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 8th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1403613/v2

Abstract

The Paris Agreement was signed by 192 Parties, who committed to reducing emissions. Reaching such commitments by 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. The 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 with the aim of reducing global warming associated with greenhouse gases emissions [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 that 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 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 essential 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 types. 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. Their aim is to promote 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, the 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 critical 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 and Manipulation Method for Starter Data Kits models
Name and reference of original method	Not applicable
Resource availability	Links to Zenodo Repositories in Annex A

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 (DaCoMaTool). This allows for data creation 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 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.
CCG 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.
CCG 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.
CCG 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 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 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.
CCG 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.
CCG	Several csv files	https://zenodo.org/record/6142375	[20]	These files

Starter Kits - Technology-specific Data for Base SAND File				contain the Capacity Factors and Residual Capacity values for all countries needed to fill in the base SAND file information.
CCG 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 specific 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 presented in Figure 1. 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 the region- and country-specific data. Once all data have been collected and put into a readable 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, in a manual or automated way, 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. Section 6 is an addition to this methodology which describes future work and potential improvements to the Starter Data Kits collection.

1. Data Collection And Manipulation Tool (Dacomatool)

The 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, it is presented, 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, by way of example, Africa's 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 B includes this information for the Asian and South 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 OSeMOSYS 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 calculations in this file [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 calculations in this file [35]. The price was increased by 10% for imported oil, and the prices of imported Heavy Fuel Oil (HFO) and Light Fuel Oil (LFO) were estimated by calcuSouthg 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 IRENA (2018) Planning & Prospects for Renewable Power: West Africa [28], and efficiencies of stoves were taken from Okolo O, Teng H. [31].

Capacity to Activity Unit, Operational Life & Capacity Factors

Default OSeMOSYS 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 the IPCC report [40].

Country-specific Data Collection and Manipulation

Box 1 presents an example of creating a Starter Data Kit for an African country. The box is organized into a series of instructions, or a "How-to Guide" to allow for the method to be applied by users to other country analyses.

[Box 1 is in the supplementary files section.]

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

After all the data have been collected in DaCoMaTool (based on the instructions of Box 1), the next step is to transfer all the data from the DaCoMaTool to the modelling tool selected for the analysis. Here, the example of transferring this dataset to clicSAND software is presented, 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 Cannone et al. [3]. A step-by-step course is freely available on the OLC platform of the Open University [23]. 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 [51] and [52].

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 [53]. The following instructions describe how to transfer data into clicSAND software. Option 2.1 describes how to do this manually using copy and paste without any programming language; Option 2.2 describes how to do this more efficiently using Python. The authors suggest Option 2.2 only if multiple Starter Data Kits are being created simultaneously: for example, if the user creates 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 2.1 should be used.

Manual Transfer

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

1. Download a blank Excel SAND Interface from this online repository [54] and build the model from scratch, involving defining new technologies and sets, linking them as per OSeMOSYS convention, and adding the data.
2. 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 user is building a Starter Data Kit for a country not in Africa, Asia, or South America. Using one of the pre-filled files will dramatically speed up the process as the entire model is already built. The user will need to update the region- and country-specific data rather than add all the scratch information.

However, if the user is interested in creating a model that is not meant to be a Starter Data Kit, 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, [22] and [23] respectively.

Instead, if the user wants to create a Starter Data Kit using option (b), the following steps should be performed to transfer the data, as summarized in Table 3 and Table 4. 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 (Table 3) and country-specific (Table 4) data. The description of each code used in the Tables below and the OSeMOSYS parameters can be found in the first tab called "Naming" in the SAND Interface Excel file. However, the authors strongly encourage users who want to create a Starter Data Kit to 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 the 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 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

In Box 2 below are presented the step-by-step instructions to transfer the data from the DaCoMaTool to clicSAND software in an automated way using a Python code.

[Box 2 is in the supplementary files section.]

