Horizon 2020 Programme

LEAP-RE

Research and Innovation Action (RIA)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 963530

Start date : 2020-10-01 Duration : 63 Months
http://www.leap-re.eu/

LEAP-RE
Long-Term Joint EU-AU Research and Innovation Partnership on Renewable Energy

Joint EU-AU report on the platform testing and validation activity

Authors : Dr. Manfred HAFNER (HEAS), Gregory Ireland (UCT), Alison Hughes (UCT), Muhammad Awais (IIASA), Edward Beyers (IIASA), Giacomo Falchetta (IIASA), Vittorio Giordano (PoliTo), Manfred Hafner (HEAS), Simon Rukeratabaro (UR), Francesco Semeria (IIASA), Marta Tuninetti (PoliTo), Adriano Vinca (IIASA), Ackim Zulu (UNZA)

LEAP-RE - Contract Number: 963530

Project officer: Bernardo Luis ABELLO GARCIA

Document title	Joint EU-AU report on the platform testing and validation activity
Author(s)	Dr. Manfred HAFNER, Gregory Ireland (UCT), Alison Hughes (UCT), Muhammad Awais (IIASA), Edward Beyers (IIASA), Giacomo Falchetta (IIASA), Vittorio Giordano (PoliTo), Manfred Hafner (HEAS), Simon Rukeratabaro (UR), Francesco Semeria (IIASA), Marta Tuninetti (PoliTo), Adriano Vinca (IIASA), Ackim Zulu (UNZA)
Number of pages	62
Document type	Deliverable
Work Package	WP12
Document number	D12.5
Issued by	HEAS
Date of completion	2022-12-29 18:53:39
Dissemination level	Error (13) !

Summary

This report is the third scientific deliverable of the RE4AFAGRI (?Renewables for African Agriculture?) project (or WP12 of LEAP-RE - Long-Term Joint Research and Innovation Partnership on Renewable Energy between the European Union and the African Union - Pillar 2). The RE4AFAGRI project seeks to support sub-Saharan African (SSA) smallholder farmers and communities to grant community-wide access to energy services and water for crop irrigation and human use with the ultimate goal of fostering rural development. RE4AFAGRI aims at providing African research institutions and public and private decision-makers with the tools and expertise necessary to operate a multi-scale modelling platform that can support the design and implementation of integrated solutions for the energy and water nexus in rural areas. In parallel, RE4AFAGRI aims at setting the ground for a multi-stakeholder discussion platform about the business models and enabling environment (policy and regulation) to promote the involvement of the private sector in water-energy-agriculture integrated solutions.

Approval

Date	By
2022-12-29 18:54:15	Dr. Manfred HAFNER (HEAS)
2023-01-03 12:11:07	Mr. Léonard LéVêQUE (LGI)

LEAP-RE

Long-Term Joint EU-AU Research and Innovation Partnership on Renewable Energy

Joint EU-AU report on the platform testing and validation activity as a deliverable of Task 12.3

Field data collection, calibration, and validation

Deliverable D12.5

WP12 of LEAP-RE (RE4AFAGRI)

http://www.leap-re.eu/

https://www.leap-re.eu/re4afagri/

Authors: Gregory Ireland (UCT), Alison Hughes (UCT), Muhammad Awais (IIASA), Edward Beyers (IIASA), Giacomo Falchetta (IIASA), Vittorio Giordano (PoliTo), Manfred Hafner (HEAS), Simon Rukeratabaro (UR), Francesco Semeria (IIASA), Marta Tuninetti (PoliTo), Adriano Vinca (IIASA), Ackim Zulu (UNZA)

Horizon2020
European Union Funding
for Research & Innovation

This project has received funding from the European Commission's Horizon 2020 Research and Innovation Programme. The content in this presentation reflects only the author(s)'s views. The European Commission is not responsible for any use that may be made of the information it contains.

Contents

Executive summary

This report is the third scientific deliverable of the RE4AFAGRI (*"Renewables for African Agriculture")* project (or WP12 of LEAP-RE - *Long-Term Joint Research and Innovation Partnership on Renewable Energy between the European Union and the African Union* - Pillar 2). The RE4AFAGRI project seeks to support sub-Saharan African (SSA) smallholder farmers and communities to grant community-wide access to energy services and water for crop irrigation and human use with the ultimate goal of fostering rural development. RE4AFAGRI aims at providing African research institutions and public and private decision-makers with the tools and expertise necessary to operate a multi-scale modelling platform that can support the design and implementation of integrated solutions for the energy and water nexus in rural areas. In parallel, RE4AFAGRI aims at setting the ground for a multi-stakeholder discussion platform about the business models and enabling environment (policy and regulation) to promote the involvement of the private sector in water-energy-agriculture integrated solutions.

This deliverable jointly refers to Tasks 12.2 "Integration of modelling infrastructure" and Tasks 12.3 "Field data collection, calibration and validation", which are two of the core research tasks in the development of the nexus modelling platform for RE4AFAGRI. This report is a technical document detailing the implementation of four models that are used in the RE4AFAGRI nexus modelling platform (WaterCROP, M-LED, OnSSET and MESSAGE-NEST), their adaptation to the specific topics and research questions inquired in RE4AFAGRI, as well as their soft-linking. This document does not present results or findings, but rather represents a technical account of the progress in the production of model outputs and the development of open-source modelling tools that are instrumental to achieving the general aims of RE4AFAGRI and that will feed into the forthcoming capacity building and dissemination activities of RE4AFAGRI.

All the models used are open source, free to use, and peer-reviewed in previous applications and works. WaterCROP is an evapotranspiration model used to estimate the water demand for crop irrigation considering agricultural and irrigation practices, climatic conditions, and the potential for agricultural and irrigation expansion. M-LED is a bottom-up electricity demand assessment platform that produces monthly demand estimates and demand projections for different sectors (including agriculture and household uses) at a high spatial resolution. OnSSET is an energy access planning optimization tool used to assess spatially explicit electrification planning pathways at a high spatial resolution over large geographic areas. OnSSET can provide a detailed plan for electrifying settlements, identifying which areas are best supplied from standalone or mini-grid systems or through an existing or expanded grid network. NEST determines the cheapest technology mix to satisfy the national energy demand with a particular focus on grid electricity of the populated areas and industries, and combines assessments and balances of agricultural water needs, household energy needs and possibilities for supplying electricity to both. A detailed overview of the models can be found in Section 2 of this report

Amongst the socioeconomic and biophysical datasets and techno-economic parameters that are needed as inputs to the models there are common data needs as well as data needs that are model-specific. In addition, the models rely, to varying degrees, on historic and simulated data. Calibration and validation of the model data was carried out for the datasets used as input to each model including the data transferred between models, where outputs of one model in the framework is used as input to another model. For example, national electricity consumption in M-LED was calibrated to a known base year value according to recent national statistics. The demand for the population clusters in OnSSET is an output of the M-LED model, and thus it was necessary to ensure that M-LED and OnSSET are using the same population clusters. OnSSET validation also included a review of the grid network data, a parameter that has a large influence on OnSSET results and is required by NEST. This type of data transfer was necessary between all the models in the modelling framework and the calibration has therefore also ensured that data transferred between models with different spatial and temporal resolutions is consistent.

Ensuring the consistency of data and the calibration of data transferred between models has been achieved through (i) holding regular modelling meetings and workshops to ensure consistency of input data and modelling assumptions among the different modelling teams and drawing information and parameters collected from experts and focus groups through Task 12.1 into the modelling frameworks, (ii) the creation of data inventories for each model, (iii) the validation of input data via desktop methods and through consultation with in-country partners and (iv) collecting data and techno-economic parameters from stakeholders (e.g. ministry, statistical offices, and smallholder consortia), among others. A detailed description of each model's data needs and the focus of the data validation for each individual model is provided in Section 3. Full data inventories for the models used including brief data descriptions, their sources, and their relevance to other models in this project can be found in the Appendices.

The RE4AFAGRI platform follows a scenario-based logic, where a set of storylines representing different potential (or desirable) future developments happening both within the country-studies and at a global scale are designed and integrated in a coherent way into the RE4AFAGRI platform models. A limited set of "storyline"-based scenarios have been developed rather than considering a full matrix of combinations of assumptions. This choice is justified by the objective of providing public and private stakeholders (the key target users of the RE4AFAGRI modelling outputs) with easy-to-interpret results, which allow comparing a "baseline" future evolution with increasingly ambitious goals or challenging conditions.

Three key scenarios have been constructed, a *Baseline Scenario* - which represents an extrapolation of recent trends into the future based on current policies and agricultural production based on historical trends; a *Moderate Development Scenario* - where efforts are made to improve the quality of living by increasing electricity, water and sanitation access and improving food security and; an *Improving Access and Sustainable Targets Scenario* - a more ambitious scenario where universal electricity, water and sanitation access is achieved by 2030, domestic agricultural production is increased to meet the nutrition standards required by the EAT Lancet diet in 2030, and the renewables share of electricity production is targeted at 100% by 2050.

One of the key purposes of the modelling activities carried out in RE4AFAGRI Tasks 12.2 and 12.3 is to produce an online interactive model output visualisation dashboard that can be used by private and public national and regional stakeholders interested in making data-informed decisions (such as investment targeting, infrastructure sizing, or development policies design). The dashboard will be used both in capacity building activities as part of project Task 12.5, as well as in dissemination activities as part of project Task 12.6. The dashboard has been designed to allow the user to select a country, and review the results that are of interest to them, for example agriculture and water requirements in different scenarios, energy demand and electricity access planning in different scenarios and irrigation and crop processing needs in different scenarios. Users will be able to select the scenario and explore results through maps and other visual tools. Users will also be able to download results for specified geographical areas. A more detailed description and examples of the concept for the interactive dashboard can be found in Section 5 of this report.

1 Introduction

This report is the third scientific deliverable of the RE4AFAGRI (*"Renewables for African Agriculture")* project (or WP12 of LEAP-RE - *Long-Term Joint Research and Innovation Partnership on Renewable Energy between the European Union and the African Union* - Pillar 2). The RE4AFAGRI project seeks to support smallholder farmers and communities in sub-Saharan Africa gain access to the energy and water services they need for crop irrigation and human use with the ultimate goal of fostering rural development. Through research and capacity building, RE4AFAGRI aims to: build a modelling platform that can support the design and implementation of integrated solutions for the energy and water nexus in rural areas of Africa; provide African research institutions, and public and private decision makers with the tools and expertise necessary to operate the multi-scale modelling platform; and disseminate the results in an accessible web-based platform.

To achieve these overarching aims, the core actions that are pursued by the RE4AFAGRI project consist of:

- **Advancing the state-of-the-art of energy-water nexus modelling** in rural areas of developing countries to bridge the current gap between large-scale and local-scale frameworks and agricultural and electrification modelling.
- **Delivering a set of open-source validated tools** developed through both analytical and empirical approaches that can be exploited by African stakeholders in future applications.
- **Designing business models** that allow the implementation of integrated energy-water solutions in rural areas, based on identified best practices, regulatory hurdles, and opportunities through the participation of African businesses, public authorities, and smallholder consortia.
- **Providing capacity building** activities for: African research institutions to enable the use of the integrated water-energy assessment platform developed; African entrepreneurs to enable the implementation of technological solutions through the identified business models; African public administrations to establish the required policy frameworks needed for a successful replicability, scalability, and transferability.
- **Supporting policymakers** in establishing policy frameworks that ensure that the business models identified can be successfully implemented by local entrepreneurs while also meeting the technical and environmental requirements to fulfil development objectives in rural areas.

The development of an interconnected modelling platform, leveraging the insights from a range of Nexus assessment tools that can capture needs, assess opportunities, and produce results at different temporal and spatial scales, is a cornerstone of the RE4AFAGRI project. The models that are combined in the Nexus assessment tool can access agricultural water needs, provide an electricity demand estimate, and assist with planning for distributed generation of electricity at a high spatial resolution as well as for a larger scale grid-based supply network. To capture the climate-water-energy-agriculture-development dimensions four individual models are combined in the nexus assessment tool namely:

- **WaterCROP:** an evapotranspiration model to: estimate the crop water demand by source (rainfall plus irrigation) as a function of the soil moisture available in the soil; assess potential irrigation expansion (by source, surface water or groundwater bodies, and based on current yield gap).
- **M-LED:** an electricity demand assessment platform covering all main demand sectors relevant to electricity system planning, modelling these in high spatial resolution and with future projections, while also targeting communities where currently electricity supply infrastructure is lacking.
- **OnSSET:** a supply-side electricity access analysis tool to assess least-cost electrification technologies and investment requirements based on electricity demand from different sectors, high-resolution population distribution, local energy potentials and distances to existing infrastructure.
- **NEST:** a framework for optimizing long-term, multi-scale energy–water–land system transformations and achieving sustainable development objectives

Figure 1 below presents schematically the individual models and their interconnections, showing the transfer of data and contribution of each model towards the nexus assessment and planning tool.

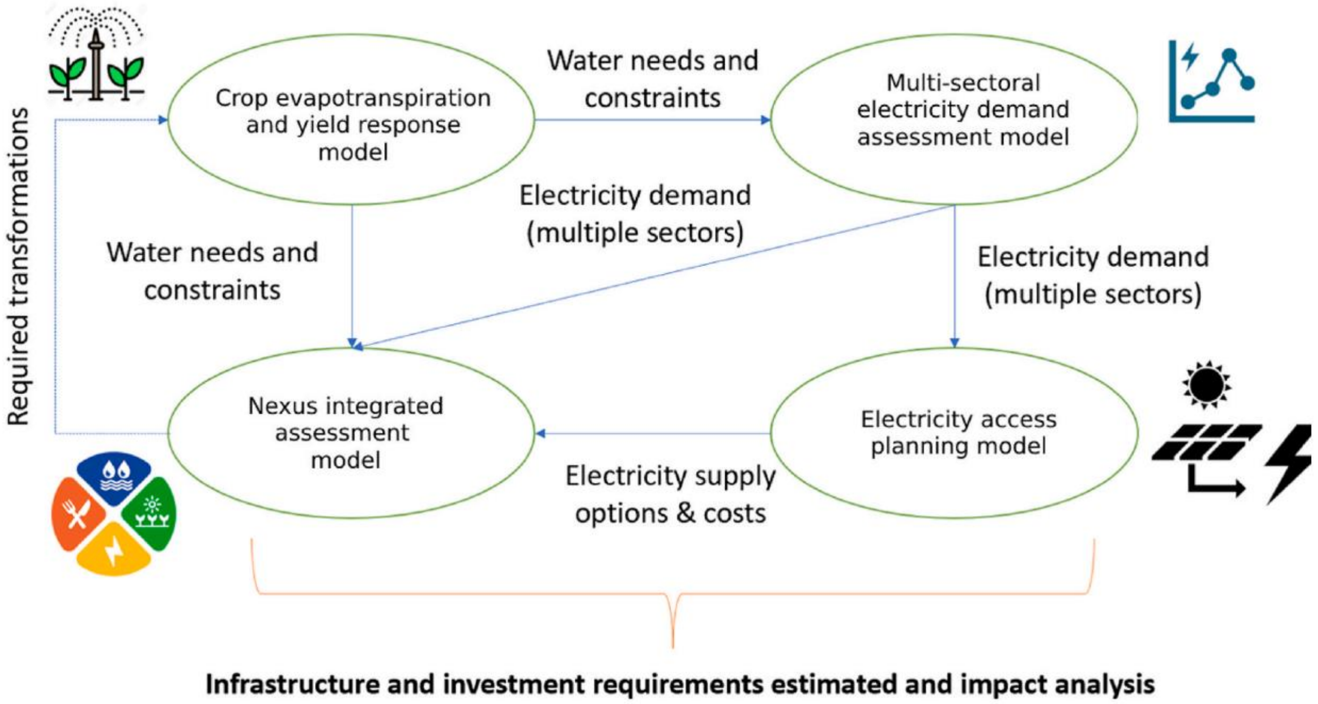


Figure 1: Models and modelling interconnections in the nexus analysis and planning tool (Source: Falchetta et. al., 2022)

The nexus assessment tool has been developed with the objective of being able to be adjusted and calibrated to be useful at various scales and throughout Africa. The focuses of Task 12.3 under the RE4AFAGRI project is on ensuring that the models and methodologies employed in the nexus assessment tool are fit for purpose and that they can achieve the ultimate goals of the project. Evaluating the accuracy and the level of insight provided by the integrated modelling platform is an important component of this process and includes ensuring that all model data are aligned between models and correctly calibrated; reviewing the sensitivity of model results to the input data in each model; and adjusting parameters and approaches to tailor the modelling framework to the project's aims. To facilitate this and ensure that the platform can be applied and produce relevant and reliable outputs in various contexts several countries have been adopted as case studies in the model development and four have been included in the stakeholder interviews to inform the structure and data within the models.

As the models analyse multiple overlapping, adjacent, or distinct fields, the data used is varied and obtained from a wide range of sources. The focus of Task 12.3 has also been to ensure consistency between the data used in the individual models and across the scenarios, and that the insights and results developed are useful in achieving the ultimate aims of the project. This requires the use of common data sources and parameters for socioeconomic and biophysical datasets and techno-economic parameters, and ensuring that these are calibrated to correctly capture a known baseline for the area under analysis as accurately as possible; verified to ensure that the models are all working well together and as expected by the project teams; and validated to ensure that they are providing meaningful results in service of the shared ultimate goals of the project.

This task has been achieved through (i) holding modelling meetings and workshops to ensure consistency of input data and modelling assumptions among the different modelling teams and drawing information and parameters collected from experts and focus groups through Task 12.1 into the modelling frameworks), (ii) the creation of data inventories for each model, and (iii) the validation of input data via desktop methods and through consultation with in-country partners (iv) collecting data from stakeholders (e.g. ministry, statistical offices, and smallholder consortia).

This report provides (i) a detailed characterisation of the RE4AFAGRI integrated modelling framework; (ii) individual summaries of the four models - including their functioning, spatio-temporal resolution, input data, main assumptions, and outputs; and (iii) an overview of the models integration, both in terms of principles and of actual technical implementation

The report then illustrates the RE4AFAGRI Approach to Model Data Calibration, Verification and Validation, both within each individual model and across the platform, i.e., in terms of cross-model consistency.

Next, the report focuses on scenario development, namely it introduces the matrix of scenarios considered in the first round of model runs, including both the socio-economic and climate change pathways assumed in each scenario, and the policy targets in terms of sustainable development objectives embedded in those scenarios.

Finally, a section highlights the expected outputs and relevance of the modelling platform (including the online dynamic visualisation interface currently under development) both within the projects aims (i.e., feeding into the activities of the other project tasks), and for the target users (in research, practice, and policymaking spheres) that the project is seeking to build capacity for.

2 Integrated modelling framework: descriptions of individual models and integration overview

2.1 WaterCROP

2.1.1 Background and fundamentals

WaterCROP is a high-resolution crop water footprint model developed by Tuninetti et al. (2015), which evaluates the water footprint of multiple crops at the spatial resolution of 5 × 5 arc min. The model performs separated water footprint computations by rainfed and irrigated production conditions and by growing season. WaterCROP performs spatially-explicit estimates of daily crop-specific actual evapotranspiration ($ET_{a,j}$) that a crop loses via evapotranspiration throughout the growing season (Tuninetti et al., 2015). For rainfed production, WaterCROP computes the water stress coefficient ($k_{s,j}$) through a daily steady state water balance. In this case, every time water from precipitation is not sufficient for optimal evapotranspiration, the crop becomes stressed, and the water stress coefficient drops below 1. For irrigated production, the model assumes that the crop receives all the water required to optimally evapotranspire via irrigation, even when water is not available from precipitation. Hence, the water stress coefficient is equal to 1 throughout the growing period (Tuninetti et al., 2015). Taking the sum of the daily $ET_{a,j}$ values for the entire growing season gives the annual actual evapotranspiration (ET_a) estimate for a crop in a grid cell. Therefore, ET_a is evaluated over different time intervals and for each study crop, thus providing multiple scenarios of crop water requirement. Furthermore, WaterCROP computes the green ($ET^R_{a,g}$ and $ET^I_{a,g}$) and blue ($ET^I_{a,b}$) shares of the crop actual evapotranspiration over the growing period, in order to evaluate the distinct contributions of precipitation (green) and irrigation (blue) water to the crop water footprint. Thus, the crop water footprint is provided as disaggregated in its green (uWF_g), blue (uWF_b), rainfed (uWF_{rf}), irrigated (uWF_{irr}) components and it is evaluated both in volume (m^3) (WF) and per unit of production ($m^3 \cdot ton^{-1}$) (uWF).

Present day cropland distribution data is derived from the *SPAM 2017 v2.1 Sub-Saharan Africa* product, providing spatially-disaggregated maps of agricultural land and yield for 42 crops (IFPRI, 2020). In addition, field size data, used to filter cropland cultivated by smallholder farmers and excluding large-scale farming, is sourced from Fritz et al. (2015). Historical wholesale farmgate crop prices data by country are obtained from the FAOSTAT database (FAO, 2017), parsed to the 42 crops in the MapSPAM database, and the median price (in USD) in the 2010-2021 period is calculated for each country. For countries with missing data for certain crops a regional average of the most recent price among countries where such data are available is considered. In addition, nutritional content tables for each crop are derived from the FAO (Becker et al., 2001) and are used to estimate potential calories, protein and fats generation potential. For the historical and future climate data refer to the Hydroclimate data section.

In addition to the hydrological and crop modelling, the following additional data sources are considered: crop coefficients and the proportional length of each growing stage came from Chapagain et al. (2004). Crop-specific and spatially explicit planting dates and length of growing period came from Portmann et al. (2010). The MIRCA2000 dataset was adopted for growing period lengths, sowing and harvesting dates; although an interpolation procedure was performed to match the dataset spatial extension to that of the 2017 harvested areas. Global gridded available water content came from the Harmonized World Soil Database (Nachtergaele et al,. 2010). Crop-specific irrigated and rainfed rooting depths came from Allen et al. (Allen et al., 1998).

2.1.2 Data: spatial and temporal resolutions (present scenario)

Spatially distributed (5 arc minute; 1/12°; ~ 10 km at the equator), crop-specific information on rainfed and irrigated yields (ton · ha) and harvested area (ha) were sourced from the Global - Agro Ecological Zones (GAEZ v4) database. The AEZ methodology estimates potential yields for historical, current and future climatic conditions, the last of which are simulated by five GCMs - GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC-ESM-CHEM, NorESM1-M - and four RCPs (FAO, IIASA, 2021). While GAEZ provides actual statistics as fresh/harvest weight, attainable yields are available as dry weight for most crops, as sugar weight for sugar cane or as lint weight for cotton. Therefore, conversion factors, as indicated in the model documentation (FAO, IIASA, 2021), have been applied to convert actual statistics, relative to the 2010 scenario, to dry weight. The conversion coefficients depend on moisture content of harvested products and, in some cases such as sugar crops, are derived from technical extraction rates (FAO, IIASA, 2021). The same procedure was applied to FAOSTAT crop statistics as they are provided at fresh weight. The MIRCA2000 (Portmann et al., 2010) dataset was adopted for growing period lengths, sowing and harvesting dates; however, a discrepancy was observed between the spatial coverage of this data set and the crops' harvested areas extension in 2010. Therefore, an interpolation process was performed to extend planting dates and growing period information to the coverage of GAEZ harvested areas data set. Crop coefficients, soil properties, soil available water content, root zone depth and the depletion fraction were derived from a previous study (Tuninetti et al., 2015).

2.1.3 Estimates of crop actual evapotranspiration.

The crop water requirement, or actual crop evapotranspiration (ET_a) quantifies the amount of water (in mm·day^{-1}) that a crop loses via evapotranspiration throughout the growing season. It is a function of climatic and phenological properties as well as agricultural practices (e.g., irrigated versus rainfed agriculture). We developed a gridded algorithm derived from Tuninetti et al. (2015) to perform spatially-explicit estimates of daily $ET_{a,j}$ (where j runs from the planting day to the harvesting day), incorporating important updates in the computational efficiency (to accommodate a time-varying assessment of crop-specific evapotranspiration). Following Tuninetti et al., daily reference evapotranspiration values (ET_0) were determined through a linear interpolation of monthly ET_0 data (Harris et al., 2020), with monthly averages assigned to the middle (i.e., the 15th day) of each month.

$ET_{a,j}$ was calculated following the FAO56 method (Allen et al., 1998; Pereira et al. 2020), where the $ET_{a,j}$ estimate is equal to the product of: the daily water stress coefficient ($k_{s,j}$) - a proxy for the daily water deficiency in the unsaturated soil layer; the daily crop coefficient ($k_{c,j}$) which integrates the effects of crop height, crop-soil surface resistance, and albedo of the crop-soil surface (Pereira et al. 2020); and the daily $ET_{0,j}$ from a hypothetical well-watered grass surface with fixed crop height, albedo and canopy resistance.

