Designing a zero-order energy transition model: a guide for creating a Starter Data Kit

Carla Cannone (C.Cannone@lboro.ac.uk)
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom; Imperial College London, United Kingdom https://orcid.org/0000-0002-1214-8913

Lucy Allington
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom https://orcid.org/0000-0003-1801-899X

Karla Cervantes Barron
University of Cambridge https://orcid.org/0000-0001-9185-3022

Flora Charbonnier
University of Oxford https://orcid.org/0000-0003-3174-0362

Miriam Zachau Walker
University of Oxford https://orcid.org/0000-0002-4757-3688

Claire Halloran
University of Oxford https://orcid.org/0000-0002-0308-5623

Rudolf Yeganyan
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom; Imperial College London, United Kingdom https://orcid.org/0000-0001-9325-8147

Naomi Tan
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom; Imperial College London, United Kingdom https://orcid.org/0000-0001-7957-8451

Jonathan M Cullen
University of Cambridge https://orcid.org/0000-0003-4347-5025

John Harrison
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom https://orcid.org/0000-0002-6434-5142

Long Seng To
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom https://orcid.org/0000-0003-4676-5810

Mark Howells
Centre for Sustainable Transitions: Energy, Environment & Resilience (STEER), Loughborough University, United Kingdom; Imperial College London, United Kingdom https://orcid.org/0000-0001-6419-4957

Method Article

Keywords: Energy System Modelling, Data Collection Tool, OSeMOSYS, clicSAND, U4RIA

Posted Date: March 1st, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1403613/v1

Abstract

192 Parties joined the Paris Agreement, which includes commitments to reduce emissions. However, developing national decarbonisation strategies requires significant analyses and investment. Analyses for such strategies are often delayed due to a lack of accurate and up-to-date data for creating energy transition models. Starter Data Kits address this issue by providing open-source, zero-level country datasets to accelerate the energy planning process. There is a strong demand for replicating the process of creating Starter Data Kits because they are currently only available for 69 countries in Africa, Asia, and South America. Using an African country as an example, this paper presents the methodology to create a Starter Data Kit from data collection to the creation of tool-agnostic data repositories. The paper illustrates the steps involved, provides additional information for conducting similar work in Asia and South America, and highlights the limitations of the current version of the Starter Data Kits. Future development is proposed to expand the datasets, including new and more accurate data and new energy sectors.

Value Of The Methodology

Decarbonisation strategies have been committed to by 192 countries as part of the Paris Agreement [1]. Approximately 72% of those emissions are attributed to energy. The energy system is complex: it has long-term infrastructure and is subject to physical constraints. Further, it can be expensive. Modelled scenarios – based on quantified data that are thermodynamically consistent – can be used to estimate emissions and costings. However, the data can be time-consuming to collect and can come from proprietary sources; and similarly, the software used to model the energy system can require specialist skills. Further, the data are dynamic. Once compiled in the correct form, it can be onerous to recompile a year later if the source data has changed, particularly if the methods have been documented. The Starter Data Kits solve this problem by ensuring a repeatable method which uses open data to produce ready-to-use outputs. These are further treated and entered into an open-source model with an easy-to-use interface.

The methodology presented here for creating a Starter Data Kit has been adopted for Africa, Asia, and South America [2]. The added value of this data pipeline is that it can be easily adapted for any other country and aggregated to any region; therefore, anyone using this blueprint can create a new Starter Data Kit for a country of their choice.

This novel technique has been adapted for use with the clicSAND software [3] for the Open Source energy Modelling System (OSeMOSYS) [4]. However, it is important to note that the result is an open collection of tool-agnostic data. Users can use the country datasets with tools other than the one presented here and improve the accuracy of the data by pointing to new sources or adding new data. To date, the methodology has been used in connection with several other tools and studies, for example:

- By academics at the University of Mauritius to develop new Starter Data Kits for their country with the aim of promoting regional collaboration between the African Small Island Developing States (SIDS) for collective effort to encourage decarbonisation and sustainable development, and investigate explorative energy scenarios [5][6][7][8];
- In combination with other tools such as the European Climate Calculator model (EUCalc) for a country model of Qatar [9];
- In China, Starter Data Kits will be combined with the China 2050 Calculator (jointly developed by China and the UK) to deduce China's energy consumption in 2050, providing a basis for Chinese policymakers to formulate environmental policies and help China achieve its carbon neutrality goals;
- In Kenya, the Starter Data Kits have provided the foundation Reference Energy System (RES) for two energy planning tools for the Kenyan government: a standalone power system model indicating new investment requirements in powerbase generation capacity and an integrated whole energy system model. Applying this methodology resulted in saved computational time but ensured the key dynamics of supply and demand variability were maintained;
- Together with an investigation of the potential for accelerated Electric Vehicle deployment in Kyrgyzstan [10];
- To analyse the benefits and differences in energy pathways of various scenarios, including a Business as Usual, Fossil Fuel, and Net Zero scenario in Vietnam; and
- To build a zero-order model for Armenia and analyse alternative development strategies.

Specifications Table

Subject Area	Energy
More specific subject area	Energy System Modelling
Method name	Data Collection Method for Starter Data Kits models
Name and reference of original method	Not applicable
Resource availability	Links to Zenodo Repositories in the Annex

Method Details

This document provides instructions on the steps and materials required to develop a Starter Data Kit. Currently, 69 Starter Data Kits have been developed for Africa, East Asia, and South America. The methodology presented here is intended to encourage practitioners to apply it to new countries and expand the current Starter Data Kits library. It is a novel process that creates data pipelines that feed into a single Data Collection and Manipulation Tool. This allows for the creation of data in a consistent format ready for a modelling analysis using one of the available tools.

Data Downloading Requirements

The methodology described in this paper refers to different databases and scripts. For ease of use, all the databases and scripts needed are presented in Table 1, in case the reader wishes to download all relevant files for the region of interest before working on an example.

Table 1: List of databases and scripts needed for to repeat this methodology

Name	File(s)	Link	Source	Description
CCG Starter Kits - Base SAND file for Africa	Africa_base_SAND.xlsm	https://zenodo.org/record/6011229	[11]	This file is the Base SAND file for Africa.
Starter Kits - Base SAND file for South America	South_America_base_SAND.xlsx	https://zenodo.org/record/6108078	[12]	This file is the Base SAND file for South America.
Starter Kits - Base SAND file for Asia	Asia_base_SAND.xlsx	https://zenodo.org/record/6109104	[13]	This file is the Base SAND file for Asia.
CCG Starter Kits - Base SAND file for Africa Coal and Natural Gas	Africa_COA_NGS_base_SAND.xlsm	https://zenodo.org/record/6033764	[14]	This file is the Base SAND file for Africa with coal and natural gas.
Starter Kits - Base SAND file for Africa Natural Gas Scenario	Africa_NGS_base_SAND	https://zenodo.org/record/6036135	[15]	This file is the Base SAND file for Africa with natural gas.
CCG Starter Kits - Base SAND file for Asia – Coal and Natural Gas	Asia_COA_NGS_base_SAND	https://zenodo.org/record/6036147	[16]	This file is the Base SAND file for Asia with coal and natural gas.
CCG Starter Kits - Base SAND file for South America – Coal and Natural Gas Scenario	South_America_COA_NGS_base_SAND	https://zenodo.org/record/6036192	[17]	This file is the Base SAND file for South America with coal and natural gas.
CCG Starter Kits - Script for CSV Creation with Base SAND Data for Starter Kits	ccg_data_to_sand.py	https://zenodo.org/record/6036155	[18]	Script to create comma-separated-value (CSV) file for base SAND file from the data collection file.
Starter Kit - List of Countries and Regions	Starter Kit - List of Countries and Regions	https://zenodo.org/record/6036252	[19]	Mapping of countries to their relevant region.
Starter Kits - Technology-specific	Several csv files	https://zenodo.org/record/6142375	[20]	These files contain the Capacity Factors and

Data for Base SAND File				Residual Capacity values for all countries needed for filling in the information in the base SAND file.
Starter Kits - Scripts for SAND Scenarios	CCG_SAND_Scenario_XXX.py	https://zenodo.org/record/6136904	[21]	These files contain scripts to run the scenarios used in the Starter Kits. Each script creates a CSV file with the parameters for a certain scenario (corresponding to the script name) for a country or list of countries. The scenarios included are Fossil Future (FF), Least-Cost (LC), and Net Zero (NZ).

