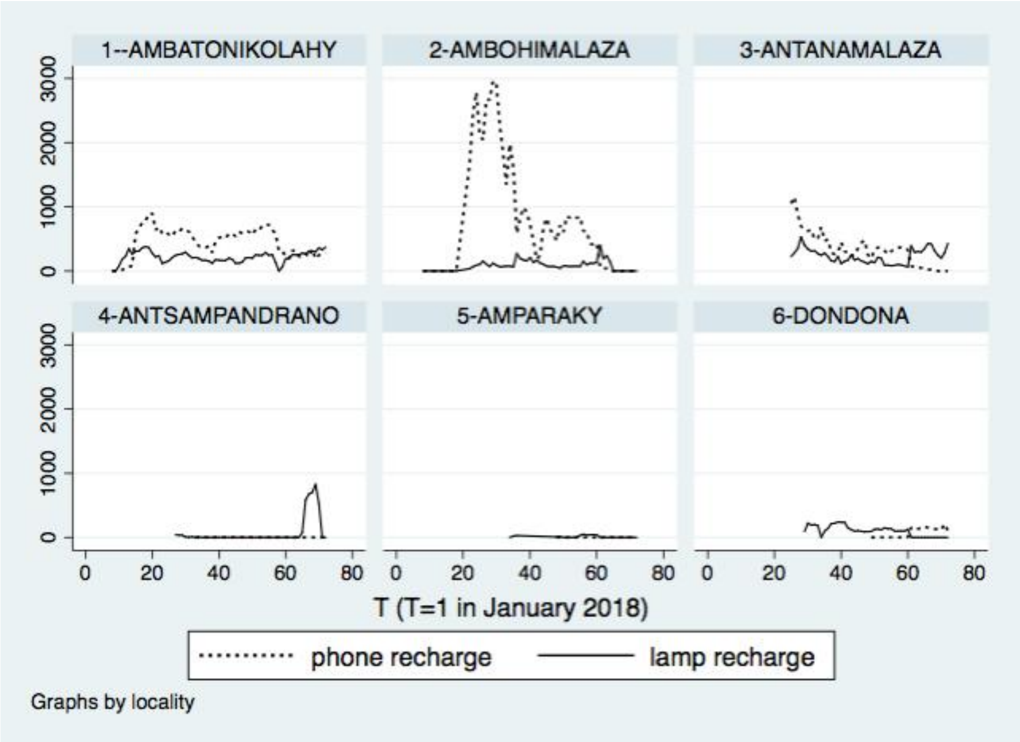


Figure 4: Phone & lamp recharges by the Café Lumière boutiques



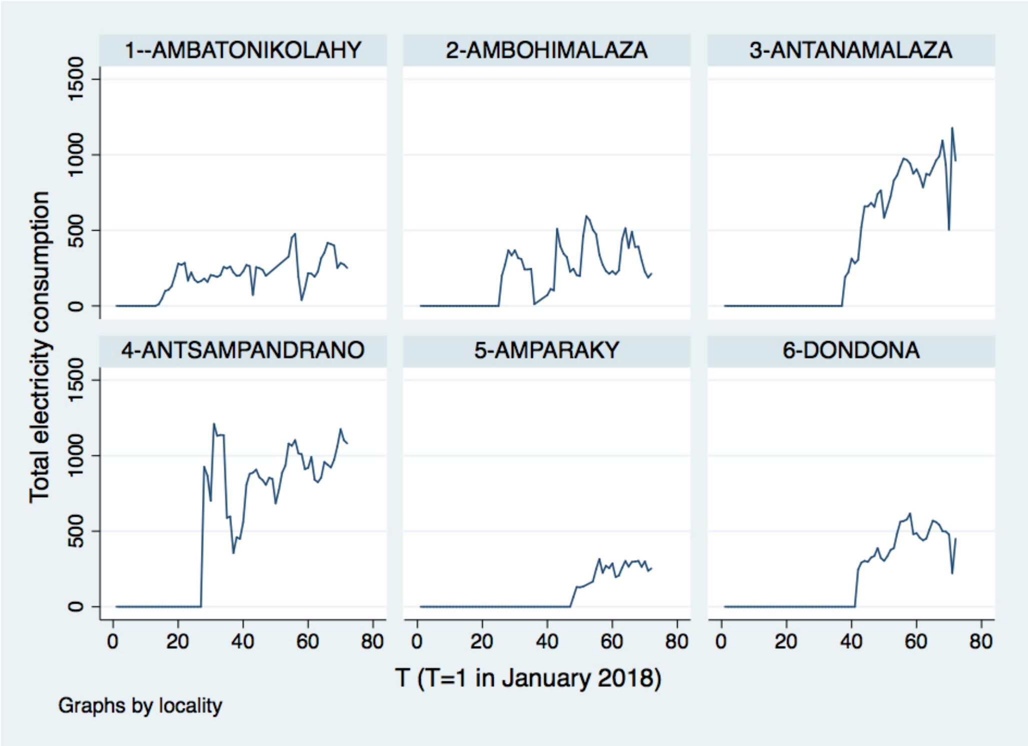
Improved public social services:

Regarding electricity supplied to collective services, we obtained from the operator a detailed list of electricity subscriptions by collective services providers as of 2022. Most health centers have received a connection to the micro-grid (5 connected health centers in 6 villages). Schools have received much less access to the micro-grids, with 3 public primary schools electrified in 6 villages. Public street lighting has been installed in all villages, but only 2 out of 6 villages provide light all over the night, and one village has decided to stop using their street lamps, for cost-saving reasons. The most prevalent category of collective services connected to the micro-grid concerns churches, which are financed by their members (10 connected churches in 6 villages), suggesting that demand for electrified collective services exists but may be better served by civil society organizations than by public services.

Improved economic development:

Regarding the economic development objective. Figure 5 reveals uneven results on total electricity consumption on the micro-grid showing only 2 villages with electricity consumption growing above 1000 kWh per month and a clear slowing down of the process.

Figure 5: Total electricity consumption (kWh per month)



There are both supply and demand possible explanations for this slowing down.

On the demand side, the principal aspect to consider is electricity demand by businesses, which corresponds by far to the highest share of electricity consumption. The project implementation has not triggered many creations of new businesses yet. Setbacks in relation to the project design reported in the project's theory of change may explain this slow dynamic. The micro-finance partner withdrew from the project and has not been replaced. In few instances, Anka has developed a few short-term credit initiatives to sell electric tools for shelling and crushing of cereals as well as for carpentry. On capacity building, efforts have been equally limited, beyond from training of Anka's staff by ESF. Anka has provided some technical training to the clients of equipment that they sold, but training of potential new VSEs has not been significant. Initiatives in this direction have started only recently, in 2023, in the poultry feed milling sector.

According to Anka's information, there was only a dozen of VSEs using electricity in 2023 in total among the six Café Lumière villages. There were however twice as much households providing commercial electricity-based services (e.g. refrigeration, recharges of mobile phones, computer services. They consumed much less electricity than VSEs but their development suggests that the possibility to access quality electricity may trigger small-scale economic diversification. Scaling-up this process might have been facilitated in presence of micro-finance and more ambitious capacity building initiatives.

On the supply side, the micro-grid capacities are saturated in several villages, not necessarily for lack of power capacity but for lack of adaptability of the load curve to demand across the day, electricity use being highly concentrated at the end of daylight, after agricultural daily activities.

Overall, the implementation data provides evidence of clear evolutions of the theory of change of the project, illustrated by smaller emphasis given to the universal access objective in most recently equipped villages, unresolved issues in promoting public social services and insufficient initiatives aiming at promoting entrepreneurship among VSEs interested in investing in electricity-using activities.

3. Data and Methods

3.1 Survey data on project impacts

Original survey data were collected in May 2017 or May 2018, before implementation, and May 2023 (to respect the seasonality), after project implementation, from 50 households in each equipped village as well as in 6 comparable control villages. Data were collected at the household and/or individual levels, allowing computation of statistical significance level of the observed impacts, which define the main results of our evaluation study.

The survey sample is composed of 599 households for the first survey wave and 595 households for the second wave. From the first wave survey 133 households[27] were not found and were replaced by 129 new households[28]. Our tests of impact are implemented on the balanced panel of 466 households present in both survey waves. As mentioned previously in the project description, these households are scattered around a total of 12 villages (6 treated and 6 non-treated). Within each village the survey was implemented in two areas: the core area, which corresponds to the Fokontany center[29], where the micro-grid infrastructure is built, and the periphery area, which corresponds to its catchment area (hamlets within a radius of 5 km around the Fokontany center). Our stratification strategy aimed at collecting about 60% of households in the core and 40% households in the periphery. In treated villages, the core is the area potentially positively impacted by the whole project, while the periphery is outs of reach of the micro-grid itself but may be impacted by the other components of the project and by spill-over effects.

Our survey has multiple levels questions: localities, households and individuals (adults and children). In this paper we focus our main impact analysis on questions at the household and

[27] 49 located in control villages and 84 in project villages.
[28] 47 located in control villages and 82 in project villages.
[29] A Fokontany is an administrative sub-division which usually groups small villages or hamlet in a common administrative entity.

individual levels. Collected individual data are relatively scarce, except for adult males. We have 423 adult males (99.5% of which are household heads) and only 290 adult females (6.2% of which are household heads).

Universal access to basic electricity services:

Private household electricity access is used to test the net impact of the project on household direct electricity access; which encompasses access through the micro-grid or through standalone systems.

In absence of private access, the objective of universal access to basic electricity services may be met by the direct supply of such services. To evaluate achievements in this direction we consider first the sources of lighting in dwellings. The light sources in dwellings can take several modalities, and we focus specifically on electricity, solar lamps and kerosene lamps. We combine electricity and solar lamps into a single category since both represent electricity-based lighting solutions.

We also consider the development of mobile phones. Mobile phones are becoming increasingly important in rural areas as they can provide a wide array of services, such as modern banking solutions for instance. Given their extensive dissemination in rural areas, recharging phone batteries is more and more considered as a basic electricity service. We analyze the number of phones possessed, the frequency of their recharges and of their usage. Such data are reported at individual level.

Improved public social services:

We consider three major social development dimensions: health, education and security.

To assess the Incidence of illnesses at household level, we utilize the declaration of symptoms of illness experienced by any household member over the past 30 days. The incidence of illnesses could be reduced in Café Lumière villages, notably through improved health care facilities accessing to electricity provided by the micro-grid, used for lighting and refrigeration. Additionally, we examine the quality of water access, as clean water is a key health determinant. Electricity can facilitate water-pumping solutions and boiling for disinfection. For this, we use variables about private access to water and water treatment, which indicate whether households have their own water source and whether they treat their water, either by boiling or using chemicals.

A topical health issue is child and maternal health. Health center electrification may have a positive impact in this respect because it facilitates childbirth in electrified healthcare facilities, providing safer delivery conditions than delivery at home. To test this impact, we consider the birth of children under five occurred in a room with electric lighting. We also consider childbirth at local health center to assess any potential changes in the utilization of local health centers.

To evaluate children vaccination, which could be facilitated by refrigeration availability in healthcare centers, we use the mandated vaccines status, which reflects the vaccination log status of children.

Indoor electrical lighting at school and at home could positively impact school attendance and completion. To test this impact, we rely on a set of three variables: missing school for at least a week, had to retake a school year, and completed the last school year. The first variable provides information on attendance and any significant events that might have hindered the child's school participation during the school year, while the second and third variables pertain to the child’s performance and their ability to successfully complete and graduate from the school year.

Finally, street lighting could improve security. To analyze this, we examine variables related to burglaries experienced by households, including home burglary, home burglary at night, harvest stolen and livestock stolen. These variables indicate whether households suffered such events during the six months preceding the survey.

Improved economic development:

Household wealth is often used as a relevant synthetic indicator to assess the impact on economic development. We built a Wealth index for each household, following the methodology proposed by the Demographic and Health Surveys (DHS). The wealth index reflects the living conditions of the households as well as assets owned. We ran a Multiple Correspondence Analysis (MCA) on dwelling characteristics available in annexes table A1, as well as assets owned listed in table A2. Following DHS practices, we use as the Wealth Index the value of the projected scores on the principal axis from the MCA, which in our sample captures 51% of the inertia. This analysis on the wealth index is complemented by a partial analysis of individual incomes, focusing on secondary incomes, given that the primary activity is predominantly in agriculture and is not directly impacted by electricity (no electricity used for irrigation).

We evaluate the possession of high-power appliances by households, defined by the number of appliances owned requiring a stable and high enough amperage electricity supply from the micro-grid. These appliances include refrigerators, cookers, televisions, DVD or VCR players. They can be used to improve households’ living conditions but they can be used also for economic purposes. To assess the latter usage, we consider a variable available only in the second wave of the survey, which asks households whether they use electricity solely for personal use or also for economic activities.

To crosscheck this analysis, we use data available on individual adult activities, which inform us on the reliance on electricity of primary activities and the number of secondary activities relying on electricity.

To complement this analysis tracking the nexus of improved electricity supply – use of more electric equipment for economic purposes – we analyze the different methods of cereal shelling to which households' resort. Cereal shelling is a very important activity in highland Malagasy rural villages as this transformation is required to provide households with their staple food: shelling can be mechanized or done manually. Mechanized shelling provides significant productivity gains, as manual shelling is otherwise a time-intensive activity. To better understand the context, we also consider the location of shelling (in or outside the village) and the distance traveled to access a shelling machine.

Finally, we evaluate access to information through major media sources using TV use, which measures the frequency of TV usage per week, and information from TV, which tracks households reporting TV as a source of information, and "Radio use," which records the weekly frequency of radio usage.

3.2 Estimation strategy

With RCT designs, treatment effects could be theoretically tested through simple mean comparisons between treated and control units only after treatment. However, since we have both a baseline and a follow-up survey, we can mobilize baseline data to control for potential remaining discrepancies and shortfalls from the randomization. One common estimation method based on two round surveys is a difference-in-differences estimator but it has been demonstrated in the literature that this approach can be too restrictive to detect effects, in particular with a limited sample size (McKenzie, 2012).. Indeed, ANCOVA can increase statistical power compared to difference-in-differences, while also controlling for baseline values which could be sources of bias due to randomization imperfections. Such approach is quite commonly used in RCT designs in rural electrification (Lee et al., 2020) or with solar lighting solutions (Chen et al., 2017; Kudo et al., 2019a), in particular with clustered treatment (Özler et al., 2018).

We therefore decided to apply the ANCOVA approach, which can be modeled as:

$$Y_{i,t} = \beta.Treated_{i,t} + \theta.Y_{i,t-1} + \varepsilon_{i,t} \, (1)$$

Where, $Y_{i,t}$ is a variable of interest (continuous or categorical), β is the Average Treatment effect on the Treated (ATT), $Treated$ is a dummy variable denoting the treatment status, and

$Y_{i,t-1}$ is the value of Y of household *i* (or individual i for questions at the individual level) before intervention (i.e., lagged dependent). We estimate this model at time *t* = 1, which is the post-treatment period. Standard errors of our estimations are all clustered at the randomization level, which is in our case the village level, according to Abadie clustering recommendations for RCT designs (Abadie et al., 2023). Due to the low number of clusters, we also include in the tables p-values computed using a wild cluster bootstrap with Webb ponderation for regressions where we find significant effects using standard clustering. We also explore the impact of clustering at the village level by providing results without clustering in robustness checks.

Given the local distribution of a Café Lumière infrastructure, which is located in the central area of the Fokontany, we surveyed households living in this core area but also households living in nearby hamlets (called periphery area). Since households living in a periphery area cannot be considered as fully treated, we define a specific variable of treatment for them. This allows to investigate any heterogeneity in the results and potential spillover effect from the range of services offered by Café Lumière. Applying this estimation strategy to the ANCOVA model we then have:

$$Y_{i,t} = \beta_1 . Treated_{core_{i,t}} + \beta_2 . Treated_{periphery_{i,t}} + \theta . Y_{i,t-1} + \varepsilon_{i,t} \qquad (2)$$

Where "*Treated_core*" denotes the treatment variable for households living in core areas, taking the value of 1 if the household is located in a Café Lumière village and 0 otherwise. Then, "*Treated_periphery*" is defined as taking the value of 1 for people living in periphery areas of Café Lumière villages and 0 otherwise.

As robustness checks we estimate the same model without the "*Treated_periphery"* variable to only compare households, or individuals, living in the core villages to all other units. This alternative simpler model represents a more classical approach with only one treated group against a control group and make no assumption of any spill-over effects. Given the fact that localities in the periphery areas can only benefit from a subset of the offered services and that our results don't show any signs of spillover effects, we deem this hypothesis quite reasonable and it seems adequate to consider this second model without "*Treated_periphery"* used as a second treatment variable. Considering this simpler model with only one treated group, we are also able to conduct robustness checks using another estimator, in particular the classical difference-in differences estimator. Although this estimator might not be the most optimal estimator for detecting any significant effects given the reasons mentioned previously, we are still able to use it as a robustness given our RCT design.

4. Results

4.1 Universal access to basic electricity services

Table 2 provides information on the impact of the project on private household access to electricity (column 1) and on access to basic electricity services, observed through electricity-based and kerosene-based lighting (columns 2 and 3). We do not see a significant increase in private electricity access in Café Lumière villages compared to control villages due to a substitution effect reducing the use of solar panels as principal source of electricity when household subscribe a micro-grid connection in treated villages, while we saw a strong uptake of individual solar panels in control villages during the period.

Table 2: Private household electricity access and sources of lighting in dwellings

VARIABLES	(1) Private access to electricity	(2) Light source: electricity or solar lamp	(3) Light source: kerosene lamp
Treated_core	0.0996	0.107**	-0.0495
	(0.0594)	(0.0480)	(0.0538)
Treated_periphery	-0.0850*	-0.0518	0.0588
	(0.0408)	(0.0468)	(0.0589)
Lagged dependent	0.411***	0.395***	0.277***
	(0.0533)	(0.0457)	(0.0412)
Constant	0.297***	0.323***	0.150***
	(0.0229)	(0.0198)	(0.0441)
Observations	466	466	466
R-squared	0.178	0.159	0.106
Question level	Households	Households	Households
Clustered p-value Treated_core	0.122	0.0470	0.377
Wild cluster p-value Treated_core	0.172	0.0771	0.417

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, *p<0.1

There is however a positive impact of the Café Lumière on household access to basic electricity services measured by the energy solutions used for lighting, as households living in Café Lumière locations use significantly more electric means of lighting such as electricity supplied by the micro-grid or individual solar panels, or solar lamps. This impact is present in core areas but not in periphery areas of treated villages. Although the use of kerosene lamps is lower in

Café Lumière locations, we do not find a significant difference in their use compared to control locations.

Mobile phone development indicators provide also informative insights on impacts on basic electricity services access (Table 3). The number of phones used by individuals is similar in Café Lumière villages and control villages (column 1). Individuals living in core areas of Café Lumière villages tend to charge their mobile phones more frequently, but this is not the case in periphery areas (column 2). They also use their mobile phone more frequently (columns 3). The impact is unfortunately only significant for male users, although the effect for women is higher than men, it is not significant due to more variation in use and a smaller sample size (columns 4 and 5).

Table 3: Mobile phones possession, recharging and usage

VARIABLES	(1) Number of phones owned (men and women)	(2) Number of charges per week (men and women)	(3) Use of phone more than once a week (men and women)	(4) Use of phone more than once a week (men)	(5) Use of phone more than once a week (women)
Treated_core	0.0803	1.145**	0.188*	0.173*	0.211
	(0.0665)	(0.383)	(0.103)	(0.0936)	(0.157)
Treated_periphery	0.0424	0.689	-0.0296	-0.0357	-0.0285
	(0.0510)	(0.602)	(0.0534)	(0.0509)	(0.0837)
Lagged dependent	0.256***	0.170	0.311***	0.281***	0.317***
	(0.0602)	(0.113)	(0.0568)	(0.0604)	(0.0855)
Constant	0.228***	1.653***	0.345***	0.422***	0.248***
	(0.0303)	(0.267)	(0.0309)	(0.0340)	(0.0409)
Observations	697	110	465	278	187
R-squared	0.068	0.096	0.136	0.120	0.145
Question level	Individuals	Individuals	Individuals	Individuals	Individuals
Clustered p-value Treated_core	0.253	0.0123	0.0952	0.0921	0.222
Wild cluster p-value Treated_core	0.257	0.0270	0.0751	0.101	0.256

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, *p<0.1

4.2 Improved public social services development

Table 4 provides results concerning health of the households and we found no evidence of less symptoms of illness in Café Lumière locations (column 1). We can however notice that during the baseline survey households living in Café Lumière location had significantly more symptoms of illness and they have recorded levels of symptoms similar to other villages during the second wave of the survey. One link usually mentioned in the literature between electricity and health is the use of electric solutions for cooking. However, we did not find any households using electricity as a mean of cooking during the second wave survey and all surveyed households are using charcoal or firewood. Electricity supplied by Café Lumière did not seem either to impact the quality of water access. The distribution of water access sources between private (wells or private taps) and public (springs or public fountains) sources has not changed (column 2). Treatment of the water, which concerns around 20% of households, is almost entirely done by boiling the water, but we do not see any difference in behavior between Café Lumière villages and others after project implementation (column 3).

We investigated the child vaccinations status and we did not find any difference between locations. In both locations we find that around 98% of under 5 children have their vaccination record up to date. This vaccination coverage rate was similar during the first and second survey waves.

Table 4: Incidence of illness and water access

	(1)	(2)	(3)
VARIABLES	Symptoms of illness	Private water access	Water is treated
Treated_core	0.0545	0.157	-0.0833
	(0.0491)	(0.102)	(0.0544)
Treated_periphery	-0.0169	0.0900	0.0297
	(0.0294)	(0.103)	(0.0803)
Lagged dependent	0.0517	0.639***	0.125**
	(0.0419)	(0.0872)	(0.0551)
Constant	0.0764***	0.0887*	0.171***
	(0.0155)	(0.0485)	(0.0329)
Observations	466	466	466
R-squared	0.015	0.447	0.032
Question level	Households	Households	Households

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 5 presents results related to childbirth and in particular the impact of electricity on the conditions of childbirth. Since the questions are related to under 5 children, we do not have panel data and therefore do not take into account the levels of the outcome variable from the previous period. Although we find no difference in the number of childbirths at the local health center between Café Lumière and control locations (column 1), we found a significant difference with higher number of childbirths that benefited from electricity in the Café Lumière locations, but only in the core areas (column 2).

Table 5: Birth delivery with electricity

VARIABLES	(1) Childbirth at local health center	(2) Childbirth at an electrified health center
Treated_core	0.138	0.517***
	(0.189)	(0.151)
Treated_periphery	-0.162	0.145
	(0.207)	(0.169)
Constant	0.612***	0.400**
	(0.149)	(0.140)
Observations	93	96
R-squared	0.044	0.184
Question level	Individuals	Individuals
Clustered p-value Treated_core	0.290	0.151
Wild cluster p-value Treated_core	0.390	0.195

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Concerning impact on education, table 6 reveals no significant differences between Café Lumière and control locations in children's school attendance (column 1), completion of school years (column 2), or completion of the previous schooling year (column 3).

Table 6: School attendance and completion

VARIABLES	(1) Missing school for at least a week	(2) Had to retake a school year	(3) Completed the last school year
Treated_core	-0.0190	0.0392	0.00699
	(0.0362)	(0.0930)	(0.00662)
Treated_periphery	0.0186	0.0536	-0.0518
	(0.0429)	(0.0808)	(0.0377)

Constant	0.0800**	0.327***	0.993***
	(0.0266)	(0.0543)	(0.00662)
Observations	303	303	291
R-squared	0.002	0.002	0.032
Question level	Individuals	Individuals	Individuals

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Incidence of thefts

VARIABLES	(1) Home burglary	(2) Harvest stolen	(3) livestock stolen	(4) Home burglary at night
Treated_core	0.0477**	0.0960	0.150	-0.200
	(0.0208)	(0.0792)	(0.0901)	(0.155)
Treated_periphery	0.0313*	0.128	-0.231***	-0.250
	(0.0147)	(0.102)	(0.0735)	(0.179)
Lagged dependent	0.0852	0.0119	0.107	
	(0.0582)	(0.0614)	(0.137)	
Constant	0.00127	0.192***	0.433***	1***
	(0.00436)	(0.0513)	(0.0582)	(7.45e-09)
Observations	466	464	465	15
R-squared	0.033	0.017	0.020	0.021
Question level	Households	Households	Households	Households

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7 does not reveal any relative improvement in security in Café Lumière locations, measured by occurrences of thefts. We even notice a higher occurrence of home burglary in Café Lumière locations (column1) but no difference in theft of harvest (column 2) or livestock (column 3). If we look further in the cases of home burglary, we see that the few cases of burglary do not appear to have incurred more or less frequently at night in Café Lumière locations (column 4).

4.3 Improved economic development

Table 8 presents results related to the wealth of the households and individual secondary incomes. The wealth index shows an increase in wealth for households living in Café Lumiere core area's locations. This increase in wealth seems however uncertain as it is only significant at the 10% level (column 1). We also investigate individual incomes; we find a slight increase of income from secondary activity in core areas of Café Lumière villages, (column 2) in particular for male individuals (column 3). It should be noted that results on income are based on a very limited sample of individuals due to the low number of respondents to this question. Regarding secondary activity only 119 males and 51 females declared to have a secondary activity out of respectively 407 males and 290 females, but few of them were able or willing to declare revenue generated from this secondary activity.

Table 8: Household wealth index and secondary individual income

	(1)	(2)	(3)
VARIABLES	Wealth index	Income from secondary activity (male and female)	Income from secondary activity (male)
Treated_core	0.221*	80,885*	99,153*
	(0.105)	(40,344)	(50,147)
Treated_periphery	-0.0897	58,810	58,120
	(0.0840)	(37,893)	(39,941)
Lagged dependent	0.786***	0.692***	0.876***
	(0.0418)	(0.190)	(0.0581)
Constant	0.110*	55,409*	49,570*
	(0.0542)	(29,479)	(23,053)
Observations	466	43	31
R-squared	0.536	0.323	0.444
Question level	Households	Individuals	Individuals
Clustered p-value Treated_core	0.0599	0.0760	0.0794
Wild cluster p-value Treated_core	0.0901	0.0571	0.0791

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 9 displays results about electricity use for economic purposes at household level. Connection to the micro-grid allows household to own and use electric appliances requiring sustained electric high power, which could hardly be powered by individual solar panels. Those increases in high-power electric usages are only significant for people living within the Café Lumière core area locations and not those living in the periphery areas (column 1). Correlatively, households living in the Café Lumière core area locations declare to have significantly more use of electricity for economic activity (column 2), and on the contrary households living in the periphery areas do not use electricity for an economic purpose.