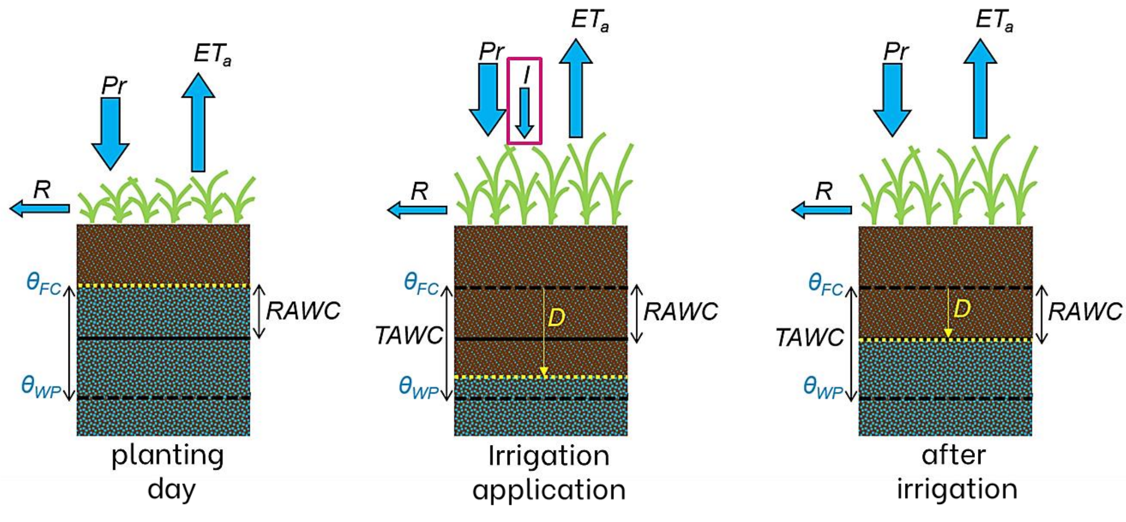


Figure 2: The WaterCROP model. Scheme of the main water flows and soil water contents (Pr: precipitation, R: surface and sub-surface runoff, I: irrigation, ETa: actual evapotranspiration, D: depletion, θFC: soil water content at field capacity, θWP: soil water content at wilting point, RAWC: readily water content that can be used by the plant, TAWC: total available water content in the root zone.

Figure 2 shows a schematic representation of the WaterCROP model. On the planting day, the water content in the root zone is equal to the field capacity meaning that there is enough water for the crop to evapotranspire and grow at its maximum potentials. From day 2 on, some water stress can compromise the ability of the crop to grow. The condition of water stress happens when the soil moisture in the root zone decreases below field capacity and become less than the readily available water content (*RAWC*). To determine the magnitude of the water stress, we estimate the water stress coefficient ($k_{s,j}$) in each grid cell and per crop. For rainfed production, the computation of the water stress coefficient was performed through a daily steady state water balance as in Tuninetti et al. (2015). In this case, every time water from precipitation is not sufficient for optimal evapotranspiration, the crop becomes stressed and the water stress coefficient drops lower than 1. For irrigated production, we assumed that the crop receives all the water required to optimally evapotranspire every day (via irrigation), even when water is not available from precipitation. Hence, the water stress coefficient is equal to 1 throughout the growing period. Taking the sum of the daily $ET_{a,j}$ values for the entire growing season gave the annual ET_a estimate for a crop in a grid cell.

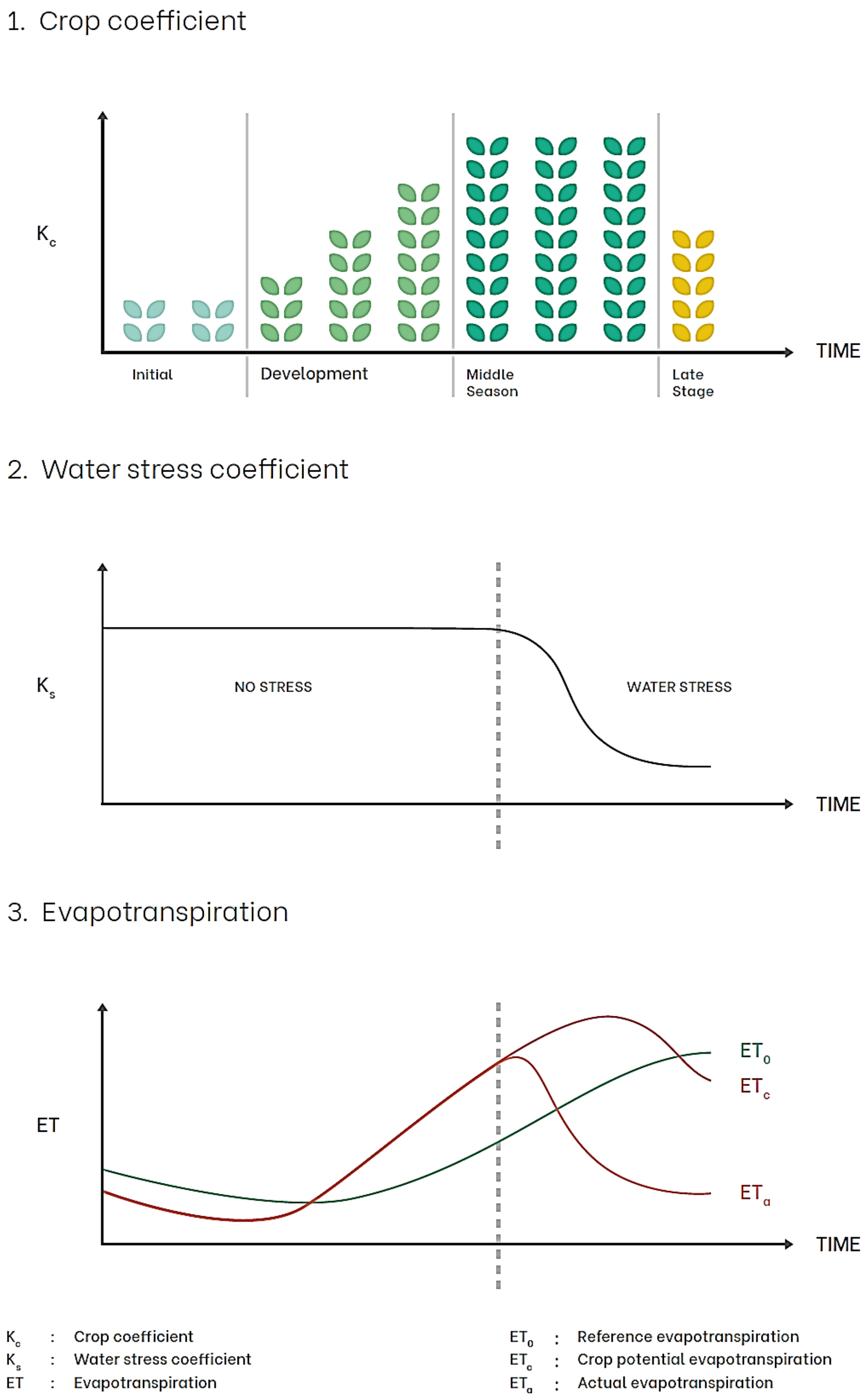


Figure 3: The key factors of crop actual evapotranspiration

2.2 M-LED (Multi-Sectoral Latent Electricity Demand assessment platform)

2.2.1 Background and fundamentals

M-LED is a bottom-up electricity demand assessment platform. Originally introduced in Falchetta et al. (2021), M-LED exploits many geospatial data sources and (sub)-national statistics to produce monthly estimates of electricity demand in a set of sectors across a country. The key benefit of the bottom-up methodology is that the output data can both be used at the native local level of analysis - i.e. communities and settlements (also called population clusters) - and be aggregated to produce sub-national or national estimates of trends in electricity demand.

M-LED is designed to operate at the country-level - calibrating current electricity consumption levels with recent national statistics and downscaling them at the local population cluster level.

M-LED is written in the R scientific computing programming language, and at the time of the writing of this report it is available for public download and use in its v1 version at https://github.com/giacfalk/M-LED/branches (the current update being used for RE4AFAGRI (v2) is available at https://github.com/giacfalk/RE4AFAGRI_platform/tree/main/M-LED, and will be made publicly available in line with the Deliverable 12.4 in Q1 2023) at the official RE4AFAGRI platform repository, i.e. https://github.com/iiasa/RE4AFAGRI_platform

M-LED is largely relying on publicly available datasets, which can be downloaded through an internal wrapper function upon the first run of M-LED on a local computer.

As a bottom-up assessment platform, M-LED performs calculations at the most granular level allowed for by the input data. M-LED demand projections are both drivers-driven, i.e. determined e.g. by the projected growth in population and economic affluence levels, and objective (or policy) driven, i.e. aimed at ensuring that sufficient electric energy to cover needs such as powering pumps for irrigation, crop processing machinery, or appliances in schools and healthcare facilities is provided.

2.2.2 Spatio-temporal characterisation

M-LED estimates electricity demand at the population settlement (namely, community) level. Population settlements are polygons enclosing a settlement such as a city or a village. There are different approaches to generate population clusters for a country, while a set of different data products of this type are also available, such as from GRID3 (https://data.grid3.org/) or from Khavari et al. (2021). Clusters include residential buildings where populations are living, as well as other sectors such as SMEs and healthcare and educational facilities. Clusters are then linked to the surrounding agricultural land (mostly not inhabited land devoted to agriculture) through a Voronoi polygonisation of land area approach based on each population cluster centroid, see https://en.wikipedia.org/wiki/Voronoi_diagram**).** This approach allows linking demand for agriculture from cropland surrounding population clusters to the estimated demand in population clusters (**Figure 4**). A maximum distance parameter is available in M-LED (with default value of 5 km) beyond which the load for irrigation is classified as “off-grid demand”, i.e. demand that is unlikely to be served from the main electricity supply system(s) serving the population cluster community.

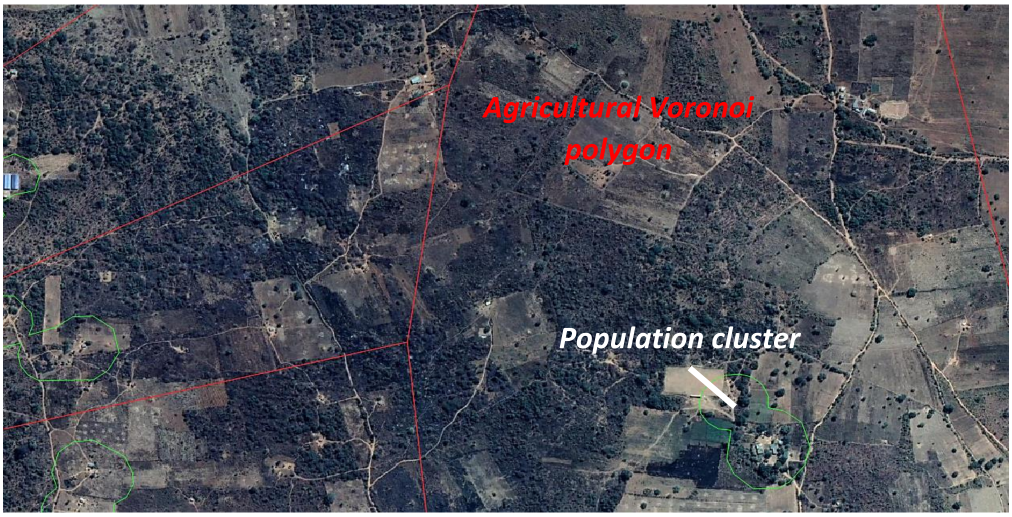


Figure 4: example of population clusters (in green) inside agricultural cluster Voronoi polygons (in red)

Also, for crop processing a similar approach is considered, where travel-time based catchment areas around cities (i.e. crop markets) are calculated (as exemplified in **Figure 5**), thus generating polygons that describe the area located within a user-defined number of minutes of travel time (considering the fastest route and mean; see Weiss et al. 2019 and 2017 for the underlying approach) from the city. In particular, a threshold of 180 minutes (3 hours) is considered. Clusters whose centroid is falling outside of those urban catchment areas are not deemed suitable to have a crop processing electricity load, as their distance to market makes it likely economically unprofitable to perform crop processing / storage at those locations.

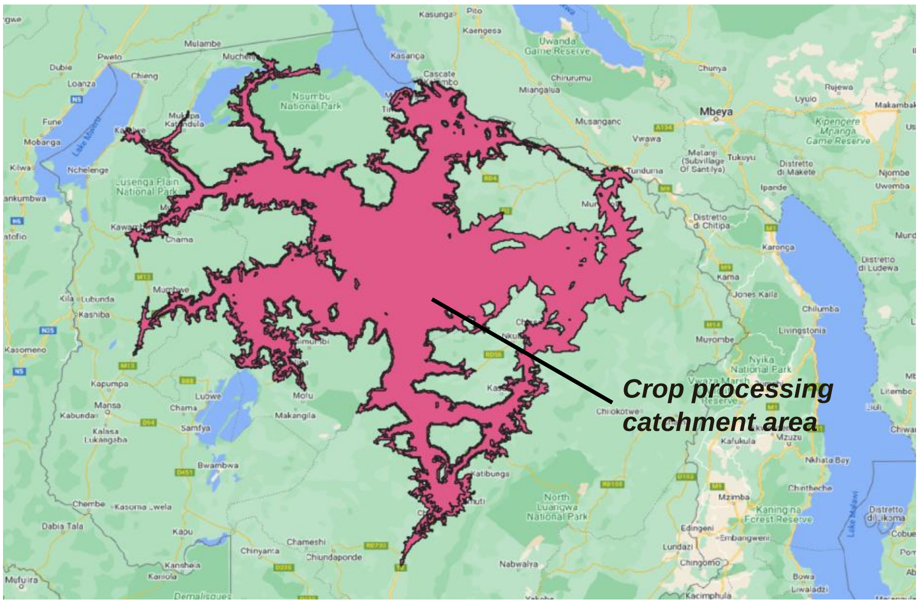


Figure 5: A crop processing catchment area (in pink). The polygon represents an approximation of the area within 180 minutes of travel time from major cities. Cropland area outside these buffers is classified as not eligible for crop processing energy demand as access to markets is too limited.

With regards to the time characterisation of M-LED, the model operates at 10-year timesteps starting from the base year (2020) at which the current electricity demand is calibrated, and reaching the target year (2060) by recursively projecting demand across time-steps. For every sector, demand is estimated both at the yearly and at the monthly level, to incorporate the role of seasonality and inter-annual variations. Irrespective of M-LED outputs being harmonised to the monthly and yearly scales, electricity demand in certain sectors is first defined at higher temporal resolution (e.g. water pumping or residential, healthcare and school demand are calculated bottom-up, starting from the hourly profiles of utilisation) and then aggregated to the monthly scale, but also at lower temporal resolutions, such as in the case of mining demand, which is then equally redistributed among months of the year.

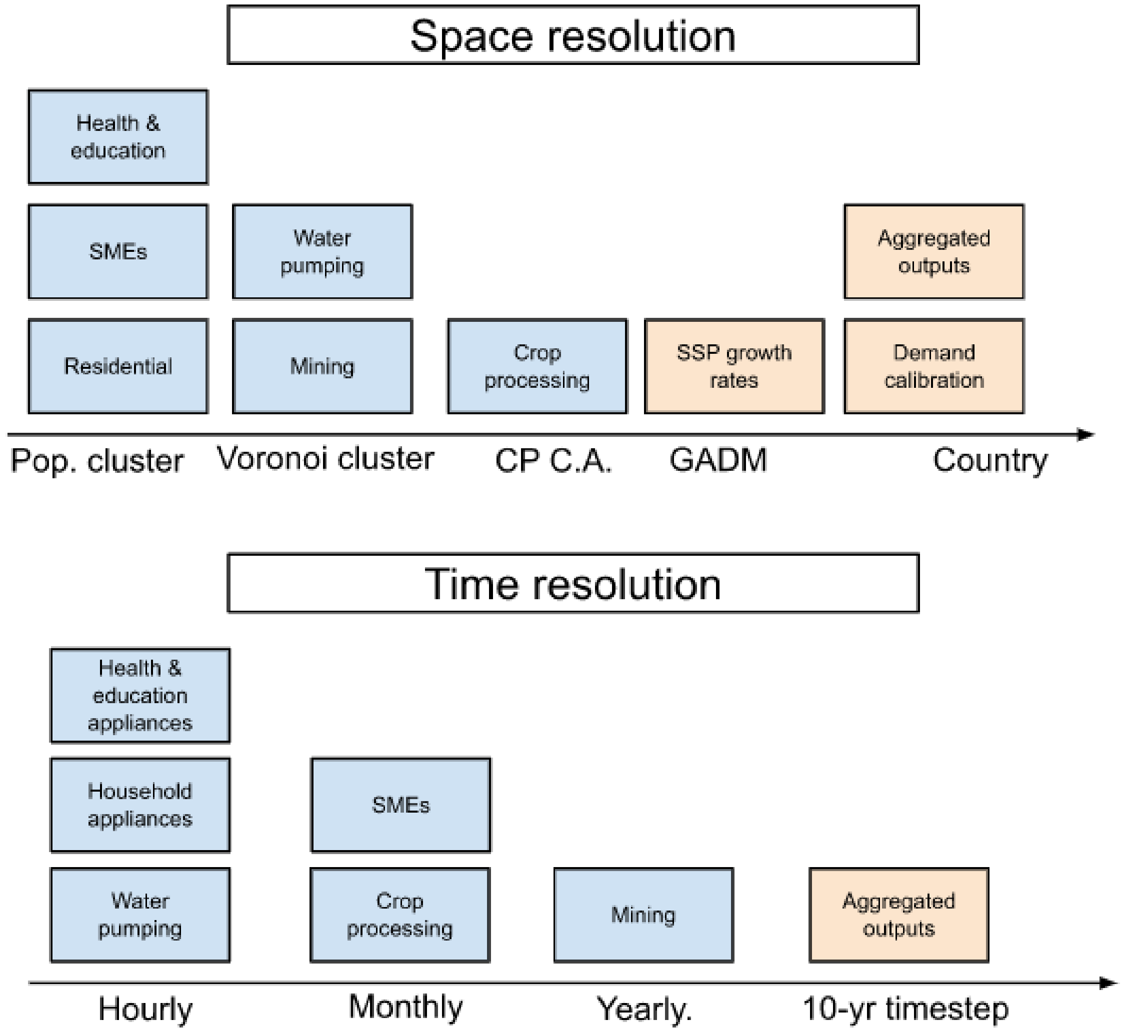


Figure 6: Conceptual representation of the sectoral spatio-temporal representation in M-LED

2.2.3 Structure

M-LED has a modular structure, disaggregated into three main types of modules:

1. Backend modules (libraries, working directories, and technical parameters definition)
2. Scenario module (specific to the country in question, also containing the specifics of the scenarios that the user wants to run)
3. Modelling modules (the actual code performing data and model operations to produce electricity demand estimates).

2.2.4 Description of modules

- **M-LED.R** is the wrapper of M-LED, where the basic technical and run-specific parameters are user inputted, the scenarios are defined, and the actual model modules are launched for the desired scenarios.
- **backend.R** is the module responsible for installing and loading the required libraries to run M-LED, as well as to download and/or collect all the input data for the specified country run. The user is not required to intervene on this file.

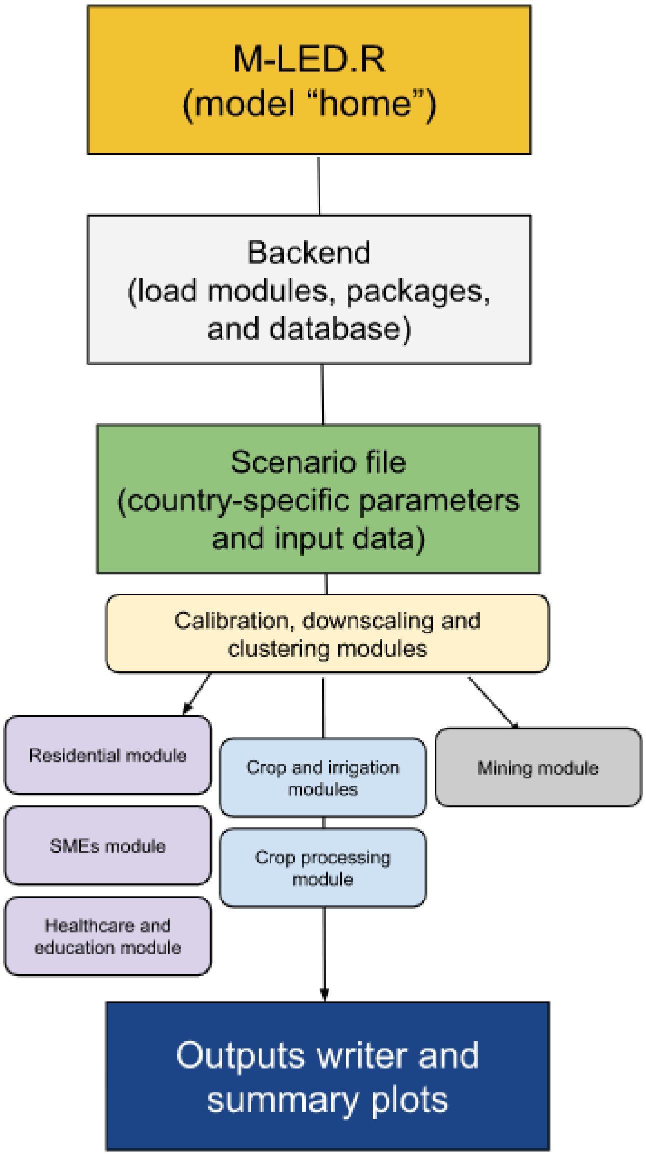


Figure 7: Structure of M-LED v2

- ***calculate_rates_for_baseline.R*** is the module responsible for calculating what are the projected future trends in the electricity access rate and power irrigated cropland share if the baseline trend is followed, i.e. if no policy target is defined.
- **scenario_*countryname*.R** is the main file where the user is required to manually intervene. At this location, all the input data subsequently called by the different M-LED modules are called; country-specific parameters for model calibration and affecting demand estimation are defined, such as national electricity demand in different sectors; technological parameters are defined; and, country-specific input data are read. The current v2 implementation of M-LED includes scenario files for four countries:
    - Zambia (operational)
    - *Rwanda (under development, ETA Q1 2023)*
    - *Zimbabwe (under development, ETA Q1 2023)*
    - *Nigeria (under development, ETA Q1 2023)*
    - *Kenya (under development, ETA Q2 2023)*

Capacity building activities in the forthcoming activities of Task 12.5 will be focused on both expanding the pool of autonomous users from these countries and in enlarging the number of supported countries by training interested researchers and stakeholders to develop their own country modules.

The scenario_*countryname*.R is itself **wrapping three additional sub-modules**, all fully automated and thus requiring no user intervention:

  - The **projector.R** script, which project the key macro drivers (population, GDP and GDP per capita, urbanisation) within each population cluster using SSP-consistent downscaled datasets and the inferred growth rates (calculated at the GADM2 administrative level boundaries).
  - The **urbanisation_calibration.R** script which calibrates clusters urban status to national SSP urbanisation projections up to the target year 2060.
  - The **rwi_to_gdp_capita.R** script which converts the Relative Wealth Index information to an Absolute Wealth Estimate following procedures described in Hruschka et al. (2015), which in turn is used to calibrate the current GDP per capita at the cluster level.

