Geospatial Cost Drivers in Computer-aided Electrification Planning: The Case of Rwanda

by

Cailinn Drouin

B.S., Mechanical Engineering, The University of Kansas (2012)
B.A., French, The University of Kansas (2007)
Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degree of
Master of Science in Mechanical Engineering
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
February 2018

Author…………………………………………………………………………………………....
Department of Mechanical Engineering
January 15, 2018

Certified by…………………………………………………………………………………..
José Ignacio Pérez-Arriaga
Visiting Professor, Sloan School of Management
Professor, Electrical Engineering, Comillas University
Thesis Supervisor

Certified by…………………………………………………………………………………..
Konstantin Turitsyn
Associate Professor, Department of Mechanical Engineering
MechE Faculty Reader

Accepted by…………………………………………………………………………………..
Rohan Abeyaratne
Professor, Department of Mechanical Engineering
Chairman, Committee on Graduate Students

Geospatial Cost Drivers in Computer-aided Electrification Planning: The Case of Rwanda

by
Cailinn Drouin

Submitted to the Department of Mechanical Engineering
on January 15, 2018 in partial fulfillment of the
requirements for the degree of
Master of Science in Mechanical Engineering

Abstract

This body of work builds upon that of others in the Universal Access Lab at the Massachusetts Institute of Technology who have endeavored to contribute to the resolution of a complex set of global issues by exploring possible solutions to the challenges of achieving universal energy access. To this end, a techno-economic electricity infrastructure modeling software referred to as the Reference Electrification Model, or REM, has been conceived as a tool of potential use in navigating those particular obstacles which lie at the nexus of technology and policy. This thesis explores the incorporation of topography-related cost driver information into the least-cost technology selection, network design, and costing processes of REM. The objective of this initiative is to incorporate geospatial information in such a way as to accurately reflect the challenges associated with construction, operation, and maintenance of electrical infrastructure in the presence of such geographical features as mountains, waterbodies, and wetland areas, as well as human-designated areas which may be considered sensitive to development, as in the case of national parks and wildlife reserves. This thesis also endeavors to provide a critique of REM in its current capacity through the lens of past electrification planning efforts in the country of Rwanda in order to formulate recommendations regarding both the use and further development of this planning tool.

Thesis Supervisor: José Ignacio Pérez-Arriaga
Title: Visiting Professor, Sloan School of Management, MIT
        Professor, Electrical Engineering, Comillas University

Acknowledgments

I would like to thank:

- Matt and the Walnut Street cohort for their unending support.

- Turner, Stephen, Vivian, Roxanne, Matt, and Olamide for the solidarity and good times.

- Claudio in his role as mentor and friend, and for his patience in accommodating my interminable stream of questions.

- Pedro and Andres for their contributions as teammates and all those from IIT Comillas who were so welcoming when we visited.

- Carlos and Fernando for the development assistance they provided.

- My supervisor, Ignacio, for giving me this opportunity.

- Rob for making this experience possible and for being supportive.

Abbreviations

AHP	Analytic Hierarchy Process
ASCII	American Standard Code for Information Interchange
BESS	Battery Energy Storage System
EARP	Early Access Rollout Programme
EDCL	(Rwanda) Energy Development Corporation Limited
ESRI	Environmental Systems Research Institute
FAHP	Fuzzy Analytic Hierarchy Process
GIS	Geographic Information System
ICE	Internal Combustion Engine
IEA	International Energy Agency
LCOE	Levelized Cost of Electricity
MCDM	Multi-criteria Decision Making
MININFRA	(Rwanda) Ministry of Infrastructure
MIT	Massachusetts Institute of Technology
NASA	National Aeronautics and Space Administration
O&M	Operations and Maintenance
PV	Photovoltaic
REM	Reference Electrification Model
RNM	Reference Network Model
SE4All	Sustainable Energy for All
SRTM	Shuttle Radar Topography Mission
TIN	Triangulated Irregular Network
TPDDL	Tata Power Delhi Distribution Limited
UN	United Nations
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WEM	World Energy Model

Table of Contents

List of Figures

List of Tables

Chapter 1

1 Introduction

This body of work builds upon that of others in the Universal Access Lab at the Massachusetts Institute of Technology (MIT) who have endeavored to contribute to the resolution of a complex set of global issues by exploring possible solutions to the challenges of achieving universal energy access. To this end, a techno-economic electricity infrastructure modeling software referred to as the Reference Electrification Model, or REM, has been conceived as a tool of potential use in navigating those particular obstacles which lie at the nexus of technology and policy. As such, the research questions explored herein were selected with the intent of complementing the existing and ongoing work related to both the development and use of REM, and reference to the work of others will be made as required to contextualize the topics explored in this thesis. Previously published theses treating the subject of REM include [1] [2] [3] [4] [5] and [6].

1.1 Background and Motivation

The direct relationship between energy access and poverty is now widely recognized by the global community. The following introductory statements of the International Energy Agency (IEA) publication *Energy Access Outlook 2017: From Poverty to Prosperity* make this point unequivocally:

> Energy access is the "golden thread" that weaves together economic growth, human development and environmental sustainability. Energy has long been recognised as essential for humanity to develop and thrive, but the adoption in 2015 by 193 countries of a goal to ensure access to affordable, reliable, sustainable and modern energy for all by 2030, as part of the new United Nations Sustainable Development Goals (SDGs), marked a new level of political recognition. Energy is also at the heart of many of the other SDGs, including those related to gender equality, poverty reduction, improvements in health and climate change. [7]

Without universal access to food, clean water, adequate health care, and educational resources, the cycle of poverty is impossible to break and economic development is stunted. And without access to adequate supplies of electricity, citizens in developing countries around the world will continue to struggle to meet these basic needs. According to [7], 1.1 billion people around the globe are still without access to electricity. It goes without saying that one of the primary challenges to achieving universal energy access is the absence of sufficient and sustained funding, but that is only part of the battle; the other is figuring out how to use that funding to an effective and efficient end. For this, proper planning resources are required.

Just as the IEA World Energy Model (WEM) has become the trusted standard for world energy demand and supply projections, so would a standard tool, or set of tools, for electricity access planning in developing nations be of great use. The universal energy access agenda would, in principle, be greatly facilitated if all stakeholders were able to evaluate solutions by a consistent means, with the same concrete data. Adoption of a universally respected toolkit for planning purposes could help to convey a sense of security to would-be investors in electrical infrastructure in the developing world. In that way, it would function much as the IEA WEM, which is "...a large-scale simulation model to replicate how energy markets function. It is data intensive, covering the whole global energy system...updated every year and developed over many years..." [8]. Maintained in this way, it is widely referenced and has taken on the status of an accepted standard for reliable information. The value in widespread adoption of such a tool can be seen in Table 1. Estimated investment needs for the attainment of "universal energy access"

vary significantly because they are developed by independent entities, using independent definitions and a wide variety of methods. Table 1 shows estimates developed by nine separate sources for the total investment needed to ensure universal energy access and, even if the extensive range of $12-279 billion USD per year cited as being provided by Bazilian et al (2014) is excluded, values in the table still range widely, from $15 billion USD per year to approximately $55 billion USD per year.

Table 1: Estimates of Investment Needed to Achieve Universal Energy Access (source: [8])

Goal	Investment needs estimates ($ billion/year)		Period	Source
	Electricity	Cooking		
Universal electricity access	45		2011–30	SE4All (2014) Finance Committee
Universal energy access	12–279	18–41	2010–30	Bazilian et al. (2014)
Universal energy access (incremental)[a]	65–86		2011–30	Pachauri et al. (2013)
Universal energy access	44.5	4.5	2011–30	IEA (2012)
Universal energy access	15	71	2010–30	IIASA (2012)
Universal energy access	48		2010–30	Dobbs et al. (2011)
Universal energy access	35–40[b]	39–64[c]	2010–30	AGECC (2010)
Universal electricity access	~55			Saghir (2010)
Universal electricity access	42.9		2005–30	World Bank Group (2006)

1.2   Research Questions

Pursuant to the notion that a decision-support tool could be of potential use in addressing the need for adequate planning resources in developing regions, the MIT Universal Access Lab set about developing REM, ca. 2013, using India as the case study country with support from the Tata Trusts[1]. In April of 2014, the director of the MIT Energy Initiative, Robert Stoner, met with Rwandan president Paul Kagame on his visit to MIT [9] to pursue a collaboration on the study of climate change. Out of this meeting was born a pilot project for the use of REM in planning the electrification of a village in Rwanda [10], with the idea that REM may ultimately be of use for the purposes of a wider electrification planning initiative taking place at the national level. With the advent of this pilot project came a trove of digital information, collected to facilitate previous electrification planning exercises in Rwanda, which was utilized by the REM development team in efforts to understand the ways in which the model may need to be adapted for use in this new case study country. It became immediately clear that Rwanda differs from India in several notable ways, not the least of which being the terrain. The team had long-since realized that the inclusion of geographical features was a capacity wanting for development in REM, but the burgeoning involvement with the Rwandese coupled with the country's unique topography, known widely as the "land of a thousand hills", brought this research issue to the forefront. Thus, efforts were undertaken in 2015 to investigate the possibilities for including topography-related cost drivers in REM, a research initiative which has culminated in the body of work presented in this thesis.

Subsequently, in the fall of 2017, the Universal Access Lab was awarded funding made possible by the World Bank and the Government of Rwanda, to serve as planning consultant, in partnership with Tata Power Delhi Distribution Limited (TPDDL), to the Rwanda Energy Development Corporation Limited

---

[1] See the MIT Tata Center website for further information: https://tatacenter.mit.edu/mission/.

(EDCL) for the purposes of developing a national master electrification plan for Rwanda[2]. In this consulting capacity, and working in concert with TPDDL, the first project task was to review prior Rwandan electrification planning efforts with the objective of isolating the strengths and weaknesses of those strategies. This information was then used to inform a constructive critique of the Reference Electrification Model so that it may be further refined and used in a manner of maximum benefit to planning entities from any region of the world.

It is from these events that the following guiding research questions, which the content of this thesis was developed to explore, were conceived:

1. How can topography-related cost driver information be effectively incorporated into the design processes of REM?
   o What are the existing topography-related capabilities of RNM, a standalone software that functions as the system design algorithm of REM? What adaptations of the interface between RNM the other algorithms which collectively constitute REM are required in order for these capabilities to be utilized in the REM system design process?
   o How is topography-related data utilized within RNM to inform the design process? Can an understanding of these mechanisms provide insight into where, and in what manner, topography-related cost driver information might be incorporated into the consumer clustering algorithms of REM?
   o Can the incorporation of topography-related cost driver information into the REM clustering process be shown to reliably reduce the global system cost output by REM?
2. What can be learned from past electrification master plans developed for Rwanda that can be used to inform both the use and further development of REM?
   o What, if any, shortcomings can be identified in past electrification plans developed for Rwanda?
   o Are there ways in which a techno-economic planning tool such as REM could help to address these shortcomings?
   o What are the limitations of techno-economic planning tools such as REM?

1.3   Overview

In addition to this introductory chapter, this thesis is composed of three other chapters. Chapter 2 will provide a brief introduction to the architecture of the Reference Electrification Model accompanied by a literature review of similar techno-economic computer-aided electrification planning tools. Chapter 3 will begin by exploring the state of the art in the application of geospatial considerations to the general domain of network routing problems. It will then address the methodology used for incorporation of similar capabilities into the REM modeling process. The thesis will conclude with Chapter 4, which includes a brief overview of the country of Rwanda and its current electrification status, a detailed appraisal of past electrification planning efforts, and a critique of the capabilities of REM, including notes on best planning practice and suggested avenues for future development research.

---

[2] Comillas Univeristy project webpage:
https://www.iit.comillas.edu/proyectos/mostrar_proyecto.php.en?nombre_abreviado=Rwanda%20NES%2017

Chapter 2

2 The Reference Electrification Model

Rural electrification planning is ultimately an exercise in how to minimize cost and maximize social impact with the most effective timeline. With so many moving parts, it is impossible to develop a single optimized solution fit for every country. Therefore, REM is being developed under the premise that to be effective in a wide range of contexts, a high level of customizability is required. This level of flexibility is intended to promote the ubiquitous adoption of REM, paving the way for it to become a standard of reference. The traditional method of obtaining this level of context-specific accuracy is to resort to brute force data collection and analysis. An example of this methodology can be seen in the Network Expansion Planning Report completed by the engineering consultant Sofreco, at the behest of the government of Rwanda, in support of the Electricity Access Roll out Program [11].

Sofreco employed a combination of excel spreadsheets, ArcGIS, and CYME power flow software to design network extensions for each individual community identified across the entire country of Rwanda. This method, while accurate, is laborious to update and doing so may even become cost preventative, should any parameter defining the scope or nature of the planning endeavor change significantly. As large-scale planning is a dynamic process, the methods and tools best-tailored to endeavors of this magnitude are those that provide the most flexibility and the greatest efficiency. The REM development team hopes to provide those juggling the unique requirements of planning in the developing world with a tool that is at once comprehensive, customizable, and efficient enough to cope reliably with this difficult and highly-variable planning environment.

The Reference Electrification Model has been conceived as an organizational tool to facilitate the efforts of planners to extend the electrical grid in an efficient and cost-effective manner. To this end, it is intended to serve as, A) an analysis tool enabling policy-makers to compare grid-expansion scenarios and the associated economic implications, and B) a planning tool with which ministries and utilities can identify the most efficient infrastructure expansion options. REM has been created to produce plans having the level of detail achieved by the aforementioned Sofreco study, but in a readily reproducible, updateable, and reconfigurable way. To accomplish these goals, REM has been configured as a single platform capable of using GIS data and user-defined criteria to identify communities, analyze generation tradeoffs, design both electrically [12] and spatially-optimized electrical networks, and produce an associated cost report. In addition to serving as a national-scale planning tool, REM is also capable of producing microgrid designs optimized for communities not scheduled for grid connection in the foreseeable future.

While the MIT Universal Access Lab seeks to offer a tool which is unique in its particular set of capabilities, it is not the first to have considered the ultimate usefulness of such a tool and endeavored to make one available to planners at the international level. What follows is a review of the literature available regarding the computer-aided electrification planning tools which appear to bear the greatest similarity to REM.

2.1 Literature Review: Contemporary Network Planning Tools

Planning for universal energy access is an extremely broad topic of research with many literature tangents which could be of use in informing the development of REM. This review will focus specifically on the availability of electricity infrastructure planning software similar in nature to REM, beginning with an overview of the current state of the art, followed by a discussion of the distinguishing capabilities of REM.

Network Planner
One tool comparable to REM in its intent is the Network Planner, developed by the Sustainable
Engineering Lab at Columbia University. The Network Planner is a web-based modeling tool, with
source code available on GitHub, which is intended for use in electrical infrastructure planning in the
developing country context. It accepts the following as inputs: coordinates of unelectrified communities
and existing network[3] infrastructure, annual interest rate, economic growth rate, population size and
growth rate, mean household size, generation fixed and variable costs, and demand values for four
customer types: household, productive, commercial, and institutional. A journal article detailing the
application of the Network Planner tool to the country of Ghana was the primary document used to
inform the following description of the capabilities of this tool. In this study [13], the objective was to
develop an electrification plan for 2600 communities in Ghana, although further information as to the size
of these communities or the criteria used to identify them does not appear to have been provided in the
publication.

Technology configurations considered included individual household solar PV systems "supplemented by
a diesel generator for productive use" [13], diesel generator-powered mini-grids, and grid electrification.
For the latter, the cost of both LV components, termed 'internal cost', which was based on a mean inter-
household value of 25 meters assumed for all communities throughout the case study country, and MV
components, or 'external cost', were considered for each grid extension. The discounted costs for the
standalone and mini-grid options for each community were compared and the least-cost option selected.
The cost of this least-cost off-grid option was then compared to the discounted cost of the internal grid
cost. If the off-grid cost was lower than that of the internal grid cost, then the grid connection option was
ruled out for that community and it was assigned the off-grid technology as its generation source.

If instead, the internal grid cost was determined to be less than the cost of the off-grid option, then the
difference in cost was divided by the MV-line cost per meter and this value, referred to as 'MVmax', was
considered as the max distance the settlement could be from a potential grid tapoff point for grid
connection to be viable. If the actual distance to a potential tapoff point was less than MVmax, then the
community was slated for grid electrification. Otherwise, it was assigned the off-grid technology
identified as least cost for that particular settlement. It should be noted that in estimation of community-
level demand, consideration was given to the number of social institutions/facilities which would
statistically be a part of a community of a given size, with estimated demand being adjusted accordingly.
The results of this process as applied to the case of Ghana are shown in Figure 1.

---

[3] It is not clear from the article whether the Network Planner is capable of accepting as input an existing electrical
network comprised of more than a single voltage level.
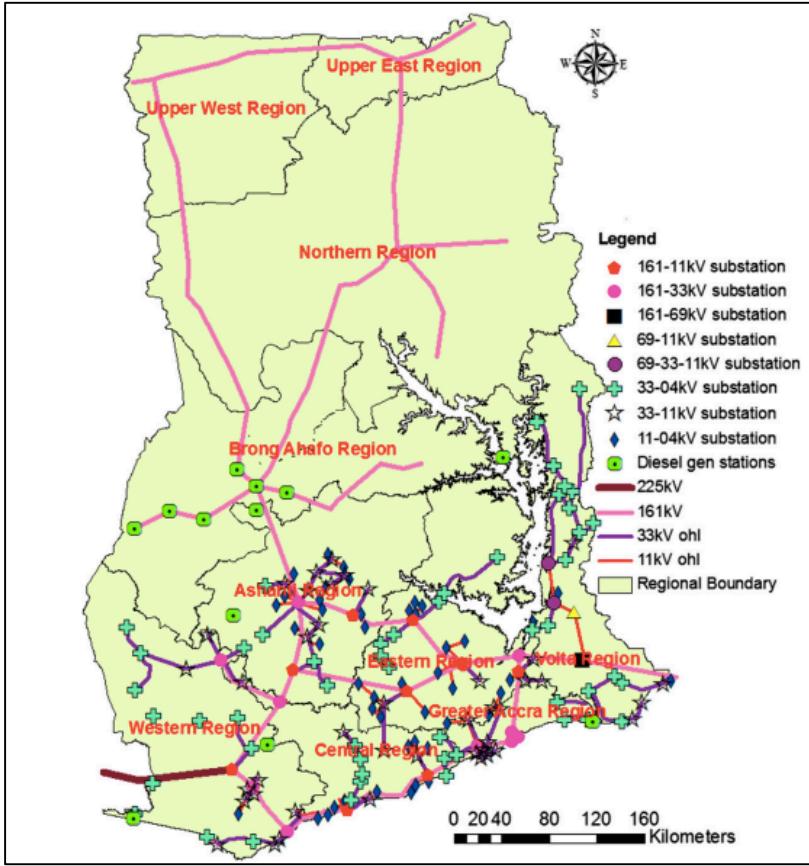
Figure 1: Results from Application of the Network Planner Tool in Ghana (source: [13])

OpeN Source Spatial Electrification Toolkit (OnSSET)

The OpeN Source Spatial Electrification Toolkit, or OnSSET, is a planning tool developed by the Division of Energy Systems Analysis at the KTH Royal Institute of Technology in Sweden. Of those tools currently available for review, OnSSET appears to have the greatest in common with REM in terms of capabilities. It has a dedicated website [14], providing links to all past publications and current projects featuring the tool, as well as a portal through which the web-based version of the tool can be accessed. This access portal indicates that "The online tool is an easy and quick way to explore ONSSET capabilities by conducting an electrification analysis for a selected country in six simple steps.", implying that the interface may actually allow the user to walk through the process from start to finish, even potentially providing his or her own data. However, all links provided on this page require a password, so this particular version of the tool could not be accessed for the purpose of this review. However, python-based source code for OnSSET is available via GitHub [15] , and a link is provided on the GitHub site leading to a United Nations (UN) webpage which provides an online interface entitled 'Universal Electrification Access', which is in some manner 'based on OnSSET' [15]. A description and introduction to the use of this tool, as well as an alternative access point for the user interface, is provided by [16]. This interface allows the user to explore pre-defined results of five least-cost technology analyses performed on each of a large suite of countries in sub-Saharan Africa and Central and South America. Each scenario assumes a uniform national consumption level corresponding to one of the Sustainable Energy for All (SE4All) consumption tiers one through five[4], which in this instance are not defined strictly by service type but rather by kWh of consumption.

---

[4] See section 4.1.3 for further discussion regarding the SE4All electricity access tiers.

Publications consulted in the appraisal of the character and capabilities of the OnSSET tool include [17], which describes application of the KTH least-cost technology assessment methodology in a case study of Nigeria, [18], entitled "A cost comparison of technology approaches for improving access to electricity services," and [19], which chronicles the application of OnSSET to the entirety of sub-Saharan Africa. The modeling process described in [17] begins with the overlay of a grid of cell size 2.5 kilometers on a map of the case study country, Nigeria in this instance. Geo-referenced population density information, projected to the analysis year using an assumed population growth rate, was then used along with similarly projected rural and urban demand estimates, to assign an anticipated load value to each cell. The publication is not explicit about how the demand values were determined, although it does go on to say that "energy access targets were in agreement with the International Energy Agency to be 170 kWh/capita/year for rural and 350 kWh/capita/year for urban areas", with urban areas in Nigeria defined as those having in excess of 20,000 inhabitants. It is thus assumed here that these target access values were in fact the values used to represent anticipated demand levels. The next step was to overlay the existing and planned high voltage network, along with select portions of the medium voltage network, to facilitate a cell-by-cell evaluation of the technology costs associated with each technology option. Technology options considered are summarized in Table 2 and include three modes of energy provision, being that of grid connection, mini-grid, and stand-alone system, and five generation sources, including the national grid, solar PV, wind turbines, diesel generators, and mini-hydro.

Table 2: Technology Options Considered (source: [17])

Category	Supply technology
Grid connection (Grid)	National grid
Mini-grid systems (MG)	Solar PV
	Wind turbines
	Diesel generators
	Mini-hydro
Stand-alone systems (SA)	Solar PV
	Diesel generators

A levelized cost of electricity (LCOE) was calculated for each technology at each cell location, taking into account the following four parameters: demand (kWh/household/year), population density (households/km^2), national cost of grid electricity ($/kWh), local energy resources available, and distance of the cell from "the closest grid connection" (km). It is not specified whether the program is constrained in any way to consider existing transformers or substations when checking for the 'closest grid connection', or rather considers only the closest geographical point to any conductor modeled for this purpose. Based on the level of granularity evident in the figures provided in the document for illustration of the process, of which Figure 2 below is exemplary, it seems most likely that the closest conductor location was considered as a potential tapoff location for the purposes of the analysis. Each cell was then evaluated to determine if the total demand for that cell justified extension of the MV grid to that cell. If a cell identified during this process as a candidate for grid extension was closer to an unelectrified cell than the existing network, then this newly 'electrified' cell was considered as the potential grid tapoff point for the unelectrified cell. All MV grid extensions were limited to a length of 50 kilometers and the minimum demand required to justify grid extension was increased linearly with every step of the grid extension. The latter was used as a mechanism intended to account for the upstream network reinforcement costs that particular extension would incur. After this process was complete, the remaining unelectrified cells were considered as candidates for off-grid electrification. LCOE values developed for off-grid technology options were described as taking into consideration such locationally-specific information as transportation cost of diesel, solar irradiance, wind power capacity factors, and hydropower resource

availability. Results of the Nigeria analysis can be seen in Figure 2. Self-declared limitations of this technology assessment process include lack of load flow analyses and the assumption of overnight costs.
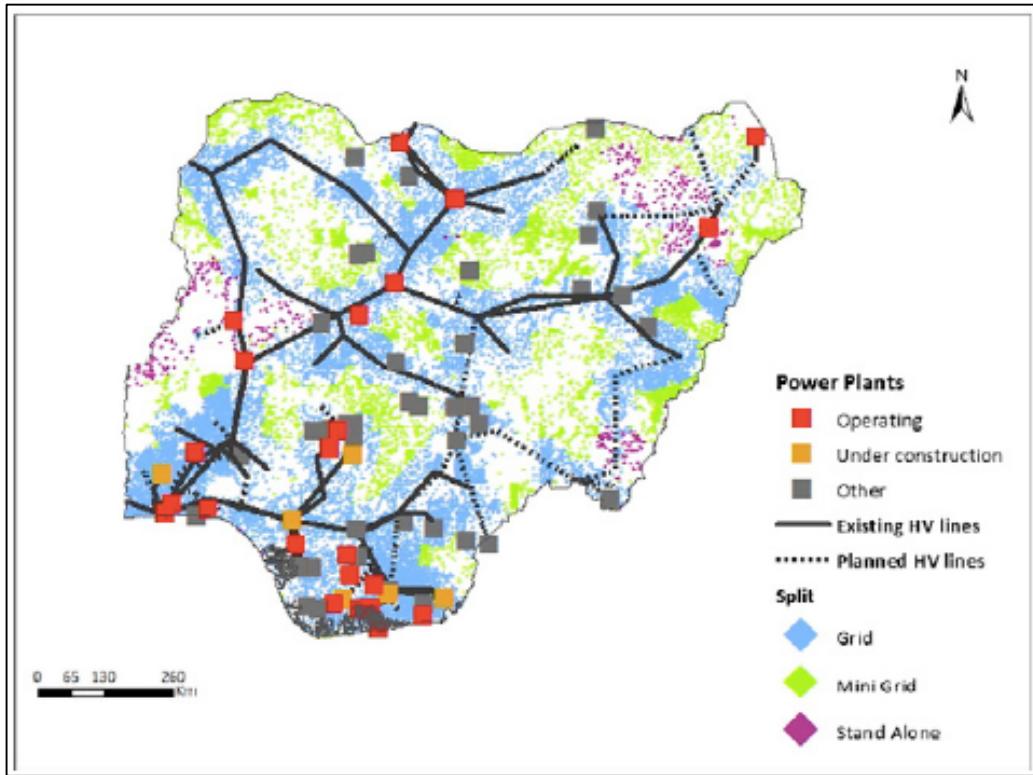
Figure 2: Optimal Nigerian Energy Mix from the Least-cost Analysis Performed in [17] (source: [17])

It is worth noting that although the subject of publication [19] was the application of OnSSET not to a single country but to the whole of sub-Saharan Africa, the technology selection process followed did not appear to differ appreciably from that described in [17], published two years prior, save for the addition of topography-related penalties to grid extension costs. Although it was not entirely clear at what stage in the process the penalties were applied, it is presumed that they were assessed in such a manner as to influence the technology selection. Factors considered in the application of penalties included elevation, proximity to roads, land cover type, slope, and distance from substations. A rubric for the values of these penalties, which were applied to the initial network investment costs, are provided in Table 3.

Table 3: Topography-related Network Investment Cost Penalty Factors (source: [17])

Geospatial factor	Weight	Class A	Class B	Class C	Class D	Class E
Digital elevation (m)	15%	0–500	501–1000	1001–2000	2001–3000	>3000
Distance to roads (km)	5%	0–5	5.1–10	10.1–25	25.1–50	>50
Slope (degrees)	32%	0–10	10.1–20	20.1–30	30.1–40	>40
Distance to sub-stations (km)	9%	0–0.5	0.6–1	1.1–5	5.1–10	>10
Land cover[a]	39%	7,9,10,14,16	2,4	1,3,5,12,13,15	6,8	0,11
Suitability index	100%	5	4	3	2	1

2.2 Conclusions: REM vs State of the Art

- Usability vs. Technical Fidelity**:** The existing web-based tools, that of Network Planner and the UN Universal Electrification Access visualization tool, have an advantage over REM in that they are easy to understand and use. The proper use of REM will require a certain level of user training[5]. This user-friendly advantage appears to come at the cost of customizability and depth of analysis, making these tools best-catered to much higher-level applications than those less superficial planning purposes for which REM is designed. It is worth noting here that although the author of this thesis has accessed the web-based version of Network Planner in the past, the link on the Columbia website which allows access to the online tool has been broken for some time and it is thus not entirely clear if this web-based version is actually still available. Also of note is that the UN plan visualization tool, while based on the KTH OnSSET model, is only available for the purposes of viewing pre-processed scenarios. This site does not provide the user with an opportunity to supply input data or otherwise customize the scenario considered. The reader may also recall from the review of OnSSET in the previous section that according to the KTH website, a web-based version of this tool is available which does accommodate user-furnished input data, but the site was not publicly accessible so no comment can be made at this time regarding the usability of this version of the tool.

- Planning-level Granularity: Based on the information available, REM is still the only tool capable of both considering household-level demand and producing network designs down to the individual consumer level. It is clear that the UN visualization tool, Network Planner, and OnSSET take population data into account, even breaking results down to average household level values, but they do not design networks and generation based on the location of individual customers. OnSSET, for instance, assumes access level targets for country populations falling into one of only two categories, that of urban and rural. REM allows for the customization of both the number of consumer types and the demand profile for each type defined, thus permitting the planner the opportunity to evaluate scenarios based either on globally applied access targets, as done by OnSSET, or on the basis of projections of consumer demand informed by up-to-date knowledge of in-country trends in economic growth and demand for electricity services by type.

- Flexibility: The UN tool does not accept user inputs but is instead a visualizer of predefined, pre-analyzed scenarios. Therefore, it is for reference only, and is not capable of being used as a tool for a user-defined context. The Network Planner and OnSSET can be applied to various case scenarios, as results are based directly on user inputs, but the lack of technical fidelity prevents the user from including many of the characteristics that define particular scenarios and may have a significant effect on planning.

- Consideration of Geospatial Cost Drivers: OnSSET is the only tool, of those reviewed, which appears equipped with the capacity to consider topography-related cost drivers when designing network extensions and when considering cost drivers which affect the identification of the least-cost technology. Although the UN visualizer is based on the OnSSET model, it is not evident from the tool description or the user interface that these capabilities were utilized in the evaluation of the scenarios available for consultation. This is a capability that until recently has been lacking in REM but which constitutes one of the primary topics of analysis covered in this thesis.

- Network Reinforcements: OnSSET is the only tool which appears to consider upstream network reinforcements, but it does so in a fairly crude manner, without the use of load flow analysis. Evaluation of network upgrades is also a capability that is considered by REM in a somewhat

---

[5] It should be noted here that the Universal Access Lab at MIT is currently working with researchers at Comillas University in Madrid, Spain and Engineers at Tata Power Delhi Distribution Limited to develop a web-based version of REM.

basic fashion, although research into the incorporation of load flow capabilities into REM is ongoing.

- Generation Technologies Considered: As REM allows for combinations of solar PV, diesel generator, and battery backup for both microgrid and standalone system configurations, it would appear to offer greater flexibility of off-grid technology choice than does the Network Planner. The UN visualization tool makes mention of additional resources, those of wind and hydropower, but does not go into detail about any combinations thereof which may be considered. As it is also not user configurable, scenarios which vary the technology options considered, prices of these technologies, or fuel prices, cannot be produced. OnSSET also considers generation resources in addition to those that are currently considered by REM and although it does not consider resource combinations in [17], it does specify in that publication that future improvements are to include technology combinations such as solar-diesel and wind-diesel hybrids.

2.3 REM: Inputs and Outputs

Inputs

The exact nature of the input file requirements of the Reference Electrification Model varies by application according to user preferences and data availability. In addition, a web-based version of the tool is currently under development which will be equipped with some level of data preprocessing capabilities, allowing the user to provide information, through a user interface, in a format other than that which will ultimately serve as input to REM. For the purposes of this paper, a summary of the input files for the version of REM currently used for development has been prepared and is available in Table 4. Further information regarding the contents of the files listed in Table 4 can be found in [20].

Table 4: Inputs Files Used with the Developer Version of REM

Folder Name	File Name	File Type	Content	Required or Optional?
loadProfiles	baseProfile_1	csv	Base consumer demand profile	Required
	profile_equivalence	csv	Consumer profile types based on linear combination of base consumer demand profile	Required
	profileLibrary	mat	Hourly demand values for each consumer type for every hour of the year	Required
	study_coordinates	csv	Points of the generation look-up table used in REM to determine off-grid generation designs	Required
RNM_input	altRaster	txt	ESRI ASCII format raster containing altitude data for study region	Optional
	catalog_ext	dat	Catalogue of grid extension network components for use in RNM	Required
	catalog_ug	dat	Catalogue of microgrid network components for use in RNM	Required
	caso	caso	RNM case file containing commands and filepaths for files called by the model	Required
	forbRaster	txt	ESRI ASCII format raster containing penalized zone data for study region	Optional
	ZonasP	txt	RNM penalized zone polygon file	Optional
	[name of pre-clustering boundary shapefile]	shapefile component files (dbf,	Village boundaries to be used if simultaneous clustering algorithm used and pre-clustering options selected	Optional

		prj, qpj, shp, shx)		
Not Applicable	buildings	csv	Projected geographical coordinates of consumers, consumer type, consumer electrification status	Required
	catalog_RNM_brief	xlsx	RNM network catalogue values for use in Clustering module	Required
	gencatalog	xlsx	Catalogue of specifications for all generation components and auxiliary equipment, including solar panels, batteries, diesel generators, converters, and charge controllers	Required
	MV_branches	csv	Reliability, cost of energy, geographical coordinates of each feeder of the existing network	Required
	Network_Data	xlsx	LV/MV/HV conductor sizes	Required
	options	csv	User preferences and user-specified run characteristics (e.g. which clustering algorithm will be used, diesel cost, minimum number of users to be considered for a microgrid, UTM zone of study region, etc.)	Required
	PV_DC	csv	Expected hourly gross production of 1 kW of installed solar PV capacity in the study region	Required

It goes without saying that the higher the fidelity of the data provided to REM, the greater the accuracy of the results obtained. Thus, wherever possible, a database containing information regarding the location, electrification status, and consumption of consumers, as well as data concerning the location and operational statistics of existing electrical infrastructure, should be developed and maintained[6]. That said, knowing that the existence of such a complete and regularly maintained database is still not the reality in many locations, the Universal Access Lab has developed methods for estimating both the location and electrification status of consumers. These methodologies are not without certain limitations, but may nevertheless be of great use in locations where records of consumer information have not been maintained. For information regarding the estimation of consumer electrification status, the reader should consult [5]. To learn more about the identification of consumer location, details of the methodology developed at MIT will be available in the forthcoming thesis by Stephen Lee.

Outputs
Result files currently output by REM include: shapefiles for the networks designed, including conductors and transformers, the consumers for which they were designed, the location of microgrid generation centers, and the location of customers receiving standalone systems; three summary spreadsheets detailing results pertaining to the entire study region (one full results file, one summary results file, and one file with results listed by customer); two spreadsheet files for each microgrid containing cash flow and dispatch information, and a dispatch file for the standalone system recipients by customer type. The contents of the primary region-level spreadsheet have been summarized in Table 5. Further information regarding the contents of all REM output files can be found in [20].

---

[6] See Chapter 4 of this thesis for a more in-depth look at the collection and use of data for the purposes of electrification planning.

Table 5: Summary of REM Output File Contents (adapted from [20])

Object	Field Title	Units	Description
Name	name	N/A	Unique identifier of the project.
Type	type	N/A	Type of project.
Cluster Type	clusterType	N/A	Type of cluster that was evaluated and selected for the electrification mode.
Number of Users	nUsers	N/A	Number of users belonging to the cluster.
Grid Energy	grid_energy	$\frac{kWh}{year}$	Energy consumed from the distribution network every on an annual basis.
Grid Energy Cost	grid_energy_cost	$\frac{USD}{year}$	Amount paid annually for the energy consumed from the grid.
PV Installed Capacity	PV_power	kW	Rated power of the PV array.
PV Annuity	PV_annuity	$\frac{USD}{year}$	Annuitized cost of the PV panel including O&M.
BESS Energy	BESS_energy	kWh	Rated battery system capacity.
BESS Annuity	BESS_annuity	$\frac{USD}{year}$	Annuitized cos of the battery including O&M and replacement.
ICE Power	ICE_power	kW	Rated installed capacity of internal combustion engine generation.
ICE Annuity	ICE_annuity	$\frac{USD}{year}$	Annuitized cost of the ICE, including O&M.
Fuel	fuel	$\frac{L}{year}$	Annual fuel consumption.
Fuel Annual Cost	fuel cost	$\frac{USD}{year}$	Annual fuel expenditure.
High Priority Non-served Energy	hiPtyNSE	$\frac{kWh}{year}$	Amount of high-priority non-served energy.
Low Priority Non-served Energy	loPtyNSE	$\frac{kWh}{year}$	Amount of low-priority non-served energy.
Peak Demand	demandPeak	kW	Maximum consumption.
Demand	demandEnergy	kWh	Yearly energy demand.

2.4   REM: Overview of Software Architecture

A process flow diagram for the Reference Electrification Model is shown in Figure 3. Section 2.3 of this thesis provided an overview of the character of both input files required by REM and output files generated at the end of the REM evaluation process. This section will discuss briefly the internal operations of REM, using Figure 3 as a guide. It should be noted at the outset that as the items labeled in Figure 3 as *Clustering*, *Off-grid Supply*, *Microgrid Network*, and *Grid Extension* can theoretically be rearranged according to the needs of the developer, these are considered as representing the basic modules in the REM electrification toolbox. See [21] for further elaboration.
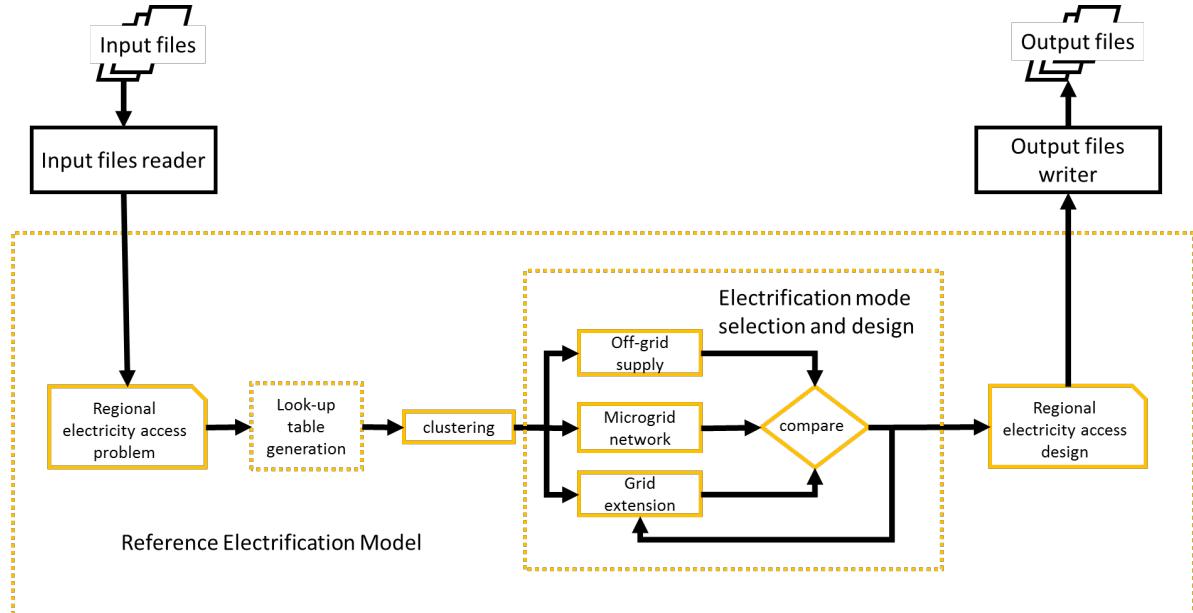
Figure 3: Reference Electrification Model Process Flow Diagram (source: [21])

The regional electricity access problem is wholly defined by the input files provided to REM. Once the input files have been processed and the information stored in Matlab structure format, the microgrid generation look-up table is created for use in the Clustering module. The generation look-up tables contain generation configurations (solar PV panels, electrochemical battery, diesel generator and any necessary power electronics) for microgrids supplying a representative set of customer combinations. Values are obtained from these tables, by means of interpolation when necessary, for each candidate cluster evaluated during the clustering process. This is done in order to avoid the costly operation of computing the optimal generation configuration for each cluster considered. Once the look-up tables have been created, all analysis region consumers are partitioned into quasi-optimal[7] groups or 'clusters' through an evaluation process which considers electricity provision from only off-grid sources. These clusters are then evaluated once again, this time with only grid extension considered. During this stage of analysis, the clusters formed originally in the off-grid clustering will be evaluated as grid-extension candidates both individually and as groups. The outcome of this process is a set of clusters recommended as candidates for grid extension.