The remainder of the paper is structured as follows: Section 1 describes the main activities to be carried out with the Data Collection and Manipulation Tool (DaCoMaTool). This first section contains two separate subsections for region- and country-specific data. Once all data have been collected and put into a coherent format, they are transferred from the DaCoMaTool to the clicSAND software. Section 2 describes two alternative ways of performing this task: a manual and an automated process. Three copies of the output, the Base SAND file, are created to develop three scenarios for each country: Fossil Future, Least Cost, and Net Zero (described in Section 3). Section 4 then explains how to run a model using the clicSAND software and how to visualise the results. Finally, Section 5 describes how to upload the results as preprint articles to Research Square and the tool agnostic Zenodo Data Repository.

1. Data Collection And Manipulation Tool (Dacomatool)

DaCoMaTool is an original tool designed to accelerate and automate the Starter Data Kit development process. It is an Excel Workbook made of multiple tabs, where the user can collect and manipulate data from various sources and organise it into an easy-to-use format in a single open collection. This data is ready for use by analysts and is energy modelling software agnostic. However, DaCoMaTool was also designed to be fully compatible with clicSAND software [3] for the Open Source energy Modelling SYStem – OSeMOSYS [4]. The user is not required to use OSeMOSYS and can instead select these datasets for use with other tools and for different analyses. In this paper, we will present, as an example, the steps to create a Starter Data Kit for a country in Africa, starting with the parameters for the Excel Workbook. Further information on the definition and the meaning of each of the following parameters described below is freely available from the OSeMOSYS course hosted on the Open Learn Create (OLC) platform for Windows and MacOS users [22] [23].

Region-specific Data Collection and Manipulation

The method and source data recorded in the Data Collection and Manipulation File for a country contains data that will remain the same for all countries in that region. Therefore, for Africa, by way of example, the method is the same for all its continental countries. Below is a summary of where the region-specific data for Africa were sourced and how they were manipulated. Annex A includes this information for the Asian and Latin American regions. Further details on the data sources can be found in the Data in Brief article [2] and in the country Data Collection and Manipulation file, under the column "Comments" in each tab.

Depreciation Method, Discount Rate and Year Split

Defaults values were used for these parameters: Depreciation Method = 1; Discount Rate = 0.1; Year Split = 1/96 for all 96-time slices so that they are of equal length.

Technologies Costs – Africa

The Electricity Model Base for Africa (TEMBA) report was used to collect the costs of refinery and transmission and distribution technologies [24][25]. Costs for renewables were taken from IRENA (2021) Planning & Prospects for Renewable Power: Eastern & Southern Africa [26] – this provides 5-yearly projected costs to 2040, with linear change assumed between datapoints and costs assumed constant after 2040. The costs of offshore wind were taken from IRENA (2019) The Future of Wind [27]. Costs for fossil power plants were taken from IRENA (2018) Planning & Prospects for Renewable Power: West Africa [28] (because the gas and coal costs were deemed less realistic in the 2021 report). Generic costs of transport and heating technologies were taken from Terpilowski Gill (2020) Decarbonising the Laotian Energy System [29]. The costs of energy efficiency technologies were estimated based on the costs of coal power plants – see [30]. Costs of stoves were taken from Okolo O, Teng H. (2017) Analysing Nigeria's Energy system in light of the UN's Sustainable Development Goals [31]. Costs of renewables with storage were estimated by combining the standard cost from the IRENA report with an estimated storage cost based on the NREL 2020 Annual Technology Baseline [32] – see [33]. The oil price was estimated based on the forecast in U.S. EIA. (2020) Assumptions to the Annual Energy Outlook 2020: International Energy Module [34], which was extended to 2070 – see [35]. The price was increased by 10% for imported oil, and the prices of imported HFO and LFO were estimated by calculating 80% and 133% of the domestic oil price, respectively, as done in TEMBA [24]. Prices of other fuels were taken from [28].

Input & Output Activity Ratios

Region-specific efficiencies for power transmission and distribution were used. For Africa, efficiencies of power plants were taken from [28] and efficiencies of stoves were taken [31].

Capacity to Activity Unit, Operational Life & Capacity Factors

Default values were used for the Capacity-to-Activity unit; for more information on default data see [36]. Operational Lifetimes for the facilities were taken from the same reports used for technology costs in each region. Capacity factors for fossil, biomass, and nuclear power plants were also taken from those reports. Country-specific capacity factors for wind, solar, and hydropower technologies were used, sourced from the PLEXOS dataset [37] (based on Renewables Ninja [38]) and global NREL datasets [39] where necessary (detailed in the following sections).

Emissions Factors

Emissions factors were sourced from Table 1.3 of IPCC (2006) Guidelines for National Greenhouse Gas Inventories [40].

Country-specific Data Collection and Manipulation

Here we will present the example of a country in Africa.

Country Selection

Create a copy of the Africa Base Data Collection File (without opening the file) for the selected country in the same Data Preparation and Manipulation folder. Name the file: **New *countryname* Data Collection**. Then open the file, which may take some time to upload.

a) The links to the file need to be updated so all the linked data is available. To do so, click on the *Data tab -> Edit Links* (Figure 2).

b) This will display a menu with the different linked worksheets. Select each and click *Change Source*. This will open up an explorer tab which allows the user to navigate to the folder with all the capacity factor and residual capacity data. The appropriate file can then be selected, and some time for links to update may be needed.

Once all the links are working, the user can select the country in Tab 1. Model Initiation. Again, some time may be needed to update the linked values in the workbook.

After the relevant data for the country of interest has been selected, the user can decide to break the links to the different worksheets updated earlier (instructions a–b) by clicking on the option of *Break Link*. This should make the spreadsheet work much faster; however, the user must be careful to select the right values to avoid having to return to the spreadsheet to change the country at a later stage.

Power Transmission and Distribution output activity ratios

Copy the output activity ratio rows for power transmission and distribution from the **TEMBA_21_10_Refer** dataset found in [25] to the rows in the **3.3 Output Activity Ratios tab**. In this dataset, the codes for power transmission and distribution technologies are ***XXEL00T00X*** and ***XXEL00TDTX,*** respectively, where XX is the country code, which can be obtained from the file **PLEXOS Country Codes.xlsx** from the same repository as the output activity ratio [20].

- In the TEMBA Output Activity Ratios tab, please copy the entire row for XXEL00T00X from column D to column BG and paste it into row 41 of tab 3.3 Output Activity Ratios, starting in column F
- In the TEMBA Output Activity Ratios tab, please copy the entire row for XXEL00TDTX from column D to column BG and paste it into row 42 of tab 3.2 Output Activity Ratios, starting in column F

PV, onshore & offshore wind and hydro capacity factors

This process is automated. The data for the selected country should be inserted in the Raw PV/Onshore Wind/Hydro/Offshore Wind CFs tabs (not the (auto) tabs) and output capacity factors for these technologies should be in the output capacity factors tab. The data can be found manually in the Zenodo repository [20].

Off-grid capacity

Copy and paste the Cumulative/Additions Label for off-grid hydropower, and off-grid solar PV for the country from the **IRENA Installed Capacities data.xlsx** (from the Zenodo repository [20]) into the Off-Grid Capacity tab. Note: if the IRENA file has the country but not the relevant off-grid technologies, the values can be assumed to be zero and no user action is required.