- The **demand_growth_weights.R** module calibrates the pace at which policy objectives that are imposed in each scenario for a given target year (e.g., universal electricity access; % of irrigated cropland; % of crop yield locally processed) are reached. The evolution of demand towards these policy goals is thus non-linear and depends on the evolution of per-capita GDP in the socio-economic scenario (SSP) of reference.
- The **electricity_access.R** module estimates the current share of people with or without electricity in each local population cluster using satellite-based nighttime light information (following the methods described in Falchetta et al. 2019).
- The **create_clusters_voronoi.**R module creates Voronoi polygons around the centroid of each input population cluster polygon. These polygons are used to perform cropland-related spatial data extraction operations, under the assumption that cropland from which electricity demand is driven (for water pumping and irrigation and consequent crop processing) at each cluster is the cropland that is closest to each cluster centroid than any other cluster centroid. The maximum distance of cropland to be considered as potential driver of electricity demand associated with population clusters is capped by the *m_radius_buffer_cropland_distance* parameter (see **Table A 1** below). Cropland beyond this distance is classified as eligible for standalone water pumping and is estimated as a separate sector.
- The **crop_module.R** module calculates the monthly and total crop-specific irrigation needs based on the WaterCrop estimates entering M-LED as input data and on the MapSPAM 2017 SSA rainfed and irrigated cropland areas and yield data, as well as constraints such as environmental flows preservation.
- The **pumping_module.R** module then estimates the monthly power and energy requirements to pump water from either surface water bodies and rivers or from groundwater aquifers net of the crop-specific irrigation technology efficiency as well as technological assumptions and parameters of the pumping model. The module then selects whether at each cluster it is optimal to develop surface or groundwater irrigation.
- The **crop_module_solar_pumps.R** module is a twin file of the **crop_module.R,** except it quantifies irrigation water demand from cropland that is more distant from each population cluster centroid than the distance determined by the *m_radius_buffer_cropland_distance*
- The **pumping_module_solar_pump.R** module is also a twin file of the **pumping_module.R,** quantifying the monthly power and energy requirements to pump water for irrigating cropland that is more distant from each population cluster centroid than the distance determined by the *m_radius_buffer_cropland_distance.*
- The **mining_module.R** module downscales the national industry demand from the mining sector (calibrated exogenously through the *industry_final_demand_tot* parameter) based on mining sites data and nighttime light radiance. It then matches mining sites to the closest population clusters. It projects demand growth in the sector proportional to GDP per capita growth rates.
- The **residential.R** module estimates household electricity demand from both:
  - People already benefiting from electricity access in the base year (2020) or at each future timestep, for whom future electricity consumption grows proportionally to locally projected GDP

per capita growth rates mediated by an income elasticity of electricity demand coefficient that is itself proportional to the GDP per capita level.
    - People who will gain electricity access at a given timestep, whom are first allocated to a given tier of electricity access, estimated through a statistical model based on current estimated tiers of electricity access among populations with electricity access, and then are attributed a given demand based on the tier-specific, urban-rural differentiated appliances-basked generated load profiles from the RAMP model outputs (Lombardi et al. 2019). See Falchetta et al. (2020) for a more detailed description of the RAMP model linkage into M-LED and the definition and calculation of appliance baskets and utilisation patterns in the context of electricity demand estimation.
- The **health_education_module.R** module estimates electricity needs for achieving given appliance use standards in currently existing healthcare and educational facilities. Proportional to the number of beds (for healthcare facilities) or to the number of pupils (for schools), facilities are mapped to a given tier/size. Then, similarly to the residential sector, facilities are linked to a given demand based on tier-specific differentiated appliances-basked generated load profiles from the RAMP model outputs. In addition, facilities are densified in the future based on local population growth rates across time steps and critical population thresholds for facilities uptake or densification.
- The **crop_processing_catchment_areas.R** module creates polygons surrounding each urban centre in the country under analysis based on the minutes_cluster parameter. This parameter defines the area within the given number of minutes of road travel time for each city, and it considers all the crop processing demand occurring within this area as feasible, while nulling that outside of those areas. In other words, it forces crop processing electricity demand to be estimated in areas that are sufficiently close to markets where products can be sold.
- The **crop_processing.R** module estimates electricity demand for crop processing and vegetables cold storage. Energy needs are depending on the crop processing energy needs database csv file, which then are multiplied by current and future potential crop yields in the corresponding crop-specific harvest period(s). Crop processing demand is then constrained to only occur in clusters within the urban markets crop processing catchment areas discussed above. A schematic representation of the local machinery requirements (crop-specific) is also provided.
- The **other_productive.R** module estimates non-farm SMEs electricity demand based on the estimated residential electricity demand at each cluster and a markup range (*range_smes_markup* parameter) varying based on the roads density and distance to nearest city of each cluster.
- The **cleaner.R** module
- Finally, the **calculate_yield_growth_potential.R** module calculates the potential growth in the yield thanks to the input of irrigation at each timestep.

Eventually, M-LED goes back to the M-LED_hourly.R script, where outputs at different spatial scales are written (also serving as input data for the other soft-linked models in the RE4AFAGRI platform), and the summary csv files and results are generated for each scenario.

2.2.5 Model calibration and demand projection

M-LED is designed to operate at the country-level. Thus, before projecting future demand, it is calibrated to current electricity consumption levels with recent national statistics and downscaling them at the local population cluster level. In particular, calibration occurs based on the following values and statistics:

- National electricity access level (ESMAP)
- National electricity demand (IEA)
- National electricity demand, residential (IEA)
- National electricity demand, industry (IEA)
- National electricity demand, other sectors (IEA)

The rationale behind how demand projections are made depends on the sector and typology of final users. For the **residential sector**, demand is divided between:

- Households that are already consuming electricity in the base year (2020) – with demand following GDP per-capita assumption of each scenario conditioned on a flexible income elasticity of electricity demand schedule
- Household that will gradually gain access to electricity after 2020 – with demand being based on representative tiers of access estimated through the RAMP appliance-based stochastic model (Lombardi et al. 2019), and tiers being parsed to each population settlement based on a set of determinants

In the non-residential sectors, demand is projected as follows:

- **SMEs:** the demand from non-farm SMEs is proportional to the demand in the residential sector at each cluster, modulated by a range factor (30%-60%) which - at each cluster - is proportional to the PCA indicator of employment rate (from DHS survey data) and roads density (from GIS road maps) of each cluster, a proxy of the labour and economic situation of a given community. Future SMEs demand then evolves at the same growth rate of residential demand at each cluster.
- **Mining:** the demand from mining is spatially downscaled from the national electricity demand from the industry sector (IEA) onto the mining sites geodatabase (Maus et al. 2020) proportional to the measured nighttime light radiance (Colorado School of Mines, 2022) from those sites. Mining demand at each mining site - parsed to nearby population clusters - is then projected to grow at the local (administrative unit of level 2 of belonging) growth rate of per-capita GDP according to the specific SSP scenario considered in the run.
- **Health & education:** health and education demand is projected to converge to a universal electrification of all facilities by a given target year, inclusive of a densification and enlargement of existing facilities following local (administrative unit of level 2 of belonging) population growth projections according to the specific SSP scenario considered in the run.
- **Crop processing:** crop processing demand is projected to converge to a given target of crop yield throughput processing by a given target year, subject to growth in crop yields thanks to growing irrigation as well as subject to the spatial constraints imposed by the crop processing catchment areas.
- **Irrigation water pumping:** water pumping energy demand follows pathways of irrigation water demand derived from the WaterCROP model.

2.2.6 Model outputs

In the v2 implementation, for each scenario run, M-LED writes output data at four levels of aggregation, serving both model interlinkage purposes as part of the RE4AFAGRI modelling platform, interactive visualisation purposes in the RE4AFAGRI dashboards under development, and as static output summary files:

1. OnSSET output geopackage, containing demand for all the original population cluster, the unit of analysis of M-LED
2. NEST output geopackage, aggregating the population clusters results at the NEST nodes level. In particular two files are written as NEST outputs, one total and one urban/rural stratified output for each NEST node.
3. GADM level 2 output geopackage, aggregating the population clusters results at the second level of administrative boundaries; useful for visualisation of aggregated results in the online dashboards and for informing policymakers
4. Summary CSVs and figures of results aggregated at the country level, disaggregated by sector, scenario, and year

The output geopackages are reporting monthly demand for each timestep and each sector in kWh/month, while the summary csv files are reporting units in TWh/year.

Figure 8: Example national summary of M-LED v2 output (Zambia, alpha runs)

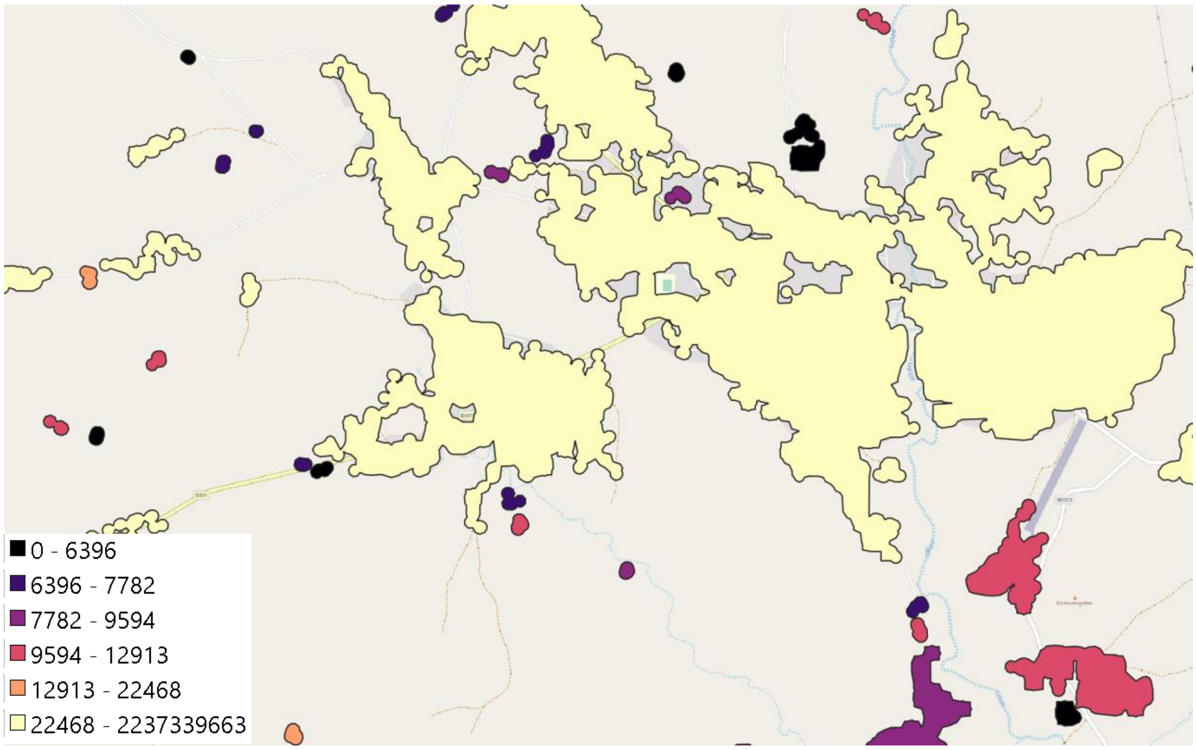


Figure 9: Example of local M-LED v2 output: residential demand in 2050 (kWh/yr in each cluster)

2.3 OnSSET (Open-Source Spatial Electrification Toolkit)

2.3.1 Background and fundamentals

OnSSET (the Open-source Spatial Electrification Toolkit) is an energy access planning optimization tool written in Python that allows the user to estimate, analyse and visualize, cost-effective, electrification pathways in high spatial detail. As it is a spatially explicit model, when calculating the optimal pathway, OnSSET considers datasets that include information on population distribution, the availability of local energy resources such as solar, wind, and small hydropower, the distance of each population cluster to the current grid networks, landcover, and technology investment and operating costs, among others. The least-cost solution presented in the model compares the cost of supplying electricity to settlements either through a central grid connection, stand-alone "solar home systems" providing energy to a single dwelling or electrifying a community using mini-grid systems (here hybrid combinations of solar, battery, diesel, hydro and/or wind). The options are visualised in **Figure 10**.

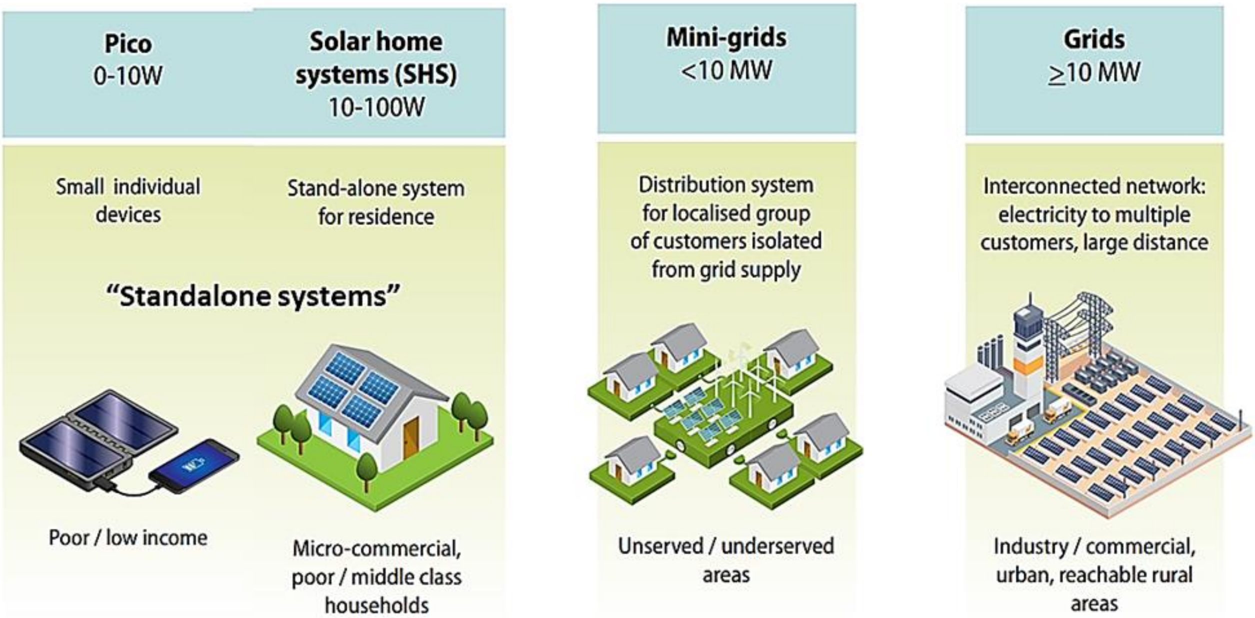


Figure 10: Different energy technology options for electricity access and their typical end-users

2.3.2 Modelling process in OnSSET

The selection of the cheapest electrification solution for each populated area requires first the collection of a significant amount of location-specific data across the area of interest. This data relates to the population and demand in the settlement, the availability of energy resources, landcover, and protected areas, as well as the distance to existing infrastructure (such as roads and existing grid network).

In the first step of the analysis, it is determined where the people with current access to electricity are: if a populated area is close to the existing electric network and there is light during the nighttime, it is considered as currently electrified via the electric network. Furthermore, locations of existing mini-grids and stable nightlights in isolated areas are used to identify additional settlements that are already electrified with mini-grids. A simplified example is illustrated below in **Figure 11**.

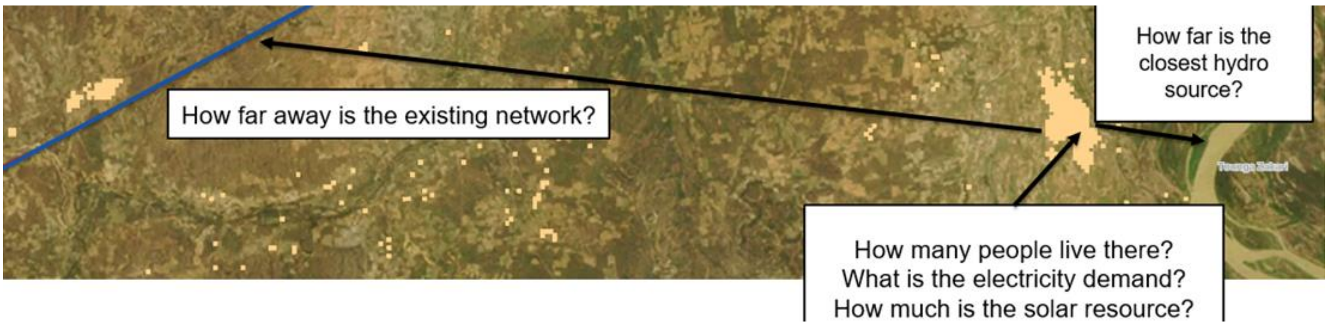


Figure 11: Data collection and calculation for each population settlement (Khavari, 2022).

In the next step shown in **Figure 12**, an electrification technology is selected for each populated area to meet the electricity demand (which is estimated as described in the previous section). This means that all households in each populated area are allocated the same electrification technology.

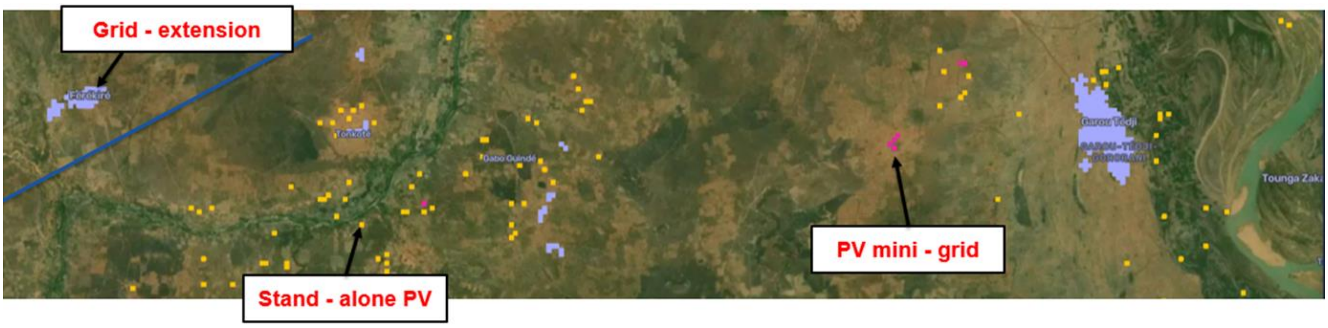


Figure 12: OnSSET least cost electrification technology option selection per settlement (Khavari, 2022).

The technology solution selected depends on the cost of providing electricity, or more precisely, the levelized cost of energy (LCOE). This is the cost that each unit of electricity would need to be sold for to cover all potential costs e.g., capital costs, interest on loans, insurance, installation and logistical costs, fuel, repairs and breakdowns, maintenance, and salaries, etc.) over the life of the project. The technology (grid extension, mini-grid, or solar home system) that can meet the demand at the lowest LCOE is selected in each area.

The total costs in a populated area are the sum of the costs of all individual connections and generating technologies. For solar home systems, each customer will have their own system, operating in isolation from each other. One larger generation source is used for mini-grids instead of providing solar panels for every single household. A mini-grid could be powered by solar energy, hydro or wind. Additionally, a distribution network is required to connect the customers to the energy sources. Depending on the size and demand of the area, economies of scale may lead to lower generation costs than for Solar Home Systems, which can make up for the additional costs from the distribution network. Finally, the last option is the extension of the centralised grid network. This considers the cost of the distribution network needed in the settlement, the new lines needed to connect the area to the existing network, as well as the cost of generation of grid electricity (which comes from the NEST model in this study). This overall processing flow is demonstrated schematically in **Figure 13** below.

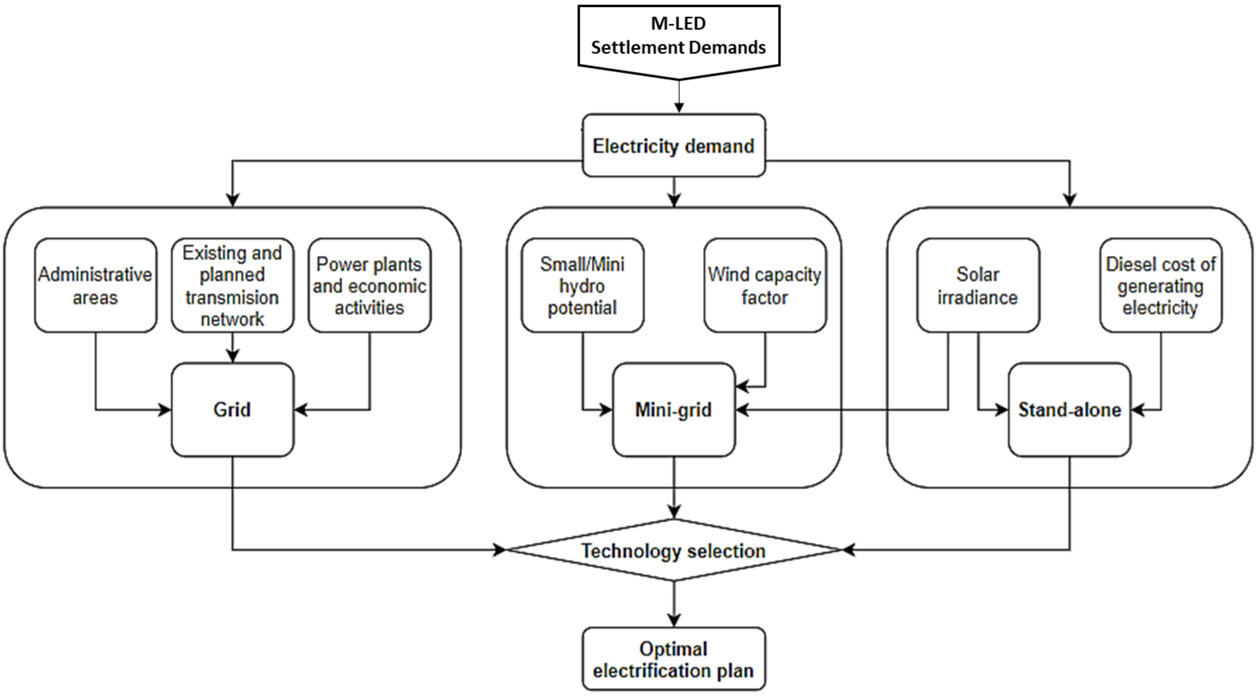


Figure 13: Processing flow of the Open-Source Spatial Electrification Toolkit (OnSSET). Adapted from (OnSSET, 2022)

This calculation and optimal technology selection is made for an intermediate and a final year, 2040 and 2060 respectively, under certain assumptions for several parameters. These parameters include the electrification target at the end year, population growth, demand growth from households, heavy industries and other businesses, and solar panel costs. When this target is not 100%, a criterium must be chosen to select which populated areas are electrified first. The default of the model prioritises households whose per capita investment cost is the lowest. Other possible criteria could be to electrify first the poorest or richest households, or those which are closer to the electric grid, or a custom prioritisation implementation.

OnSSET determines if settlements should be supplied from the grid or by off-grid technologies (and which ones), and NEST determines the cheapest technology mix to satisfy the demand for grid electricity and the estimated cost of that electricity.

2.3.3 Model outputs

OnSSET's final output is detailed information for every settlement in the country, and allocating a least-cost electrification technology, for all the scenarios to be modelled. They can be processed as large data files and summarized into key indicators of interest. The most common examples are included in **Table 1** below, and visualised dynamically on maps as show below in **Figure 14**.

Table 1: OnSSET example model output statistics and values available for each population settlement cluster

Variable	Description	Unit
Population	The population served by each technology in each cluster, in each year.	people
Start year electrification status	OnSSET estimates the percentage of a population cluster that is likely to have electricity access in the start year based primarily on nightlight data coverage of population, but also distances to existing electricity infrastructure or roads if nightlight data is insufficient. These values are calibrated to match national statistics, and for rural and urban areas separately.	% electricity access
New Connections	The number of newly electrified population by each technology in each year.	people
Capacity	The additional capacity required to fully cover the targeted demand in each year.	kW
Investment	The capital upfront investment required by each technology to reach the electrification target in each year.	USD
Technology mix	Generation mix contributions per technology as calculated by the OnSSET analysis.	% per tech
LCOE	The LCOE achieved in each location, for each technology, as calculated by the OnSSET analysis.	USD/kWh

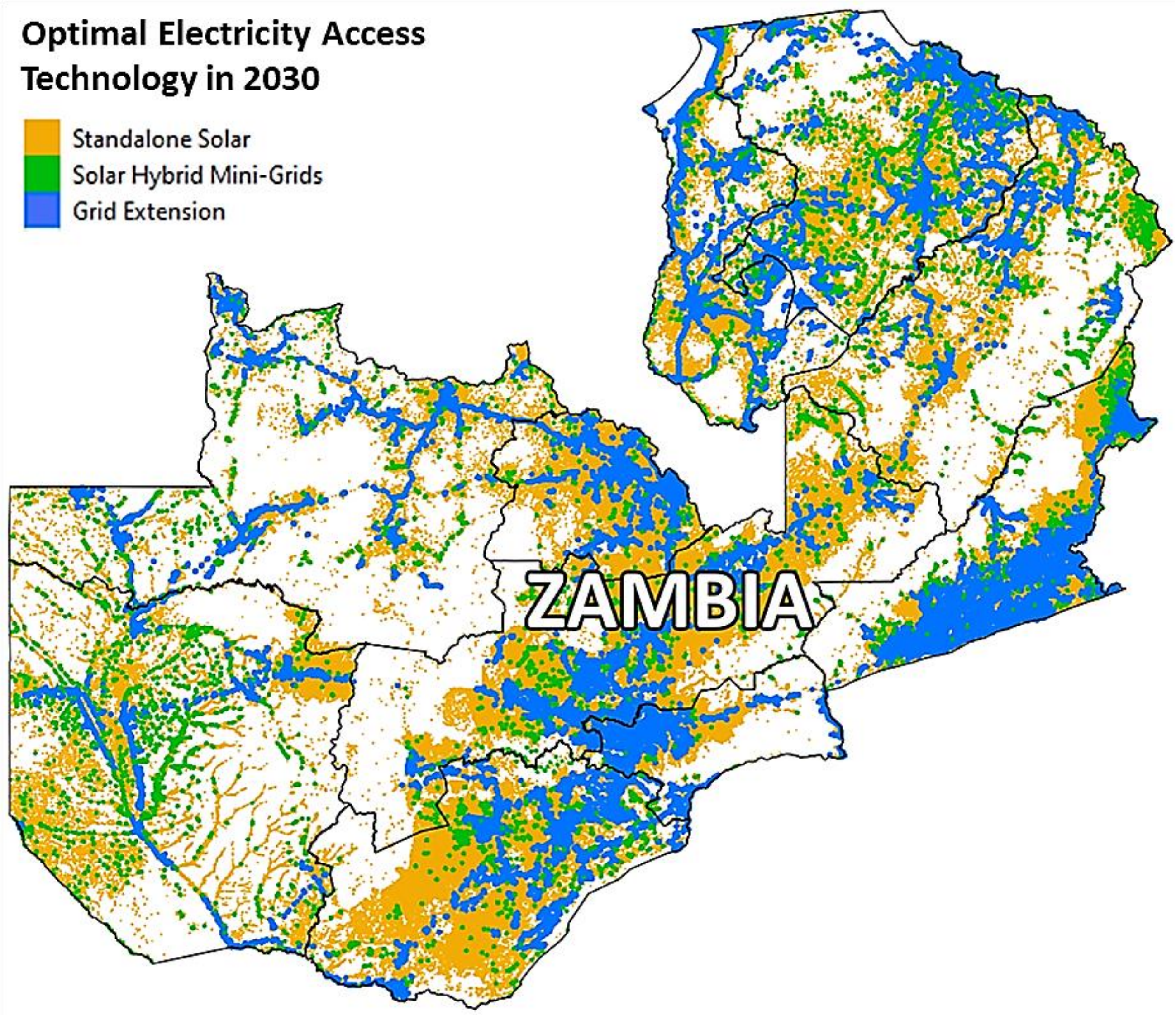


Figure 14: Optimal technology options are shown for 750 thousand individual population settlement clusters in Zambia. Blue shows where grid extension is found as cheapest, orange showing standalone solar systems, and green showing solar hybrid mini grids.