With the clustering process complete, the three modules *Off-grid Supply*, *Microgrid Network*, and *Grid Extension*, are called. These create, respectively, the off-grid generation designs, microgrid network designs, and grid-extension designs for the two sets of candidate clusters. Whereas the generation and network designs were previously approximated in order to facilitate the clustering decision-making process, both are now designed in detail specific to the demand makeup of each cluster – including the geographical location of each consumer – for the purpose of electrification mode comparison. The comparison step involves the following three analyses:

1. For each off-grid candidate cluster, the best off-grid electrification mode, either standalone system or microgrid, is selected.

2. For each grid-extension candidate cluster, the cost of connecting it to the grid is compared with the sum of the off-grid electrification costs which were determined for each constituent cluster

---

[7] The clusters are referred to as 'quasi-optimal' because, as the clustering algorithms are based on heuristic processes, the results are not guaranteed to represent the mathematically optimum configuration.

(i.e. for each cluster originally issued from the off-grid clustering stage which was included in the formation of this grid-extension candidate cluster). If the sum of the off-grid electrification costs is less than the grid-extension cost, all constituent clusters are assigned to off-grid electrification mode.

3. For each constituent cluster relegated to off-grid electrification mode in the previous step, compare the cost of off-grid electrification with that of grid extension for this single cluster. (This step is represented by the arrow in Figure 3 which indicates a post-comparison process return to the Grid Extension module).

Following the design and compare stage, designs corresponding to the least-cost electrification modes identified are saved to the run results file and are ready for planner evaluation.

Chapter 3

3 Geospatial Awareness

This chapter chronicles efforts made to date to render the Reference Electrification Model (REM) least-cost technology selection, network design and costing processes geospatially aware. Geospatial awareness, in this instance, refers to the incorporation of topography-related cost-driver information in such a way as to accurately reflect the challenges associated with construction, operation, and maintenance of electrical infrastructure in the presence of such geographical features as mountains, waterbodies, and wetland areas, as well as human-designated areas which may be considered sensitive to development, as in the case of national parks and wildlife reserves.

3.1 Problem Definition and Working Hypotheses

The task of making REM geospatially aware has two distinct components: 1) harnessing the existing capabilities of the Reference Network Model (RNM), and 2) incorporating topographical cost-driver information into the clustering decision-making process. As described in section 2.4, the modules responsible for final system design are Off-grid Supply, Microgrid Network Design, and Grid-extension Design, the latter two of which are based on a standalone distribution network design software program referred to as RNM. The RNM software is already equipped with the capacity to recognize topography, but this capability has not previously been used within REM.

Activating this capability will serve to correct the designs currently being produced by REM for realism; that is to say, designs are currently being developed under the assumptions that all terrain is flat and that there are no topographical obstacles or sensitive zones that would incur additional costs if traversed. As will be discussed further in the remainder of this chapter, RNM can recognize the existence of such geographical features and utilize user-specified penalties associated with these features to make network routing decisions which minimize cost. However, it is important to understand that once RNM has received the generation clusters, these clusters may no longer be reconfigured. Thus, such instances may arise as RNM designing a network that connects consumers which are separated by a zone or obstacle that is costly enough to traverse or circumnavigate that it would have been less expensive to electrify these consumers by separating them into two or more clusters.

It is for this reason that an attempt is being made to render the REM clustering module topography aware as well. Unlike RNM, the clustering module is not yet equipped with the capacity to recognize topography, so this capacity must be developed. In order to accomplish this, it is first important to understand the nature of the clustering decision-making process and to consider the relationship between the clustering and final design modules.

The clustering algorithm makes decisions based, in part, on approximations of the network costs which would be calculated by RNM, were it called to design full networks during the clustering decision-making process[8]. It also uses generation costs which are approximated by means of interpolation and considers only a subset of all possible clustering combinations. We therefore recognize, given the heuristic nature of the clustering process, that REM is not guaranteed to produce mathematically optimal results, and that there exists the potential for improvements in the heuristic which may lead to lower cost results than those currently being produced. From this we establish the following working hypotheses:

---

[8] These network cost approximations are justified by the computational time savings they are presumed to afford; the actual advantage of this heuristic has not been verified.

o Incorporating topography-related cost driver information into clustering will result in network cost approximations which more closely resemble the true network costs RNM would calculate if called to design networks during the clustering process, as both modules would then have access to the same cost-driver information.

o The more closely the network cost approximations made in clustering resemble the network costs that RNM would calculate, the greater the probability of arriving at the global optimal least-cost solution for the entire analysis region.

3.2 Literature Review: Geospatial Considerations in Least-cost Route Selection

The objective of this literature review is to identify and describe methodologies used by others to make consideration of topographical attributes a feature of their network route selection processes. The domain of network route optimization is expansive, as is the pool of literature available for review on this topic. While those references considering explicitly the geospatial aspects of the routing problem constitute a subset of the routing literature available, those available for consultation are nevertheless also numerous. An effort has been made here to select, from among those in this geospatial subset, a collection of articles which together serve to paint a holistic picture of the state of the art.

Review
Of the sources consulted, those cited in this review include [22] [23] [24] [25] [26] [27] [28], all of which concern least-cost network path identification using GIS platforms to incorporate geospatial information into the route selection process. The network routing problem concerns not only electrical infrastructure but other transport systems as well, including natural gas and oil pipelines and roadways, which constitute the networks of primary concern in several of the sources considered. The nature of the network does effect the set of topographical regions or 'obstacles' considered to a certain extent, due to domain-specific concerns such as the minimum turning radius required for pipelines. However, by and large the features considered as having the potential to incur added cost or induce a preference for a circuitous route were the same and fell into the three general categories defined by [27]: environmental, economical, and sociological. Typical topographical attributes of concern included slope of terrain, waterbodies, land cover type, land use, geological makeup, and protected areas. All sources consulted made use of geographic information system (GIS) software, ArcGIS in the majority of the cases, to compile GIS data from various sources into a single database for use in identification of the aforementioned zones of concern. In all cases, this information was combined in order to create a raster of cost weights which could then be used in the selection of the least-cost route from point A to point B. Weighting criteria were often in the form of values determined by and applied at the discretion of the planner, as was the case with [22] [23] [24] [25] and [28]. In the case of [26] and [27], multi-criteria decision-making methods (MCDM), including the analytic hierarchy process (AHP) and fuzzy analytic hierarchy process (FAHP), were employed to determine how information was used in the application of weights.

Excepting the cases of [25] and [27], for which manual route selection was performed by the planner, all researchers employed variations of a discrete optimization method whereby a finite set of predefined nodes and branches was used, along with Djikstra's shortest path algorithm, to construct a least-cost network, with branches weighted to account for topographical features considered as incurring additional investment. One unique application of this discrete node and branch optimization approach included the method used by [24], whereby the geomorphology of an area was represented with a triangular irregular network (TIN), a digital terrain model in the form of a Delaunay triangular mesh, with nodes including both source-sink pairs and Steiner points (non-source-sink waypoints) and Delaunay branches representing network branch candidates. Another method, variations of which were employed by [22] and [23], involved the use of raster files, where the centers of the cells were considered as nodes, each representing either a source, sink, or waypoint, and the distance between the centers of any two adjacent cells was considered a branch. Movement was only allowed to take place in the eight cardinal directions,

with a standard cost assigned to each branch and weights applied to the nodes. In [26], the least-cost path selection process was performed within ArcGIS, using the Cost Distance and Cost Path tools from the Cost Distance Toolbox, tools which according to [26] are based on the Djikstra algorithm, although the author of this thesis was unable to find information on the ArcGIS website to corroborate this statement.

Conclusions
It is evident that the use of GIS software and cost optimization methods has become commonplace in the realm of network design. Over the course of the 21st century, methods have evolved from the informed manual manipulation of routes in a trial and error quest for a 'cost-optimized' route to the use of more sophisticated, and at least quasi-exhaustive, computer-aided dynamic optimization programs. It would appear from the literature consulted that discrete optimization methods, with particular preference for Djikstra's shortest path algorithm, have come to dominate as a result of the ease with which they are integrated with the GIS software raster format and for the reduction in solution space they afford. Where geographically-driven cost-weighted values or penalties are concerned, there would seem to be little dispute that these play an instrumental role in effective route selection. It is also clear, however, that the exact values and combinations thereof to be assigned to various spatial attributes is still more of an art than a science, and that although there are various methods suggested for the assignment of these cost weights, it is still ultimately up to the informed planner to adjust these based on a common-sense appraisal of the design results.

3.3 Problem Solving Approach

Following from the discussion in the introductory section of this chapter, the topography-recognition capabilities of RNM will be activated for use in REM and similar capabilities will be incorporated into the Clustering module, which encompasses both the sequential and simultaneous versions of the clustering algorithm. As apparent from the literature review of least-cost route selection precedents, there are a multitude of ways to approach this problem. However, as the reigning hypothesis ventures that incorporating the same cost driver information into the clustering process as is available to RNM will bring the clustering network cost approximations closer to the network costs which would be approximated by RNM, it follows that the manner in which this information is used in RNM should also, to the extent possible, be replicated in clustering. It should be noted that although this approach might seem obvious to the reader, it may be that the clustering algorithms are not well-conditioned to utilize cost driver information in the same manner as it is employed in RNM. Nevertheless, it would seem a reasonable place to begin.

Thus, the first step is to develop a thorough understanding of what the relevant capabilities of RNM are and how to use them. These mechanisms will then be harnessed for use in REM and the resulting system designs appraised. Once this is accomplished, we will identify the locations in both clustering algorithms where it is appropriate to incorporate the mechanisms used by RNM, discuss any challenges encountered in the incorporation process, and revisit the working hypotheses through a review of the resulting clustering configurations and system designs. Sample data from the country of Rwanda will be used to aid discussion regarding input files, the development process, and process results.

3.3.1 Harnessing Topographical Capabilities of RNM

This section will explore the use of geospatial cost drivers within RNM by first describing the topography-related capabilities of this software and the corresponding input files required. This will be followed by a demonstration of the effects produced when these capabilities are employed in the generation of REM system designs. The section will conclude with a deep dive into the configuration of the cost-adjustment mechanisms within RNM.

Provided with the correct input files, as described hereafter, RNM is capable of adjusting both network costs and network routing based on altitude, slope, and zones with penalties associated. The two input files required by RNM in order to perform the aforementioned functions are an altitude raster and a zonal polygon file. RNM can accept altitude information in Environmental Systems Research Institute (ESRI) ASCII raster format[9]. National Aeronautics and Space Administration (NASA) Shuttle Radar Topography Mission (SRTM) digital elevation data covering all countries is available in either 30-meter[10] or 90-meter resolution from the United States Geological Survey (USGS)[11] website. An example of the 30-meter SRTM data obtained for the country of Rwanda is displayed in Figure 4. This data can be downloaded in TIFF format, projected into the appropriate Universal Transverse Mercator (UTM) coordinate system, and converted to ESRI ASCII raster format using a GIS software such as ArcGIS.
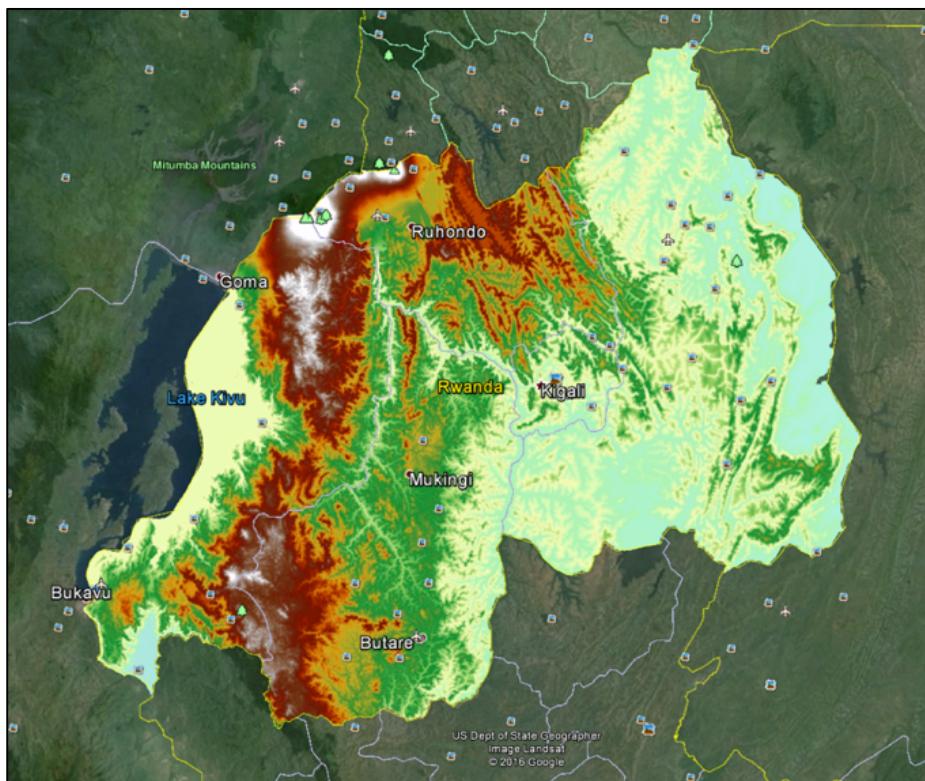
Figure 4: NASA Shuttle Radar Topography Mission Rwanda Altitude Data – 30m Resolution

If the planner would like geographical features such as lakes and wetlands, or other designated zones such as national parks, to be considered in the network route design process, then a polygon file must be created using an executable function entitled 'shp2rnm' which is executed through the computer command window. This function requires as inputs the shapefiles having the extension .dbf and .shp and containing information for all zones to which the user wishes to apply penalties in polygon feature format. Once created, the RNM input polygon file will contain, for each polygon, a set of defining coordinates

---

[9] A description of this format can be found at
http://resources.esri.com/help/9.3/ArcGISdesktop/com/gp_toolref/spatial_analyst_tools/esri_ascii_raster_format.htm
[10] It should be noted that the spatial resolution of the 30m and 90m data is actually one and three arc-seconds, respectively, which is approximately 30 (90) meters at the equator.
[11] The USGS SRTM Repository is available here: https://lta.cr.usgs.gov/SRTM.

accompanied by a header where the user may specify an investment penalty multiplier value. The default value is ten, but may be changed to any value, including decimals. This value will be used to avoid penalized areas in network routing decisions. However, when the actual cost of the line is calculated, the cost penalty multiplier is capped at ten. In other words, if a value less than ten is entered, the line cost will be multiplied by that value. But, if a value greater than ten is entered, then the cost will be multiplied by ten. Further notes on input file creation are available for reference in Appendix I.

3.3.1.2   REM Results

The following figures were produced with REM to illustrate the effects of providing the altitude and penalized zone input files to RNM. Consumers are represented by beige circles, stepdown transformers by larger yellow circles, LV conductors in blue, and MV conductors in red. Penalized zones[12], in this instance comprised of lake, river, and wetland areas, are shown in light blue and altitude ranges are represented by a gradation of colors ranging from mauve, starting at 1,364 meters, to green, with the highest altitude displayed in the figure falling in the range of 2,455-2,815 meters. The network design shown in Figure 5 was produced by RNM in the absence of altitude and zone input files. Conductors follow straight line paths in almost all cases, traversing water bodies and changes in altitude without discrimination.
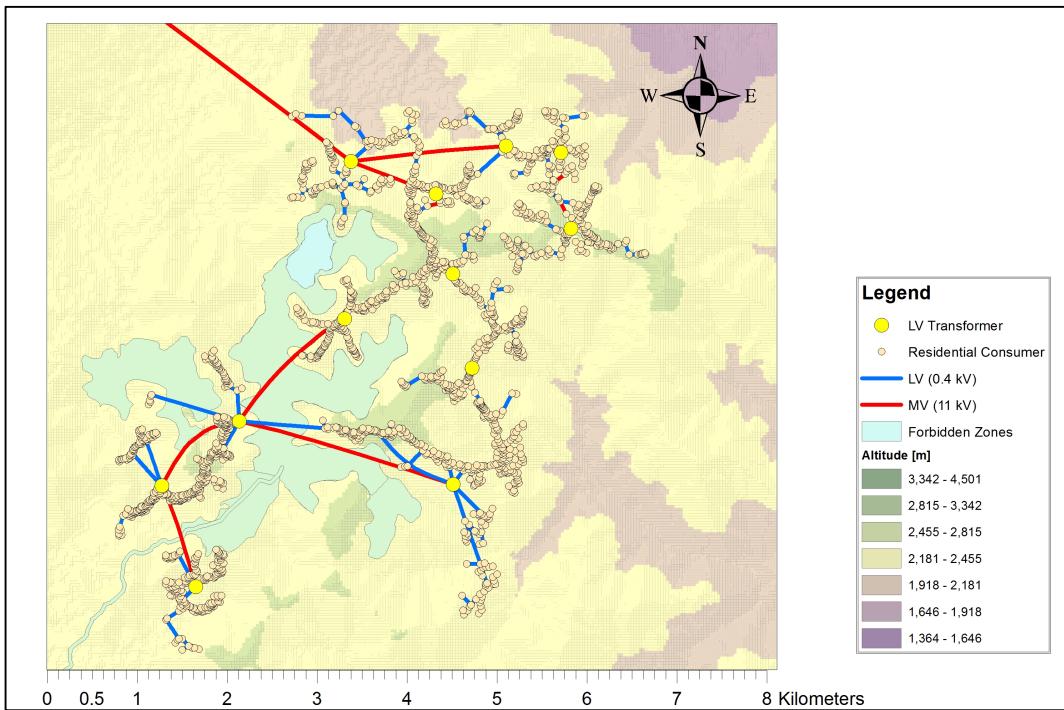
Figure 5: RNM System Design Produced without Altitude or Penalty Zone Information

Figure 6 shows a transformation of the network design due to inclusion of the altitude input file, with conductors routed so as to avoid, wherever possible, changes in elevation which are considered as typically incurring some level of additional construction and or maintenance cost.

---

[12] The legend accompanying all figures from Figure 5 through Figure 8 denote penalized zones as 'Forbidden Zones'. This was a moniker used at the time these figures were created but which is not accurate in all cases so has since been replaced with 'Penalized Zones'.
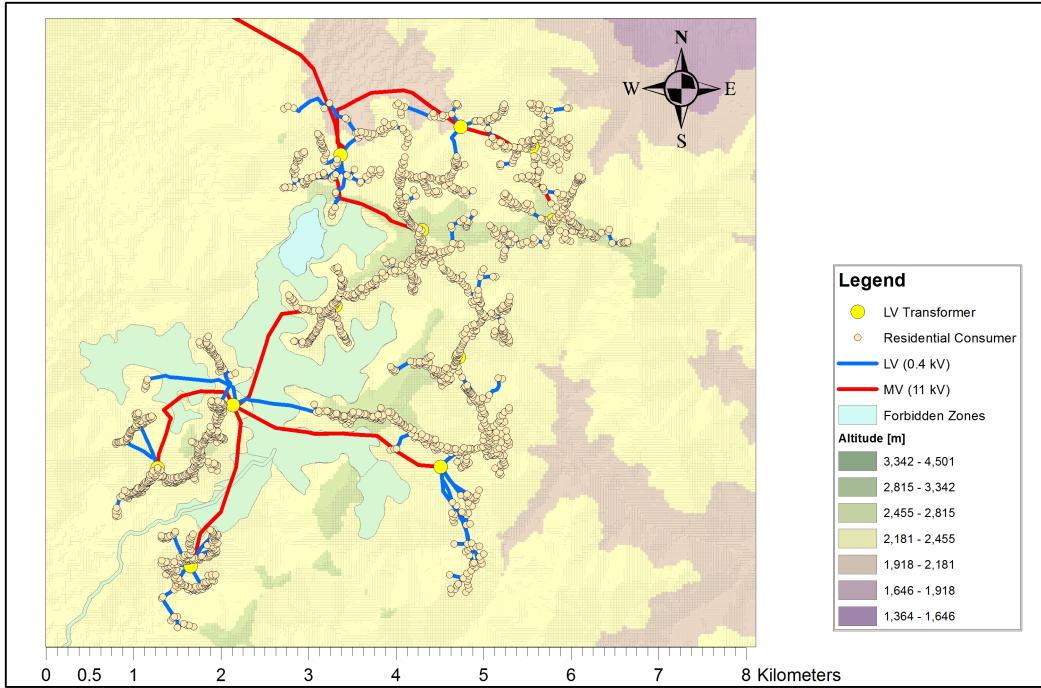
Figure 6: RNM System Design Produced with Altitude Information Provided

The network design shown in Figure 7 is the result of making both the altitude and penalized zone input files available to RNM. It is readily apparent that conductors have now been routed in such a manner as to avoid traversing penalized zones, insofar as it is economically advantageous to do so. In this instance, a penalty value of ten was used.
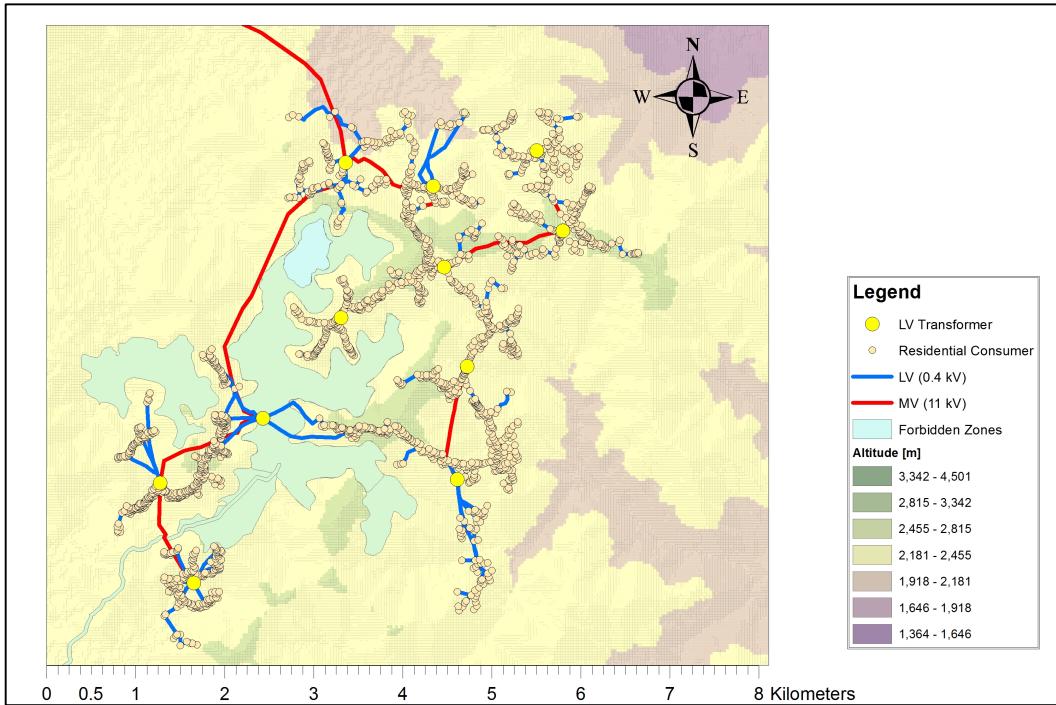
Figure 7: RNM System Design Produced with Altitude and Zone Information – Zone Penalty of 10

The design results displayed in Figure 8 were created based on a penalty value of 100. As mentioned previously, this penalty value of 100 would have been used in this instance as an increased deterrent to constructing lines which cross penalized zones, but the penalty value reflected in the reported cost of the lines which still cross penalized zones will be ten, not 100. A comparison of this figure with the last shows that the increased penalty value resulted in the removal of one of the two low voltage lines, which in Figure 7 crossed the central penalized area, and the institution of a new medium voltage line connecting the communities in the lower portion of the figure.
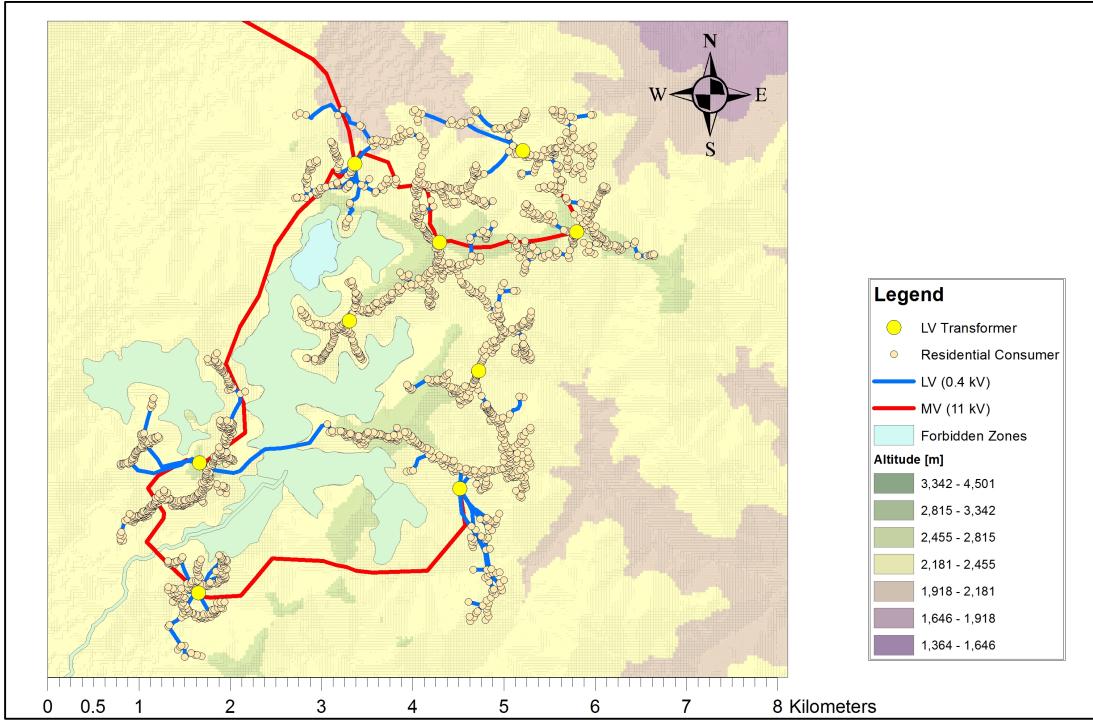
Figure 8: RNM System Design Produced with Altitude and Zone Information – Zone Penalty of 100

3.3.1.3   Network Cost Adjustment in RNM

Details of the network cost calculations used by RNM can be found in Appendix II, the contents of which have been synthesized from [29]. For the purposes of the present discussion, we will refer to elements of Appendix II as required to illustrate the topography-related network cost adjustment process employed by RNM. All equation numbers beginning with the letter 'A' correspond to those used in the appendix.

RNM considers three penalty types when adjusting network costs to reflect topographical elements: altitude, slope, and zone. Each network line evaluated is divided into segments based upon the cell size of the altitude raster input file. Penalties, which are assessed on a per-segment basis, for both altitude and slope are captured in the term *dKinv(i,j)*. If the gradient of the line segment is greater than 0.57[13], then the term *dKinv(i,j)* takes the value 20. Otherwise, *dKinv(i,j)* is calculated per equation (A15).

$$dKinv(i,j) = (1 + (dKinvMax - 1) \cdot NG(i,j)) \cdot dC(i,j) \qquad \text{(A15)}$$

---

[13] Where 0.57 is equal to the inverse tangent of 37°

28

The variable *dKinvMax* takes the value 3, which is a hard-coded value for which no definition or justification has been provided in the RNM documentation[14]. The term *NG(i,j)* represents the gradient of the segment and *dC(i,j)* is the altitude penalty which take the value 1 for altitudes 0-500 meters, 1.02 for 500-1000 meters, and 1.05 for altitudes greater than 1000 meters[15]. Equation (A15) thus effectively reduces to the following:

$$dKinv(i,j) = (1 + 2 \cdot segment\ slope) \cdot (1 + altitude\ penalty\ percentage) \qquad (1)$$

The user-provided zone penalties are referred to in [29] as modified prohibited area factors[16], or *MPAFs*. The *MPAF* value for a given segment is used, along with the term *dKinv(i,j)*, in equation (A14) to calculate the length multiplier *CM(i,j)*:

$$CM(i,j) = (dRm \cdot dKinv(i,j) + 1 - dRm) + MPAF \qquad (A14)$$

As the term *dRm* is currently hard-coded within RNM to equal 0.5[17], equations (1) and (A14) combine to become:

$$CM(i,j) = \left(0.5 \cdot \left((1 + 2 \cdot segment\ slope) \cdot (1 + altitude\ penalty\ percentage)\right) + 0.5\right) \qquad (2)$$
$$+ MPAF$$

The term *CM(i,j)* is a per-unit value which is multiplied by the corresponding segment of the line being evaluated. Based on equation (2), this penalty multiplier can be considered as being composed of two main terms; the *MPAF* term represents the portion of the penalty term related to zones, and the remainder of the equation is the portion of the penalty related to slope and altitude collectively. In order to develop intuition for the behavior of equation (2), the following table of sample values has been developed.

---

[14] Unless a justification can be provided as to why a value of one is subtracted from the variable *dKinvMax*, as shown in equation (A15), it is the recommendation of the author of this thesis that the "minus one" value be removed altogether and the variable *dKinvMax* be configured to take a user-specified value. The variable *dKinvMax* could then be justified in the REM user manual as a linear multiplier for the slope, to be used as the slope penalty. One step further would be to reconsider the linear nature of the penalty altogether. See Appendix III for a discussion regarding the use of slope in RNM.

[15] These altitude values are currently hard-coded but are not appropriate for non-coastal countries such as Rwanda, whose lowest altitude is 925m. These values should be user provided and not hard-coded in future applications.

[16] This is a misnomer as the areas are not necessarily prohibited, only penalized (a steep enough penalty may serve to functionally prohibit and area but no areas are strictly prohibited).

[17] According to Dr. Carlos Mateo at Comillas University, the justification for this value is that 50% of the length-dependent portion of the line cost is considered to be affected by topography, so only 50% of the cost is to be penalized by topography penalties.

Table 6: Sample Values for Equations (1) and (2)

Altitude: 0-500m				Altitude: 500-1000m				Altitude: >1000			
Slope	dKinv	CM (MPAF = 0)	CM (MPAF = 10)	Slope	dKinv	CM (MPAF = 0)	CM (MPAF = 10)	Slope	dKinv	CM (MPAF = 0)	CM (MPAF = 10)
0	1	1	11	0	1.02	1.01	1.01	0	1.05	1.025	11.025
0.1	1.2	1.1	11.1	0.1	1.224	1.112	1.112	0.1	1.26	1.13	11.13
0.2	1.4	1.2	11.2	0.2	1.428	1.214	1.214	0.2	1.47	1.235	11.235
0.3	1.6	1.3	11.3	0.3	1.632	1.316	1.316	0.3	1.68	1.34	11.34
0.4	1.8	1.4	11.4	0.4	1.836	1.418	1.418	0.4	1.89	1.445	11.445
0.5	2	1.5	11.5	0.5	2.04	1.52	1.52	0.5	2.1	1.55	11.55
> 0.57	20	10.5	20.5	> .57	20	10.5	10.5	> .57	20	10.5	20.5

If we consider that the variable *dKinv* represents a single percentage penalty value for both slope and altitude, it can be concluded from Table 6 that equation (2) serves only to divide the percentage value in half before adding it to the zone penalty value, *MPAF*. If we take for instance a slope value of 0.1 for the first altitude range, the penalty value represented by *dKinv* is 20%, and the penalty value represented by *CM*, when the *MPAF* value is set to zero, is 10%. This holds true in the remainder of the cases. We can further conclude that if the default *MPAF* value of ten is used, the slope/altitude portion of the resulting penalty multiplier, *CM*, is of relatively little significance. The results of this appraisal will be used in the following sections as the basis for developing a similar penalty application approach for use in the Clustering module.

Once the term *CM(i,j)* has been calculated for each segment of a given line, the resulting values for each segment are summed to form what is termed in [29] the 'Height Cost Factor'[18] (HCF), which is in units of kilometers and is used in equations (A3), (A5), and (A6) in the calculation of the net present values (NPV) of line investment, line preventative maintenance, and line corrective maintenance costs, respectively. These equations are provided below for reference. We shall refer to HCF_{line} as the effective length of a line.

$$NPV_{line,investment} = IC \cdot HCF_{line} + NPV_{equipment} + NPV_{DFP} + NPV_{CVR} \qquad (A3)$$

$$NPV_{line,maintenance,prev} = \left(\sum_{t=0.5}^{t \leq YUL} \frac{1}{dTd^t} \right) \cdot C_{line,maintenance,prev} \cdot HCF_{line} \qquad (A5)$$

$$NPV_{line,maintenance,corr} = \left(\sum_{t=0.5}^{t \leq YUL} \frac{1}{dTd^t} \right) \cdot C_{line,maint,corr} \cdot FR_{average} \cdot HCF_{line} \qquad (A6)$$

For the purposes of this report, it is important to note how the value of HCF_{line} is used in these three equations. In equation (A3), which determines the NPV of line investment, this effective line length term is multiplied only by the variable *IC*, which represents the investment cost of a line in units of $/km. It does not affect the remaining terms in equation (A3), which include, according to [29], equipment for "improving line reliability, ditches, facades, and posts, capacitors, and voltage regulators". Transformers

---

[18] This is also a misnomer as it has not only to do with the height but also the slope.

also presumably figure among the equipment considered. In equations (A5) and (A6), the effective line length is multiplied by all costs associated with line maintenance. It should be noted that the total NPV of a line is calculated in RNM according to equation (A1), or the sum of the NPV values of line maintenance, investment, and losses. The maintenance and investment terms are governed by equations (A3), (A5), and (A6), as discussed above. The NPV of line losses is governed by equation (A2) which does not include the HCF_{line} term. Equations (A1) and (A2) are shown below for reference.

$$NPV_{line} = NPV_{line,losses} + NPV_{line,investment} + NPV_{line,maintenance} \qquad (A1)$$

$$NPV_{line,losses} = \left(\sum_{t=0.5}^{t \leq YLG} \frac{dC^{2(t-1)}}{dTd^t} + \sum_{t=YLG+1}^{t \leq YUL} \frac{KFD^2}{dTd^t} \right) \cdot C_{PL,0} \qquad (A2)$$

3.3.1.4 *Alternate Path Algorithm in RNM*

In the event that at least one of the two input files discussed in the previous section is provided, the Reference Network Model will make use of a route selection algorithm during the process of designing distribution networks[19]. This route selection, or alternate path, algorithm works by first identifying the midpoint of the straight line joining two consumers, then successively moving that midpoint along a vector perpendicular to that of the original line. Each time the midpoint is moved, the cost of the new path, defined by the endpoints and the new midpoint location, is determined. The number of times this operation is repeated is a function of the modulus of the original line. We will refer to this as the 'midpoint analysis'. Of the alternate paths considered, the path of least cost is selected. If this path is any other than the original straight-line route, then this path becomes the new 'base route' and is considered for further refinement. The segment of this path which contributes the greatest amount to the total path cost is now subjected to the midpoint analysis. An example of this process is illustrated in Figure 9, with a sample 'summary' of the set of paths considered shown in Figure 10[20].
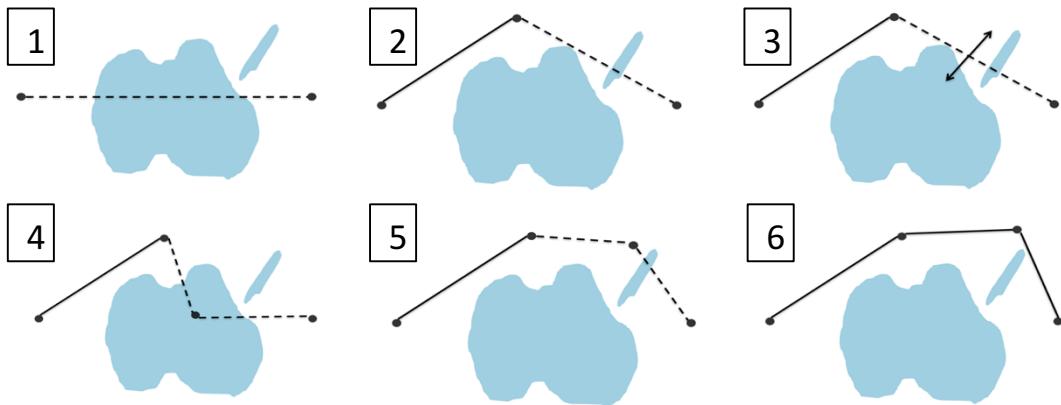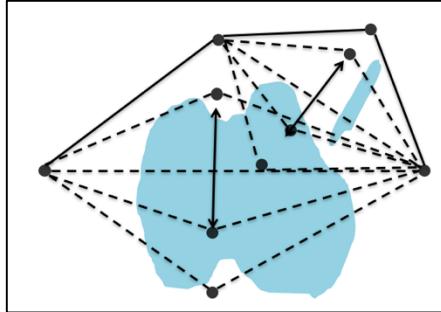

Figure 9: Illustration of the RNM Alternate Path Algorithm Midpoint Analysis

---

[19] RNM may also use this in the case of settlements, a topic broached in Chapter 4.
[20] Images shown in Figure 9 and Figure 10 were provided by Dr. Carlos Mateo of Comillas University.