- In the IRENA datasheet, ensure the 'Type' column (column AC) is filtered to '**off-grid**'
- Filter to the selected **country** in the 'IRENA Menu' column (column A)
- In the '**Sub-technology**' column, filter to Solar Photovoltaic
- Order by Years and then copy the 'Cumulative/Additions Label' column (column AL) into tab **3.7 Off-Grid Capacity**, cells C5–C24 – if there is no data for a year, then put 0, ensure the years match up with the years in column A

- Repeat the process but filter out for Renewable Hydropower as the sub-technology, copying the Cumulative/Additions Label column into cells B5–B24 in the tab 3.7 Off-Grid Capacity

Check that the residual capacity rows for PWRSOL002 and PWRHYD004 (rows 920 and 1085) are updated in the **3.7 ResCap data (auto) tab**. If not in the IRENA list, the **TEMBA_21_10_Refer** file can be used, go to the residual capacity tab:

- Copy the rows in the residual capacity tab for XXSOV1F01X and XXSOV2F01X from cell B to cell BE. If 0 in all years, no user action is required.
- Paste those rows into tab 3.7 ResCap data (auto) tab in rows 1065 and 1066 (PWRSOL002), starting in column I (to the right of the yellow cells).
- Ensure that the orange Residual Capacity row for PWRSOL002 (row 1085) is updated if adding values.
- Off-grid hydropower is excluded in this case since TEMBA does not include this technology.

On-grid residual capacity

This step is automated. Check that on-grid residual capacity data for the selected country is in the 3.7 ResCap data (auto) tab. Focus on checking between rows 724–1197 as this is where the power plant technologies are (all begin with PWR). Find the raw on-grid residual capacity data for the selected country in the file **residual capacity.csv** from [41] and check that there are no more than 11 power plants of each type for the selected country, as this is the maximum number permitted by the spreadsheet. If there are more than 11 rows, extra rows can be manually added and extra data copied directly from the CSV file.

A few manual adjustments are needed:

- For hydropower (rows 852–920), move any plants in the PWRHYD001 group to PWRHYD002 if they have a capacity between 0.01–0.1 GW and to PWRHYD003 if less than 0.01 GW (capacity is in column H). To move a plant, copy the entire row for that plant from column E to column AR, and paste it in the corresponding cells of the intended technology group.
- For oil power plants (rows 985–1016), all residual capacity is automatically inserted as PWROHC001 (LFO plant). However, you can do a quick Google search using the Power Plant name and move any plants to PWROHC002 (HFO gas turbine) using the method above if needed. If it is unclear or you cannot find anything, leave the plant where it is.
- The same should be done for gas power plants (rows 921–952), moving plants from PWRNGS001 (CCGT) to PWRNGS002 (OCGT) if you can find out that the plant is an OCGT. If it is unclear or you can't find anything, leave the plant where it is.

Useful sites for checking the type of power plant are [42] [43]; these can be searched by power plant name or by country and cover both oil and gas.

Refinery and Transmission & Distribution residual capacity

Copy and paste the residual capacity for the following technologies from the **residual capacity tab** of the TEMBA_21_10_Refer dataset [25] into the **3.7 ResCap data (auto) tab** as below. If values of 0 are seen for all years, no user action is required. If values are non-zero, copy the entire row in TEMBA from column B to BE:

- The row for XXCRUDRE1X in TEMBA (where XX is country code) should be copied to the Residual Capacity row for UPSREF001 in row 3181 in tab 3.7 ResCap data (auto), starting in column I
- The row for XXCRUDRE2X in TEMBA (where XX is country code) should be copied to the Residual Capacity row for UPSREF002 in row 3197 in tab 3.7 ResCap data (auto), starting in column I
- The row for XXEL00T00X in TEMBA (where XX is country code) should be copied to the Residual Capacity row for PWRTRN in row 1117 in tab 3.7 ResCap data (auto), starting in column I
- The row for XXEL00TDTX in TEMBA (where XX is country code) should be copied to the Residual Capacity row for PWRDIST in row 835 in tab 3.7 ResCap data (auto), starting in column I
- Note, the rows that are pasted into should be the ones shaded in orange.

Make a note of the PWR technologies (rows 724–1197 in tab 3.7 ResCap data (auto)) that do not have any residual capacity in the country, as this is needed for Step 7. Do not include the following technologies: PWRTRN, PWRDIST, PWRTRNIMP, PWRTRNEXP. Repeat these steps for the refinery technologies: UPSREF001 and UPSREF002 (rows 3166 to end) (XXCRUDRE1X & XXCRUDRE2X in TEMBA respectively).

Capacity Constraints

In the 3.8 Capacity & Inv Constraints tab, set the **Total Annual Max Capacity Investment** for 2015–2020 to 0 (column E to column J) for power generation and refinery technologies that have 0 residual capacity in the country, using the list made in Step 7. Do not do this for the following technologies: PWRTRN (row 439), PWRDIST (row 440), PWRTRNIMP (row 437), PWRTRNEXP (row 443).

If the country has no offshore wind potential, put 0 for the **Total Annual Max Capacity** for offshore wind all years (PWRWND002, row 37, 0 should be input from column E to BH). Highlight the rows where these constraints have been have added in green, as they will be pasted to SAND later.

Demands

This is partly automated. Check rows 7–11 in the 4. TEMBA Demands Data tab have been filled in with data (rows 7–11, columns C–BF). Make sure this process is done in the **4. TEMBA Demands Data tab** (<u>not</u> the 4. TEMBA Demands Data (auto) tab).

For IEA countries, insert the final consumption (in PJ; convert from TJ if needed) for each fuel in each sector in the country for 2015–2018 into the table (rows 15–10, columns B to E) in the 4. TEMBA Demands Data tab from the IEA Sankey Diagram [44], marking 0 if there is no consumption for that fuel/sector. Key points:

- For this step, ensure that the 'Final Consumption' option is selected rather than 'Balance' in the navigation pane on the left of the IEA website
- Ensure the unit is changed to PJ/TJ at the top of the diagram, and be sure to convert to PJ if it is in TJ (divide by 1,000)
- The consumption in each sector can be seen by clicking on the sector on the diagram, which opens a pie chart
- Change the year by dragging the slide along the bottom

For non-IEA countries:

- Find the UN Energy Balance for the selected country on the UN website [45] (there are PDFs for groups of countries in alphabetical order).
- For these countries, we will only insert data in the 2018 and 2017 columns of the table in the 4. TEMBA demands data tab (columns D and E, rows 15–30). The UN energy balances are in TJ; so divide by 1,000 to convert to PJ when inserting data. The top section of the UN energy balance is usually marked as 2018, then 2017 data are below – but check this for the selected country.
- Look at the data in the 'Final energy consumption' sections of the energy balance for 2017 and 2018. For industry, use the values for 'Manufacturing, const, mining'. For transport use the values for 'Transport'. For commerce use the values 'Commerce and public services'. For residential use the values for 'Households'. Use 'All Oil' for oil products; sum the values for 'Primary biofuels/Waste' and 'Charcoal' for biofuels and waste.
- Leave the columns for 2015 and 2016 (columns B and C) blank, so they will not be considered in the average calculated in column F.

Demands will then be automatically calculated: check in the 4.1 Accumulated Annual Demand tab that the rows for TRAMCY, TRACAR, TRABUS, INDHEH, INDHEL, RESCKN, COMHEL, and RESHEL have been filled in (shaded in green, between rows 19–29), and in the 4.2 Specified Annual Demand tab that the rows for INDELC, RESELC, and COMELC (rows 22, 25, & 27) have been filled in (shaded in green). These demands consider the input activity ratios of the technologies used to deliver them; for example, the demand for motorcycles is initially calculated in terms of demand for oil based on the input data, and this is then converted into the final energy demand for motorcycles considering the efficiency of the oil motorcycle technology.