2.4 NEST (The NExus Solutions Tool)

2.4.1 Background

The NExus Solutions Tool (NEST) is an open modelling platform that integrates multi-scale energy–water–land resource optimization with distributed hydrological modelling. The approach provides insights into the vulnerability of water, energy and land resources to future socioeconomic and climatic change and how multi-sectoral policies, technological solutions and investments can improve the resilience and sustainability of transformation pathways while avoiding counterproductive interactions among sectors. NEST can be applied at different spatial and temporal resolutions, and is designed specifically to tap into the growing body of open-access geospatial data available through national inventories and the Earth system modelling community. A case study analysis of the Indus River basin in South Asia demonstrates the capability of the model to capture important interlinkages across system transformation pathways towards the United Nations' Sustainable Development Goals, including the intersections between local and regional transboundary policies and incremental investment costs from rapidly increasing regional consumption projected over the coming decades. The documentation and code for the initial case studies is published by Vinca et al., 2020, and can be found online at: https://github.com/iiasa/NEST

For this deliverable a newer version of NEST has been developed and used, which uses the MESSAGEix framework developed at IIASA and allows the generation of country models with representation of the Energy-water-land system similar to the previously developed NEST model in Vinca et al., 2020.

2.4.2 Structure

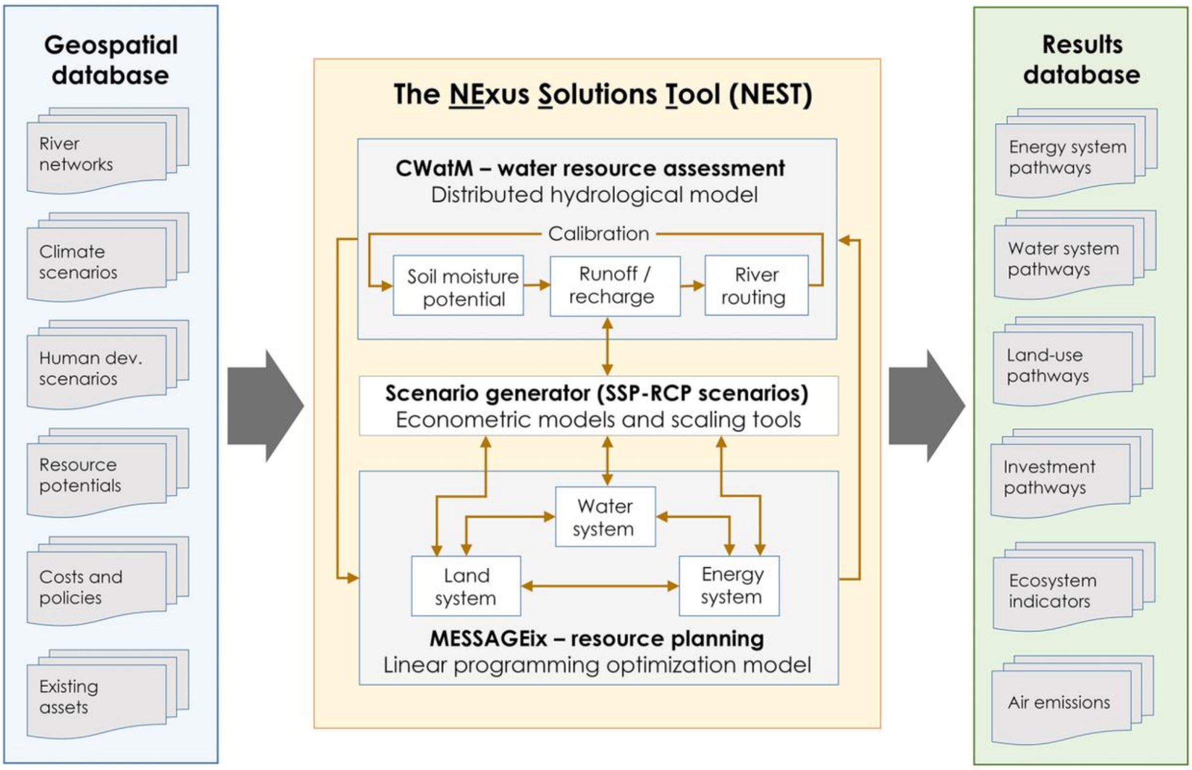


Figure 15: Generic scheme of the NEST framework in Vinca et al., 2020

Figure 15 provides a schematic representation of NEST, showing the input data, assessment and solution tools and outputs. The following steps describe how this new version of NEST (or MESSAGEix-Nexus) is generated (depicted in **Figure 16**):

1) downscaling the features of the global MESSAGEix model to a country of choice by using the Country Model Generator (CMG). This results in a single-node energy system country model. A case study of a similar approach have been published by Orhthofer et al. 2019.

2) connecting to the M-LED projections on electricity demand for different sectors at the sub-national spatial definition depicted in **Figure 17**. OnSSET outputs related to centralized vs off-grid electrification are also linked to the model at this stage.
3) adding the water system composed by water availability, demand and infrastructure (wastewater treatment and distribution systems) as described in NEST (with the exception of the river flow, canals and water storage representations) by using the general functions of the 'water' message_data module (which will become open source under https://docs.messageix.org/projects/models/en/latest/). Data used for hydrological variables come from global ISIMIP 2b Global Hydrological Model outputs (Frieler et al., 2017)
4) linking to Water Crop by including crop water requirements and irrigation technologies.

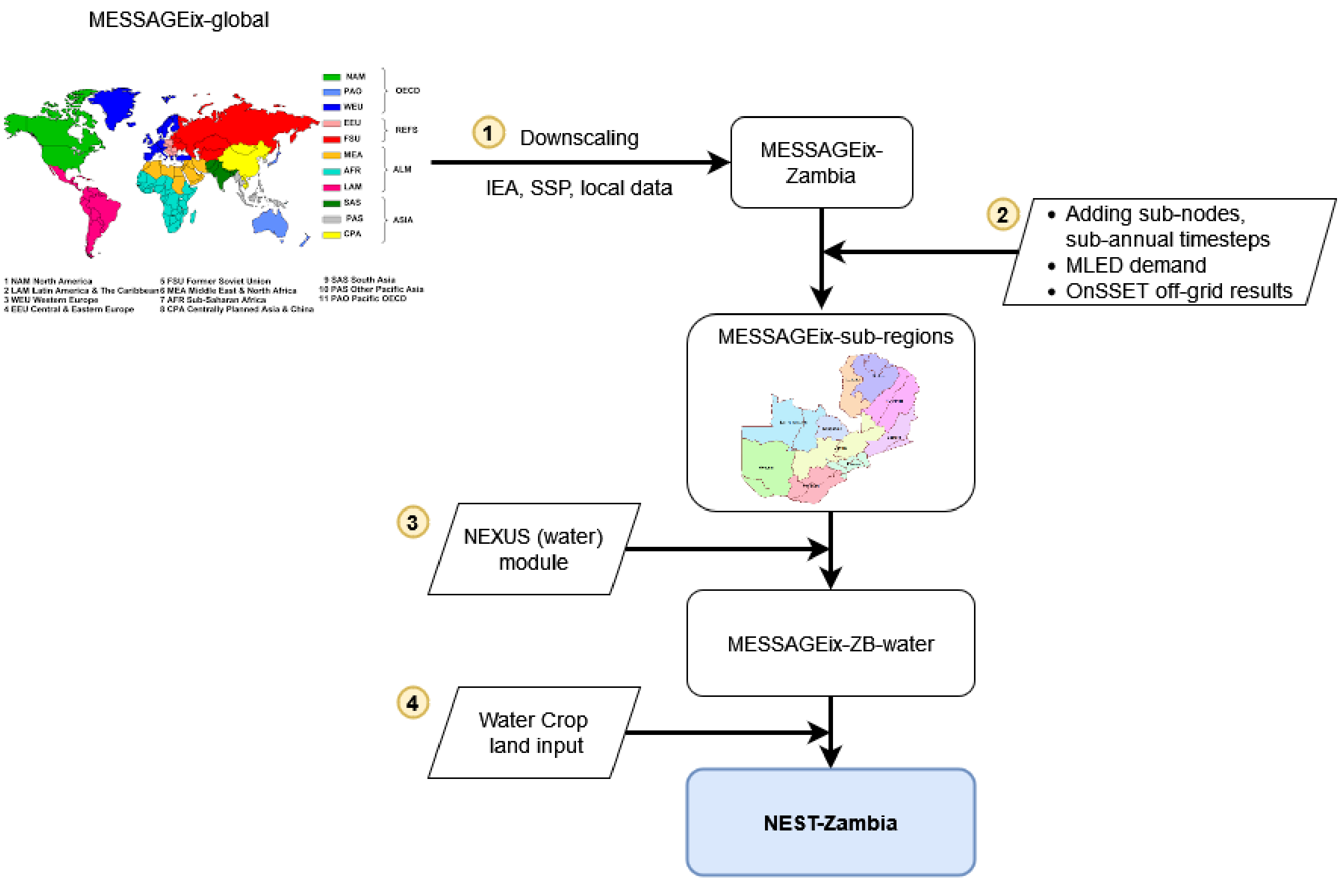


Figure 16: Process for developing NEST for a country, case study for Zambia

2.4.3 Spatio-temporal characterisation

The spatial delineation of the NEST Zambia model is illustrated in **Figure 17**, where the spatial units are the result of the geo-intersection of the hydrological sub-catchments (HYDROBASIN level 4 definition https://www.hydrosheds.org/products/hydrobasins) with the official administrative boundaries of the country resulting into 24 spatial units also referred to as *nodes* in the report. This approach is used to have a correct water balance representation and also socio-political representation for policy analysis and helps the integrated solutions to be balanced across different dimensions (Energy, Water, political, land). Primary & Secondary energy sectors in the energy component of the model are represented in a centralized manner and linked through the electricity grid to the various nodes.

The temporal resolution of the model is annual for the energy supply part of the model and monthly for the end electricity demand, rural electricity generation, water, and crop system. The model has a time horizon up to 2050 defined by 10-year time-steps.

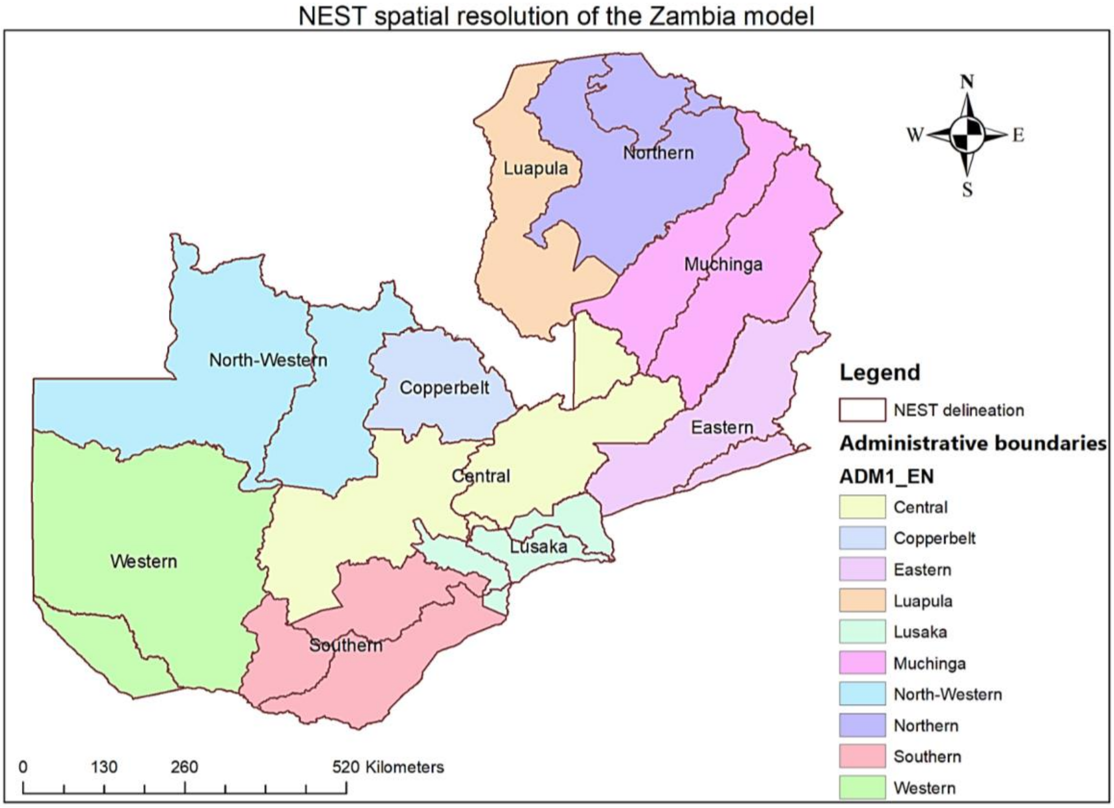


Figure 17: NEST spatial delineation and resolution

2.4.4 Parameters

The mathematical formulation of NEST is based on the MESSAGEix model developed at IIASA, which offers a versatile structure to model the energy-water-land systems with different spatial and temporal resolution, within an optimization framework that minimize total system cost under future constraints and development assumptions (e.g. growing demands). The detailed description of the input parameters, the output variables and equations of the model are openly available at https://docs.messageix.org/en/stable/ .

The details of the different sectors follow the structure described in Vinca et al., 2020. Further improvements and modifications have been done to better integrate NEST within the RE4AFAGRI framework.

In particular, the energy demand estimates from M-LED have been aggregated at the node level as monthly values for each of the modelled years. NEST also includes the expected share of grid vs off-grid electricity generation provided by OnSSET runs under common scenario assumptions.

2.4.5 Model Outputs

The NEST model is a multi-sector integrated assessment model with flexible configuration. Depending on the level of detail of the sectoral representation, different set of outputs can be produced by each sector.

In the RE4AFAGRI modelling platform, NEST receives output data from other models, such as off-grid electricity generation from OnSSET or crop and land-use from Water Crop. These output are therefore not direct output of the NEST solution. The following output indicators are outputs that NEST can produce in the RE4AFAGRI configuration.

Energy

- Primary and secondary energy use
- electricity generation mix (**Figure 18**)
- energy prices
- CO_2 emissions related to the energy sector
- future capacity requirements of energy technologies

- future required investment and expenditures in energy technologies

Water

- water availability from different sources: groundwater, surface water, recycled water, desalination (if suitable), as in **Figure 19**
- water withdrawals/demands from different sectors, also in **Figure 19**
- capacity requirement of water infrastructure technologies
- investment in the water infrastructure sector
- water prices

Land

- water and electricity use for irrigation
- investment requirements in irrigation

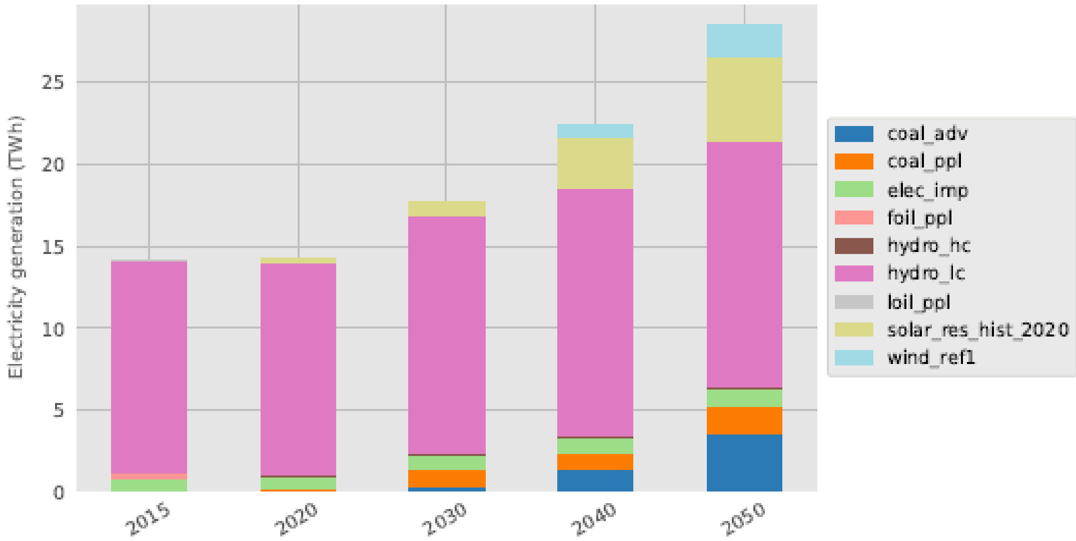


Figure 18: Output electricity generation mix from the NEST model for the Zambia case study, including different sources: coal (coal_adv, coal_ppl) electricity import (elec_imp), oil generation (foil_ppl, loil_ppl), hydropower (hydro_ch, hydro_lc), solar and wind

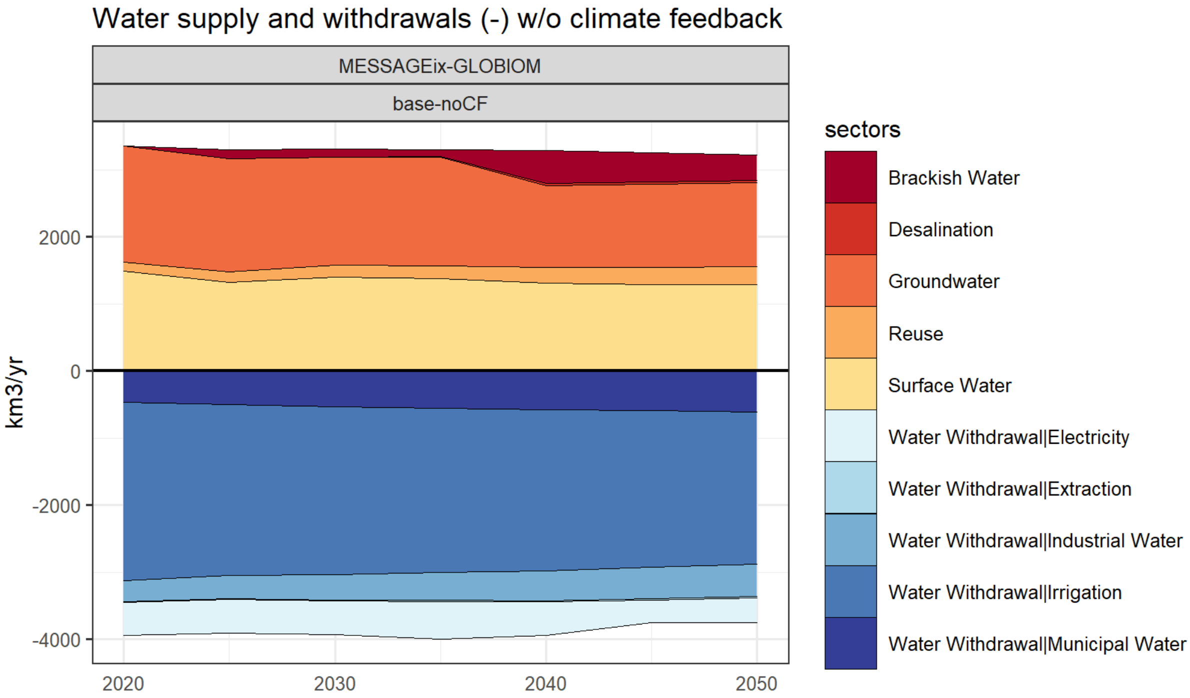


Figure 19: Sample output from NEST, showing different water sources and withdrawals.

3 Model framework: calibration, verification, and validation

Many diverse efforts have been made to maximize the accuracy and alignment of all data and assumptions used across the integrated modelling framework and to ensure that the methodologies applied are answering the right questions and ultimately providing insights that are specifically useful to the project's stakeholders. Some of these efforts will be described in general detail, while others will be explained in more detail in different sub-sections such as alignment to the previously completed Task 12.1 *"Tailoring research and methods to improve the climate-energy-water-agriculture nexus for African realities"*.

3.1 Integration of focus group discussion findings of Task 12.1 into the modelling frameworks

Task 12.1 of this project reviewed agricultural value chains, irrigation practices and crop processing in Nigeria, Zambia, Zimbabwe and Rwanda, providing important input for both the water, energy, land nexus modelling as well as the business cases. The review included a literature review, the administration of questionnaires, and focus group discussions to elicit both qualitative and quantitative information from local experts in Nigeria, Zambia, Zimbabwe and Rwanda who have experience in agricultural practices and land use, water needs and irrigation practices, and energy access constraints in rural areas in these countries. 12.1 provided a validation for the scope of the modelling work, analytical assumptions, research questions and scenario relevance.

The review and stakeholder discussions highlighted several important areas for consideration in the models, the key points are summarised below:

3.1.1 Spatial resolution of the models and representation of population densities

- Small scale subsistence farms are the majority typically - smallholder farmers typically have a 0.5 ha plot size in Nigeria, Rwanda and Zimbabwe and less than 2ha in Zambia
- Structural challenges related to the supply of water and energy for irrigation are often consistent between countries - such as the dominance of smallholder farming in the absence of irrigation, however there are also important differences such as the distance lying between farmers and their villages, which is related to the density of the population settlement in the country. These geographical differences play a key role in the economic trade-off between different electricity supply and irrigation systems options and therefore need to be factored into the modelling.
- Nexus models that explore access to energy and water in rural areas require high spatial resolution given the high sparsity and heterogeneity of settings affected by these issues (Almulla et al., 2020).

3.1.2 Irrigation practices and their interaction with the energy system

- There is currently very limited use of irrigation in all countries and only in larger farms, crops are rainfed and farmers are therefore susceptible to droughts and crop losses.
- Irrigation needs differ between countries and crops.
- Irrigation methods differ between locations, and the source of water also differs between locations, this affects pumping needs. Often these methods are based on price, for example surface water may be preferred where boreholes are too expensive.
- Timing of irrigation is typically in the early morning.
- Irrigation (energy) needs to take into account the distance of farms from villages.
- Generally, irrigation systems that are currently powered by fossil fuels pump water as needed. Solar powered systems will pump and store water during sunlight hours, to be released from storage as needed.
- The supply of energy for Irrigation should consider opportunities for storing energy for example, if gravity fed irrigation is adopted it provides a way of using/storing electricity generated from PV and allows irrigation to happen at any time of the day.

3.1.3 Crop processing and cold storage

- When optimising the local energy supply, processing of crops can offer opportunities to flatten the load profile. Additional PV use during the day for processing can make energy more affordable and provide additional value to the local farmers.
- Crop rotation throughout the year also offers an opportunity to manage the energy needs of crop processing throughout the year.
- Many of the opportunities for adding value to crops such as drying, milling, pressing and refrigeration will require additional energy and electricity, but may also be supplied by the national grid, where the village is connected to the national grid.
- Individual ownership of crop processing equipment is rare.
- Energy efficiency of crop processing offers opportunities to improve the service.