Figure 10: Example of Alternate Path Sets Considered

If no alternative path of lesser cost is identified for that segment, then the alternate path algorithm stops. If one is identified, then the most expensive of the three segments of the new base route will be subjected to this process anew. This methodology repeats until either no alternatives of lesser cost are identified, or the max number of user-specified path vertices has been reached, whichever comes first. It should be noted that in addition to the set of paths identified via the midpoint analysis, two other sets of paths are also considered at each stage of this process: a 20/80 point analysis, and an 80/20 point analysis. In these instances, the point at which the straight line is segmented is first moved to a location which represents 20% of the line modulus with respect to consumer A, then the same but with respect to consumer B.

3.3.2 Rendering the Clustering Process Geospatially Aware

Before considering the methodological approach to incorporating topography-related cost-driver information into the clustering decision-making process, it is important to note that the clustering algorithms themselves remain topics of active research. It is thus prudent to begin the discussion by first delineating the geospatial subproblem from the clustering parent problem.

The intention of the clustering process in REM is to identify subgroups within a given population and assign these groups generation sources in such a way that the grouping-generation configuration selected minimizes the cost of electrifying the entire planning region. The methodology by which cluster pairs are selected differs according to clustering algorithm version, but for the two versions currently available (i.e. simultaneous and sequential), the decision-making process that follows once a cluster pair is identified is fundamentally the same. When considering two clusters, the decision to be made is whether or not combining these two clusters to form a single cluster is economically advantageous. The combination of these clusters is considered beneficial to the system if the generation cost savings would exceed the increase in network cost. In the simultaneous version of clustering, the increase in network cost is approximated by the cost of a single conductor spanning from the weighted demand center of one cluster to that of the other cluster. This is referred to as the center-to-center distance. In the sequential version of clustering, the increase in network cost is considered to be the shortest of either the center-to-center distance or, the shortest customer-to-customer distance multiplied by a scalar which currently takes the value of fifty. The shortest customer-to-customer distance is defined as the shortest distance between a consumer belonging to cluster A and a consumer belonging to cluster B.

For the purposes of topography cost driver integration, the artifacts of the current clustering methodologies as described above are taken for granted. The 'topography integration' process concerns itself only with modification of the way the costs of the lines are calculated and the route that each line follows, with the endpoints of all lines taken as defined by the clustering process. One exception to this involves a manipulation of the manner in which customer-to-customer distance is used by the sequential clustering algorithm, a modification necessary for accommodation of the topography integration process. The details of this modification will be addressed in the ensuing discussion.

32

In section 3.3.1, dissection of the methodology by which geospatial attributes are considered within RNM revealed two distinct mechanisms: the adjustment of line costs and the selection of least-cost network routes. In similar fashion, the integration of topography into the clustering process will be conducted with respect to these two separate mechanisms, with the topic of network cost adjustment covered first.

Definition of Cost Drivers
The set of topography-related cost drivers to be considered for incorporation into the clustering decision-making process has been defined to be consistent with those which are used in RNM and which also appear to constitute those used by the majority of the sources consulted in the literature review. These cost drivers include slope of terrain, altitude, and geographic zones as described below:

- **Slope of terrain:** The slope of terrain on which network infrastructure is installed is positively correlated with the cost of network installation and maintenance; that is, when one increases, so does the other. This is due to the increased labor and equipment costs associated with accessing, and installing infrastructure on, steep terrain.
- **Altitude:** The altitude at which network infrastructure is installed is positively correlated with the cost of the network installation and maintenance. This is due to the increased labor and equipment costs associated with accessing, and installing infrastructure at, high altitudes. The cost of the infrastructure installed may also increase at higher altitudes (e.g. more robust poles may be required to withstand weather conditions which are harsher at higher altitudes than at lower altitudes).
- **Geographic Zones:** The nature of the geographic zone traversed by a network may affect the installation and maintenance costs incurred. Factors which may impact the costs associated with installation and/or maintenance might include, but are not limited to, additional permitting fees, increased equipment and labor costs (e.g. building in rocky or marshy areas may require reinforced infrastructure; building through a forest may require cost of equipment and labor to fell trees), and increased construction time (increased cost of construction loan, etc.).

Methodology - Use of Cost Driver Information in Clustering
As the network costs calculated in clustering are approximations of those which RNM would calculate were it called for this purpose, it was determined that incorporating this cost driver information into the clustering modules in such a manner that it is used as it would be by RNM would be a logical place to begin. The methodology used by RNM was not followed to the letter in the clustering topography integration process but rather taken as a guide. Following is a discussion of the modifications made to the clustering algorithms, starting with a description of the input files required to accommodate this process.

Input File Creation
As is the case with RNM, two input files are required in order to support topography-related functionality in the clustering module: one containing altitude data, the other penalized zone information. The altitude raster described in section 3.3.1 can be used as-is for the clustering module, so there is no need for a separate altitude input file. This is not the case for the penalized zone input file, which must be supplied in ESRI raster format, with the same header information as used for the altitude raster. Notes on creating the zone raster such that it is consistent with the zone raster which is created internally by RNM from the zone polygon input file can be found in Appendix I. Both raster files are read by Matlab using the *arcgridread* function, which saves the primary data for each to dedicated matrices and stores the metadata in structure format. The two matrices and all pertinent metadata is then saved to a single structure for ease of use. This structure is called *rr* and is stored within the primary structure *localData*, descriptions for the contents of which can be found in [21].

<u>Application of Penalties</u>
The altitude matrix is used to create a slope matrix, which has the same dimensions as the altitude matrix and where the value contained in each cell is the worst case (i.e. steepest) slope given the center cell altitudes of the eight neighboring cells.[21] Each network line to be evaluated is then broken into segments based on its intersection points with a grid, which is itself based on the cell size and dimensions of the input rasters. The z-coordinate of each segment endpoint is then determined according to its location on the grid and the altitude values in the adjacent cells[22]. The altitude-adjusted, or '3D', segment lengths are then calculated and the sum of the segment moduli saved as the true line modulus.

Slope penalties[23] are assessed to each segment in the form of a length extension, or 'line length penalty'. Line length penalties are determined for each segment based on the slope of that segment, where the actual slope value translates directly as the percentage of total length added on as additional length. As in RNM, if the slope exceeds approximately 37°, then a multiplier with the value of 20 is used. It is worth noting that since length-dependent costs are applied on a per-kilometer basis, a percentage increase in length is tantamount to a percentage increase in cost (i.e. The calculation of the line-length-dependent portion of the cost is equal to the total length multiplied by the cost-per-kilometer. Therefore, assessing a percentage increase to the length is mathematically equivalent to assessing a percentage increase to the cost, as multiplication is commutative.) Zone penalties are applied in the same manner as slope penalties, with the zone penalty located in the raster cell corresponding any given segment being the value that is multiplied by that segment, with the resulting value used directly as a line length penalty. Once length penalties have been calculated for all segments of a given line, the penalized lengths are summed and saved as the line penalty length. All segment information, including endpoint coordinates, true moduli, and length penalties, are saved in a single dedicated structure, entitled *routes*, which is then used by the function responsible for the calculation of line cost.

In the clustering module, line cost is calculated for a given conductor length and power flow for both a medium voltage and a low voltage conductor, each selected from the RNM catalog such that the thermal limit of the conductor is respected and the voltage drop incurred would not exceed 10%. Costs associated with the expected losses of each line are calculated and added to the cost of the conductor. The cost of two MV/LV transformers may be added to the cost of the MV line, depending upon the particular application of the line cost function. The lesser of the two total costs is then output as the cost of the line. For the incorporation of line length penalties into this cost calculation process, it is important first to recognize that line length is used for three purposes in the line cost function. It is used first to calculate the maximum per-kilometer impedance the line can have with respect to the voltage drop limit. It is also multiplied by the per-kilometer conductor and losses costs in order to determine the total cost associated with each conductor option. For determining both the maximum impedance value and the total cost associated with losses, the true line modulus must be used. If an altitude raster was provided, the 3D length is selected for this purpose, otherwise the 2D length is used. For calculation of the total conductor cost, independent of costs associated with losses or transformers, the length used is the true line modulus plus the sum of the zone length penalties. In this way, penalties are assessed as a function of the base cost of the conductor, which includes both corrective and preventative maintenance costs, but does not factor into the costs associated with line losses or transformers. This is consistent with the application of penalties in RNM, as described in section 3.3.1.

---

[21] This is also the methodology used for determination of slope in ArcGIS, so a developer could choose instead to have the user provide a slope raster as a third input file, were that found to be advantageous for any reason.

[22] This value is currently taken as the average of the altitude values in the adjacent cells. In a future version, this should either be adjusted to use multilinear interpolation or segment endpoints should be defined as coincident with gridlines passing through the center of each raster cell instead of along the edge of each cell (in which case the z-coordinate of any given segment endpoint would simply be the altitude value found in the host cell).

[23] See Appendix III for further discussion regarding the definition of slope penalties in RNM and Clustering.

<u>Location of Penalty Application</u>
The stages of clustering relevant to the discussion at hand include Delaunay line generation, off-grid clustering, and grid extension clustering. In the initial stages of clustering, the Delaunay lines are sorted by length from shortest to longest and this line order is used to prioritize cluster pairs for evaluation. In the case of the sequential clustering algorithm, each time two clusters are joined, the cluster with the lower index number 'absorbs' the cluster with the higher index number. When this happens, the low-number cluster inherits all Delaunay lines which linked the high-number cluster to other candidate clusters. In instances where both the high and low-number clusters had links to the same candidate cluster, the low-number cluster now has two Delaunay lines linking it to that candidate cluster. All of the instances where this occurs are reviewed, the shortest Delaunay line is saved as the shortest customer-to-customer distance, and the set of cluster pairs is resorted based upon the updated Delaunay line values. The simultaneous clustering algorithm conducts this process in a slightly different manner, with multiple cluster pairs being evaluated in parallel, but the fundamental use of the Delaunay lines is the same. In this capacity, only the Delaunay line length is utilized, since there is no power demand associated with these lines and thus no cost can be calculated. Assuming, however, that sorting the cluster pairs based on the length of these lines has merit[24] in the clustering decision-making process, their lengths are first adjusted by addition of the penalty lengths. They are therefore sorted based not on their true moduli, but on their effective lengths (i.e. moduli plus penalized lengths).

In off-grid clustering, as mentioned previously, the center-to-center distance of a cluster pair is used as the length of a conductor, the cost of which serves as a proxy for the network upgrade cost that would be incurred if the two clusters were joined. This length and cost are modified using the topography penalty methods described above. In the case of the sequential clustering algorithm, this center-to-center length is compared to a length equal to fifty times the shortest customer-to-customer distance for that cluster pair, and the line cost is then based on shortest of the two lengths. This mechanism was instituted when it was realized that in instances where the weighted generation centers of two large clusters are a considerable distance apart, and the settlement between them relatively sparse, the center-to-center distance lead to what was most likely an overestimation of network costs. However, this methodology posed an issue for calculation of line length penalties because the quantity '50 x customer-to-customer distance' has no physical representation. After consulting with the development team, it was decided that instead of basing cost on the minimum length, costs would first be calculated for both the customer-to-customer distance and center-to-center distances, then the *cost* of the customer-to-customer distance would be multiplied by fifty and the minimum of the two costs taken to be the final cost of the line. The results of this change will be discussed further in section 3.4.

In the grid extension clustering section, seven scenarios are evaluated to determine whether grid extension or an off-grid solution is economically advantageous for any given cluster. These scenarios are illustrated in Figure 11, with the attributes of each conductor listed in Table 7. As all seven scenarios are considered for each cluster pair evaluated, costs must be calculated for a total of ten conductors for every iteration, including four grid extension lines and six center-to-center lines. Center-to-center distances are defined as in the off-grid clustering section; grid lines extend from the cluster weighted demand center to a tapoff point of an existing network. For each cluster pair, costs for all ten conductors are calculated based on the penalty application methodology described above.

---

[24] It is not a foregone conclusion that sorting the clustering pairs by Delaunay line length has an effect on the outcome of the clustering process. That this process has an effect on the results of the clustering process, and that it is in some respect an advantageous one, has not been verified.
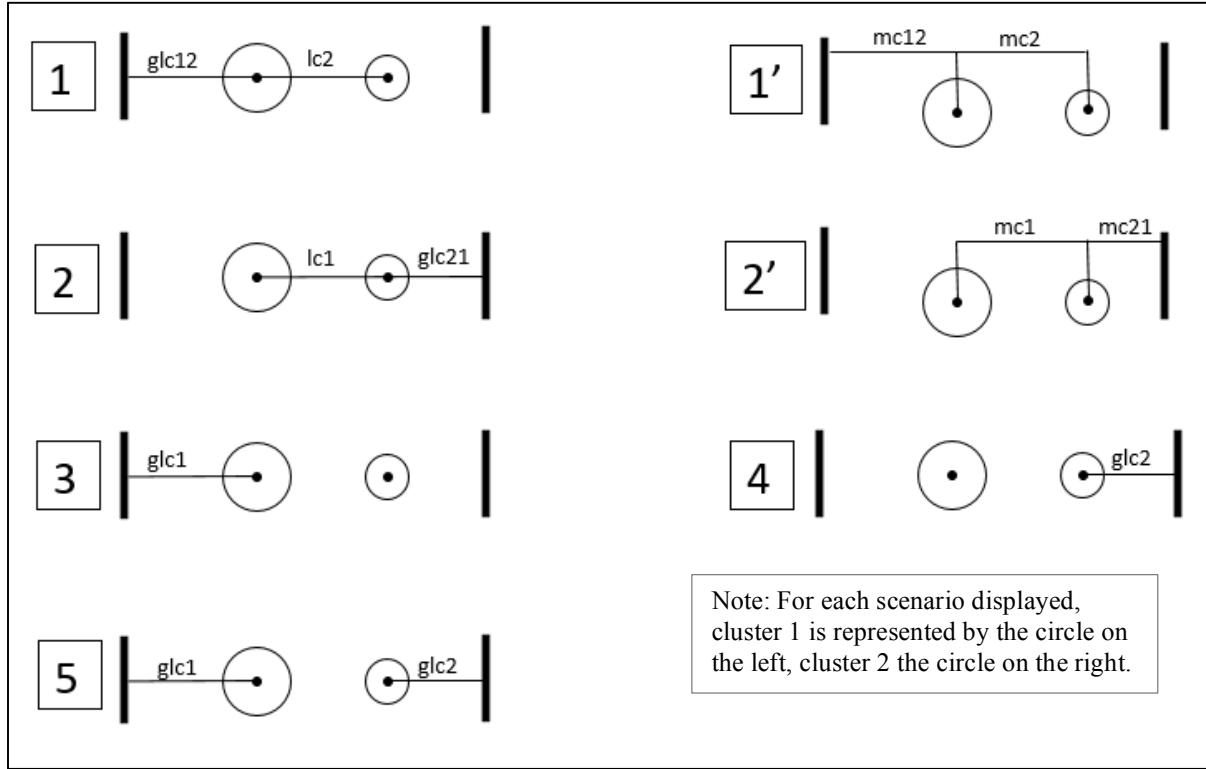
Figure 11: Grid Extension Scenarios

Table 7: Grid Extension Scenario Conductor Summary (ref. Figure 11)

Conductor Name	Voltage Designation	Load	Gridline Tapoff Location
lc1	LV	Cluster 1	NA
lc2	LV	Cluster 2	NA
mc1	MV	Cluster 1	NA
mc2	MV	Cluster 2	NA
mc12	MV	Clusters 1 and 2	NA
mc21	MV	Clusters 1 and 2	NA
glc1	LV or MV	Cluster 1	Location 1
glc2	LV or MV	Cluster 2	Location 2
glc12	LV or MV	Clusters 1 and 2	Location 1
glc21	LV or MV	Clusters 1 and 2	Location 2

A summary of the locations where length and cost adjustments occur to reflect topography-related cost drivers in the clustering module is provided for reference in Table 8.

Table 8: Network Cost Adjustments by Location

Clustering Stage	Network Object	Attribute Affected
Off Grid	Delaunay lines	Length
	Center-to-center* Line	Length and Cost
	Customer-to-customer** Line	Length and Cost
Grid Extension	Grid extension line	Length and Cost
	Center-to-center* Line	Length and Cost

* Length of line spanning from the weighted demand center of one cluster to that of another
** Length of shortest Delaunay line linking a customer in cluster A of a pair to a customer in cluster B of that pair

3.3.2.2   Alternate Path Algorithm in REM

As there is no written documentation of the alternate path algorithm used in RNM, and the information recounted in section 3.3.1 regarding this algorithm was only recently conveyed to the MIT team, development of a similar algorithm for use in the clustering module is still underway. Therefore, no results have yet been obtained using this version of the alternate path algorithm. However, in the interim, a continuous optimization version was developed using Matlab's nonlinear programming solver *fmincon*. Provided with the endpoints of a line, four initial waypoints, evenly spaced along the line, and an objective function which performs the cost adjustment described earlier in this section, the solver searches for the least cost path by varying the y-coordinates of the waypoints. The x-coordinates remain fixed in order to limit the search space. Additional measures taken to limit the search space include the imposition of upper and lower bounds[25] and, in the case of off-grid clustering, the use of an objective limit and threshold value.

The objective limit and the threshold value are numerically equal but serve two separate purposes. The value, represented by the variable *maxLineCost*, is defined as the maximum cost which may be incurred by the additional network requirements of joining two clusters in order for the decision to join those clusters to be economically advantageous. As the network 'upgrade' cost approximations calculated in the off-grid clustering algorithm are used only to inform the decision of whether or not two clusters should be joined, it is not necessary to find the lowest possible value of this network 'upgrade'. It suffices to determine whether or not a value can be obtained for the network upgrade cost which is below the decision threshold value, represented by *maxLineCost*. This value is therefore used first to determine if it is necessary to use the alternate path algorithm (if the line cost is already less than the threshold value, then there is no need to evaluate it further), and then as an objective limit for the solver (i.e. execution is halted once any value below the objective limit is found).
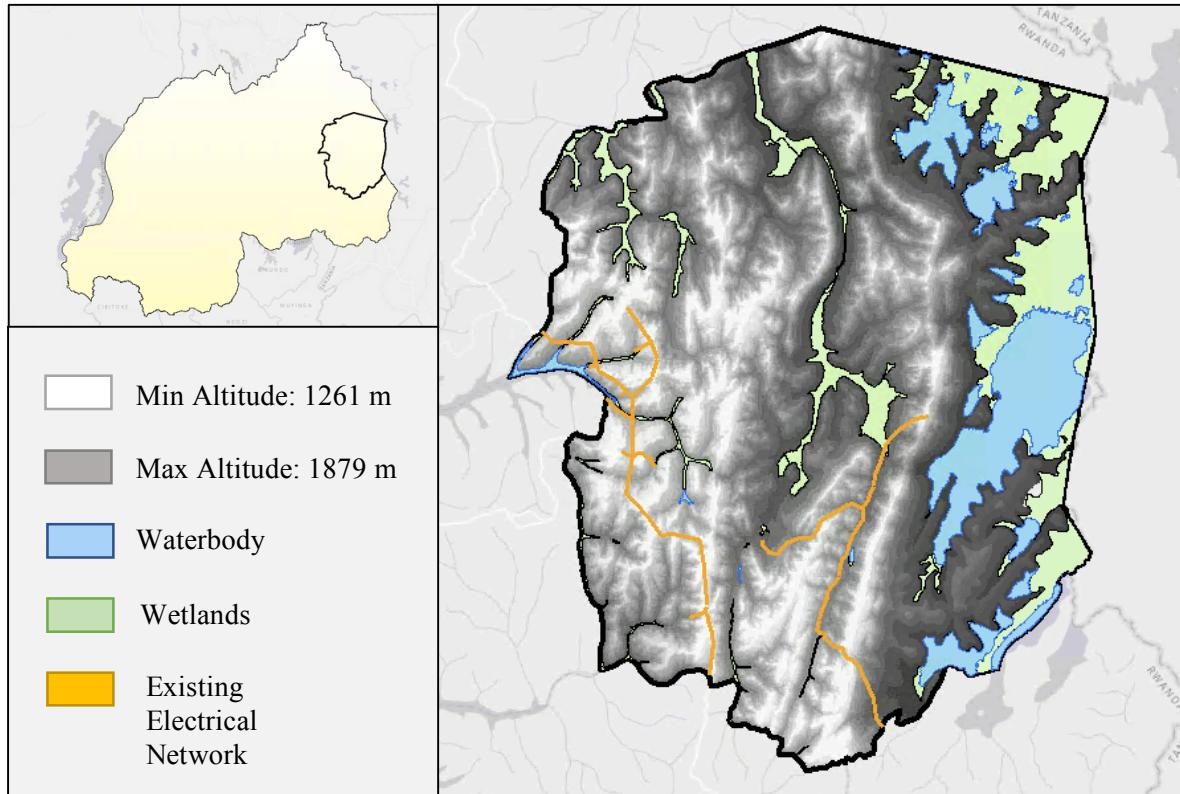
All of these mechanisms are used in an effort to limit the computational time required by the alternate path algorithm, while still exercising the presumed benefits of exploring alternative network routes. Unfortunately, due to the nature of the grid extension clustering algorithm, it was not possible to implement a similar 'maxLineCost' mechanism. System cost and configuration results obtained using this algorithm, and those cost adjustment algorithms described herein, are discussed in the following section.

3.4   Results and Discussion

The algorithms described in section 3.3.2 were tested by performing REM software runs using input data from the Kayonza district of Rwanda, a region with approximately 40,000 consumers. An image of the test region is shown in Figure 12. To the right of the figure is an image of Kayonza including both the existing electrical network and the features for which penalty values were applied. A legend for this map

---

[25] The upper and lower limits are currently equal to the maximum(minimum) endpoint y-coordinate value plus(minus) a distance equal to fifty percent of the difference in the endpoint y-coordinates.

can be found in the lower left of Figure 12, and in the upper left, an image of the country of Rwanda with the district border of Kayonza indicated in bold. For reference, the district measures approximately 45 kilometers east-west and 58 north-south.


Figure 12: Kayonza District Including Penalized Features and the Existing Electrical Network

The results of these test runs, which are shown in Table 10, include those obtained using both the sequential and simultaneous clustering algorithms. Three scenarios were run for each version of the clustering algorithm, including a base run, a straight-line cost correction run, and an alternate-path cost correction run. The base runs included all geospatial cost correction capabilities of RNM, but none in clustering. The straight-line runs included the addition of the cost correction algorithms in clustering, in addition to the RNM functionalities, and the alternative path runs implemented the *fmincon* alternative path algorithm in addition to this. In the case of sequential clustering, these three runs were performed using the line multiplier adjustment made in off-grid clustering to accommodate the new algorithms. In light of this, a fourth run was included which used the original mechanism of the scalar fifty multiplied by the shortest customer-to-customer distance, so that the effect of the modification to this mechanism could be isolated and reviewed. A brief summary of select terminology in the form of Table 9 has been provided to remind the reader of the significance of the expressions used herein for software run scenario identification.

Table 9: Summary of Select Terminology

Term	Definition
sequential clustering algorithm	The consumer clustering algorithm used in the Clustering module whereby cluster pairs are evaluated one at a time to in order determine whether or not it is more economical to electrify them separately or together.
simultaneous clustering algorithm	The consumer clustering algorithm used in the Clustering module whereby multiple cluster pairs are evaluated in parallel in order to determine whether or not it is more economical to electrify them separately or together.
straight-line path	The euclidean distance between any two endpoints selected to denote the extremes of a conductor used to approximate a newtwork cost, whether it be from the weighted demand center of one cluster to that or another, or from the weighted demand center of one cluster to a grid tapoff location.
alternate path	The distance between any two endpoints (see straight-line path definition above) measured along a path of least cost resistance as identified by the alternative path algorithm described in section 3.3.2 of this thesis.

Table 10 lists first a name identifying the nature of the scenario, followed by the total elapsed runtime, the number of consumers assigned to each of the three generation categories (MG – microgrid, IS – isolated system, GE – grid extension), the number of clusters appearing in the final design (GE – grid extension cluster, OG – off-grid cluster), and finally the net present value of the total investment required to support the regional design. The only change occurring between the first and second run entries in the table was the modification of the off-grid clustering length selection mechanism, which was altered to function as a cost selection mechanism instead.

Table 10: REM Network Design Results with Geospatially-aware Algorithms in Clustering and RNM

Scenario Name	Elapsed Time	MG	IS	GE	GE clusters	OG clusters	NPV
Clustering Algorithm Version: Sequential							
Base Case: 50*Length	224 Min. (3.7 hr)	20,939	112	19,759	89	308	$34,613,121
Base Case: 50*Cost	237 Min. (4 hr)	23,553	114	17,143	90	319	$35,435,367
Straight Line Cost Correction	249 Min. (4.2 hr)	26,767	180	13,863	100	346	$36,522,964
Alternate Path Cost Correction	2266 Min. (37.8 hr)	17,597	146	23,067	46	294	$33,499,730
Clustering Algorithm Version: Simultaneous							
Base Case	128 Min. (2.1 hr)	31,569	174	9,067	41	141	$41,357,209
Straight Line Cost Correction	136 Min. (2.3 hr)	31,028	164	9,618	41	170	$39,583,968
Alternate Path Cost Correction	1076 Min. (18.9 hr)	8,228	1,800	30,782	559	494	$36,577,445

It is the understanding of the author of this thesis that the "50*length mechanism" was originally incorporated into the sequential clustering algorithm based solely on an aesthetic appraisal of the clustering configuration and not on a cost analysis of any sort. Due to the subjective nature of this decision, it is difficult to superficially ascertain whether modifying this mechanism to be fifty times cost, instead of fifty times length, does or does not achieve the desired effect. What can be concluded, based on the numerical results shown in Table 10, is that the effects of the change on the resulting clustering configuration were relatively minimal, increasing the number of grid extension clusters from 89 to 90 and the off-grid clusters from 308 to 319. In addition to this, the modification resulted in a net present value increase of approximately $800,00, or just under 2.4%. Based on what appear to be relatively modest perturbations of the results, the scalar value of fifty was retained for used in subsequent runs. An objective methodology for exploring the value of this multiplier is discussed later in this section.

Thus we have, for the sequential version of clustering, an effective base case scenario which requires four hours to run and returns a system design consisting of 90 grid extension clusters, 319 off-grid clusters, and a total NPV of $35,435,367. The addition of straight line cost correction increased this run time to

4.2 hours, the grid extension clusters to 100, the off-grid clusters to 346, and the NPV to $36,522,964, a cost increase of approximately 3%. Although these changes would, again, appear to be relatively innocuous, the increase in NPV does run counter to the hypothesis set forth in section 3.1 that the addition of cost driver information into clustering would actually lead to a reduction in system cost. It should be noted, however, that upon the addition of the alternate path functionality, the cost is reduced from the base case NPV by approximately $1.94M dollars, which is a decrease of just under 5.3%. This would appear to support the validity of the working hypotheses and justify the implementation of these algorithms. However, it cannot be overlooked that the addition of this alternate path algorithm caused the runtime to increase nine-fold, from 4.2 hours to 37.8 hours, which most likely renders the use of this particular algorithm untenable.

With respect to the simultaneous clustering run results, the base case cost was found to be $41,357,209, which is approximately $5.9M greater than that returned by the sequential version, representing an effective increase in the base cost for the region of approximately 16.7%. While this is not an insignificant cost increase, use of the simultaneous algorithm did reduce the runtime by almost 50%, taking it from 4 hours to 2.1 hours. These differences aside, the system cost trend resulting from the use of algorithms introduced in this thesis was similar to that seen with the sequential clustering scenarios. The addition of straight line cost adjustment alone reduced the NPV in this instance by just under $1.8M, a decrease of approximately 4.3%, with a total decrease in NPV from the base case of almost $4.8M, or approximately 11.6%, following implementation of the alternate path algorithm. The runtime underwent a substantial increase as well, going from 2.1 hours in the base case, and 2.3 in the straight-line scenario, to 18.9 hours with the addition of the alternate path algorithm. The change in clustering configuration from the straight line to the alternate path scenarios, however, was entirely dissimilar from that seen with the sequential algorithm runs. In that case, the number of both the off-grid and grid-extension clusters decreased; with the simultaneous runs, however, the number of clusters appears to have actually increased, and somewhat dramatically so. The number of grid extension clusters was the same in both the base and straight-line cases, remaining at 41, but this number jumped to 559 with the addition of the alternate path algorithm. That represents an almost 14-fold increase. A similar, if somewhat less pronounced, effect was seen with the off-grid clusters, which increased from 141 in the base scenario, to 170 with straight-line cost adjustments, and finally jumping to 494 clusters.

It would seem that, at least from a cost reduction standpoint, the combination of cost penalty and alternate path mechanisms explored here have merit and are of potential benefit to the REM software. However, considering results from this singular case example, the evidence is all but conclusive, especially given the somewhat contradictory behavior seen with both the straight-line cost adjustments and clustering configuration results from one clustering algorithm to the next. I would suggest that the REM-adapted version of the RNM alternate path algorithm be finished and tried to see if similar results are obtained and if any improvement can be made with respect to the computational time required. I also think that in order to draw confident conclusions about the behavior of these algorithms, intuition needs to be built by running a suite of cases in addition to the Kayonza district because, unlike with RNM, where a visual assessment is sufficient to verify that using the geospatially-aware algorithms is advantageous, images are of little use in this case.

In addition, it occurs to me that the only subjective way to vet the clustering approximation of network costs is to perform what I will refer to as an 'RNM comparison'. For this suggested analysis, a set of N cluster pairs will have to be defined, with generation and operations and maintenances costs calculated and saved for each pair for both the individual clusters and for each cluster combination. Then network costs must be calculated using both the current clustering approximation and by calling RNM. With this set of information, a simple spreadsheet could be used to identify the clustering decision which would be made for each pair of clusters using first the network approximation, and then the true network costs. From this it can be determined what percentage of the time the network approximation leads to the same clustering decisions as the RNM network costs. Performing this both with and without geospatially-

aware algorithms in the clustering module would provide an indication as to whether or not these algorithms actually serve to improve the network cost approximation. This would not only give us important insight into the integrity of the clustering approximation, both with and without the algorithmic additions outlined in this thesis, but would also afford the opportunity to explore the effects of the hard-coded length multiplier (which currently takes the value of 50) used in the sequential off-grid clustering algorithm. It may be that the current value used is not the optimal multiplier for the intended purposes, or that a function would better serve our goals than a static scalar.

In addition, in light of the discussion regarding computational time, it would be prudent at this juncture to go through the exercise of using RNM to calculate the true additional network costs during the entire clustering process so that a concrete value can be assigned to the computational cost savings that using the network cost approximation in clustering actually affords. This value can then be compared to the level of accuracy the approximation was found to exhibit in cluster merging decisions and the accuracy-vs.-runtime tradeoff can be considered and used to inform the prioritization of future research activities.

Chapter 4

4 Electrification Master Planning

As introduced in section 1.2, the MIT group has been invited, along with Tata Power Delhi (TPDDL), to assist the current electrification planning initiative in Rwanda with the initial task of performing a review of past electrification plans, for which a collection of supporting documents has been provided. Two technical plans have been identified as the primary candidates for review, one written in 2009 by the consulting company Castalia, the other by Sofreco, dating to 2013. This chapter will be devoted to the review of these electrification plans in an effort to draw conclusions that may help inform the use and further refinement of REM. We will begin with a brief look at the history, geography, and energy profile of Rwanda, followed by detailed review and comparison of the Castalia and Sofreco master plans. We then consider the evolution of the Rwandan energy sector strategy, both during the development of these two plans and in the wake of their publication, and conclude with a detailed assessment of the role to be played by the Reference Electrification Model and a critique of its current capabilities.

4.1 Development of a Master Electrification Plan: Case Study Rwanda

The purpose of this section is review and critique the Castalia and Sofreco electrification master plans in the context of the evolving Rwandan regulatory landscape.

4.1.1 Rwandan context

Rwanda is a small landlocked country characterized by persistent rolling hills, located in eastern sub-saharan Africa, bordered by Uganda to the north, Tanzania to the east, Burundi to the south, and sharing a western border, dominated by Lake Kivu, with the Democratic Republic of the Congo. Country administrative boundaries include: 5 provinces (Northern, Western, Eastern, Southern, and Central/Kigali City), which are divided into 30 districts (outlined in red in Figure 13), and subdivided into a total of 416 sectors (outlined in grey in Figure 13).
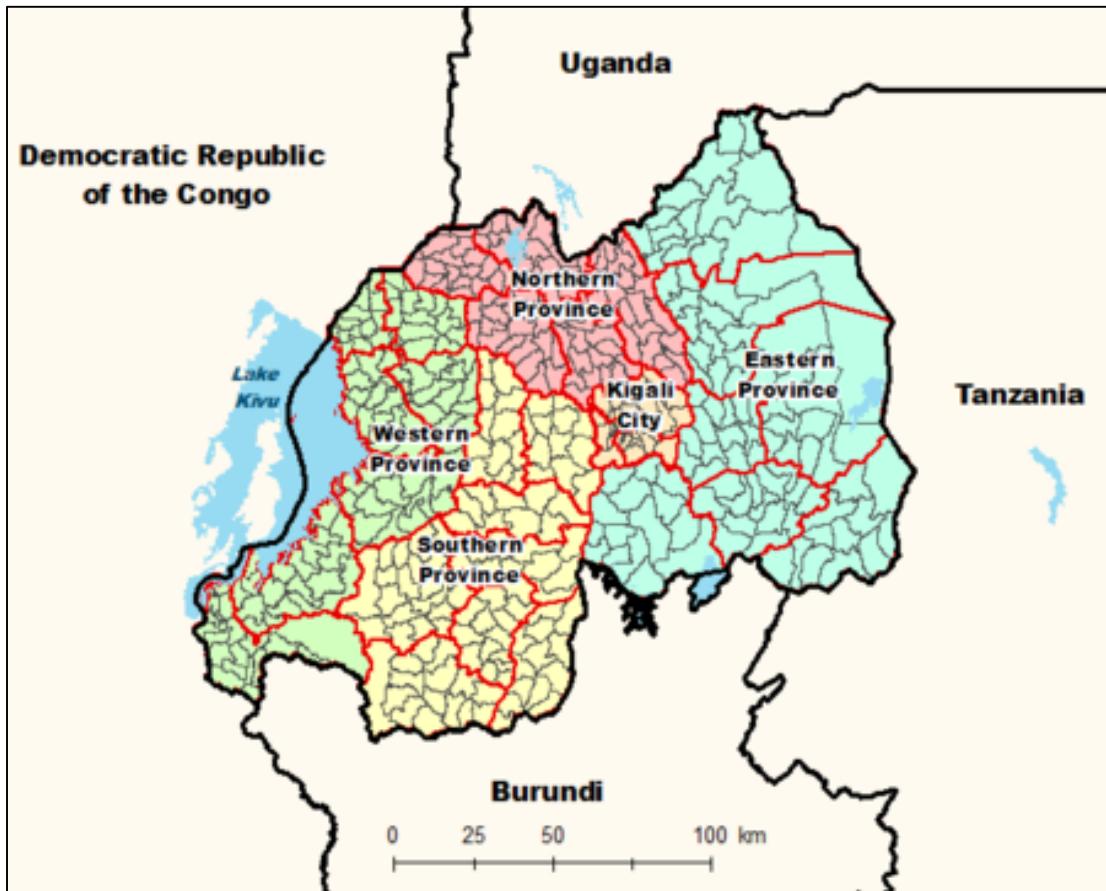
Figure 13: Political Boundaries of Rwanda (source: [30])

4.1.1.1   Current Electrification Status and Energy Profile

According to the USAID Rwanda Power Africa Fact Sheet [31], the current electrification rate in Rwanda is 31% (3% off-grid, 28% on-grid), a rate that has more than tripled since 2009 according to Figure 14, but which still leaves approximately 7 million without access to electricity.
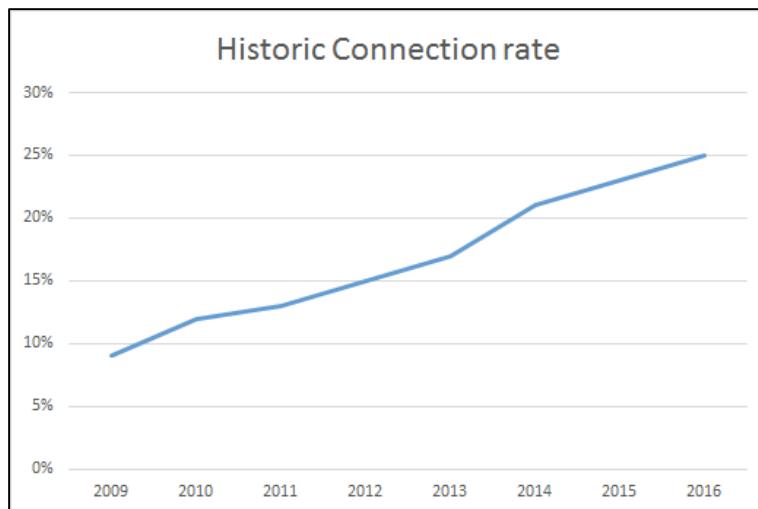

Figure 14: Rwanda Electrification Rate (grid access only) 2009-2016 (source: [32])

43

Rwanda has a total installed generation capacity of 190MW, with a generation source mix dominated by hydro power, which accounts for 44% of the overall generation mix, followed by methane from Lake Kivu peat, accounting for 28% of the country's total generation. The complete breakdown of generation capacity by source is shown in Table 11.