Electricity demand profile

Copy and paste the hourly electricity demand profile for the selected country from the PLEXOS All Demand UTC 2015.tab dataset downloadable from [41] into the 4.2 Elc Demand Profile Raw Data tab.

- In PLEXOS the countries are along the tab, with the region code (AF for Africa), followed by the country code), copy the whole column starting from row 2 to row 8761.
- Paste the column into tab **4.2 Elc Demand Profile Raw Data** starting in cell B4 (marked in yellow).

Go to the **4.2 Specified Dem Profile Calc** tab and to the rows for RESELC, COMELC, and INDELC (rows 21, 24, and 26). Adjust the value in column L (Bennet Factor) until the value in column M is exactly equal to 1. Only small adjustments are needed: e.g., if the value in column M is 1.007, first try adjusting the value in column L to 0.98, then make further small adjustments if needed. Check that Specified Demand Profiles have been calculated for RESELC, COMELEC, and INDELC in the 4.2 Specified Demand Profile Output tab (columns W, Z, AB).

Import & Export activity limits

For IEA countries: Insert the amounts of imported and exported electricity (PJ) from the IEA Sankey diagram [44] for the country for 2015–2018 into the TotalTechnologyAnnualActivityUpperLimit rows for PWRTRNIMP (row 238) and PWRTRNEXP (row 244) in the tab 5.1 Activity in columns F (2015) to I (2018). Columns beyond column I are automatically calculated based on the values entered in columns F to I. Important points:

- For this Step ensure the **'Energy Balance'** option is selected in the left-hand navigation pane on the IEA Sankey website.
- Ensure that the unit is set to PJ/TJ as in Step 8, and make sure to carry out unit conversions if needed.
- Data can also be obtained from the IEA's energy balance tables .
- If there is no data, set to 0.

For non-IEA countries: open the UN energy balance for the selected country used in the earlier demands step. UN data are in TJ, which need to be divided by 1,000 to convert to PJ. Go to the tab 5.1 Activity. Insert the amounts of imported and exported electricity from the energy balance in 2017 and 2018 into the TotalTechnologyAnnualActivityUpperLimit rows for PWRTRNIMP (row 238) and PWRTRNEXP (row 244) in column H for 2017 and column I for 2018. Electricity imports and exports are found in the UN energy balance in the top section for each year in the rows for 'Imports' and 'Exports' under 'Electricity'. Columns beyond column I are automatically calculated based on the values entered in columns F to I. Important points:

- The values for 2017 must be inserted into the columns for 2015 and 2016 (columns F and G) - we will assume imports & exports remain similar across years.
- Do not include the minus sign found before the values for exports in the UN energy balance data.
- Remember that the UN data are in TJ and need to be divided by 1,000 to convert to PJ.

Renewable and fossil resources

Insert the estimated renewable energy potentials in the selected country into the table in the tab Data in Brief Tables 8 & 9 from the sources indicated in the table below. Some notes:

- If the table does not contain a potential for geothermal, insert a value of 0 for the geothermal potential (geothermal potentials are only included in the report for Eastern/Southern Africa and can be assumed to be 0 elsewhere).
- If there is a dash in the report, assume 0.
- When using the World Small Hydropower Development Report tables, select the value from the far-right column titled "Potential (<10MW)", or "Potential Capacity" for Middle Africa.
- When using the **IRENA SAPP and WAPP** reports for hydropower potential, the potential is sourced from the column titled 'Identified Projects (MW)' but also add the Existing Capacity (MW) as we want overall potential (instructions above tables in the links too). Ignore the small hydropower potentials in these reports, as these are sourced from the report above. When using the **IRENA Eastern & Southern Africa report**, hydropower potential is taken from the column titled 'Hydropower (MW), Potential'.

Table 2: Data sources for Renewable and fossil fuel resources for African countries

Country	Small Hydro	Hydro & Geothermal (where applicable)	PV, CSP, Wind
Angola, Botswana, DRC, Eswatini/Swaziland, Lesotho, Malawi, Mozambique Namibia, South Africa, Zambia, Zimbabwe	World Small Hydropower Development Report (Pages 9, 83, 119, 146, 175) [46]	IRENA Southern African Power Pool (SAPP) report (Table 2, page 21) [47]	Hermann et al. 2014 (Table 10 pages 35–36) [48]
Benin, Burkina Faso, Cote D'Ivoire, Gambia, Ghana, Guinea, Liberia, Mali, Niger, Nigeria, Senegal, Sierra Leone, Togo	World Small Hydropower Development Report (Pages 9, 83, 119, 146, 175) [46]	IRENA (West African Power Pool) WAPP report (Table 4, page 31) [28]	Hermann et al. 2014 (Table 10 pages 35–36) [48]
Burundi, Ethiopia, Kenya, Rwanda, Sudan, Uganda, Tanzania, Djibouti, Egypt, Libya	World Small Hydropower Development Report (Pages 9, 83, 119, 146, 175) [46]	IRENA Eastern & Southern Africa report (tables here)	Hermann et al. 2014 (Table 10 pages 35–36) [48]
Cameroon, Central African Republic, Chad, Republic of Congo, Equatorial Guinea, Eritrea, Gabon, Mauritania, Somalia, Guinea-Bissau	World Small Hydropower Development Report (Pages 9, 83, 119, 146, 175) [46]: note Chad & Eritrea are absent, assume 0.	TEMBA report (Table 12 on page 89) [49]	Hermann et al. 2014 (Table 10 pages 35–36) [48]

Insert the estimated fossil fuel reserves in the country into the table in the tab Data in Brief Tables 8 & 9 from Table 11 on page 88 of the TEMBA report [49]. If there is a dash, assume 0. If the country is not in the table, this means it is assumed there are no domestic reserves. In this case, insert 0 for coal, gas, and oil. Check that total technology model period activity upper limits have been added for MINOIL, MINNGS, and MINCOA in the 5.1 Activity tab (row 601, 605, 611), and that total annual max capacity limits have been updated for PWRGEO (row 24) and PWRHYD001-004 (rows 33–35) in the tab 3.8 Capacity & Inv Constraints if applicable.

2. Data Transfer From Data Collection And Manipulation Tool To Clicsand Software

At this point, following the instructions above, all the data have been collected in DaCoMaTool. The next step is to transfer all the data from the DaCoMaTool to the modelling tool selected for the analysis. Here we present the example of transferring this dataset to clicSAND software, which is a user-friendly interface for OSeMOSYS users. More information on the architecture, the specific functionalities and the installation process for this software can be found in [3]. A step-by-step course is freely available on the OLC platform of the Open University. Through theoretical classes and practical exercises, this course guides the user through the steps required to build a simple OSeMOSYS model using clicSAND software [22]. At the end of the course, the user receives a certificate of completion. clicSAND software is available also for MacOS users here and [50], and the respective open-source course can be found at this address [23].

In clicSAND software, the user adds the data in an Excel workbook – called SAND Interface – which contains all the OSeMOSYS parameters needed to create a model that the solver can process to find the optimal solution [51]. The following instructions describe how to transfer data into clicSAND software. Option 1 describes how to do this manually using copy and paste without using any programming language; Option 2 describes how to do this more efficiently using Python. The authors suggest Option 2 only if multiple Starter Data Kits are being created at the same time: for example, if the user is creating a Starter Data Kit for each country of a selected region. Using the Python code will accelerate the transfer and reduce the risk of human error while copy-pasting. On the other hand, if the user is working on a single country Starter Data Kit and has no previous knowledge of Python, Option 1 should be used.