3.1.4 Other considerations

- Small scale farmers face several challenges when it comes to improving crop production. Apart from the need to irrigate due to uncertainty of rainfall and droughts, fertilizer use needs water/irrigation to work properly.
- Diesel prices have been increasing, affecting the affordability of pumping with diesel, and making PV more attractive.
- There is currently a year-to-year variation in yields and revenues.
- Skills are needed.

3.1.5 Reporting and sharing of results

- Reliable estimates of water and energy demand
- Reliable estimates of costs
- Impacts of climate change
- Results at the local scale

3.2 General cross-cutting validation efforts

Several detailed workshops and meetings were held between consortium members and modelling teams to ensure that the modelling activities, their interconnections and outputs, and the associated data requirements meet the ultimate aims of the project. Among these, a set of joint intensive project technical workshops and "hackathons" have been organised, with in-person meetings when possible. Developing co-created and jointly understood documents such as the current report has also represented a crucial building block to ensure agreement on key aspects influencing the modelling outputs.

Data inventories have been set up for each model describing the data sets - including their categorisation, sources, formats, and text descriptions. These inventories also serve to support transparency and reproducibility of the project outputs by recording the sources and metadata in a consistent way. An online data repository is eventually planned to be set up to preserve and make available all the necessary data sources needed for the operation of the models developed, which will also be needed to enable capacity building and knowledge transfer in future project activities. Data used by each model is described and sources are listed in the Appendix of this report.

Finally, a jointly published position paper has been recently published in the *Energy Strategy Reviews* journal (Falchetta, et. al., 2022). The scientific paper includes an introduction on the key principles followed by the RE4AFAGRI project, including an overview of the modelling aspects which have then been presented in more detail in the current report. The position paper also explores the key interconnections between the modelling platform and the business model research being carried out in parallel as part of Task 12.4.

3.3 Common data sources and inputs

To ensure cross-model consistency, some significant common data sources and parameter choices are used by various models across the modelling framework. These include socioeconomic and biophysical datasets and will be described here.

3.3.1 Hydroclimate data

Climate and hydrological data used in the models in the project are primarily sourced from the ISIMIP project. These datasets are mostly used by WaterCROP and NEST. This data includes climate variables for temperature and precipitation, as well as outputs from the global hydrological models (GHMs), for variables for potential evaporation, runoff, discharge, among others. These primary datasets are publicly available for download from the ISIMIP repository.

Climate models have variable performance over different regions. Thus, to understand and inform selection of suitable models for use, performance of the ISIMIP GCMs over the study area may be evaluated against observed data from the CRU repository. The following method, which has previously been done for the ISIMIP 2b data, will be applied for the ISIMIP 3b datasets.

We used historical observations and simulated climatic data of precipitation (P) and reference evapotranspiration (ET0), under baseline conditions with historical and future (RCP2.6 and RCP6.0) climate forcings. Historical P observations for the period 1961-1990 were obtained directly from the University of East Anglia's Climate Research unit (CRU CL v. 2.0) as long-term monthly average gridded data at 10x10 arc min resolution. ET0 data were provided for the same period by the Food and Agriculture Organization of the United Nations (FAO), which computes evapotranspiration according to the Penman - Monteith method (Allen et al. 1998), using climatic variables from CRU CL v. 2.0, as 5x5 arc min resolution monthly long-term averages. Historical estimates and future climate projections of monthly P and ET0 were sourced from the ISI-MIP repository - simulation round ISI-MIP2b as 30x30 arc min gridded data. Specifically, monthly ET0 data were provided as the output of PCR-GLOBWB global hydrological model, driven by four different Global Climate Models (GCMs) - GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5 - as climate forcings, while monthly P data at 30x30 arc min resolution were obtained from the same four different GCMs as ISIMIP input data. Intervals covering future time periods were simulated under both RCP2.6 and RCP6.0 - available for all simulation periods and GCMs; the choice of the RCPs included depends on the availability of ISIMIP climatic data. In order to prepare the climatic variables describing each scenario for which results are presented, we calculated the difference between each present or future time interval of P and ET0 and their historical estimates for the period 1961-1990. Then, the anomaly obtained was added to the reference observed P ET0 data (1961-1990). The practice of adding the perturbation - consisting in the difference between the modelled climate projection and the modelled historical climate - to an observed reference climate is a standard in the analysis of climate model results (Rosa et al. 2020), since, by 'anchoring' the modelled climate change to a common observed reference, greater confidence can be attributed to the results. This is necessary since historical climate simulations differ among models, which also vary in their assumptions and in the representation of some processes.

3.3.2 Socio-economic developments assumptions

Where possible models will use socioeconomic scenario data that are based on the Shared Socioeconomic Pathways (SSPs) framework, which provides five narratives of socioeconomic development primarily characterised by challenges to mitigation and adaptation (O'Neil et al. 2014). Subsequent work by the global and climate change community has developed consistent national level projections for key aspects such as population, GDP, urbanisation, education, governance. The original SSP scenarios have the base year of 2010, whilst forthcoming updates in the next year are expected to start at 2020.

In this project it is expected that at least the population (KC et al 2017), GDP (Dellink et al 2017) and urbanisation (Jiang et al 2017) data outputs of the SSPs will be used - where necessary, baseline reported values will be calibrated to recently reported values (e.g. GDP and population), e.g. from World Bank. Based on recent project scenario developments and data availability, is has been decided to use primarily SSP2 – "Middle of the Road".

3.4 M-LED – model specific calibration and validation efforts

3.4.1 Input datasets

The M-LED database combines and harmonises an array of data sources at different native spatio-temporal resolution, ranging from gridded raster datasets to discrete spatial feature vector data, up to national statistics. **Table A 1** (in the Appendix) illustrates the input data sources included in the M-LED database, along with the unit of each database, the type of data and its spatio-temporal resolution, a short description and a bibliographical reference or URL source. In addition, **Table A 1** also reports instances in which the input data is either derived from another model in the RE4AFAGRI platform, or feeding into another model in the RE4AFAGRI platform, or, finally, also used in another of the model in the RE4AFAGRI platform. This section will also discuss some of the specific alignment of inputs and outputs between the models.

3.4.2 Parameters and assumptions

Besides datasets, each M-LED scenario file (specific to each country) also includes a set of numerical parameters and assumptions, the with the main values listed in **Table A 3** (in the Appendix). Note that many of these parameters have been tailored through focus group discussions with stakeholders involved in the RE4AFAGRI Task 12.1 focus groups, which are reported in Deliverable D12.1 of the RE4AFAGRI project.

3.4.3 Model results validation

The output electricity demand of the Zambia pilot-country RE4AFAGRI model is compared with a recent independently-run "*Integrated resource plan (energy) (IRP) for Zambia"* study shared by the RE4AFAGRI consortium partner University of Zambia. In that study, four sectoral electricity demand were separately modelled. Quoting the IRP methodology (also see **Table 1** and **Figure 20** below):

> *Residential demand was initially performed making direct use of outputs from the Least-Cost Geospatial Electrification Plan (LCGEP); LCGEP outputs were combined with detailed analysis of recent household demand trends in Zambia, which indicates rapidly falling demand per connection as the grid expands to peri-urban and rural communities.*
>
> *Commercial and industrial (C&I) demand for electricity is closely related to GDP. This component of demand is much more straight-forward to model than residential. There is a strong link between demand and GDP.*
>
> *It was possible to obtain more detailed electricity demand projections from the sector stakeholders.*
>
> *Otherwise, analysis could be performed with the aim of identifying econometric relationships that might be used to project electricity demand from the mining sector.*
>
> *As with residential demand, demand from the agriculture sector was also informed by the LCGEP. The LCGEP is focused on electricity demand for irrigation. It is assumed that demand from value-add downstream activities are incorporated in the C&I regression.*

Table 2: Zambian IRP demand modelling methodology summary

Component of demand	Type of approach	Outline of approach
Residential	Econometric combined with bottom-up information from LCGEP	• Residential demand growth driven primarily by new connections. • Incremental demand from new connections varies with the cumulative number of residential connections. • Number of new connections informed by LCGEP, which provides analysis down to 'site' level, of which there are 3,248 across Zambia.
C&I (other than mining)	Econometric	• C&I demand varies with real-terms GDP. • Note that this component of the projection also includes public sector buildings such as schools and clinics. These facilities, while very important for Zambia's development, account for a very small portion of total electricity demand.
Mining	Bottom-up	• Projections provided at the mine level by individual mines and/or Copperbelt Energy Corporation (CEC).
Agriculture	Econometric combined with bottom-up information from LCGEP	• Overall agricultural demand for electricity assumed to vary with GDP (to be confirmed). • Potential location of agricultural loads informed by LCGEP analysis of Food and Agriculture Organization of the United Nations (FAO) data.

As seen, the Zambia IRP follows a different and independent approach in estimating future electricity demand, and thus is suitable for validating the M-LED output projections. **Figure 20** below shows the results of the national electricity demand (by sector) according to the IRP plan:

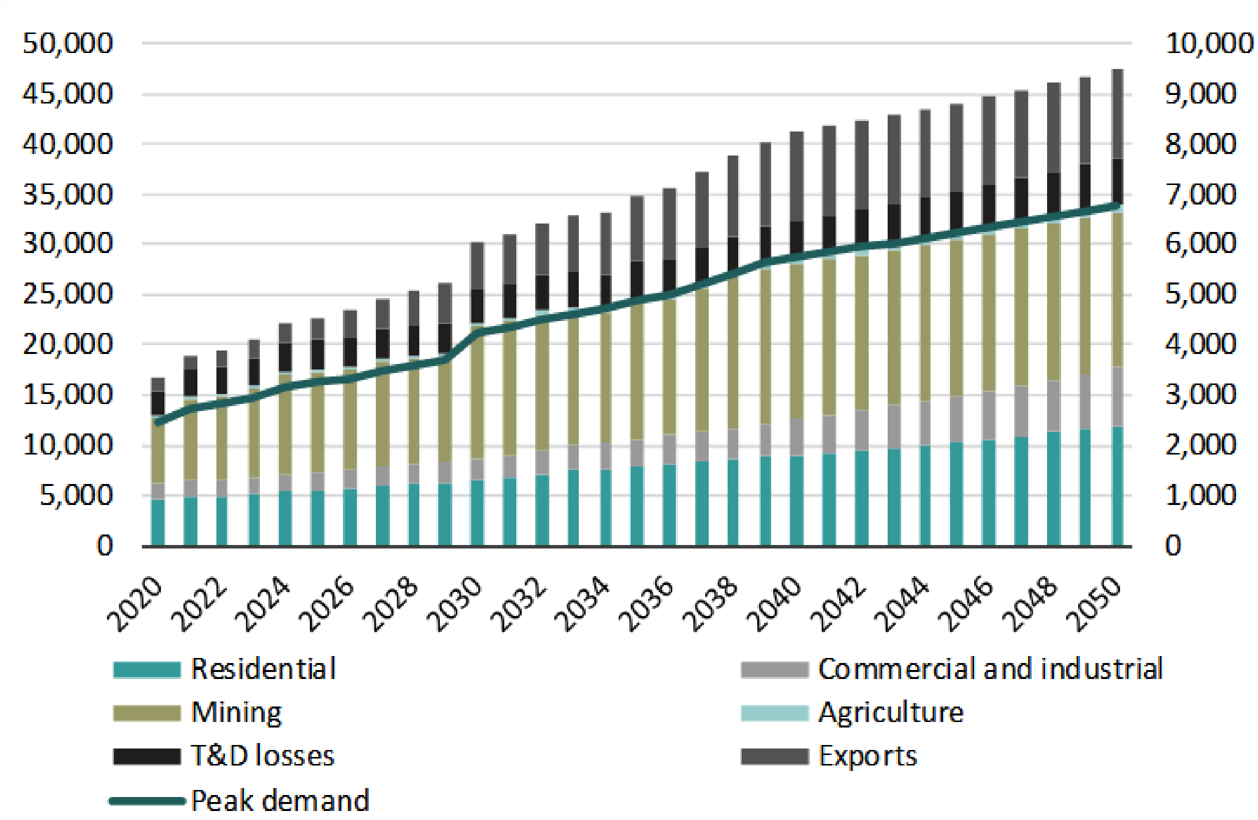


Figure 20: Projected electricity demand according to the Zambia IRP (values in GWh/yr)

The IRP forecasts the Zambian final electricity demand (net of exports and T&D losses) to reach about 35 TWh/yr, of which about 12 TWh/yr from the residential sector (roughly one third), 6 TWh of commercial and industrial demand, and the vast majority taken by the mining sector, with the agricultural sector representing only a marginal share of the total.

Comparing these results with the results of a preliminary baseline scenario M-LED run which was run independently and with no awareness of the Zambian IRP results (reported in **Figure 21** below) shows that electricity demand is projected to sum at about 36 TWh/yr in 2050, which is very similar to the IRP value. Also, the sectoral split is rather consistent, with the residential and mining sector each taking roughly one third of the total demand, and the remaining being split between commercial, agriculture and other sectors.

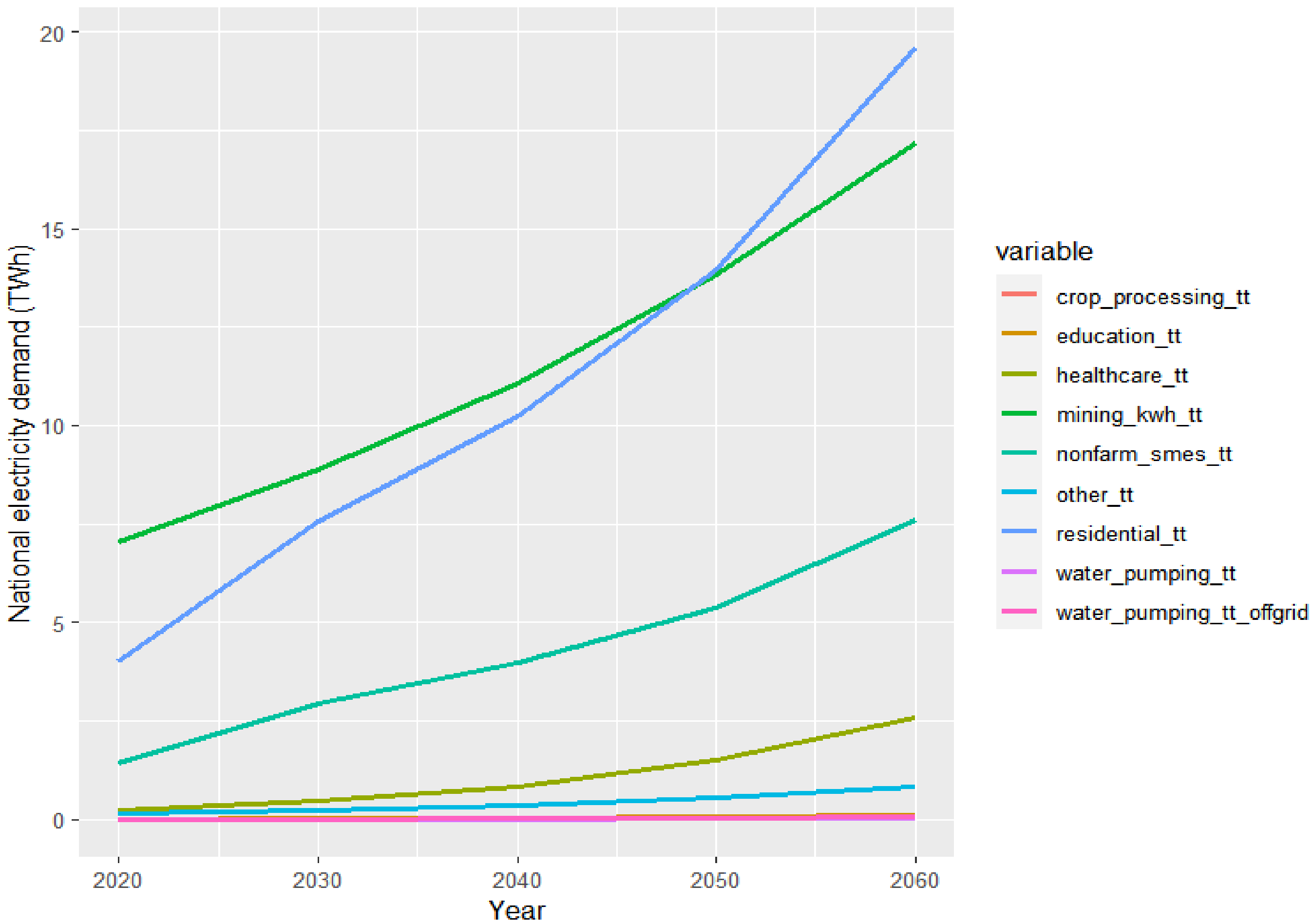


Figure 21: Projected electricity demand according to an M-LED baseline preliminary run

3.5 OnSSET - specific calibration and validation efforts

3.5.1 Input data improvements and benchmarking

Several input datasets have been improved in this work, before full linking of inputs and outputs of the different models, and these improvements have been benchmarked against similar studies for Zambia - specifically the *Global Electrification Platform* (GEP) version 2 update (GEPv2, 2021). The datasets which received the most attention for improvement were the high-resolution population distribution datasets and the HV and MV electricity grid network. A combination of methods were used to improve these which are described below along with the benchmarking exercises completed. These validated population and grid datasets are also both used as direct inputs into NEST and M-LED.

The validation and benchmarking presented includes a review of the improvements made to the datasets as well as the differences observed in electrification solutions from OnSSET. The population data was updated using the high-resolution building detection settlement boundary population clusters published by GRID3 (2022) combined with the new bottom-up WorldPop population distributions (WorldPop, 2022). The electricity network data was improved by creating an improved/validated grid dataset which combines OpenStreetMap data and gridfinder.org (2021) to remove non-existing grid lines or add missing lines.

A high-level National scale view of the improvements/comparison of both the population settlement clusters and the grid network datasets is demonstrated broadly in **Figure 22** below.

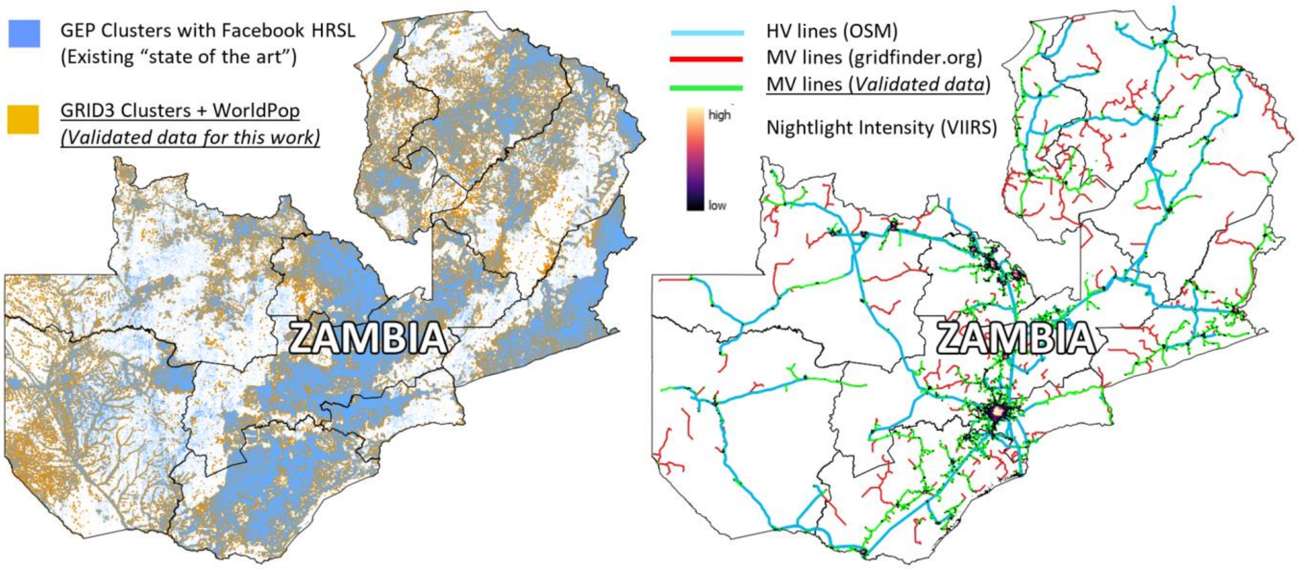


Figure 22: (left) Shows the existing GEP clusters in Blue overlaid on the improved GRID3 clusters, demonstrating that everywhere that orange can be seen are clusters that have been missed by the GEP dataset and it's underlying Facebook HRSL population rasters. (right) shows the OSM HV lines in blue which is used by both models, and the updated grid data used in this study in green, compared to the gridfinder.org (2020) data in red demonstrating that everywhere that red can still be seen on the map are false positive grid lines that do not exist.

We then directly benchmarked our national OnSSET optimization results against the GEP High-demand scenario which utilizes the World Bank Multi-Tier Framework scenario; (Urban Tier 5 and Rural Tier 3) scenario data for residential demand (ESMAP, 2015). All other costs and parameters including solar, hydro and diesel costs and connection costs are kept the same. This allows for any differences observed to be attributed to the improvement of the 2 input datasets. The validity of this specific scenario was not analysed in detail; however, it was the most straightforward scenario to match all associated input variables against.

3.5.2 Population and Settlement Cluster Improvements

Having accurate population data is a critical input in OnSSET because it determines the demand for electricity in any scenario. The size, location, density, and local resources of the population as well as whether households are electrified or un-electrified are all key factors that influence the outlook of the results of the electrification analysis.

In this analysis we have improved the population datasets by using a combination of sources and methods. We use the population clusters in the latest released data from Geo-referenced Infrastructure and Demographic Data for Development (GRID3, 2022) as the basis for the building detection and settlement shapes and groupings. To our knowledge this is the most accurate and complete open-access settlement detection data set for the African continent. The GRID3 settlement boundaries are also more aesthetically pleasing and realistic with very natural shapes compared to the squared-off or otherwise sharp boundary edges seen in the GEP clusters. These settlement clusters are then “populated" using the latest bottom-up) population distributions (WorldPop, 2022. Finally, the total population values, electrification rates, and rural/urban ratios are all calibrated using the same methods and values as in OnSSET and GEPv2 to allow comparable benchmarking (Khavari, 2021; GEPv2, 2021).

This approach aimed in part to minimize the number of false negatives and false positives in the data. A false positive is a cell that appears populated in the dataset but are uninhabited. False positives will lead to population settlements appearing larger than they are, as well as indicating population clusters where there are none – this usually occurs when rocks, trees, shadows, or water edges are mistaken for buildings or if temporary structures have been moved. Inversely, false negatives are when settlements and populations that do exist in an area are excluded or missed – this is usually due to a failure of the image processing algorithms or if old satellite imagery was use and new settlements have since been built in those areas (Khavari, 2021).

In **Figure 23** below two examples are shown comparing the updated population clusters used in this study (outlines coloured in orange) to the clusters used in the GEP (outlines coloured in blue). In the left of the figure is an example of many false positives in West Lunga National Park where after inspection all of the blue GEP clusters in this area are not structures but either rocks or trees. In the right of the figure several orange outlined settlements are completely absent in the GEP data where after inspection with satellite imagery are found to have many permanent structures built there. In the right of the figure the more natural shape of the GRID3 clusters can be seen compared to the square outlined shapes of the GEP clusters.