Table 11: Installed Generation Capacity by Source - 2016 (source: [33])

	Capacity MW	
Hydro	83.1	44%
Thermal Methane	52.8	28%
Thermal Oil	29.6	16%
Solar	8.75	5%
Imports	15.5	8%
	190 MW	

The majority of the current power supply in Rwanda comes from facilities established through public-private partnerships with the government of Rwanda [34]. And despite the low level of generation capacity by developed nation standards, the legacy of electricity provision in Rwanda is not as young as one might suspect, actually dating as far back as 1939 with the establishment of REGIDESO [35] ,which was reponsible for the provision of water, electricity and gas in Rwanda and Burundi. A Rwanda subsidiary of REGIDESO was split off in 1963 and this entity became known as Electrogaz starting in 1976. The institutional diagram of the energy sector, shown in Figure 15, shows Electrogaz as a state-owned, vertically-integrated utility. This figure was taken from the report of the master electrification plan developed in 2009 for Rwanda by the consultant Castalia.
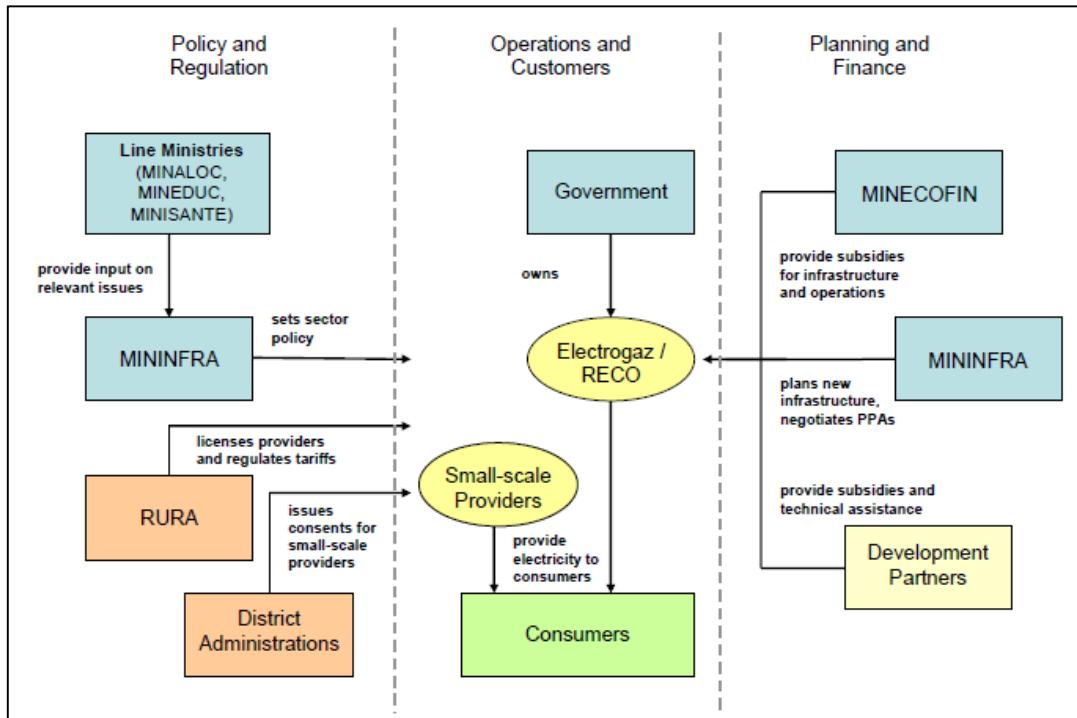

Figure 15: Rwanda Energy Sector Institutions ca. 2009 (source: [36])

- Electrogaz – Government-owned, vertically-integrated electricity and water utility responsible for generation, transmission, distribution, and retailing. Electrogaz was split in 2008 into an electric utility RECO (Rwandan Electricity Corporation) and a water utility RWASCO (Rwanda Water and Sewage Corporation), but remained under the control of the government.
- MININFRA – Ministry of Infrastructure, sets energy policy, plans sector investments for which it obtains funding from the Ministry of Economy and Finance (MINECOFIN), negotiates procurement contracts such as Power Purchase Agreements and Gas concessions.
- MINALOC – Ministry of Local Government, has authority over adminstrative offices, the electrification of which has been a priority in planning efforts, as discussed further in later sections of this report.
- MINEDUC – Ministry of Education, has a role similar to that of MINALOC but for educational institutions.
- MINASANTE – Ministry of Health, has a role similar to that of MINALOC but for health facilities.
- RURA – Rwanda Utility Regulatory Authority, independent regulator established in 2001, responsibilities include but are not limited to tariff setting, issuing development licenses, proposing legislation, approval of the national budget and oversight of consumer protection.

The general structure of the institutional diagram displayed in Figure 15 is still valid as of 2017, although the titles of a few of the entities listed have changed. The year following the publication of this figure in the Castalia planning report, the subsidiaries of Electrogaz, RECO and RWASCO, were dissolved, and the name of the utility was changed to Energy, Water, and Sanitation Authority, or EWSA. In 2014, a law was issued repealing EWSA, and in its stead, The Rwanda Energy Group Ltd. (REG) was formed [34]. REG is the reigning institution, as of 2017, and has two subsidiaries: The Energy Utility Corporation (EUCL) and The Energy Development Corporation (EDCL). Water and sanitation governance responsibilities have been seperated from energy services and reside under the authority of WASAC, the Water and Sanitation Corporation. According to the REG website [37], the primary responsibilities of EUCL include operations and maintenance of all generation facilities and transmission and distribution infrastructure, as well as electricity retail and customer service. The primary tasks of EDCL include managing investment in the development of generation facilities, developing transmission infrastructure, and planning and execution of universal energy access projects comensurate with meeting the national electrification targets.

4.1.1.2   Chronology of Electrification Planning Policies and Reports

The first electrification master plan which appears to have been developed for Rwanda was drafted by Hydro Québec International in August of 1991, with a planning horizon reaching to 2010 [38]. Implementation of this plan was interrupted by the genocide in 1994, a massive upheaval of the country's social, political, and economic systems from which it would spend the following decades trying to recover. A series of national infrastructure development legislation was issued starting with the National Urban Housing Policy in 2004 followed by the Economic Development & Poverty Reduction Strategy published in 2007, which appears to have been the first national proclamation in recent decades expressing the intention of the administration to prioritize electrification. The electrification plans, policies, and programs which followed in the years leading up to present day Rwanda will be discussed in further detail in the following sections of this report, but as the references are many, a timeline has been provided to assist the reader in following the narrative. Items referenced in this report are shown in blue text in Figure 16, with the two reports which constitute the main focus of the discussion displayed in bold font. Any items shown in grey have been included for context only and do not figure materially into the report.
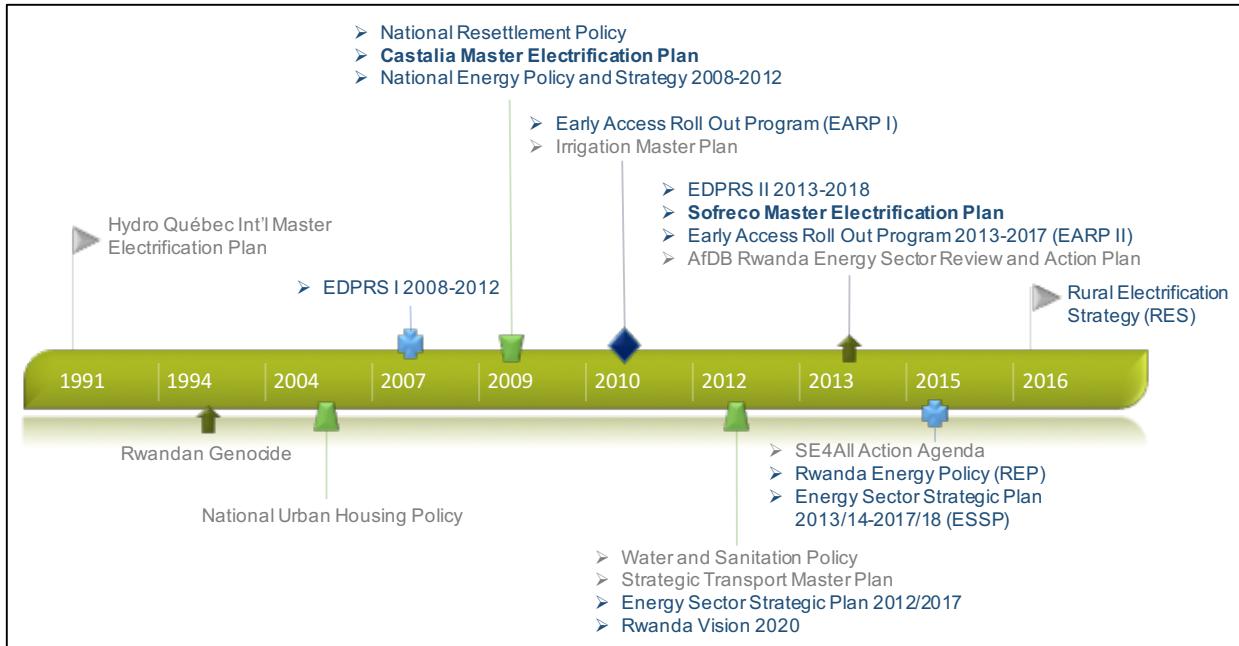
Figure 16: Timeline of Electrification Plans, Policies, and Programs in Rwanda

It is worth noting here that among those items shown in grey are the 2013 African Development Bank (AfDB) *Rwanda Energy Sector Review and Action Plan* [39] and the 2015 *SE4All Action Agenda* [40] reports. The first was produced by the AfDB 'in partnership' with the Rwandan government with the objective of providing input to the 2013-2017 Economic Development and Poverty Reduction Strategy (EDPRS II). The latter was authored by MININFRA with support of the AfDB and, according to the publication, "…has been integrated into the workplan of the country's Sector Working Group, the main coordination and advisory body bringing together stakeholders in the sector including the key ministries, government agencies, development partners, private sector representatives, civil society and NGOs. The Action Agenda reflects inputs from all of these groups." Both of these reports were drafted with the intention of formulating a strategic 'roadmap' and action plan in support of Rwandan energy sector development and would make excellent resources for a reader wishing to consider, in further detail than is covered in this thesis, the institutional and regulatory dynamics which influenced the energy sector during this time period.

4.1.2   Review of Past Master Electrification Plans

In this section, the two master electrification plan reports identified, those written by the consultants Castalia and Sofreco, respectively, will be reviewed along with their supporting documents for the purposes of identifying the methodology followed by each in the creation of a national electrification plan for Rwanda. The two methodologies will then be compared in order to identify common elements, unique attributes, strengths, and weaknesses. These findings will then be used to form the basis of a critique of the capabilities of the Reference Electrification Model, so that recommendations may be formed concerning both the future use and development of this techno-economic planning tool.

46

4.1.2.1   Castalia Master Plan Study (2009)

The Castalia report is comprised of two documents: the Prospectus and the Technical Annex. Both were consulted for the purposes of this review.

4.1.2.1.1   Context

The Castalia study was published in 2009, two years after the release of the Economic Development & Poverty Reduction Plan (EDPRS I), which called for an increase in the total number of electricity connections from 100,000 in 2008 to 350,000 in 2012 (~16% of households [36]) and the prioritization of establishing connections for social infrastructure (i.e. health facilities, schools, and administrative offices) with the objective of electrifying 100% of both health facilities and administrative offices and 50% of schools by 2012. A memorandum of understanding (MOU) had been signed the year before, in 2008, between the Ministers of Finance and active development partners agreeing to facilitate an increased level of coordination and cooperation among actors belonging to government institutions, utilities, development agencies, and development partners in what was being called a sector-wide approach, or SWAP[26]. The Castalia report describes this as a "move away from implementing one-off energy projects towards a systematic and coherent approach to planning, financing, and making new electricity connections."

The Castalia report also makes reference to the government having recently enacted multiple pieces of legislation geared toward reorganization of the energy sector and refinement of related development plans. Among these laws were those calling for the creation of a new electricity utility 'RECO' and the establishment of a National Energy Development Agency 'NEDA'. The independent regulator RURA was also purportedly given additional powers and responsibilities. In general, it indicated that the new legislation sought to transition away from government-managed energy provision toward a system involving a greater degree of private-sector investment and participation, specifically inviting private investment in generation. Exclusive control over the transmission system was given to Electrogaz. The new energy policy, National Energy Policy and National Energy Strategy 2008-2012 [41], called for cost-reflective tariffs and required RURA to consult MININFRA before setting tariffs in order to be clear about the level of subsidies which could be expected from the government.

4.1.2.1.2   Objectives

The Castalia Prospectus states that the intention of its planning exercise was to develop a five-year least-cost technical plan satisfying the electrification goals set forth by the EDPRS, which it referred to as "set[ting] the medium-term framework to 2012 for achieving the long-term aspirations embodied in Vision 2020". The reference being made in this instance is to the original version of Rwanda's Vision 2020 which was published in the year 2000 and which established the goal of reaching 35% countrywide electrification by 2020. In addition to developing a plan that would demonstrate how the 35% electrification goals could be met from a technical standpoint, Castalia was tasked with the complementary objective of providing a phased investment plan to instill confidence in would-be investors, whom it calls out as its primary audience: "The Prospectus seeks to raise US$250 million from development partners to fund access programme investments. The Prospectus aims to ease the analytical burden on development partners by providing credible information on Rwanda's electrification plans and presenting all relevant information and analysis on the access programme in one document."

---

[26] See the Energy Sector Management Assistance Program (ESMAP) report entitled *Rwanda | Extending Access to Energy,* [50], for further details.

4.1.2.1.3 Assumptions

There was no section in the report dedicated to the enumeration of all assumptions used by Castalia in the development of its plan, but an effort was made to formulate a list of assumed values as they became apparent over the course of reviewing the Prospectus and Technical Annex. This is not, therefore, to be taken as a necessarily comprehensive list but, may be considered at least as a suggestion of the types of values assumed.

Assumptions appearing in the Technical Annex include those pertaining to the technology selection process (e.g. discount rate, population growth rate, average number of people per household), those pertaining directly to the off-grid technologies considered (e.g. costs of generator models of varying capacities, annual maintenance costs, installation costs), and those employed in what the report refers to as the "operations and financing model".

Assumptions appearing in the Prospectus:

- Technology costs: Investment lifetime was considered to be ten years for all energy sources; the costs each technology configuration was expected to incur over a ten-year period were discounted into present value terms assuming a discount rate of 10%.

- ECIV2 survey responses were used to estimate the percentage of households per district that can afford electricity. The following inequality was used to determine affordability:

$$Total\ Expenditure * Energy\ Expenditure\ (\%) > Current\ Tariff * Minimum\ Consumption\ (kWh)$$

Where *Minimum Consumption* was assumed to be 20kWh per month (the equivalent of 2-3 energy efficient[27] light bulbs and a radio) and approximate average energy expenditures were considered as follows:

Table 12: Consumer Energy Expenditure Estimates by Location Type

Area	Max. Energy Expenditures (%)	Expenditures on Min. Consumption (%)
Urban	10	2-3
Peri-urban	15	4-5
Rural	20	8-10
Deep rural	25	10-14

- The study assumed that economic growth would translate into an average annual household income increase of 5%, which over the following 10 years would result in an aggregated increase of 1.5% in the population that could afford an electrical connection. It was not, however, made clear why a 10-year planning horizon was used, as Castalia states its objective as developing a plan that can achieve the EDPRS 5-year goals and that it is considering 20 years as the long-term planning horizon.

- It is assumed that the following system sizes will be sufficient to meet the demand (nature of demand unspecified) for social institutions receiving standalone solar PV systems in lieu of grid connection: health centers – 3kW, schools – 1.5kW, administrative offices – 1kW.

- 100% grid reliability and sufficient generation supply were assumed to be available for the duration of the planning period.

---

[27] This language is taken from the Prospectus; it is assumed that energy efficient light bulbs referred in this instance to compact fluorescent lights (CFLs).

- For the purposes of microhydro minigrid cost estimation, an average ADMD (defined as After Diversity Maximum Demand) of 300W is assumed along with a power factor of 0.99.

- New grid extension will be funded as follows: 10% of the capital cost will be covered by connections fees (although consumers are purportedly only expected to contribute 10% of this value, or ~$100USD), 10% will be paid for by the utilities through surplus tariff revenues, with the remaining 80% being covered by the government and development partners.

- The government was planning to complete a full tariff study the year this master plan was studied.

4.1.2.1.4 Inputs

- Table 13 shows the inputs considered in the development of the Castalia master plan as reported by Castalia in the Prospectus.

Table 13: Table of Inputs as Provided in the Castalia Prospectus (source: [36])

Table 4.1: Variables Considered in Planning the Access Programme

Type of Characteristic	Variable	Source	Figure Reference
Demographic	Population data and growth[11]	National ID Survey	Figure 4.2
Political	Country and administrative boundaries	National Institute of Statistics and MINALOC	Figure 4.2
Political	Presidential priorities	Electrogaz	Not shown
Social	Administrative offices	MINALOC	Figure 4.3
Social	Health centres	MINISANTE	Figure 4.3
Social	Schools	MINEDUC	Not shown
Natural features	Roads, rivers, national parks etc	MININFRA and MINIRENA	Not shown
Electricity infrastructure	Location of power lines (HV, MV and LV)	Electrogaz	Figure 4.4
Electricity infrastructure	Location of micro-hydro sites	BTC micro-hydro atlas	Figure 4.4
Demographic	Income levels and poverty data	NIS EICV2 Survey	Figure 4.5

- It was indicated in the Prospectus that income level and poverty data were only available at district level.

- Per-cell population density values were calculated from data collected by the Rwanda National Institute of Statistics (NIS) during the 2008 National ID Project and entered as cell attributes in ArcGIS.

- The geographical location of all prioritized social facilities, including health clinics, administrative offices, and schools were obtained from NUR-CGIS (the Rwanda Centre for Geographic Information Systems and Remote Sensing), and the MINEDUC and MINISANTE ministries, and were incorporated into the GIS database.

- Existing and planned electrical network information was obtained from NUR-CGIS and MININFRA and a subset of potential microhydro sites were taken from the "Hydropower Atlas", which was made available through a study conducted by BTC in 2008. It should be noted that the microhydro sites selected from the atlas for consideration in this study were only those having a potential installed capacity of 1 MW or less.

- Average household income by district was taken from the House Living Conditions Study EICV2 Survey, conducted by the National Institute of Statistics of Rwanda (NISR) and a GIS map was created indicating the percentage of the population within each district living under the poverty line of $1.25/day. The results of this exercise are shown in Figure 17.


Source: EICV2 Survey
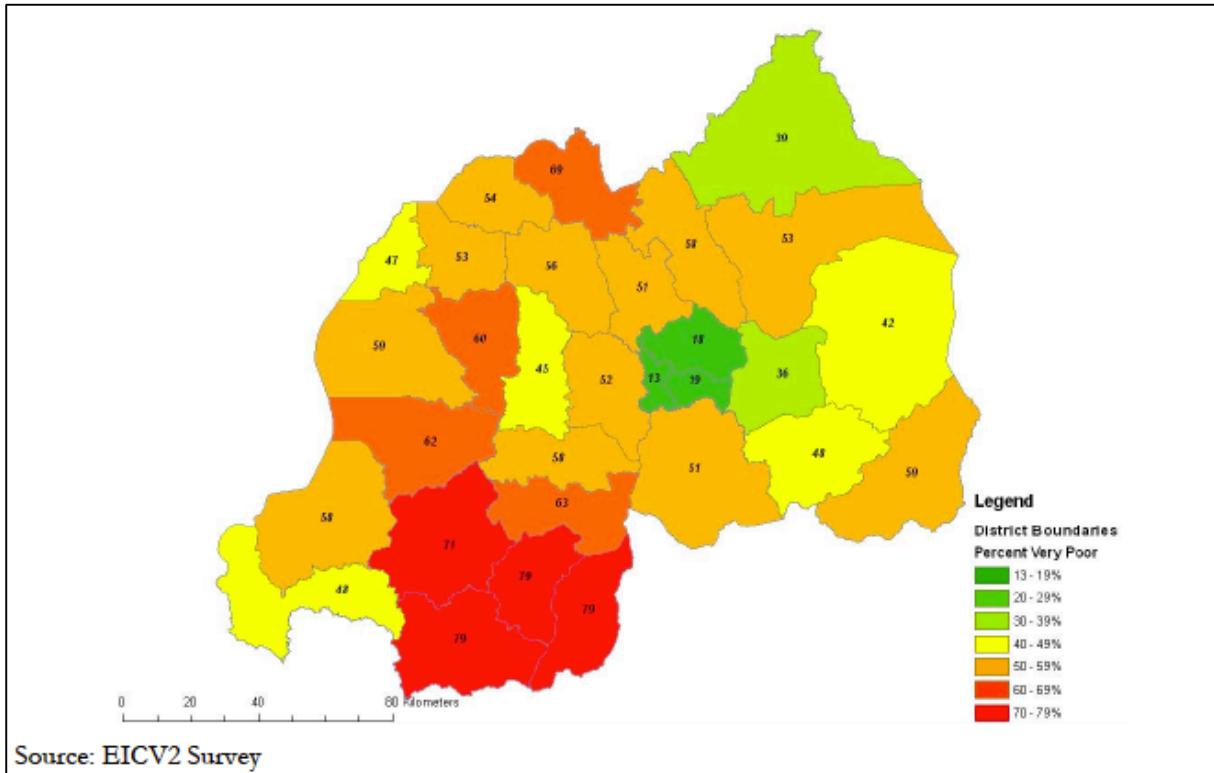
Figure 17: Percentage by District of Rwandans Living on Less Than $1.25 per Day (source: [36])

- 'Typical' Conductor Sizes, as shown in Table 14, were provided by the utility for use in assessment of the existing network, as the Prospectus indicates that "Distribution network information on line and cable types for distribution feeders in Rwanda is currently very limited."

Table 14: Standard Conductor Size Chart as Provided by RECO/Electrogaz for the Castalia Report

kV	Conductor	Current (Amps)	Capacity (MVA)
110	240/40	610	117
70	132/24	420	51
30	120/20	410	22
15	70/12	300	8
6.6	35	190	3

- The Prospectus indicates that network loading data was obtained from the National Dispatch Center, but no further details as to the nature of this data, the form in which it was provided, or how exactly it was used, were provided.

4.1.2.1.5   Methodology

Castalia states that the planning horizon used was 20 years, with a 'directed focus' on the first five years in order to address the EDPRS electrification targets. The software used included Microsoft Office Excel and ArcGIS version 9.2. Below is a figure provided in the Prospectus for the purpose of illustrating the steps followed in the planning process. The titles from the planning blocks shown in this illustration will be used herein as the framework for the description of the methodology employed.
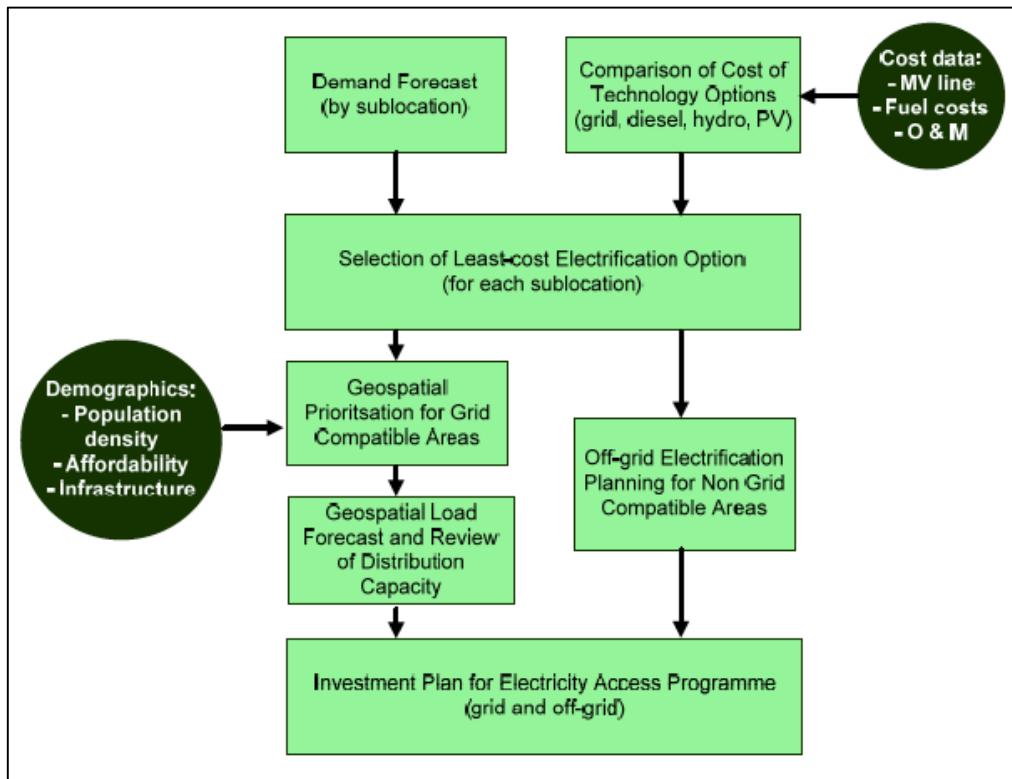

Figure 18: Planning Process Steps as Defied by Castalia (source: [36])

Demand Forecast (by sublocation)
Sublocations used for the purposes of this study were administrative divisions referred to as cells, defined by the Rwanda National Institute of Statistics (NIS) to support its 2002 census data gathering activities. It is indicated on pg. 23 of the Prospectus that "The technology selection model first conducts a simple load forecast for each planning cell, using variables on population growth, affordability and load growth for new connections (using an S-curve growth rate)." Details regarding the load forecast calculation are not provided, but average estimated annual per customer demand values are furnished in the Technical Annex and are shown in Table 15. A load forecast, for which the source listed is "Electrogaz Reports and Government Estimates", was provided on pg. 4 of the Prospectus to accompany a brief discussion regarding utility-scale generation expansion planning, but there was no indication that this load forecast information was used in the Castalia study itself.

51

Table 15: Estimated Average Consumer Demand Per Customer

General Category	Customer Type	2009	2010	2011	2012	2013
Residential and Institutional	Existing	1224	1260	1296	1332	1368
	Urban	993	1031	1074	1125	1155
	Peri-urban	649	679	720	774	795
	Rural	443	460	481	505	519
	Deep Rural	354	368	385	404	415
Commercial	Existing	80016	82016	84888	87432	87432
	New Connections	40008	41208	42444	43716	43716

*All values shown are in units of kWh.

Comparison of Cost of Technology Options (grid, diesel, hydro, PV)
The off-grid technology options considered include diesel generator set microgrid, microhydro generator microgrid (where resources are conducive), and solar photovoltaic (PV) coupled with batteries for standalone systems only. Microgrids consisting of technology combinations (e.g. microhydro or solar PV combined with diesel backup) do not appear to have been considered. The selection of electricity provision technology was performed on a per-cell basis, which means that all inhabitants of a given cell will receive power from the same technology. The cost (both fixed and variable) of each off-grid energy provision option was considered and the least cost option selected. This analysis was performed in a spreadsheet-based model that is referred to in the Prospectus as the "Technology Selection Model". Images of spreadsheet model results are provided in the Technical Annex but the calculation details used to obtain those results and the actual spreadsheet were not made available for review.

Selection of Least-cost Electrification Option (for each sublocation)
This analysis was also performed using the Technology Selection Model. Now that the least cost off-grid option has been selected, it is compared with the cost of grid extension, and the least cost of these two options is assigned as the electrification mode for the cell. The analysis is performed as follows: the cost of the selected off-grid option is divided by the cost per kilometer of medium voltage (MV) line and the resulting length is considered to be the maximum theoretical conductor length for which grid connection would be economical for the cell in question. The value used as a proxy for the actual grid extension distance the cell would require is the distance from the center of the cell to the edge of the nearest adjacent cell. If this distance is less than the max theoretical conductor distance, the cell is considered to be 'grid compatible', or best suited to grid connection. Otherwise, the cell is slated for electrification with the least-cost off-grid source.

It should be noted that it is not disclosed the capacity of the MV conductor or what cost per kilometer value was used in this procedure. Presumably the conductor size was not the same for every cell but was a function of the projected demand of that cell, but this information was not included. In addition, if the cost of the off-grid option used here is the same as that used for the off-grid technology selection, then it includes both fixed and variable costs. This would imply that the cost per kilometer of the MV conductor would also need to include variable costs, such as operations and maintenance costs. According to an entry on pg.23 of the Prospectus, a section of the report other than that in which this process is initially presented, the procedure is described as follows: "The model then compares the best decentralized electrification option with the capital and operating costs of a "generalised grid" supply, initially disregarding the location of the existing electricity network." This would seem to imply that the cost per kilometer does in fact include some manner of operating costs, although this value too remains unspecified.

Geospatial Prioritization for Grid-compatible Areas
Cells deemed 'grid compatible' are then weighted in order to determine the priority order for connection to the grid. Weights are based on the following attributes of a cell: proximity to the existing medium-voltage network, number of unelectrified households, inter-household distance (determined from population density values), number of unelectrified social facilities, access to road networks, consumer ability to pay, existence of economic development centers/markets, and presidential priorities for electrification. The methodology employed for evaluating these inputs and using them to assign weighted values to cells is extremely vague and appears to be based on a purely subjective visual assessment of the data as imported into and portrayed in ArcGIS. The extent of the information provided in the Prospectus regarding this weighting assessment process is as follows: "Each attribute receives a weight according to its relative significance in determining the likely costs and benefits of electrification. The total score for each sublocation is then used to rank the planning cells relative to each other. Cells that receive the highest scores are prioritized for electrification in the access programme, according to a realistic assessment of the number of new connections that can be achieved in each year." It then goes on to say that "This approach captures many of the elements that would form part of a project cost-benefit analysis – higher population densities, proximity to the network and access to roads will all tend to lower the costs of the programme. Prioritizing cells with more unelectrified households, social institutions, and economic activity will tend to increase the economic benefits of the programme."

Based on this cell prioritization, a least-cost annual grid extension approach was outlined for the next five years, 2009 through 2013, as was a final plan for the year 2020 claiming to achieve the Vision 2020 goal of 35% national electrification. Results are discussed in the following section. It is indicated in the Prospectus that an additional non-least-cost scenario was considered at the behest of the government which "provides grid coverage in ever sector by 2012". However, the corresponding plan illustration is for "100 Percent Grid Coverage of Administrative Sectors" and the legend indicates that the connections shown are through planning year 2013. In either case, it is not clear if 'providing grid coverage' to a 'sector' consists of electrifying all inhabitants, or extending MV line to that cell and connecting only certain consumers. Due to lack of details, the methodology by which the plan for this scenario was developed cannot be elaborated upon further for the purposes of this paper.

Geospatial Load Forecast and Review of Distribution Capacity
Although the Prospectus makes claim to development of a 'geospatial load forecast', neither the method by which this load forecast is developed nor the assumptions used in its development are discussed. With regard to the process of reviewing the distribution capacity, the Prospectus indicates only that the "typical conductor sizes have been compared with network loading data received from the National Dispatch Centre", but supporting information was not provided.

Off-grid Electrification Planning for Non-Grid-Compatible Areas
Microhydro: As described in the Castalia inputs section, section 4.2 of the Prospectus reiterates that of the potential hydro sites identified in the Hydropower Atlas, only those with a capacity equal to or less than 1MW were considered as potential mini-grid microhydro sites and entered into the GIS input file. However, the Prospectus then indicates in section 4.5 that only those sites having a potential capacity of less than 750kW were considered. Based on the microhydro generator sizes listed in the Technical Annex, combinations of the following generator sizes were considered in determination of system cost: 5,25,50,100, and 250kW. It is indicated that sites with higher expected net present values, as well as those which are not expected to be capacity-constrained within the following ten years, were prioritized. Based on this information, the top 25 sites were selected to factor into the five-year development plan (without indication as to why the number 25 was used). Of these 25 microhydro sites, it was then determined which should be recommended for interconnection to the grid based on whether or not the sum of the costs to develop and interconnect the site were less than the estimated long-term marginal grid supply cost estimation of $0.12/kWh.

Solar PV: Social institutions which lay outside of cells selected for grid connection in the 5-year plan were identified and included in the 5-year plan as recipients of solar PV units.

4.1.2.1.6   Report Conclusions and Recommendations

The results of the technology selection model are shown in Figure 19. Distribution lines shown represent the existing network. A total of 95% of the cells were identified as being grid compatible, with microhydro mini-grids representing the least-cost option for about 4.5% of the cells, and diesel mini-grids and standalone solar PV being considered the least-cost option in fewer than 1% of the cells. It was further concluded in the report that while microhydro facilities may or may not be connected to the grid, depending upon when the hydro resource is developed and when the grid arrives, the diesel mini-grid cells (shown in brown) identified by the process were close enough to the existing grid that these cells would most likely be interconnected instead of electrified with diesel, although based on their relatively low demand they were nevertheless considered as not justifying prioritization under the access program. A note to the reader to be aware that substations and grid-compatible cells were assigned the same color (bright green) in the figure below, as were microhydro cells and water bodies (light blue), and while there do not appear to be any solar PV cells, the color chosen for solar PV (yellow) is so similar to that of substations and grid-compatible cells that it may be effectively dissimulating any that are displayed.
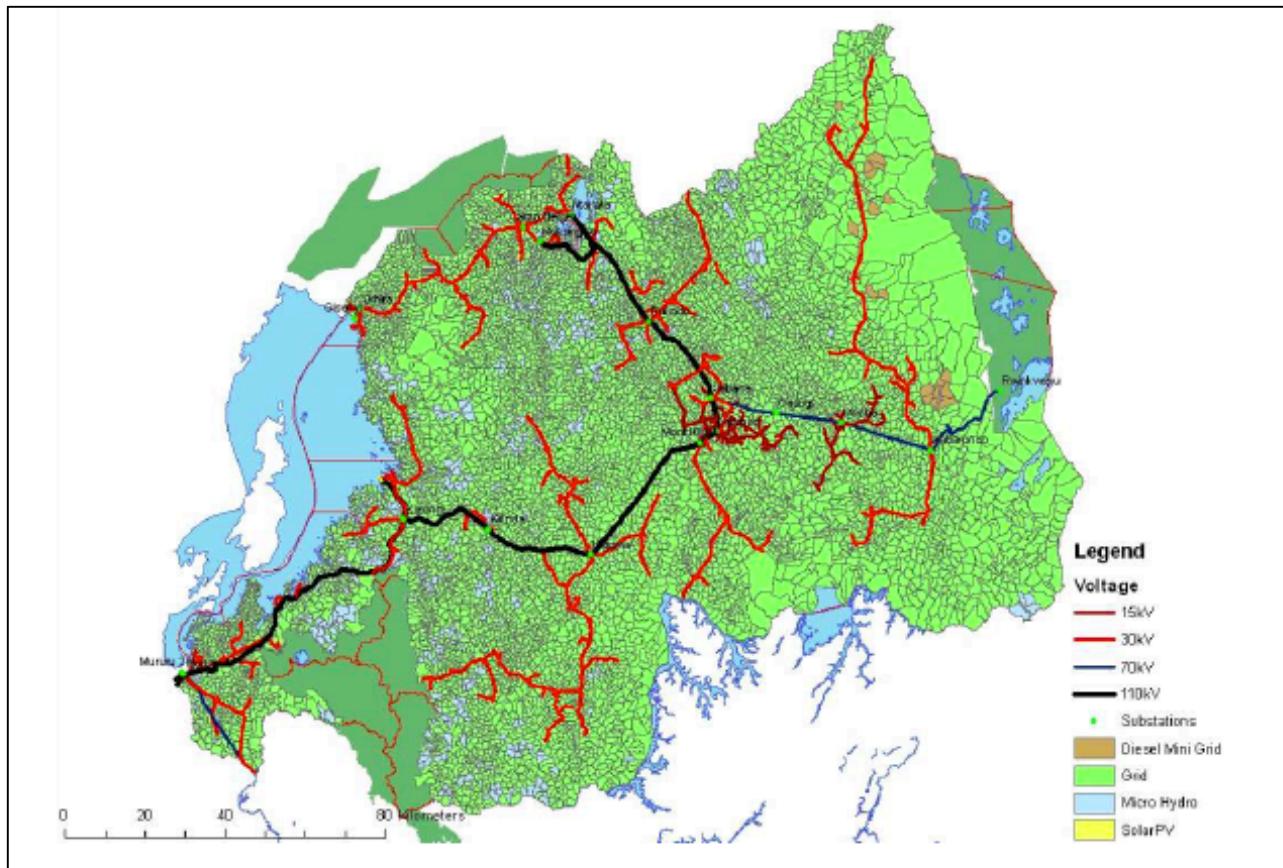

Figure 19: Technology Selection Results (source: [36])

Figures were also provided to illustrate the proposed network expansion. Based on these figures, proximity to the grid would appear to have been the attribute of greatest importance in cell prioritization. The relative contributions to cell weight made by other attributes is not readily apparent. The study

54

concludes that >60% of the population were at that time living within five kilometers of the existing grid, and that over 95% of the connections planned to be made through 2012 lay within this buffer, indicating that "the government's targets can be achieved without a massive extension of the medium voltage (MV) network". It goes on to say that a non-insignificant number of the new connections are infills, meaning that they are slated to be made in areas which already have an existing low voltage (LV) network and are concentrated in urban and peri-urban areas, which the report contends will also help to minimize program costs.

A figure illustrating the existing network and the corresponding five-kilometer buffer is available in Appendix IV, along with figures portraying the annual recommended grid connection scenarios for the ensuing five years, the plan for achieving 100% Grid Coverage of Administrative Sectors, and a plan which achieves the Vision 2020 goal of electrifying 35% of the population by the year 2020. An annual planning approach for progressing from the plan in 2013 to that of 2020 was not provided.

Other results provided in the Prospectus and Technical Annex include:

- Medium and low voltage (MV/LV) annual investment totals, although no mention was made in the Prospectus or Technical Annex as to how LV costs were estimated (no LV-level distribution designs having been shown or discussed at any point in the report). The Technical Annex does provide annual lump sum totals for the following items:
    - MV Lines: Wires & Accessories, Poles, Transport, Labor
    - MV/LV Transformers: Transformers & Accessories, Transport, Labor
    - LV Lines: Wires & Accessories, Poles, Transport, Labor
    - Service Connections: Service Cable & Meter, Transport, Labor
- The following conclusions were drawn with regard to network upgrades: 'An initial high-level assessment of current loading indicates that adequate capacity will be generally available to support the new connections during the initial years of the access programme. However, detailed network modelling and load flow analysis will be necessary in the implementation stage of the access programme to identify potential capacity constraints in specific areas." It goes on to say that a cost estimate for network upgrades has been provided with the report but that it is based only on expected demand growth.
- Sample results of the technology selection process are provided in the Technical Annex for 20 out of the 9300 cells evaluated. Included in these sample results are the technology types and sizes considered for each cell, the corresponding total system costs that were calculated for each technology, the max MV length, distance to the nearest cell, and least cost technology selected.
- The following are titles and subcomponent descriptions of additional results figures not discussed above (all values listed are provided on an annual basis for the years 2009-2013, some through the year 2020):
    - Capital Cost Components of Electricity Access Programme – per km costs for network extensions, per connections costs for low voltage connections separated into infill, expansion, and CFLs(?), network strengthening ($USD/MWh), off-grid technology connection costs for both solar PV and microhydro ($/kW), solar hot water and technical assistance.
    - Estimated System Operating Costs – Costs associated with current connections including generation, transmission and distribution, and administrative costs; The same breakdown of costs but for new grid connections; Off-grid connections; Operating contingencies.

- Summary of Annual Capital Costs Associated with Sector Investment Plans (excluding generation under PPAs) – Generation, HV Transmission, Distribution Rehabilitation, Access Program.

4.1.2.2   Sofreco Master Plan Study (2013)

The Sofreco study is comprised of eleven reports, one planning and one load flow report for each of the five planning regions and a combined report summarizing the methodology used and findings at the national level. Several supporting spreadsheet models were provided along with the report as well, in addition to a full suite of GIS shapefiles. All documents were consulted for the purposes of this review.