Manual Transfer

The first data that need to be added are the Regional Specific Data. The user has two options:

a. Download a blank Excel SAND Interface from this online repository [52] and build the model from scratch, which will involve defining new technologies and sets, link them as per OSeMOSYS convention, and then add the data.
b. Download one of the pre-filled SAND files freely available on Zenodo for Africa [11], Asia [13], and South America [12]. These files already have all the Starter Data Kit technologies defined and all the regional data added.

The authors strongly advise choosing option (b) even if the Starter Data Kit that the user is building is not for one of the available regions. Using one of the pre-filled files will dramatically speed up the process as the entire model is already built, and the user will just need to update the region- and country-specific data rather than adding all the information from scratch.

If, however, the user is interested in creating a model which is not meant to be a Starter Data Kit, then option (a) would be more appropriate. For instructions on how to compile an OSeMOSYS model from scratch using SAND Interface (as part of the clicSAND software), please refer to the freely available courses available for Windows and Mac users respectively [22] [23].

Instead, if the user wants to create a Starter Data Kit using option (b) then the following steps should be performed to transfer the data. It is important to note that data should be copy-pasted from the Data Collection and the Manipulation Tool to the SAND Excel Interface using the "paste values" function. The authors suggest renaming the downloaded SAND Excel file as "New_your_country_name_Base_SAND" and then adding the regional and country-specific data. The description of each code used in the Tables below as well as the OSeMOSYS parameters can be found in the first tab called "Naming" in the SAND Interface Excel file. However, the authors strongly encourage users that want to create a Starter Data Kit to firstly complete the online course to acquire the basic skills needed.

Table 3: How to transfer Region-Specific data to SAND Interface

Data	Location of Data in Data Collection File	How to add/update data in SAND Interface for the following Parameter
Discount Rate (Default Value)	1. Model initiation	Discount Rate
Depreciation Methods (Default Value)	1. Model initiation	Depreciation Methods
Year Split Values (Default Values)	2. Year Split	Year Split
Technologies’ Costs	3.1 Technology Costs, rows 3–80 (capital costs), rows 202–279 (fixed costs), rows 401–478 (variable costs)	CapitalCost, FixedCost and VariableCost for all the technologies
Input Activity Ratios	3.2 Input Activity Ratios, rows 3–80 (excluding N.A.)	InputActivityRatio for all the technologies where the value is not N.A.
Output Activity Ratios	3.3. Output Activity Ratios, rows 3–80	OutputActivityRatio for all the technologies listed
Capacity To Activity Unit	3.4 Capacity to Activity Unit, rows 3–80	CapacityToActivityUnit for all the technologies listed
Operational Life	3.5. Operational Life, rows 3–80	OperationalLife for all the technologies listed
Emissions Activity Ratio	6.1. Emissions Activity Ratio, rows 4–15	EmissionsActivityRatio for all the technologies where the value is different from 0.
Capacity Factors for power plants with non-variable production	3.6 Output Capacity Factors, column S, U, W, X, Y, Z, AA, AD, AE, AK, and AP	CapacityFactor for each power plant with a non-variable production (PWRBIO001, PWRCOA001, PWRGEO, PWROHC001, PWROHC002, PWRNGS001, PWRNGS002, PWRCPS001, PWRCSP002, PWRNUC, PWROHC003). Leave default value of 1 for all the other technologies.

Table 4: How to transfer country-specific data to SAND Interface

Data	Location of Data in Data Collection File	Technologies/Fuels
Output Activity Ratios for Transmission and Distribution Technologies	3.3 Output Activity Ratios, rows 41 and 42, columns F to BI	PWRTRN, PWRDIST – in SAND make sure to add the data for PWRTRN on the row marked ELC002, and for PWRDIST on the row marked ELC003.
Capacity Factors for power plants with variable energy production	3.6 Output Capacity Factors, between columns AC to BX, row 3 to row 98. *These should be extended to 2070 in SAND*	PWRSOL001, PWRHYD001, PWRHYD002, PWRHYD003, PWRWND001, PWRWND002 (only if the country has offshore wind potential) PWRHYD004, PWRSOL001S, PWRWND001S
Residual Capacity for Transmission, Distribution, and Refineries Technologies	3.7 ResCap data (auto), between rows 724–1197 for PWR techs and rows 3181 for UPSREF001 and row 3197 for UPSREF002, columns I to AR. (see note above on adding residual capacity too).	PWRTRN, PWRDIST, and all PWR technologies that have residual capacity in the country, as well as UPSREF001/UPSREF002 if they have residual capacity in the country.
Total Annual Max Capacity for Hydropower, Geothermal, and Offshore wind power plants	3.8 Capacity & Inv Constraints, rows 33, 34, 35, (37 for offshore wind), 75, and 24, columns E to BH.	PWRHYD001, PWRHYD003, PWRHYD003, PWRHYD004, PWRGEO. PWRWND002 if it has been set to 0 for the country as offshore wind is not possible.
Total Annual Max Capacity Investment	3.8 Capacity & Inv Constraints, between rows 418–473 depending on the country, columns E to BH.	For 2015–2020 for any of the PWR or UPS technologies that have been set to 0 for because there was no residual capacity (these cells should be highlighted in green when prepping the data)
Accumulated Annual Demand	4.1 Accumulated Annual Demand. Rows 19, 20, 21, 23, 24, 26, 28, and 29, columns D to BG	Fuels: TRAMCY, TRACAR, TRABUS, INDHEH, INDHEL, RESCKN, COMHEL, RESHEL
Specified Annual Demand	4.2 Specified Annual Demand. Rows 22, 25, and 27, columns D to BG	INDELC, RESELC, COMELC
Specified Annual Demand Profile	4.2 Specified Dem Profile Output. Columns W, Z and AB, rows 4 to 99. *extend these values to 2070 in SAND*	INDELC, RESELC, COMELC
Total Technology Annual Activity Upper Limit	5.1 Activity, rows 238 and 244, columns F to BI.	PWRTRNIMP, PWRTRNEXP
Total Technology Annual Activity Lower Limit	5.1 Activity, rows 39 and 45, columns F to BI.	PWRTRNIMP, PWRTRNEXP
Total Technology Model Period Activity Upper Limit	5.1 Activity, rows 601, 605, 611, column E.	MINOIL, MINNGS, MINCOA

Automated Transfer with Python code

Copy the following relevant files to a local directory.

- The data preparation Python script ccg_data_to_sand.py, which includes a list of Python packages requirements [18].
- The "Base Data Collection File.xlsx" [1].
- The 'Starter Kit - List of Countries and Regions.xlsx' file from [19].

Prepare a virtual Python environment.

- Open the terminal.
- Navigate to the relevant directory.

cd /path/to/ccg_data_to_sand_directory

- Create a virtual environment (recommended).

python3 –m venv venv

source venv/bin/activate

Note: example command lines are given using python3, though users can also complete the same steps using python.

- Download the package dependencies.

python3 -m pip install -r requirements.txt

For further help (or if you don't already have pip), please check:
https://packaging.python.org/tutorials/installing-packages/

Update user inputs. In the ccg_data_to_sand.py Python file, the user should update:

- path_input: the path to the folder containing the input data files “Data Collection File.xlsx' and 'Starter Kit - List of Countries.xlsx'.
- path_SAND_out: the path to the folder where the "{country_name} Base SAND parameters.csv" file will be saved (default same as path_input).
- countries: list of countries for which SAND files should be created. Note this should be written in CAPITAL letters.

Run Python script: run the script twice, once to obtain the link of the adequate base SAND file to download as will printed on the screen, and a second time to perform the data preparation step.