Regarding the use of electricity in economic activities of individuals (Table 10), the use of electricity seems slightly higher for the primary economic activity of individuals living in core areas of Café Lumière villages with a positive coefficient only significant at the 10% level (column 1). By contrast, although the sample size is quite limited, we found a more positive and significant effect for the use of electricity in secondary economic activity of men (column 2), but not of women (column 3).

Table 9: Households high-power appliances possession and economic uses of electricity

VARIABLES	(1) Use high power appliances	(2) Economic use of electricity
Treated_core	0.219**	0.111**
	(0.0941)	(0.0432)
Treated_periphery	0.0151	-0.0778**
	(0.0711)	(0.0324)
Lagged dependent	0.482***	
	(0.108)	
Constant	0.135***	0.102***
	(0.0392)	(0.0235)
Observations	463	264
R-squared	0.211	0.041
Question level	Households	Households
Clustered p-value Treated_core	0.0402	0.0258
Wild cluster p-value Treated_core	0.0511	0.0140

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Note: column 2 does not control for lagged dependent, as this question was not in pre-treatment survey.

Table 10: Individual activities and their electricity reliance

VARIABLES	(1) Electricity use for primary activity	(2) Electricity use for men's secondary activity	(3) Electricity use for women's secondary activity
Treated_core	0.0880*	0.238**	-0.273
	(0.0490)	(0.0884)	(0.278)
Treated_periphery	-0.00102	0.0362	0.121
	(0.0261)	(0.134)	(0.223)
Lagged dependent	0.443**	-0.167	0.727***
	(0.144)	(0.118)	(0.166)
Constant	0.0240**	-2.203***	-0
	(0.0103)	(0.244)	(0.400)
Observations	459	55	27
R-squared	0.095	0.073	0.173
Question level	Individuals	Individuals	Individuals
Clustered p-value Treated_core	0.100	0.0226	0.365
Wild cluster p-value Treated_core	0.262	0.0260	0.536

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

We also observe a positive impact in the processing and transformation of agricultural products (Table 11). We found that households living in Café Lumière locations are using significantly less manual shelling (column 1). Mechanized shelling done locally, which can be enabled by local electricity access is more frequent in core areas of treated villages (column 2). By contrast, mechanized shelling is done less frequently outside the village (column 3) and correlatively, the distance covered by a household to access a husker tend to be slightly shorter for households living in Café Lumière core area locations (column 4).

Table 11: Choices of cereal shelling by households

VARIABLES	(1) Manual shelling	(2) Local mechanized shelling	(3) Outside village mechanized shelling	(4) Distance to shelling machine
Treated_core	-0.211**	0.318***	-0.0870***	-0.279*
	(0.0929)	(0.0979)	(0.0249)	(0.136)
Treated_periphery	-0.141	0.0281	0.120	0.687
	(0.108)	(0.110)	(0.0839)	(0.587)
Lagged dependent	0.442***	0.368***	0.0922	0.195***
	(0.0710)	(0.0847)	(0.0738)	(0.0484)
Constant	0.334***	0.238**	0.0798***	0.494***
	(0.101)	(0.0791)	(0.0239)	(0.118)
Observations	445	445	445	217
R-squared	0.253	0.234	0.073	0.247
Question level	Households	Households	Households	Households
Clustered p-value Treated_core	0.0441	0.00773	0.00501	0.0644
Wild cluster p-value Treated_core	0.0601	0.0140	0.00500	0.0621

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 12 provides results concerning access to media and information, which can be an important driver for development. We can see individuals living in core areas of Café Lumière villages watch more frequently the television (column 1), while there is no difference in the frequency of radio use (column 2), which is low powered media carrier. If we investigate the type of media viewed on the television, we see that more people are getting information from television in core areas of Café Lumière (column 3).

Table 12: Access to information

VARIABLES	(1) TV use	(2) Radio use	(3) Information from TV
Treated_core	0.122**	0.0278	0.0688***
	(0.0503)	(0.0749)	(0.0122)
Treated_periphery	-0.00464	0.00688	-0.0126
	(0.0434)	(0.0645)	(0.0104)
Lagged dependent	0.293***	0.133**	0.176*
	(0.0606)	(0.0483)	(0.0849)
Constant	0.0549	0.345***	0.0161**
	(0.0307)	(0.0579)	(0.00726)
Observations	697	697	697
R-squared	0.152	0.013	0.049
Question level	Individuals	Individuals	Individuals
Clustered p-value Treated_core	0.0332	0.717	0.000152
Wild cluster p-value Treated_core	0.0501	0.734	0.00500

Standard errors clustered at the village level in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5. Discussion

Universal access to basic electricity services:

Overall, we do not find positive impact on private electricity access, due to a substitution effect between standalone systems and micro-grid connections. There are however some indications of a positive impact of the project on access to basic electricity services, which can be procured through battery recharge services supplied in the Café Lumière boutique or in other shops. This is particularly relevant for the use of mobile phones, which has significantly increased in the core areas of treated villages. Robustness checks on private electricity access however show that if we consider a model without the "*Treated_periphery*" variable, the effect becomes significant at the 5% level for households living in the treated villages. Findings concerning use of modern lighting solutions and phone charges are confirmed by the robustness model, but not by the DiD estimator.

Improved social development:

On social development impacts we observe positive impact, particularly on maternal and child healthcare in core areas of treated villages., thanks to the electrification of health care centers, offering safer conditions for birth delivery. This is consistent with our locality survey data

(Table 13), which show that the proportion of electrified primary health care centers (csb2) has increased much more in treated than in control villages, even though treated villages were disadvantaged initially. This is true both for the existence of indoor lighting and for the availability of refrigeration. These findings of improvement in conditions of childbirth are also confirmed by robustness checks on the alternative simple treatment model.

By contrast, we do not observe impact on education, neither on security.

Table 13: Electrification of csb2 by locality and survey wave

	Share of csb2 with electric lighting			Share of csb2 with refrigeration		
	1rst survey wave	2nd survey wave	Difference over time	1rst survey wave	2nd survey wave	Difference over time
Treated	33%	100%	+67%	17%	83%	+67%
Non treated	75%	75%	0%	50%	75%	+25%
Difference treated vs. non-treated	-42%	25%	+67%	-33%	8%	+42%

Improved economic development:

The limited changes in the household's wealth index and income could be explained by the short period between surveys and potential impacts on wealth could take a long time to be fully measurable. Although this changes on measures of wealth and income are mellow in magnitude, they are quite consistent across several outcomes and other changes are more pronounced. We can see for instance use of electricity for economic purposes are more prevalent in villages where the project Café Lumière has been implemented.

Although results on the increase of the Wealth index appear limited, they are confirmed in our robustness checks by both the alternative model using treated villages only and the DiD estimator, while results on secondary income are only confirmed by the DiD estimator. However, results regarding usage of high-power electric appliances, economic use of electricity, as well as its use in a secondary activity remain positive and significant with the alternative model or the DiD estimator, suggesting that although the effect is hard to measure quantitatively in terms of income, it is more robust qualitatively (Annex Tables A10 and A11).

Having the possibility to access distributed electricity helps individuals to conduct electrified non-agricultural activities. This could in turn help generating more revenue and increase wealth in a more significant manner in the longer term. Although the primary activity, which is farming for 90% of the people, is not directly impacted by electricity access, the ability of individuals to use electricity in secondary activities may be key for generating new incomes.

This is visible in the positive impact on the household use of electricity for economic purposes, and correlatively of high-power appliances.

Although our survey is limited to the observation of impacts from the viewpoint of household data, we find similar results from the viewpoint of businesses in the subsector of cereal shelling, which is central in rural villages of the Malagasy highlands region: there is a positive impact on the local supply of mechanized shelling in Café Lumière locations.

Our locality survey provides complementary descriptive information on businesses impacts. We observe that shops and dinners' electrification progress during the period in treated villages (Table 14), but also that the number of such businesses declines. A similar observation is obtained from data on services activities, which encompasses all services but shops and dinners (e.g., shelling, carpentry, small mechanical reparation, tailor, hairdressing, video club) (Table 15). The reduction of the number of observed businesses in Café Lumière villages is likely the mere result of their higher economic efficiency compared to un-electrified businesses, in a sort of Schumpeterian process of creative destruction.

Table 14: Share of electrified shops and dinners and number of observations

	Share of electrified shops and diners			Number of shops and diners		
	1rst survey wave	2nd survey wave	Difference over time	1rst survey wave	2nd survey wave	Difference over time
Treated	12%	83%	+71%	197	140	-57
Non -treated	11%	4%	-7%	87	141	83
Difference treated vs. non-treated	1%	79%	+78%	110	-1	-140

Table 15: Share of electrified services and number of observations

	Share of electrified services providers			Number of services providers		
	1rst survey wave	2nd survey wave	Difference over time	1rst survey wave	2nd survey wave	Difference over time
Treated	86%	100%	+14%	83,5	68	-15,5
Non-treated	91%	48%	-43%	11	27	+16
Difference treated vs. non-treated	-5%	52%	+57%	72,5	41	-31,5

Note: numbers are averages of dry and wet seasons

6. Conclusion

The Café Lumière project has demonstrated clear improvements in specific areas of electricity access of off-grid rural communities, with positive economic and social consequences on treated villages compared to non-treated villages. More households gained access to basic electricity services in treated villages. More household and entrepreneurs accede to higher quality electricity services in terms of reliability and power availability, leading to greater penetration of electricity in craft and services. The economic transformation process that could result from this change is however still underway, as is the Schumpeterian creative destruction process under way in VSEs. More entrepreneurs use electricity, but the creation of new activities is lagging behind. This may be due to the failure of the project to deliver on micro-finance support, and also to insufficient capacity building in entrepreneurship, as these two dimensions were identified in the theory of change as integral parts of the project but were underperformed.

Social development indicators, such as education, health and public safety, show limited impact. The only noticeable exception concerns the impact of electrification of healthcare centers. Specifically, electrified healthcare centers have increased the proportion of childbirths benefiting from electricity, contributing directly to maternal and child health development.

Overall, the jury is out concerning the expected long-term dynamics of economic and social transformation, which is still uncertain at this point. Another evidence of the incompleteness of successes obtained so far is that these successes are limited to the core area of treated villages and we did not measure any positive spillover effects on locations at their periphery. This absence of spillover effect is consistent with our analysis of the limitations of the economic transformation process triggered by mere increased and improved electricity access.

That being said, although the time interval between our first and second waves of survey is long (7-8 years), the inception of the project has been slow, and retarded by the COVID-19 crisis. Consequently, we measure with this survey actually only 2-3 years of impact of the project, which is not long enough to derive final conclusions.

References

Abadie, A., Athey, S., Imbens, G.W., Wooldridge, J.M., 2023. When Should You Adjust Standard Errors for Clustering?*. Q. J. Econ. 138, 1–35. https://doi.org/10.1093/qje/qjac038

Aklin, M., Bayer, P., Harish, S.P., Urpelainen, J., 2017. Does basic energy access generate socioeconomic benefits? A field experiment with off-grid solar power in India. Sci. Adv. 3, e1602153. https://doi.org/10.1126/sciadv.1602153

Banerjee, A., Duflo, E., Kremer, M., 2019. Understanding Development and Poverty Alleviation. Sver. Riksbank Prize Econ. Sci. Mem. Alfred Nobel 2019, The Committee for the Prize in Economic Sciences in Memory of Alfred Nobel.

Berthelemy, J.-C., Millien, A., 2018. Impact of Decentralized Electrification Projects on Sustainable Development: A Meta-Analysis. FERDI Work. Pap. P240.

Chen, A.Z., Fischer, J., Fraker, A., Shah, N.B., Shirrell, S., Stein, D., 2017. Welfare impacts of an entry-level solar home system in Uganda. J. Dev. Eff. 9, 277–294. https://doi.org/10.1080/19439342.2017.1307248

Dinkelman, T., 2011. The Effects of Rural Electrification on Employment: New Evidence from South Africa. Am. Econ. Rev. 101, 3078–3108. https://doi.org/10.1257/aer.101.7.3078

Duran, A.S., Sahinyazan, F.G., 2021. An analysis of renewable mini-grid projects for rural electrification. Socioecon. Plann. Sci. 77, 100999. https://doi.org/10.1016/j.seps.2020.100999

Duthie, M., Ankel-Peters, J., Mphasa, C., Bhat, R., 2024. The elusive quest for sustainable mini-grid electrification: New evidence from Indonesia. Energy Sustain. Dev. 80, 101454. https://doi.org/10.1016/j.esd.2024.101454

Kudo, Y., Shonchoy, A.S., Takahashi, K., 2019a. Short-Term Impacts of Solar Lanterns on Child Health: Experimental Evidence from Bangladesh. J. Dev. Stud. 55, 2329–2346. https://doi.org/10.1080/00220388.2018.1443207

Kudo, Y., Shonchoy, A.S., Takahashi, K., 2019b. Can Solar Lanterns Improve Youth Academic Performance? Experimental Evidence from Bangladesh. World Bank Econ. Rev. 33, 436–460. https://doi.org/10.1093/wber/lhw073

Lee, K., Miguel, E., Wolfram, C., 2020. Experimental Evidence on the Economics of Rural Electrification. J. Polit. Econ. 128, 1523–1565. https://doi.org/10.1086/705417

McKenzie, D., 2012. Beyond baseline and follow-up: The case for more T in experiments. J. Dev. Econ. 99, 210–221. https://doi.org/10.1016/j.jdeveco.2012.01.002

Moore, N., Glandon, D., Tripney, J., Kozakiewicz, T., Shisler, S., Eyres, J., Zalfou, R., Leon, M., Kurkijan, V., Snilstveit, B., Perdana, A., 2020. Systematic Review on the Impact of Access to Electricity on Household Welfare (ADB Independent Evaluation). ADB.

Nicolas, C., Samson, B., Rozenberg, J., 2019. Meeting the Sustainable Development Goal for Electricity Access—Using a Multi-Scenario Approach to Understand the Cost Drivers of Power Infrastructure in Sub-Saharan Africa. World Bank, Policy Research Workin Paper.

Özler, B., Fernald, L.C.H., Kariger, P., McConnell, C., Neuman, M., Fraga, E., 2018. Combining pre-school teacher training with parenting education: A cluster-randomized controlled trial. J. Dev. Econ. 133, 448–467. https://doi.org/10.1016/j.jdeveco.2018.04.004

Acknowledgements

This study has been implemented thanks to ESF, the developer of the Cafés Lumière project, who has shared with us primary information on the design and implementation of the Project and to its operator, Anka Madagascar who has shared with us its operational reports. The Cabinet Conforme, an independent survey company, has collected the survey data necessary to study the impacts of the project, under joint supervision by ESF ad FERDI. We have also benefited from the deep knowledge of these partners on the local context in which this study has been performed. We are however the sole responsible of this study conclusions, which we have developed in full independence.

Funding

This research was jointly supported by FERDI and Electriciens sans Frontière under MOU signed 24/02/2023 guaranteeing its independence.

This work was supported by the Agence Nationale de la Recherche of the French government through the program "France 2030" (grant number ANR-16-IDEX-0001).

Annexes

Table A1: Dwelling Characteristics

What is the nature of the roof?
What is the nature of the walls?
How many floors does the dwelling have?
How many rooms does the dwelling have (living rooms: bedroom, kitchen, living room)?
Is the kitchen located inside the dwelling?
What is the housing status of the dwelling?
What type of toilets does the household have?
What is the main source of water supply for the household?
Is the water of good quality (colorless, odorless, tasteless)?
Does the household have land to cultivate?
Does the household own the land it cultivates?

Table A2: Assets owned

Table	Cart
Chair	Car/truck
Bed or mattress	Radio
Wardrobe/closet/trunk	Television
Watch/clock	VCR
Sewing machine	DVD player
Stove	Computer (desktop or laptop)
Refrigerator or freezer	Tablet
Fan	Landline phone
Bicycle	Mobile phone
Moped/motorcycle	

Table A3: Robustness on private household electricity access and sources of lighting in dwellings.

VARIABLES	(1) Private access to electricity	(2) Private access to electricity	(3) Light source: electricity or solar lamp	(4) Light source: electricity or solar lamp	(5) Light source: kerosene lamp	(6) Light source: kerosene lamp
Treated_core	0.122**		0.122**		-0.0656	
	(0.0554)		(0.0434)		(0.0432)	
Lagged dependant	0.406***		0.390***		0.272***	
	(0.0555)		(0.0471)		(0.0411)	
DID		0.0498		0.0329		0.0638
		(0.0676)		(0.0676)		(0.0577)
Constant	0.276***	0.308***	0.311***	0.278***	0.167***	0.488***
	(0.0244)	(0.0501)	(0.0208)	(0.0441)	(0.0356)	(0.0412)
Observations	466	932	466	932	466	932
R-squared	0.174	0.032	0.157	0.047	0.104	0.051
Method	ANCOVA	DiD	ANCOVA	DiD	ANCOVA	DiD
SE clustered	Yes	Yes	Yes	Yes	Yes	Yes
Question level	Households	Households	Households	Households	Households	Households

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table A4: Robustness on mobile phones possession, recharging and usage.

VARIABLES	(1) Number of phones owned (men and women)	(2) Number of phones owned (men and women)	(3) Number of charges per week (men and women)	(4) Number of charges per week (men and women)
Treated_core	0.0689		0.945**	
	(0.0619)		(0.336)	
Lagged dependant	0.254***		0.180	
	(0.0600)		(0.118)	
DID		-0.00259		2.430
		(0.0573)		(2.020)
Constant	0.240***	0.292***	1.826***	4.447**
	(0.0217)	(0.0327)	(0.181)	(2.008)
Observations	697	1,394	110	446
R-squared	0.067	0.008	0.080	0.004
Method	ANCOVA	DiD	ANCOVA	DiD
SE clustered	Yes	Yes	Yes	Yes
Question level	Individuals	Individuals	Individuals	Individuals

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table A4 (continued): Robustness on mobile phones possession, recharging and usage

VARIABLES	(5) Use of phone more than once a week (men and women)	(6) Use of phone more than once a week (men and women)	(7) Use of phone more than once a week (men)	(8) Use of phone more than once a week (men)	(9) Use of phone more than once a week (women)	(10) Use of phone more than once a week (women)
Treated_core	0.196*		0.183*		0.218	
	(0.0939)		(0.0891)		(0.141)	
Lagged dependant	0.310***		0.280***		0.318***	
	(0.0570)		(0.0599)		(0.0861)	
DID		0.0852		0.0806		0.0970
		(0.0947)		(0.108)		(0.124)
Constant	0.337***	0.320***	0.413***	0.379***	0.240***	0.243***
	(0.0262)	(0.0411)	(0.0265)	(0.0376)	(0.0347)	(0.0511)
Observations	465	1,146	278	668	187	478
R-squared	0.135	0.053	0.119	0.055	0.145	0.053
Method	ANCOVA	DiD	ANCOVA	DiD	ANCOVA	DiD
SE clustered	Yes	Yes	Yes	Yes	Yes	Yes
Question level	Individuals	Individuals	Individuals	Individuals	Individuals	Individuals

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Table A5: Robustness on incidence of illness and water access

VARIABLES	(1) Symptoms of illness	(2) Symptoms of illness	(3) Private water access	(4) Private water access	(5) Water is treated	(6) Water is treated
Treated_core	0.0590		0.132		-0.0910*	
	(0.0446)		(0.0782)		(0.0465)	
Lagged dependant	0.0511		0.645***		0.125**	
	(0.0419)		(0.0885)		(0.0559)	
DID		-0.0953		0.117*		-0.0702
		(0.0643)		(0.0560)		(0.0716)
Constant	0.0721***	0.156***	0.110**	0.305***	0.179***	0.251***
	(0.0132)	(0.0292)	(0.0388)	(0.0840)	(0.0354)	(0.0389)
Observations	466	932	466	932	466	932
R-squared	0.015	0.045	0.443	0.034	0.031	0.011
Method	ANCOVA	DiD	ANCOVA	DiD	ANCOVA	DiD
SE clustered	Yes	Yes	Yes	Yes	Yes	Yes
Question level	Households	Households	Households	Households	Households	Households

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A6: Robustness on birth delivery with electricity

	(1)	(2)	(3)

VARIABLES	Child vaccination is up to date	Childbirth at local health center	Childbirth at an electrified health center
Treated_core	0.0281	0.185	0.472***
	(0.0397)	(0.146)	(0.120)
Constant	1.015***	0.565***	0.444***
	(0.0161)	(0.118)	(0.102)
Observations	88	93	96
R-squared	0.007	0.028	0.170
Method	ANCOVA	ANCOVA	ANCOVA
SE clustered	Yes	Yes	Yes
Question level	Individuals	Individuals	Individuals

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A7: Robustness on school attendance and completion

	(1)	(2)	(3)
VARIABLES	Missing school for at least a week	Had to retake a school year	Completed the last school year
Treated_core	-0.0250	0.0220	0.0237
	(0.0375)	(0.0951)	(0.0152)
Constant	0.0860***	0.344***	0.976***
	(0.0210)	(0.0427)	(0.0152)
Observations	303	303	291
R-squared	0.002	0.000	0.007
Method	ANCOVA	ANCOVA	ANCOVA
SE clustered	Yes	Yes	Yes
Question level	Individuals	Individuals	Individuals

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A8: Robustness on incidence of thefts

VARIABLES	(1) Stolen at home	(2) Stolen at home	(3) Harvest stolen	(4) Harvest stolen	(5) Animals stolen	(6) Animals stolen	(7) Stolen at home at night
Treated_core	0.0397*		0.0630		0.210**		
	(0.0197)		(0.0567)		(0.0898)		
Lagged dependant	0.0841		0.0309		0.116		
	(0.0596)		(0.0520)		(0.137)		
DID		0.0223		0.0306		0.167*	
		(0.0240)		(0.0683)		(0.0763)	
Constant	0.00937	0.0329**	0.222***	0.126***	0.373***	0.0719***	0.800***
	(0.00592)	(0.0119)	(0.0466)	(0.0279)	(0.0512)	(0.0204)	(0.155)
Observations	466	932	464	930	465	931	15
R-squared	0.027	0.009	0.005	0.022	0.012	0.067	0.000
Method	ANCOVA	DiD	ANCOVA	DiD	ANCOVA	DiD	ANCOVA
SE clustered	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Question level	Households	Households	Households	Households	Households	Households	Households

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A9: Robustness on household wealth index and secondary individual income

VARIABLES	(1) Wealth index	(2) Wealth index	(3) Income from secondary activity (male and female)	(4) Income from secondary activity (male and female)	(5) Income from secondary activity (male)	(6) Income from secondary activity (male)
Treated_core	0.244**		53,198		67,553	
	(0.106)		(46,551)		(54,021)	
Lagged dependant	0.786***		0.646***		0.842***	
	(0.0421)		(0.189)		(0.0566)	
DID		0.236*		108,712**		133,636*
		(0.110)		(45,376)		(66,136)
Constant	0.0867*	-0.0923	85,459**	129,716***	82,794**	136,502***
	(0.0448)	(0.0831)	(32,204)	(23,644)	(28,210)	(25,800)
Observations	466	932	43	306	31	225
R-squared	0.535	0.031	0.300	0.033	0.425	0.041
Method	ANCOVA	DiD	ANCOVA	DiD	ANCOVA	DiD
SE clustered	Yes	Yes	Yes	Yes	Yes	Yes
Question level	Households	Househol ds	Individuals	Individuals	Individuals	Individuals

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A10: Robustness on household's high-power appliances possession and economic uses of electricity

VARIABLES	(1) Use high power appliances	(2) Use high power appliances	(3) Economic use of electricity
Treated_core	0.215**		0.130**
	(0.0843)		(0.0445)
Lagged dependant	0.482***		
	(0.108)		
DID		0.125*	
		(0.0691)	
Constant	0.139***	0.153***	0.0824***
	(0.0317)	(0.0281)	(0.0212)
Observations	463	929	264
R-squared	0.211	0.035	0.035
Method	ANCOVA	DiD	ANCOVA
SE clustered	Yes	Yes	Yes
Question level	Households	Households	Households

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table A11: Robustness on individual activities and their electricity reliance

VARIABLES	(1) Electricity use for primary activity	(2) Electricity use for primary activity	(3) Electricity use for men's secondary activity	(4) Electricity use for men's secondary activity	(5) Electricity use for women's secondary activity	(6) Electricity use for women's secondary activity
Treated_core	0.0883		0.234**		-0.299	
	(0.0506)		(0.0918)		(0.236)	
Lagged dependant	0.443**		0.0147		0.727***	
	(0.143)		(0.0519)		(0.163)	
DID		0.0849		0.550*		0.230
		(0.0476)		(0.299)		(0.242)
Constant	0.0237**	0.0129*	-1.841***	-1.307***	0.0260	-1.882***
	(0.00996)	(0.00607)	(0.0879)	(0.292)	(0.369)	(0.132)
Observations	459	1,137	55	272	27	116
R-squared	0.095	0.036	0.071	0.097	0.168	0.080
Method	ANCOVA	DiD	ANCOVA	DiD	ANCOVA	DiD
SE clustered	Yes	Yes	Yes	Yes	Yes	Yes
Question level	Individuals	Individuals	Individuals	Individuals	Individuals	Individuals

Robust standard errors in parentheses; *** p<0.01, ** p<0.05, * p<0.1

Supplementary material: Note on sampling and data collection

Selection of the villages

The regions of Vakinankaratra and Itasy were not only selected because of their low access to electricity but also because of their proximity and relatively easy access from the capital city, Antananarivo.

First, villages were preselected based on objective criteria, national official data and meetings with the population. The selection criteria were the following:

- distance to the national grid: 10 to 15 km)
- population: minimum of 150 households in the Fokontany center (around 750 inhabitants)
- No existing electrification plan from the national electricity company (Jirama) or any other electrification projects
- Economic potential
- Spatial rationale: idea to consider cluster of villages to ease the management by the operator
- Local motivation

Based on these preliminary criteria, a short list of 18 villages was identified.

In these 18 villages, a more detailed demand analysis was conducted to sharpen the economic potential of the village, assess the electrification needs (both productive and domestic) of the local population, as well as of the public services.
This demand analysis aimed at making sure to implement only economically sustainable electrified services in the villages.

In the end, 12 villages were considered to have a full potential to host a Café Lumière.

Out of these 12 villages, 6 were randomly selected to effectively receive a Café Lumière. The remaining 6 villages are considered as control villages.

Due to administrative, budgetary and logistic reasons, baseline data collection prior to any project implementation focused on 6 villages (2 treated and 4 controls) in May 2017. The remaining 6 villages (4 treated and 2 controls) were investigated in May 2018 (to respect seasonality of the survey and comparability of the collected data).