4.1.2.2.1   Context

The Sofreco report was published in 2013, the same year that the second Rwandan Economic Development & Poverty Reduction Strategy (EDPRS II 2013-2018) was published, the Early Access Roll Out Program was updated (EARP II), and the year after the revised version of Rwanda Vision 2020 and the Energy Sector Strategic Plan 2012/2017 were published. The Vision 2020 document indicates that the government had set a goal of increasing the electrification rate to 1.5 million (or ~70% of the population) by 2017. EDPRS II called for 45% of households to receive grid connection by 2017 through EARP, with the remainder of the population to be connected through off-grid installations. The Energy Sector Strategic Plan also indicated that the "government will support programs to ensure that 100% of households get access to electricity through grid and off-grid solutions by 2017".   There is also a continuation of the initiative discussed in the Castalia Prospectus to electrify social institutions which are not slated to receive grid connection in the subsequent five years via solar PV, although goals have at this point been increased with respect to the desired electrification rate of schools, from 50% to 100%. The goal for health facilities and administrative offices remains at 100%. By this time, MININFRA has been appointed responsible for managing the off-grid electrification efforts, while EWSA has been placed in charge of the grid component. According to Figure 14, the national electrification rate in Rwanda had grown to approximately 17%.

4.1.2.2.2   Objectives

The stated objectives of Sofreco are to A) develop a plan to support the Early Access Rollout Program to achieve the goal set forth by the government to increase the national electrification rate to 70% by 2017, and B) provide the government with an estimate of how many people would need to be relocated to the Imidugudu settlements in order to achieve this goal. Sofreco describes its role as follows: "SOFRECO was appointed as a Planning Consulting Firm with its mission's main objective to assist the newly formed Planning & Design Unit to meet the electrification targets through the following: GIS, Electricity network design & planning, capacity building, and training activities."

4.1.2.2.3   Assumptions

- Consumer load values (see Table 16).
- Province-level load forecast assumptions (see Table 17).
- It was assumed that only 75% of potential consumers would connect during the planning period (although scenarios where 100% connection was achieved were also considered), and that population growth would remain at a steady 3% for the duration of the planning period.
- The cost for a fill-in connection occurring during the year 2012 was established by EWSA, independent of this report, to be $600USD (where a fill-in connection is defined as any connection made to a portion of the network in a year other than the year of its construction). The

assumption was then made in this report that the connection costs in subsequent years would be based on this value but reduced by $50 each year "as more and more infrastructure will be installed, and less low voltage has to be constructed each year to cater for all the fill-in connections."

- The phased capital investment schedule (see the results section) assumes a 5% escalation rate.

- The Sofreco study assumes 100% grid reliability and that sufficient generation supply will be available to support the developed plans. It is stated in the report that the national target at that point in time was to achieve a total of 1000MW of installed generation capacity by 2017.

4.1.2.2.4   Inputs

- Orthophotos which were several years old at the time, but which presumably covered the entire country, were provided by EWSA.

- Consumer locations were derived from these orthophotos.

- Topographical data including roads, rivers, lakes, wetlands, and national parks was obtained from various institutions.

- The locations of planned 11,782 Imidugudu settlements were provided by MINALOC.

- The 2012 EWSA connection forecast.

- Locations of social infrastructure – initial dataset provided by Rwanda government and EWSA but data considered to be outdated so efforts were made by Sofreco to update this database.

- Existing electrical network – the initial dataset, presumably a vestige of the Castalia study, was found to be incomplete and incorrect so efforts were made by Sofreco to update this database.

- Loading of existing substations and feeders and 'normal open points on the network' were provided by the National Control Centre.

4.1.2.2.5   Methodology

The Sofreco combined report states initially that electrification via extension of LV and MV lines will be "implemented based on the prioritization methodology described in the project Prospectus", which presumably refers to the Sofreco Prospectus, as it then goes on to say that, "During the inception period, the revision of the order of prioritized zones for undertaking the design and planning activities was discussed with the Director of the Planning and Design Unit at EARP. Considering available information, on-going projects, recently completed projects and the overall state of network development in each zone, it was agreed that the activities would be carried out in the following revised order:…". The report goes on to describe the reorganization of the electrification priorities by zone as follows (items taken directly from [11]):

1. South Zone: This zone has minimal activity in terms of new additional consumer connections. This has to be taken up in first priority.
2. North Zone: Three big EPC contractors are being awarded contracts to undertake approximately 14000 connections. The MV extension is expected to be 150 km. The project completion period is 18 months and the work is likely to begin in early 2012. This zone is considered priority-2 and has to be started after the South Zone.
3. Western Zone and Eastern Zone: In the Western zone, about 150 km of MV line is under construction. The LV Line component will be built by EWSA. This zone is considered priority-3 and is to be started after the North Zone. A Tunisian Company, STEG, is currently adding 500 km of line in the Eastern Zone to achieve an expected additional 50,000 connections.

4. Central Zone, Kigali: The Central Zone is considered priority-4 and to be started after completion of the other zones indicated above.

All data available from EWSA was incorporated in the geodatabase and a suite of institutions, both governmental and non-governmental, were solicited for any further information which may have been available. Extensive efforts were then undertaken to update both the network infrastructure and social institution location portions of the database. Technicians were deployed by EARP and Sofreco to collect location data on social facilities in order to update the information provided by EWSA and the government. Additional infrastructure types were added over what had originally been provided, including factories, markets, and universities. Approximately 2,500 additional infrastructure points were added to the approximately 4,000 Sofreco had initially been provided.

The country was divided into five zones which bear some similarities to the administrative provinces, in terms of region naming convention and boundary, but differ in that "the location of the source determines the zone for study purposes", which presumably refers to the generation source(s) considered to be serving that area. Figure 20 shows the Sofreco planning zones (left) as compared with the Rwanda administrative boundaries (right).
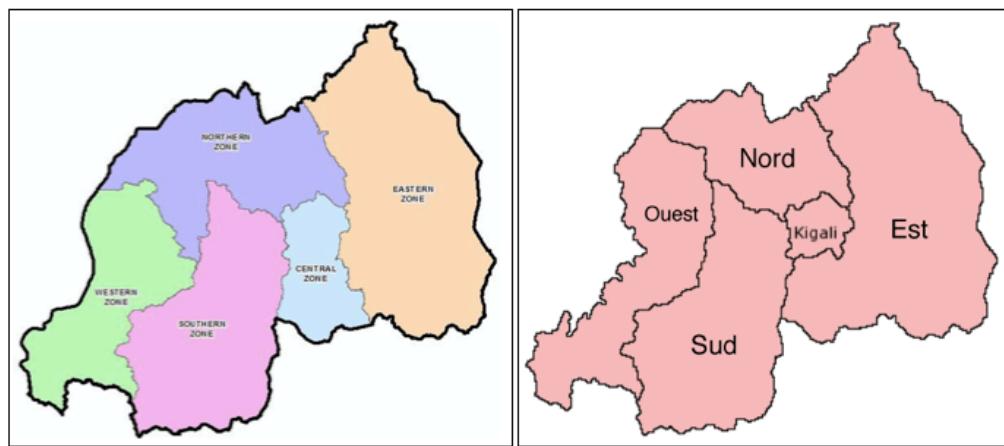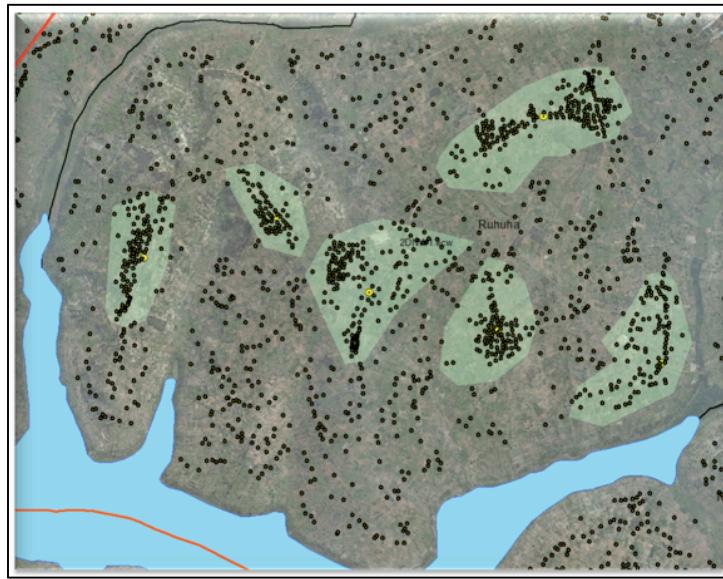

Figure 20: Sofreco Planning Zones (left) (source: [11]), Rwanda Administrative Provinces (right)

Consumer locations were identified by placing a point on top of each dwelling as detected using the orthophotos provided by EWSA. These points were then compiled into shapefiles. It is presumed that this location exercise was performed manually, as the report contains no specification to the contrary. Just over 1.7 million households were identified.

Population 'concentrations' were identified manually through GIS map visualization of the consumer points identified from the orthophotography. An example of these clusters is shown in Figure 21. It is not evident that any maximum customer-to-customer distance or other numerical grouping criteria was employed in this process. These groupings are referred to in the report as planned MV/LV transformer zones and were made to encompass not only consumer concentrations but also to include all Imidugudu settlements and as many social facilities as possible. Zones were created covering a total of 482,854 *existing potential* consumers. That figure does not account for expected population growth due to Imidugudu resettlement but the report indicates on pg. 24 that expected growth due to resettlement *was*

considered in the design phase for the purposes of determining network component capacity needs. It is worth noting here that it is not entirely clear from the order of planning operations described in the remainder of the report that this was the case.


Figure 21: Population Clusters or 'Transformer Zones' Identified by Sofreco (source: [11])

The number of customers lying within each transformer zone was then tallied and these values used to estimate both initial and forecasted load values for each zone. Distribution networks were designed for each cluster, an example of which can be seen in Figure 22, with LV conductors sized in order to accommodate expected long-term load growth. Locations of transformers and the routing of MV and LV lines appear to have been identified manually. LV lines were routed along roads with the working definition of electrification being that a potential consumer was located "within service connection length from a low voltage line".
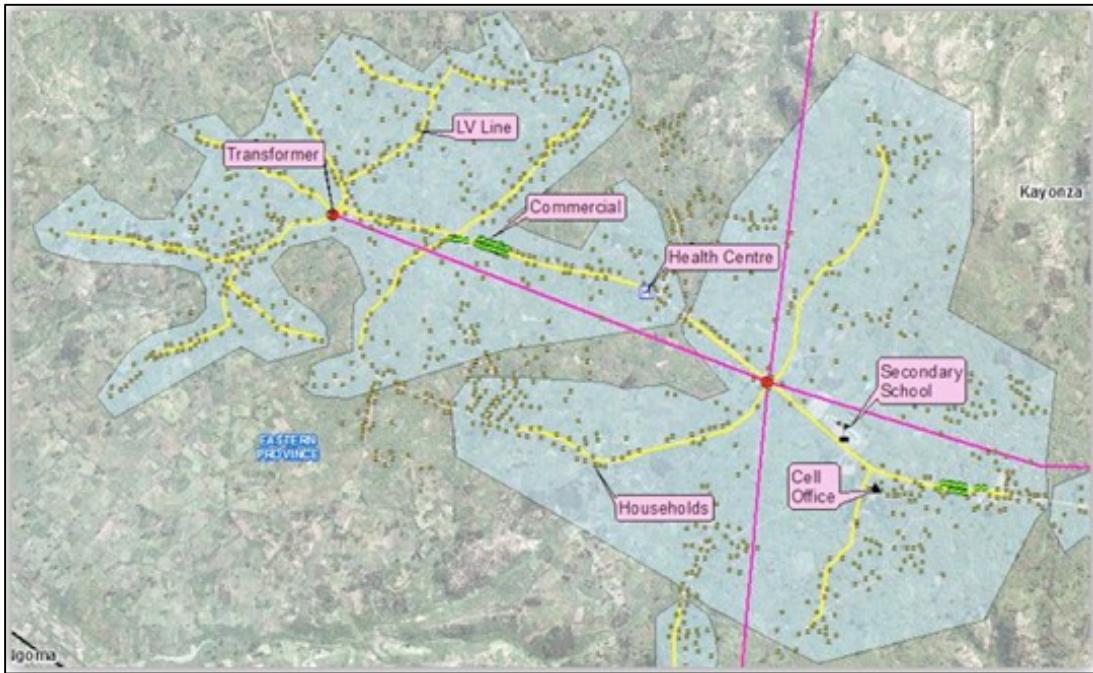
Figure 22: Illustration of Network Design for two Transformer Zones (source: [11])

The report was not explicit about what planning horizon was considered 'long term', but entries in the spreadsheet model would seem to suggest ten years (2013-2022). The report indicates that values estimated for initial demand, demand growth, and saturation for each transformer zone were "derived from statistics and with the assistance of EARP's planning department". These values can be seen in Table 16.
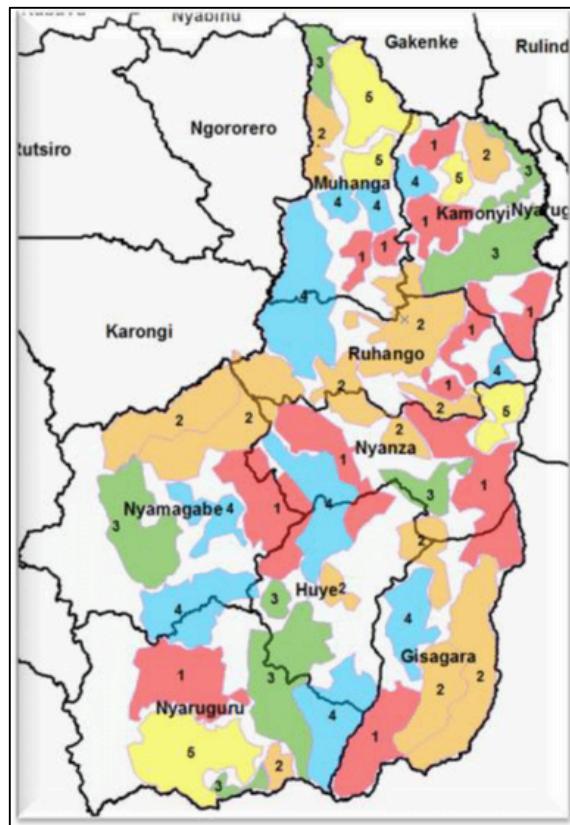
Table 16: Consumer Load Assumptions (source: [11])

Load Category	Rating(KVA)	% Growth from 2012-2017	% Growth from 2017-2022
Household growth	100 watts	300 Watts saturation after 6 years	
Business Centre	10	3	2
Primary school	1	3	1
Secondary school	2	3	1
Health centres	10	4	2
Hospitals	20	2	1
Coffee wash station	2	1	1
Cell office	1	3	1
Sector office	2	3	1
District office	4	3	1
Factories	20	2	1
Market	8	4	2

After the LV network was designed to accommodate the existing infrastructure and forecasted load for each transformer zone, the MV network ties required in order to connect these transformer zones to the existing

MV network were planned. It should be noted that care was taken to route lines with respect to topographical features, avoiding traversing sensitive areas such as wetlands and crossing areas which may incur extra construction and maintenance costs such large bodies of water. It was stated in the report that "all existing and future sources of supply, such as existing and planned substations, planned hydro sites and potential hydro sites" were considered in the planning of new MV conductor sections. "Thicker backbone lines were planned to these sources or potential sources of supply. Also, proposed substation sites were used as future sources of supply and backbone lines were planned from here. New feeders were also created to relieve the load off existing long feeders." Load flow studies were performed using a software called CYME to evaluate various buildout solutions during this line routing process. The CYME model files were purportedly provided to the Rwandan working group but only the regional load flow analysis results have been provided for MIT review.

The transformer zone designs were then grouped into what the report refers to as 'lots' for contract bidding purposes. This process resulted in a total of 146 lots which were then prioritized for construction over a period of five years, from 2013 -2017. Each lot was assigned a priority ranking based on the following attributes: the social infrastructure it contained, its distance from the existing network, government targets for the district where the lot was located, cost, and social demand, which was considered higher in areas with no pre-existing electrical connections. An example of lot prioritization in the southern planning region can be seen in Figure 23.


Figure 23: An Example of Lot Prioritization in the Southern Planning Region (source: [11])

61

A connection forecast was then created for the purposes of determining how many consumers would need to be successfully relocated and connected through the Imidugudu resettlement program in order for the government's goal of reaching a 70% electrification rate by 2017 to be reached. The forecast developed covered the years 2012 through 2017. The forecast used for year one was a connection forecast developed by EWSA. For the years 2013 through 2017, the load was calculated based on the following:

- the timeline for construction based on the aforementioned lot prioritization

- an assumed population growth rate of 3%

- that only 75% of potential consumers will be connected to a new line during the year it is constructed (the remaining being calculated as fill-in connections in the following years, although the report is not explicit as to how these numbers were incorporated)

The sum of the existing connections as of the start of 2012 and total number of connections theoretically achievable by the end of 2017 was then subtracted from the total expected 2017 country population. This remainder was considered to be the number of consumers which must be relocated and connected over the span of the planning period in order to reach the goal of 70% electrification by 2017. It was assumed for cost assessment purposes that approximately 1/5 of this total number of relocated consumers would be connected annually between 2013 and 2017. The analysis was then repeated assuming a potential consumer connection rate of 100% instead of 75%. An updated load forecast was determined including population resettlement values. The load values provided in the spreadsheet model that accompanies the report have been aggregated at the planning province level. Table 17 shows the forecast assumptions used and the resulting start load value.

Table 17: Province Level Load Forecast Assumptions (source: [11])

Load Forecast	South	North	East	West	Central	
ADMD start	0.06	0.06	0.06	0.06	0.3	kw
ADMD Saturat	0.3	0.3	0.3	0.3	0.7	kw
Duration for saturation	6	6	6	6	6	yr
Start Load	18	15.6	4.5	18	65	MVA
% Growth on Existing load	4%	3%	3%	4%	5%	

ADMD, according to [42], stands for After Diversity Maximum Demand and is defined as follows: "After diversity maximum demand is used in the design of electricity distribution networks where demand is aggregated over a large number of customers. After diversity maximum demand (ADMD) accounts for the coincident peak load a network is likely to experience over its lifetime and as such is an overestimation of typical demand."

The ADMD start values provided do not appear to be consistent with this definition, they would appear rather to be per customer average demand values (60 watts per customer in all planning zones except for the Central Zone, for which the ADMD start value is 30 watts). It should be noted that these values are hard-coded in the spreadsheet model provided and thus the more granular data from which these figures were purportedly conceived is not available, making them difficult to reproduce or update. In addition, if the ADMD start and saturation values are intended to represent consumer averages, then they would appear to be highly inconsistent with the table of consumer load values used in the estimation of loads for each transformer zone (see Table 16) whose values range from 100W to 20 kW. The long-term load forecast values, assuming relocation targets are met, are shown in Figure 24.
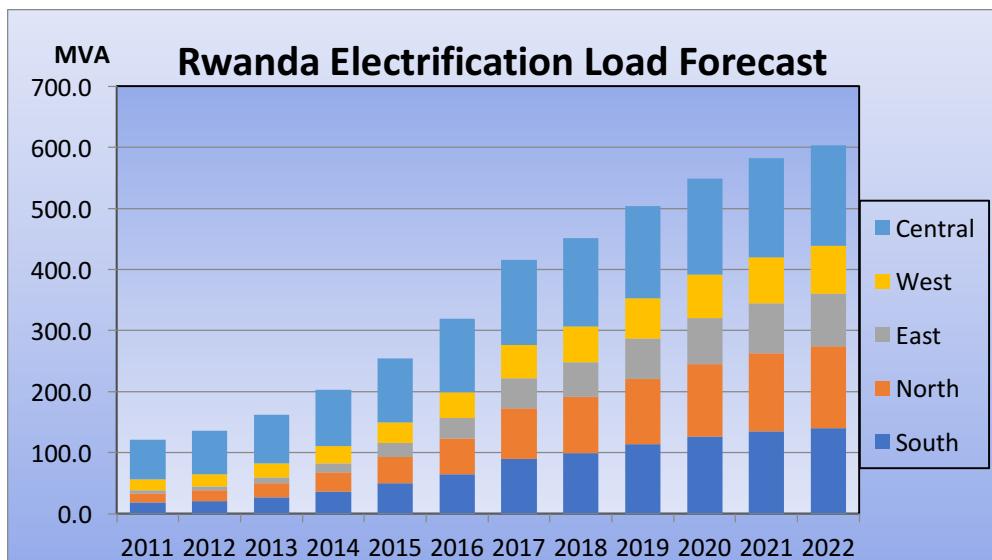
Figure 24: Long-term Load Forecast Assuming Relocation Targets Met (source: [11])

Following the load forecast adjustment, a detailed bill of quantities was developed based on quantity and length values calculated using the GIS software and a phased capital investment schedule was created. This was accomplished by first tallying the capital investment required in each region for new construction based on the lots identified and the year each was scheduled to be developed (it is not clear if this includes cost for suggested upgrades for the existing network). Then the annual costs for fill-in connections, as based on the 2012 value provided by EWSA (see the Sofreco Assumptions section for further details), were tallied. These two sets of values were then added.

4.1.2.2.6   Report Conclusions and Recommendations

It was concluded in the Sofreco report that the network update designs it had formulated left "very few places more than 2 or 3km away from a planned Medium Voltage line", the vast majority of Rwanda's population would eventually be grid connected. The final network design was comprised of 5,083 transformers, 6,594km of MV lines, and 10,058km of LV lines.

A load flow report was created for each region. Each load flow report contains two sections: 1) a detailed analysis of the network, as it existed at the time of publication, in each region on a per-substation basis, making specific recommendations for components which needed to be upgraded based on the existing level of loading and related severity of losses and voltage issues, 2) the same manner of analysis for a proposed network buildout design. Commentary was also provided as to whether or not EWSA already had plans in place to address any needed upgrades identified. It was pointed out in the load flow reports that the conditions on which the recommendations were made are to a certain extent ephemeral and should be reevaluated if buildout plans change or are delayed. "The proposed year of implementation must be carefully monitored by looking at voltage levels and loading. The connection rate of new consumers will determine the rate of load growth. The connection rate can be manipulated by outside factors such as economy, connection application cost, relocation of people, etc." The main report also included a reminder of the importance of ongoing coordination with transmission planning efforts, as these analyses were conducted on the distribution network only.

All case files developed for use in the CYME load flow software, for both existing and proposed networks, were submitted in original format to the Rwanda EARP working group as a deliverable along with the reports so that the model may be rerun and the network upgrade plans reevaluated, should

changes in forecast, load growth, and/or saturation values change. The phased capital investment program is summarized in Table 18.

Table 18: Phased Capital Investment Program (source: [11])

		COST (USD Millions)					
COST SUMMARY		2012/13	2013/14	2014/15	2015/16	2016/17	2017/18
EPC Contracts		0.00	82.14	87.72	86.52	54.63	6.25
MV_LV Contracts		0.00	22.05	19.86	21.73	17.03	3.06
Fill In Projects		68.62	88.82	85.23	80.74	75.36	69.08
YEARLY TOTAL		68.62	193.01	192.81	188.99	147.02	78.39
ACCUM TOTAL ALL (PV)		68.62	261.63	454.44	643.43	790.45	868.84
Excalated	5%	68.62	202.66	212.57	218.78	178.70	100.05
Accum	5%	68.62	271.28	483.85	702.63	881.33	981.38

For all new connections made from 2013 to 2017 other than fill-in connections (where 'new connections' here is defined as all connections made to newly constructed lines in the year they are constructed, including connections resulting from population growth and relocation), and assuming that only 75% of potential customers connect annually, the average connection cost was determined to be $1102USD, decreasing to $1065USD if 100% connection was assumed. The average cost for all customers connected over the roll-out period, including fill-in customers, was calculated to be $631USD.

4.1.2.3   Master Plan Comparison: Sofreco vs. Castalia

The purpose of this section is to use the information identified during review of the Castalia and Sofreco plans to compare the objectives, approaches, assumptions and any other aspects of the two reports which may help to inform the future use and or development of the Reference Electrification Model.

Objectives/Priorities/Major Assumptions
Priorities appear to have been reversed from one report to the next, from A) an attempt by the Castalia plan to maximize the cost-benefit tradeoffs by prioritizing connection of those closest to population centers and the existing network, and targeting regions where the ability to pay appeared to be the highest, to B) the prioritization regime followed by the Sofreco study of making those areas which had seen the fewest connections in the electrification process up to that point top priority and relegating fill-in connections in the Kigali region to the back burner. Furthermore, the possible modes of electrification differed from one plan to the next. Castalia claimed to be formulating a least-cost plan, where off-grid vs. grid extension technology costs were compared for given demand level estimates. Sofreco, on the other hand, did not claim to be striving for a least-cost approach, only a thorough one which aimed to meet the stated national electrification goals. By the time of the Sofreco report, MININFRA had been put in charge of off-grid electrification and EWSA of grid extension. It is not clear if the EARP Planning & Design group was affiliated only with/or under the direction of EWSA, but Sofreco was only asked to look at grid extension. The only off-grid element mentioned in the Sofreco report was the government goal of electrifying social facilities not slated for grid connection in the subsequent five years with standalone solar PV, although the report makes no mention of this beyond the introduction section. The Sofreco plan was also created after the Resettlement Act had gained momentum, so included this information in its plan whereas it was not considered in the Castalia report.

Overall Structure and Organization of the Documents
One thing to note with regard to both reports is the lack of clarity regarding order of planning operations and the manner in which assumptions were used. The planning methodology followed by Castalia was particularly difficult to reconstruct for the purposes of reporting it in this paper. Elements of the planning process followed were discussed out of chronological order with assumptions provided in some locations

but not in others and terminology that was often vague or descriptions which were seemingly conflicting. The Sofreco report was much easier to follow as it was, for the most part, laid out in chronological order, but it still required a relative amount of effort on the part of the reader to piece information together. For instance, it is unclear whether or not the load forecast including the resettled consumers was the one used to inform the creation of the network design. The way the report reads, the forecast including resettlement was determined after the network upgrades were designed. It is thus not evident upon which assumptions the transformer zone load forecasts, and thus ultimately the network design, were based.

Quality of the Input Data
It was the conclusion of both reports that a more comprehensive and up-to-date geodatabase than what was presented to them was ultimately required for the types of planning exercises they were attempting to carry out. Sofreco put a great deal of work into gathering data for and updating GIS files of the existing network, and commented in its report that it discovered during this process that the network files it was initially provided (presumably those created during the Castalia planning process) were not only outdated but also incorrect. The figure below was taken from the Sofreco report, and although no legend was included with the images, they nevertheless give an indication of the amount of work required of Sofreco in order to update the database for use in their planning process.
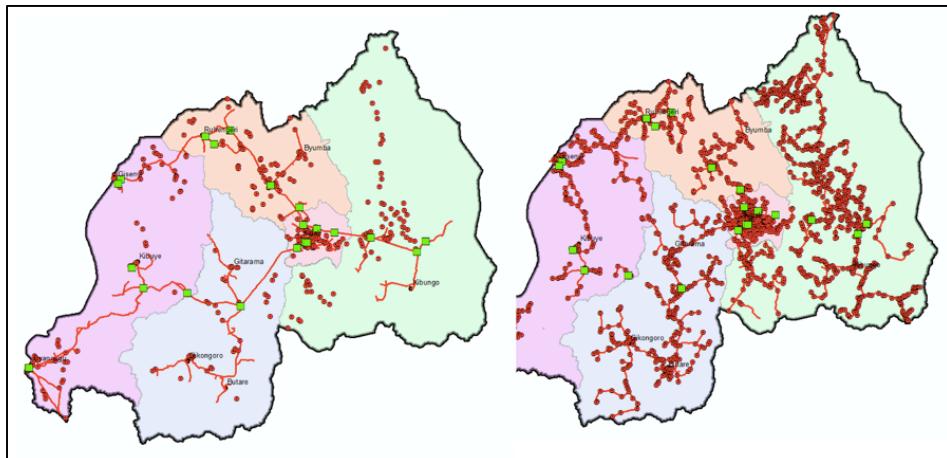

Figure 25: Sofreco Network Data Before (left) and After Sofreco Updates (right) (source: [11])

Methodologies
Several different areas of the methodologies employed by these two companies are worth commenting on, so the discussion has been itemized by subtopic below.

- Distribution Line Routing and Topographical Features: The Castalia report does not appear to have utilized any topographical information in the distribution line routing process, in fact the Prospectus makes no mention of this process at all, although LV and MV component cost estimates are provided. Lack of evidence that LV designs were actually created casts a certain level of doubt on the validity of these cost estimates. Sofreco, on the other hand, created very detailed MV and LV designs and in the process, made use of ArcGIS files which included data on waterways, wetlands, national parks, and roads to ensure that networks were planned such that they circumnavigated undesirable zones and were accessible by road.

- Population Clustering: Both consultants used some manner of population grouping to establish clusters of consumers who would be evaluated as a group, although each did so for differing reasons. Castalia grouped the population by administrative cell boundary for generation technology selection purposes (i.e. every citizen belonging to a particular cell was destined,

according to that planning methodology, to be assigned the same generation type). In the Sofreco study, population 'concentrations' were identified manually through GIS map visualization of the consumer points identified from the orthophotography. It is not evident that any maximum customer-to-customer distance or other numerical grouping criteria was employed in this process. Sofreco referred to these groupings as 'planned transformer zones' and included all of the Imidugudu resettlement areas as well as social institutions, wherever possible.

- Assigning consumers to off-grid vs grid extension: Sofreco did not consider off-grid modes of electrification as it appeared that group was tasked with developing a grid-extension-only plan. Castalia did consider both grid extension and off-grid options, but the off-grid technology configurations it considered were somewhat limited, at least by more recent standards, and the equivalent MV length methodology of selecting between grid extension and the least-cost off-grid technology was somewhat crude, using the closest adjacent cell as a proxy for the nearest network tapoff location.

- Developing a load forecast: The Castalia report indicated that the planning team developed both a load forecast for each cell (for the purposes of least-cost generation technology selection) and then later a 'geospatial load forecast' based on the resulting network development plan, but in neither case did it explain how exactly the forecast was developed. It is presumed that the cellular load forecasts were based on the estimations provided in the Technical Annex of average demand by consumer type, but it is not evident from any information provided in what way the 'geospatial load forecast' differed from the cellular forecast. Sofreco provided a greater level of assumption information than did Castalia, but the report failed to provide a holistic explanation of how the various assumptions were used and at which stages of the development process (with regard, in particular, to the Imidugudu resettlement back calculation and its influence on the network design), leaving the reader to question the results and detracting from the integrity of the claim that this model was readily repeatable.

- Assessing the existing network for needed upgrades: Castalia used a load forecast and existing network loading data to conclude that "adequate capacity will be generally available to support the new connections during the initial years of the access programme", but the assertion was vague and unsupported by numerical details. No load flow was performed. A cost estimate for future upgrades was provided based on expected load growth, although no methodology was described in support of this either, and 'load growth' was not defined (i.e. Did it include fill-in, new connections, and increase in average consumer demand?). Sofreco, on the other hand, did perform load flow analyses in order to identify upgrades required by the existing network, and to evaluate possible electrical infrastructure buildout solutions. Detailed load flow reports were provided for each planning region.

- Identification of Planning Stages: In the case of Sofreco, the Planning Director of the EARP design group dictated the order of planning zone priority, with those regions having received the fewest connections to that point taking top priority. From there, Sofreco created lots in each region, each including one or more transformer zones, for bidding purposes and ranked them 1 through 5 (according to somewhat subjective criteria), with these values corresponding to the year within the five-year planning period during which they were to be constructed. In the Castalia report, priority for electrical connection via grid extension was given to those consumers located within five kilometers of the existing network and which inhabited districts considered as having the greatest level of affordability. Social institutions were also included, even if located in areas not meeting these criteria. Once enough cells were selected to meet the 35% electrification target, they were assigned to one of each of the five years of the planning timeframe, with those closest to the existing network and closest to large population centers prioritized for the first years.

Study Replicability

It is not apparent that Castalia made the technology selection spreadsheet it used available to any particular Rwandan planning group for later use, making the results of this study static and difficult to replicate. Sofreco left both the spreadsheet model it used for demand and cost forecasting as well as the CYME software load flow files to the EARP planning group for to update and reuse as necessary. However, it is important to note that off-grid technologies were not considered in the Sofreco planning process so no tool was provided which can assist in making decisions involving off-grid generation options.

Usefulness of Report Results

The intention of the Castalia report was to elicit funding by instilling confidence in donors with what was intended to be a thorough analysis. But in an attempt to show that all possible costs had been considered, the report seems to have succeeded in providing a great deal of unsupported, static values, many of which have little practical use (for instance, Table 4.8 in the Technical Annex provides expected costs for each year from 2009 to 2013 for such things as "Lowering Programme Cost" and creating "Joint sector performance reports"). And this seems to have been done at the expense of providing procedural details for the planning process, which may have ultimately been of much greater use to Rwandan planners. Sofreco's study, on the other hand, left a substantial legacy of planning details, which were seemingly of considerably greater use from a planner's perspective, with specific upgrade recommendations, designs based on load flow reports, and exact locations for placement of infrastructure.

4.1.3 Evolution of the Rwandan Energy Sector Strategy

The Electrification goal as of the original Vision 2020 publication (ca. 2010) was 35% by 2020. The political agenda to increase the national electrification rate went from a relatively modest goal of adding 250,000 connections over the course of five years (from 100,000 in 2008 to 350,000 in 2012), when the Castalia plan was being drafted, to the more ambitious goal of achieving an electrification rate of either 70% (~1.5 million connections) or even 100% by 2017[28], by the time the Sofreco report was published. EDPRS II, published in 2013, stated that "approximately 48% of the total population will be within feasible range of the grid" and went on to specify that EARP "will target around 45% of households with direct connections by 2017, while for the remainder of the households off-grid solutions represent an attractive cost-effective option". The Energy Sector Strategic Plan 2013/14 – 2017/18, published in 2015, set as one of its 'high-level target objectives' to increase household electricity access by 70%: 48% grid connection, 22% off-grid.

Then in 2016, the Rural Electrification Strategy (RES) was published by MININFRA, recognizing that the consumption levels seen with the EARP connections which had been made by that point (totaling 226,000 households, representing an increase in the national electrification rate from just under 15% in 2012 to approximately 24% by June 2016) were too low to ensure cost recovery for the utility. Almost half of those connected under EARP were using less than 20 kWh per month when, according to the ESSP published in the same year, cost recovery was only possible if each consumer used no less than approximately 130 kWh per month. The report deduced from this that off-grid technologies needed to play a bigger role, pointing out that "Two emerging technologies, standalone solar systems and mini-grids, have recently started to be delivered at large scale across Africa, and there is significant scope for these technologies to be used in Rwanda to rapidly scale up electricity access." The document proposes a change in the working definition of energy access, suggesting that instead of continuing to measure energy access as the percentage of grid-connected households, the levels of access defined in the SE4All

---

[28] The national electrification rate goal appears to change depending on the document consulted; see the discussion from the Sofreco report *Context* section of this report for further details.

Multitier Framework should be incorporated into the national lexicon. These access levels can be seen in Table 19 below.

Table 19: SE4All Consumption Tiers (source: [32])

Level	Tier 1	Tier 2	Tier 3	Tier 4	Tier 5
Energy usage	Household lighting, radio and phone charging	Household lighting, radio, phone charging and basic appliances (TV or fan)	Tier 2 plus medium appliances such as low power refrigeration	Tier 3 plus high power appliances such as pumping	24/7 high power suited to commercial and industrial uses

It is suggested in the RES that the ESSP target of electrifying 48% of households through grid-connection by 2017/18 be revised to include any source (either on-grid or off-grid) which can provide access levels corresponding to the SE4All Tier 2 or higher. To put the descriptions provided in Table 19 in perspective, below is a figure showing monthly consumption of existing EARP consumers (as of 2013) by tier, including a suggested alternative form of energy supply for each tier.

4.2 Conclusions: The Role of a Techno-economic Electrification Planning Tool

Discussion in this section will be centered around reflections on the capabilities of the Reference Electrification Model in light of the planning documents reviewed. The critique has been split up into three subsections, the first of which is *REM Limitations and Areas for Further Development*. Under this heading, both the current avoidable limitations and intrinsic limitations of REM are considered; where the limitations are avoidable, consideration is given as to what measures might be required for improvement to be achieved. The second category is *Notes on the Use of REM*, where thought is given to strategies for using REM to accommodate an interest on the part of the user to replicate certain methods used in the reviewed planning exercises. The final subsection is entitled *Advantages of Using REM* and explores opportunities for improvement identified in the reviewed master plans and the ways in which REM may be of assistance in addressing these in future planning efforts.

4.2.1 REM Limitations and Areas for Further Development

- Recognition and Accommodation of Existing Infrastructure: REM can design using both LV and MV conductors but can only accept a single voltage value for all existing network inputs, therefore it does not discriminate between conductor sizes when identifying tapoff locations. It also does not recognize existing substations or transformers when considering potential tapoff locations[29]. In addition, the final system design assumes that a substation is located at each tapoff location so does not include the costs associated with constructing or maintaining these substations. The user does have the option of entering a tapoff 'connection fee' in the RNM component catalog, the value of which is added to the overall MV network cost each time a new grid extension connection is made. However, this same fee is added for each grid connection, regardless of the load to be served by that particular tapoff location, and irrespective of the existence, or lack thereof, of a substation at that location.

---

[29] Such infrastructure may be represented in REM by creating a small length of conductor with the geographic coordinates of the substation, although the functional value of this 'trick' is quite limited. REM currently considers any location on an existing electrical conductor to be a potential tapoff location, so there is no guarantee that it will select this disguised substation over any other point on a neighboring line.

- Planning for Public Infrastructure: REM demand doesn't currently have a default public/municipal 'customer' profile to account for community-level demand such as public lighting, something which is factored into the expected demand in the Rwanda Gicumbi District planning report [43]. As customer demand profiles are customizable, a user wishing to represent public lighting, for instance, could create a demand profile for a single street lamp and associate this demand profile to a consumer coordinate for every street light planned for a specific community. Alternatively, all municipal lighting needs could be combined into a single customer demand profile and a consumer point representing that load could be placed in a central community location. However, while both of these options would be sufficient to represent the demand of municipal lighting, they both ignore the need for the design of a dedicated network by which all of these lights may be controlled independent of the rest of the local grid. An additional consideration is that the results file of REM includes system cost values given on a per-customer basis, including net present value (NPV), annual investment and operation cost, annual administrative cost, and annual non-served energy cost. It may be of interest to users who want to represent municipal lighting or other communal uses of power to have this component of the per-consumer costs listed separately for accounting and/or decision-making purposes. This may be a cost that local governments would be interested in collecting through taxes, as opposed to having it considered as part of the per customer cost for household connection.