- Open the terminal.
    - Activate the relevant virtual environment (if applicable)

cd /path/to/ccg_data_to_sand.py

source venv/bin/activate

- Run the script.

python3 ccg_data_to_sand.py

Copy values into parameter sheet

- Create a new empty base SAND Excel file by copying 'Base SAND.xlsm' [52] and renaming it to "New_{country_name}_Base_SAND.xlsm".
- Ensure all rows are visible and no filtering is active in country Base SAND.xlsm. Macros for the file may need to be enabled.
- Open the "{country_name} Base SAND parameters.csv" file that was generated by the Python script.
- Copy the contents from the entire CSV file.
- Paste values only into the Parameters tab of the "New_{country_name}_Base_SAND.xlsm" file.
  Note: This might only work in the desktop Excel version, not the web app.

3. Scenario Creation

At the end of Section 2, the user will have completed transferring the data from the Data Collection and Manipulation file to their "New_your_country_name_Base_SAND", which will include all the regional and country-specific data. On this file, no additional constraints are applied; therefore, if the user runs this model, the OSeMOSYS code will identify the optimal solution of the problem. This section will explain the steps needed to recreate the three scenarios available for each Starter Data Kit, namely the Fossil Future, Least Cost and Net Zero Scenario. For more information on the definition and the idea behind each of these scenarios, we refer the reader to the Data in Brief publication on the Starter Data Kits [2].

To speed up the process of applying constraints on the Base SAND file, we used a support Excel template called "Scenarios Constraints", freely available for download on Zenodo [53]. Instructions are available in the workbook. This workbook has four tabs:

- Input Data – add initial data, such as Specified Annual Demands and the grid and off-grid extension split.
- Fossil Future – add constraints using TotalAnnualMaxCapacityInvestment and TotalTechnologyAnnualActivityUpperLimit parameters.
- Least Cost – add constraints and run the model two times to constraint wind and solar
- Net Zero – add a constraint to nuclear, primary fossil fuel and biomass production technologies. The model needs to be run two times.

Hereafter, we will share the instructions on applying constraints to the Base SAND file to create the three scenarios. However, an intermediate run should be performed to complete a Least Cost and Net Zero Scenarios. Section 3.1. presents the manual way of performing this task while section 3.2. the automated alternative using Python codes. We refer the user to Section 4 for more instructions on running and visualising results using clicSAND software for OSeMOSYS.

Manual scenario creation

Fossil Future

- Create a copy of the 'Scenario Constraints' file [53] and name it 'Your_Country_Name_Scenario_Constraints'.
- Follow the instructions in the 'Input Data' tab of the 'Your_Country_Name_Scenario_Constraints' file to add a few bits of data for the country from the Data Collection and Manipulation File.
- The Data Collection and Manipulation file can now be closed.
- Go to the Fossil Future tab in the Country Scenario Constraints file and follow the instructions there to set up the Fossil Future scenario in a copy of the Base SAND file named "Your-Country_Name_FF_SAND (all instructions needed should be in this tab). Then download the Country FF SAND file to a local computer – we suggest creating a folder called 'Runs' on the Desktop and putting this (and everything else from now on) in here.

Least Cost Scenario

- Go back to the Country Scenario Constraints file and follow the instructions to set up the Least Cost scenario in the Least Cost tab.
- Once set up, download the Country LCv1 SAND file and run it for the FF scenario. Repeat the process above to set up the results database and results excel file for LCv1 – creating a copy of the results database and Results Template previously downloaded and naming them Country LCv1 Results Database/Results. Remember to change the file name of the results file produced by clicSAND to remove the spaces before importing to the database and select all the relevant technologies for each graph in the Results Excel file.
- Then go back to the Country Scenario Constraints file and follow the instructions to set up Country LCv2 before running it and creating the results files in the same way.

Net Zero Scenario

- Go back to the Country Scenario Constraints file and follow the instructions to set up the Net Zero scenario in the Net Zero tab.
- Once set up, download the Country NZv1 SAND file and run it as done for the FF and LC scenarios. Repeat the process above to set up the results database and results excel file for NZv1.
- Then go back to the Country Scenario Constraints file and follow the instructions to set up Country NZv2, before running it and creating the results files in the same way.

Automated Scenario Creation with Python Code

If the user is creating scenarios for multiple countries, the authors suggest automating scenario creation with the following process.

1. Copy the following data preparation scripts (found on this Zenodo repository: [21]) to a local directory:

    - CCG_SAND_Scenario_FF.py
    - CCG_SAND_Scenario_LCv1.py
    - CCG_SAND_Scenario_LCv2.py
    - CCG_SAND_Scenario_NZv1.py
    - CCG_SAND_Scenario_NZv2.py

2. Create a copy of the 'Scenario Constraints' file [53] and name it 'Your_Country_Name_Scenario_Constraints'. Save it to a local directory.

3. Update user inputs: in the CCG_SAND_Scenario_FF.py Python file, the user should update:

    - path_countries: path to base folder that contains other folders. Need not be used if folder structure is different.
    - path_SAND_out: path to folder with SAND CSV files
    - path_scenarios: path to folder with Scenario file
    - countries: list of countries for which SAND CSV files should be created

4. Run Python scripts.

    - Open the terminal.
    - Activate the relevant virtual environment created in the previous section to create the model (if applicable). This environment should already have the necessary packages installed.

cd /path/to/ccg_data_to_sand.py

source venv/bin/activate

- Type python3 cd /path/to/ CCG_SAND_Scenario_FF.py, then press ENTER.

5. Copy values into parameter sheet.

    - Create a new empty FF SAND.xlsm file by copying the country_base_SAND.xlsm and renaming country_FF_base_SAND.xlsm.
    - Open the new country_FF_SAND.xlsm file in Excel and ensure all rows are visible and no filtering is active.
    - Open the County FF SAND parameters.csv file in Excel.
    - Copy the contents of the entire CSV file.
    - Paste using the "values" option into the Parameters tab of country_FF_SAND.xlsm file. Note: This might only work in desktop Excel version, not web app.

6. Repeat Steps 3–5 for the Least Cost version 1 and 2 scenarios and Net Zero version 1 and 2 scenarios using the following python files instead of the CCG_SAND_Scenario_FF.py file:

    - CCG_SAND_Scenario_LCv1.py
    - CCG_SAND_Scenario_LCv2.py
    - CCG_SAND_Scenario_NZv1.py
    - CCG_SAND_Scenario_NZv2.py

4. Running Models And Visualising Results

clicSAND software was used to run the three scenarios and SAND Excel data files. clicSAND software includes two free powerful software, namely GLPK and CBC. The following resources can be used to learn how to run a model and visualise results using these software:

- The most updated version of the OLC free online course called "Energy and Flexibility Modelling" OLC for Windows and Mac Users.
- SoftwareX paper on clicSAND software functionalities, architecture, and documentation material.

5. Generation Of Articles, Figures, And Data Repositories

Manual Uploads

We created a Zenodo repository for each country which includes the following documents:

- Table1_<COUNTRY>.csv: Installed Power Plants Capacity in The Country

- Table2_<COUNTRY>.csv: Techno-economic parameters of power generation technologies
- Table3_<COUNTRY>.csv: Projected costs of renewable power generation technologies for selected years to 2050.
- Table4_<COUNTRY>.csv: Techno-economic parameters for transmission and distribution technologies
- Table5_<COUNTRY>.csv: Techno-economic parameters for refinery technologies
- Table6_<COUNTRY>.csv: Fuel price projections to 2050
- Table7_<COUNTRY>.csv: Fuel-specific CO2 Emission Factors
- Table8_<COUNTRY>.csv: Estimated Renewable Energy Potentials
- Table9_<COUNTRY>.csv: Estimated Fossil Fuel Reserves
- A $references.bib$ file which contains all the data sources
- A SAND file: "<COUNTRY> Base SAND.xlsm"
- A SAND file: "<COUNTRY> FF SAND.xlsm"
- A SAND file: "<COUNTRY> LCv2 SAND.xlsm"
- A SAND file: "<COUNTRY> NZv2 SAND.xlsm" – if this scenario was feasible for the selected country
- Data Collection and Manipulation File: "New <COUNTRY> Data Collection"
- Template to create scenarios: “Country_Name_Scenario_Constraints” File

It is important to note that Tables 1–9 in each repository are tool and interface agnostic, meaning that these data can be used with other tools. The user is therefore not locked into using OSeMOSYS or clicSAND software. Nevertheless, the SAND files have been added to the repository to speed up the uptake of the Starter Data Kits already created.