Table S1. List of villages

Villages	Regions	Latitude	Longitude	Status	Year of installation
Ambatonikolahy	Vakinankaratra	-19.766667	46.921111	Treated	2018
Ambohimalaza	Vakinankaratra	-19.931944	46.858611	Treated	2019
Antanamalaza	Vakinankaratra	-19.427500	47.627500	Treated	2020
Antsampandrano	Vakinankaratra	-19.545000	47.592778	Treated	2020
Amparaky	Itasy	-19.308056	46.630000	Treated	2021
Dondona	Itasy	-19.198889	46.874167	Treated	2020
Ambatoasana	Vakinankaratra	-19.341111	46.763611	Control	-
Ambatolahy	Vakinankaratra	-19.732222	47.286389	Control	-
Ambatomiady	Vakinankaratra	-19.661667	47.425556	Control	-
Anivosaha	Vakinankaratra	-19.968611	46.957222	Control	-
Fenoarivo	Vakinankaratra	-19.552222	47.250556	Control	-
Mandrosohasina	Vakinankaratra	-19.577500	47.178333	Control	-

Selection of households

A household was considered to be all the members (related or not) who "permanently" shared the same living space under the authority of a head of household. We therefore considered as members people who sleep and eat at least once a day in the accommodation when they are present in the locality.
We considered "permanently" the fact to be present for more than 6 months of the year in the household.

A sensitivity analysis coupled with time and budget constraints led us to consider a sample of 50 households per village (hence a total target sample of 600 households). Among these 50 households, 30 were randomly selected in the center of the Fokontany and 20 were randomly selected in its catchment area (any hamlets within a radius of 5 km around the Fokontany center). This stratification strategy aimed at capturing any spill-over effects within the surrounding communities.

Selection of the households of the center of the Fokontany

The choice of the 30 households surveyed in the center of the Fokontany was made randomly and stratified according to the distance from the middle of the center of the Fokontany, so that households were distributed equitably between the different zones of the center of the Fokontany.

The itinerary method was used to select the households surveyed. It consists of defining the route followed by the investigator as follows:

1. All the investigators place themselves in the center of the Fokontany, each chooses the direction indicated by the pen that they have previously thrown into the air.
2. Each investigator moves to the limit of the center of the Fokontany following this direction. He/she counts the number of households arranged along this axis between its starting point (the middle of the center of the Fokontany) and its arrival point, the limit of the center of the Fokontany (last dwelling, see Figure S1).
3. He/she then divides the number of households counted by the number of households he/she must survey. Take the case where there are 20 households between the middle of the center of the Fokontany and the boundary and where an interviewer must interview 5 households: 20/5 is equal to 4.
4. Starting from the limit of the center of the Fokontany where he/she is located, the interviewer goes back towards the middle and surveys the 4th household encountered. Then, he/she counts 4 again and investigates the 8th household encountered, etc. stopping when he/she has surveyed his five households.

Figure S1 illustrates by example the selection of households to be surveyed in the case where there are 3 interviewers and they each have 5 households to interview.

Interviewer A counts 14 households up to the limit of the center of the Fokontany, 14/5 is equal to 2 in absolute value, he/she interviews every second household on the way back to the middle.

Interviewer B counts 20 households up to the boundary of the center of the Fokontany, 20/5 equals 4. He/she surveys the 4th household, then counts 4 to the next, then 4 to the next, etc., until he has interviewed five households.

Interviewer C counts 16 households up to the limit of the center of the Fokontany, 16/5 is equal to 3 in absolute value, he interviews the 3rd household then counts 3 until the next one.

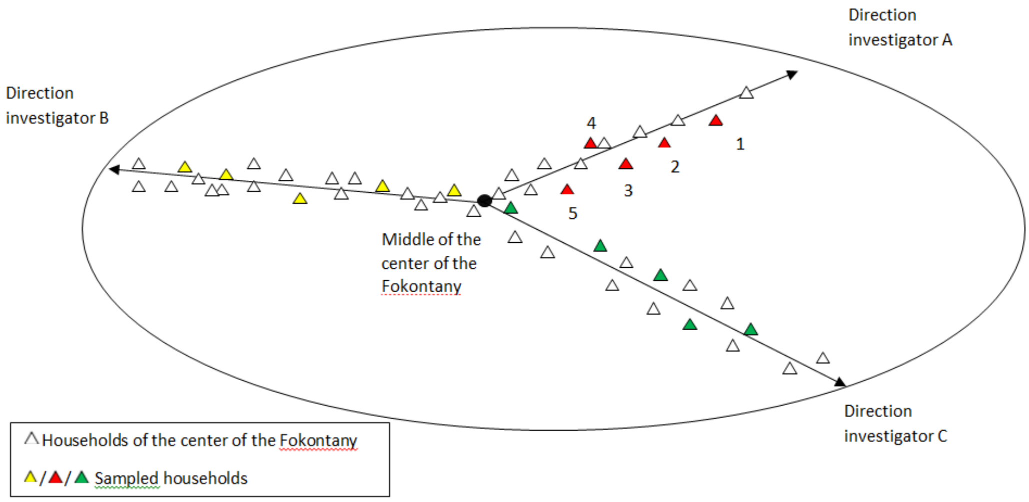


Figure S6. Itinerary method to select households in the center of the Fokontany

Selection of the households of the catchment area

As for the households in the center of the Fokontany, the choice of the 20 households surveyed in the catchment area was made randomly and stratified according to the distance from the center of the Fokontany, so that the households were distributed at different distances from the center of the Fokontany.

A preliminary mapping exercise was conducted by a FERDI cartographer prior to the baseline surveys. For each village, using satellite images, a map of the center of the Fokontany and its catchment area (5 km) highlighting habitation zones was prepared (see example below).

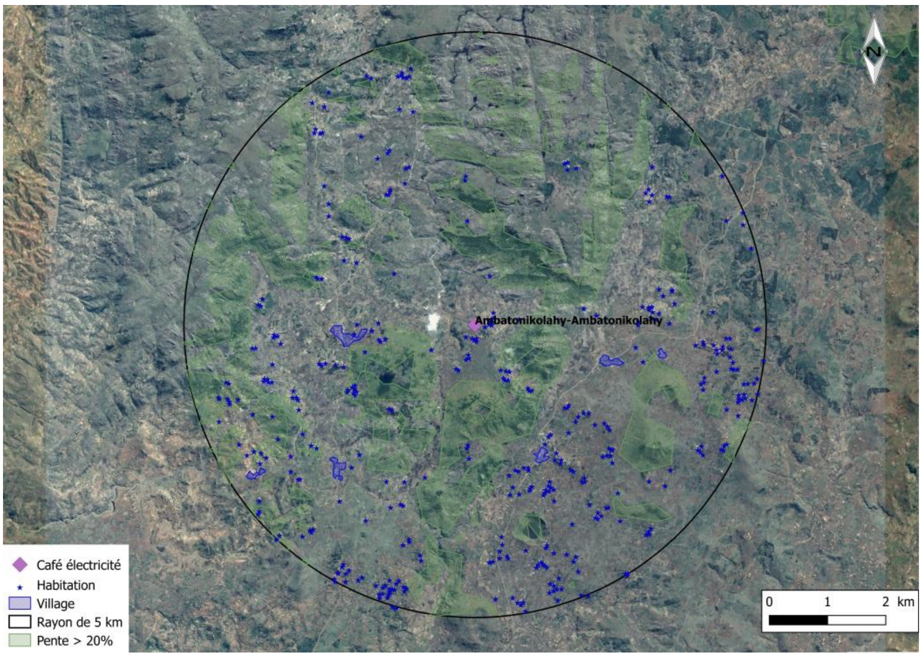


Figure S7. Map of the settlement areas of the village of Ambatonikolahy and its catchment area

Once in the village, the team of investigators validated the accuracy of these maps with the local leaders.

Then, as for the center of the Fokontany, each of the investigators randomly chose a direction marked by a road, a track or a path between the directions where they were more likely to find habitation zones at different distances from the center of the Fokontany.

For instance, for the village of Ambatonikolahy above, the North, North East directions were not considered in the random selection, since the likelihood to find settlements was limited.

<u>Maps of other Fonkotanys surveyed during 2017</u>

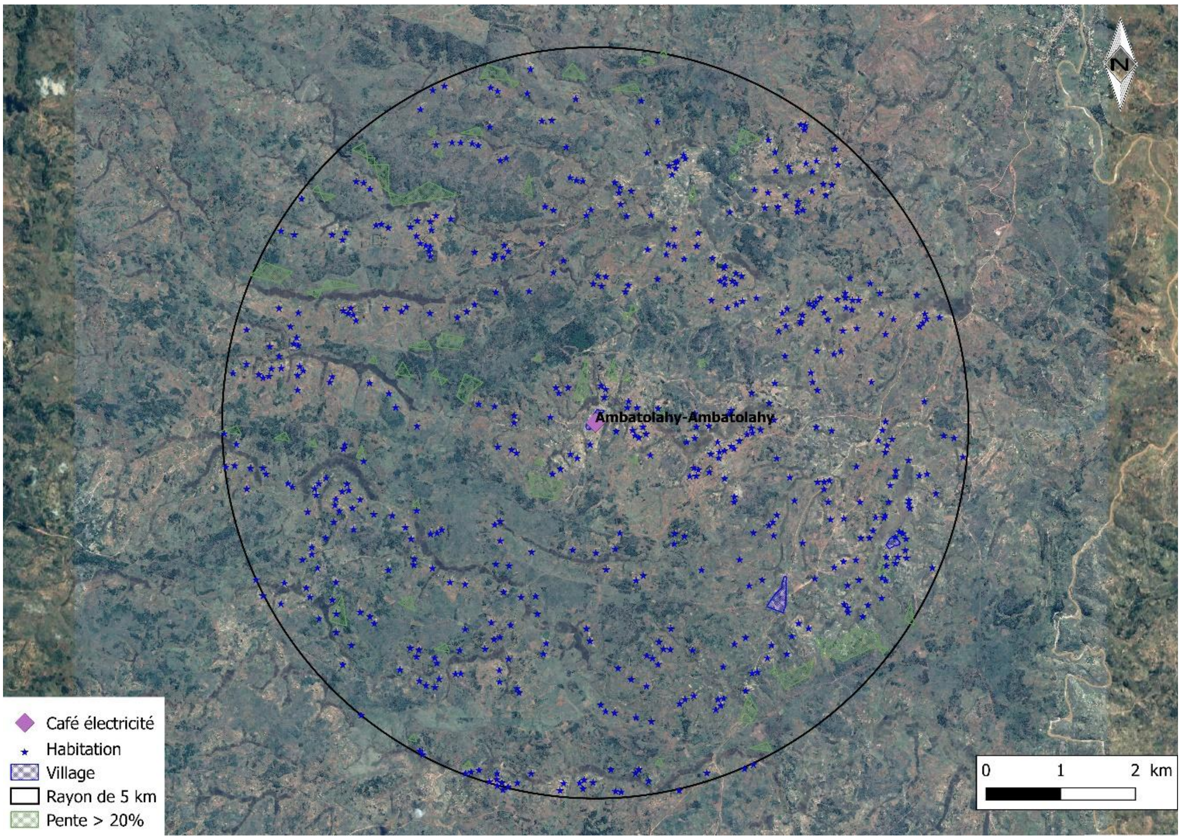


Figure S3: Map of the settlement areas of the village of Ambatolahy and its catchment area.

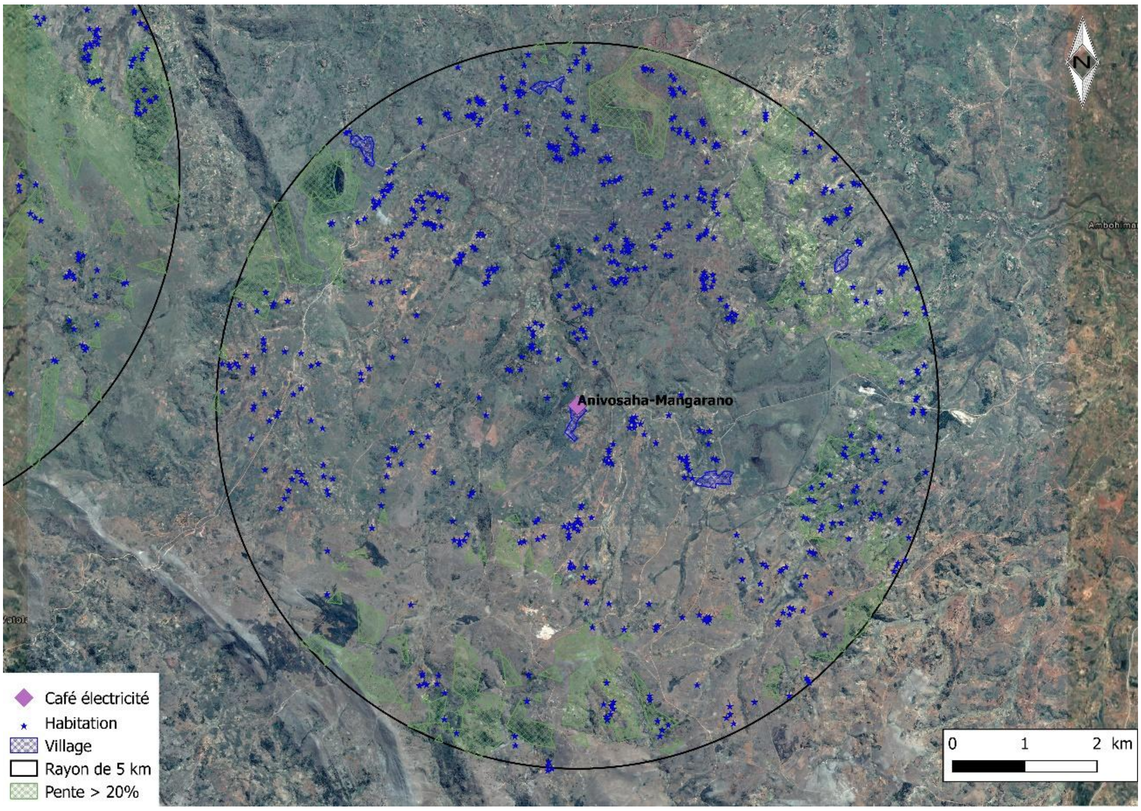


Figure S4: Map of the settlement areas of the village of Anivosaha and its catchment area.

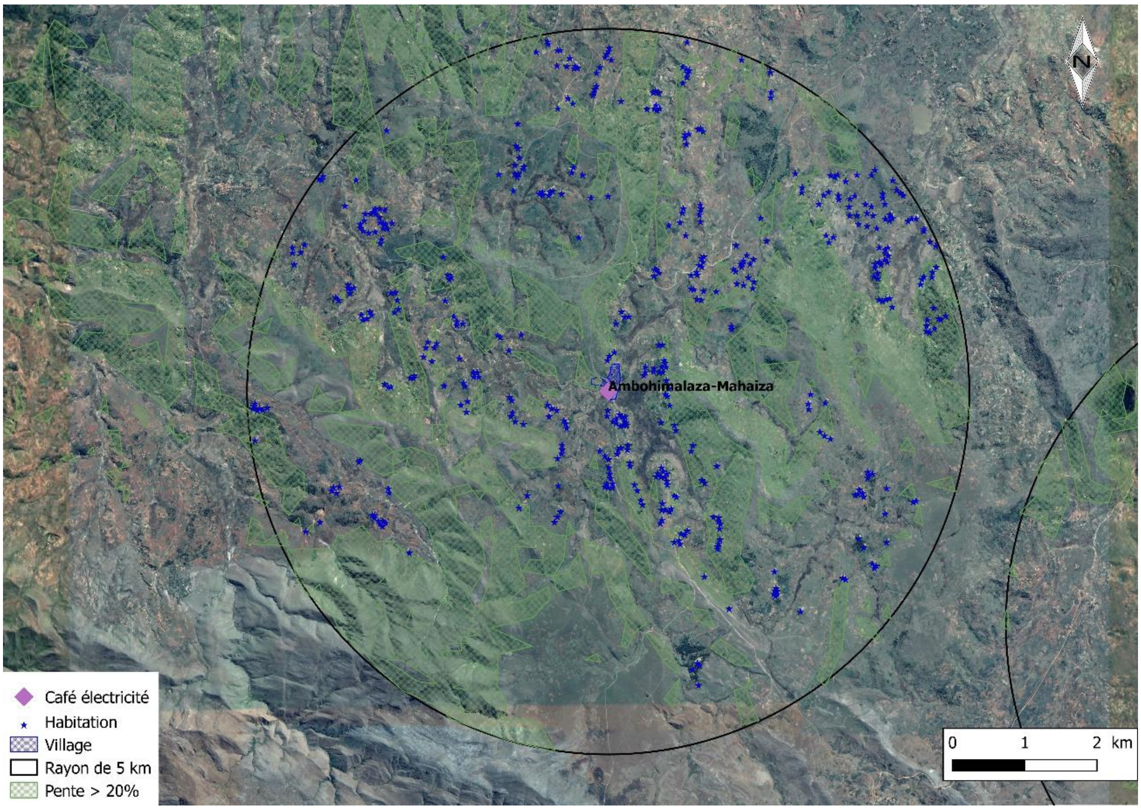


Figure S5: Map of the settlement areas of the village of Ambohimalaza and its catchment area.

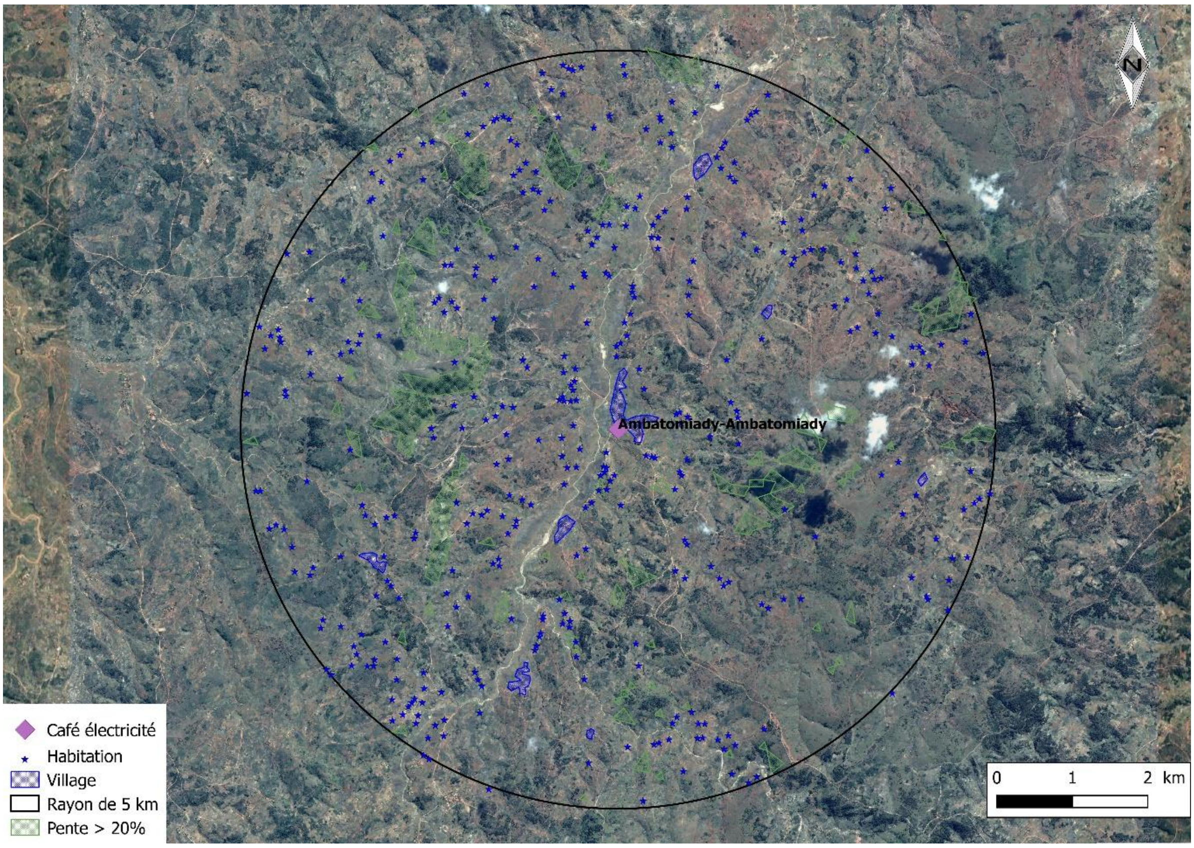


Figure S6: Map of the settlement areas of the village of Ambatomiady and its catchment area.

Questionnaires

Household questionnaire

The household questionnaire was designed to identify and measure as many transmission channels (beneficial or harmful) of electrification by Café Lumière on the living conditions of households as possible.
For this, the questionnaire covered a wide range of the household characteristics while emphasizing the aspects directly impacted by the installation of Café Lumière: health, education and economic activity.

The questionnaire was sequenced in 5 stages:

- The head of household and/or any other member of the household likely to answer correctly were questioned about the general characteristics of the household and its members as follows:
    - Identification section listing all the household members with key demographic information (gender, age, relation to head of household) as well as literacy and education levels;
    - Maternity and vaccination section dedicated to gather information on the children below 5 years old;

  - Education section dedicated to all the school-aged children (5-15 years old) whether they were at school or not;
  - Health section listing all the household members who suffered from any diseases in the last 30 days;
  - Housing, Land and Security section to investigate the characteristics of the house (roof, walls, floor), its equipment (water, toilets, etc.), as well as the land use of the households and the security issues in the last 6 months;
  - Durable goods section listing the durable assets of the household (chair, table, bed, television, motorbike, etc.) and their cattle and farm animals;
  - Energy sections detailing all the domestic energy use of the households (including cooking), and the productive energy use of the households (transforming cereals and information and technology services);
  - Food consumption section detailing the ingredients and their origins (purchase, own production etc.) of the regular meals eaten by the households in the morning, lunch and dinner times;
- The head of household was interviewed privately (in an isolated manner) about his personal living conditions as follows:
  - income generating activities and their revenues;
  - community and citizen life (member of community groups, participation in elections etc.);
  - savings;
  - access to credit;
  - access to information;
  - use of mobile phone;
  - personal security issues in the last 6 months;
  - timetable of a common day listing all activities from waking up to going to bed.
- The spouse of the head of household was interviewed privately (in an isolated manner) about her personal living conditions. The questions were the same as for the head of the household, except for some additional questions on maternity and contraception, as well as on the decision-making power within the household for different expenses;
- If the composition of the household permitted it, a school-aged boy (5-15 years), whether he was in school or not, was interviewed privately (in an isolated manner) about his personal living conditions (access to information, school homework, security, consumption of meat, fish or fresh drinks, timetable), a schooling test was also administered to him;
- If the composition of the household permitted it, a school-aged girl (5-15 years), whether she was in school or not, was interviewed privately (in an isolated manner) about her personal living conditions (access to information, school homework, security, consumption of meat, fish or fresh drinks, timetable), a schooling test was also administered to her.

At baseline, 599 households were interviewed representing 3,086 people. The following people were interviewed individually (in an isolated manner):

- 519 male head of households
- 547 women (30 being head of households and 516 being spouses to the head of households)
- 205 school-aged boys (5-15 years)
- 180 school-aged girls (5-15 years)

595 households were interviewed during the second wave of surveys. From the baseline survey 133 households were not found and were replaced by 129 new households living close to where the not found households of the baseline used to live.
In the second wave, the following people were interviewed individually (in an isolated manner):

- 527 male head of households
- 369 women (25 being head of households and 344 being spouses to the head of households)
- 214 school-aged boys (5-15 years)
- 185 school-aged girls (5-15 years)

Locality questionnaire

The locality questionnaire was intended to assess the dynamism, infrastructure, economy, etc. of the center of the Fokontany and its catchment area. It was administered to local resource people. Resource people in the locality were the president of the fokontany, the school director, the director of the health center, the teacher or any influential person who had good knowledge of the situation in their village.

The locality questionnaire was structured as follows:

- Localization, accessibility and electrification section;
- Demographic and commercial activities section;
- Community life (groups, associations etc.) section;
- External partners (present or past funded projects) section;
- Water infrastructures section;
- Education infrastructures section;
- Health infrastructures section;
- Services offered in the locality (cereal transformation, hair salon, tailor etc.) section;
- Security section listing all the security issues in the last 3 years;
- Catchment area section listing all the hamlets within a radius of 5 kilometers and their key characteristics (demographics, economic activities, infrastructures, etc.).

Questionnaire outline (french) :

- Informations générales :
  - Identification des membres permanents du ménage
  - Maternité et Vaccination
  - Éducation
  - Santé
- Ménages :
  - Logement Terre Sécurité
  - Statut d'occupation
  - Téléphone - eau – assainissement
  - Accès à la terre – valorisation
  - Sécurité
  - Actifs et biens durables
  - Energie
  - Sources de lumière
  - Combustible
  - Consommation Alimentaire
  - Services productifs
  - Services informatifs
- Individus : homme/femme :
  - Religion – ethnie
  - Activités économiques personnelles
  - Vie sociale - citoyenne
  - Economie - pouvoir de decision
  - Crédit
  - Accès à l'information – ouverture
  - Téléphone
  - Sécurité
  - Emploi du temps
- Individus : garçons/filles :

Général

Emploi du temps

Mathématiques

Lecture

Chapter 3:
Estimating Electricity Access Impact of Chinese Energy Projects in Africa Using Machine Learning and Remote Sensing Data.

Vincent Nossek, CERDI, CNRS, IRD, Université Clermont Auvergne

Abstract

Access to electricity data is typically available only at a national level and often relies on non-annual survey data. To address this limitation, our study introduces a novel approach combining multi-source survey data used as ground truth data with remote sensing information to produce annual subnational estimates of electricity access. We train a machine learning (ML) model to predict electricity access rate for the African continent plus Madagascar and then predicts yearly electricity access rates for each 0.07-degree grid cell (approximately 7 km) for the period 2013–2022. These predicted raster maps of electricity access across Africa inform us on remote and underserved rural areas that are typically sparsely surveyed due to the dispersed population. These subnational estimates offer valuable insights for policymakers and energy-sector decision-makers. To demonstrate the utility of this data, we conduct a case study evaluating the impact of geolocated energy infrastructure projects financed by the Chinese government and a subset of World Bank's projects. The analysis finds positive and significant effects of these projects on electricity access in treated grid cells, with an effect increasing over time up to at least 8 years after commitment. We estimate energy projects to have a local impact ranging between +2 to +8 percentage points, depending on the financers, in electricity access five years after commitment. The dynamic observed indicate a slow and steady adoption of electricity in areas where projects where implemented. This case study highlights how combining ML and remote sensing data can inform progress made toward achieving Sustainable Development Goal 7 and support the evaluation of policies and investments aimed at improving energy access in specific regions.

Keywords: Machine learning; Remote sensing; Electricity access; Infrastructure; Geospatial data; Impact evaluation; Africa; Chinese investments; World Bank investments.