- Generation Capacity: REM currently assumes infinite generation capacity, or in other words, that sufficient generation capacity will be available to support any plan developed. The reader may have noted that the Castalia study cites the national goal of achieving 1000MW of installed generation capacity by 2017 and the load forecast on which its network design plan was based was one exceeding 400MW by 2017. However, as was indicated in the introduction to the Rwandan context in this report, the current generation capacity in Rwanda remains significantly below this value, at 190MW. It might therefore be prudent to create an input category in REM whereby the user has the option to specify a generation capacity limit so that any network extension design created will be capped at serving only as much demand as can reasonably be supplied.

- Off-grid Technology Options: Microhydro is not currently a technology considered in off-grid design in REM, but it is one which appeared in both the Castalia and Sofreco reports, indicating that hydro resource potential in Rwanda is great. Castalia considered potential microhydro sites under 1 MW (or 750kW, there were conflicting references) as candidate generation options for microgrid. As for Sofreco, even though off-grid options were not considered in its planning process, Sofreco did mention in its report introduction that "There is a plan to fast track the roll out of ±85 mini and micro hydro plants." So having this available as a generation option in REM would be desirable in locations with resource profiles such as that of Rwanda. However, just as consideration is given in REM to locational solar resources in the form of expected solar production per kW of installed capacity, so too would it be necessary to obtain hydrology resource data and configure REM to be capable of interpreting this information appropriately. As was done in the Castalia report, if a study was performed and hydro resources identified, the potential generation resources could be input into REM with geographical coordinates and the off-grid design module of REM could be configured to consider these on a conditional basis, such as if a cluster of consumers is located within a certain maximum distance from the potential resource.

- True Cost of Diesel: The Castalia report had a methodology for dealing with diesel cost which took into account transportation cost, a feature which REM used to have and which was removed for technical reasons, not theoretical ones, so should be implemented anew. Castalia ranked each cell as one of three diesel cost zones, based on its distance from Kigali, adding 5% for those between 25km and 85km from Kigali and 10% for those between 85km and 145km. The diesel cost function which was formerly a part of REM adjusted for transportation costs by using Open

Street Map to calculate distances from population hubs to rural destinations along with estimates of transportation cost per mile (fuel cost) to estimate diesel cost in any given location.

- Consumer Grouping: In principle, one of the primary benefits that the REM software has to offer is that it has a greater level of granularity than similar tools and other planning methods. As of yet, it is the only planning tool of its kind that offers least-cost designs based on individual consumer-level demand and location. However, it is possible that this level of granularity may not be desired in all situations. For instance, under certain conditions, starting the clustering heuristic decision-making process with individual consumers as the cluster seeds can lead to grouping and technology assignment decisions that may appear overly contrived. It is one thing to assume that household–level analyses are not typically made because they are prohibitively time-consuming in the absence of an automated evaluation tool, but it is also quite possible that at least for some purposes, a higher–level analysis might be preferred.

  I believe the protoclustering mechanism in REM may provide such optionality in this instance. Protoclustering is an option whereby the user is able to elect that consumers be pregrouped based on regional boundaries, as done in the Castalia study, or according to a maximum user-specified customer-to-customer distance, which would effectively lead to the identification of population concentrations, as done in the Sofreco study. While this is already an option in REM, it is only currently integrated into one of the two clustering versions available for use, the version referred to as simultaneous clustering. Further testing needs to be done in order to determine if these two clustering versions, simultaneous and sequential, are distinct enough in their respective capabilities that keeping both of them as options for use with REM is warranted. But until such testing is done and a decision reached, it might be prudent to adopt the pre-clustering option for use with the sequential clustering algorithm as well, so that the user is not constrained to using the simultaneous clustering version only to have access to the protoclustering option.

- Network Reinforcements: REM is currently only capable of considering the upstream reinforcement costs which might be incurred by the network additions it designs via inclusion of an estimated average reinforcement cost adder to the volumetric grid energy retail cost. It is not, however, capable of identifying the actual network reinforcements which would be required. For this, the network additions designed by REM, along with the existing network, would need to be analyzed using a separate load flow software, as done by Sofreco. This is not the ideal order of operations, however, considering that if REM had knowledge during the technology selection process of upstream reinforcement costs which would be incurred as a result of the grid extension designs it is considering, the technology selection decisions it makes might differ from what they would have been otherwise, in the absence of this locationally-specific cost adder information. Thus, a priori, it would be an improvement to REM if the load flow analysis could be performed in concert with the technology selection and grid design process. One possible way of achieving this would be by harnessing the capabilities of the Reference Network Model (RNM) module used in REM for creating all final network designs. The brownfield version of the RNM module is capable of performing load flow assessments to assess infrastructure buildout options given an existing network, as was performed by Sofreco using the CYME software. However, REM is currently only configured to use the greenfield version of RNM which is not equipped with this capability.

- Result Files: The intention of Castalia was to provide a comprehensive breakdown of each and every type of cost which might be incurred in all domains of development and operation throughout the national energy sector. Many of the cost estimates provided are outside the realm of what REM is intended to provide, such as those for line items such as operational costs of generation facilities, engineer and administrator salaries, and even performance report drafting and lightbulbs. That said, it would certainly not be a futile exercise to solicit feedback from Rwandan planning groups as to whether or not there are additional pieces of cost or statistical

information that REM is does not currently output as a results file but could reasonably be configured to do so. That said, one thing that would be advantageous to add to the headings of the results file issued by REM are the units for all outputs. This is a somewhat minor point in light of the others that have been made, but it is nevertheless still an attribute that is lacking and one which would greatly facilitate interpretation of the results file.

4.2.2   Notes on the Use of REM

- Road Recognition: While the RNM module is capable of identifying roads in densely populated areas based on the settlement patterns of inhabitants, it has been shown that this capability is not well-adapted to more sparsely populated areas or settlement patterns of long and narrow and or somewhat irregular compared to more traditional European housing configurations. Thus, a roads input file should be provided[30] to RNM in instances where consideration of roads with respect to new electrical infrastructure layout is considered important.

- Resettlement Areas: REM is a highly consumer-centric planning software meaning that in order for it to consider an area for electrification, there must be consumers present in that area. In instances where investment in electrical infrastructure is expected to precede the existence of consumers, as in the case of Imidugudu resettlement areas in Rwanda, one of several options will need to be selected in order to accommodate this situation. One option is to create additional consumer points which represent the anticipated future demand. Alternatively, the user may choose to post-process the results in order to account for the additional consumer fill-in he or she expects to see in a given area. If the latter option is selected, the user is encouraged to make sure full consideration is given to the effects the post-processing additions may have on the individual system cost results.

- Database Maintenance: It was emphasized in both the Castalia and Sofreco reports that the database each was given was insufficient to effectively carry out the planning exercises they were asked to perform. When attempting to create and routinely update a cost-optimized plan for a large area, especially one subject to as many financial, political, and social uncertainties as tend to be at play in the developing country context, it is of the utmost importance that a comprehensive database be developed and kept up-to-date. Information used by one or both of the plans reviewed and by REM include the following: locations and basic technical details of existing network components including distribution and transmission-level conductors, substations, and transformers, consumer location and type (e.g. household, school, health facility, administrative office, market, cell phone tower, water pumping station), feeder load profiles, and GIS files of all zones considered sensitive for the purposes of transmission distribution route planning and off-grid generation siting (e.g. lakes, rivers, wetlands, national parks). It is important to ensure that capacity data is included with each and every electrical component entered and that careful attention be paid to the coordinate systems used for shapefile creation and manipulation. The GIS shapefiles provided to MIT for analysis appear to have been obtained from various sources and not only contained widely varying levels of component specifications but were also projected into inconsistent coordinate zones. Files created by Sofreco proved to be particularly difficult as they were not in a conventional coordinate system but had been assigned a custom projection, presumably created by Sofreco, which was created in order to accommodate the fact that Rwanda straddles two UTM (universal transverse Mercator) zones. However, this unconventional modification was neither documented nor referenced in any of the Sofreco reports, which can

---

[30] As of the publication of this thesis, REM has not been configured to provide the filepath information to RNM which is necessary to enable this functionality. It is anticipated, however, that this could be accomplished without much difficulty.

lead to an undue and burdensome amount of user confusion, especially when combined with files of dissimilar coordinate systems.

- Capacity Building: The Sofreco report indicates on several occasions throughout the report that it is critical that the planning process be a dynamic one. To this end, the deliverables included along with the final combined and regional reports were the spreadsheet model and the CYME load flow software model files. It would undoubtedly be informative for the REM development team to have a frank conversation with the planning entities in Rwanda with whom it is currently in contact regarding the extent to which the model files left behind by Sofreco were put to use as intended. If they have not used them, why not? If they have, do they have any feedback that may help the REM team to better cater the model to their planning process? If their efforts were stymied by institutional issues instead of technical ones, understanding this history would also be informative.

  It has been noted by the REM team through experience interacting with both host country planners and development agencies such as the world bank, that capacity building is a critical need in the context of most developing countries. It has proven unrealistic to expect that making available a model, tool, or methodology of any sort will prove tantamount to encouraging progress in the absence of thorough and continued efforts to understand the institutional context and facilitate planning. So one might assume that if modeling instruments left behind by Sofreco ultimately went unused, it may have been in no minor part due a deficit in the institutional capacity required to use them. But whatever the post-delivery experience was with the Sofreco models, it was likely influenced by a confluence of complex institutional, political, and social factors that must truly be understood through on the ground partnerships and sustained efforts if what can sometimes seem like a pattern of white paper futility is not to be repeated and true progress is to be made.

- Anchor Loads: It was indicated in the 2016 Rural Electrification Strategy that sites with anchor loads such as cell phone towers or water pumping stations would be prioritized in the selection of microgrid candidate sites. It is possible in REM to define multiple customer types, each with its own demand profile. If one wanted anchor loads to be considered in the REM least-cost generation technology evaluation process, a customer profile could be created for each separate type of anchor load and the locations of these loads could be included in the customer input file.

- Consumer Ability to Pay: Consumer affordability was considered in the Castalia planning process. Consumer ability to pay or willingness to pay is not explicitly factored into the decision-making process of REM but could be used as a post-processing decision-making criteria in a similar fashion as Castalia.

4.2.3 Advantages of Using REM

- Repeatability: Failing to provide an updateable model or even explicit details as to the exact nature of assumed values, calculations used, and procedures followed results in a static study which quickly becomes outdated, especially if execution of plan implementation is delayed, as it is unable to adapt to changing circumstances. A case in point is the technology prices on which the least-cost generation source selections were predicated in the Castalia plan. The values for solar panels and batteries used in the Castalia report as follows:

72

Table 20: Castalia Solar PV Cost Assumptions (source: [44])

General Data and Common Assumptions - SolarPV							
Assumed Lifespan - Panel		years	20				
Assumed Lifespan - Battery		years	3				
Assumed Lifespan - Other equipment		years	10				
SolarPV costing assumptions			**SolarPVunits**				
	SolarPV Code	SPV1	SPV0.4	SPV0.2	SPV0.1	SPV0.05
	Panel size [kVA]	1	0.4	0.2	0.1	0.05
Cost of Panel	US$	7,000	3,500	3,000	2,350	1,750
Cost of Battery	US$	3,500	1,750	880	600	180
Cost of Other Equipment (regulators, lamps, accessories	US$	3,500	1,750	880	600	180
Capital investment cost						
Initial Solar PV installed cost	US$	14,000	7,000	4,760	3,550	2,110
Battery Replacement (every 3 years)	US$	3,500	1,750	880	600	180
Annual recurrent MCH Maint costs	%	5.0%	5.0%	5.0%	5.0%	5.0%
Load Lost Through Intermitency	%	15%	15%	15%	15%	15%

It can be seen in Table 20 that a 200W solar panel was valued at $3,000 at the time the Castalia report was published in 2009. The 250W panel used in the REM catalog is currently valued at $125. Batteries have undergone a similar market transformation, with notable decreases in cost over the last decade coupled with increased longevity. Costs are difficult to compare in this instance as battery specifications are not provided in the table of assumptions.

- Technology Selection: As discussed earlier in this report, the national electrification strategy in Rwanda has shifted since the Sofreco plan was developed toward greater inclusion of off-grid technologies including microgrids and standalone solar home systems. However, as pointed out in the Rural Electrification Strategy published by MININFRA in 2016, microgrids are "only viable in very specific circumstances"[31]. Although modeling deliverables used in the Sofreco study were intended to serve as tools allowing for dynamic updating of the plans as the electrification effort moved forward, the Sofreco planning methodology was not created to consider economic tradeoffs between grid extension and off-grid technology options. REM, however, is equipped with this capability. In addition, if a user does wish to consider a scenario that does not include off-grid options, as done in the Sofreco report, he may either force 100% grid extension in REM by assigning artificially elevated cost values to the off-grid components, or use instead the RNM module of REM directly, bypassing the consumer clustering option altogether. Thus, REM is effectively capable of allowing the user to consider both strictly least-cost scenarios, where tradeoffs between grid extension and off-grid technologies are evaluated, and scenarios where all connections are made through grid extension.

- Consumer Location: There are two ways to identify customer locations: visit each site with a GPS device and record manually or identify the coordinates via digital images. For the purposes of performing this identification across an entire country, the latter is clearly the preferable approach as it reduces both the time and the manpower required to accomplish the effort. But identifying buildings via digital images is nevertheless a relatively time-consuming task in and of itself, it the identification must be performed by hand, as done by Sofreco. What REM can offer as an

---

[31] It should be noted that while this may be the case in Rwanda, in countries where distribution companies are financially strapped, microgrids may be the only way to make progress in electrification efforts in the near to medium term.

improvement to the process is a household identification software which can greatly expedite the activity through automation.

- Consumer Clustering: A priori, if the user is seeking to identify population clusters of any sort (such as the transformer zones identified by hand in the process used by Sofreco), an automated system for doing this would be preferred. REM can perform this task with a more thorough assessment (trying various MV/LV options, clustering permutations, and technology configurations) than is typically reasonable or even possible to do by hand, and it can do it in less time. And, as discussed previously, the user has several options for dictating the way clusters will be formed, using either individual consumers as clustering seeds or pre-defining maximum customer to customer distance or boundaries, such as village limits or another other type of administrative boundary.

- Distribution Line Routing: All LV/MV line routing, in instances where it was performed in these studies, was by all evidence done by hand and REM offers an automated alternative which recognizes topographical features such as mountains, wetlands, national parks, lakes, and rivers, and allows for the assignment of penalties to those features and any other zones the user wishes to avoid traversing with electrical infrastructure.

- Accounting for Assumptions: One potential advantage of using REM is that with all network design steps contained within a single platform, there is no question as to what assumptions were made, where these values were used, and in what order tasks were performed. The process of reconstructing the order of operations followed by these consultants using only the reports and supporting documents as guides, was a surprisingly non-obvious task. The order of in which tasks were performed in the plan development process was often obscured by items being discussed out of chronological order and the appearance in of conflicting descriptions of the same task placed in different report or supporting document sections. Assumptions were similarly difficult to track and in more than one case, appear to not have been documented at all. Having an automated planning process for which all input files are in a standard format with value definitions thoroughly documented in a user's manual would go a long way in alleviating any confusion when it comes to questions pertaining to the process followed or the need arises to recreate it.

- Accuracy of Per Customer Costs: REM would allow for Rwandan planners and policymakers to determine for themselves what the per customer costs would be for off-grid systems based on up-to-date technology costs instead of relying on a figure from a static study provided by a third party, as in the case of the average microgrid cost per connection estimate used in the 2016 Rural Electrification Strategy report.

- Facilitated Scenario Testing: Based on the findings in the 2016 RES report that many of the customers connected to the grid during the EARP grid rollout had consumptions too low to ensure cost recovery for the utility, one could infer that the load forecast developed by the Sofreco report was overly optimistic, or overly generalized. REM may be able to help planners avoid learning this only after construction efforts have been underway for several years by facilitating scenario testing. With the option of entering multiple customer types and varying the hourly demand profiles of certain segments of the population, assumptions of consumer willingness to pay can be tested, allowing the user to select the grid extension scenario that provides the greatest level of confidence that cost recovery will be possible.

- Grid Reliability: The plans developed in both the Castalia and Sofreco reports assumed implicitly that grid supply would be 100% available. It may be that this is of less concern in Rwanda than it is in other countries with persistent blackouts, such as India, but the REM group has found it to be the case that microgrids may sometimes be better suited to meeting the demand profile of certain consumers if the segment of the network to which they are or would be connected has a less than

74

desirable history of consistency of service. Based on the priority uses of the funding available and the timeline of generation facility build completion, (several of the documents consulted in the drafting of this report referred to the possibility of delays in the introduction of new utility-scale generation sources) it may prove of interest at some point to planners that REM allows the user to specify hourly reliability values on a per-feeder basis.

4.3   In Summary

Based on this assessment, it would seem that, globally, REM is well-positioned to meet existing planning needs, while providing the added advantages of speed and thorough least-cost analysis, and remaining flexible enough to adapt to the widely varying conditions of the developing world context.
It is nevertheless important to recognize that certain potential areas of improvement have been identified that might help bridge the gap between the way planning has been done previously and adoption of the REM tool, increasing the relative usefulness of the tool as a standalone modeling software in such a way as to better its chances of large scale adoption. Possible improvements include: more accurate accommodation of existing infrastructure, the addition of methodologically rigorous consideration of network upgrade implications, expansion of off-grid technology options, and application of a utility-scale generation limit.

5 References

[1] D. Ellman, "The Reference Electrification Model: A Computer Model for Planning Rural Electricity Access," MIT, Boston, 2015.

[2] Y. Borofsky, "Towards a Transdisciplinary Approach to Rural Electrification Planning for Universal Access in India," MIT, Boston, 2015.

[3] P. Levi, "Feasibility of Grid Compatible Microgrids," MIT, Boston, 2016.

[4] V. Li, "The Reference Electrification Model: A Comprehensive Decision-Making Tool for the Design of Rural Microgrids," MIT, Boston, 2016.

[5] T. Cotterman, "Enhanced Techniques to Plan Rural Electrical Networks Using the Reference Electrification Model," MIT, Boston, 2017.

[6] P. C. Cutillas, "Clustering-related improvements in the Reference Electrification Model," Madrid, 2016.

[7] International Energy Agency, "Energy Access Outlook 2017: From Poverty to Prosperity," OECD/IEA, 2017.

[8] WBG,ESMAP,IEA, "SE4ALL Progress Toward Sustainable Energy 2015: Global Tracking Framework Report.," World Bank Group, 2015.

[9] Massachusetts Institute of Technology News Office, "Rwandan President Paul Kagame Visits MIT," 23 April 2014. [Online]. Available: http://news.mit.edu/2014/rwandan-president-paul-kagame-visits-mit. [Accessed 23 12 2017].

[10] Universidad Pontifica Comillas IIT, "Pilot project for electrification of a village in Rwanda taking the local school as the anchor load," August 2014. [Online]. Available: https://www.iit.comillas.edu/proyectos/mostrar_proyecto.php.en?nombre_abreviado=2014%20UA%20Rwanda.

[11] Sofreco, "Combined and Zonal Design Reports(s)," 2013.

[12] C. M. Domingo, "A Reference Network Model for Large-Scale Distribution Planning with Automatic Street Map Generation," IEEE Transactions on Power Systems, vol.26, No.1, February 2011.

[13] Kemausuor et al, "Electrification planning using Network Planner tool: The case of Ghana," *Energy for Sustainable Development,* 2014.

[14] KTH Royal Institute of Technology et al., "OnSSET: Open Source Spatial Electrification Tool," [Online]. Available: http://www.onsset.org/. [Accessed 11 12 2017].

[15] KTH-dESA, 2017. [Online]. Available: https://github.com/KTH-dESA/PyOnSSET. [Accessed 12 December 2017].

[16] United Nations, "DESA Modelling Tool: Universal Access to Electricity," [Online]. Available: http://un-desa-modelling.github.io/electrification-paths-presentation/. [Accessed 2017].

[17] Mentis et al, "A GIS-based approach for electrification planning - A case study on Nigeria," *Energy for Sustainable Development,* 2015.

[18] Nerini et al, "A cost comparison of technology approaches for improving access to electricity services," *Energy,* 2015.

[19] Mentis et al, "Lighting the World: the first application of an open source, spatial electrification tool (OnSSET) on Sub-Saharan Africa," *Environmental Research Letters,* July 2017.

[20] C. Vergara, "REM_API_20170514," 2107.

[21] C. Vergara, "Rural Electrification Toolbox Documentation," 2017.

[22] Ju Kang and Buyng Lee, "Optimisation of pipeline route in the presence of obstacles based on a least cost path algorithm and laplacian smoothing," *International Journal of Naval Architecture and Ocean Engineering,* 2017.

[23] Monteiro et al, "GIS Spatial Analysis Applied to Electric Line Routing Optimization," *IEEE Transactions on Power Delivery,* vol. 20, no. 02, April 2005.

[24] Petzold et al, "Network Planning using Geomorphology," in *9th ACM International Symposium on Advances in Geographic Information Systems*, Atlanta, Georgia, USA, 2001.

[25] Luchmaya et al, "Using Terrain Information in an Electrification Planning Tool," in *Transmission and Distribution Conference and Exposition, 2001 IEEE/PES*, Atlanta, Georgia, USA, 2001.

[26] Hasan Eroglu and Musa Aydin, "Optimization of electrical power transmission lines' routing using AHP, fuzzy AHP, and GIS," *Turkish Journal of Electrical Engineering & Computer Science,* 2013.

[27] Yildirim et al, "A Raster Based Geospatial Model for Natural Gas Transmission Line Routing - Chapter 5," October 2013. [Online]. Available: https://www.intechopen.com/books/natural-gas-extraction-to-end-use/a-raster-based-geospatial-model-for-natural-gas-transmission-line-routing. [Accessed December 2017].

[28] Ergun et al, "Technology and Topology Optimization for Multizonal Transmission Systems," *IEEE Transactions on Power Systems,* vol. 29, no. 5, 2014.

[29] Pedro Ciller and Carlos Mateo, "Greenfield Reference Network Model (RNM): modelling of investment, energy losses and maintenance costs," 2015.

[30] Wikipedia, "Sectors of Rwanda," [Online]. Available: https://en.wikipedia.org/wiki/Sectors_of_Rwanda. [Accessed November 2017].

[31] USAID, "Rwanda Power Africa Fact Sheet," 05 October 2017. [Online]. Available: https://www.usaid.gov/powerafrica/rwanda. [Accessed 02 December 2017].

[32] Republic of Rwanda, Ministry of Infrastructure, "Rural Electrification Strategy," 2016.

[33] Energy Development Corporation Limited of Rwanda, "Current Status of Power System in Rwanda," EDCL, June 2016.

[34] USAID, "Rwanda Power Africa Fact Sheet," 5 October 2017. [Online]. Available: https://www.usaid.gov/powerafrica/rwanda.

[35] Wikipedia, "Water Supply and Sanitation in Rwanda," [Online]. Available: https://en.wikipedia.org/wiki/Water_supply_and_sanitation_in_Rwanda. [Accessed 02 December 2017].

[36] Castalia Strategic Advisors, "Rwanda Electricity Sector Access Programme: Investment Prospectus," March 2009.

[37] Rwanda Energy Group Limited, "Rwanda Energy Group," [Online]. Available: http://www.reg.rw/index.php/about-us/history. [Accessed 03 December 2017].

[38] Fichtner and Republic of Rwanda EWSA, AEPE Program: Actualisation Study of the Electricity Masterplan, Kigali, May 2011.

[39] African Development Bank Group, "Rwanda Energy Sector Review and Action Plan," 2013.

[40] Republic of Rwanda, Ministry of Infrastructure, "Sustainable Energy for All Action Agenda, Final Draft for Cabinet Approval," Kigali, 2015.

[41] Republic of Rwanda Ministry of Infrastructure, "National Energy Policy and National Energy Strategy 2008-2012," January 2009.

[42] C. Barteczko-Hibbert, "After Diversity Maximum Demand (ADMD) Report," Northern Powergrid Ltd, 2015.

[43] Republic of Rwanda, Northern Province, Gicumbi District, "District Development Plan (2013-2018)," 2012.

[44] Castalia Strategic Advisors, "Rwanda Electricity Sector Access Programme: Technical Annex," March 2009.

[45] D. G. P.C. Maithani, Achieving Universal Energy Access in India: Challenges and the Way Forward, New Delhi: Sage Publications Pvt Ltd, 2015.

[46] Republic of Rwanda, Ministry of Infrastructure, "Towards universal energy access by 2020 in Rwanda," Powerpoint presentation, 2017.

[47] Y. Uwamahoro, "Rural Electrification in Rwanda," Powerpoint presentation, Energy Sector Coordinating Engineer, 2008.

[48] Republic of Rwanda, Ministry of Infrastructure, "Forward Looking Joint Sector Review FY 2014/15," May 2015.

[49] Energy Sector Management Assistance Program (ESMAP), "Rwanda|Extending Access to Energy: Lessons from a Sector-Wide Approach (SWAp)," 2013.

[50] EuropeAid, "Technical Assistance FAcility for the Sustainable Energy for All Initiative West and Central Africa: Rwanda," June 2015.

[51] Dutch Ministry of Foreign Affairs, "Access to Energy in Rwanda: Impact evaluation of activities supported by the Dutch Promoting Renewable Energy Programme.," August 2014.

[52] Innovation Energie Devéloppement (IED), "GeoSim Planning Tool," [Online]. Available: http://www.geosim.fr/index.php?page=home. [Accessed 2017].

[53] KTH Department of Energy Technology, "OpeN Source Spatial Electrification Toolkit," [Online]. Available: https://www.kth.se/en/itm/inst/energiteknik/forskning/desa/projects/open-source-spatial-electrification-toolkit-onsset.com. [Accessed 2017].

[54] Quadracci Sustainable Engineering Lab, University of Columbia Earth Institute and School of Engineering, "Network Planner," [Online]. Available: http://qsel.columbia.edu/network-planner/. [Accessed 2017].

[55] Republic of Rwanda, Ministry of Infrastructure, "Energy Sector Strategic Plan," Kigali, 2012.

[56] Peters et al, "Impact evaluation of Netherlands supported programmes in the area of Energy and Devleopment Cooperation in Rwanda|The provision of grid electricity to households through the Electricity Access Roll-out Programme.," June 2014..

[57] International Energy Agency, "Energy Access Outlook 2017," International Energy Agency, 2017.

[58] F. e. a. Nerini, "The World Energy Outlook 2014, Inputs to the Africa Chapter Working Team. KTH working paper series.," KTH Division of Energy Systems Analysis and The Swedish International Development Agency, 2015.

[59] V. e. a. Modi, "Electrification Planning Using Network Planner Tool: The Case of Ghana," Energy for Sustainable Development, 2014.

[60] Republic of Rwanda, Ministry of Finance and Economic Planning (MINECOFIN), "Economic Development and Poverty Reduction Strategy II 2013-2018," MINECOFIN, May 2013.

[61] Republic of Rwanda, "Economic Development & Poverty Reduction Strategy 2008-2012," September 2007.

[62] World Bank Group, "Scaling Up Access to Electricity: The Case of Rwanda," World Bank Group Live Wire, 2014. [Online]. Available: http://documents.worldbank.org/curated/en/621551468059083947/pdf/88703-REPF-BRI-PUBLIC-Box385194B-ADD-SERIES-Live-wire-knowledge-note-series-LW22-New-a-OKR.pdf. [Accessed 2017].

[63] Republic of Rwanda, Ministry of Infrastructure, "Energy Sector Strategic Plan 2012/2017," Kigali, 2012.

[64] Republic of Rwanda, Rwanda Vision 2020: Revised 2012, 2012.

[65] Republic of Rwanda, Ministry of Infrastructure, "Rwanda Energy Policy," 2015.

[66] Republic of Rwanda, Ministry of Infrastructure, "Energy Sector Strategic Plan 2013/14-2017/18," 2015.

[67] Energypedia, "Rwanda Energy Situation," [Online]. Available: https://energypedia.info/wiki/Rwanda_Energy_Situation. [Accessed 02 December 2017].

6   Appendix I

The document below was originally drafted for use internal use by Universal Access Lab developers and/or any other user with interest in the development of these files. It is for this reason that the content is fairly informal and devoid of figure and table numbers[32]. It has been included here for reference as much of the information contained complements the descriptions provided in Chapter 3 of this thesis and may prove of some use to the reader.
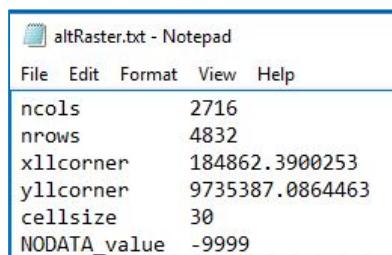
Input File Generation

There are three input files required to utilize all topography-related capabilities: 1) an altitude raster, 2) a penalized zone raster, and 3) a penalized zone polygon file. Files #1 and #2 are used in clustering, files #1 and #3 are used by RNM. Below is a recipe for creating each file. Note that there are many file creation and conversion options in both ArcGIS and Matlab, so there are multiple approaches to raster file creation and manipulation. I am not familiar with QGIS, so am not aware of what the counterpart processes may be in that software, or if they exist. If you find that there is an alternative to any of the steps below which is superior in terms of processing time or otherwise, please feel free to update this document with that knowledge. The raster file creation steps provided below involve the use of manual manipulations in ArcGIS and notepad. Python scripts may be written to automate processes in ArcGIS, which may facilitate file creation if input files have to be created for many regions.

1) Altitude Raster

Please note that while the RNM manual indicates that this file "can be generated by the model automatically or can be given as input", it has been confirmed by the Comillas researchers that the model is not actually capable of generating this file automatically, thus it must be provided as input by the user if altitude values are to be considered.

This file must be in ESRI ASCII Raster format. The header of the file resembles the following:

By default, the altitude values contained within the body of the file are provided in meters, as are the cell size and UTM values (for which the standard unit is meters). REM retrieves this information from the metadata and converts it to km internally for use in the clustering algorithms. Please note that the RNM manual states that the cell size value is required to be in km, and the accompanying example shows 0.25 as the cell size value, which seems to corroborate

---

[32] It should be noted that the references cited within this appendix are listed at the end of this appendix and do not correspond to the primary thesis reference list.
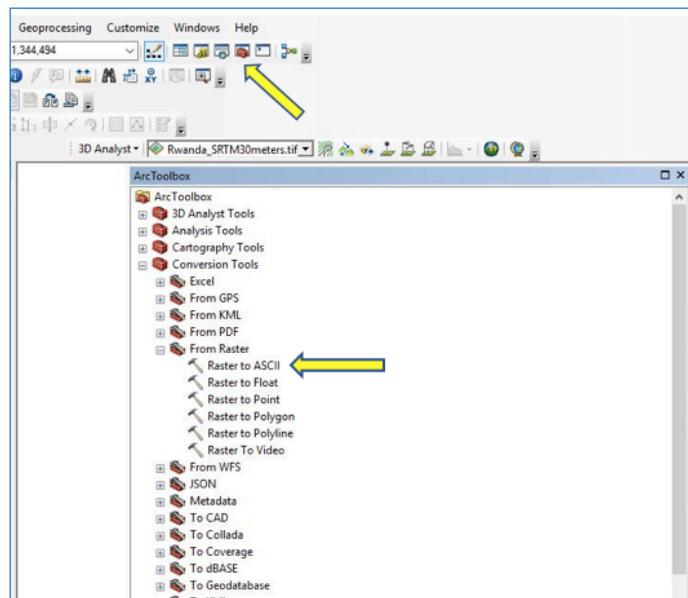
this, but Carlos Mateo has confirmed for me by reviewing the RNM source code that RNM expects this value to be input in meters, <u>NOT km</u>.

All header values appearing in the altitude raster file must match those in the penalized zone raster file, if one is included. The coordinates 'xllcorner' and 'yllcorner' are the x and y-coordinates of the lower left corner of the raster, which is considered to be the origin. The coordinate system in which these values are given is, by default, the coordinate system of the shapefile from which this raster is made. These coordinates need to be in the UTM coordinate system values. As UTM coordinates are denoted in meters, the cell size value will be in meters as well. The data source for creating this file is currently the SRTM (Shuttle Radar Topography Missions) dataset; see [1] for more detailed information. This data is available for almost every region of the world in either 1 arcsecond (approx. 30m at the equator) or 3 arcsecond (approx. 90m at the equator) resolution. This data is available in many formats. The raster for the Kayonza region was obtained from [2] and was downloaded in the .tiff file format. Matlab has a function called 'geotiffread' which could most likely be used to manipulate this data into a form usable for our purposes, but I have not explored this functionality. The .tiff file obtained from [2] contains data with resolution of 1 arc-second and metadata in the wgs1984 geographic coordinate system. Please consider the following note on ArcGIS and coordinate systems provided in the figure below (taken from [3]).
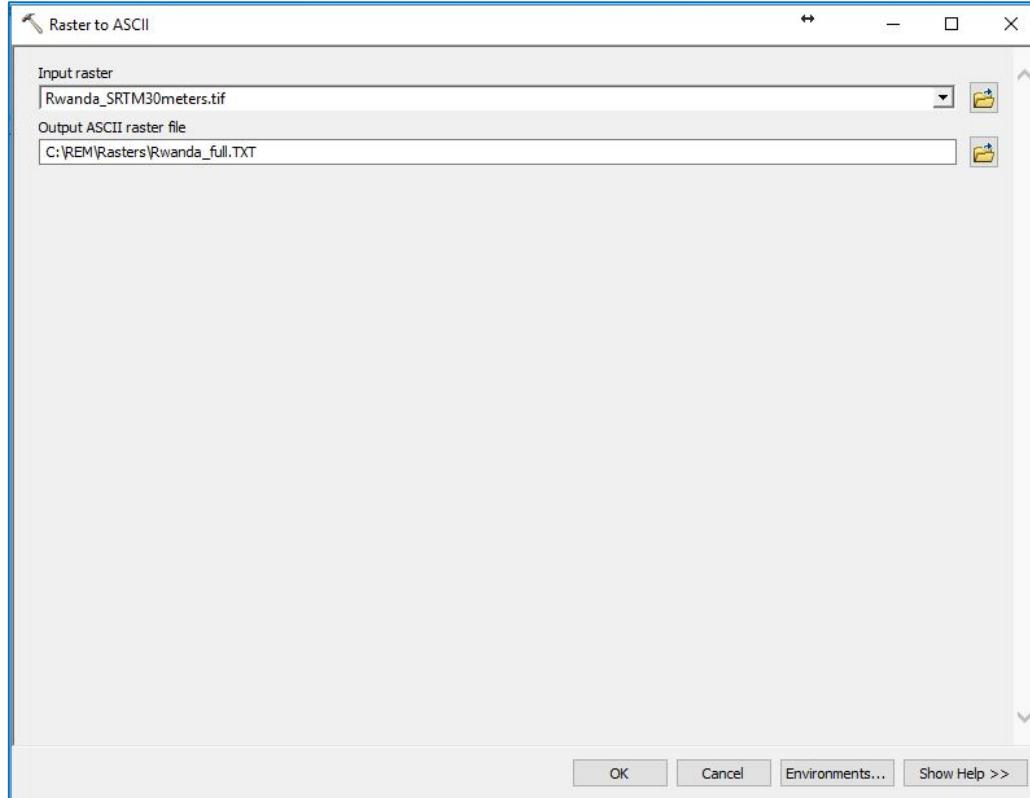
Empty maps and scenes take their coordinate systems from the first layer added to them. When you add additional layers to a map or scene, they are automatically displayed using the same coordinate system of the map or scene. If the map or scene's coordinate system is not the same as the native coordinate system of the layer's source data, the data is projected on the fly using a transformation. Be aware, however, that projecting on the fly can take longer to draw and is not advisable if you are editing data or performing analysis. It is far preferable to ensure that all the data is in the same coordinate system. Use the Project or Project Raster tool to project spatial data from one coordinate system to another.

When converting the .tiff file to ESRI ASCII Raster format, the conversion to UTM coordinates will need to be specified. A method for doing both is outlined below.