To extend an existing Starter Data Kit, it is recommend to download the SAND files available in the Country repository and start building on these. For example, suppose the renewable constraints that have been applied are not consistent with the local reality in the country or the most updated government plans. In that case, the Scenario Constraints support file can be used to adjust that constraint quickly; and this can then be added to a new version of the Scenario file, rerun, and the results compared.

This is just one of many adjustments that can be made to the Starter Data Kits. These Starter Data Kits are zero-order energy models that accelerate the energy modelling process by avoiding time-consuming activities such as model creation, design, data collection, and scenario development. In the next section, some suggestions for extending the Starter Data Kits will be presented as well as future developments for new kits.

The complete Zenodo repositories and preprint papers per country are available in the Annex.

Automated Uploads with Python code

There is an option for users who have created model results for several countries and want to share them on Zenodo via batch upload. This requires running a script to craft the batch upload. The scripts can be found in a Github repository [54], which contains its own instructions for use and can be forked in case adjustments to the files included or file paths are needed.

6. Limitations Of The Starter Data Kits And Future Developments

Potential extensions to the Starter Data Kits to improve their accuracy are detailed below. Works using the Starter Data Kits should always cite the Data in Brief paper, the MethodsX, and Zenodo repository for the selected country. Those wishing to conduct their own analyses using these data as a starting point should consult the country-specific SAND files available on Zenodo for the most accurate and final input assumptions for each country. Opportunities to build on the Starter Data Kit model through extensions could include modifying input data, the model structure, and the main assumptions taken. The authors suggest that a new piece of data could be added in the next iteration of the Starter Data Kits.

Modifying the Input Data

- Currently, the same demand profile is used across all sectors – adjust this for each sector to account for peaks occurring at different times in different energy sectors.
- Currently, we use generic data for the costs of cooking, transport, and heating technologies – it would add value if more country- or region-specific costs could be found.
- We only consider CO2 emissions – adding in emissions factors for methane or other GHGs would be an improvement.
- Residual capacity is based on a global power model (PLEXOS [55]) - these sources will certainly have gaps which could be reduced by comparing the data with government data or other in-country sources. Planned projects should also be added.

Modifying Model Structure

- It would be beneficial to split demands within sectors further and potentially add others such as agriculture.
- Demands are treated simply – it would be valuable to improve these by converting the transport demand to consider final transport demand in vehicle km. Regarding transport, a wider range of transport technologies could be added, and country-specific data on modal share could be sourced (currently, an assumption is made for each country based on regional data).
- There is a uniform approach to time-slicing with 4 seasons used, each split into day and night. Increasing the accuracy of time representation to allow for consideration of the seasonality in different countries would be an improvement. This would require adjustment of the capacity factor and specified demand profiles, and potentially the year split depending on the approach.
- Currently, only one region is used – perhaps for countries with significant geographic variability a multi-region approach could be employed to improve accuracy.
- Storage and flexibility are considered very simply. Specific storage technologies could be added to the model to improve this.
- Other technologies could be added: interconnector projects (existing and planned); LNG terminals.

Modifying Assumptions

- There are gradual investment constraints on demand-side technologies and renewable energy potentials in the scenarios. However, in-country knowledge could be used to limit yearly investment in power plant technologies and make capacity expansion more realistic. It would probably be equally essential to improve the consideration of the time taken to approve and construct each power plant technology, most likely using total annual max capacity investment in OSeMOSYS.
- Imports and exports are modelled very simply. It is assumed that the country can continue importing and exporting power at the same levels as in 2020, with no costs considered. To improve this, the option could be provided to scale imports and exports with demand and to add an electricity price that would allow the model to import and export as is economical. Interconnectors are also not considered.
- It would be valuable to improve the consideration of the country’s policies and plans. For example, the model currently allows investment in any technologies that are technically feasible in the country (although nuclear is turned off in all scenarios), but it may be that certain technologies are more or less likely to feature in the country due to policy goals, conflicts with land and water use, and so on.

Declarations

Acknowledgements

We would like to acknowledge data providers who helped make this and future iterations possible; they include IEA, UNSTATS, APEC, IRENA, UCC, KTH, UFOP and others.

CRediT Author Statement

Carla Cannone: Conceptualisation; Data curation; Investigation; Methodology; Formal analysis; Validation; Visualisation; Writing – Original Draft. **Lucy Allington**: Conceptualisation; Data curation; Investigation; Visualisation; Validation; Writing – Secondary Draft. **Karla Cervantes Barron:** Data curation; Validation; Writing – Secondary Draft. **Flora Charbonnier:** Data curation; Validation;

Writing – Secondary Draft. **Miriam Zachau Walker:** Data curation; Validation; Writing – Secondary Draft. **Claire Halloran:** Data curation; Validation; Writing – Secondary Draft. **Rudolf Yeganyan:** Writing – Review. **Naomi Tan:** Writing – Review. **Jonathan M Cullen:** Writing – Review. **John Harrison:** Writing – Review. **Long Seng To:** Writing – Review. **Mark Howells:** Writing – Review.

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships which have or could be perceived to have influenced the work reported in this article.

Funding

This paper has been written with the support of the Climate Compatible Growth Programme (#CCG) of the UK's Foreign, Commonwealth & Development Office (FCDO). The views expressed in this paper do not necessarily reflect the UK government's official policies.

References

1. “2050 Calculator | Faculty of Natural Sciences | Imperial College London.” https://www.imperial.ac.uk/2050-calculator (accessed Feb. 23, 2022).
2. B. Kondev, N. Nurmanbetova, A. Aidarbekova, S. Zhenishbekova, and J. Djakypbay kyzy, “Electric Vehicles (EVs) in Kyrgyzstan: Status quo, potentials and strategies for accelerated deployment,” 2018.
3. C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for Africa,” CCG Starter Kits Supporting Data and Scripts, 2022.
4. C. Cannone *et al.*, “Starter Kits - Base SAND file for South America,” CCG Starter Kits Supporting Data and Scripts, 2022.
5. C. Cannone *et al.*, “Starter Kits - Base SAND file for Asia,” CCG Starter Kits Supporting Data and Scripts, 2022.
6. C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for Africa Coal and Natural Gas Scenario,” CCG Starter Kits Supporting Data and Scripts, 2022.
7. C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for Africa Natural Gas Scenario,” CCG Starter Kits Supporting Data and Scripts, 2022.
8. C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for Asia - Coal and Natural Gas Scenario,” CCG Starter Kits Supporting Data and Scripts, 2022.
9. C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for South America - Coal and Natural Gas Scenario,” CCG Starter Kits Supporting Data and Scripts, 2022.
10. F. Charbonnier *et al.*, “CCG Starter Kits - Script for csv creation with base SAND data for Starter Kits,” CCG Starter Kits Supporting Data and Scripts, 2022.
11. C. Cannone *et al.*, “Starter Kit - List of Countries and Regions,” CCG Starter Kits Supporting Data and Scripts, 2022.
12. C. Cannone *et al.*, “Starter Kits - Technology-specific data for Base SAND file,” CCG Starter Kits Supporting Data and Scripts, 2022.
13. C. Halloran *et al.*, “Starter Kits - Scripts for SAND scenarios,” CCG Starter Kits Supporting Data and Scripts, 2022.
14. “OLCreate: PUB_5111_1.0 Energy and Flexibility Modelling: OSeMOSYS & FlexTool,” https://www.open.edu/openlearncreate/course/view.php?id=8105 (accessed Jan. 18, 2022).
15. “OLCreate: PUB_5398_1.0 (MacOS Users) Energy and Flexibility Modelling: OSeMOSYS & FlexTool,” https://www.open.edu/openlearncreate/course/view.php?id=8409 (accessed Feb. 03, 2022).
16. L. Allington *et al.*, “Selected ‘Starter kit’ energy system modelling data for selected countries in Africa, East Asia, and South America (#CCG, 2021),” Feb. 2022. doi: 10.21203/RS.3.RS-1178306/V1.
17. I. Pappis *et al.*, “Energy projections for African countries,” 2019. doi: 10.2760/678700.