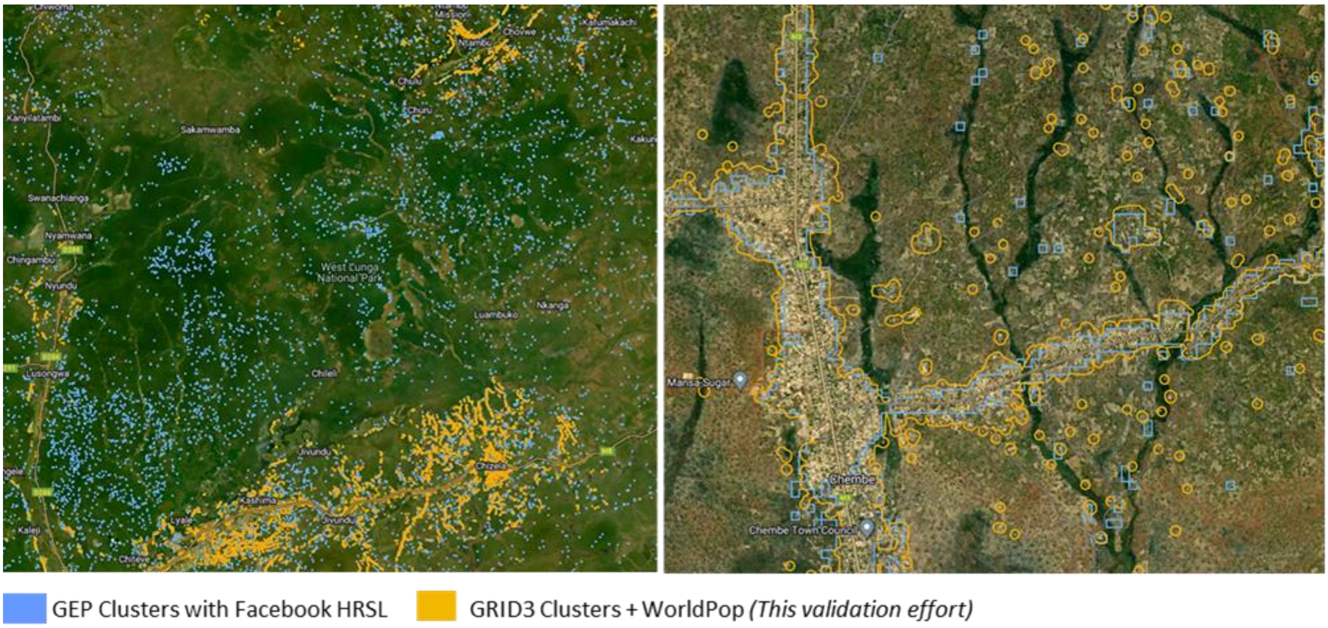


Figure 23: Comparison of population cluster datasets between GEP clusters (based on Facebook HRSL data) and the GRID3 settlement clusters combined with WorldPop bottom-up population counts. On the left is an example of many false positives in a National park and on the right many false negatives where settlements are not detected in the HRSL data. Imagery by Mapbox (2022).

3.5.3 Grid Network Improvements

Like the population cluster approach, this grid network validation tries to minimize the number of false positive (detecting grid networks which do not exist) and false negatives (missing grid networks that exist) in the grid network represented in OnSSET. This study validates the GEPv2 grid dataset using energy utility data from the energy regulation board of Zambia (ERB). Further, the GEPv2 dataset is improved by combining OpenStreetMap and Nightlight data from the VIIRS satellite to remove many of the gridfinder.org (2020) lines that are incorrect (false positive) by manually inspecting all lines with the multiple overlaid data layers.

Figure 24 below shows the OSM HV lines in blue which is used by both models, and the updated grid data used in this study in green, compared to the gridfinder.org (2020) data in red demonstrating that everywhere that red can still be seen on the map are false positive grid lines that when inspected do not exist.

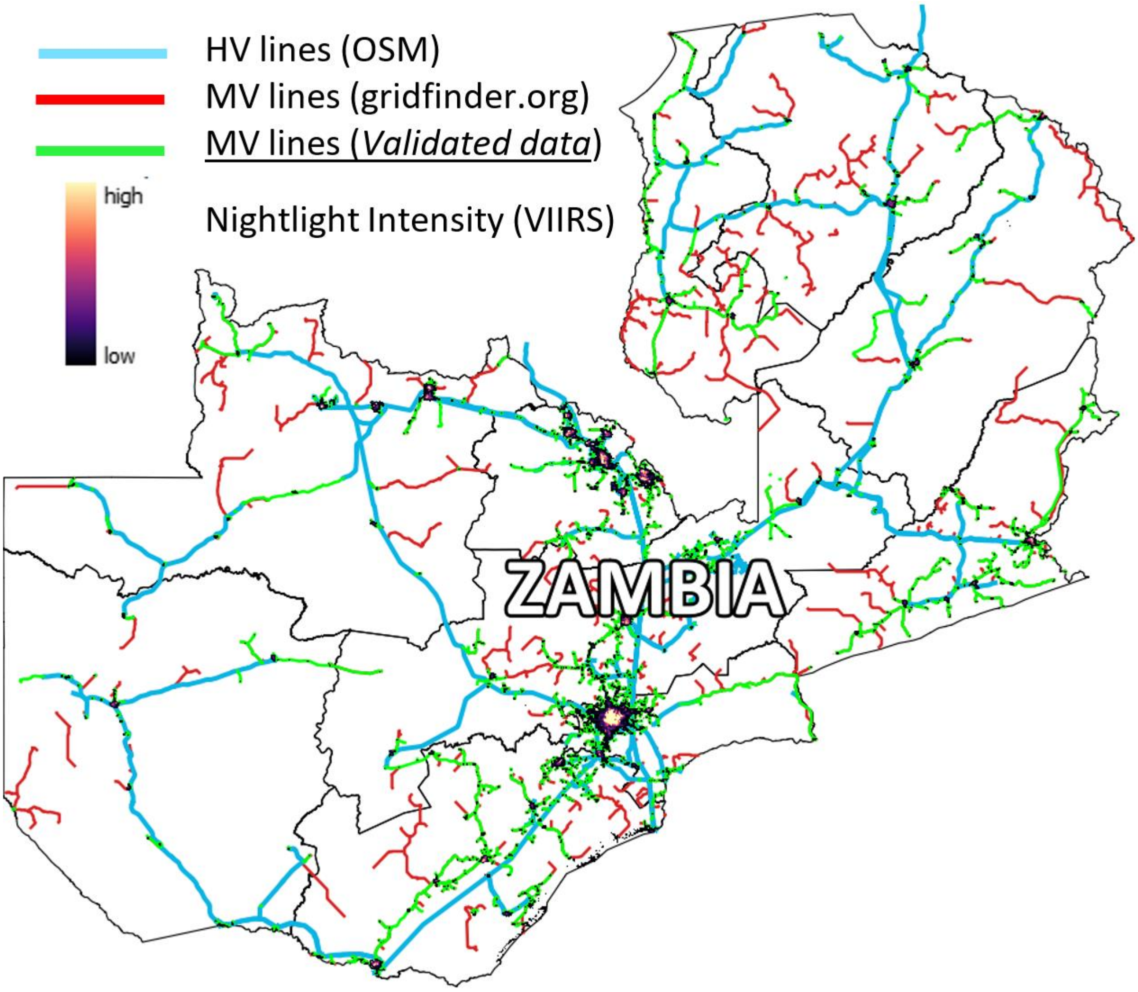


Figure 24: Shows the OSM HV lines in blue which is used by both models, and the updated grid data used in this study in green, compared to the gridfinder.org (2020) data in red demonstrating that everywhere that red can still be seen on the map are false positive grid lines that when inspected do not exist.

3.5.4 Example OnSSET outputs from benchmarking against Global Electrification Platform v2:

The OnSSET model was run with the updated population and grid datasets, with a 100% universal access target in 2030 and Tier 5 demand in urban areas and Tier 3 demand in rural areas and matching all other input datasets and assumptions. The least-cost technology options, determined by the model, for 2030 for this study are displayed on a map in **Figure 25** below.

The results show a mixed penetration of standalone solar, solar hybrid mini-grids and grid extension throughout the country. A very small number of hydro mini grids were also discovered as optimal; however, these provide electricity to less than 1000 people in both the GEPv2 and our study which is negligible compared to the other options and were therefore excluded visually from the map and charts shown here for clarity.

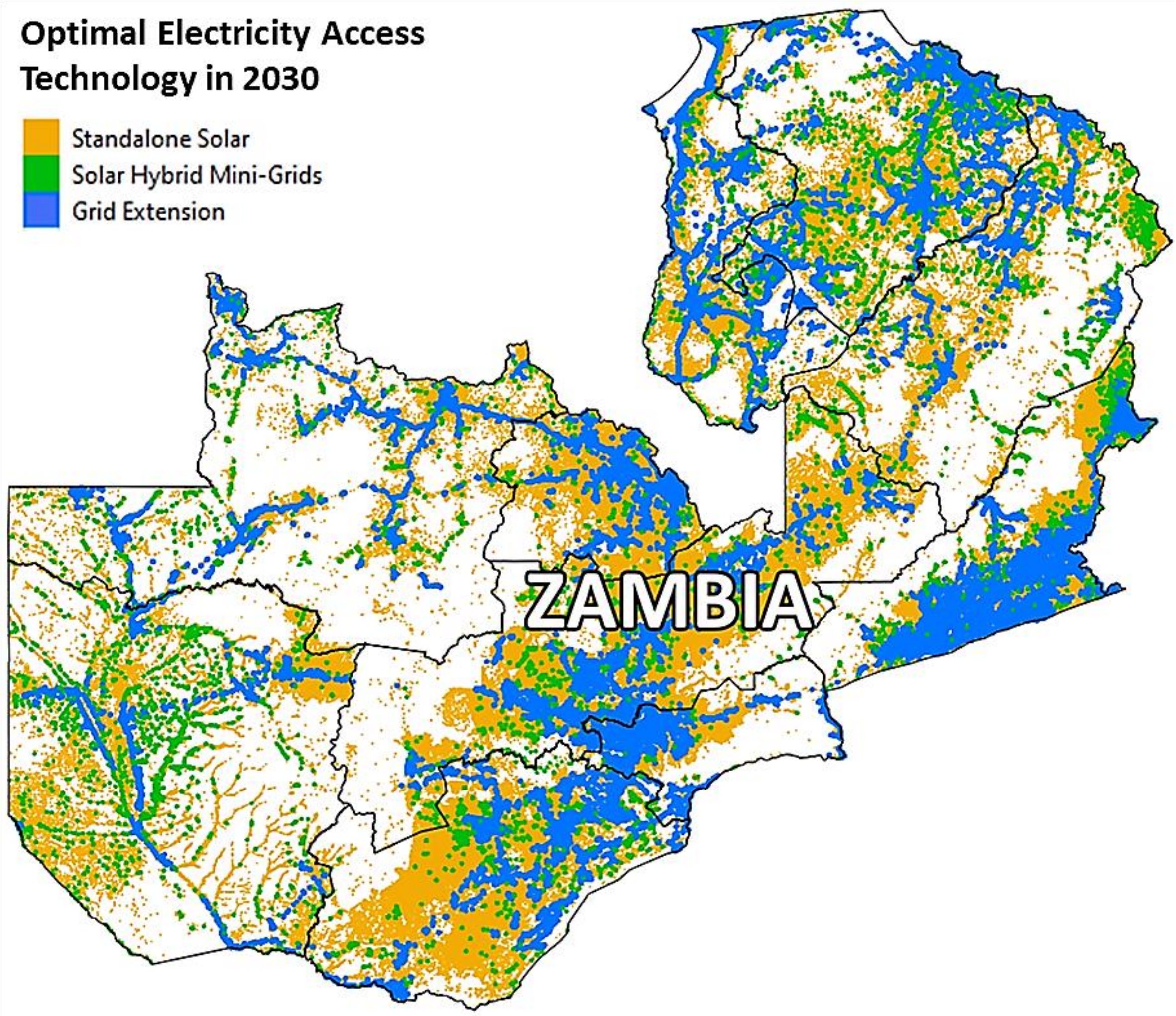
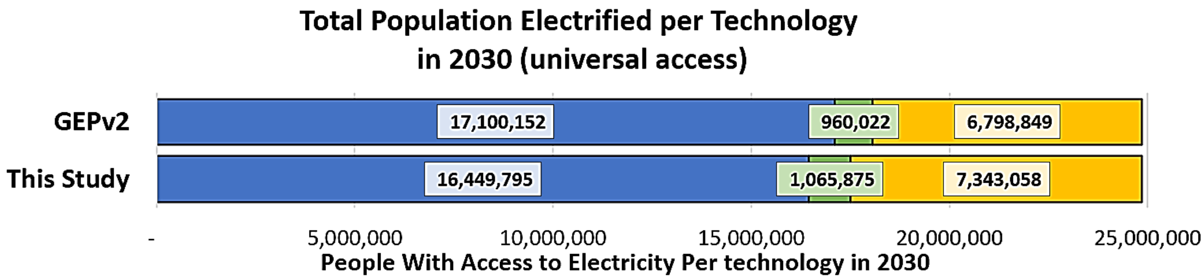


Figure 25: Optimal technology options are shown for 750 thousand individual population settlement clusters in Zambia. Blue shows where grid extension is found as cheapest, orange showing standalone solar systems, and green showing solar hybrid mini grids.



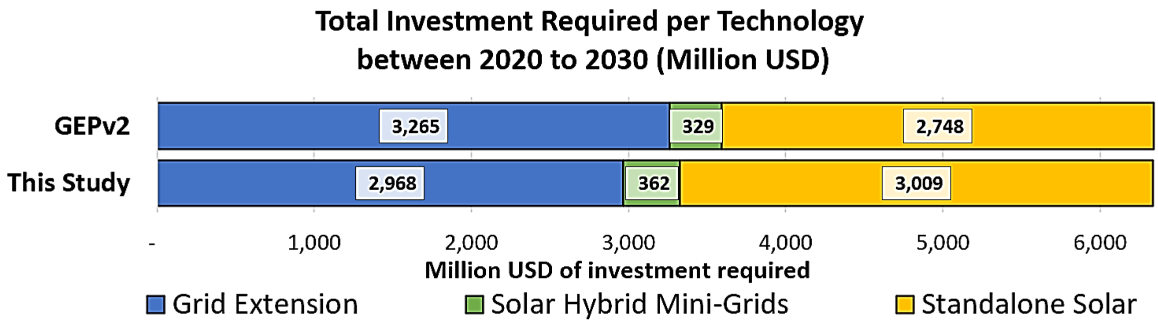


Figure 26: Total population and investment cost breakdowns per technology between the GEP and this study. (note: these are not 100% stacked charts)

The total investment cost results can be seen to be surprisingly similar between the GEPv2 and this study as shown in **Figure 26** (0.047% difference). However, a reduction of 10.0% in grid investments is seen combined with 8.7% increase in standalone solar systems, and a 9.3% increase in solar mini-grid investments. Similarly, for total population per technology a reduction of 10.0% in grid investments is seen combined with a 7.4% increase in standalone solar systems, and a 9.9% increase in solar mini-grid investments.

These differences can be mainly attributed to the gridfinder.org dataset finding significant amounts of false-positive grids (the red lines in **Figure 24**) making grid extension seem viable for incorrectly short distances to population clusters. The differences demonstrate the importance of the validation performed for OnSSET.

3.6 NEST – specific calibration and validation efforts

3.6.1 Input data sources

To parametrize the model for Zambia in terms of resources, technological processes, socio-economic constraints and demands of different sectors, we used the data sources outlined in **Table A 5** in the Appendix. It is however pertinent to note that much of the data required to run NEST can be obtained from open-source databases with global coverage to allow less dependence on data related challenges from local partner countries. Considering the importance of parameterization of these global datasets, we also calibrate the data with local sources.

3.6.2 Input data and results validation

The output of the electricity generation mix for Zambia pilot-country RE4AFAGRI model (**Figure 18** previously) is compared with a recent independently-run "*Integrated resource plan (energy) (IRP) for Zambia"* study shared by RE4AFAGRI consortium partner University of Zambia, showing alignment with the NEST results.

Table N1: Energy supply assumptions

Table 9: Energy supply and supply-demand balance using Antares (Baseline Case)

Energy Generation (GWh)	Existing 2020	Generation (GWh)		
		2026	2030	2040
Hydro Storage(H_STOR)	10122	8617	10300	12753
Hydro Run-of-River(ROR)	2222	10541	14024	17458
Thermal - Steam (Coal/other fuels)	2020	1589	2802	6008

Electricity demand for NEST come directly from the M-LED model and are validated as described in the M-LED section. Similarly, land use and crop distribution are dependent on the data and methodology used by Water Crop and not directly validated in the NEST model.

Water demand 2010 values and 2020 estimates from Wada et al,. 2014 are compared to AQUASTAT, World Bank estimates in **Figure 27**. The figure shows that for 2010 NEST is underestimating domestic water demands, but for 2020 shows higher demand than WB data. To be noted that the WB data seems to be constant over time, without showing any increase in water demand in the last decade.

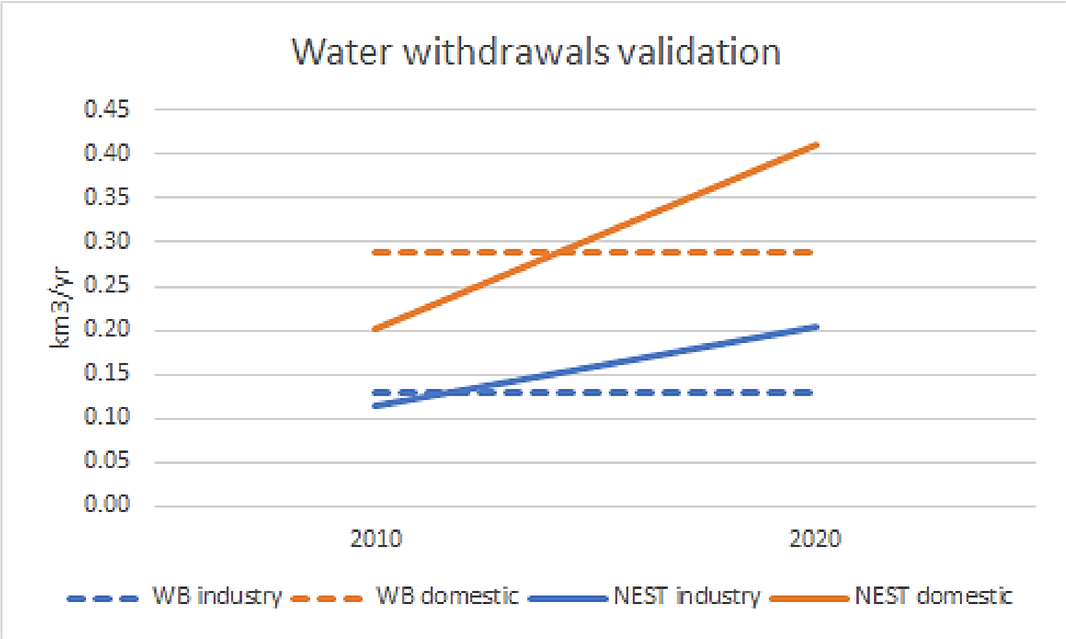


Figure 27: Comparison of NEST input data for Zambia on water demand and World Bank estimates (Food and Agriculture Organization, AQUASTAT data)

As mentioned in the previous sections, water availability is used from the outputs of the Global Hydrological Model outputs from the ISIMIP project. We compared the Zambian renewable surface & ground water availability with the values reported by AQUASTAT (FAO, 2022) and found that the estimates match with groundwater while the surface water estimates are close. However, AQUSTAT estimates were constant over the time while the data we are using are the result of assumptions by the hydrological model. The underlying assumptions for different scenarios and boundary conditions varies the results across time series. We plan to further validate these results with local stakeholders in the future and adjust our hydrological data accordingly.

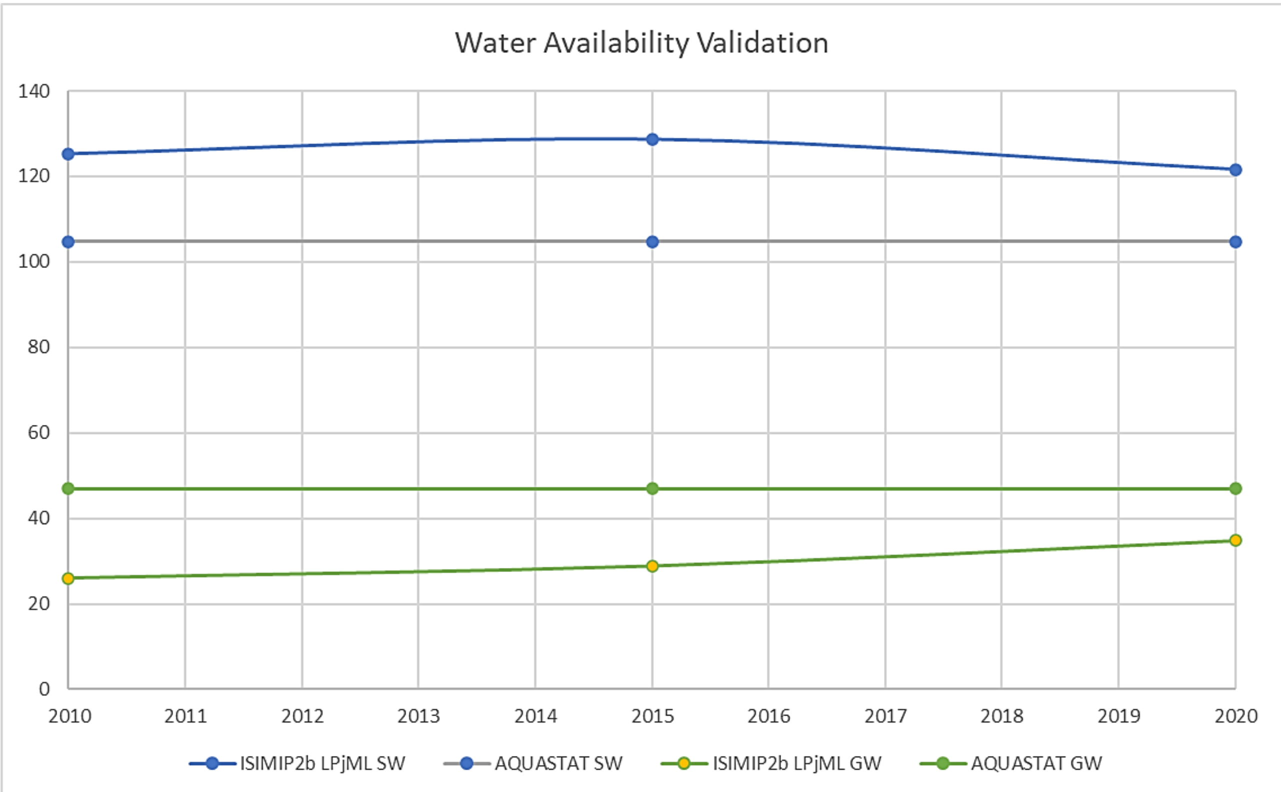


Figure 28: Comparison of NEST input data for Zambia on renewable Surface Water (SW) availability and Groundwater (GW) availability and World Bank estimates (Food and Agriculture Organization, AQUASTAT data)

4 Scenario development

The RE4AFAGRI platform follows a scenario-based logic, where a set of storylines representing different potential (or desirable) future developments happening both within the country-studies and at a global scale are designed and integrated in a coherent way into the RE4AFAGRI platform models.

Scenarios represent pathways towards the horizon year of 2050 considered in the RE4AFAGRI platform models, and each model generates and presents results for "milestone years" of each decade – i.e. 10-year time steps.

An explicit decision to develop a limited set of "storyline"-based scenarios rather than considering a full matrix of combinations of assumptions and targets was made by the RE4AFAGRI modelling group. This choice is justified by the objective of providing public and private stakeholders (the key target users of the RE4AFAGRI modelling outputs) with easy-to-interpret results, which allow comparing a "baseline" future evolution with increasingly ambitious goals or challenging conditions.

In the first modelling round (piloted in the country-study of Zambia), three key scenarios are constructed, as described in **Table 3** and Section 4.1 below.

The scenario development effort is broadly divided into the following steps:

1. Developing overarching future scenarios and narratives
2. Choice of baselines
3. Selection of key macro drivers and their future evolution
4. Setting future targets for key development metrics
5. Linking and/or aligning to existing related or equivalent policies already in place stated as accepted

Table 3: Core RE4AFAGRI scenarios

Scenario / Narrative	**Climate Expectations**	**National goals, policies, and targets**	**Socioeconomic Future Projections**
"Baseline"	RCP 7.0	Current policies, Agriculture: project historical trend of production.	SSP2
Moderate development	RCP 7.0	Electricity access: halving the gap by 2030. Water access & sanitation: halving the gap by 2030. Food security: increasing water supply to meet domestic food crops production demand by 2030 (given current diet, no extensification, fertilization, trade)	SSP2
Improving access, sustainable targets	RCP7.0 (SSP1, RCP2.6 on land)	Universal electricity access. Universal water access & sanitation. Nutrition security: improve yields to meet future food crop production demand to meet the EAT Lancet diet by 2030. Renewable electricity share = 100% + climate constraints/commitments.	SSP2

4.1 Developing future scenarios and narratives

Three key scenarios are developed:

- **Baseline Scenario:** the baseline scenario represents an extrapolation of recent trends into the future, and it is useful to highlight potential challenges in absence of changes in trends (e.g., remaining gap in energy, water access and adequate nutrition).
- **Moderate Development Scenario:** some efforts are made to improve the quality of living in the case-study country, by increasing energy, water, and sanitation access so that the access gap estimated in the baseline (percentage of population remaining without access) is at least halved by 2030. A food nutrition target also aims at ensuring domestic food production by improving crop yields through irrigation
- **Improving Access, Sustainable Targets Scenario:** this scenario includes ambitious and ideal targets of universal access to electricity, water, and sanitation by 2030, and domestic agriculture production improving to meet decent living nutrition standards (EAT Lancet diet). In addition, measures to guarantee 100% renewable electricity generation are in place.