1. ArcToolbox > Conversion Tools > Raster to ASCII
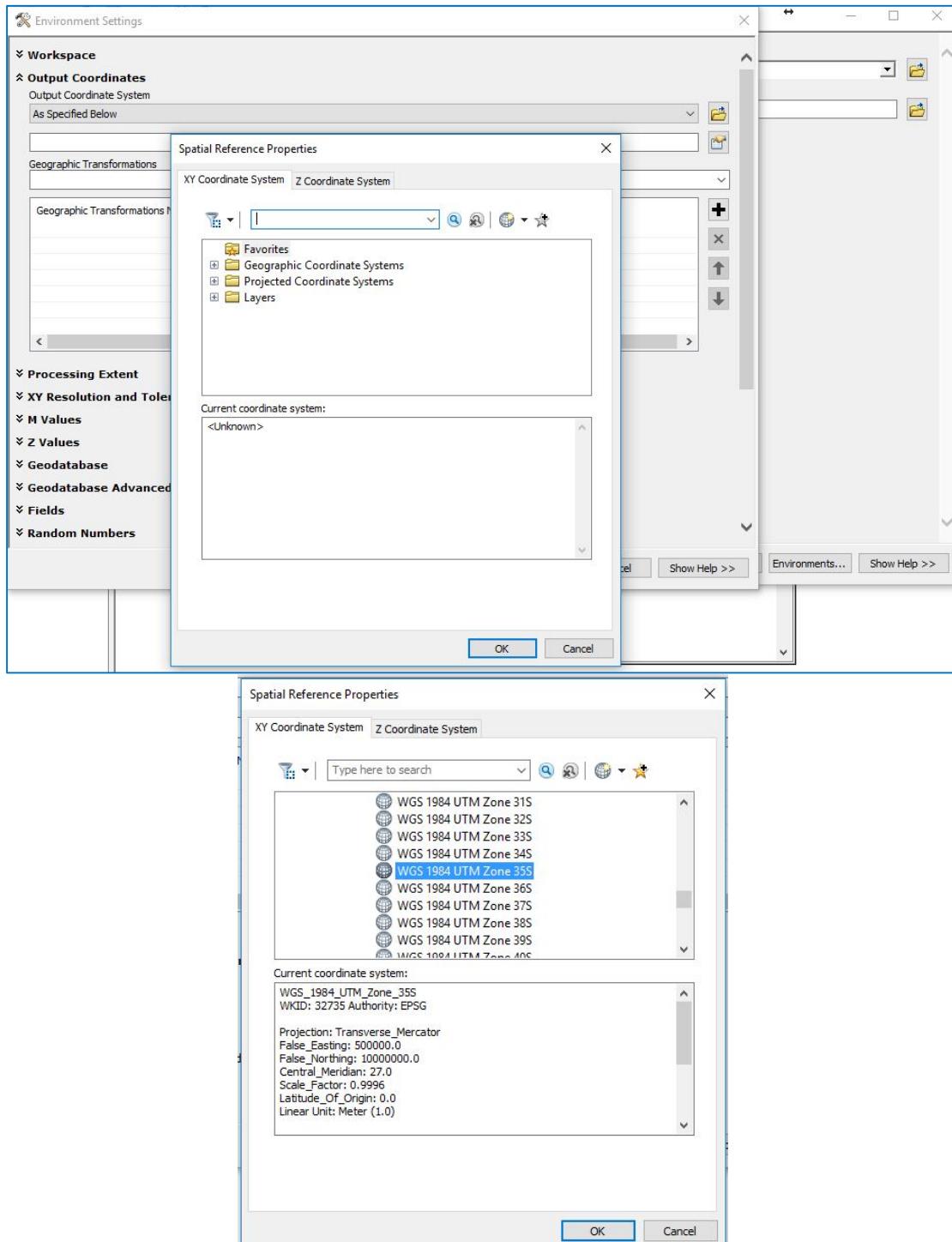
2. Select SRTM file with the .tiff file extension as the input raster (drag and drop or select file through folder icon at right). Assign name and location to output file. Click on the 'Environments' icon at the bottom of the window.
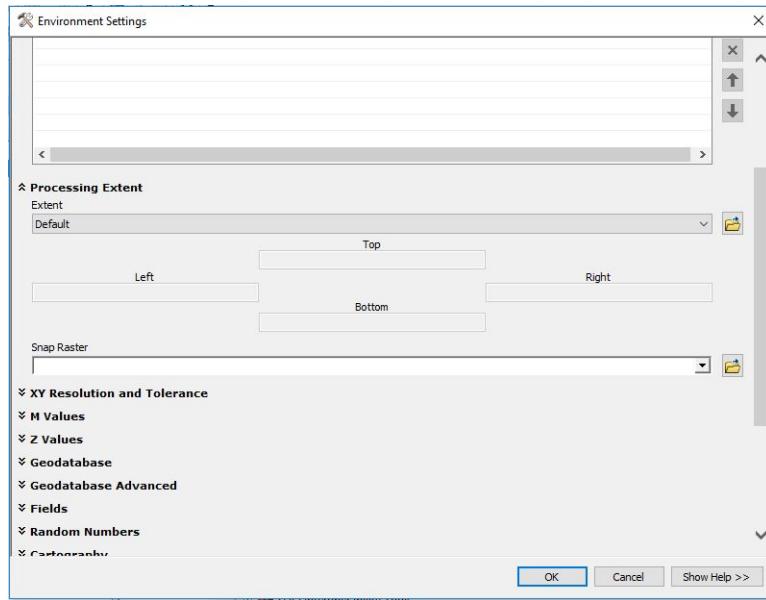

3. Select Output Coordinates
    Enter "As specified below" in the dropdown menu
    Click on the icon to the right of the line below this and select: Projected Coordinate Systems >
    UTM > WGS 1984 > Southern Hemisphere (or Northern, as appropriate) > the UTM zone
    appropriate for your study region
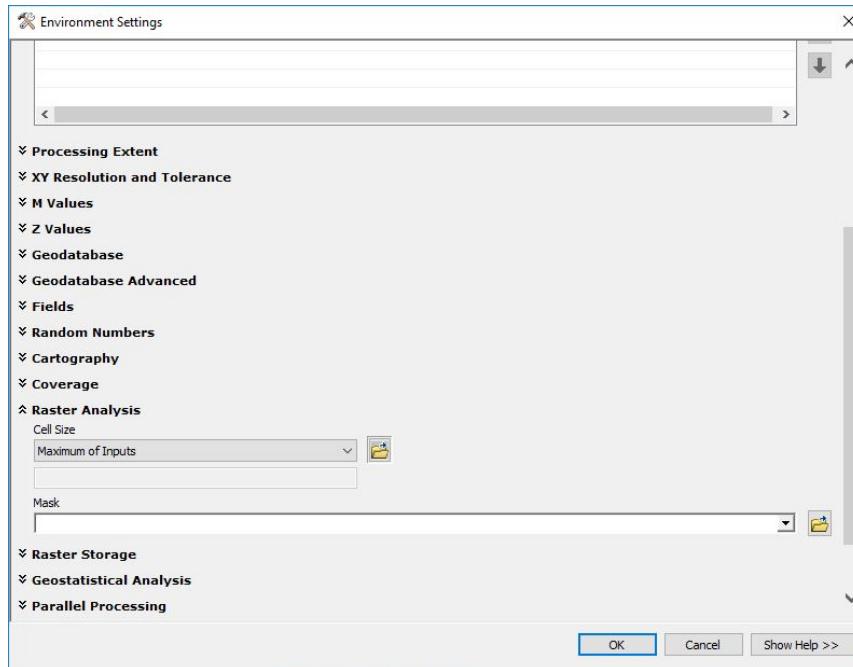
Note that in the Environments Settings window, under the "Output Coordinates" option there are several other dropdown menus which may be of use in dictating the characteristics of the raster output file. One which may be of particular interest is "Processing Extent", which allows you to specify the northern, southern, eastern, and western coordinate limits of the raster you wish to create. This is helpful if you wish to create a raster for a subregion of a country.
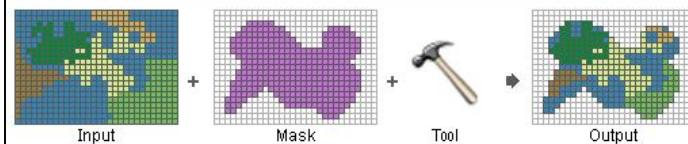
Another which may prove useful is "Raster Analysis", under which are the options to dictate cell size (may be useful when trying to ensure that the cell size values for both the altitude and forbidden zone raster files are exactly the same) and to 'mask'.

The masking functionality may be useful if it becomes necessary to create raster files for individual subregions from larger parent files. See [4] for further details.

Illustration

Input   Mask   Tool   Output
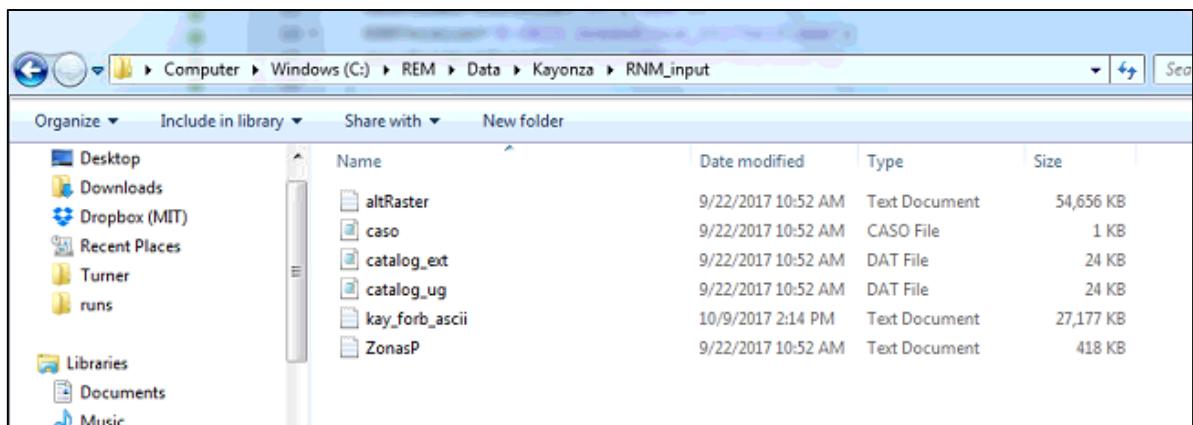
Mask identifies the areas in the analysis extent to be included in the tool execution.

Usage notes

- Setting an analysis mask means that processing will only occur on locations that fall within the mask, and all locations outside of it will be assigned to NoData in the output.

4. Once all elections have been specified, select OK to create the raster file. Once the raster text file has been created, navigate to the folder where you indicated the file was to be saved and open the file using notepad. Use search and replace* to change all of the no data values of -9999 to zero – EXCEPT FOR the value in header. The value assigned to 'NODATA_value' in the header must remain as -9999 (or at least any number which does not match a value found within the body of the file). Place the file in the RNM_input file that your case will be calling. The file must be named 'altRaster.txt'.

   *When performing a find and replace operation of this magnitude, I have found that Notepad++ will crash consistently and excel will often freeze up as well. Notepad works quickly and reliably without freezing or crashing. So please be congnisant of the text editor you are using when processing these files. If you have trouble, you may need to try another editor.
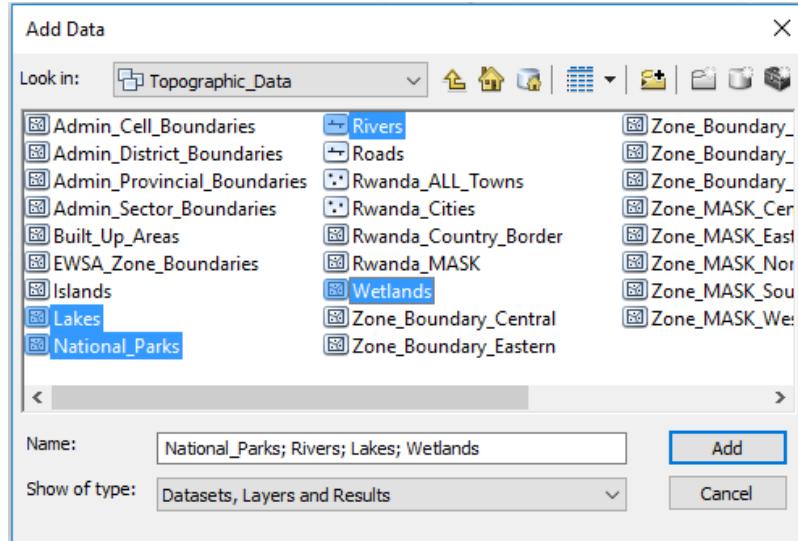
2) **Penalized Zone Raster**
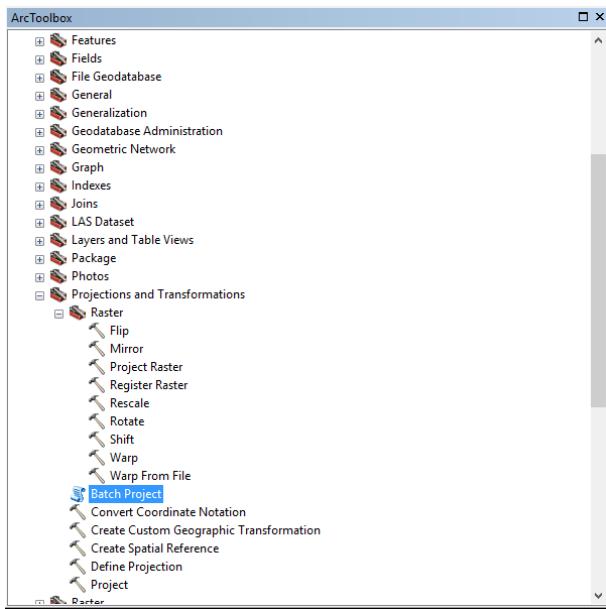
   This file must be in ESRI ASCII Raster format and follow the same rules for the header values as outlined in the Altitude Raster section. The no data values of -9999 in the body of the text file must also be replaced with zeros (except for the value in the header). Please refer to the previous section for details. Below is an outline of the steps that are unique to generation of the penalized zone raster.

The goal is to load all of the shapefiles containing zones to which we wish to apply penalties, ensure that all features contained within these files are polygons, combine them into a single shapefile, convert this shapefile to an ascii raster text fill as done in the previous section, and post process the contents of the text file as need be. I will use the data we have for Rwanda as an example.
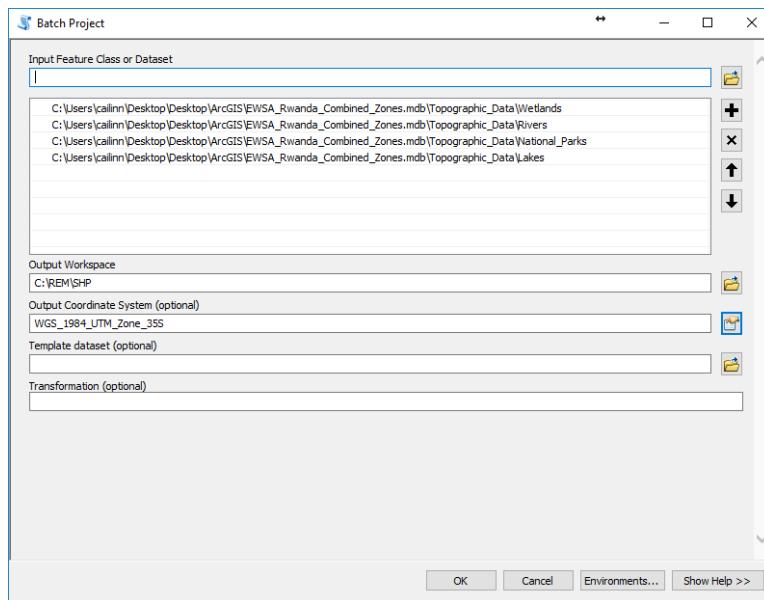
1. Identify the shapefiles containing zones slated to be penalized. In this case, we will choose bodies of water, wetlands, and national parks. It should be noted that this data was provided to us directly but much of it is available online in various data repositories for countries around the globe.

2. If necessary, project the shapefiles into the appropriate coordinate system. Note that these files are already in a projected coordinate system and that reprojecting them into another system will result in some level of compounded loss of fidelity to the original data source. See [5] for notes on good practice when handling digital elevation models (DEMs). Nevertheless, as we have chosen to use UTM zone 35S as our projected coordinate system in this particular instance, it is necessary to convert all four of these shapefiles. For this we can use the batch projection functionality: ArcToolbox>Projections and Transformations>Batch Project.

Identify the files to be projected, indicate the location where the new files are to be saved, and select the preferred output coordinate system. Be aware that the files will be output with the same name, or a filename extension of "_1" if they are saved to the same folder as the originals. For this reason, it is advised that you be sure to either save the new files to a folder with a name that clearly differentiates them from the originals, or rename them after they have been created.

3. Next it is important to verify the feature type of the features contained in each of these shapefiles by looking at the shape field in the attribute table. All features must be polygons, as exemplified by the Lakes attribute table shown below on the left, but NOT by the Rivers attribute table shown on the right. The rivers in this case are represented by polylines.

OBJECTID *	Shape *	ID_	NA
1	Polygon	DNNETR53	Lac Burer
2	Polygon	DNNETR290	Lac Kivun
3	Polygon	DNNETR317	Lac Cyam
4	Polygon	DNNETR320	Lac Nashi
5	Polygon	DNNETR321	Lac Mpan
6	Polygon	DNNETR316	
7	Polygon	DNNETR310	Lac Rwak
8	Polygon	DNNETR285	
9	Polygon	DNNETR284	Lac Hago
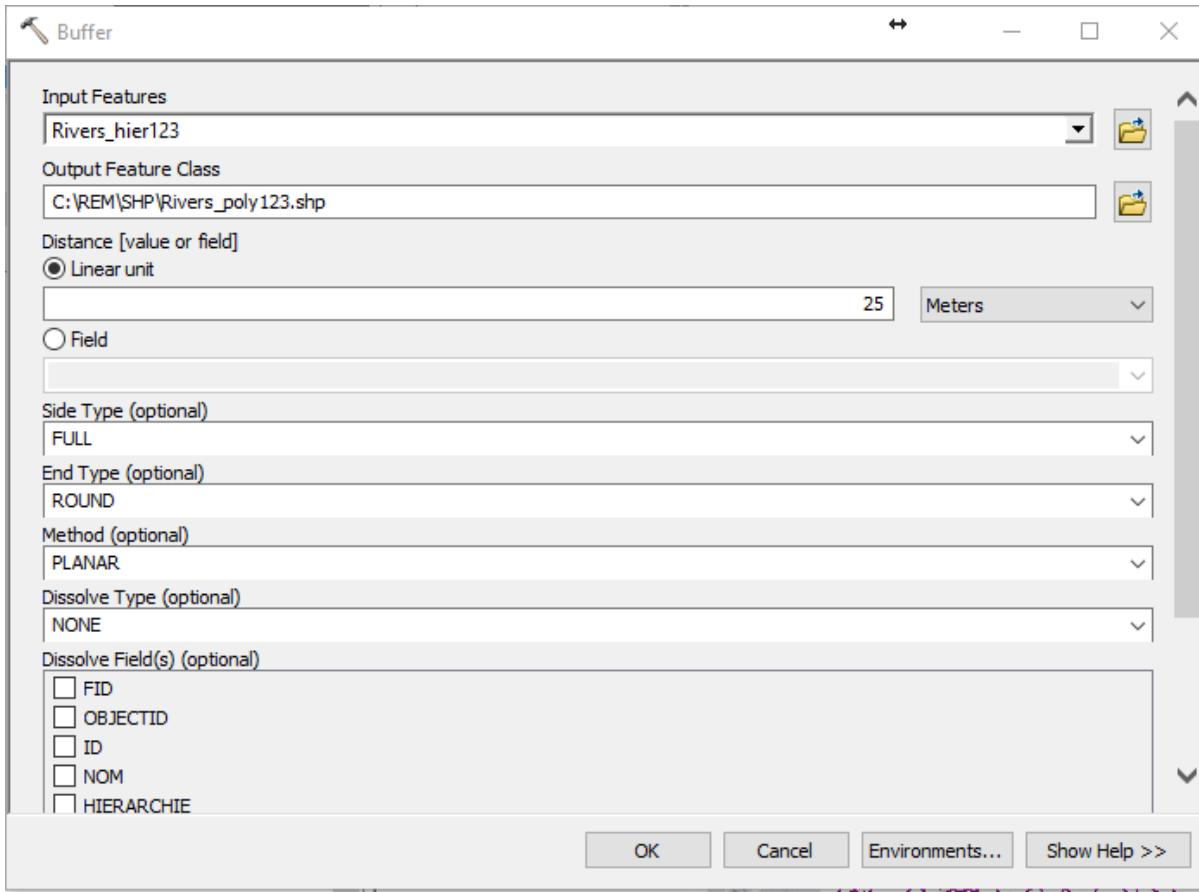
OBJECTID *	Shape *	ID	NOM
1	Polyline	0	Satinsyi
2	Polyline	0	Satinsyi
3	Polyline	0	Mugogwe
4	Polyline	0	Satinsyi
5	Polyline	0	Rutemba
6	Polyline	0	Rutemba

To rectify this, go to Geoprocessing>Buffer, select the Rivers shapefile as the input feature, specify the location where the new shapefile is to be saved, select 'Linear unit' and specify the distance and units of the buffer zone (note this value is measured from the centerline out, so specifying a buffer of 25 meters, as done below, will result in rivers which span 50 meters across). Make sure the dissolve type is set to NONE, otherwise all rivers will be merged into a single feature.

Note that arbitrarily assigning a single width to all rivers is obviously inaccurate and far from ideal, but we can only be as accurate as the data we are given. The attribute table of the river file we have contains a field called 'herarchy' where every object (each object being a section of a river) is assigned a value of 1 through 5. However, I was unsuccessful in finding a definition of these categories in either the metadata associated with this file on the RCMRD GeoPortal (see [2]), or the website of the Rwanda Ministry of Agriculture and Animal Resources (which the RCMRD GeoPortal cites as the author of the file). Opening a KML of this river file in google earth and measuring the width of various rivers labeled with different hierarchy values seemed to indicated that rivers with hierarchy values of 1,2, or 3, were on the order of 30-70m across. Rivers of hierarchy 4 or 5 were significantly less substantial. As rivers of the first sort would seem to represent a possible deterent to infrastructure routing, whereas rivers of the latter are most likely of little to no consequence when it comes to line routing, only the former were used. Data for the rivers of hierarchy 1,2, and 3 were exported to a new shapefile, converted to polygons, each having a total width of 50m, and subsequently merged into the single penalized zone raster for inclusion in the penalized zone raster file. If a feature file which more accurately depicts these bodies of water can be obtained, either from the client or via another perhaps publicly-available repository, that would be preferred. But for the time being, we will proceed with this data as described.
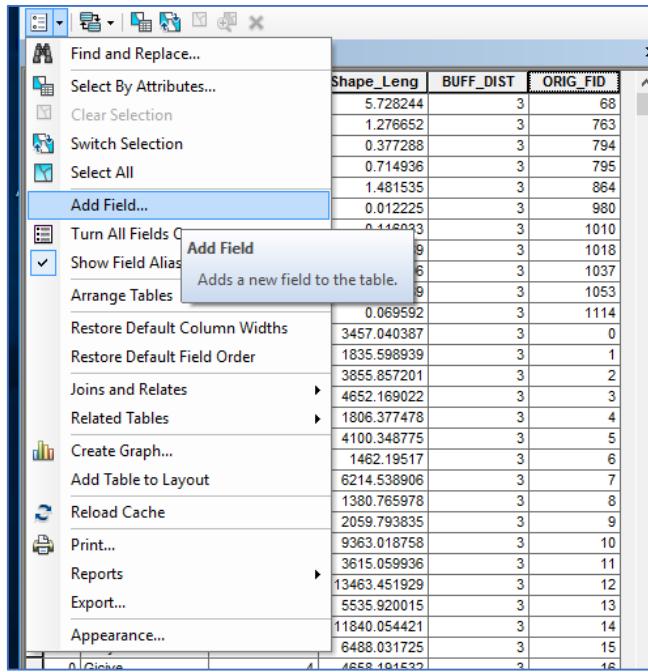
You can check the attribute of the new table to verify that the river features are now polygons.
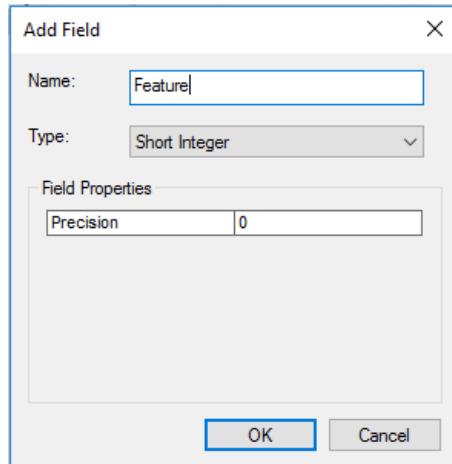
4. Now we want to merge these four shapefiles into a single shapefile, but before doing so, we are going to assign each type of feature a unique identifier so that we can distinguish them once they have been combined. We may either change the values within an existing field of each attribute table or create a new field. The value inserted may be a name, such as 'River', or a numerical value. If you know what penalty values you would like to assign, you can enter those values here. If you would like to be able to add penalty values at a later by using 'find and replace' in a text editor, then add a unique numerical identifier here such as '1' for rivers, '2' for lakes, etc. You may choose to use names, but a number will have to be

assigned when the file is converted to a raster, so perhaps preferable to just assign a number here and note which number corresponds to which feature. For this example, we are going to assign values as follows: 1-Rivers, 2-Lakes, 3-Wetlands, 4-National Parks.
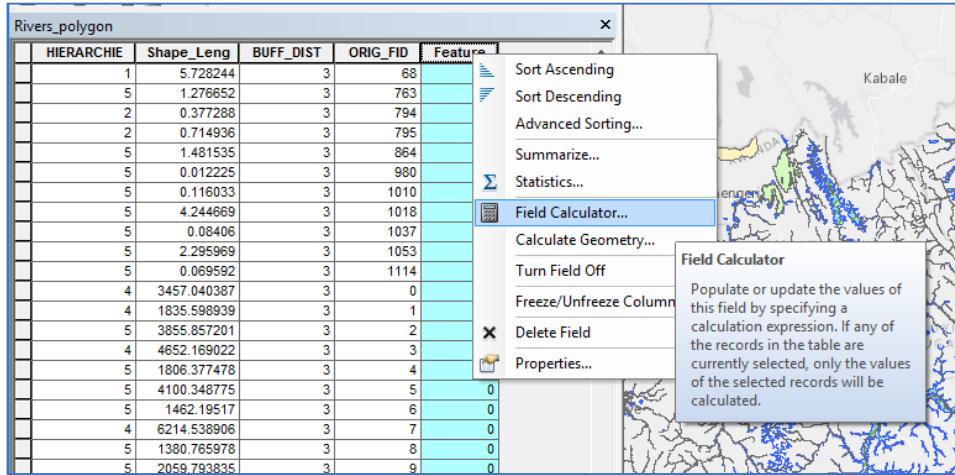
In the attribute table, click the drop-down menu at the left of the menu bar and select 'Add Field…'.
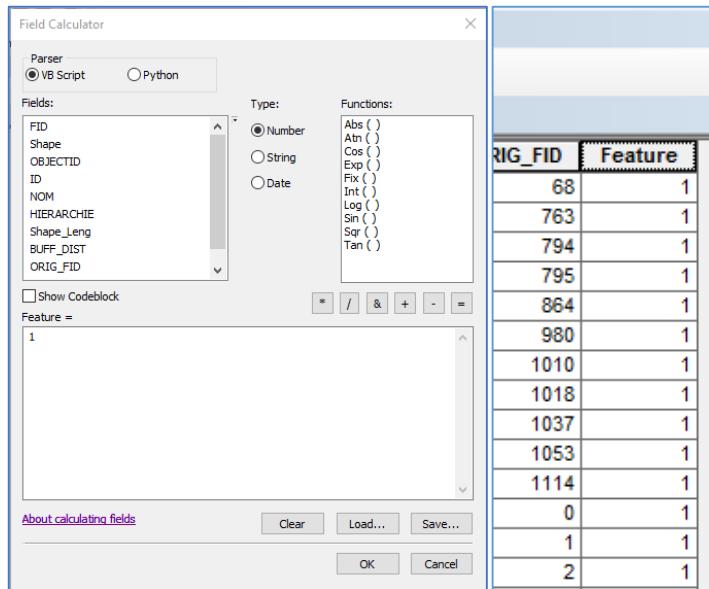
Enter field name and type.
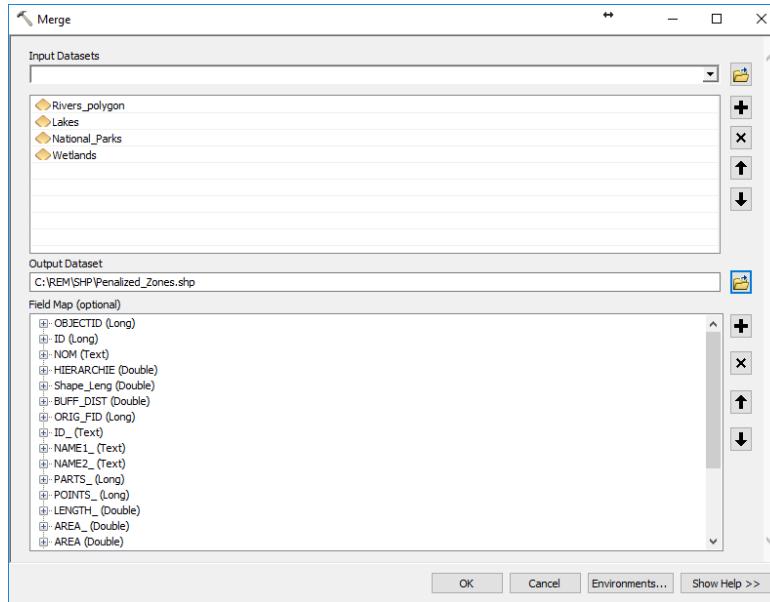
Right click on the new field, select 'Field Calculator…', and click yes to the warning that pops up.
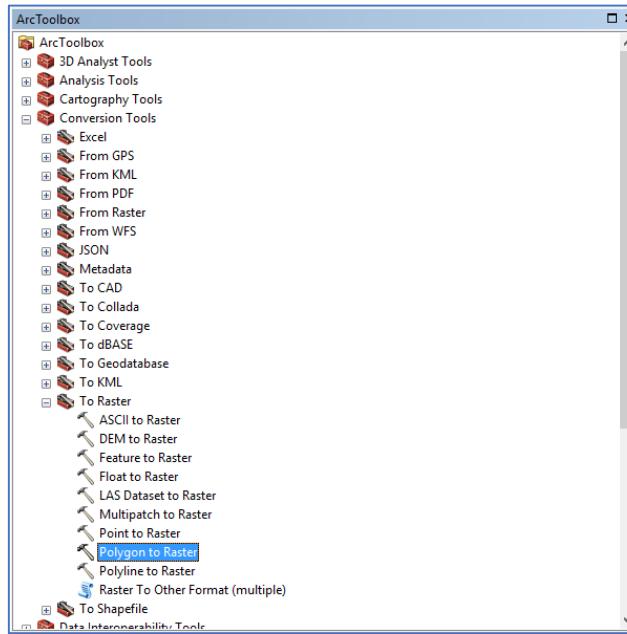
Enter the value in "Field =" and click 'ok'. Now the feature value for each river is 1.
Repeat for the other three shapefiles, naming the new field 'Feature' in each case but assigning the appropriate value according to feature type (River, Lake, etc.).

5. Merge the four shapefiles into a single shapefile by going to geoprocessing>merge.
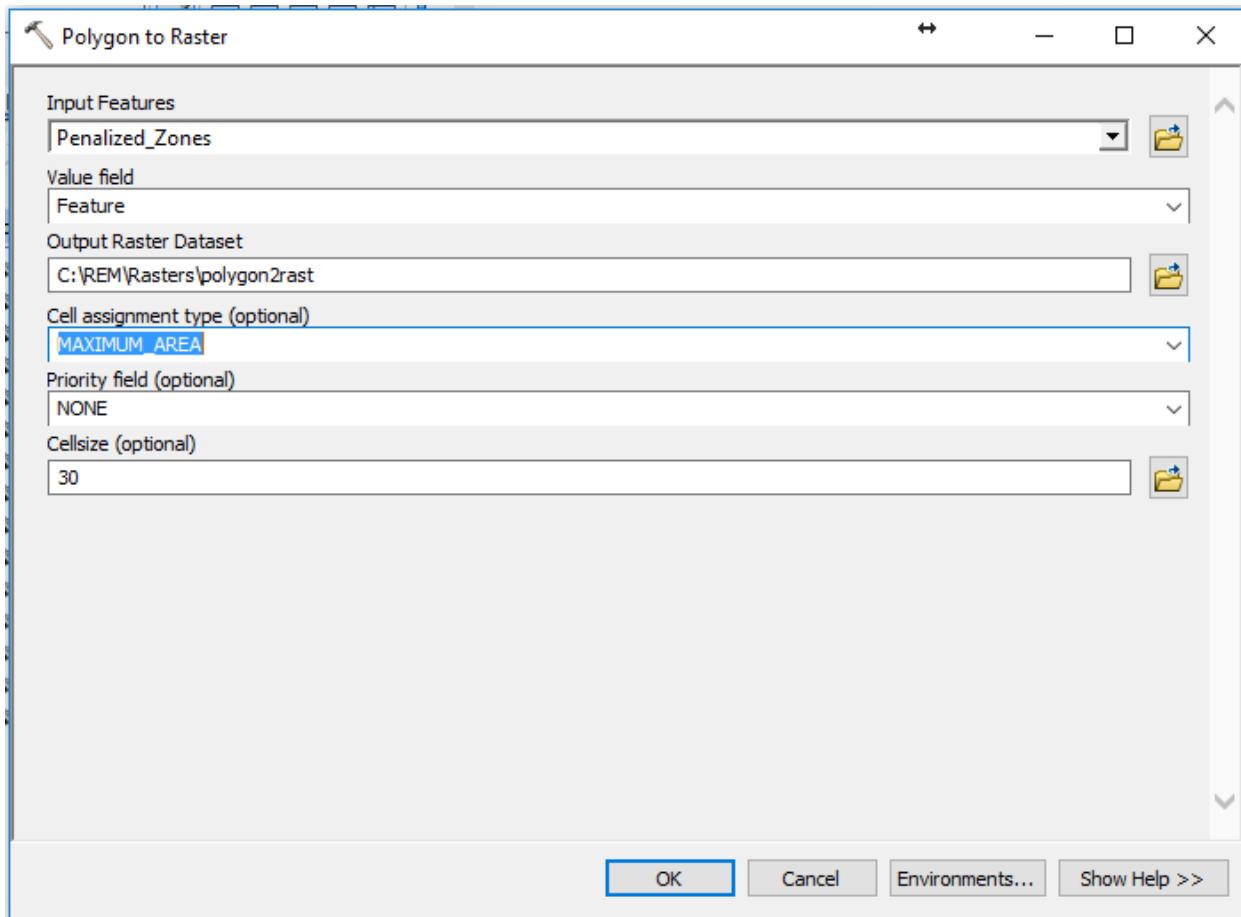
6. To convert this file to the ascii format, we must first convert it to raster format. Go to ArcToolbox>Conversion Tools>Polygon to Raster.
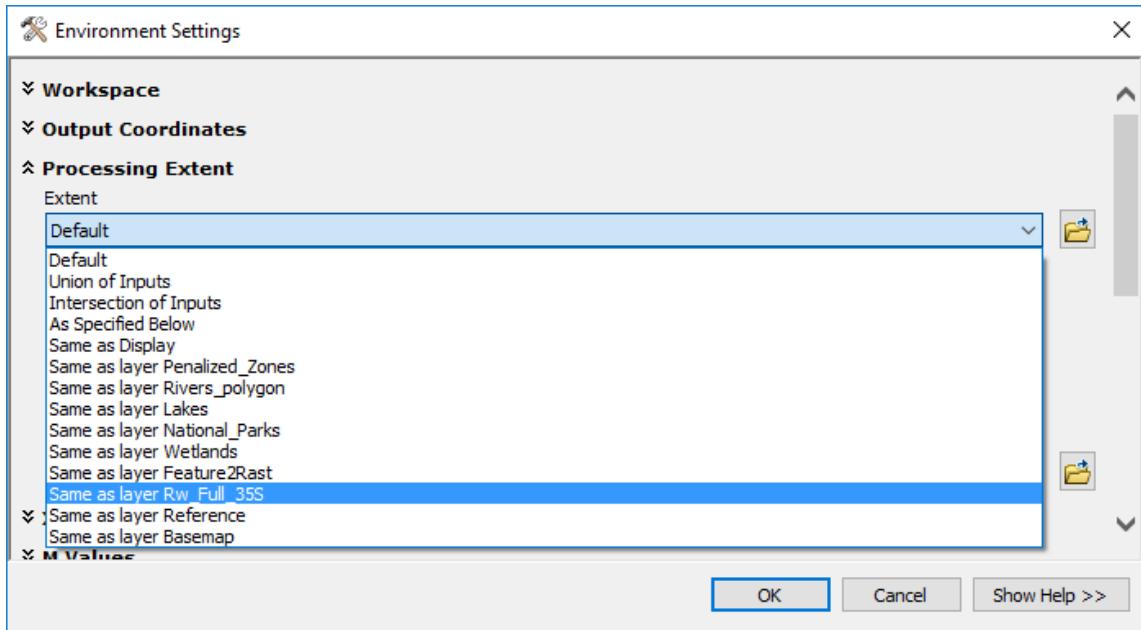
Select the new field you've created in the 'Field' dropdown menu, in this case we select 'Feature'. Select 'MAXIMUM_AREA' from the cell assignment type dropdown menu*. Be sure to specify a cell size consistent with that of the other input files.

*If more than one feature lies within a cell, selecting 'Maximum area' will give preference to the feature occupying the largest area within the cell (i.e. if the two features have different values, one must take precedence, as only a single value can be assigned to any given raster cell). If a

cell size greater than that dictated by the 1arcsec SRTM data is used, the 'MAXIMUM_COMBINED_AREA' cell assignment type may be more appropriate. See [6] for further details about the cell assignment type options and the other fields in the Polygon to Raster function.

Now convert the raster file you've created to the requisite ascii file using the Raster to ASCII function as specified in the Altitude Raster section of this document. When doing so, be sure to go to Environments>Processing Extent and select the altitude raster from which you made the altitude ascii file as the extent. This will ensure that the header values 'xllcorner' and 'yllcorner' match those of the altitude raster. As we already set the cell sizes of the two rasters equal to each other (see above), then specifying here that the raster extents match should also ensure that the header values 'ncols' and 'nrows' are the same for both rasters.

Raster text file post-processing steps include changing all of the no data values in the body of the file to zero (except for the value in the header) and using 'find and replace' to change the feature identifying numbers (1-Rivers, 2-Lakes, 3-Wetlands, 4-National Parks) to penalty values. Place the file in the RNM_input file that your case will be calling. Unlike the other two input files discussed in this document, the penalized zone raster is not subject to any mandatory name convention. Although the filepath for it specified in section 3E of the script, so must be updated with the filename as appropriate. Therefore, a mandatory naming convention should ultimately be applied to this file as it is to the other two.

3) Penalized Zone Polygon File

The RNM forbidden zones file is created using the executable file named 'shp2rnm' that is located in the RNM.EXE folder in Dropbox (MIT Universal Energy Access Group>RNM>RNM.EXE>'shp2rnm.exe'). Extract the file (so that it has the extension .exe instead of .7z) and open a command window in your computer. After the prompt (mine is C:\Users\cailinn>), type the following:

shp2rnm (shapefile name).dbf (shapefile name).shp>C:\(path where you want the new file to be saved)\(name you want new file to have).

Below is an example where 'forbidden_nw_test' is the name of the shapefile containing forbidden zones, 'C:\REM\runs\Data\RNM_input' is the folder where I want the new file to be saved, and 'testfile' is the name I want the new file to have. Note that the shapefile to be processed by shp2rnm.exe must be located in the 'C:\Users\cailinn' file in this example for this to work.

```
C:\Users\cailinn>shp2rnm forbidden_nw_test.dbf
```
…
```
forbidden_nw_test.shp>C:\REM\runs\Data\RNM_input\testfile
```
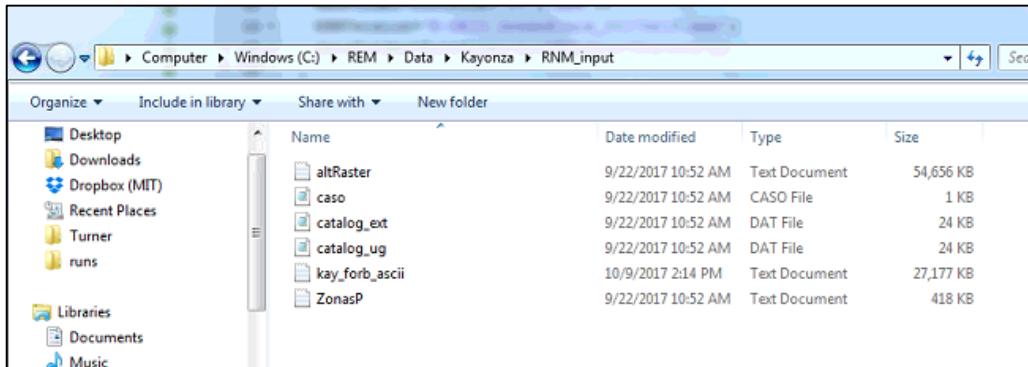
Take care that the shapefile used is in a projected coordinate system and that that coordinate system is consistent with those of the other input files used. Be aware that shp2rnm.exe converts the projected coordinates from meters to kilometers, so the coordinates will appear in the polygon file as reduced by a factor of 1000, as can be seen in the figure below. The resulting file will resemble the image below. The header contains three penalty factors, 'FACTOR_CINV', 'FACTOR_CMANP', and 'FACTOR_CMANC'; however, RNM executable files created on or later than January 27,2017 use ONLY the penalty factor FACTOR_CINV. The penalty factor FACTOR_CINV is assigned the value of 10* by default, but may be changed to any value, including decimals. This value will be used to avoid penalized areas in network routing decisions. However, when the actual cost of the line is calculated, the cost penalty multiplier is capped at 10. In other words, if a value less than 10 is entered for FACTOR_CINV, the line cost will be multiplied by that value. But if a value greater than 10 is entered, then the cost will be multiplied by 10. Note that this header repeats over and over again in the file, one time for each zone polygon; so if you want to change the penalty values, you need to repeat the change for each header.