18. I. Pappis, V. Sridharan, W. Usher, and M. Howells, “KTH-dESA/jrc_temba: TEMBA 2.0 (Version v2.0.3) [Data set],” 2021. doi: 10.5281/zenodo.4633042.
19. International Renewable Energy Agency, “Planning and Prospects for Renewable Power: Eastern and Southern Africa,” 2021, Accessed: Feb. 21, 2022. [Online]. Available: .
20. International Renewable Energy Agency, “Future of wind: Deployment, investment, technology, grid integration and socio-economic aspects (A Global Energy Transformation paper)” 2019, Accessed: Feb. 21, 2022. [Online]. Available: .
21. International Renewable Energy Agency, “Planning and prospects for renewable power: West Africa,” 2018, Accessed: Feb. 21, 2022. [Online]. Available: .
22. E. Terpilowski-Gill, “Decarbonising the Laotian energy system,” Imperial College London, 2020.
23. “Energy Efficiency Technologies Costs - Support File for Data Manipulation Starter Data Kits | Zenodo.” https://zenodo.org/record/6206363#.YhO1xujMKUk (accessed Feb. 21, 2022).
24. O. Okolo and H. Teng, “Analysing Nigeria’s Energy system in light of the UN’s Sustainable Development Goals,” 2017.
25. “Data | Electricity | 2021 | ATB | NREL.” https://atb.nrel.gov/electricity/2021/data (accessed Feb. 21, 2022).
26. “Storage Technologies Costs - Support File for Data Manipulation Starter Data Kits | Zenodo.” https://zenodo.org/record/6206365#.YhO1v-jMKUk (accessed Feb. 21, 2022).
27. U.S. EIA, “Assumptions to the Annual Energy Outlook 2020: International Energy Module,” 2020.
28. “Oil Prices - Support File for Data Manipulation Starter Data Kits | Zenodo.” https://zenodo.org/record/6206215#.YhO1yejMKUk (accessed Feb. 21, 2022).
29. “OLCreate: PUB_5111_1.0 Energy and Flexibility Modelling: OSeMOSYS & FlexTool.” https://www.open.edu/openlearncreate/course/view.php?id=8105 (accessed Feb. 16, 2022).
30. M. Brinkerink, B. Gallachóir, and P. Deane, “Building and Calibrating a Country-Level Detailed Global Electricity Model Based on Public Data,” Energy Strateg. Rev., vol. 33, p. 100592, Jan. 2021. doi: 10.1016/j.esr.2020.100592.
31. “Renewables.ninja.” https://www.renewables.ninja/ (accessed Feb. 21, 2022).
32. NREL, “Annual Technology Baseline 2020 Data,” 2020.
33. “Publications - IPCC-TFI.” https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol2.html (accessed Feb. 21, 2022).
34. M. Brinkerink and P. Deane, “PLEXOS-World 2015,” 2020.
35. “Global Energy Monitor.” https://www.gem.wiki/Main_Page (accessed Feb. 21, 2022).
36. “List of Gas PowerPlants - GEO.” http://globalenergyobservatory.org/list.php?db=PowerPlants&type=Gas (accessed Feb. 21, 2022).
37. International Energy Agency, “IEA Sankey Diagram,” 2019.
38. “UNSD — Energy Statistics.” https://unstats.un.org/unsd/energystats/pubs/balance/ (accessed Jun. 03, 2020).
39. United Nations, “World Small Hydropower Development Report 2019,” 2019.
40. IRENA, “Southern African Power Pool: Planning and Prospects for Renewable Energy,” 2013. https://www.irena.org/documentdownloads/publications/sapp.pdf (accessed Oct. 23, 2020).
41. S. Hermann, A. Miketa, and N. Fichaux, “Estimating the Renewable Energy Potential in Africa,” IRENA-KTH working paper, International Renewable Energy Agency, Abu Dhabi, 2014.
42. I. Pappis *et al.*, “Energy projections for African countries service contract 936531.”
43. C. Cannone *et al.*, “clicSAND for OSeMOSYS: a user-friendly interface using open-source optimisation software for energy system modelling analysis | Research Square,” 2022, Accessed: Feb. 16, 2022. [Online]. Available: https://www.researchsquare.com/article/rs-1338761/v1.
44. C. Cannone, N. Tan, A. Kell, de W. Nicki, M. Howells, and R. Yeganyan, “ClimateCompatibleGrowth/clicSAND for Mac,” Jan. 2022. doi: 10.5281/ZENODO.5925647.
45. C. Cannone, “Towards evidence-based policymaking: energy modelling tools for sustainable development,” Universitat Politècnica de Catalunya, Barcelona, 2020.

46. “Blank SAND Interface for clicSAND Software | Zenodo.” https://zenodo.org/record/6203284#.YhOM4OjMJPY (accessed Feb. 21, 2022).
47. “Scenario Constraints Excel Template for Starter Data Kits | Zenodo.” https://zenodo.org/record/6204654#.YhOcv-jMJPY (accessed Feb. 21, 2022).
48. K. Cervantes Barron, W. Usher, C. Cannone, and L. Allington, “CCG_Starter_Kits: Automatic Data in Brief article creation,” CCG_Starter_Kits: Automatic Data in Brief article creation, 2021.
49. M. Brinkerink, B. Gallachóir, and P. Deane, “Building and Calibrating a Country-Level Detailed Global Electricity Model Based on Public Data,” Energy Strateg. Rev., vol. 33, Jan. 2021. doi: 10.1016/J.ESR.2020.100592.

Figures

Figure 1

Description of the complete process to develop a Starter Data Kit for any country. The paper reads well as a step-by-step guide.

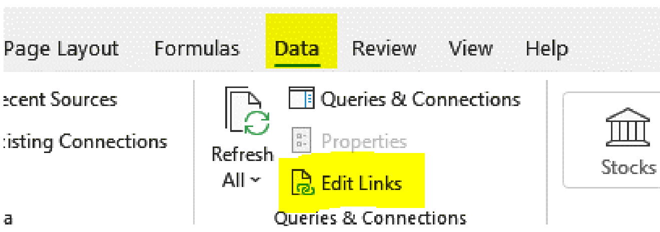

Figure 2

Updating links on Excel spreadsheet of the Base Data Collection File.

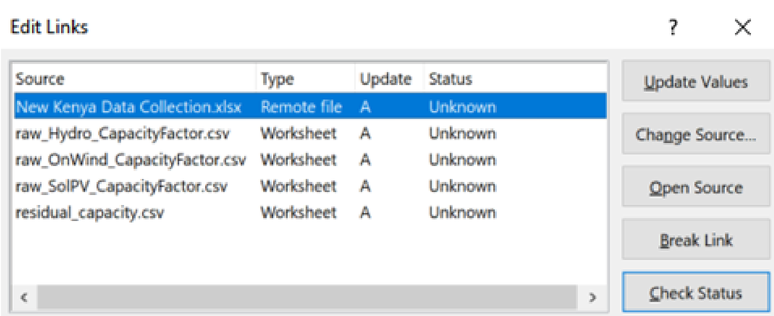

Figure 3

Changing the source of the different spreadsheets linked to the Base Data Collection File according to the user file paths.