1. Introduction

Electricity has been a critical driver of economic activity since the Industrial Revolution. However, access to electricity remains limited in Sub-Saharan Africa, where only 50%[30] of the population had access in 2021. This aggregate figure masks significant disparities between urban and rural areas: while 80% of urban residents have access to electricity, this figure drops to just 30% in rural areas. The United Nations Sustainable Development Goal (SDG) 7, which aims to "ensure access to affordable, reliable, sustainable, and modern energy for all," has highlighted the urgent need to improve electricity access while addressing the environmental constraints facing the world today. Achieving SDG 7 presents unique challenges for African countries, including a high rural population—approximately 58%[31] in 2022—which necessitates extensive deployment of new grid and off-grid infrastructure. However, efficient infrastructure planning and impact evaluation of electrification initiatives require robust data and evidence-based decision-making.

At the heart of this paper is the question of how to evaluate the impacts of electrification projects effectively, notably energy projects financed by China or the World Bank, two main providers of infrastructures financing in Africa. Overall Chinese aid impact have been studied in the literature (Dreher et al., 2021) but sector specific and local level impact is much more rare due to lack of comprehensive local data, although project data has been geocoded (Goodman et al., 2024). The challenge to evaluate energy projects impact lies in the scarcity of detailed subnational data on electricity access in the case of energy projects, which hinders accurate tracking of progress toward SDG 7 and limits the ability to assess the effectiveness of energy investments. Addressing this data gap is critical to enable tracking and evaluation of these projects and optimizing future interventions.

Efforts to address this data gap include the work of Falchetta et al. (2019), who produced high-resolution (1 km) electricity access datasets for Sub-Saharan Africa using nighttime lights (NTL) and population data (Falchetta et al., 2019). However, their approach relied on a number of strict assumptions about the relationship between NTL intensity and electrification rates and used field surveys for validation rather than calibration. Other researchers have combined geo-referenced survey data with satellite imagery to calibrate predictive models for wealth estimation. Advances in machine learning (ML), particularly in ease of use and implementation, have enabled significant progress in this area. Jean et al. (2016), for example, demonstrated the potential of combining field data with remote sensing to generate new datasets for unsurveyed areas or time periods. Much of the literature employing this approach has focused on wealth estimation, using DHS surveys and finding strong predictive

[30] https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS?locations=ZG
[31] https://data.worldbank.org/indicator/SP.RUR.TOTL.ZS?locations=ZF&view=chart

performance for ML algorithms based on daytime and nighttime satellite imagery (Chi et al., 2022; Ratledge et al., 2022; Yeh et al., 2020).

Raw satellite imagery, while valuable, presents challenges due to its high resolution, RGB (red, green, and blue) structure, and large volume. High-resolution daytime imagery is also costly, especially for extensive geographic coverage over multiple time periods, and requires computationally intensive algorithms such as convolutional neural networks (CNNs), including architectures such as ResNet-18 algorithm (Li et al., 2022; Sherman et al., 2023). Although many studies adopt complex methodologies and datasets, they often show that simpler approaches can achieve comparable predictive performance (Yeh et al., 2020). For instance, Yeh et al. (2020) demonstrated that NTL data alone can predict local wealth nearly as effectively as a combination of NTL and daytime imagery. This suggests that the added complexity of daytime imagery may not always be necessary, particularly when focusing on specific measurements like electricity access.

Based on this literature and the approach proposed by Dhorne et al. (Dhorne et al., 2021), which trains a ML model to link ground truth measurement of electricity access from the Multi-Tier Framework (MTF) surveys with remote sensing data, our approach also calibrates ML models using electricity access rates as the target variable. We however rely on a wide range of geo-localized surveys such as Demographic and Health Surveys (DHS), Living Standards Measurement Study (LSMS), Afrobarometer, and Multiple Indicator Cluster Surveys (MICS) and a slightly different set of remote sensing data. Once calibrated, the ML model predicts yearly electricity access rates for each 0.07-degree grid cell (approximately 7 km) for the period 2013–2022, generating detailed raster maps of electricity access across Africa. We apply this generated data in a case study evaluating the impact of electrification projects financed by the Chinese government and a subset of World Bank projects. Using geolocated data on project sites, we identify treated and control grid cells and employ matching and difference-in-differences estimators to assess the local-level impacts of these initiatives.

This paper contributes to the literature on remote sensing and machine learning used to predict ground truth data at a local level and demonstrates how ML can be used to generate data on electricity access and enable impact evaluations of electrification projects. Our approach simplifies data requirements and computational demands while maintaining predictive accuracy. Building on the work of ESMAP researchers (Dhorne et al., 2021), we focus exclusively on African countries, leveraging an extensive set of household surveys and remote sensing variables for model calibration. Our findings highlight reduced systematic biases in predicted electricity access values compared to earlier studies and emphasize the importance of generating data for rural areas, where gaps in subnational electricity access data are particularly pronounced. By filling these gaps, we enable the generation of consistent panel datasets for previously unsurveyed areas, years, and countries not covered by existing surveys. This paper also contributes to the impact evaluation literature for the energy sector by documenting the effects of projects financed by the Chinese government or the World Bank thanks to the generated data on electricity access. We assess projects effects on electricity access rates as well as its dynamic over time, highlighting a slow but steady and increasing adoption of electricity access in areas close to the deployment of electric infrastructures. Overall, this application demonstrates how ML-driven datasets can directly inform the evaluation of energy projects, bridging the gap between data generation and actionable insights for policymakers.

This paper is organized as follows: Section 2 describes the data sources used for ML training and prediction and the case study on project infrastructures. Section 3 outlines the methodology for ML modeling, prediction performances of the trained models and the external validation of predicted results. Section 4 presents the methodology and results of the impact evaluation on infrastructure projects. Section 5 discusses the findings from the previous results sections and section 6 conclude on the findings.

2. Data

2.1 Machine learning data

Our target variable (i.e. dependent variable) is the electricity access rate computed at the cluster level. To compute local access rates, we mobilize multiple geo-localized survey sources namely DHS (US AID), LSMS (World Bank), MICS (UNICEF) and Afrobarometer. We collected a total of 155 surveys covering the period 2013-2022.

All these surveys combined amount to a total of 1 825 099 surveyed households distributed over 58 736 geo-localized cluster-year units over the whole period. We compute the average access rates at the geo-localized cluster-year unit level. Figure 1 presents a heatmap of the geographical distribution of the clusters, all periods considered. We can clearly see that areas with the most clusters are primarily capitals and major cities[32]. Although we cover most of the countries over the period no geo-localized surveys were found for 8 countries, namely Algeria, Libya, South Sudan, Somalia, Djibouti, Eritrea, Central African Rep. and Rep. of Congo[33]. As displayed in Table 1, DHS surveys are the most represented over time, allowing us to cover the all period. For LSMS surveys most of the clusters that we could retrieve containing questions concerning electricity access where from year 2018, when LSMS launched a harmonized survey on households living standards in multiple countries. UNICEF have been providing geo-localized data for their MICS surveys quite recently. Only 3 countries are available through MICS for our study, Malawi, Madagascar and Nigeria. Finally, with

[32] High number of clusters in Malawi is due to the presence of all four surveys sources present in this country. Burundi and Rwanda appear to have a high number of clusters but this can be attributed to the small size of those countries and therefore the high concentration of clusters.

[33] We provide maps for each data sources in annexes.

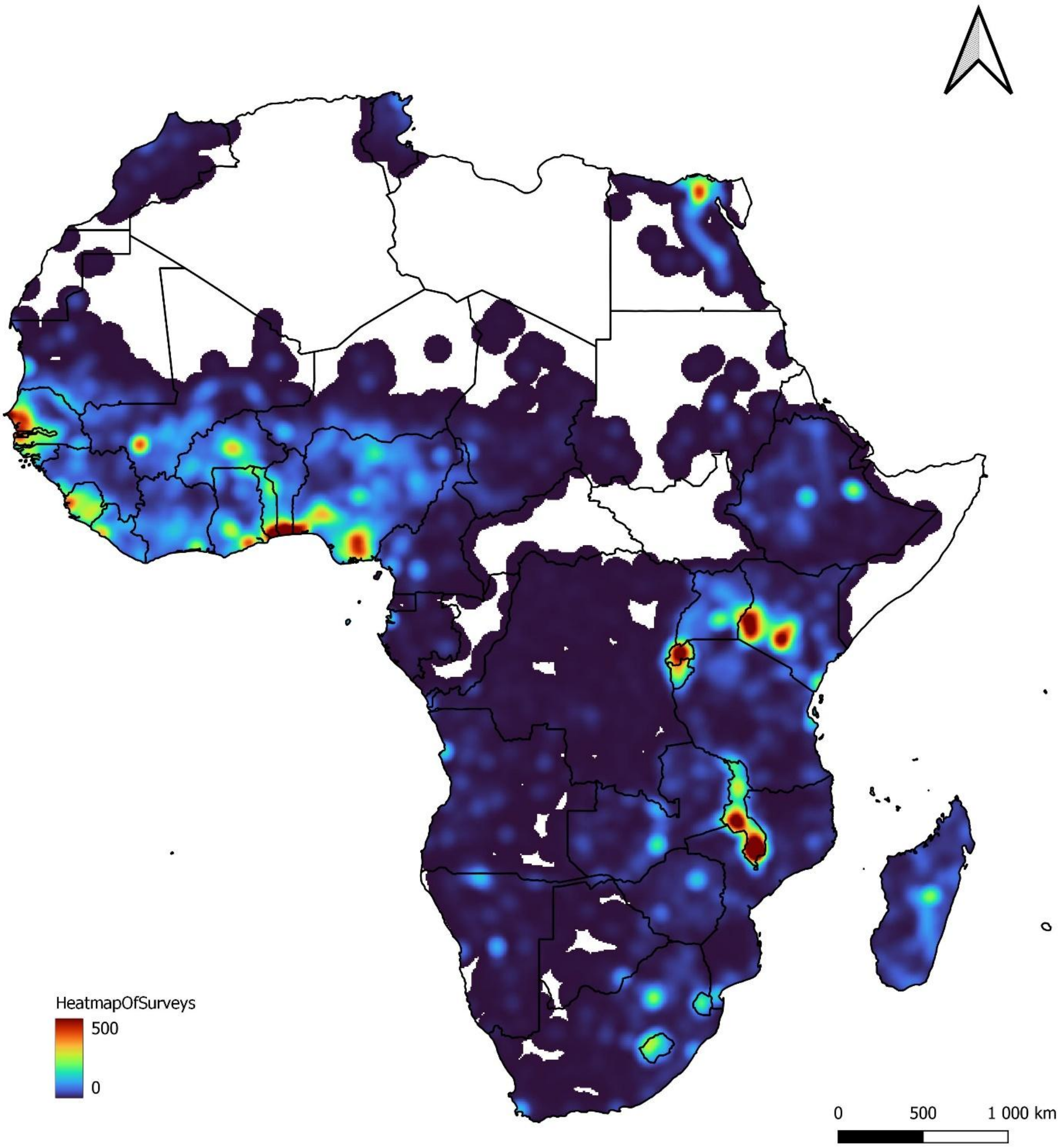
Afrobarometer data we mobilized all surveys from round 7 and round 8 which covers years from 2017 to 2022, with 2017 and 2021 being the main years of surveys round for respectively round 7 and 8.

Table 1: Temporal distribution of geo-localized survey clusters

Year	Survey source				Total
	DHS	LSMS	MICS	Afrobarometer	
2013	4 300	0	0	0	**4 300**
2014	6 418	0	0	0	**6 418**
2015	3 772	518	0	0	**4 290**
2016	3 151	780	0	0	**3 931**
2017	3 629	953	0	4 259	**8 841**
2018	3 892	5 447	774	1 200	**11 313**
2019	2 412	717	151	1 196	**4 476**
2020	1 732	516	965	450	**3 663**
2021	2 869	0	2 079	3 965	**8 913**
2022	1 691	0	0	900	**2 591**
Total	**33 866**	**8 931**	**3 969**	**11 970**	**58 736**

If we look in total over all the survey sources, years 2017, 2018 and 2021 are the ones with the most observations, and DHS is by far the main source of observations for our study, followed by Afrobarometer and LSMS.

Figure 1: Heatmap of the survey clusters from DHS, LSMS, MICS and Afrobarometer.



Concerning our predictor variables, we use a range of variables mostly derived from remote sensing and geospatial information. Table 2 displays all those variables with their characteristics and sources. The specific variables set is composed of 13 variables selected for their respective relevance on the domains of infrastructures development and electricity access. This predictor variables are the following: VIIRS Nighttime Light (NTL) indicative of luminosity emission; Land Cover, such as the share of built-up, croplands, grasslands, forests and water areas; mean of the Normalized Difference Vegetation Index (NDVI) indicating vegetation health; distance to main roads; distance to grid-network; mobile network availability (2G and 3G); gas flares locations; industrial mines locations; distance to cities (cities defined as locations with more than 50k habitants), altitude; gradient; population and urban or rural areas. We selected this variable set based on the literature on electricity access and infrastructure deployment. For each of the variables we selected sources that would give us the best spatial and temporal coverage.

Table 2: Spatial variables

Predictor variables	Resolution	Temporality	Variable type	Source(s)
Nighttime Light VIIRS	500 m	2013-2022	Continuous	EOG, Colorado School of Mines
Land Cover	100 m	2015-2019	Discrete	Copernicus
NDVI	250 m	2013-2022	Continuous	MODIS (MOD13Q1), USGS
Distance from main roads	1 km	2021	Continuous	OpenStreetMap
Distance to Grid-network	1 km	2018	Continuous	World Bank
Mobile network availability	250 m	2013-2021	Discrete	GSMA
Gas Flares locations	(vector)	2012-2022	Discrete	EOG, Colorado School of Mines
Industrial mines locations	(vector)	2010-2019	Discrete	Minex & US Geological Survey
Distance to cities	1 km	2015	Continuous	Nelson et al., 2019
Altitude	90 m	2018	Continuous	SRTM, CGIAR-CSI
Gradient	90 m	2018	Continuous	SRTM, CGIAR-CSI
Population	1 km	2013-2021	Continuous	LandScan, Oak Ridge Lab.
Urban or Rural areas	1 km	2015	Discrete	SEDAC, NASA

Nighttime light is widely used in the literature on electricity access due to its very strong correlation with electricity consumption(Elvidge et al., 1997; Shi et al., 2014). Also VIIRS data, compared to DMSP, allows for detection of very low level of lights in rural areas (Elvidge et al., 2013, 1997; Falchetta et al., 2019) which is quite important to produce accurate mapping access over the African continent. Finally VIIRS data have demonstrated temporal variability allowing to help in predict the evolution in economic activity and therefore electricity consumption (Beyer et al., 2022; Elvidge et al., 2020). NDVI and land cover classification data, also called land use, informs on the land characteristics and is a pre-processed version of raw satellite imagery that synthetizes the main relevant information on land characteristics. We also rely on the existing infrastructures such as main roads, grid networks and mobile coverage, in particular 2G and 3G coverage, which are critical determinants for physical or informational connectivity of a given area. We also consider the presence of mining and extraction activities by using the location of known and referenced industrial mines as well as gas flares generated at oil extraction sites. Industrial mining and extracting activities are usually dependent on energy access and can emit quite high levels of NL while not reflecting the true population access to electricity living in the area. Gas flares emit luminosity unrelated to electricity use and can introduce bias in predicted values of electricity access if not considered. We also consider distance to cities, defined as urban areas of more than 50,000 people, to consider the impact of human activities and network effects around major cities.

Altitude and gradient are variables that inform on the local geography and the difficulty to install and maintain infrastructures, such as in hilly or mountainous regions. Finally, population and rural or urban areas inform on the density of population and potential for human infrastructures and activity levels.

2.2 Case study data

To investigate the impact of Chinese's energy infrastructure projects we use the AidData's Geospatial Global Chinese Development Finance Dataset, Version 3.0 (Custer et al., 2023; Goodman et al., 2024). This dataset contains geolocated Chinese projects supported by official financial and in-kind commitments (or pledges) from China between 2000 and 2021, and features a total of 9405 projects across 148 low and middle-income countries supported by Chinese grant and loan commitments. We retrieve from this dataset only projects related to the energy sector. We also remove projects that are not located precisely and keep only projects with a precision level identified at the ADM6 level or under. After those filtering steps and taking into account the temporality of our electricity access data, namely panel data over the period 2013-2022, we are able to evaluate 108 energy projects, located over 36 countries. We consider as a treatment date the commitment year of the project[34]. Using commitment rather than completion year ensure no missing data and allows for comparability with World Bank projects, which only document commitment dates. All Chinese financed projects amount to a total of 17 709 million USD of investment during the period.

Although AidData did provide a similar dataset for World Bank projects, it only contains projects implemented before 2013, which is the beginning of our predicted dataset. We therefore had resorted to use another dataset to carry a comparison by mobilizing geolocated projects data collected by the World Bank's Geo-Enabling Initiative for Monitoring and Supervision (GEMS) team[35]. Although this data is not comprehensive at a continental or a country level and does not encompass the whole universe of World Bank's projects, it allows us to compare to some extent the effects of energy projects financed by the World Bank. Our sample from the GEMS data provides 95 locations of projects located in 4 countries.[36] We can notice that although we have a big difference in the number of countries covered, the overall number of project locations evaluated is quite similar. We consider as treatment date the commitment year. All World Bank projects amounts to a total of 1 070 million USD in

[34]We provide in robustness results where we consider completion date as treatment for Chinese funded projects. For a few projects with a missing completion date, but with a commitment date, we compute the average project duration commitment and completion (3 years) and add it to the commitment date to extrapolate the expected completion year.

[35] https://www.worldbank.org/en/topic/fragilityconflictviolence/brief/geo-enabling-initiative-for-monitoring-and-supervision-gems

[36] Projects are located in Ethiopia, Côte d'Ivoire, Kenya and Rwanda.

investment. This huge disparity in the amounts of projects evaluated reflect the fact we are working on a subset of World Bank projects which have been geolocated by the GEMS team.

Table 3 presents the classification of projects based on keywords present in the project title and description.

Table 3: Projects characteristics

	Project types	Number of projects	Amount (million USD)
Projects financed by China	Power production non-renewable	36	7 686
	Power production renewable	53	7 912
	Transmission and distribution	19	4 110
	Total	*108*	*19 709*

Note: For Chinese projects each project corresponds to a given location.

	Project types	Number of projects	Number of locations	Amount (millionUSD)
Projects financed by the World Bank	Power production renewable	2	29	650
	Transmission and distribution	2	66	420
	Total	*4*	*95*	*1 070*

To give some examples of energy projects financed by China we can cite: the co-financing of Dadinkowa Hydropower Plant Project in Nigeria, with a power capacity of 39MW representing an investment of USD 32.6 million by Powerchina Huadong Engineering Corporation (Aid data ID: 97786); the Mpika Power Supply improvement project sought to improve the supply of power in Mpika district within Muchinga Province in Zambia, representing an investment of $29.6 million by the Industrial and Commercial Bank of China (ICBC) (Aid data ID: 57556) or a donation of donates solar-powered street lights to three villages in Kgalagadi district (Khuis, Gakhibane, and Makopong) in Bostwana, estimated at $550 thousands (Aid data ID: 91639). On the side of the World Bank, we can give as an example the project Off-grid Solar Access Project for Underserved Counties implemented in Kenya. This project is composed of 4 components, the first component being mini-grids for community facilities, enterprises, and households, the second component is stand-alone solar systems and clean cooking solutions for households, the third component is the stand-alone solar systems and solar water pumps for community facilities and finally the fourth component is the implementation support and capacity building. This project financed by the World Bank amounts to $325 million (World Bank project ID: P157055).

3. Machine Learning and data generation

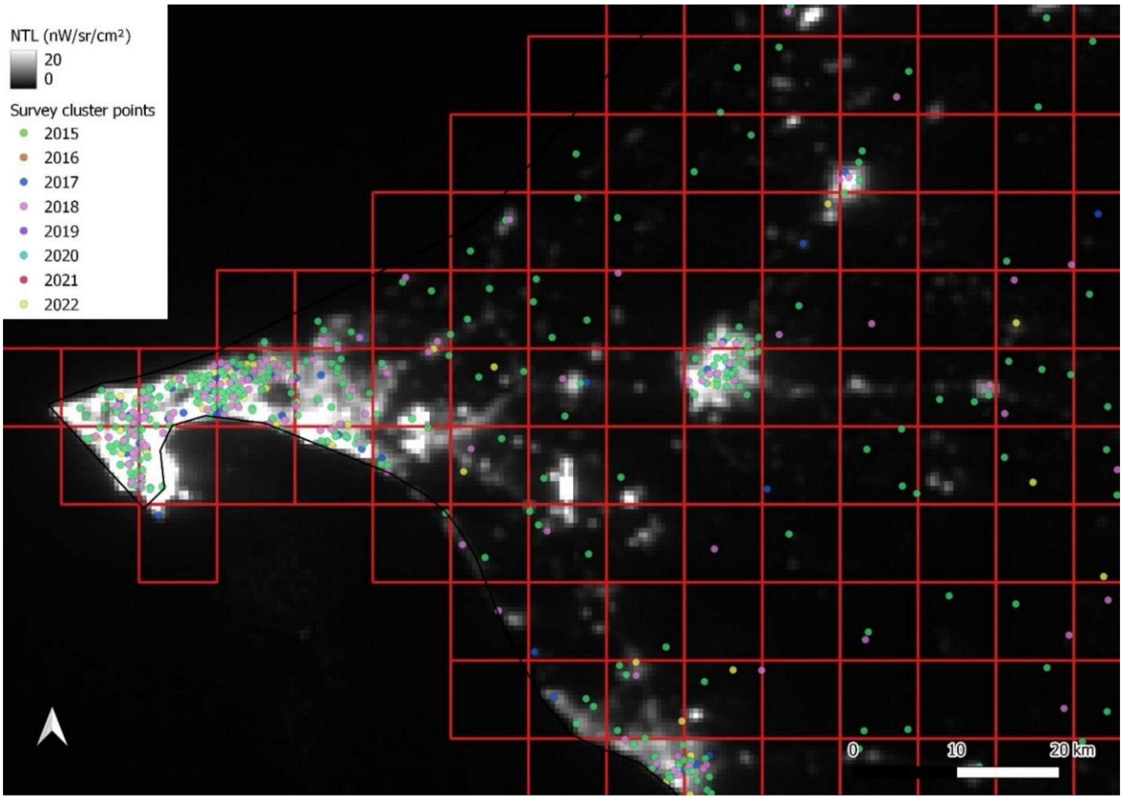
3.1 Methodology

To train the machine learning (ML) models, we first constructed a dataset using a raster grid over the African continent, with a resolution of 0.07° (approximately 7 km at the equator). The use of geo-localized surveys imposes technical constraints on the minimum grid resolution due to the anonymization process applied to DHS/LSMS/MICS surveys, which introduces random noise to the true GPS cluster locations. However, this random noise is mitigated when using larger grid cells. For instance, Michler et al. (2022) demonstrated minimal to no impact on estimates when using LSMS surveys in conjunction with weather raster data at a resolution of 4x4 km. Similar studies have employed grid cells of 6.72x6.72 km, showing strong predictive performance in wealth estimation using DHS data and satellite imagery (Ratledge et al., 2022; Yeh et al., 2020). Based on this literature, we adopted a grid resolution of 7x7 km to sufficiently mitigate anonymization noise while ensuring comparability with existing work. Figure 2 illustrates the 7x7 km grid over the Dakar area, with survey clusters depicted as points, and nighttime lights from 2022 used as a base layer.

We rasterized both predictor and target variables to a common resolution using our grid to create a unified dataset. For each grid cell and year with available survey data, we computed the target variable—the electricity access rate—by averaging all survey data points located within the grid cell.[37] Predictor variables included all spatial variables listed in Table 2, as well as the corresponding survey years, the cell's country ISO code, and the latitude and longitude of the cell centroid. As seen in Figure 2, some survey cluster-year units from a same year can be located within a common grid cell, and are therefore aggregated at the grid cell-year level, leading to a slightly lower number of cell-year units. This process yields a dataset structured as an unbalanced panel of cell-year units with 39,302 observations, which is used as our ML training dataset.

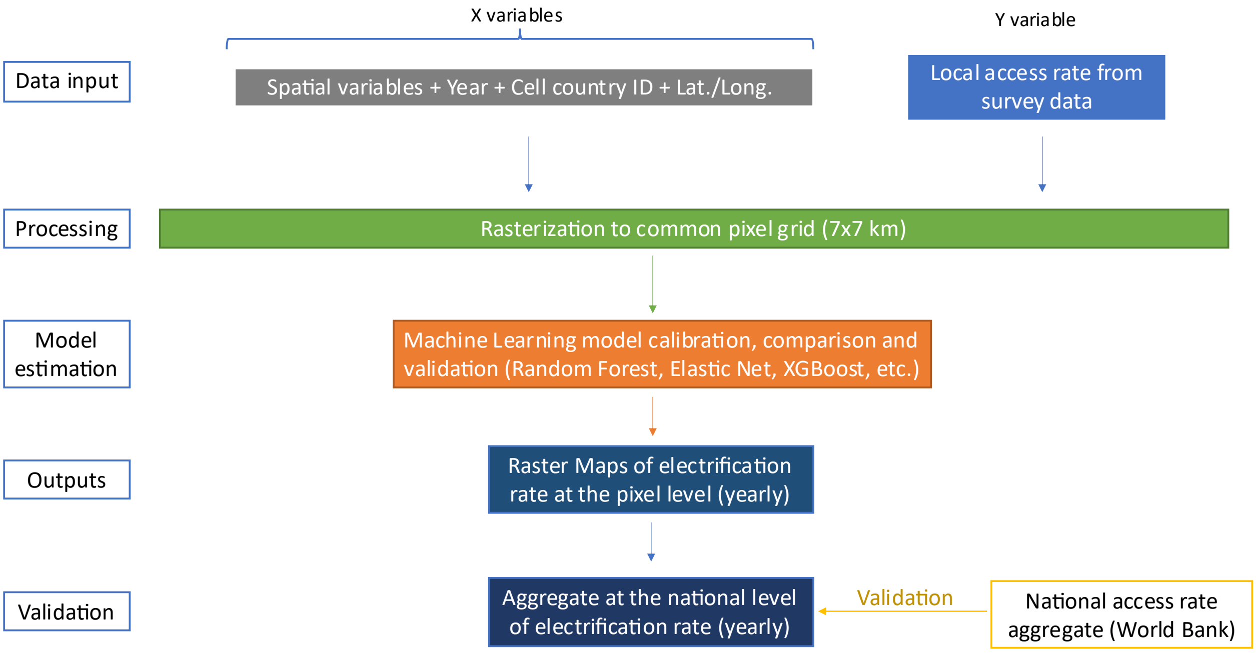
[37] If for one given cell-year unit we have multiple cluster points within a cell, we aggregate the access rates by taking the mean of all available cluster points within a cell.

Figure 2: Grid map around Dakar, Senegal



Notes: grid cells of 7x7km are in red, geolocated cluster are points colored relative to the survey year, NTL is in gray scale.