This scenario ensemble includes different levels of ambitions in improving the production and access to electricity, water infrastructure and water for agriculture, with different grades of investment requirements or secondary impacts on natural resources (e.g., coal or water withdrawals).

Additional scenarios can be run upon interest from the modelling team and stakeholders to assess the sensitivity to specific model parameters (e.g., technology costs) or to address specific questions (e.g., intensification vs extensification of agriculture, or achieving targets in 2030 and or 2050 etc.)

Example of questions that the framework will be able to answer:

- What is the water required to increase crop yields in year 2030 to achieve domestic food security?
- Share of off-grid or grid connection when increasing demand in rural areas?
- How do long-term investment portfolios look given different levels of development ambition?
- What are the economic implications of reaching targets in 2030, 2040, or 2050?
- How costly would it be to achieve 100% renewable electricity (i.e., no diesel/coal)?

4.2 Policy dimensions or targets specifically implemented in each model.

1. M-LED

- *el_access_share_target* -> target share of population with electricity in each planning year
- *crop_processed_share_target* -> target share of crop yield locally processed in the last planning year
- *irrigated_cropland_share_target* -> target share of irrigation water demand met in the last planning year
- *groundwater_sustainability_contraint* -> impose limit on groundwater pumping based on monthly recharge

2. WaterCROP

- Constraint: yearly ground water must be recharged
- Objective: target on yield gap.

3. OnSSET

- *el_access_share_target* -> target share of population with electricity in each planning year
- *100_percent_renewables* -> disallow diesel generation for hybrid mini-girds and/or standalone systems

4. NEST

- Constrain specific power plants, including share of renewables
- future exogenous demand trends, including the energy access target
- Share of rural/grid generation
- Constraints to water extraction (surface or groundwater)
- water access and sanitation targets, as share of population with access, water efficient infrastructure, environmental flows, and share of irrigation technology use

5 Interactive online dashboard concept

One of the key purposes of the modelling activities carried out in RE4AFAGRI Tasks 12.2-12.3 is to produce an online interactive model output visualisation dashboard that can be used by private and public national and regional stakeholders interested in making data-informed decisions (such as investment targeting, infrastructure sizing, or development policies design). The dashboard will be used both in capacity building activities as part of project Task 12.6, as well as in dissemination activities as part of project Task 12.7. Moreover, the online website will also contain results from the business model research activities carried out under Task 12.4 (results of which will be published in a dedicated deliverable in Q2 2023).

Figure 29 illustrates the current concept (under development and subject to change) of the landing page of the 'dashboards' section of the website, upon which the user is prompted to select a country of interest.

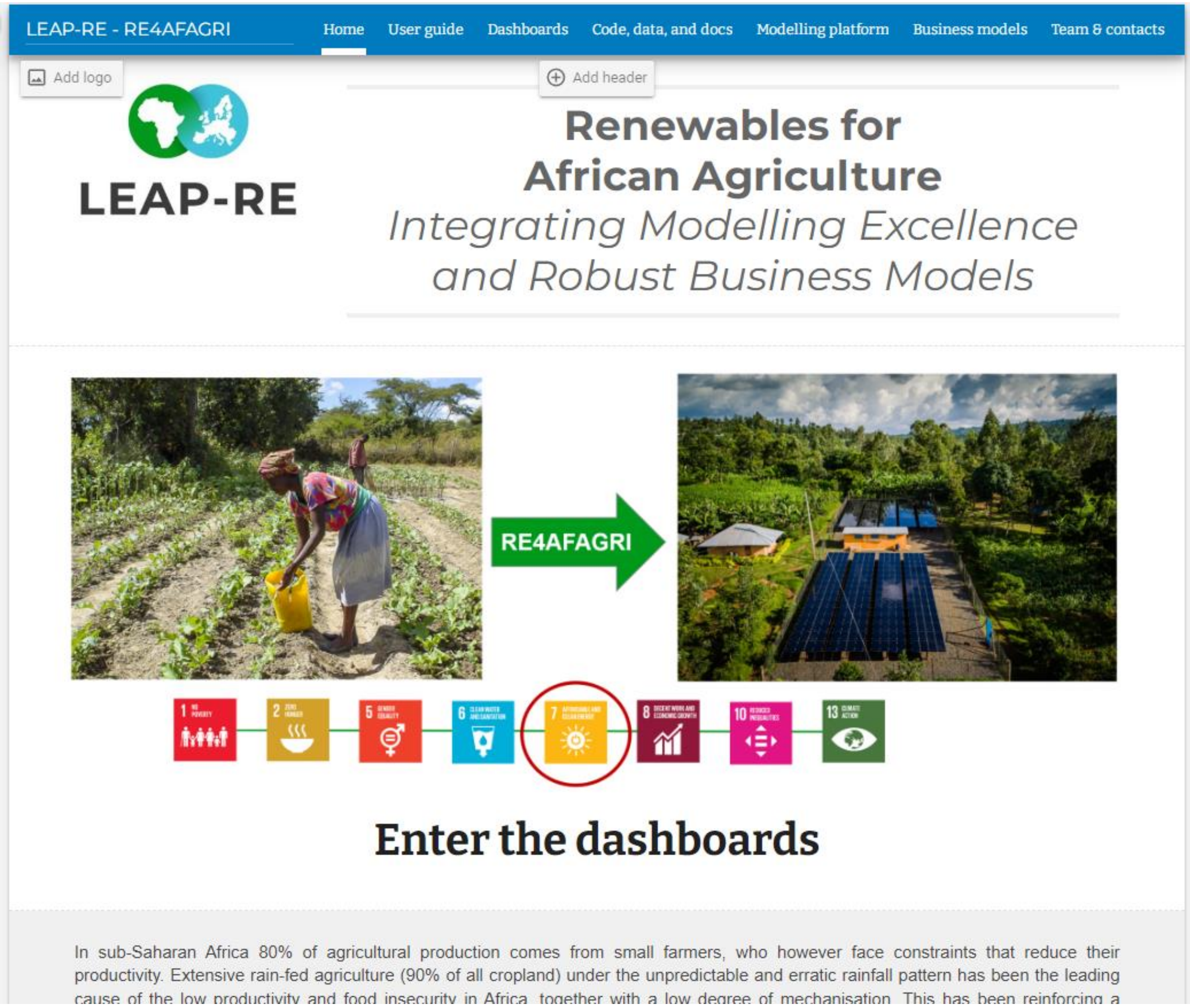


Figure 29: concept of the RE4AFAGRI dashboard website - landing page

Figure 30: concept of RE4AFAGRI dashboard website - country dashboards selection

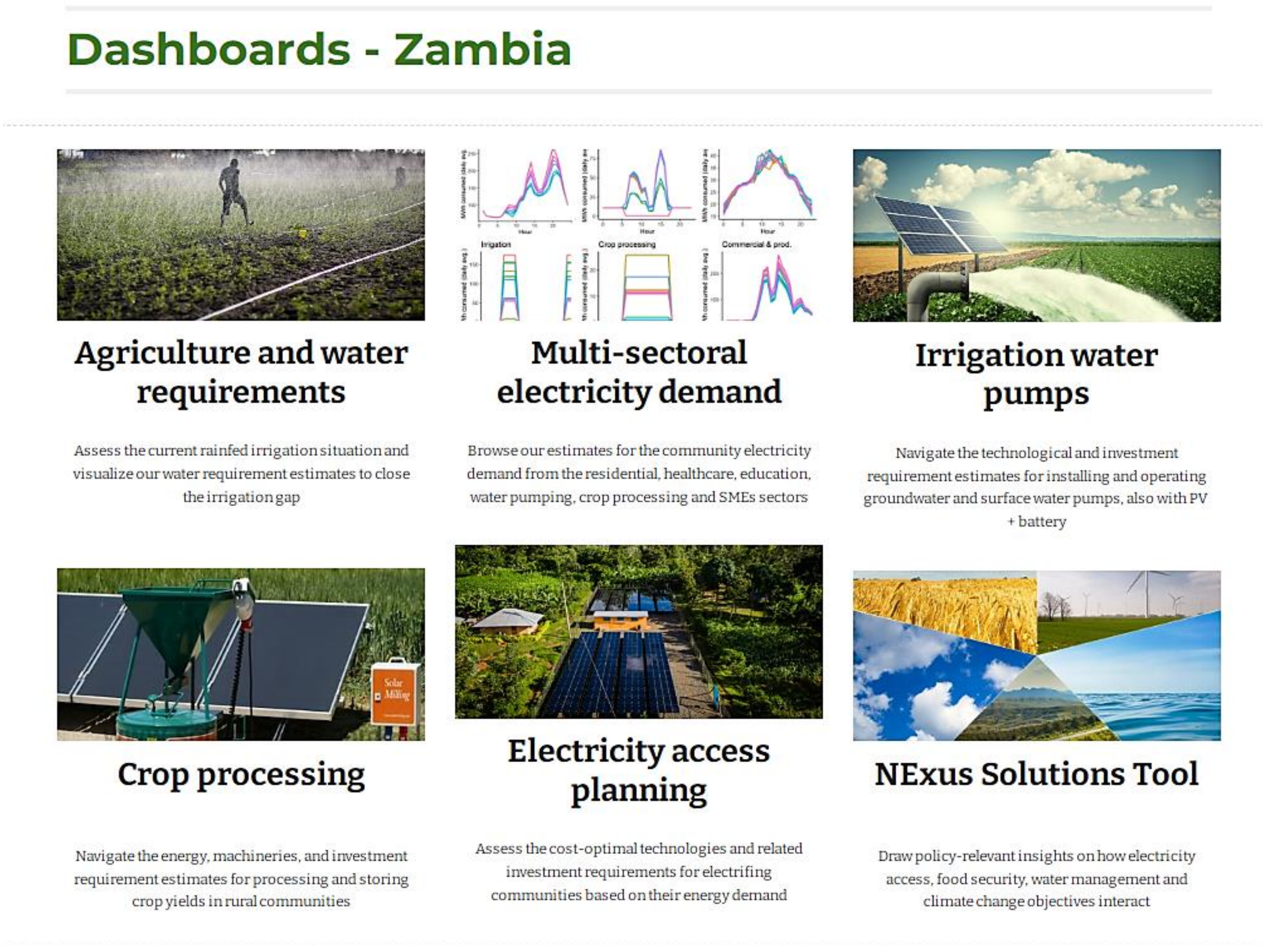


Figure 31: concept of RE4AFAGRI dashboard website - dashboard selection page sample (under development)

Subsequently, the user is prompted to select the specific dashboards of their interest (**Figure 31**). Currently, six dashboards are under development or planned to be developed:

- Agriculture and water requirements (WaterCROP outputs and MapSPAM data)
- Multi-sectoral energy demand (M-LED outputs)
- Irrigation water pumps (supplementary M-LED outputs)
- Crop processing (supplementary M-LED outputs)
- Electricity access planning (OnSSET and NEST-ENERGY outputs)
- NExus solution tools (NEST outputs)

Eventually, by accessing an individual dashboard (see example in **Figure 32** below), the user can select scenarios, timestep, as well as different dimensions and explore the maps and the box heatmaps to assess different variables of interest and their projected evolution over the modelled scenarios. The outputs are displayed at the second administrative level unit, but by clicking on a given province, the user can download a cluster-level geopackage file containing high spatial resolution outputs for all the clusters located inside a given administrative unit. Additionally, the user will also be able to download the whole country clusters geopackage to perform in depth analysis using desktop GIS software.

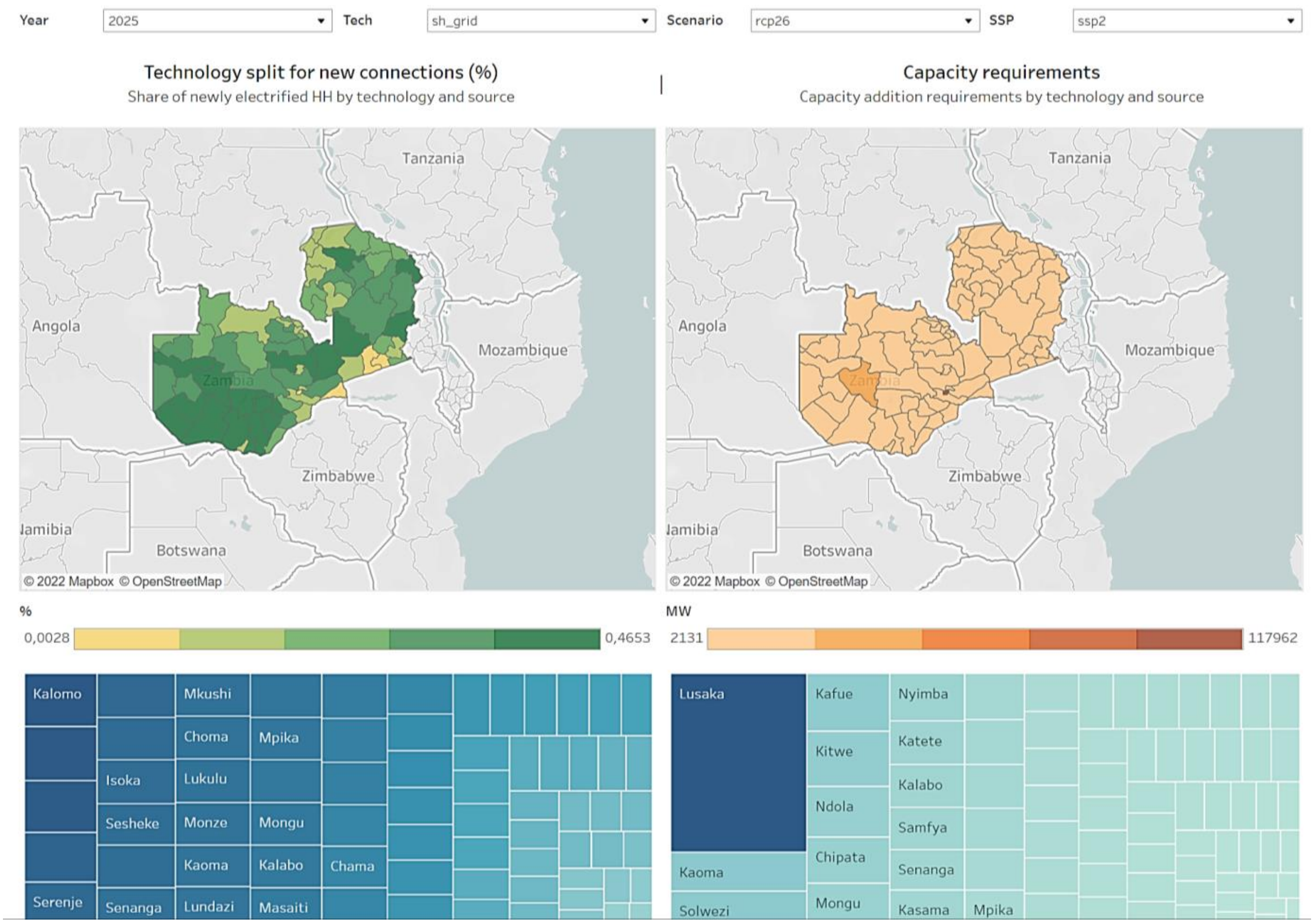


Figure 32: concept of RE4AFAGRI dashboard website, electricity access planning dashboards

6 Conclusions and further work

This report provides a detailed description of the interconnected models that are used in the *RE4AFAGRI nexus modelling platform* namely: WaterCROP, M-LED, OnSSET and MESSAGE-NEST. It describes their adaptation to the specific topics and research questions elaborated on in the "RE4AFAGRI" project, including stakeholder engagements and contextual alignment. It also describes the details of the technical model soft-linking integration. The technical and qualitative aspects of calibration, verification, and validation of model data, assumptions, and interconnections has been a focus of this work. The specific data used as separate inputs to the models, the intermediate data and results integrated between models, and the specific research questions targeted in the project have been evaluated and presented here.

The activities described in this report fall under Task 12.2 and Task 12.3 of the RE4AFAGRI project. The models used in the nexus modelling platform are all open source, and an important aspect of the work carried out under Task 12.2 and 12.3 has been ensuring that the models and results can be interrogated and used to support real-world public and private decision makers, interested stakeholders, while also being useful to build scientific capacity among the research community of AU and EU countries. Documented here for Task 12.3 are the many diverse efforts that have been made to maximize the accuracy and alignment of all data and assumptions used across the integrated modelling framework and to ensure that the methodologies applied are answering the right questions and ultimately providing insights that are specifically useful to the project's stakeholders.

This task has been achieved through (i) holding modelling meetings and workshops to ensure consistency of input data and modelling assumptions among the different modelling teams and drawing information and parameters collected from experts and focus groups through Task 12.1 into the modelling frameworks), (ii) the creation of data inventories for each model, and (iii) the validation of input data via desktop methods and through consultation with in-country partners (iv) collecting data from stakeholders (e.g. ministry, statistical offices, and smallholder consortia).

In addition to the model and data descriptions, the report provides an overview of the scenarios that will be used in the analysis undertaken in Task 12.2. The three scenarios developed and described in this report will allow an interrogation of needs for electricity and water firstly in a "business-as-usual" case, as well as two scenarios with increasing levels of ambition for access to electricity, water, sanitation, agricultural production.

An overview of the interactive results dashboard and visualisation platform that will be used to disseminate the results and support capacity building activities is also presented here. A concept design demonstrating its technical capabilities and prototype visual design is shown with several illustrative examples.

The eventual modelling of scenarios and full integration of the model suite are currently undergoing following the work presented in this report and will be described in preprints and eventually scientific articles to be drafted starting in 2023. An online data repository will be set up to preserve and make available all the necessary data sources needed for the operation of the models developed, which will also be needed to enable capacity building and knowledge transfer in future project deliverables and the activities in Tasks 12.5 and 12.6.

In particular, in the first half of 2023 Deliverables 12.3 "*Integrated platform source code on GitHub as main deliverable of Task 12.2*" and 12.4 "*User guide for the modelling platform on GitHub as second deliverable of Task 12.2*" will ensure the source code is made openly accessible online, along with the replication data, and a user guide for the RE4AFAGRI platform. All the materials will be made available at:
https://github.com/iiasa/RE4AFAGRI_platform

References

Becker, K. H., Gillin, E. & Kabat, L. Food balance sheets: a handbook. Food Agric. Organ. U. N. Rome Italy (2001).

C. Rosenzweig, J. Elliott, D. Deryng, et al. Assessing agricultural risks of climate change in the 21st century in a global gridded crop model intercomparison, PNAS 111 (9) (2014) 3268–3273. doi:www.pnas.org/cgi/doi/10.1073/pnas.1222463110.

Chapagain, A. K. & Hoekstra, A. Y. Water footprints of nations. (2004).

Colorado School of Mines (2022), NIGHTTIME LIGHTS, URL: https://payneinstitute.mines.edu/eog/nighttime-lights/

E. H. Sutanudjaja, R. van Beek, N. Wanders, et al. Pcr-globwb 2: a 5 arcmin global hydrological and water resources model, Geosci. Model Dev. 11 (2018) 2429–2453. doi:https://doi.org/10.5194/gmd-11-2429-201

ESMAP, "Multi-Tier Framework for Measuring Access to Electricity," 2015. https://mtfenergyaccess.esmap.org/methodology/electricity.

F. T. Portmann, S. Siebert, P. D¨oll, Mirca2000—global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling, Global biogeochemical cycles 24 (1) (2010).

Falchetta, G., Stevanato, N., Moner-Girona, M., Mazzoni, D., Colombo, E., & Hafner, M. (2021). The M-LED platform: advancing electricity demand assessment for communities living in energy poverty. Environmental Research Letters, 16(7), 074038.

FAO, Fao map catalog, [Accessed: November 2021] (2014).

FAO, IIASA, Global Agro-Ecological Zones (GAEZ v4) – Data Portal user's guide, FAO, 2021.

FAOSTAT, R. FAOSTAT database. Food Agric. Organ. U. N. (2017).

Food and Agriculture Organization, AQUASTAT data, 2022, https://www.fao.org/aquastat/en/databases/maindatabase https://data.worldbank.org/indicator/ER.H2O.FWTL.K3?locations=ZM

Fritz, S. et al. Mapping global cropland and field size. Glob. Change Biol. 21, 1980–1992 (2015).

GEP v2, "Zambia," 2022. Global Electrification Platform. World Bank ESMAP. https://electrifynow.energydata.info/explore/zm-2?year=2030&scenario=2_1_0_0_0_0 (accessed May 29, 2022).

GRID3, "Geo-Referenced Infrastructure and Demographic Data for Development," 2022. https://data.grid3.org/datasets/GRID3::grid3-zambia-settlement-extents-version-01-01-/about (accessed May 12, 2022).

Harris, I., Osborn, T. J., Jones, P. & Lister, D. Version 4 of the CRU TS monthly high-resolution gridded multivariate climate dataset. Sci. Data 7, 109 (2020).

Hruschka DJ, Gerkey D, Hadley C. Estimating the absolute wealth of households. Bull World Health Organ. 2015 Jul 1;93(7):483-90. doi: 10.2471/BLT.14.147082. Epub 2015 May 15. PMID: 26170506; PMCID: PMC4490812.

International Food Policy Research Institute. Spatially-Disaggregated Crop Production Statistics Data in Africa South of the Sahara for 2017. (2020) doi:10.7910/DVN/FSSKBW.

B. Khavari, A. Korkovelos, A. Sahlberg, M. Howells, and F. Fuso Nerini, "Population cluster data to assess the urban-rural split and electrification in Sub-Saharan Africa," Sci. Data, vol. 8, no. 1, pp. 1–11, 2021, doi: 10.1038/s41597-021-00897-9.

Khavari, Sahlberg, Korkovelos, Mentis, & Moksnes. 2022. OnSSET teaching material. Zenodo. https://doi.org/10.5281/zenodo.4574031

K. Frieler, S. Lange, F. Piontek, et al. Assessing the impacts of 1.5 °c global warming – simulation protocol of the inter-sectoral impact model intercomparison project (isimip2b), Geosci. Model Dev. 10 (2017) 4321–4345. doi:https://doi.org/10.5194/gmd-10-4321-2

Leiwen Jiang, Brian C. O'Neill, Global urbanization projections for the Shared Socioeconomic Pathways, Global Environmental Change, Volume 42, 2017

L. Rosa, D. D. Chiarelli, M. Sangiorgio, A. A. Beltran-Pe˜naa, M. C. Rulli, P. D'Odorico, I. Fung, Potential for sustainable irrigation expansion in a 3 °c warmer climate, PNAS 117 (47) (2020) 29526–29534. doi:https://doi.org/10.1073/pnas.2017796

Lombardi, F., Balderrama, S., Quoilin, S., & Colombo, E. (2019). Generating high-resolution multi-energy load profiles for remote areas with an open-source stochastic model. Energy, 177, 433-444.

M. New, D. Lister, M. Hulme, I. Makin, A high-resolution data set of surface climate over global land areas, Climate Research 21 (2002) 1–. doi:10.3354/cr02100

M. Tuninetti, S. Tamea, P. D'Odorico, F. Laio, L. Ridolfi, Global sensitivity of high-resolution estimates of crop water footprint, Water Resour. Res. 51 (2015) 8257–8272. doi:doi:10.1002/2015WR017148

Mentis, D., Howells, M., Rogner, H., Korkovelos, A., Arderne, C., Zepeda, E., Siyal, S.H., Taliotis, C., Bazilian, M., De Roo, A., Tanvez, Y., Oudalov, A., Scholtz, E., 2017. Lighting the World, The first global application of an open source, spatial electrification tool (OnSSET), with a focus on Sub-Saharan Africa. Environmental Research Letters Special Focus on Energy Access for Sustainable Development.

Maus, V., Giljum, S., Gutschlhofer, J., da Silva, D. M., Probst, M., Gass, S. L., ... & McCallum, I. (2020). A global-scale data set of mining areas. Scientific data, 7(1), 1-13.

Nachtergaele, F. et al. The harmonized world soil database. in Proceedings of the 19th World Congress of Soil Science, Soil Solutions for a Changing World, Brisbane, Australia, 1-6 August 2010 34–37 (2010).

Orthofer CL, Huppmann D and Krey V (2019) South Africa After Paris—Fracking Its Way to the NDCs? Front. Energy Res. 7:20. doi: 10.3389/fenrg.2019.00020

O'Neill, B.C., Kriegler, E., Riahi, K. et al. A new scenario framework for climate change research: the concept of shared socioeconomic pathways. Climatic Change 122, 387–400 (2014). https://doi.org/10.1007/s10584-013-0905-2

OnSSET, 2022, OnSSET Documentation, https://onsset.readthedocs.io/en/latest/references.html

Pereira, L. S., Paredes, P., & Jovanovic, N. (2020). Soil water balance models for determining crop water and irrigation requirements and irrigation scheduling focusing on the FAO56 method and the dual Kc approach. Agricultural water management, 241, 106357.