*In the event that a zonal polygon is contained within another zonal polygon, such as in the case of a lake that lies within a larger area labeled as wetlands, RNM will create three entries with three separate headers – one to add the wetland polygon, one to add the lake polygon, and one to subtract the lake polygon. To subtract the lake polygon, RNM will make the penalty factors (of which only one is used, see above for details) negative values. i.e. If the FACTOR_CINV value for the wetland is 5, and for the lake is 10, there will be be three polygon headers with FACTOR_CINV equal to 5 in the first, 10 in the second, and -10 in the third. Please be cognisant of this effect when adjusting FACTOR_CINV values.

```
Numero_de_nucleos:  6
NOMBRE:             95;0.906;7118.31;0;0;0;7118.33;lakes;0;223.868;0;0;0;0;7122.82;907129;1;1
TIPO:               0
FACTOR_CINV:        10
FACTOR_CMANP:       10
FACTOR_CMANC:       10
VERTICES:           247
XMin=               778.6517
XMax=               9818.8796
YMin=               780.2025
YMax=               9820.1578
Perim=              0
Area=               0
CdmX=               779.4271
CdmY=               9819.5187
CdmZ=               0
Radio=              0
 778.6856739 9819.9396526 0
 778.7147846 9819.9316668 0
 778.7518259 9819.9157267 0
 778.7968176 9819.9050697 0
 778.8391815 9819.9076541 0
 778.8656841 9819.9261473 0
 778.8816163 9819.9578937 0
 778.8869428 9819.9790657 0
```

Place the file in the RNM_input file that your case will be calling. The file must be named 'ZonasP.txt'.

Please refer to the *Reference Network Model Manual_v18_Greenfield* for further details regarding the values contained within this file. For details regarding line cost calculation and penalty application in RNM, refer to the document entitled *RNM_Greenfield_Economical_Description*, updated October 6, 2017 to reflect changes made to the way FACTOR_CINV is used. Both of these files are available in the Dropbox folder MIT Universal Energy Access Group>RNM>doc. The manual is in the folder 'Manual', the economical description is in the folder entitled 'Other documentation'.

Sources
[1] https://lta.cr.usgs.gov/SRTM1Arc
[2] http://servirportal.rcmrd.org/layers/servir%3Arwanda_srtm30meters
[3] http://pro.ArcGIS.com/en/pro-app/help/mapping/properties/specify-a-coordinate-system.htm
[4] http://pro.ArcGIS.com/en/pro-app/tool-reference/environment-settings/mask.htm
[5] https://blogs.esri.com/esri/ArcGIS/2008/09/26/combining-raster-dem-tiles-the-right-way/
[6] http://pro.ArcGIS.com/en/pro-app/tool-reference/conversion/polygon-to-raster.htm

7 Appendix II

Penalty Application in RNM[33]

Please note that this appendix is organized such that the final line cost equations are given first, followed by a successive breakdown of the terms provided in these initial equations.

Line Cost

$$NPV_{line} = NPV_{line,losses} + NPV_{line,investment} + NPV_{line,maintenance} \tag{A1}$$

$$NPV_{line,losses} = \left(\sum_{t=0.5}^{t \leq YLG} \frac{dC^{2(t-1)}}{dTd^t} + \sum_{t=YLG+1}^{t \leq YUL} \frac{KFD^2}{dTd^t} \right) \cdot C_{PL,0} \tag{A2}$$

$$NPV_{line,investment} = IC \cdot HCF_{line} + NPV_{equipment} + NPV_{DFP} + NPV_{CVR} \tag{A3}$$

$$NPV_{line,maintenance} = NPV_{line,maintenance,prev} + NPV_{line,maintenance,corr} \tag{A4}$$

$$NPV_{line,maintenance,prev} = \left(\sum_{t=0.5}^{t \leq YUL} \frac{1}{dTd^t} \right) \cdot C_{line,maintenance,prev} \cdot HCF_{line} \tag{A5}$$

$$NPV_{line,maintenance,corr} = \left(\sum_{t=0.5}^{t \leq YUL} \frac{1}{dTd^t} \right) \cdot C_{line,maint,corr} \cdot FR_{average} \cdot HCF_{line} \tag{A6}$$

$$dTd = \frac{WACC}{100} + 1 \tag{A7}$$

$$C_{PL,0} = P_{PL,0} \cdot 8760 \cdot loss\ factor \cdot cost\ of\ losses \tag{A8}$$

$$loss\ factor = \frac{P_{average}}{P_{max}} = \frac{1}{P_{max}} \cdot \frac{\int_0^T P(t)dt}{T} \tag{A9}$$

---

[33] Contents adapted from [29]. It should be noted that although some of the constituent equations differ slightly from those presented in [29], all deviations are such that the primary equations are mathematically equivalent to those presented in the original source.

$$P_{max} = P_{PL,0} \cdot dC^{2 \cdot YDG} \tag{A10}$$

$$dC = \frac{PIPD}{100} + 1 \tag{A11}$$

Where:
- YLG – years of losses growth: 10
- YUL – years of useful life: 40
- KFD – k-factor for demand until end of useful life (p.u.): 0.01
- WACC – weighted average cost of capital or discount rate: 6%
- YDG – years of demand growth: 5
- cost of losses ($/kWh): 0.05
- $P_{PL,0}$ – the total power losses in the peak demand hour of a line in the starting year (kW)
- $PIPD$ - percentage increase in power demand: 3%
- $C_{PL,0}$ represents the costs associated to power losses in the line in the starting year
- P_{max} - the total power losses in the peak demand hours of a line in the year of maximum demand, which is defined as the terminal year in the years of demand growth
- $P_{average}$ - average power losses
- P(t) - the instantaneous demand power losses
- $\int_0^T P(t)dt$ - the energy losses of the system during T
- IC are the investment costs of a line per km, which is an input parameter of the catalogue ($/km).
- HCF_{line} is the height cost factor of a line (km). Its value is equal to 1.0 if the influence of orography is neglected when calculating costs associated to lines.
- $MPAF_{line}$ is the modified prohibited area factor. Its value is equal to 0 if there is not a phohibited area.
- NPVequipment = is the net present value of equipment for improving reliability that is located in the line.
- NPVditches = is the net present value of ditches, façades and post associated to the line.
- NPVcapacitors & voltage regulators = is the net present value of capacitors and voltage regulators that are located in the line.
- $C_{line,maintenance,prev}$ are the maintenance preventive costs per km and per year of a line ($/(km · year)).
- $C_{line,maintenance,corr}$ are the maintenance corrective costs per failure of a line ($/failure).
- $FR_{average}$ is the average failure rate of a line per km and per year (failure/(km · year)). This is calculated using the input parameters TF_min, TF_med and TF_max of the RNM catalogue, obtaining $FR_{average}$ as the mean of the triangular fuzzy number given by (TF_min, TF_med, TF_max). $FR_{average}$ usually takes values from the interval [0.1, 0.2] in lines.

Height Cost Factor

$$HCF_{line} = LH \tag{A12}$$

HCF_{line} is the height cost factor of a line (km). Its value is equal to the length of the line if the influence of orography is neglected when calculating costs associated to lines. The line height (LH) of the line whose starting and ending points are A and B is calculated as:

$$LH = \sum_{i=1}^{p-1} CM(r_i|_{XY}) \cdot dist(s_i, s_{i+1}) \tag{A13}$$

Where $dist(s_i, s_{i+1})$ is the Euclidean distance between s_i and s_{i+1} (km) for a line broken into segments as shown in Figure AII - 1, and CM(i,j) is the (i,j)-th element of the 'cost matrix' and is calculated as per equation (A14).
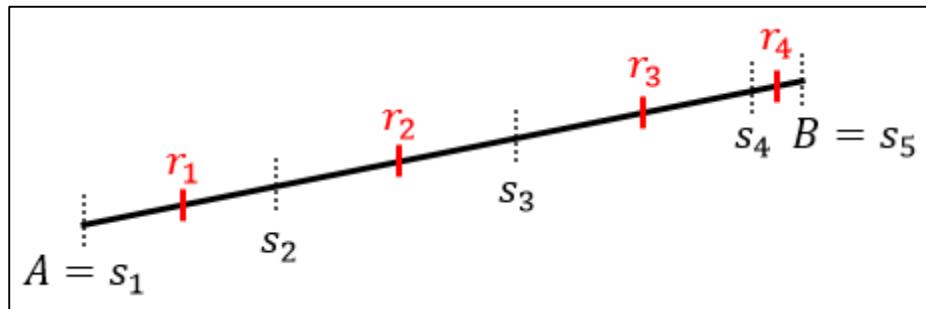
Figure AII - 1: Line Divided into Segments Based on Raster Cell Resolution

$$CM(i,j) = (dRm \cdot dKinv(i,j) + 1 - dRm) + MPAF \tag{A14}$$

Where:
- $dRm = 0.5$ (hard-coded value)[34]
- *MPAF* is the 'modified prohibited area factor', which is equal to FACTOR_CINV, the penalty value entered into the penalized polygon zone input file (discussed further below)

---

[34] According to Dr. Carlos Mateo at Comillas University, the justification for this value is that 50% of the length-dependent portion of the line cost is considered to be affected by topography, so only 50% of the cost is to be penalized by topography penalties.

Altitude-related component of CM(i,j) – 'dKinv'

If NG(i,j), or the normalized gradient of a segment, is less than 0.57, then dKinv(i,j) is calculated as follows. Otherwise, dKinv(i,j) = 20.

$$dKinv(i,j) = (1 + (dKinvMax - 1) \cdot NG(i,j)) \cdot dC(i,j) \qquad \text{(A15)}$$

Where:
- dKinvMax = 3 (hard-coded value, no justification provided for this value)
- $NG(i,j)$ is the (i,j)-th element of the normalized gradient matrix.
- dC(i,j) = a value which depends upon the value of RM(i,j)
- $RM(i,j)$ is the (i,j)-th element of the GIS input file matrix (m)
- dC(i,j) =

        1 for altitude 0-500m
        1.02 for altitude of 500-1000m   **IF normalized slope ≤ 0.57[35]**
        1.05 for altitude of >1000m

Zone Penalties
The modified prohibited area factor is calculated as follows.

$$MPAF = \begin{cases} IPAF & IPAF < 10 \\ 10 & IPAF \geq 10 \end{cases}$$

Where:

- $MPAF_{line}$ is the modified prohibited area factor in per-unit. Its value is equal to 0 if there is not a prohibited area.
- $IPAF_{line}$ is the input prohibited area factor in per-unit. Its value is equal to 0 if there is not a prohibited area.

---

[35] Where 0.57 is equal to the inverse tangent of 37°

8 Appendix III

The slope penalty currently applied in RNM is a linear function of the slope gradient. In order to test the hypotheses set forth in section 3.1, the slope penalty currently applied in the clustering algorithms is also a linear function of the slope gradient. The purpose of this appendix is to suggest an alternative definition of slope penalty, which if adopted, should be considered for implementation in both RNM and the clustering algorithms.

It would seem reasonable to argue that the slope of a section of terrain has little to no effect on the cost of installing or maintaining electrical infrastructure until the lower limit of a certain range of grade is reached. Within this range, the cost presumably changes noticeably with incremental changes in slope, until it reaches the upper limit of this range. This upper limit would be defined by a value such as the 37 degrees currently used in RNM to denote a degree of incline beyond which it is impractical to erect infrastructure. The penalty would increase sharply during the intermediate range and then level off after the upper limit is reached. This effect could be represented by a function such as that shown equation (B1).

$$y = A * \tan^{-1}(Bx + C) + D \qquad \text{(B1)}$$

An example of the shape of this function is illustrated in Figure III - 1, where the value of the scalars A, B, C, and D take the values 7, 0.5, -15, and 10.5, respectively.
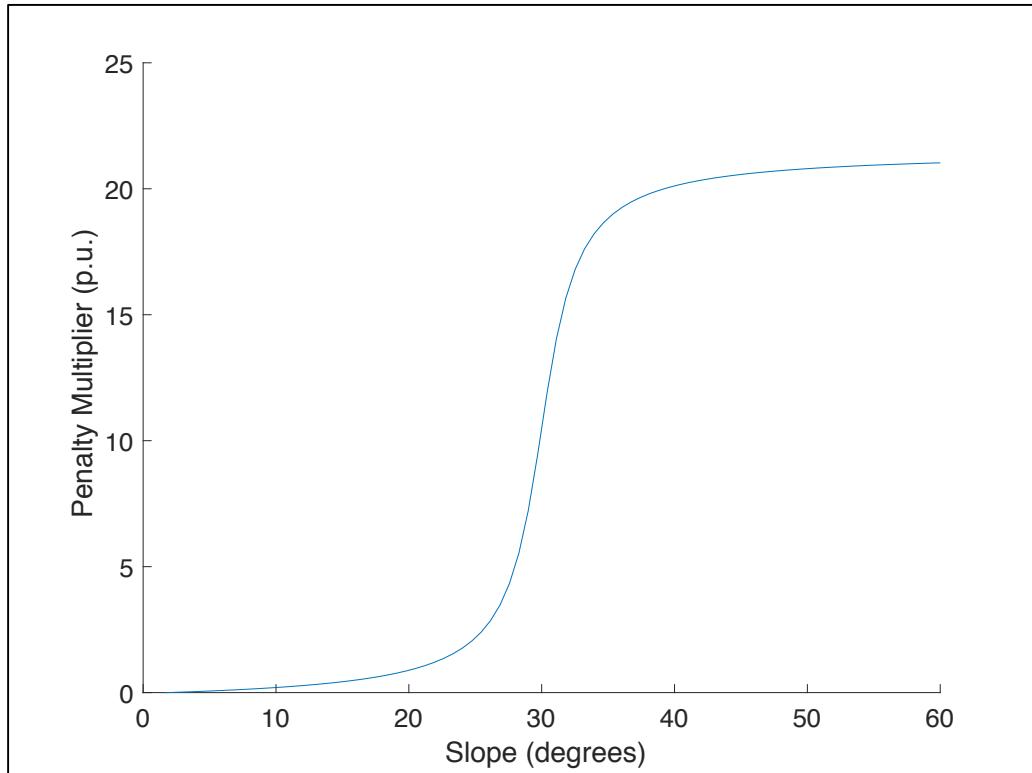

Figure III - 1: Alternative Slope Penalty Definition Using Arctangent Function

The function whose graph is shown in Figure III - 1 represents an evolution of slope penalty assignment which transitions smoothly as slope increases, as opposed to the current application in RNM and clustering whereby the slope penalty is equal to the slope until the latter exceeds 37 degrees, at which point there is a discontinuous jump to a slope penalty multiplier of 20.

It should be noted that the RNM software was developed in conjunction with electrical distribution utilities in Spain and is currently used as the distribution network design standard by which all electrical distribution utilities in Spain are remunerated. It therefore stands to reason that the current methodology used for assessing slope penalties was developed based on informed end-user input and has since functioned in a manner sufficient to satisfy all parties involved in the remuneration process. In light of that fact, it might be that the level of slope penalty definition refinement suggested in this appendix would be of little functional consequence. Nevertheless, it may prove useful to reconsider at some juncture the manner in which slope penalties are defined and applied.

9 Appendix IV
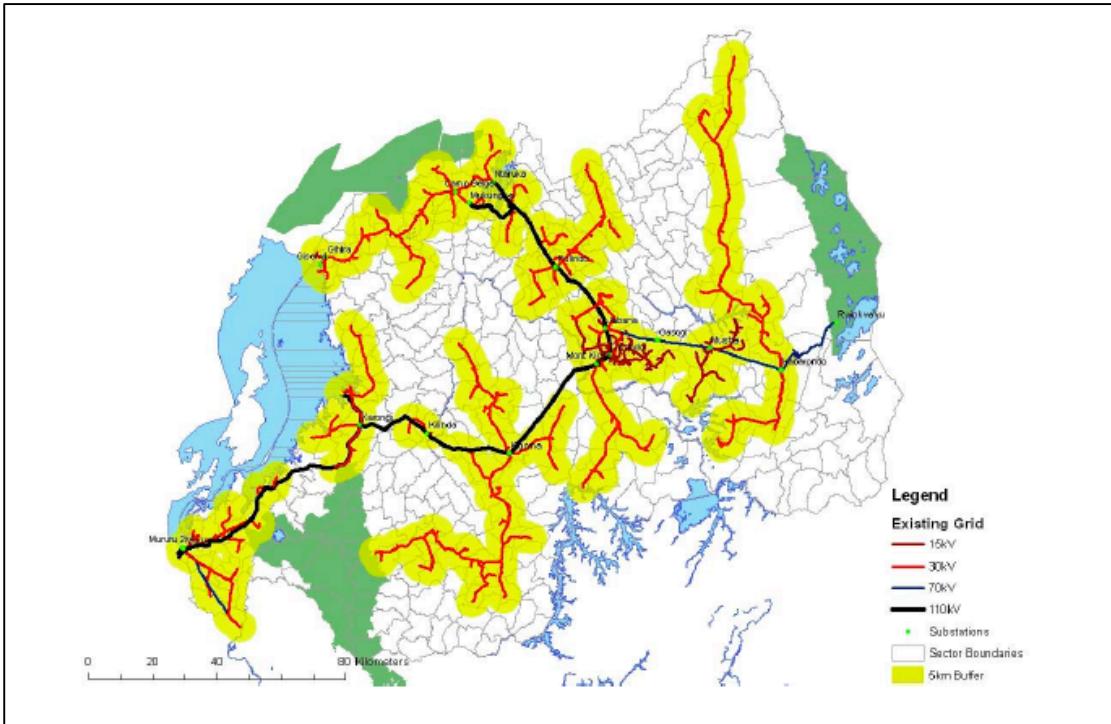
*Figure IV - 1: Existing Network Shown with 5km Buffer Zone (source: [36])*
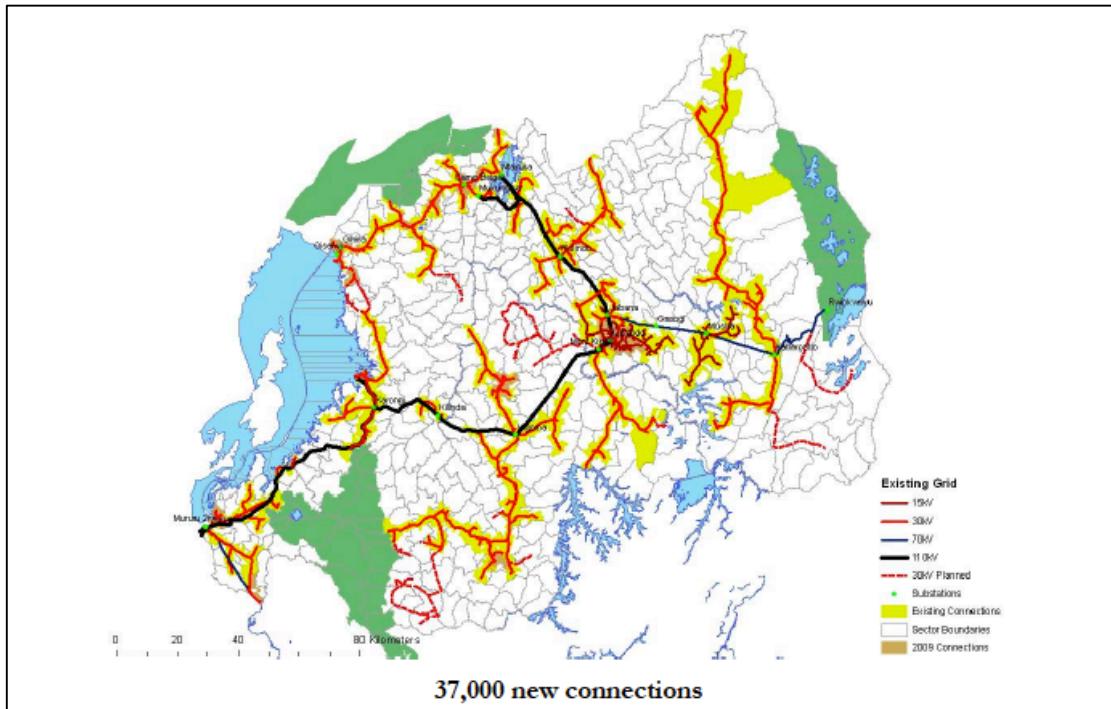
*Figure IV - 2: Least-cost Grid Connection Scenario: 2009 Connections (source: [36])*
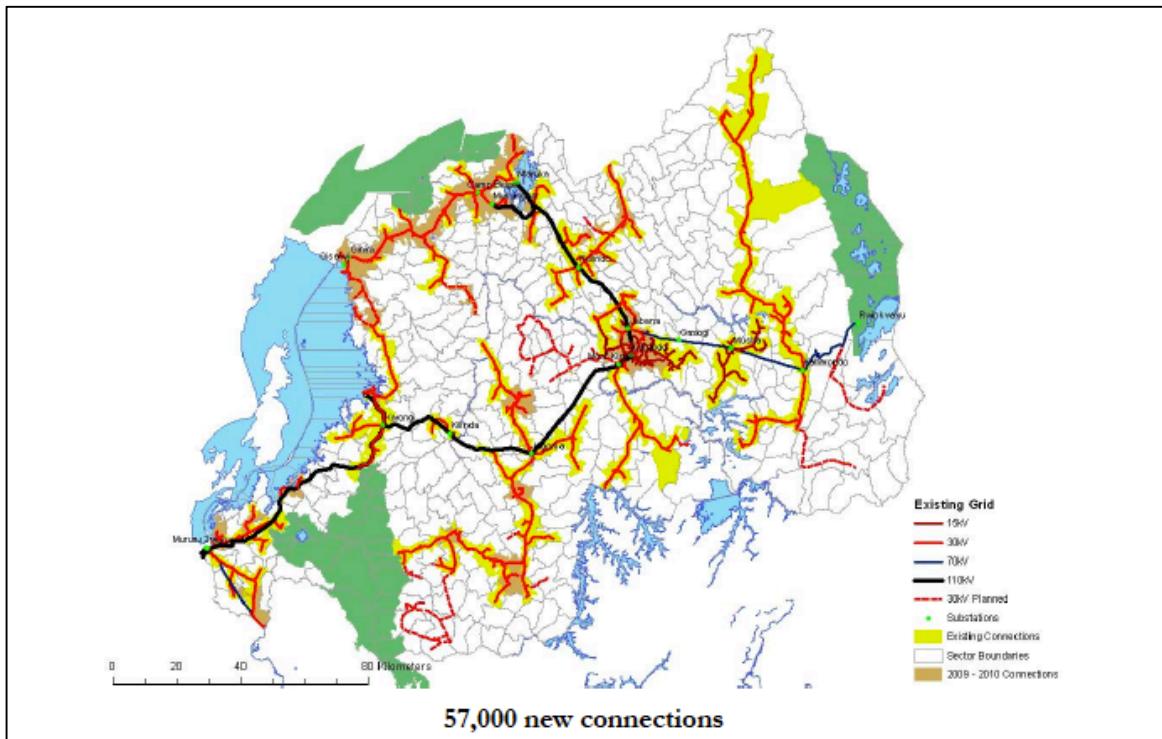
*Figure IV - 3: Least-cost Grid Connection Scenario: 2009-2010 Connections (source: [36])*
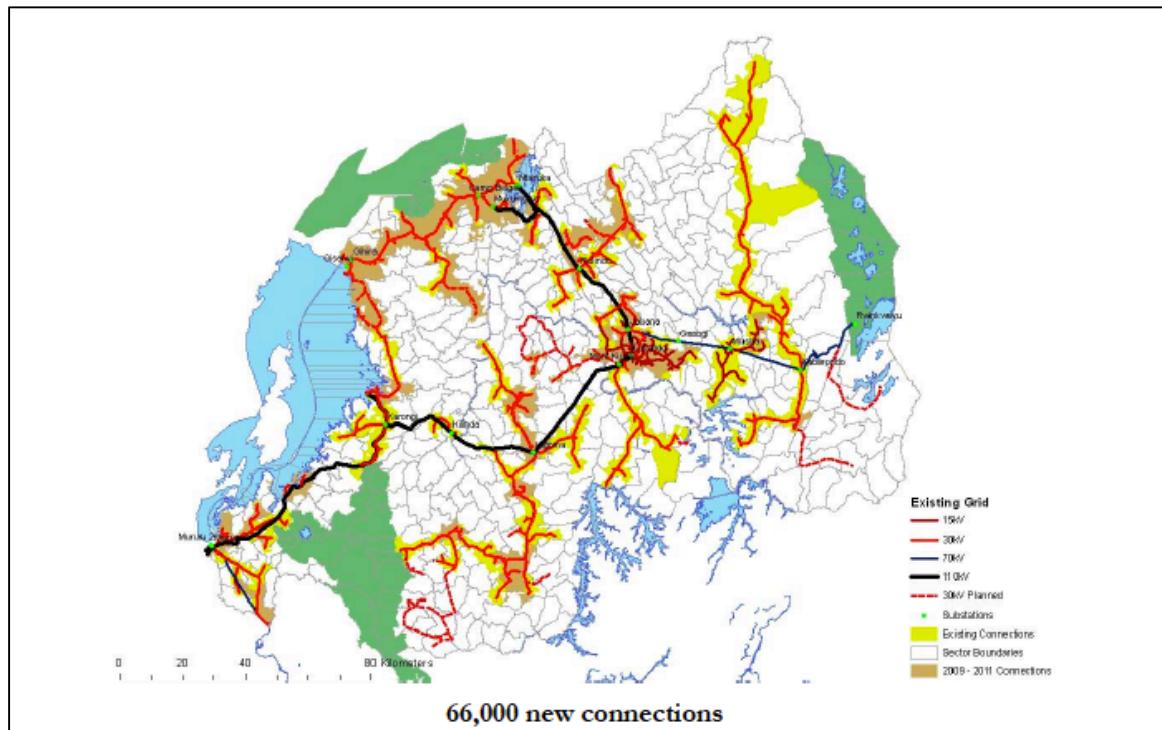
*Figure IV - 4: Least-cost Grid Connection Scenario: 2009- 2011 Connections (source: [36])*
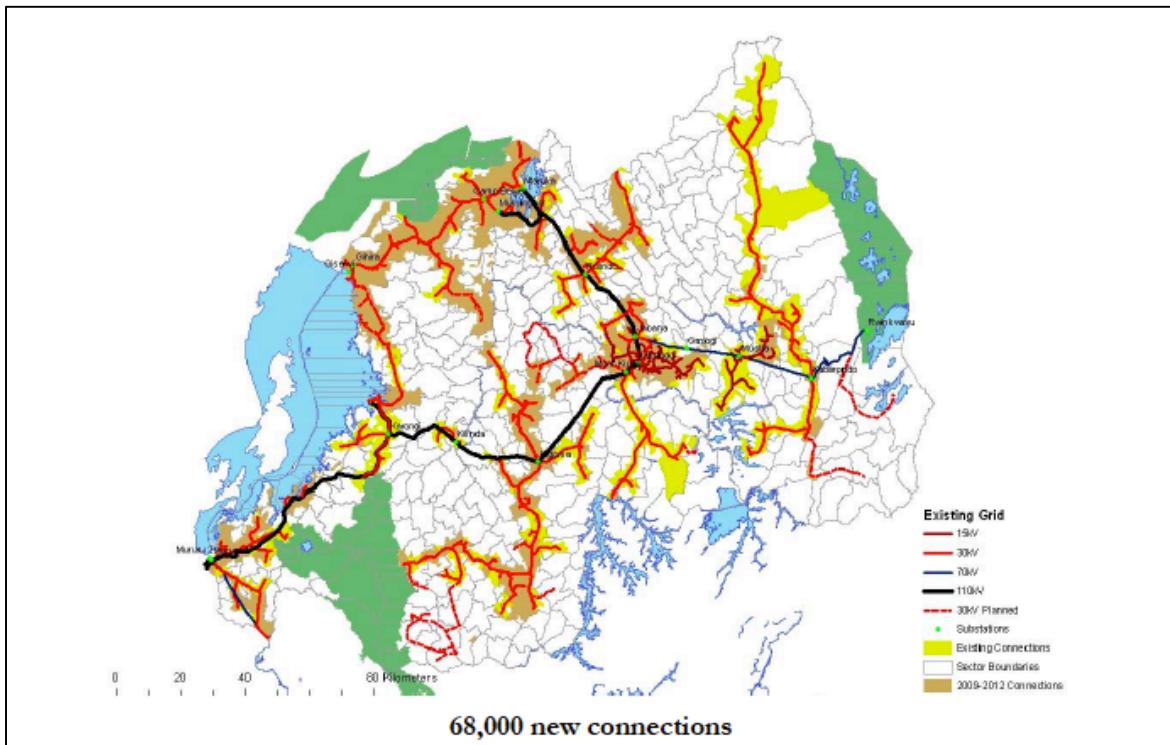
*Figure IV - 5: Least-cost Grid Connection Scenario: 2009 -2012 Connections (source: [36])*
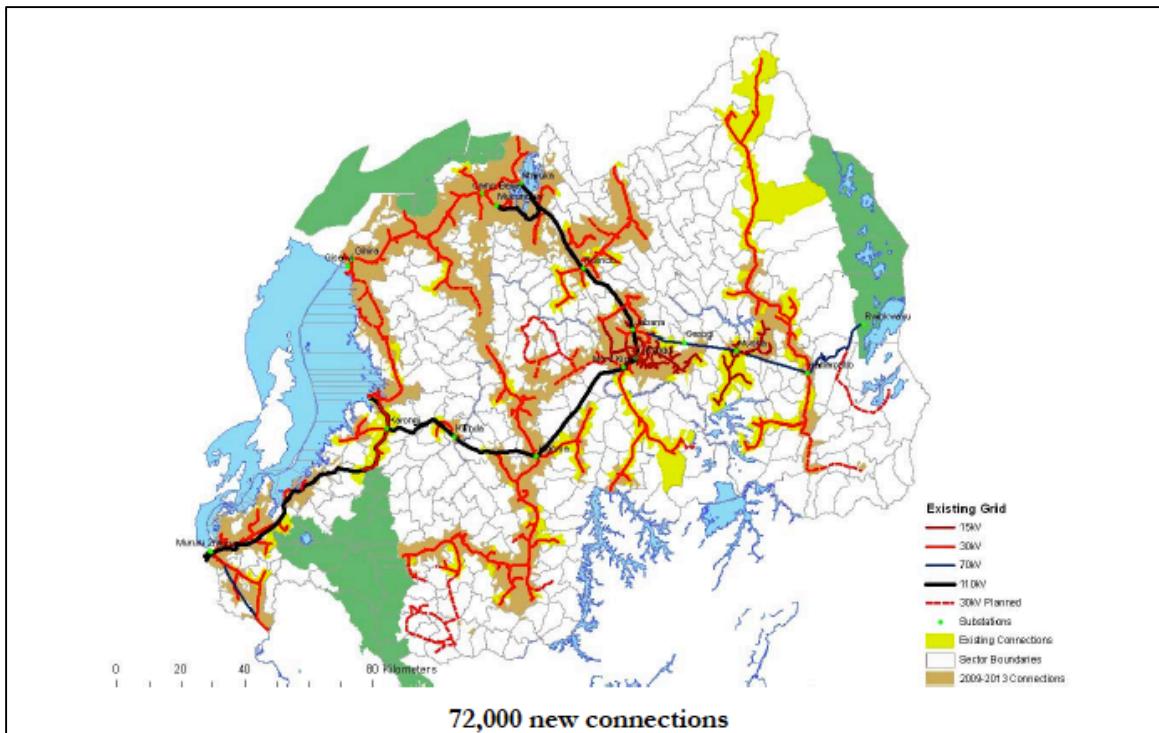
*Figure IV - 6: Least-cost Grid Connection Scenario: 2009-2013 (source: [36])*

105
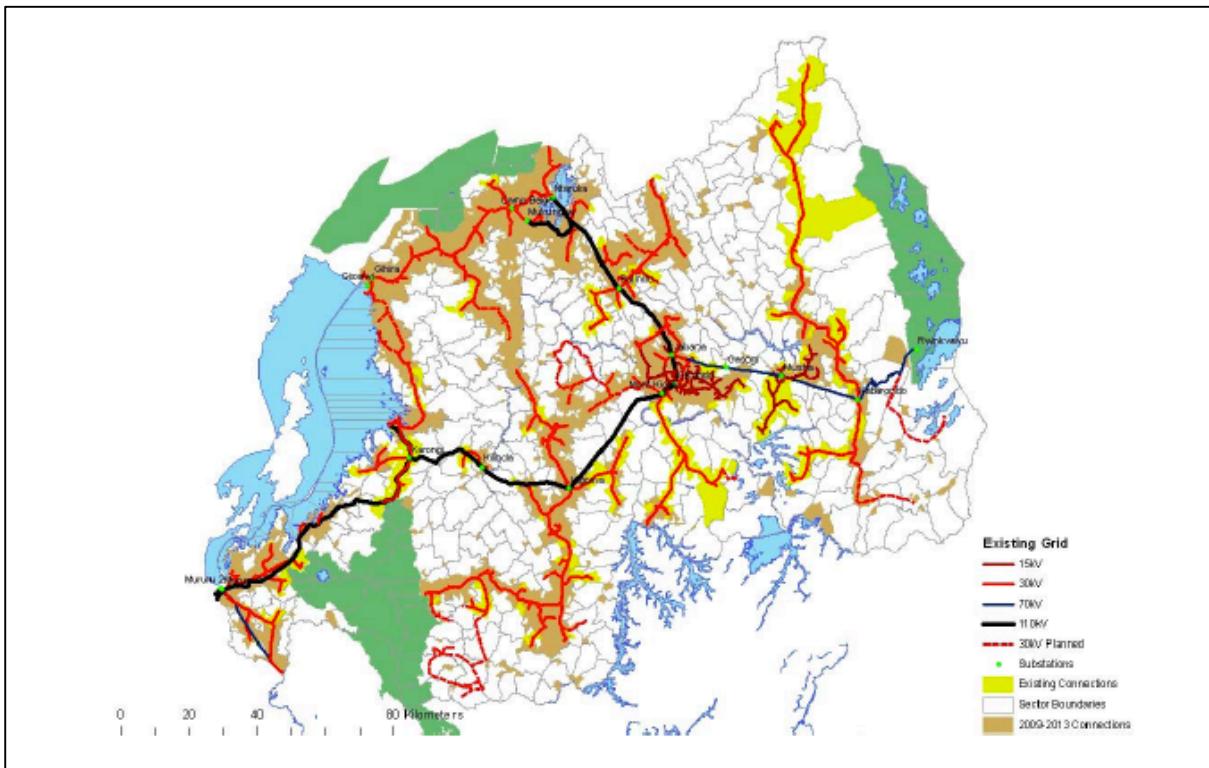
Figure IV - 7: Scenario Whereby 100% of Administrative Sectors are Grid-connected by 2012 (2013?) (source: [36])
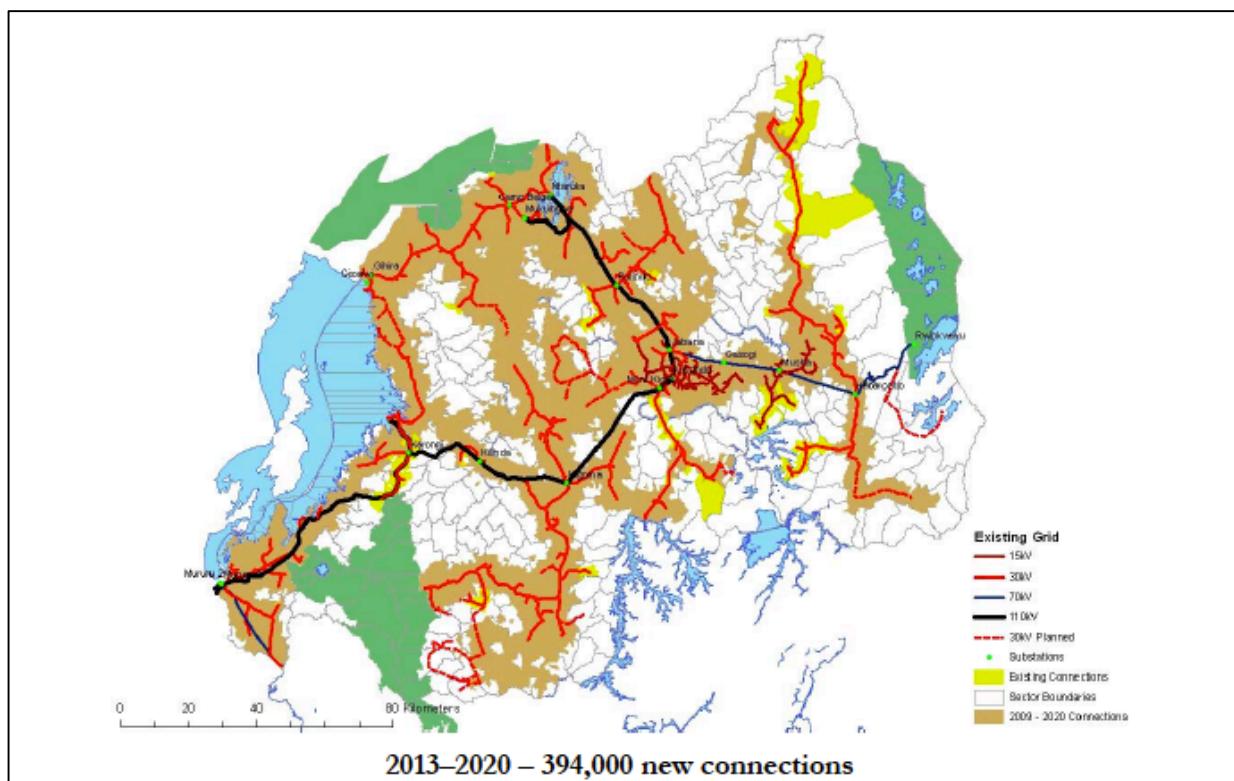

Figure IV - 8: Least-cost Grid Connection Scenario: 35% National Electrification Rate by 2020 (source: [36])