Figure 3: Methodology schematic



We trained multiple ML models, including Elastic-Net regularization, Random Forest, Gradient Boosted Trees (XGBoost), k-Nearest Neighbor (k-NN), Neural Networks (NN)[38], Deep Neural Networks (DNN)[39], and Support Vector Machines (SVM). We applied repeated cross-validation with 5 folds, repeated 10 times, to ensure robust performance evaluation. Hyperparameters for each model were optimized using a grid search approach. The model with the highest prediction accuracy for local electricity access rates was selected for subsequent analysis. Using this retained model, we predicted electricity access rates for each year, producing annual maps of electricity access across the African continent.

To validate the predicted electricity access values, we aggregated the subnational predictions to the national level and compared them to World Bank national aggregates. This external validation is used as a quality check at the macro-scale that the model outputs are aligned with independently reported data.

The variables used in this study are more flexible and computationally efficient compared to raw satellite images, which are usually used in the literature mentioned previously. Using raw high-resolution imagery introduces significant hardware constraints, such as high memory and processing power requirements, as well as the cost of acquiring multi-year data over large geographic areas. Furthermore, raw imagery limits the range of applicable ML models, as only convolutional neural networks (CNNs) are known to perform well in computer vision tasks (Rolf et al., 2021). By relying on pre-processed spatial data, we avoided these constraints, enabling the use of a broader range of resolution scales and ML algorithms and reduces the computing power required to train models and infer data.

3.2 Machine learning results and performance

Based on the literature on local wealth index predictions, which represents a similar regression task, we use R-squared as our primary performance metric, specifically the mean R-squared values after k-fold (k=5) cross-validation[40]. R-squared is a widely used metric in regression tasks as it measures the proportion of variance in the target variable explained by the model, providing an intuitive indication of predictive accuracy. It is particularly useful for comparing models, as higher values signify better alignment between predictions and actual values. For comprehensiveness, we also report and comment on Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), two additional standard performance metrics commonly used to evaluate model accuracy.

[38] Single hidden layer

[39] Two hidden layers

[40] We use a 5 folds design for training and testing (4 folds for training, 1 for testing performance). This 5-fold design is repeated 10 times with different assignation of cell-year units to each fold to insure robustness of the performance metrics.

Figure 4 displays the R-squared results for the trained models using the specified variable set after hyperparameter tuning. The results indicate that non-linear models clearly outperform linear models in capturing the relationships between electricity access and the selected explanatory variables. Non-linear models (k-NN, NN, DNN, SVM, XGBoost, and Random Forest) demonstrate significant improvements in prediction power, with R-squared values increasing from approximately 0.4 for linear models to at least 0.6 for most non-linear models.

Elastic-Net regularization does not substantially improve the predictive power compared to a standard linear regression model. This suggests that most of the selected variables contribute relevant information for predicting electricity access and that the standard linear model does not overfit the training set. These findings also indicate that relying solely on nighttime lights to predict electricity access would be insufficient, as Elastic-Net regularization would have improved performance by eliminating unnecessary variables if this were the case.

For non-linear models, Figure 4 highlights the significant performance gap they achieve compared to linear models. Among the tested models, the Random Forest model performs best in predicting electricity access across the sample, with a stable R-squared value of 0.67 over repeated cross-validation, as displayed in Table 4. XGBoost is the second-best performing model, followed by Neural Networks. The Random Forest model, using the specific set of variables, achieves similar or even superior performance on pooled samples compared to recent literature that employs raw satellite imagery and deep learning for predicting local wealth (Ratledge et al., 2022; Yeh et al., 2020).

Figure 4: R-squared on pooled predictions by model types

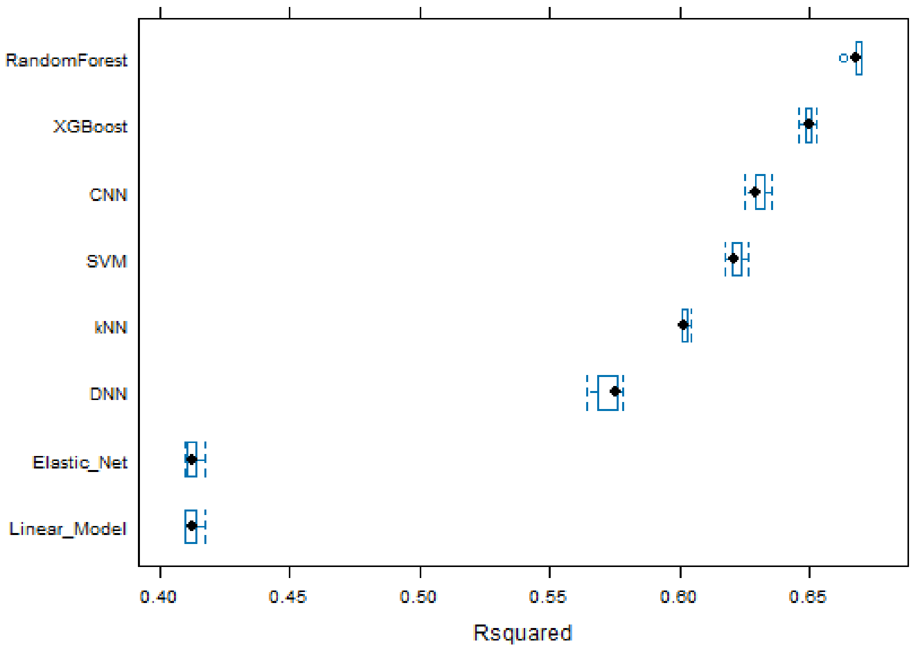


Table 4: R squared from repeated cross validation

Model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Linear_Model	0.41	0.41	0.41	0.41	0.41	0.42
Elastic_Net	0.41	0.41	0.41	0.41	0.41	0.42
DeepNeuralNet	0.56	0.57	0.58	0.57	0.58	0.58
kNN	0.60	0.60	0.60	0.60	0.60	0.60
SVM	0.62	0.62	0.62	0.62	0.62	0.63
NeuralNet	0.62	0.63	0.63	0.63	0.63	0.64
XGBoost	0.65	0.65	0.65	0.65	0.65	0.65
RandomForest	0.66	0.67	0.67	0.67	0.67	0.67

Tables 5 and 6 present the results for Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE), two additional widely used performance metrics for ML models. These metrics yield results consistent with the R-squared findings, confirming that the Random Forest model outperforms the others across all three metrics. Specifically, the Random Forest model achieves the lowest RMSE and the second-lowest MAE, with a mean absolute error of approximately 15 percentage points on average.

Given the R-squared, RMSE and MAE performance metrics, we retain the trained Random Forest model as our best model and use it for inference of the electricity access rates for cell-year units in the following sections.

Table 5: Root mean squared error from repeated cross validation

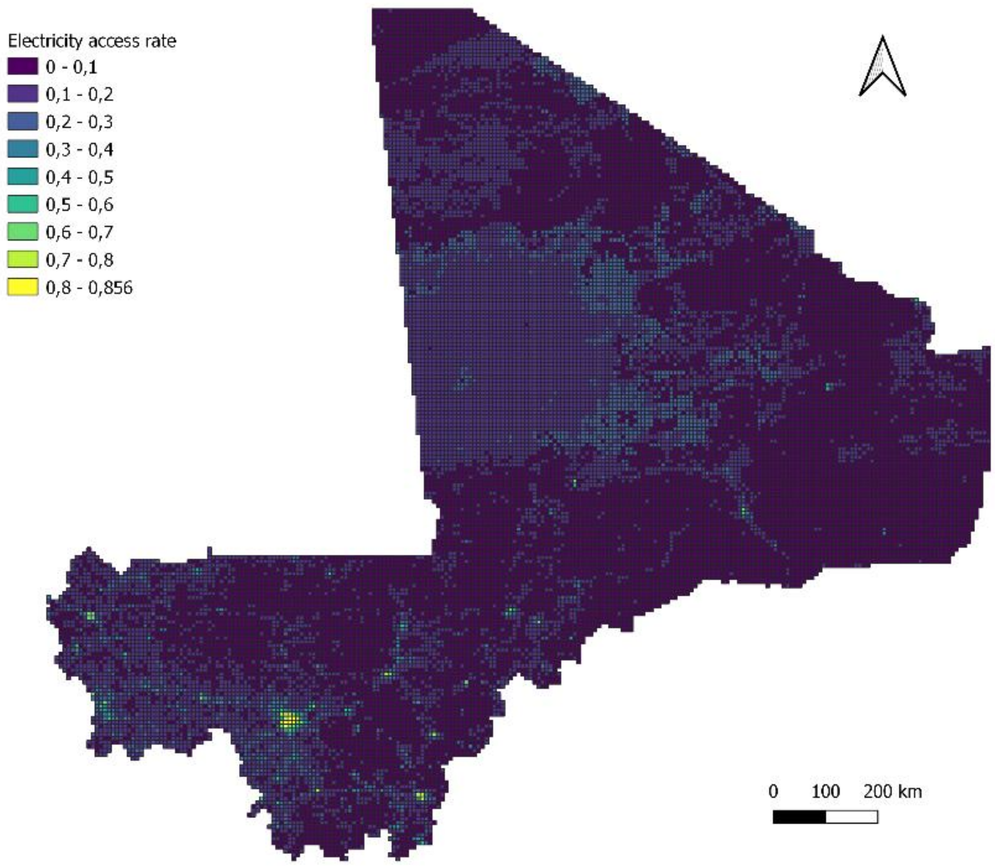
Model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Linear_Model	0.29	0.29	0.29	0.29	0.29	0.29
Elastic_Net	0.29	0.29	0.29	0.29	0.29	0.29
DeepNeuralNet	0.24	0.25	0.25	0.25	0.25	0.25
kNN	0.24	0.24	0.24	0.24	0.24	0.24
SVM	0.23	0.23	0.24	0.23	0.24	0.24
NeuralNet	0.23	0.23	0.23	0.23	0.23	0.23
XGBoost	0.22	0.22	0.22	0.22	0.22	0.22
RandomForest	0.22	0.22	0.22	0.22	0.22	0.22

Table 6: Mean absolute error from repeated cross validation

Model	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Elastic_Net	0.23	0.23	0.23	0.23	0.23	0.23
Linear_Model	0.23	0.23	0.23	0.23	0.23	0.23
DeepNeuralNet	0.18	0.18	0.19	0.19	0.19	0.19
kNN	0.16	0.17	0.17	0.17	0.17	0.17
NeuralNet	0.16	0.16	0.16	0.16	0.16	0.16
XGBoost	0.16	0.16	0.16	0.16	0.16	0.16
SVM	0.15	0.15	0.15	0.15	0.15	0.15
RandomForest	0.15	0.15	0.15	0.15	0.15	0.15

Figure 5 visualizes the output of the Random Forest model for Mali at a 0.07-degree pixel resolution. As expected, urban areas exhibit higher electricity access rates. Additionally, the visualization reveals that the southeastern part of the country shows higher electricity access levels compared to the northern region (Azawad).

Figure 5: Map of predicted electricity access values for Mali in 2021

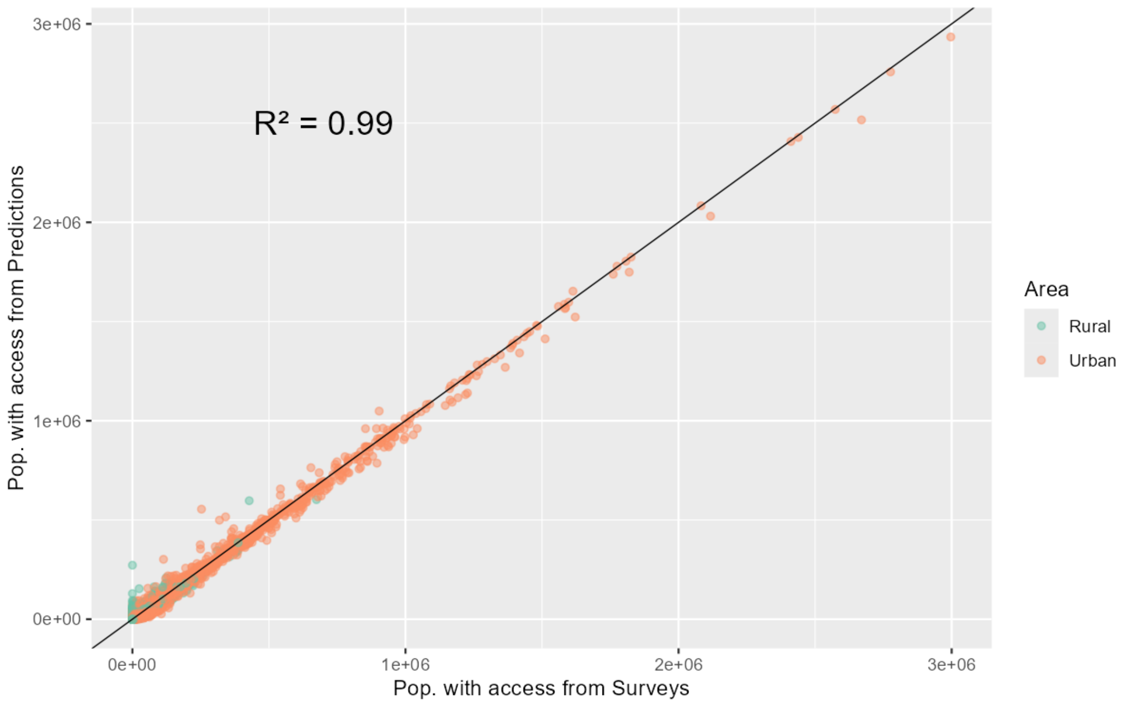


3.3 Machine learning external validation

We evaluate predictions at the grid cell level by comparing survey sources with predictions from our trained Random Forest (RF) model. For each grid cell, we calculate the total number of people with access to electricity, derived either from the survey source or the model prediction, by multiplying the grid cell population by the respective access rates. This approach stretches the distribution over a broader range of values, allowing us to detect potential distortions in predicted access rates, particularly at extreme values.

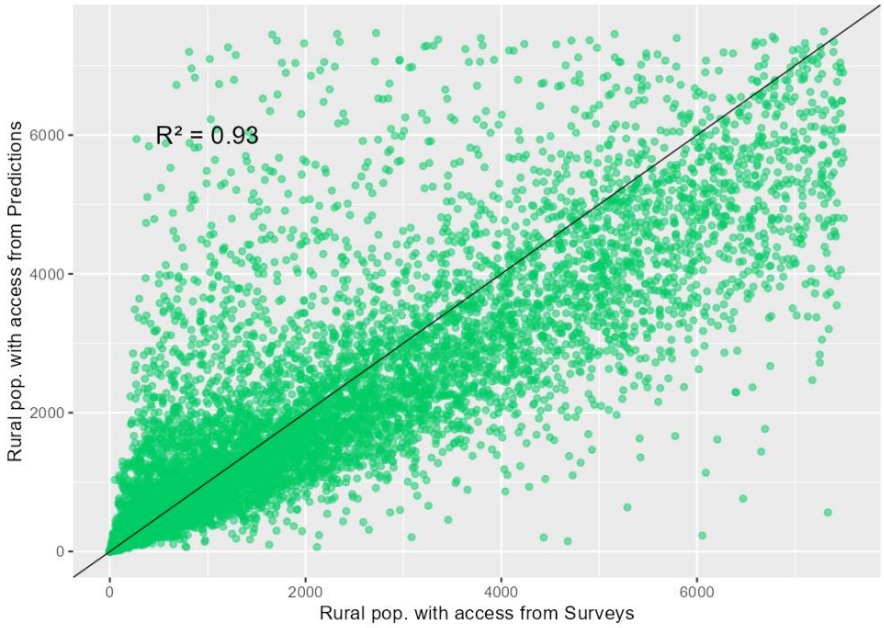
Overall, we observe no particular distortions in the distribution between the number of people with electricity access as reported by surveys and as predicted by the ML model (see Figure 6). The relationship between survey and predicted access rates is linear, with no apparent upward or downward bias, even at the distribution's extreme values. When focusing specifically on rural areas (Figure 7), we maintain a linear relation without any strong distortions.

Figure 6: Estimated population at the grid cell level with access to electricity from surveys vs RF predictions



Notes: 45° line in black.

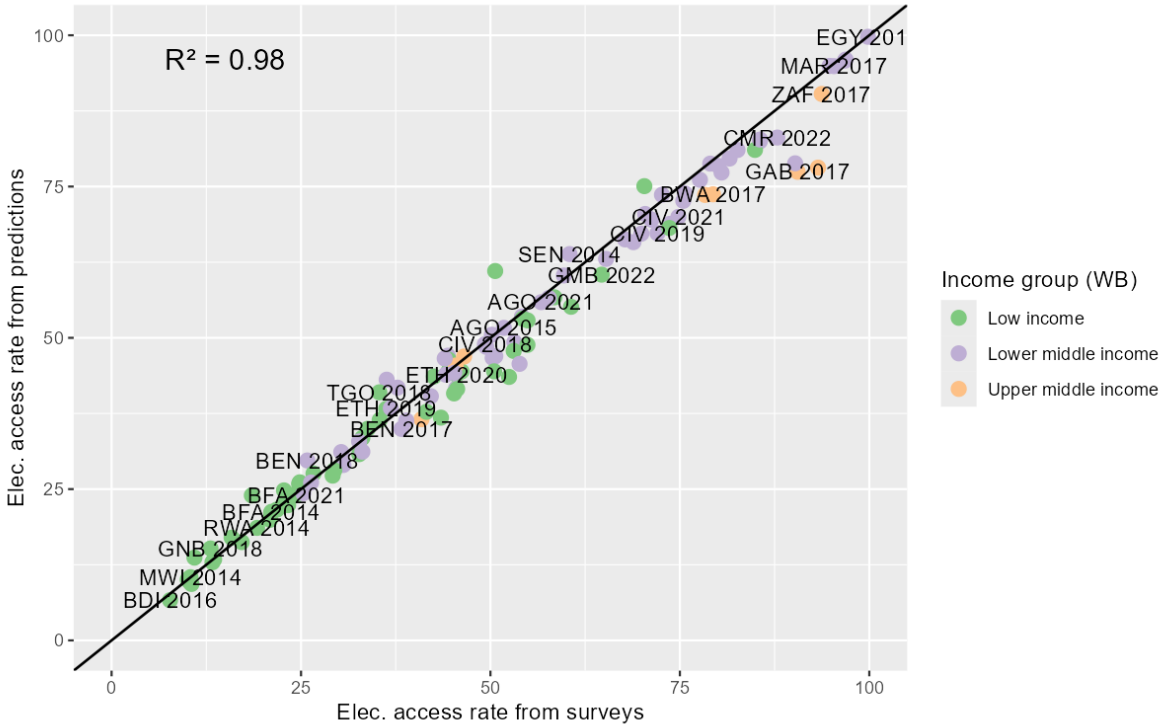
Figure 7: Estimated population with access to electricity from surveys vs RF predictions in rural areas



Notes: 45° line in black.

Next, we aggregate the results at the country level by computing national electricity access rates using both local survey data and predicted values. Figure 8 displays these aggregated results, combining all surveys available for each country and year. The predicted national rates exhibit a strong fit with survey-based rates, with no discernible upward or downward bias, even for countries near the minimum or maximum values of the distribution. As expected, countries belonging to lower-income groups generally have lower electricity access rates. The strong correlation observed at the local level persists at the national scale, further validating the model's performance.

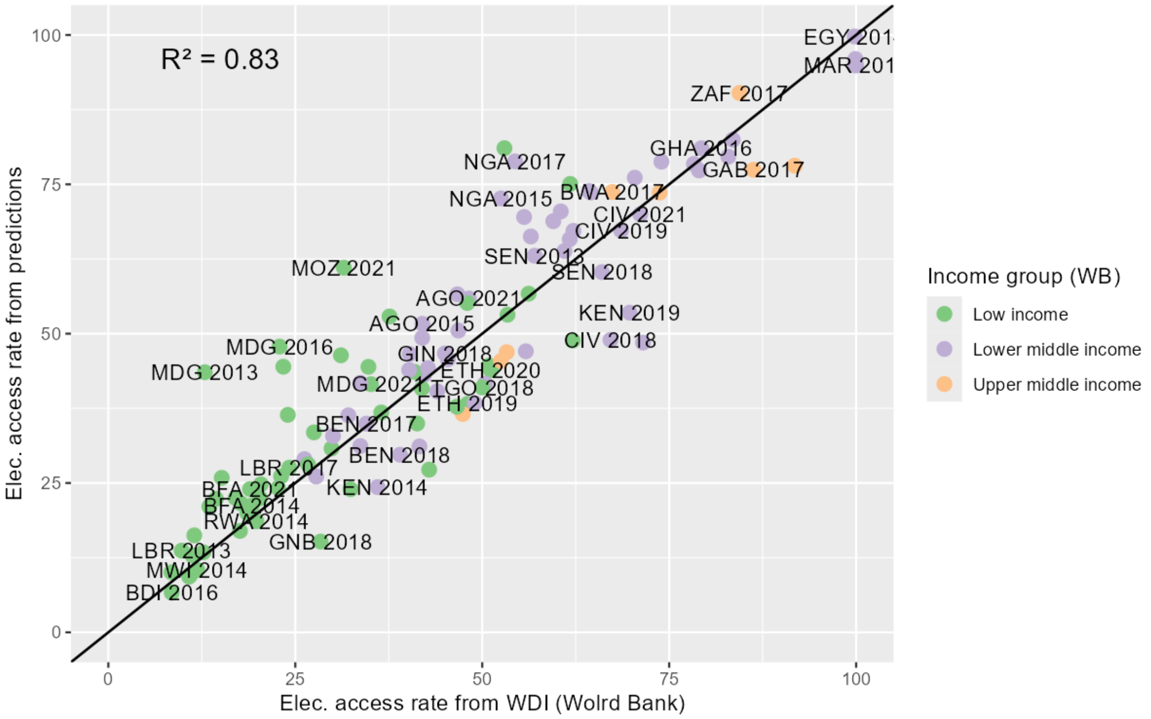
Figure 8: Aggregated country access rates (in %)



Notes: 45° line in black.

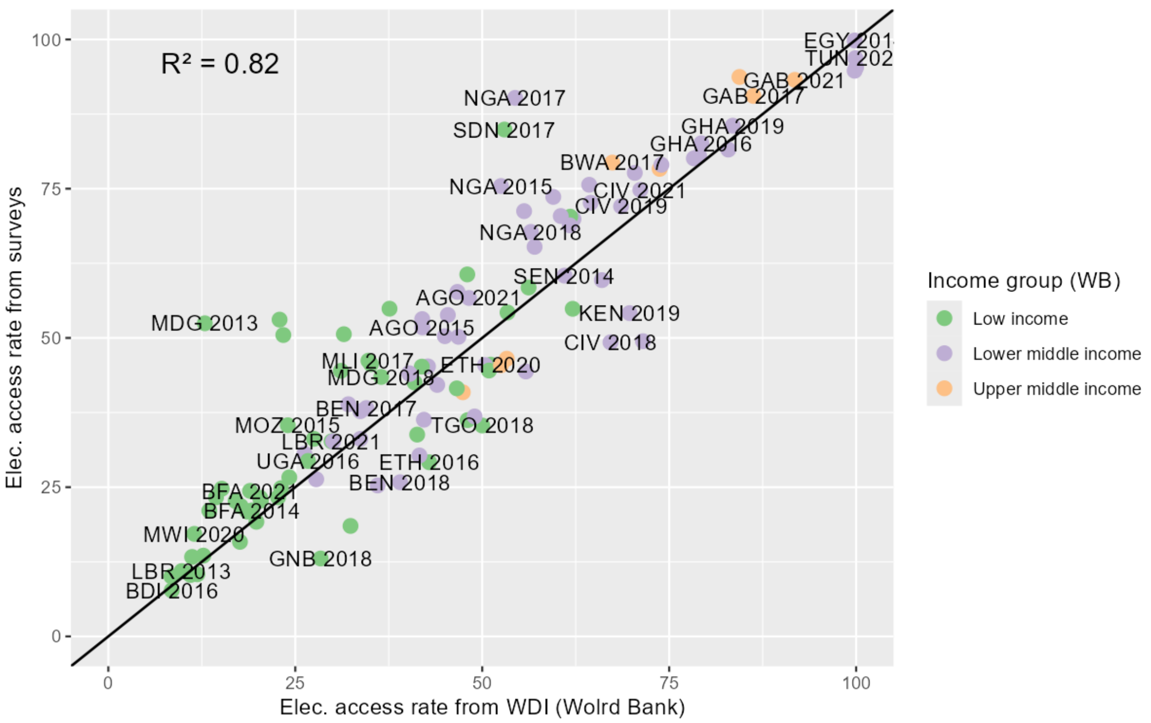
We then compare our national estimates to an external source, specifically the World Development Indicators (WDI) from the World Bank. As shown in Figure 9, the predicted access rates are strongly correlated with World Bank estimates, achieving an R^2 of 0.83, and exhibit no significant bias. This represents an improvement over Falchetta et al. (2019), whose estimates showed a downward bias at the national level, and Dhorne et al. (2021), who reported upward biases for low-income countries and downward biases for higher-income countries. Although the R^2 between predicted values and World Bank data is lower than that obtained when comparing survey sources, this discrepancy likely reflects differences in survey sampling methodologies and aggregation techniques employed by the World Bank. To explore this further, we compare national survey-based estimates with the World Bank data in Figure 10 and observe similar discrepancies. This finding supports the conclusion that most differences arise from sampling and aggregation methodologies rather than from poor predictive performance of the model. Thus, the observed deviations are attributable to data source differences rather than model limitations, further reinforcing the reliability of the Random Forest model predictions.

Figure 9: Aggregated country access rates (in %) from predictions vs WDI



Notes: 45° line in black.

Figure 10: Aggregated country access rates (in %) from Surveys vs WDI



Notes: 45° line in black.

4. Case study: Impact evaluation of energy projects

4.1 Methodology

We apply the predicted data from the trained ML model discussed previously to a case study aiming at measuring the impact of energy infrastructures projects. To conduct this impact study, we rely first on matching tools to select relevant control grid cells and then a Difference-in-Differences (DiD) estimator to estimate the causal effect of an infrastructure project. Following previous work on the impact of electricity access (Ratledge et al., 2022), we consider as a treated cell any cell that is within a 2 km buffer area of a geolocated energy project, and as controls all cells that are never treated by any infrastructure projects from Chinese or World Bank projects. We select control cells with matching using a nearest-neighbor algorithm based on Mahalanobis distance. To ensure similarity in between treated and control cells in our matching we retained three main variables: urbanization level, average population and average NTL level[41]. Once the best control cells selected, we conduct a causal inference estimation on electricity access with a DiD estimator. The key assumption underlying the DiD approach is the parallel trends assumption, which posits that in the absence of treatment, the difference between the treatment and control groups would have remained constant over time. However, traditional DiD models can be challenged when treatments are staggered or heterogeneous across units and time. To address these limitations, the estimator proposed by de Chaisemartin and d'Haultfœuille offers a more robust approach (de Chaisemartin and D'Haultfœuille, 2020)[42]. The de Chaisemartin and d'Haultfœuille estimator improves on traditional DiD methods by better accounting for treatment dynamics, reducing biases, and ensuring more reliable causal inferences in staggered empirical settings with panel data. Since it is difficult to ensure that a cell is not treated by another energy program outside the scope of our dataset, estimates could however be bias downwards or upwards as respectively control cells or treated cells are actually also treated by another unknown program.