Portmann, F. T., Siebert, S. & Döll, P. MIRCA2000—Global monthly irrigated and rainfed crop areas around the year 2000: A new high-resolution data set for agricultural and hydrological modeling. Glob. Biogeochem. Cycles 24, (2010).

R. G. Allen, L. Pereira, D. Raes, M. Smith, Crop evapotranspiration: Guidelines for computing crop water requirements, FAO Irrig. Drain. Pap. 56 (1998) 26–40.

Rob Dellink, Jean Chateau, Elisa Lanzi, Bertrand Magné, Long-term economic growth projections in the Shared Socioeconomic Pathways, Global Environmental Change, Volume 42, 2017,

Samir KC, Wolfgang Lutz, The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100, Global Environmental Change, Volume 42, 2017, Pages 181-192, ISSN 0959-3780, DOI:10.1016/j.gloenvcha.2014.06.004

Tuninetti, M., Tamea, S., D'Odorico, P., Laio, F. & Ridolfi, L. Global sensitivity of high-resolution estimates of crop water footprint. Water Resour. Res. 51, 8257–8272 (2015).

URL https://data.apps.fao.org/map/catalog/srv/eng/catalog.searc

Vinca, A., Parkinson, S., Byers, et al. The NExus Solutions Tool (NEST) v1.0: an open platform for optimizing multi-scale energy–water–land system transformations, Geosci. Model Dev., 13, 1095–1121, https://doi.org/10.5194/gmd-13-1095-2020, 2020.

Wada, Y., Wisser, D., and Bierkens, M. F. P.: Global modeling of withdrawal, allocation and consumptive use of surface water and groundwater resources, Earth Syst. Dynam., 5, 15–40, https://doi.org/10.5194/esd-5-15-2014, (2014)

WorldPop, "Open Spatial Demographic Data and Research," 2022. https://wopr.worldpop.org/?ZMB/ (accessed May 13, 2022).

Appendix: Data tables and sources per model

WaterCROP input data lists and sources

The table below lists and describes the different datasets and their sources required as inputs to WaterCROP

Table A 1: WaterCROP input data sources

Data	Units	Type & resolution	Description	Source	Relation to RE4AFAGRI platform models
Climate					
Simulated Potential Evapotranspiration	mm/day	Raster layers 30x30 arc min	Daily potential evapotranspiration (2006-2099); - RCP2.6 & RCP6.0; - 2005 land use, fertilizer input and nitrogen deposition; - Regular CO2 experiment;	ISIMIP 2b - hydrological model: PCR-GLOBWB - GCM: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5	Consistent with NEST
Historical Potential Evapotranspiration	mm/day	Raster layers 30x30 arc min	Daily potential evapotranspiration (1861-2005); - GCMs based historical climate; - varying direct human influence; Regular CO2 experiment;	ISIMIP 2b - hydrological model: PCR-GLOBWB - GCM: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5	Consistent with NEST
Observed historical Potential Evapotranspiration	mm/day	Raster layers 10x10 arc min	Daily potential evapotranspiration (1961-1990);	FAO map catalogue, derived form CRU CL v2.0	Consistent with NEST
Simulated Precipitation	mm/day	Raster layers 30x30 arc min	Daily precipitation (2006-2099); - RCP2.6 & RCP6.0; - 2005 land use, fertilizer input and nitrogen deposition; - Regular CO2 experiment;	ISIMIP 2b - GCM: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5	Consistent with NEST
Historical Precipitation	mm/day	Raster layers 30x30 arc min	Daily Precipitation (1861-2005); - GCMs based historical climate; - varying direct human influence; Regular CO2 experiment;	ISIMIP 2b - GCM: GFDL-ESM2M, HadGEM2-ES, IPSL-CM5A-LR, MIROC5	Consistent with NEST

Observed historical Precipitation	mm/day	Raster layers 10x10 arc min	Daily Precipitation (1961-1990);	CRU CL v2.0	Consistent with NEST
Land					
Historical crop areas	ha	Raster layers 10x10 arc min	Area of irrigated and rain-fed crops in the most recent year	SPAM 2017 v2.1 Sub-Saharan Africa	Consistent with M-LED
Historical crop yields	kg/ha	Raster layers 10x10 arc min	Current local yields, per crop and irrigation technique	SPAM 2017 v2.1 Sub-Saharan Africa	Consistent with M-LED
Growing period lengths	days	Raster layers 5x5 arc min	Relative to year 2000	MIRCA2000 Portmann et al. (2010)	Consistent with M-LED
crop coefficients & proportional length of each growing stage		Crop specific values		Chapagain et al. (2004).	Consistent with M-LED

M-LED input data lists and sources

The tables below lists and describes the different datasets and their sources required as inputs to M-LED as well as their relation to other RE4AFAGRI platform models.

Table A 2: M-LED input data sources

Data	Units	Type & resolution	Description	Source	Relation to RE4AFAGRI platform models
Socioeconomic & geophysics					
Country administrative boundaries	-	Vector polygons	Level 0, 1, 2, 3 of national boundaries	GADM v4.1, Hijmans et al. (2022), https://gadm.org/data.html	Consistent with administrative boundary definition of other models in the RE4AFAGRI platform
Roads density	Density in polygon (GIS calculated)	Vector polygons	Roads shapefile	GRIP global roads database, https://www.globio.info/download-grip-dataset	-
DHS wealth and employment distribution	%	Vector polygons, sub-national units resolution	% of men and women employed	DHS Spatial Data Repository	-
Travel time to nearest city/market	minutes	Raster layer	Minutes of travel time via the fastest travel mode to reach the nearest city of at least x,000 thousand inhabitants. (GIS calculated from cities point shapefile and friction layer raster)	World Cities database; friction layer raster: Weiss et al. (2020)	Consistent with **OnSSET**
Gridded population and population clusters vector file	pop/km2	Raster layer and polygons shapefile	Daytime population distribution and extent of settlements/cities	GRID3 (https://data.grid3.org/)	Consistent with gridded population and population clusters used in **OnSSET**
Urban rural categorisation	% of population living in urban areas	Downscaled national statistic	Category of settlement; SSP-consistent	Jiang and O'Neill (2017)	Consistent with **OnSSET** calibration approach

Electricity access	%	Vector polygons	Electrification status of each population cluster		Derived from **OnSSET** calibration
BCU (river Basin Catchment Units)	-	Vector polygons	Obtained by intersecting Hydrosheds water basin polygons and administrative boundaries polygons	HydroSheds (2021) and GADM (2022)	Derived from **NEST**
Other infrastructure					
Healthcare facilities	Tier	Vector points	Different levels of healthcare (for social infrastructure energy demand)	Maina et al. / WHO (2019)	-
Schools	-	Vector points	Schools (for social infrastructure energy demand)	Humanitarian Data Exchange	-
Energy					
Diesel price	USD/l	Raster layer	Based on recent prices from https://data.worldbank.org/indicator/EP.PMP.DESL.CD?locations=ZG and travel time to city	World Bank and Szabo et al. (2011) adjustment for transport	Consistent with **OnSSET**
Crop-specific energy requirements for crop processing	kWh/ton	CSV file	kWh/ton, based on literature review	Falchetta et al. (2020)	-
Prevalent energy access tier	Category	Raster layers	Connected to WB-MTF of energy access	Falchetta et al. (2019)	-
Climate					
Climate zones	Category	Raster layer	Coppen-Geiger classification	GAEZ	-
Dawn/dusk times	Hour of the day, per month	Raster layer	To design appliances use patterns	suncalc R package	-
Land					
Historical crop areas	ha	Raster layers	Area of irrigated and rain-fed crops in the most recent year	SPAM 2017 v2.1 Sub-Saharan Africa	Consistent with **WaterCROP**
Historical crop yields	kg/ha	Raster layers	Current local yields, per crop and irrigation technique	SPAM 2017 v2.1 Sub-Saharan Africa	Consistent with **WaterCROP**

Irrigation statistics	%	Country-level statistics	% of agricultural land equipped for electricity-powered irrigation	FAO Aquastat	-
Crop prices	USD/kg	Country-level statistics	Wholesale prices per crop	FAOStat	-
Field size	Category	Raster layer	To identify smallholder farmed land	Fritz et al. (2015)	-
High-resolution cropland extent	Cropland presence	Raster layers	To downscale coarse crop-specific maps (e.g. MAPSPAM)	GFSAD30	-
Yield gain potential thanks to irrigation	% change on top of current yields	Raster layers	Yield growth potential if irrigation is supplied (estimated)	Tuninetti et al.	Derived from **WaterCROP**
Crop calendar	Date of the year	Country-level statistics	Crop specific and country specific plantation and harvest period(s)	FAO Crop Calendar	Derived from **WaterCROP**
Water					
Crop-specific irrigation demand	m3/ha/month	Raster layers	Crop specific, location-specific, evapotranspiration-based calculation	Tuninetti et al.	Derived from **WaterCROP**
Depth, productivity, size of groundwater	m. mm/h, m	Raster layers	Groundwater depth to calculate pumping energy needs	MacDonald et al. (2011)	-
Distance to surface water	m	Raster layers	Permanent surface water basin distance to evaluate extraction potential	HydroSheds (2021)	-
Groundwater recharge sustainability constraints	mm/hour/month	Raster layers	Recharge should not be exceeded by withdrawals to preserve the hydrological balance	ISIMIP 3b	Consistent with **NEST**

Table A 3: M-LED parameters and main numerical assumptions (country-specific)

Parameter	Unit(s)	Description	Source
Socio-economic and energy statistics			
national_official_population	People	National population in baseline year (2020).	World Bank Data https://data.worldbank.org/indicator/SP.POP.TOTL
national_official_elrate	% of population	National share of the population classified as "with access to electricity" in baseline year (2020).	World Bank Data https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS
ppp_gdp_capita	2011 constant USD (PPP)	GDP per capita based on purchasing power parity (PPP) in baseline year (2020).	World Bank Data https://data.worldbank.org/indicator/NY.GDP.PCAP.PP.KD
gini	-	Gini index measures the extent to which the distribution of income (or, in some cases, consumption expenditure) among individuals or households within an economy; most recent available country value.	World Bank Data https://data.worldbank.org/indicator/SI.POV.GINI
electr_final_demand_tot	kWh/yr	Yearly total final electricity demand in the country (electricity produced + imports - exports - losses) in baseline year (2020).	IEA World Energy Balance https://www.iea.org/countries
residential_final_demand_tot	kWh/yr	Yearly final electricity demand in the country in the residential sector in baseline year (2020).	IEA World Energy Balance https://www.iea.org/countries
industry_final_demand_tot	kWh/yr	Yearly final electricity demand in the country in the industry sector in baseline year (2020).	IEA World Energy Balance https://www.iea.org/countries
cropland_equipped_irrigation	% of agricultural land	% of agricultural land equipped for electrical powered irrigation systems in baseline year (2020).	FAO Aquastat https://tableau.apps.fao.org/#/views/ReviewDashboard-v1/country_dashboard
urban_hh_size	people/HH	Average size of an urban household in the country in baseline year (2020).	Global Data Lab (DHS surveys) https://globaldatalab.org/
rural_hh_size	people/HH	Average size of a rural household in the country in baseline year (2020).	Global Data Lab (DHS surveys) https://globaldatalab.org/
Economic parameters			
discount_rate	%	Discount rate to estimate CBA of standalone	Assumption

		solar pumping systems installation	
Boundary conditions and thresholds			
pop_threshold_productive_loads	people / population cluster	Minimum number of people to make a cluster eligible for crop processing activities to take place (local crop processing electricity demand >0)	Assumption
m_radius_buffer_cropland_distance	metres	Maximum distance of cropland to include load in community load (radius buffer in meters from cluster centroid)	Assumption
threshold_surfacewater_distance	metres	Distance threshold which discriminates if groundwater pumping is necessary or a surface pump is feasible	Focus group discussions (Deliverable D12.2)
threshold_groundwater_pumping	metres	Maximum aquifer depth at which the model allows for groundwater pumping	Focus group discussions (Deliverable D12.2)
maxflow_boundaries	m3/hour	Max and min boundaries for the flow of irrigation pumps, in m3/s	Focus group discussions (Deliverable D12.2)
Technological and energy parameters			
lifetimepump	years	Lifetime of the pump, for solar pumping CBA	Focus group discussions (Deliverable D12.2)
eta_pump / eta_motor	%	Efficiency parameters of the water pump	Focus group discussions (Deliverable D12.2)
range_tank	m3	Water storage tank range	Focus group discussions (Deliverable D12.2)
tank_usd_lit	USD/liter	Water storage tank cost	Assumption
slope_limit	%	Maximum average terrain slope threshold for surface pumping being considered at cluster i	Assumption
fuel_consumption	l/h	(l/h) (for CBA modules)	OnSSET
fuel_cost	USD/l	Retail price of petrol (for CBA modules)	OnSSET
truck_bearing_t	ton/truck	Truck bearing for transporting additional yield to nearest market (for CBA modules)	OnSSET
minutes_cluster	minutes	Minutes of travel time around each settlement to create crop processing clusters (areas of cropland around communities eligible for crop processing to take place)	Assumption
load_curve_cp	% of daily consumption per hour of the day	Hours of the day when crop processing is assumed to take place	Focus group discussions (Deliverable D12.2)
load_curve_irrig	% of daily consumption per hour of the day	Hours of the day when irrigation is assumed to take place	Focus group discussions (Deliverable D12.2)

range_smes_markup	%	Maximum and minimum size of the non-farm SME demand relative to the cluster's residential electricity demand. Proportional on roads density and distance to nearest city	Moner-Girona et al. (2019)
nhours_irr	#hours	Average number of hours of irrigation days when the pump is in operation	Focus group discussions (Deliverable D12.2
load_curve_prod_act	% of daily consumption per hour of the day	Hours of the day when non-farm SMEs is assumed to take place	Focus group discussions (Deliverable D12.2
battery_buffer	%	Buffer size of the solar pumps battery pack	Assumption
battery_efficiency	%	Efficiency of the solar pumps battery pack	Assumption
depth_discharge	%	Depth of discharge of the solar pumps battery pack	Assumption
usdperkwhbattery	USD/kWh	Marginal cost of the solar pumps battery pack	Assumption

OnSSET input data lists and sources

The table below lists and describes the different datasets and their sources required as inputs to OnSSET as well as any validation efforts already applied to them in this project.

Table A 4: OnSSET input data sources for Zambia

Parameter or Dataset	Unit	Type or Value	Description	Source	Validation or model linkage
Geospatial Datasets (GIS)					
Administrative boundaries (National and sub-national borders)	-	vector geometry	Delineates the boundaries of the analysis.	GADM	Consistent with other RE4AFAGRI models.
BCU (river Basin Catchment Units)	-	vector geometry	Obtained by intersecting Hydrosheds water basin polygons and administrative boundaries polygons	HydroSheds (2021) and GADM (2022)	Derived from NEST delineations. Matched between M-LED, NEST, and OnSSET for consistency.
Elevation	meters above sea level	raster	Elevation maps are used in several processes in the analysis (Energy potentials, restriction zones, grid extension suitability map etc.).	NASA Shuttle Radar Topography Mission	90x90m resolution elevation maps used. Assumed to be adequate. Further improvements considered minimal.
Hydropower	various	vector points	Points showing potential mini/small hydropower potential. Provides power availability in each identified point.	Korkovelos, et. al. 2018 "A Geospatial Assessment of Small-Scale Hydropower Potential in Sub-Saharan Africa" https://www.mdpi.com/1996-1073/11/11/3100	Korkovelos, et. al. 2018 (default of OnSSET) used without further adjustments
Land Cover	category	raster	Land cover maps are use in a number of processes in the analysis (Energy potentials, restriction zones, grid extension suitability map etc.).	https://lpdaac.usgs.gov/products/mcd12q1v006/ or https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MCD12Q1	Dataset updated to use USGS "mcd12q1v006" categorical landcover maps. Updated from previously discontinued data.

Night-time Lights	-	raster	Dataset used to, identify and spatially calibrate the currently electrified/non-electrified population.	Payne Institute - Earth Observation Group https://eogdata.mines.edu/products/vnl/#annual_v2 (Processed from VIIRS satellite)	Updated from previous 2016 maps to use composite of 2015-2020 data from eogdata.mines.edu
Population	various	vector geometries and raster surfaces	Spatial identification and quantification of the current (base year) population.	GRID3 Settlement Extents: https://data.grid3.org/datasets/GRID3::grid3-zambia-settlement-extents-version-01-01-/about & WorldPop Population Counts: https://wopr.worldpop.org/?ZMB/	Extensive recreation and revalidation of population clusters using GRID3 and WorldPop and updated future projections to match SSPs. More details in the text. Updates benchmarked against existing GEPv2 Data.
Roads	-	vector lines	Current road infrastructure is used to specify grid extension suitability.	GRIP global roads database https://www.globio.info/download-grip-dataset	Updated to match inputs with M-LED and OnSSET
Solar irradiation	kWh/m2/year	raster	Provide information about the Global Horizontal Irradiation over an area. This is later used to identify the availability/suitability of Photovoltaic systems.	https://globalsolaratlas.info/download	Updated to use latest data from globalsolaratlas.info (improved resolution from 1km x1km to 250m x 250m)
Wind speed	m/s (at 100m) annual average	raster	Provide information about the wind velocity over an area. This is later used to identify the availability/suitability of wind power (using Capacity factors).	https://globalwindatlas.info/en/download/gis-files	Updated to use latest data from globalwindatlas.info (improved resolution from 1km x1km to 250m x 250m)
Total electricity demand of settlements	kWh/year	Settlement boundary geometries	Provides the total electricity consumption estimate for the settlement per sector	M-LED model output	M-LED values are used instead of previous methodology.

Travel-time	minutes	raster	Minutes of travel time via the fastest travel mode to reach the nearest city of at least 50 thousand inhabitants.	World Cities database; friction layer raster: Weiss et al. (2020)	Default from Weiss et. al. used – same dataset as in M-LED
Electricity Substations	kVa	vector points	Current substation infrastructure is used to specify grid extension suitability	-	Not used in analysis for Zambia. Validated lines data supersede this.
Electricity grid (existing)	kV	vector lines (Separate for HV and MV lines)	Current grid network. High voltage and medium voltage lines. Low voltage lines not mapped.	Gridfinder: https://gridfinder.org/ Open Street Map Export: https://overpass-turbo.eu/ Nighttime lights: https://eogdata.mines.edu/products/vnl/#annual_v2	Extensive cross validation and improvement with multiple data sources. Benchmarked against GEPv2 OSM+gridfinder data.
Electricity grid (planned)	kV	vector lines (Separate for HV and MV lines)	Planned/committed grid network extensions	-	Not currently used in analysis.
Other techno-socio-economic parameters or assumptions					
National Electrification Rate	%	National: 44%	National share of the population classified as "with access to electricity" in baseline year (2020).	World Bank Data https://data.worldbank.org/indicator/EG.ELC.ACCS.ZS	Updated to match values used in M-LED with latest available data.
Average rural household size Average urban household size	people	Different per zone based on national survey data	Gives the total estimated average persons per household in the rural and urban areas.	Global Data Lab (DHS surveys) https://globaldatalab.org/	Updated to match values used in M-LED with latest available data.

Expected Hydro mini-grid cost:	$/kWp	3000	These costs are used to calculate the total system costs of minigrids of various technologies and combinations	International Renewable Energy Agency, "Renewable Power Generation Costs in 2017," IRENA, Innovation Outlook: Renewable Mini-Grids. 2016. IRENA, "Solar PV in Africa: Costs and Markets," 2016. A. Korkovelos et al., "The Role of Open Access Data in Geospatial Electrification Planning and the Achievement of SDG7. An OnSSET-Based Case Study for Malawi Apr. 2019,	Unchanged from previous benchmarked values. Same values as in GEPv2.
Expected Hybrid mini-grid cost:	$/kWp $/kWh $/kWp $/kW $/kW $/kW	· PV panels: 503 · Batteries: 139 · Inverter: 80 · Charge controller: 142 · Diesel generator: 261 · Wind turbine: 2800			Unchanged from previous benchmarked values. Same values as in GEPv2.
Expected PV stand-alone (or SHS) costs:	$/kWp	~9620 $/kWp if kW < 0.02 ~8780 $/kWp if 0.02< kW < 0.05 ~6380 $/kWp if 0.05< kW < 0.1 ~4470 $/kWp if 0.1< kW < 1 ~6950 $/kWp if kW > 1	These costs are used to calculate the cost of SHS of different sizes (economies of scale)		Unchanged from previous benchmarked values. Same values as in GEPv2.
Grid electricity wholesale generation cost	$/kWh	Taken from each scenario run of the NEST model	The grid electricity generation cost is the wholesale total cost of electricity generation levelized over the operating lifetime of the energy system.	NEST model outputs	Updated to use values from NEST modelling. Previous GEPv2 values no longer used.

Transmission and Distribution (T&D) costs	$/km $/km $/km $/unit $/unit $/unit	HV line (69-132 kV): ~53000 MV line (11-33 kV): ~7000 LV line (0.2 – 0.4 kV): ~4250 HV to MV substation (1000 kVA): ~25000 MV to V substation (400 kVA): ~10000 Service transformer (50 kVA): ~4250	T&D costs of each appropriate type, size, and distance are calculated to determine the total electric transmission and distribution cost from the nearest suitable existing infrastructure to each settlement in the area of study.	Energy Sector Management Assistance Program (ESMAP), “Model for Electricity Technology Assessment (META).” 2014. R. Karhammer et al., “Sub-Saharan Africa: Introducing Low-Cost Methods in Electricity Distribution Networks,” ESMAP October 2006.	Unchanged from previous benchmarked values. Same values as in GEPv2.

NEST input data lists and sources

The tables below list and describe the different datasets and their sources required as inputs to NEST and indicates where inputs are coming directly from another associated model.

Table A 5: NEST input data table

Data & description	Units	Type or Value	Source
Socioeconomic & geophysics			
Country administrative boundaries	-	polygon	Database of Global Administrative Areas (GADM)
Basin and sub-basin boundaries	-	polygon/15 arc-second resolution	HydroSheds
Urban and rural population	million people	gridded (1 km)	Population Grids
Urban and rural GDP	$	country	M-LED
Energy, data needed for country prototype and rural energy			
Historical energy supply and demand by sector	GWh/year or month	national scale	IEA
import-export of energy commodity	GWh/year or month	national scale	IEA
Installed Power plant Capacity	MW	points spatial	World Resource Institute Power plant database
Non-hydro-power-plant capacity, age and location	MW	points spatial	World Electric Power Plant (WEPP) Database 2018
Power plant cooling technologies: cost and water, energy consumption, historical capacity	various	points spatial	Raptis et al. (2016)
Power plant cost and performance: cost of technologies and power plants	$	country	Parkinson et al. (2016); Fricko et al. (2016) technology level 2014
Historical transmission capacity and roads	MW	gridded/shapefile	OpenStreetMap
Resource extraction potential: values and estimate of growth rate to 2050 for coal, lignite, crude, gas	GJ	country	local sources
Resource extraction costs	$	country	MESSAGEix-GLOBIOM assumptions
Rural electricity generation			

Sectoral electricity demand (off-grid)	kWh/km2	gridded	M-LED
Access to Energy: % people needing access to energy	%	gridded	M-LED, Assumption
mini-grid and grid extension costs	$/kw/km	country values	OnSSET
small generation potential, solar, wind, small hydro	kW	gridded	OnSSET
small generation costs, solar, wind, small hydro	$/kW, $/kWh	province, country	OnSSET assumptions
Land			
Historical crop areas	km2	gridded	waterCROP
Historical crop yields	ton/ha	gridded	waterCROP
Historical crop production	ton/month	gridded or per province	waterCROP
Historical irrigation water supply by crop/water source	km3/month	gridded	waterCROP
Irrigation cost and performance	$, km3/ton DM	country values	Vinca et al. 2020
Water			
Projected Daily Runoff	kg/m2/sec	0.5°	ISIMIP 2b
Projected Daily Discharge	m3/sec	0.5°	ISIMIP 2b
Projected Monthly Groundwater recharge	kg/m2/sec	0.5°	ISIMIP 2b
Depth to groundwater	m	gridded	Fan et al. 2013; UN-IGRAC
Access to Water , sanitation, recycling	% population	country data	WHO, WRI aqueduct
Historical water demand by sector	km3/month	gridded	Wada et al 2014,2016
Projections for water demand	km3/month	gridded	Wada et al 2014,2016
Surface and groundwater performance: costs and energy consumptions	$/m3, kwh/m3	local/country	Vinca et al 2020
Costs and performance of water distribution & treatment	$/m3	local/country	Vinca et al 2020