We estimate the model defined as follows:

$$AccessRate_{i,t} = \alpha + \sum_{e=-4}^{-2} \beta_e^{lead} D_{i,t}^{e} + \sum_{e=0}^{8} \beta_e^{lag} D_{i,t}^{e} + C_i + Y_t + \varepsilon_{i,t} \qquad (1)$$

Where i is a grid cell, t a time period and e the event study time (i.e. elapsed time since treatment). $AccessRate_{i,t}$ is the access rate to electricity of grid cell i at time t, $D^{e}_{i,t}$ a dummy variable denoting the treatment status of grid cell i being e periods after or before treatment

[41] This set of variables are selected as they are quite indicative of treatment and provide good matching characteristics. For both population and NTL, we collapse the temporal dimension by taking the average before treatment for treated cells, and the whole period for control cells.

[42] https://github.com/chaisemartinPackages/did_multiplegt_dyn

at time t. $D^{e}_{i,t}$ takes the value 1 when $e >= 0$ and the cell is treated. C_i is a cell fixed effect, Y_t a year fixed effect and and $\varepsilon_{i,t}$ the error term. β_e^{lag} represents the Average Treatment on the Treated (ATT) at period e after treatment, which is our coefficient of interest reported in the following regression tables. All standard errors are clustered by default at the cell level, as recommended by de Chaisemartin and d'Haultfœuille.

We provide estimates for projects financed by China and the World Bank, as well as a heterogeneity analysis based on projects characteristics like the type of project (i.e. production or distribution) or location (i.e. urban or rural). We also investigate the robustness of our results by modifying the cell's treatment status based on their distance to a project. We provide in annex results where we removed the buffer zone of 2km around projects location, considering as treated only cells intersecting with the geolocation of a project. We also investigate the robustness of our estimates by using an alternative DiD estimator provided by Callaway and Sant'Anna (2021). Callaway and Sant'Anna provide an alternative estimator to de Chaisemartin and d'Haultfœuille for staggered difference-in-differences. It is a comparable estimator as it allows to compute dynamic effects and is robust to effects heterogeneity. The key difference between their approaches relies on how they handle heterogeneity, where Callaway and Sant'Anna explicitly estimate effects by subsampling different cohorts using a never-treated control group, whereas de Chaisemartin and d'Haultfœuille is based on the notion of treatment effect decompositions and switching status ensuring that comparisons are only made between groups that experience treatment at different times.

4.2 Results

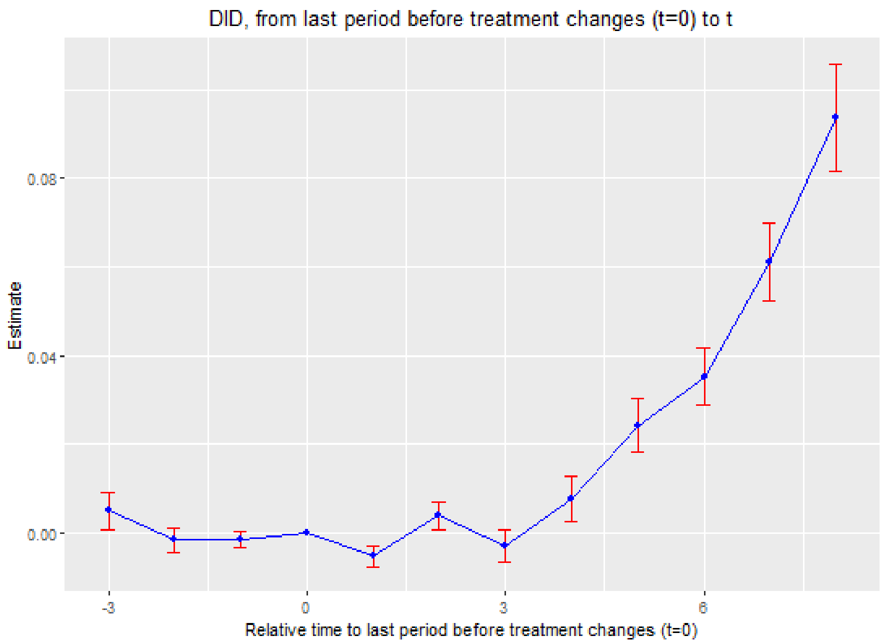
Using the predicted electricity access data, we first assess the impact of energy infrastructure projects financed by China. We estimate the effects for cells within a 2 km radius of project sites compared to control cells as defined previously. Figure 11 and Table 7 present the estimates from the dynamic difference-in-differences (DiD) for these these projects. The analysis reveals a positive and significant impact on electricity access on treated cells, defined as cells within the 2 km buffer area surrounding the project locations. The overall average of the effect over the entire period (i.e., the first 8 years post-implementation) is approximately an increase of 2 percentage points. Furthermore, the results show a continuous growth in the effect over time, culminating at a 9 percentage-point increase eight years after implementation. This sustained growth suggests that within treated cells, more households gradually gain access to electricity due to the infrastructure deployed, with no evidence of a plateau within the first eight years**.** This trend indicates that, while infrastructure is made available, economic, social, or behavioral factors may influence the pace at which households adopt electricity. The observed slow and steady increase in electricity access also implies that measuring the socio-economic impacts of such projects may require a longer timeframe to capture significant and measurable outcomes. We also notice that there is no significant effect during the first three years after commitment, which is in line with the average delay of three years computed between commitment and completion year, corresponding to the building or implementation phase of the project. When using completion year as treatment date, we then notice no delay with a significant effect the first year after implementation (see Table A10 in annex).

These findings highlight the capacity of Chinese energy projects to contribute meaningfully to local electrification, particularly in areas directly surrounding the projects. However, the average effect over eight years suggests a moderate impact, emphasizing the need for a deeper understanding of the project types and technologies driving these outcomes.

Table 7: DiD estimates for Chinese projects

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	-0.00522	0.00121	-0.00759	-0.00285	11232	1071
Effect_2	0.00397	0.00157	0.00089	0.00705	9521	1067
Effect_3	-0.00299	0.00186	-0.00663	0.00064	8958	1067
Effect_4	0.00775	0.00258	0.00268	0.01281	7557	1063
Effect_5	0.02428	0.00311	0.01818	0.03038	6442	1052
Effect_6	0.0352	0.00325	0.02883	0.04157	5338	1038
Effect_7	0.0612	0.0045	0.05238	0.07001	3966	745
Effect_8	0.09362	0.00619	0.08147	0.10576	2654	508
Av_tot_eff	0.02088	0.00219	0.01658	0.02518	18847	7611

Figure 11: China projects DiD

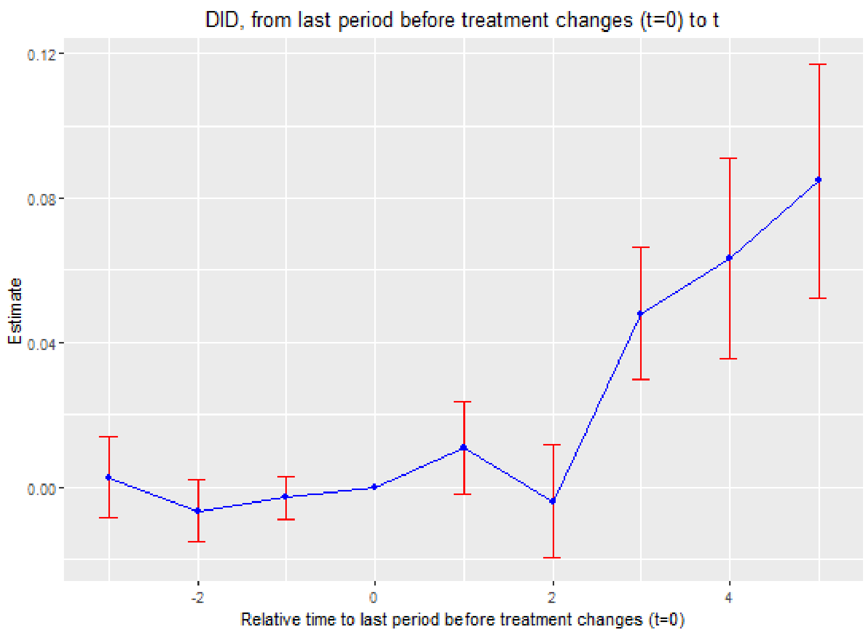


Results of the DiD analysis for World Bank-financed energy projects are presented in Figure 12 and Table 8. Due to the more restricted sample of World Bank projects, we are only able to compute robust dynamic effects up to five years after commitment. We find a positive and significant impact of these projects on electricity access, with the average effect over the first five years estimated at 4 percentage points increase, which is twice magnitude observed for Chinese-financed projects. The dynamics of the effect are broadly similar to those of Chinese projects with a stagnation phase and then a sharp dynamic raise in access rate. We notice however that World Bank projects begin to have a significant effect on the third year, which is one period earlier than what is found on Chinese projects. In terms of magnitude, we also notice that at 5 years after implementation World Bank projects yield an increase of 8 percentage points, against only +2 points for their Chinese counterparts.

Table 8: DiD estimates for World Bank GEMS projects

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	0.01067	0.00653	-0.00212	0.02346	551	110
Effect_2	-0.00393	0.00793	-0.01947	0.0116	551	110
Effect_3	0.04794	0.00934	0.02964	0.06624	332	108
Effect_4	0.06334	0.01419	0.03553	0.09115	332	108
Effect_5	0.08472	0.0165	0.05239	0.11705	330	108
Av_tot_eff	0.04028	0.00929	0.02208	0.05847	1650	544

Figure 12: DiD estimates for World Bank GEMS projects



These findings suggest that World Bank projects may achieve a greater local impact in a relative shorter time span. However, these results must be interpreted with caution, as they are based on a relatively small sample of World Bank projects. Consequently, we cannot ensure that this sample is representative of the entire World Bank portfolio of energy projects.

To verify the robustness of our results, we conducted additional checks using the alternative estimator proposed by Callaway and Sant'Anna (2021). Results remain significant with similar magnitudes of the effect and we find the similar dynamic (see annex tables A11 and A12). We also check robustness of the results by removing the buffer zone used to consider the treatment these variations yield consistent results with previous findings, confirming the reliability of our estimates (see annex tables A8 and A9).

Results heterogeneity

The observed differences between Chinese and World Bank projects may be driven by variations in context or project types and technological focus. For example, Chinese projects include a substantial proportion of renewable power production initiatives, while in contrast World Bank projects are mostly transmission and distribution initiatives, which may yield quite different effects due to technological differences.

The heterogeneity analysis (tables and figures available in annex) further supports this interpretation, showing that renewable energy projects financed by China drive the majority of the observed positive impacts. For the World Bank, transmission projects account for most of the positive effects. Both Chinese renewable energy projects and World Bank transmission projects yield an increase of approximately 20 percentage points at their last estimable period, which are respectively 8 years after commitment for Chinese projects and 5 years for World Bank projects. However, the robustness of these findings is limited by the reduced sample sizes in the heterogeneity analysis. Moreover, only transmission projects funded by the World Bank show no significant pre-trend bias, allowing for a reliable causal interpretation. Consequently, conclusions about other project types should be treated with caution.

Focusing on Chinese projects, we also examine the heterogeneity of results based on the type of location treated, such as urban versus rural areas. The majority of projects are located in rural areas (1,003 rural treated cells compared to 68 urban treated cells), indicating a clear priority in financing infrastructure projects in rural regions. The measured effects in rural areas are consistent with the overall sample estimates, which is expected given their large representation in the sample. However, for projects located in urban areas, the observed effects are not significant. Unfortunately, this finding is based on a very small sample, as shown in Table 9 and Figure 13, resulting in large standard errors and no statistically significant effects for urban areas.

Table 9: DiD estimates for Chinese projects located in urban areas

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	0.00946	0.00587	-0.00204	0.02096	636	68
Effect_2	0.01023	0.00792	-0.00529	0.02576	571	68
Effect_3	0.01424	0.01137	-0.00805	0.03653	531	68
Effect_4	0.01804	0.01305	-0.00753	0.04361	492	68
Effect_5	0.02667	0.0158	-0.00429	0.05762	403	63
Effect_6	0.02746	0.01948	-0.01071	0.06564	319	51
Effect_7	0.01213	0.02586	-0.03855	0.06281	230	29
Effect_8	0.00024	0.02621	-0.05113	0.05162	162	28
Av_tot_eff	0.01574	0.0125	-0.00877	0.04025	1212	443

Figure 13: DiD estimates for Chinese projects located in urban areas

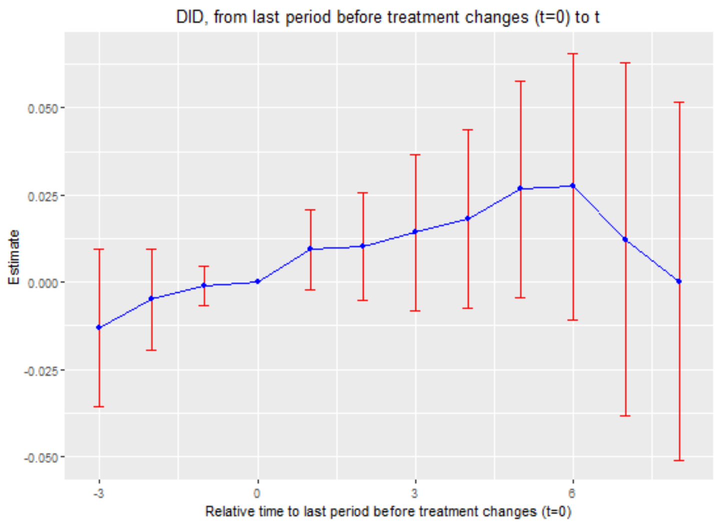
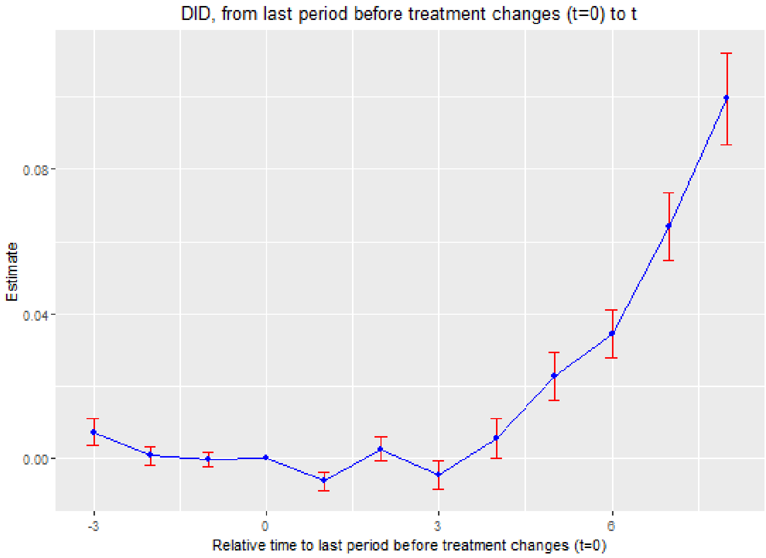


Table 10: DiD estimates for Chinese projects located in rural areas

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	-0.00624	0.00128	-0.00875	-0.00373	9250	1003
Effect_2	0.0026	0.00168	-0.00069	0.00589	7811	999
Effect_3	-0.00464	0.00196	-0.00848	-8,00E-04	7279	999
Effect_4	0.00566	0.00278	0.00021	0.0111	6122	997
Effect_5	0.02289	0.00333	0.01636	0.02942	5237	986
Effect_6	0.03453	0.00347	0.02773	0.04133	4363	973
Effect_7	0.064	0.00478	0.05464	0.07337	3232	692
Effect_8	0.09956	0.00646	0.08691	0.11221	2163	471
Av_tot_eff	0.02032	0.00235	0.01571	0.02494	16215	7120

Figure 14: DiD estimates for Chinese projects located in rural areas



Cost-effectiveness analysis

To estimate the cost-effectiveness of energy projects financed by either China or the World Bank, we compute the number of people gaining access to electricity living in respective treated cells and then compare it to their respective investment amounts. Although those estimates rely on a number of assumptions, it allows us to compare efficiency between the two financiers. We consider the year 2022, which is the last year of our sample, and multiply treated cells' predicted access rates by cells' population to retrieve the number of people with access to electricity where projects were deployed[43]. Once population gaining access to electricity thanks to projects are computed, we then divide the total of persons with access by the total of respective investments to retrieve the cost per person of the infrastructure projects[44].

Maintaining the previously made assumption that the effect of each project is confined to cells within a 2 km radius of the project site, Chinese-funded projects enabled approximately 746,000 people to gain access to electricity during the period 2013–2022. Based on the total investment costs over this period, the cost per person is estimated at USD 26,429. Given the observed dynamics of increasing electricity access over time for individual projects, this cost estimate is likely to decrease in the long run as more households gain access until full electrification is achieved within the treated areas.

[43] We make here the assumption that people with access to electricity within a treated cell all benefited from the implemented project. This assumption seems reasonable in rural areas but might be more debatable for urban areas where electricity access might be less dependent of one given project.

[44] Concerning investment amounts we rely on the assumption that overall, co-financing does not play a different role between projects financed by China or the World Bank. If one of the two relies more heavily on co-financing we are then underestimating the projects' total cost.

For World Bank-funded projects, we estimate that 76,193 people gained access to electricity within treated cells during the same period, resulting in an investment cost of USD 14,043 per person. These estimates suggest that World Bank projects demonstrate significantly higher cost-effectiveness compared to Chinese-funded projects.

However, it is important to interpret these results with caution, as our analysis does not encompass the full portfolio of World Bank-financed energy projects. Additionally, variations in project types, scales, and implementation contexts could further influence these cost-effectiveness estimates. Also, as we saw in the previous section, the effect of a project grows over time, meaning that over the years more and more people gain access to electricity, which reduces the cost per person in the long run[45].

5. Discussion

While our model's performance is on par with existing studies using machine learning (ML) and remote sensing data for wealth prediction, there is room for improvement. Our model explains approximately 70% of the variability in electricity access, leaving 30% unexplained. Such limitations are common in studies leveraging ML approaches with remote sensing data, whether based on raw imagery or pre-processed inputs, as in this study. Prediction errors can lead to imprecise assessments of electricity access trends, potentially skewing policy conclusions about access improvements and disparities.

A critical area for enhancement lies in improving the precision of spatial data used for model training. Current practices by major survey organizations often involve anonymizing spatial data to protect respondent privacy, which introduces significant challenges for finely resolved geospatial analysis. For example, the random displacement of survey cluster locations, up to 2.5 kilometers in urban areas and 5 kilometers in rural areas (Burgert et al., 2013), reduces the reliability of location-based analyses. This issue is particularly pronounced in rural zones, where precise geolocation could enable smaller grid cells and better capture the relationship between household survey data and remote sensing inputs. Addressing this limitation could lead to improved model accuracy, especially in remote and sparsely populated areas.

In evaluating the impact of energy projects financed by the Chinese government and the World Bank, our results demonstrate that we see a significant increase in electricity access in areas surrounding project installations. This highlights the potential of leveraging ML and remote sensing for systematic and scalable project evaluation. If geospatial data on project locations were consistently recorded, this approach could enable standardized evaluations of

[45] On average Chinese projects were committed in 2013 and completed in 2017, while World Bank projects were committed in 2017.

development interventions at a relatively low cost and enable a more systematic tracking of cost-effectiveness analysis.

However, a notable limitation of our empirical analysis is the restricted representation of World Bank projects in the sample. This limits the generalizability of our findings regarding the comparative impact of Chinese versus World Bank-funded projects. Despite this, our results suggest that the impact on electricity access per dollar spent may be higher for World Bank-funded projects, if we make the assumption our sample is representative of the whole portfolio. Further research with larger and more diverse datasets is needed to validate this observation and better understand the factors driving differences in project outcomes.

The study predominantly focuses on Chinese-financed projects due to the relative availability of geolocation data for these interventions. Unfortunately, most Development Finance Institutions (DFIs) do not systematically provide geolocated data for their projects, significantly constraining the scope of research. This limitation may lead to conservative impact estimates, as some grid cells classified as "controls" may, in fact, have been influenced by other energy projects outside the scope of this study and other potential unobserved confounders and data limitations remain to be considered. While averaging results across many projects reduces the bias caused by misclassified control cells, such misclassification could distort conclusions in difference-in-differences analyses, particularly when focusing on a small number of projects.

The systematic use of geospatial data by DFIs could significantly enhance project monitoring and evaluation tasks. Remote sensing and geolocated survey data provide a cost-effective foundation for assessing development interventions, allowing organizations to leverage existing data to implement rigorous impact evaluations. A coordinated effort to integrate geospatial data into project reporting and evaluation frameworks could lead to a more data-driven, transparent, and standardized approach to understanding the effectiveness of energy and other development projects.

In addition to improving geospatial data availability and precision, future research should explore incorporating additional data sources, such as near real-time satellite imagery, socio-economic indicators, and crowdsourced data, to capture more nuanced drivers of electricity access. Developing more ML models that can account for temporal dynamics and causal relationships could also improve the understanding of electrification trends and policy impacts. Furthermore, expanding this approach to evaluate other development sectors, such as health, education, and water access, could enhance the broader application of ML and remote sensing in global development research. Finally, applying these methods in other regions beyond Africa would test the robustness and scalability of the approach, contributing to the global agenda for achieving equitable and sustainable development outcomes.

6. Conclusion

This study demonstrates that electricity access rates can be predicted at the local level with good confidence, achieving performance comparable to the standards in the literature on wealth, while employing a less resource-intensive approach. By utilizing Random Forest models with pre-processed data rather than raw satellite imagery, we achieve results equivalent to advanced computer vision models, but with significantly reduced computational and data processing demands. Our approach also addresses a critical limitation of methods relying on raw imagery, which is to relieve the constraints of working with cross-sectional data, due to the complexity, size, and cost associated with processing high-resolution images across time, by being able to produce multi-year predictions.

We advance the literature by generating electricity access predictions for multiple years and across the whole African continent without particular downward or upward biases as it was observed in other large-scale mapping efforts, such as those by Falchetta et al. (2019) or Dhorne et al (2021). Our model achieves precise predictions of electricity access levels at a granular scale (7 x 7 km) and fills critical data gaps for un-surveyed areas, particularly in rural regions representing the largest share of population living in Africa. These predictions enable the creation of comprehensive electricity access maps, which can be aggregated to larger spatial scales, such as administrative units, providing actionable insights for policymakers and planners.

Finally, we illustrate the utility of our predicted electricity access data in evaluating the impacts of energy infrastructure projects. Our analysis reveals that energy projects financed by China and the World Bank have positive and significant measurable impacts on electricity access at the local level with a dynamic growing effect over time. Specifically, Chinese-financed projects increase electricity access by around +2 percentage points in treated cells after five years of commitment, reaching up to +9 points after eight years. On the other hand, World Bank-financed projects demonstrate a higher average increase of +8 percentage points after only five years, although these findings should be interpreted cautiously due to the limited sample size of World Bank projects. These effects are averages across the energy sector and encompass varying outcomes depending on the technology, type or location of the projects.

Despite the promising results, this study has several limitations. The model relies heavily on the availability and quality of survey, which may vary significantly across countries and regions. The limited spatial and temporal resolution of the input data might also lead to oversights in areas where electrification processes occur rapidly or at finer scales. Furthermore, while we demonstrate the utility of our data for impact evaluation, the heterogeneity in project types and technologies warrants more granular analysis to better understand potential variations in outcomes.

In conclusion, our study highlights the potential of machine learning models to bridge critical data gaps in electricity access. This approach offers significant advantages for infrastructure planning and impact evaluation, particularly in underserved rural areas, and provides a foundation for future research and policy interventions in the energy sector. Research could also extend the evaluation of infrastructure impacts to include socio-economic outcomes, such as education, health, and income improvements, providing a more holistic assessment of electrification projects. Finally, applying similar methodologies to other regions beyond Africa could test the robustness and scalability of the approach, contributing to the global agenda for universal electricity access.

References

Beyer, R., Hu, Y., Yao, J., 2022. Measuring Quarterly Economic Growth from Outer Space. Policy Res. Work. Pap. https://doi.org/10.1596/1813-9450-9893

Burgert, C.R., Colston, J., Roy, T., Zachary, B., 2013. Geographic displacement procedure and georeferenced data release policy for the Demographic and Health Surveys. DHS Spat. Anal. Rep. No. 7.

Callaway, B., Sant'Anna, P.H.C., 2021. Difference-in-Differences with multiple time periods. J. Econom., Themed Issue: Treatment Effect 1 225, 200–230. https://doi.org/10.1016/j.jeconom.2020.12.001

Chi, G., Fang, H., Chatterjee, S., Blumenstock, J.E., 2022. Microestimates of wealth for all low- and middle-income countries. Proc. Natl. Acad. Sci. 119, e2113658119. https://doi.org/10.1073/pnas.2113658119

Custer, S., Dreher, A., Elston, T.B., Escobar, B., Fedorochko, R., Fuchs, A., Ghose, S., Lin, J., Malik, A., Parks, B.C., Solomon, K., Strange, A., Tierney, M.J., Vlasto, L., Walsh, K., Wang, F., Zaleski, L., Zhang, S., 2023. AidData | Tracking Chinese Development Finance: An Application of AidData's TUFF 3.0 Methodology. AidData William Mary.

de Chaisemartin, C., D'Haultfœuille, X., 2020. Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects. Am. Econ. Rev. 110, 2964–2996. https://doi.org/10.1257/aer.20181169

Dhorne, M., Nicolas, C., Arderne, C., Besnard, J., 2021. Tracking Advances in Access to Electricity Using Satellite-Based Data and Machine Learning to Complement Surveys (Brief). World Bank, Washington, DC.

Dreher, A., Fuchs, A., Parks, B., Strange, A., Tierney, M.J., 2021. Aid, China, and Growth: Evidence from a New Global Development Finance Dataset. Am. Econ. J. Econ. Policy 13, 135–174. https://doi.org/10.1257/pol.20180631

Elvidge, C., Baugh, K., Zhizhin, M., Hsu, F.-C., 2013. Why VIIRS data are superior to DMSP for mapping nighttime lights. Proc. Asia-Pac. Adv. Netw. 35, 62–69. https://doi.org/10.7125/APAN.35.7

Elvidge, C.D., Baugh, K.E., Kihn, E.A., Kroehl, H.W., Davis, E.R., Davis, C.W., 1997. Relation between satellite observed visible-near infrared emissions, population, economic activity and electric power consumption. Int. J. Remote Sens. 18, 1373–1379. https://doi.org/10.1080/014311697218485

Elvidge, C.D., Ghosh, T., Hsu, F.-C., Zhizhin, M., Bazilian, M., 2020. The Dimming of Lights in China during the COVID-19 Pandemic. Remote Sens. 12, 2851. https://doi.org/10.3390/rs12172851

Falchetta, G., Pachauri, S., Parkinson, S., Byers, E., 2019. A high-resolution gridded dataset to assess electrification in sub-Saharan Africa. Sci. Data 6, 110. https://doi.org/10.1038/s41597-019-0122-6

Goodman, S., Zhang, S., Malik, A.A., Parks, B.C., Hall, J., 2024. AidData's Geospatial Global Chinese Development Finance Dataset. Sci. Data 11, 529. https://doi.org/10.1038/s41597-024-03341-w

Jean, N., Burke, M., Xie, M., Davis, W.M., Lobell, D.B., Ermon, S., 2016. Combining satellite imagery and machine learning to predict poverty. Science 353, 790–794. https://doi.org/10.1126/science.aaf7894

Li, Z., Liu, F., Yang, W., Peng, S., Zhou, J., 2022. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Trans. Neural Netw. Learn. Syst. 33, 6999–7019. https://doi.org/10.1109/TNNLS.2021.3084827

Ratledge, N., Cadamuro, G., de la Cuesta, B., Stigler, M., Burke, M., 2022. Using machine learning to assess the livelihood impact of electricity access. Nature 611, 491–495. https://doi.org/10.1038/s41586-022-05322-8

Rolf, E., Proctor, J., Carleton, T., Bolliger, I., Shankar, V., Ishihara, M., Recht, B., Hsiang, S., 2021. A generalizable and accessible approach to machine learning with global satellite imagery. Nat. Commun. 12, 4392. https://doi.org/10.1038/s41467-021-24638-z

Sherman, L., Proctor, J., Druckenmiller, H., Tapia, H., Hsiang, S.M., 2023. Global High-Resolution Estimates of the United Nations Human Development Index Using Satellite Imagery and Machine-learning. Working Paper Series. https://doi.org/10.3386/w31044

Shi, K., Yu, B., Huang, Y., Hu, Y., Yin, B., Chen, Z., Chen, L., Wu, J., 2014. Evaluating the Ability of NPP-VIIRS Nighttime Light Data to Estimate the Gross Domestic Product and the Electric Power Consumption of China at Multiple Scales: A Comparison with DMSP-OLS Data. Remote Sens. 6, 1705–1724. https://doi.org/10.3390/rs6021705

Yeh, C., Perez, A., Driscoll, A., Azzari, G., Tang, Z., Lobell, D., Ermon, S., Burke, M., 2020. Using publicly available satellite imagery and deep learning to understand economic well-being in Africa. Nat. Commun. 11, 2583. https://doi.org/10.1038/s41467-020-16185-w

Annexes

Figure A1: DHS clusters location

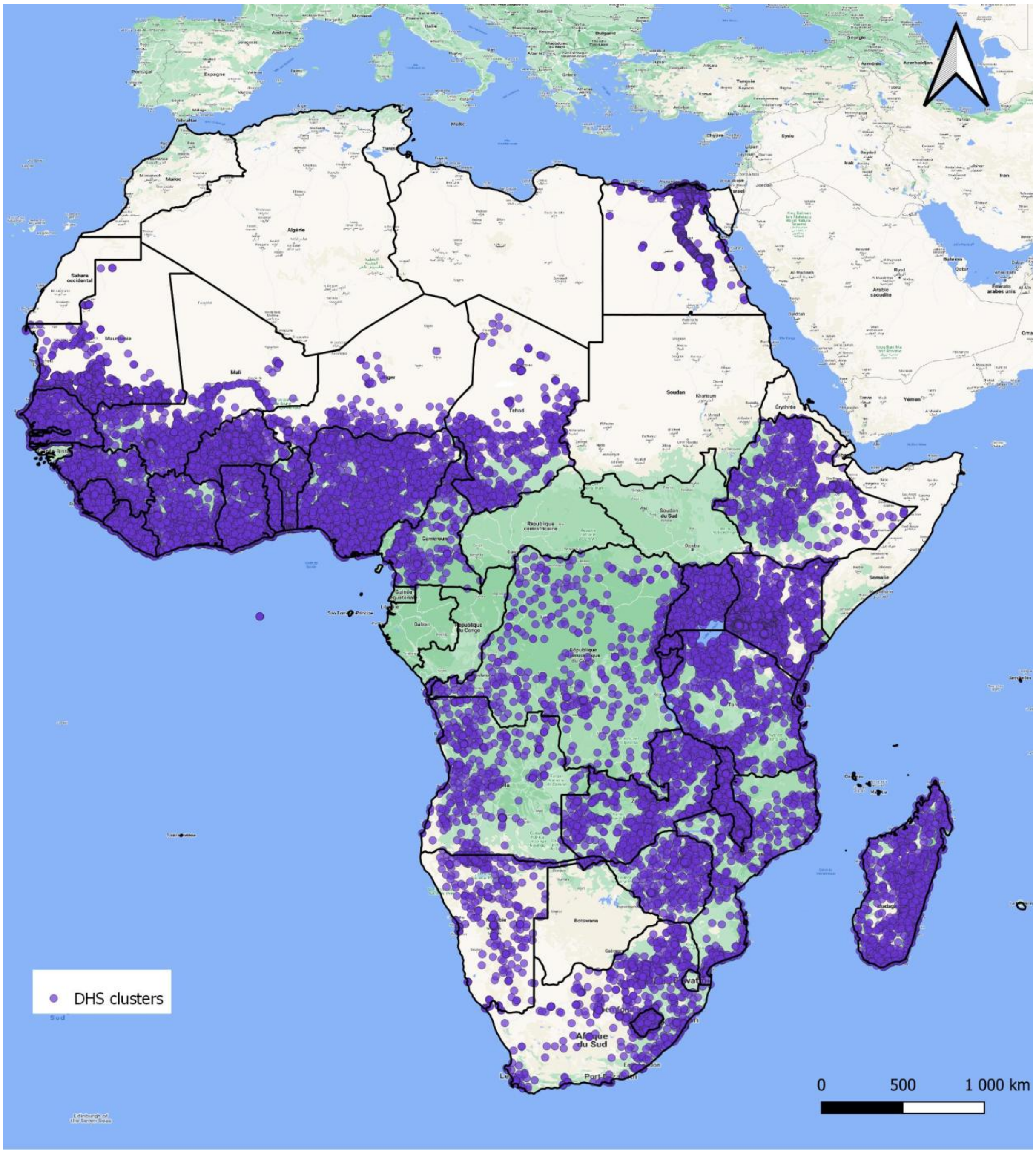


Figure A2: LSMS clusters location

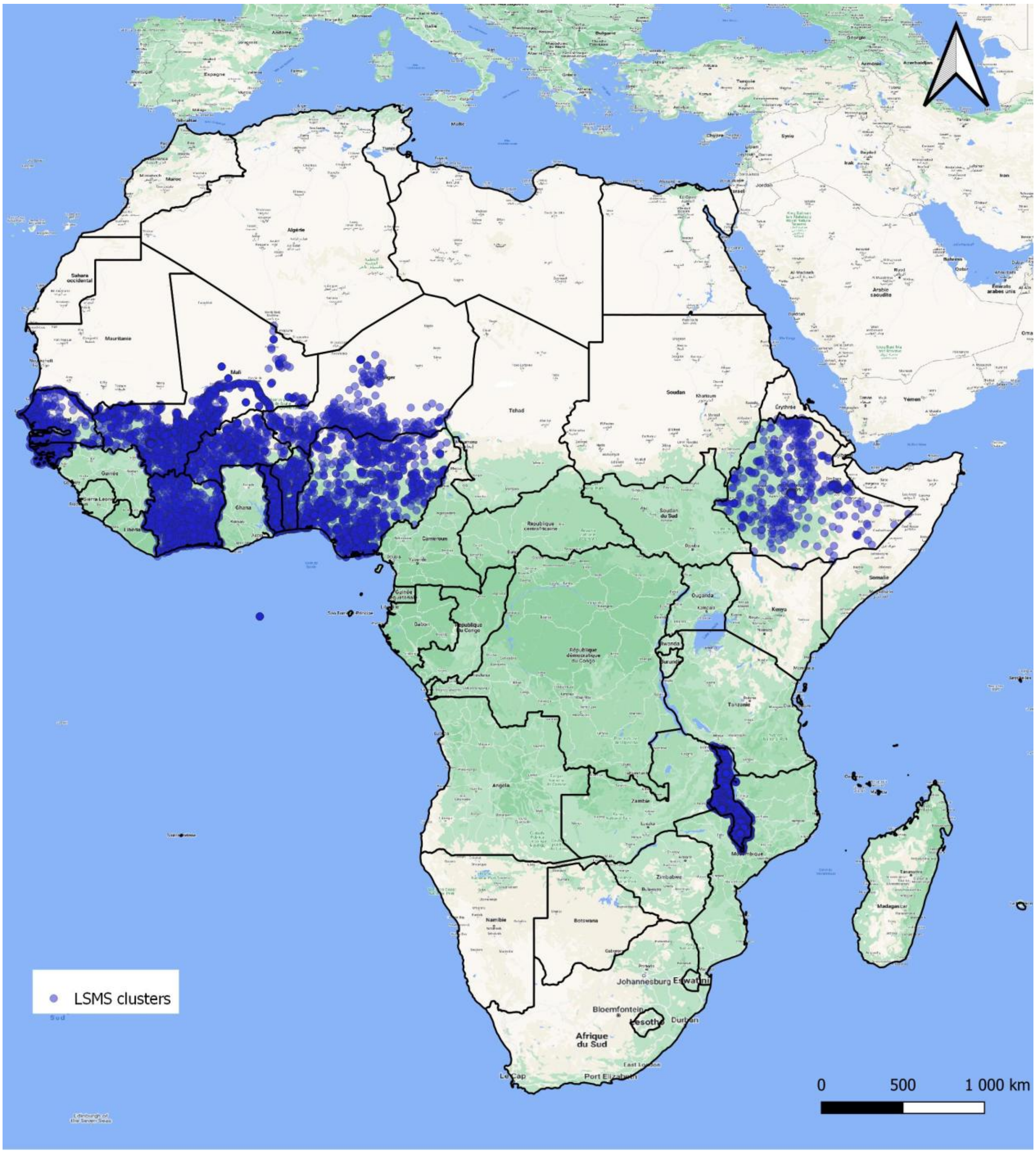


Figure A3: MICS clusters location

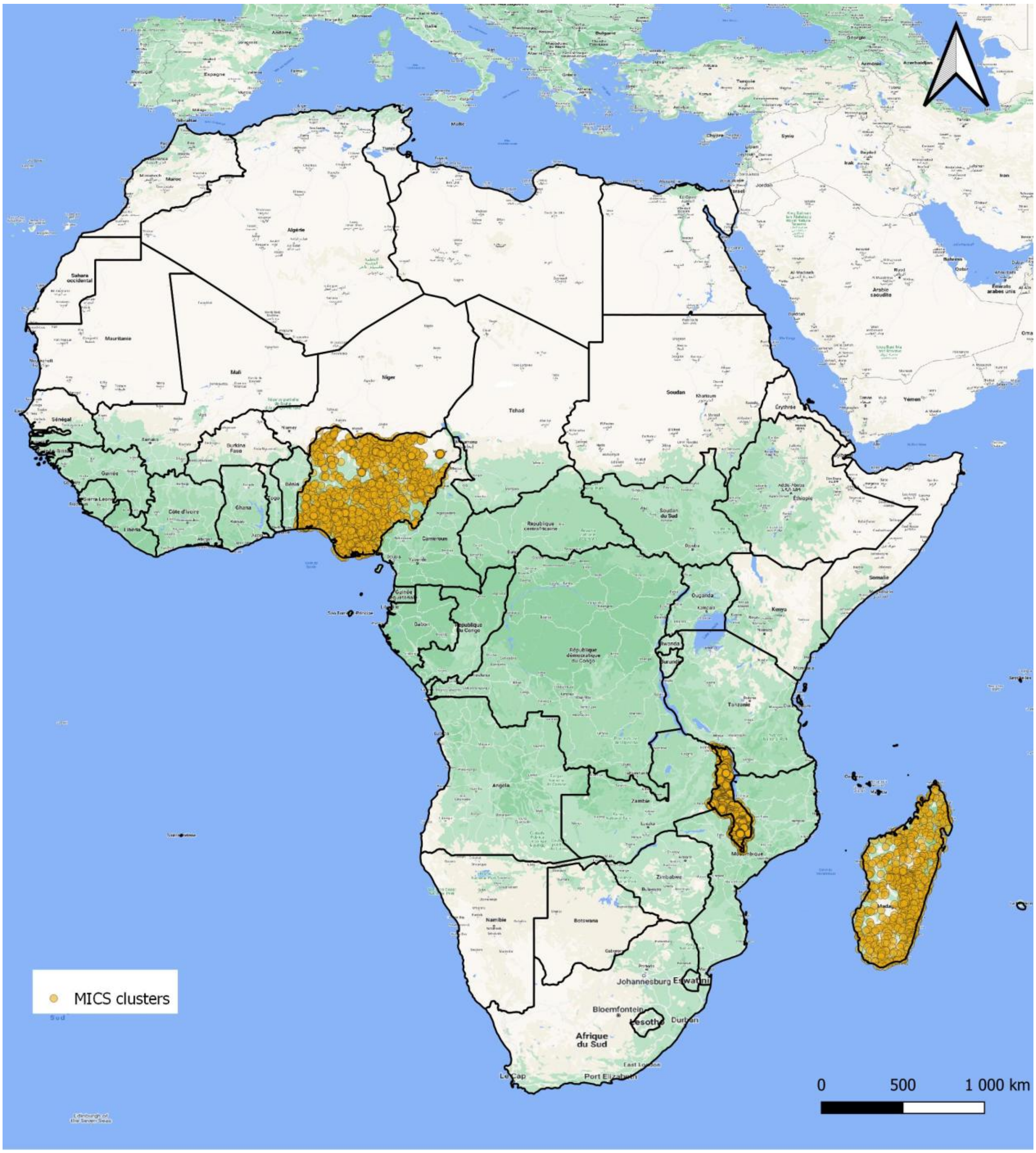


Figure A4: Afrobarometer clusters location

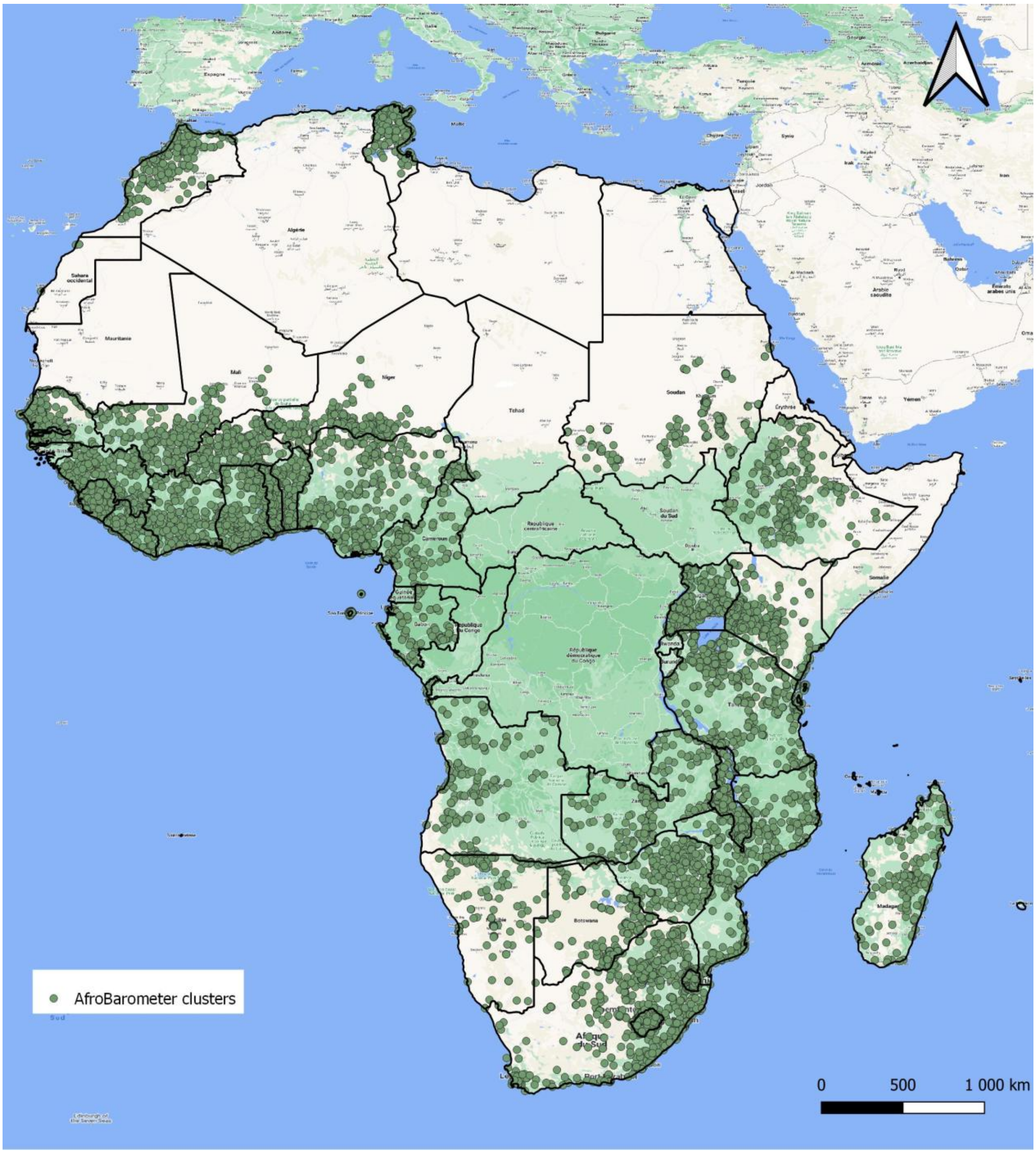


Table A1: Chinese location projects by recipient countries

Country	Number of locations
Angola	7
Botswana	1
Burundi	2
Cabo Verde	2
Cameroon	3
Chad	2
Comoros	2
Congo	2
Cote d'Ivoire	2
Democratic Republic of the Congo	1
Egypt	3
Equatorial Guinea	5
Eritrea	1
Ethiopia	13
Gabon	3
Ghana	6
Guinea	4
Kenya	7
Lesotho	1
Malawi	1
Mali	3
Mauritius	1
Morocco	2
Nigeria	2
Sao Tome and Principe	1
Senegal	1
Sierra Leone	4
South Africa	4
South Sudan	1
Sudan	3
Togo	1
Tunisia	1
Uganda	1
West Bank and Gaza Strip	1
Zambia	13
Zimbabwe	1
Total	108

Table A2: World Bank location projects by recipient countries

Country	Number of locations
Federal Democratic Republic of Ethiopia	1
Republic of Cote d'Ivoire	62
Republic of Kenya	28
Total	91

Table A3: Temporal distribution of energy projects

	China		**World Bank**
Year	Projects commited	Projects completed	Projects commited
2013	17	6	
2014	10	17	
2015	9	18	1
2016	9	15	
2017	8	17	2
2018	3	11	
2019	3	11	
2021	2	4	1

Table A4: DiD estimates for Chinese projects on power production projects based on renewables.

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	-0.00632	0.0035	-0.01318	0.00054	1394	156
Effect_2	-0.00233	0.00359	-0.00937	0.00471	1350	156
Effect_3	-0.01239	0.00478	-0.02176	-0.00301	1315	156
Effect_4	0.03721	0.00807	0.02139	0.05303	1130	152
Effect_5	0.04792	0.01044	0.02746	0.06838	940	141
Effect_6	0.10109	0.0118	0.07796	0.12423	761	133
Effect_7	0.15663	0.01394	0.12932	0.18395	599	131
Effect_8	0.19334	0.01648	0.16104	0.22564	433	121
Av_tot_eff	0.05802	0.00688	0.04453	0.07151	2696	1146

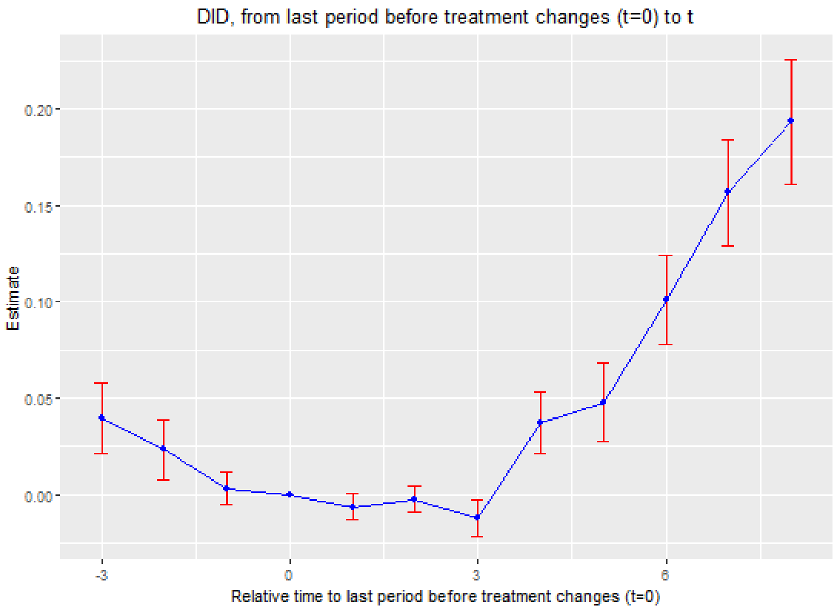


Table A5: DiD estimates for Chinese projects on transmission and distribution projects

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	-0.0052	0.00175	-0.00862	-0.00178	3147	634
Effect_2	0.00016	0.00216	-0.00407	0.00438	3092	634
Effect_3	-0.01003	0.00217	-0.01428	-0.00578	2814	634
Effect_4	0.00626	0.00324	-9,00E-05	0.01261	2536	634
Effect_5	0.01094	0.00361	0.00386	0.01802	2536	634
Effect_6	0.01253	0.00335	0.00596	0.01911	2536	634
Effect_7	0.04844	0.0054	0.03787	0.05902	1624	356
Effect_8	0.06665	0.0055	0.05588	0.07742	1624	356
Av_tot_eff	0.01113	0.00256	0.0061	0.01616	11111	4516

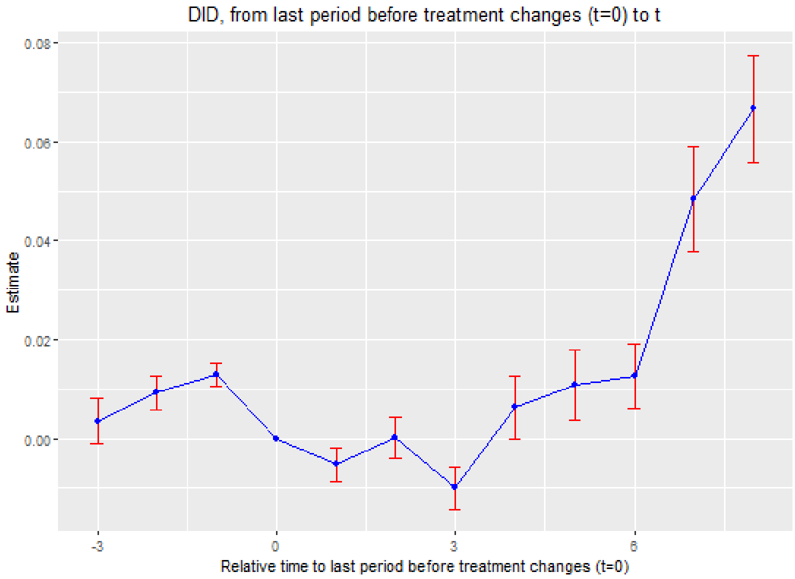


Table A6: DiD estimates of Wold Bank projects on power production based on renewables

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	0.02637	0.00534	0.01591	0.03684	173	57
Effect_2	0.00651	0.00718	-0.00756	0.02057	173	57
Effect_3	-0.01877	0.00639	-0.0313	-0.00625	114	55
Effect_4	-0.03272	0.01385	-0.05986	-0.00558	114	55
Effect_5	-0.02832	0.01478	-0.05729	0.00065	112	55
Av_tot_eff	-0.00902	0.00869	-0.02604	0.00801	629	279

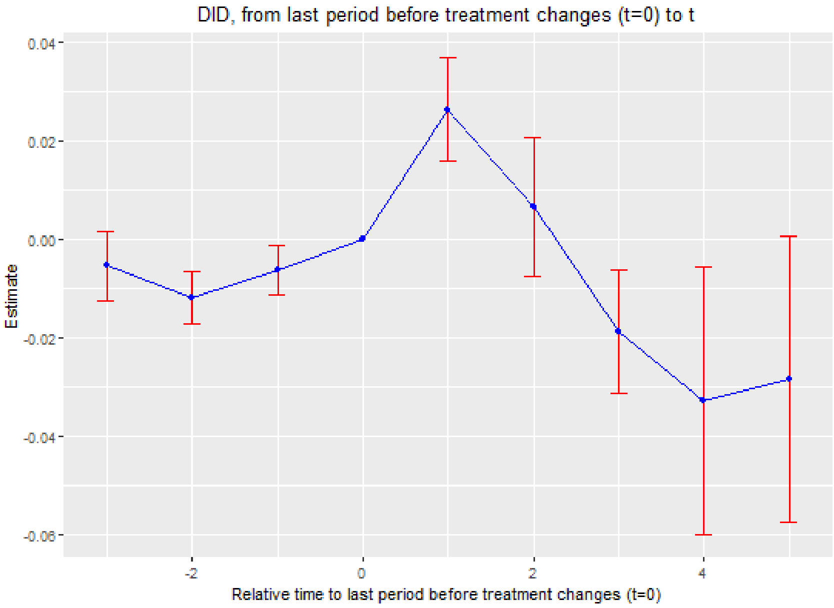


Table A7: DiD estimates of World Bans projects on transmission and distribution

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	-0.00674	0.01226	-0.03076	0.01729	211	53
Effect_2	-0.0155	0.01438	-0.04369	0.01268	211	53
Effect_3	0.11669	0.01342	0.09039	0.14299	159	53
Effect_4	0.16286	0.01861	0.12638	0.19934	159	53
Effect_5	0.20158	0.02171	0.15904	0.24413	159	53
Av_tot_eff	0.09178	0.01429	0.06376	0.1198	740	265

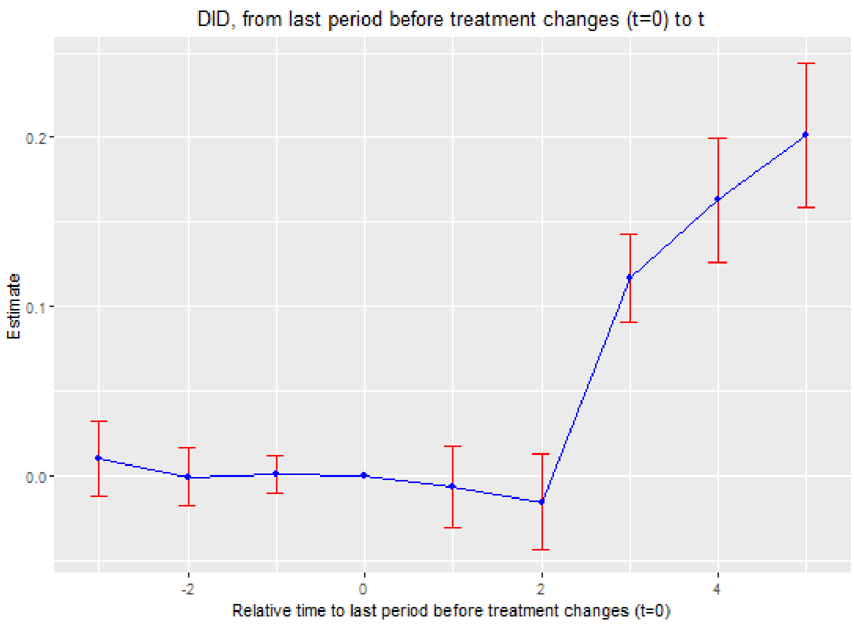


Table A8: DiD estimates of Chinese projects without 2km buffer zone

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	-0.00514	0.00154	-0.00816	-0.00212	10406	720
Effect_2	0.00375	0.00193	-3,00E-05	0.00753	8889	719
Effect_3	-0.0051	0.00224	-0.00948	-0.00071	8497	719
Effect_4	0.0063	0.003	0.00043	0.01217	7174	716
Effect_5	0.0192	0.00353	0.01228	0.02612	6081	710
Effect_6	0.03467	0.00367	0.02748	0.04187	4994	704
Effect_7	0.06021	0.00539	0.04965	0.07077	3686	471
Effect_8	0.09588	0.00757	0.08104	0.11071	2471	328
Av_tot_eff	0.0192	0.00252	0.01426	0.02415	15845	5087

Table A9: DiD estimates of World Bank projects without 2km buffer zone

	Estimate	SE	LB CI	UB CI	N	Switchers
Effect_1	0.0059	0.00722	-0.00824	0.02005	444	57
Effect_2	-0.00898	0.009	-0.02662	0.00866	444	57
Effect_3	0.04824	0.0102	0.02825	0.06823	278	56
Effect_4	0.06235	0.01599	0.03101	0.09369	278	56
Effect_5	0.08251	0.01934	0.04461	0.12042	277	56
Av_tot_eff	0.03772	0.00985	0.01843	0.05702	1278	282

Table A10: DiD estimates of Chinese projects when using completion date as treatment date.

	Estimate	**SE**	**LB CI**	**UB CI**	**N**	**Switchers**
Effect_1	0.01439	0.00163	0.0112	0.01759	14130	1070
Effect_2	0.01674	0.00214	0.01255	0.02093	12016	1067
Effect_3	0.02325	0.0027	0.01795	0.02854	9984	1056
Effect_4	0.0461	0.00344	0.03935	0.05284	7972	1050
Effect_5	0.05463	0.0042	0.04639	0.06287	6467	958
Av_tot_eff	0.03048	0.0025	0.02558	0.03539	18261	5201

Note: Due to the fact we are using completion date we are only able to compute dynamics effects up to 5 periods after T0.

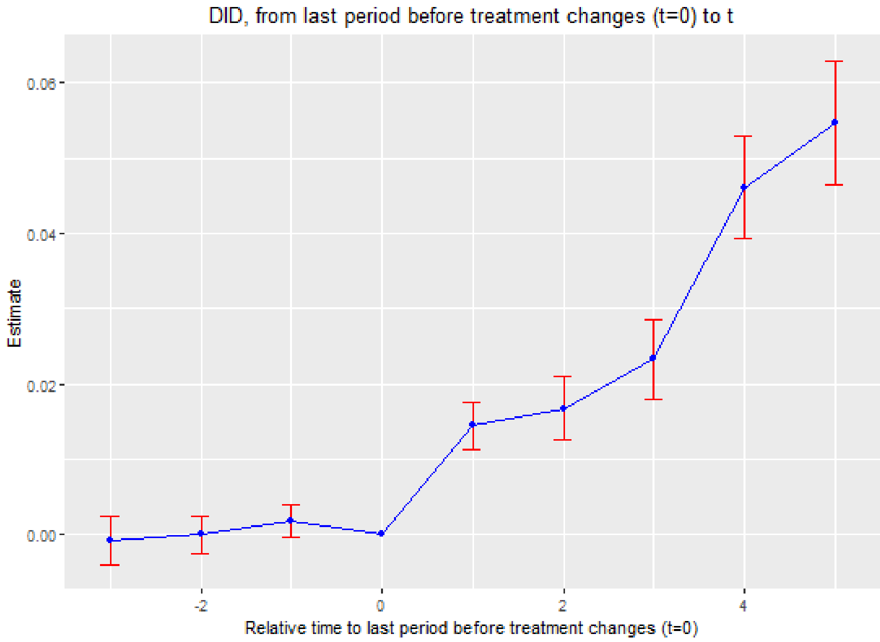


Table A11: DiD estimates of Chinese projects with Callaway and Sant'Anna estimator

event_time	estimate	std_error	conf_low	conf_high	N
0	0.002	0.0012	-4,00E-04	0.0044	
1	-0.0022	0.002	-0.0061	0.0016	
2	0.0079	0.0028	0.0024	0.0134	
3	0.0027	0.0036	-0.0043	0.0098	
4	-0.0091	0.0038	-0.0164	-0.0017	
5	0.0121	0.0046	0.0032	0.0211	
6	0.036	0.0047	0.0267	0.0453	
7	0.0437	0.0058	0.0324	0.0551	
Post-Treatment Avg	0.0117	0.0031	0.0056	0.0178	1551

Note: due to methodological differences in notation between de Chaisemartin and d'Haultfœuille (dCdH) and Callaway and Sant'Anna (CS) event time 0 correspond to different periods, respectively the year before implementation (i.e. reference year for the estimates) for the former and year of implementation for the latter. We therefore display effects from 0 to 7 for CS estimates, which corresponds to effects from 1 to 8 showed in dCdH.

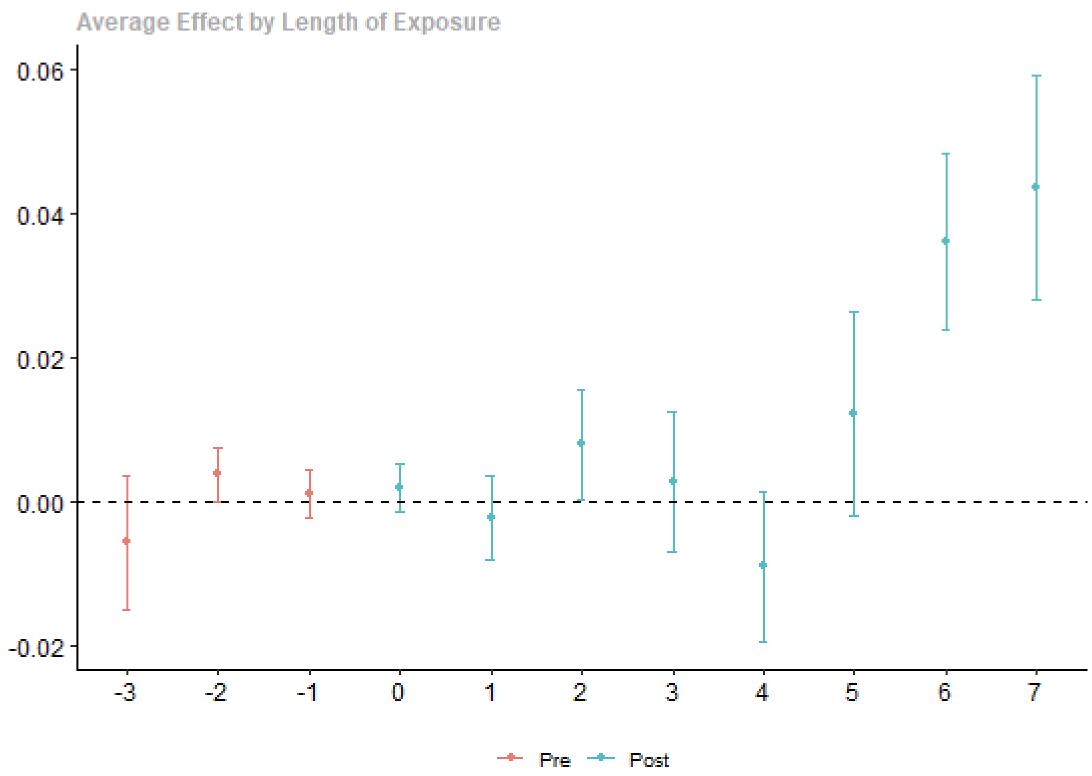
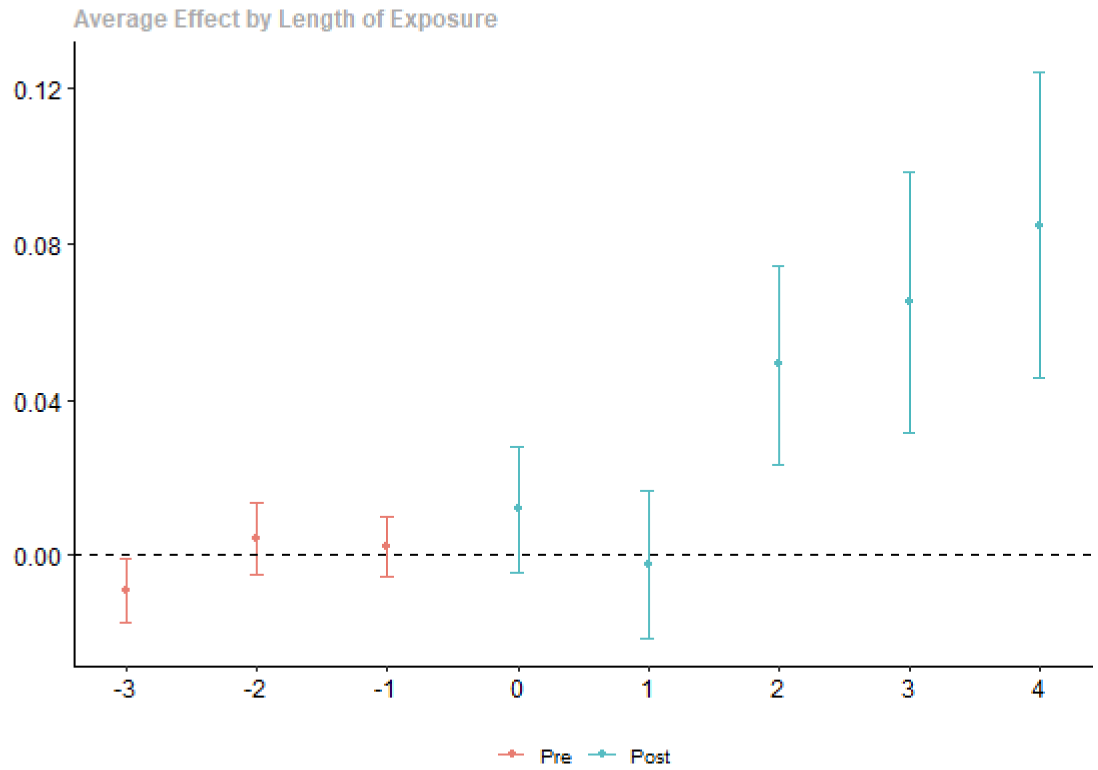


Table A12: DiD estimates of World Bank projects with Callaway and Sant'Anna estimator

event_time	estimate	std_error	conf_low	conf_high	N
0	0.0118	0.0065	-9,00E-04	0.0246	
1	-0.0024	0.0076	-0.0172	0.0124	
2	0.0488	0.0094	0.0303	0.0673	
3	0.065	0.014	0.0375	0.0925	
4	0.0847	0.0161	0.0532	0.1163	
Post-Treatment Avg	0.0416	0.0095	0.0231	0.0601	220

Note: due to methodological differences in notation between de Chaisemartin and d'Haultfœuille (dCdH) and Callaway and Sant'Anna (CS) event time 0 correspond to different periods, respectively the year before implementation (i.e. reference year for the estimates) for the former and year of implementation for the latter. We therefore display effects from 0 to 4 for CS estimates, which corresponds to effects from 1 to 5 showed in dCdH.



General conclusion

Access to electricity is widely recognized as a cornerstone of economic and social development and underpins modern economies by enabling productivity, fostering innovation, and improving living standards. Despite its importance, electricity access remains unevenly distributed, particularly in developing regions where infrastructure constraints and financial limitations hinder progress. This thesis has explored the multifaceted challenges and opportunities associated with electrification, with a particular focus on the role of mini-grids, multiservice energy platforms, and large-scale energy infrastructure projects.

The objective of this thesis was twofold: first, to assess the impact of different electrification models on economic and social outcomes, and second, to develop new analytical approaches to measure and evaluate electricity access more accurately. To achieve this, the thesis combined causal inference methods, randomized controlled trials (RCTs), and machine learning techniques, providing a comprehensive framework for understanding the effectiveness of electrification interventions.

The first essay examined the causal impact of mini-grid electrification on local economies, highlighting its potential to foster income-generating activities and improve social indicators. The second essay presented experimental evidence from an RCT evaluating multiservice energy platforms, shedding light on how access to multiple energy services can enhance household welfare and entrepreneurship. The final essay leveraged machine learning and remote sensing to predict electricity access at a granular level and evaluate the effectiveness of Chinese and World Bank-funded electrification projects in Africa. Together, these studies offer new insights into the determinants of electricity adoption, the effectiveness of different electrification strategies, and the role of data-driven methodologies in policy evaluation. As we turn to the general conclusions of this work, we synthesize the key findings from each essay, draw policy implications, and identify promising avenues for future research in the field of development economics and electrification.

1. Synthesis of Main Findings

1.1 Impact of Mini-Grids on Economic and Social Development (Chapter 1)

The first essay of this thesis investigates the causal effect of mini-grid electrification, addressing a crucial question in the electrification debate: To what extent do mini-grids contribute to improving electricity access and generating meaningful impacts for local communities? Through a robust empirical approach, the study provides causal estimates of mini-grid dynamic effects, offering new insights into the role of decentralized energy solutions.

A key finding is that mini-grids have a significant impact on increasing electricity access in rural and underserved areas. By providing a stable and localized power source, mini-grids successfully bridge the gap for communities where the extension of the national grid remains financially or technically unfeasible. This is particularly relevant for remote areas, where mini-grids serve as a cost-effective alternative to centralized electricity infrastructure.

The analysis also highlights the heterogeneity of mini-grid impacts, showing that their effectiveness depends on several project characteristics and reliability. The study underscores that while mini-grids can significantly enhance electricity access rates, their long-term sustainability depends on both technical and economic viability. One of the key challenges identified is the reliability of mini-grid projects. We found that around half of the projects evaluated seem to have some failure during the evaluation period, which severely impact negatively the overall results.

From a policy standpoint, these findings suggest that mini-grids can play a crucial role in achieving universal electricity access, particularly in rural areas, but their implementation must be carefully designed to align with local energy needs and financial constraints, while addressing reliability concerns.

In conclusion, this study provides clear causal evidence that mini-grids are an effective tool for improving electricity access, especially in off-grid and rural areas. However, the findings also emphasize that their sustainability remains a challenge, requiring further consideration of demand-side constraints, financial viability, technical reliability and integration with broader electrification initiatives.

1.2 Lessons from an RCT on Multiservice Energy Platforms (Chapter 2)

The second essay of this thesis evaluates the impact of multiservice energy platforms through a randomized controlled trial (RCT), focusing on the Café Lumière projects initiative. These platforms go beyond conventional electrification by providing multiple energy services. This study offers a unique experimental assessment of how bundled energy services influence household behaviors, economic activity, and broader development outcomes.

A key finding of the study is that access to a multiservice energy platform significantly increases electricity consumption and the uptake of energy-dependent services. Households and businesses located near Café Lumière sites make extensive use of the available energy services, suggesting that demand for electricity is not only constrained by access but also by the availability of productive and convenient applications. Electricity access therefore seem to drive economic activity and may increase households' wealth, however positive impacts on education, health and public safety are quite limited. The only social positive impact seems to be the improved quality of services provided by healthcare centers.

From a policy perspective, these results highlight the potential of multiservice energy platforms as a viable electrification strategy, particularly in areas where traditional grid extension deployment remains challenging. However, policymakers and project developers should manage expectations with results related to some social aspects as we find almost no significant effect on such topics.

In summary, the findings from this RCT suggest that bundling electricity with complementary services enhances adoption and economic impact. Future electrification strategies should consider how to leverage multiservice energy platforms as part of a broader effort to improve access to both energy and digital infrastructure.

1.3 Machine Learning for Predicting Electricity Access and Evaluating Infrastructure Projects (Chapter 3)

The third essay of this thesis investigates the impact of energy infrastructure projects in Africa, focusing on those financed by China and the World Bank. To overcome the limitations of survey-based data on electricity access, the study develops a machine learning (ML) model trained on remote sensing and geospatial data to generate subnational annual electricity access estimates. These estimates are then used to evaluate the effectiveness of electrification projects using a difference-in-differences (DiD) approach.

A key finding of the study is that electrification projects have a positive and statistically significant impact on electricity access, with effects that increase over time. Treated areas experience an increase in electricity access and the results show that the adoption of electricity in project areas follows a slow but steady trajectory, with continued increases observed up to eight years after project implementation.

The study also finds notable differences between Chinese-financed and World Bank-financed projects. World Bank projects tend to generate higher short-term increases in electricity access, while Chinese projects show a more moderate adoption pattern, with a lower but steady impact over time. These differences may be influenced by project type, implementation strategies, or financing structures, though further research is needed to establish causal explanations.

From a policy perspective, these results highlight the potential of machine learning as a tool for tracking electricity access and evaluating infrastructure projects. The findings also suggest that both Chinese and World Bank energy projects contribute positively to electrification efforts, but their effectiveness varies in terms of speed and magnitude of impact. Future energy policies should consider the long-term dynamics of electricity adoption and explore ways to accelerate access while ensuring sustainability.

In conclusion, the study demonstrates that electrification projects contribute meaningfully to electricity access expansion, though with varying effects depending on the financer. The findings emphasize the need for long-term monitoring of electrification projects and further research into the contextual factors influencing project effectiveness.

2. Implications for Policy and Development Strategies

The research findings outlined in this thesis have significant implications for policymakers and stakeholders involved in the electrification and broader development sectors.

Context-specific Electrification Strategies: The evidence presented strongly suggests that no single electrification model universally fits all contexts. Mini-grids emerge as particularly effective in rural, isolated, or economically disadvantaged communities where extending the national grid is neither feasible nor economically viable. Policymakers should adopt flexible, context-sensitive strategies, considering local geographic conditions, community energy needs, population density, and economic factors when choosing between decentralized mini-grids and centralized grid extensions. Given the observed reliability challenges, policymakers and project planners must prioritize the technical and financial sustainability of mini-grids. Policies should support robust standards and regular maintenance protocols, alongside financial frameworks that ensure long-term operational viability, possibly through public-private partnerships, subsidies for initial infrastructure investments, or innovative microfinancing models for rural energy providers.

Sustainable Project Design and Reliability: Ensuring rigorous technical and reliability standards across all electrification initiatives is critical. Policymakers should mandate regular monitoring and maintenance frameworks, alongside financial sustainability measures, including community-based management and public-private partnerships, as standard components of project implementation.

The Limited but Complementary Role of Multiservice Platforms: Multiservice energy platforms demonstrate that bundling energy services can significantly boost economic activities and energy consumption. However, given the limited observed impacts on broader social outcomes such as education or safety, policymakers should view these platforms as complementary or as a first step towards electrification rather than primary electrification solutions.

Encouraging Productive Uses of Electricity: Policies should incorporate targeted incentives and training programs aimed at promoting productive electricity use. Initiatives focusing on agriculture, small businesses, healthcare, and education can accelerate adoption rates and amplify economic and social returns from electrification investments.

Leveraging Machine Learning and Geospatial Analysis: The use of machine learning and remote sensing data presents substantial opportunities for revolutionizing electrification project planning, monitoring, and evaluation. These approaches overcome critical data limitations, offering policymakers and international donors an accurate, timely, and granular view of electricity access dynamics. Adoption of these methods can significantly enhance targeting efficiency, reducing resource wastage and improving project outcomes. Policymakers should therefore encourage investments in capacity building and data infrastructure to integrate ML-driven analytics into routine electrification monitoring systems.

Financing Models and Donor Strategies: Comparative Insights: Comparing electrification projects financed by China and the World Bank reveals notable differences in effectiveness and adoption trajectories. The evidence tends to indicate that the World Bank's financing strategy yields quicker and higher short-term impacts, potentially due to its integrated project design emphasizing transmission and distribution. Policymakers, co-investors and development practitioners should recognize the strategic implications of these financing approaches, carefully selecting funding models based on project timelines, expected speed of adoption, and desired magnitude of impact. Additionally, ensuring transparency and rigorous impact evaluation should be foundational to both donor strategies.

Addressing Challenges of Long-term Adoption and Sustainability: A key insight from this thesis is the need for sustained focus on the long-term adoption of electricity due to the quite slow and steady nature of the electricity adoption. Initial infrastructure deployment alone does not guarantee immediate, widespread use. Effective adoption often involves broader economic transformations, behavioral changes, and supporting complementary services. Policymakers must consider ongoing strategies to stimulate electricity use, such as subsidies for productive electricity uses, training in electricity-based entrepreneurship, and public-awareness campaigns, thus maximizing infrastructure investments and ensuring sustainable energy transitions.

Overall, these policy insights stress the importance of nuanced, evidence-based electrification approaches, harnessing technological innovations such as machine learning and geospatial analytics, while strategically leveraging various financing and infrastructure deployment models to achieve equitable, sustainable, and impactful electrification outcomes.

3. Limitations and Avenues for Future Research

While this thesis provides valuable insights into electrification and its impacts, several limitations must be acknowledged. These limitations highlight the need for further research to refine methodologies, expand datasets, and deepen our understanding of the long-term effects of electricity access.

Data Limitations and Challenges: A major limitation in evaluating electrification efforts is the availability and granularity of data. While remote sensing and machine learning help address some of these gaps, they are not without limitations, particularly in capturing household-level access and usage patterns. A critical data limitation is the absence of geolocation for many electrification projects, which significantly hinders impact evaluation efforts. Without precise spatial data on where projects are implemented, it becomes challenging to define treatment and control groups in causal analysis, limiting the ability to assess the effectiveness of investments. Many development finance institutions, including multilateral banks and national governments, do not systematically disclose geolocated project data, creating significant barriers to transparent and rigorous evaluation. Future efforts should advocate for a more systematic geospatial reporting for all energy infrastructure projects, which can serve development study research but also local countries administrations in their data collection and infrastructure mapping efforts.

Challenges in Causal Inference and Unobserved Factors: Causal inference in electrification studies remains challenging due to unobserved heterogeneity and confounding factors that may bias results. Although experimental and quasi-experimental methods were employed in this research, they rely on assumptions that may not always hold in real-world contexts. For instance, while difference-in-differences and matching techniques help mitigate bias, they cannot fully account for unobserved variables such as local governance quality, informal energy sources, or pre-existing economic conditions that may influence electricity access and adoption.

Long-Term Socio-Economic Impacts of Electrification: The long-term effects of electricity access on broader socio-economic outcomes remain an open area for future research. While this thesis demonstrates significant economic benefits from electrification, the persistence and depth of these effects, particularly in health, education, and social inclusion, require further investigation. Future studies should adopt longitudinal approaches, tracking households, businesses, and communities over extended periods to assess how electricity access influences income growth, educational attainment, healthcare quality, and migration patterns. Additionally, research should explore how electrification interacts with digitalization and automation trends, particularly in rural areas, to ensure that new technologies do not exacerbate inequalities but rather contribute to inclusive development.

Refinements in ML-Based Predictive Models for Electricity Access: While machine learning and remote sensing have proven valuable in estimating electricity access at a granular level, further refinements are necessary to enhance accuracy and applicability. Current models rely on publicly available remote sensing data and survey sources, but their predictive power could be improved by integrating higher-resolution satellite imagery, survey data with true location (i.e., not noised locations), real-time grid monitoring data, and mobile network usage patterns. Future advancements should focus on deep learning techniques, multi-source data fusion, and real-time updating mechanisms to improve the precision and reliability of electricity access estimates.

Comparative Studies on Financing Models and Electrification Pathways: This thesis highlights key differences in electrification impacts between Chinese-financed and World Bank-financed projects, yet broader comparative studies on financing models remain limited. Variations in investment structures, project governance, and technology deployment strategies across different financiers may have significant implications for project effectiveness. Future research should conduct comparative case studies on the cost-effectiveness, implementation speed, and socio-economic impacts of different financing models, including public-private partnerships, concessional loans, and community-based electrification schemes. Understanding how financing mechanisms influence sustainability, affordability, and adoption rates will be crucial for designing effective energy policies tailored to different national and regional contexts.

Final Thoughts on Research Directions: Addressing these limitations and advancing future research along these dimensions will enhance the understanding of electrification's role in economic development. The insights gained from this thesis provide a strong foundation for these future research avenues, emphasizing the need for interdisciplinary approaches that bridge energy economics, policy analysis, and technological innovations. Moving forward, electrification research should continue evolving alongside global development challenges, ensuring that electricity access serves as a true catalyst for inclusive and sustainable progress.

4. Final Remarks

Electricity access remains a fundamental pillar of economic and social development, yet its effectiveness depends on the mode of implementation, sustainability, and integration with broader development strategies. This thesis underscores the need for a comprehensive approach to electrification, combining mini-grids, large-scale infrastructure projects, and emerging technologies like machine learning to enhance impact assessment and project efficiency. While electrification initiatives generate clear economic benefits, their long-term success hinges on reliable infrastructure, productive energy use, and adaptive financing mechanisms tailored to local contexts. Data-driven policymaking, leveraging satellite imagery, remote sensing, and machine learning, offers new opportunities for improving electrification strategies by enhancing monitoring and evaluation. Ultimately, achieving universal electricity access requires collaborative efforts from governments, donors, and private actors, integrating rigorous evaluation techniques with innovative deployment models to ensure that electrification translates into sustained development and poverty reduction.