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: the Fossil Future, Least Cost, and Net Zero Scenario. For more information on the definition and the idea behind each of these scenarios, the reader should refer 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, a support Excel template called "Scenarios Constraints" [55] was used, freely available for download on Zenodo. 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, the instructions on applying constraints to the Base SAND file will be shared to create the three scenarios. However, an intermediate run should be performed to complete a Least Cost and Net Zero Scenarios. Box 3 presents the manual way of performing this task while Box 4 the automated alternative using Python codes. The user should consult Section 4 for more instructions on running and visualising results using clicSAND software for OSeMOSYS.

[Box 3 and 4 are in the supplementary files section.]

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 this software:

- The most updated OLC free online course version is called "Energy and Flexibility Modelling" OLC for Windows and Mac Users [22] [56].
- SoftwareX paper on clicSAND software functionalities, architecture, and documentation material [51].

5. Generation Of Articles, Figures, And Data Repositories

Manual Uploads

A Zenodo repository was created 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 CO_2 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 recommended 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 and future developments for new kits.

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

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 [57], which contains its own instructions for use and can be forked (i.e., copied to a Github user’s repository for customization) 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 and extend their data coverage 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 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. The authors suggest that new data types could be added to the methodology in future iterations of the Starter Data Kits.

6.1. Modifying the Input Data

Currently, several model assumptions use generic values. Thus, there is potential for such assumptions to be improved by sector, country, or region. Details of such improvements are the following:

- The same demand profile is used across all sectors, so this could be adjusted for each sector to account for peaks occurring at different times in different energy sectors.
- Generic costs are used for cooking, transport, and heating technologies, so finding more country- or region-specific costs could add value.
- Only CO_2 emissions were considered, so adding in emissions factors for methane, or other GHGs would be an improvement.
- Residual capacity is based on a global power model (PLEXOS [58]). These sources have gaps that could be reduced by comparing the data with government data or detailed country sources. Planned projects should also be added.

6.2. Modifying Model Structure and Assumptions

The model scope is growing regarding countries, regions and supporting technologies (e.g., storage). Thus, the assumptions in the model and the structure of the data and calculations need to include this growing complexity. Some areas where change would be beneficial, and their proposed changes are the following:

- Split demands within sectors further and potentially add others such as agriculture.
- Demands are treated in an introductory manner – it would be valuable to improve these by converting the transport demand to consider final transport demand in vehicle/km unit. A more comprehensive 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 four seasons, each split into day and night. Increasing the accuracy of time representation to consider the seasonality in different countries would be an improvement. This would require adjusting the capacity factor and specified demand profiles, and potentially the year split depending on the approach.
- Currently, only one region is used, so perhaps for countries with significant geographic variability, a multi-region approach could improve accuracy.
- Storage and flexibility are considered in a simple manner. Thus, specific storage technologies could be added to the model.
- Other technologies could be added, e.g., interconnector projects (existing and planned) or LNG terminals.
- There are gradual investment constraints on demand-side technologies and renewable energy potentials in the scenarios. However, in-country knowledge could limit yearly investment in power plant technologies and make capacity expansion more realistic. It would be equally essential to improve the consideration of the time taken to approve and construct each power plant technology, most likely using total annual maximum capacity investment in OSeMOSYS.
- Imports and exports are modelled in a simple manner. 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 country's policies and plans. For example, the model currently allows investment in any technically feasible technologies in the country (although nuclear is turned off in all scenarios). However, 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.

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; Visualisation; 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] "The Paris Agreement | United Nations." https://www.un.org/en/climatechange/paris-agreement (accessed Feb. 26, 2022).

[2] L. Allington, "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.

[3] 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.

[4] M. Howells *et al.*, "OSeMOSYS: The Open Source Energy Modeling System. An introduction to its ethos, structure and development.," *Energy Policy*, vol. 39, no. 10, pp. 5850–5870, 2011, doi: 10.1016/j.enpol.2011.06.033.

[5] D. Timmons *et al.*, "Cost minimization for fully renewable electricity systems: A Mauritius case study," *Energy Policy*, vol. 133, p. 110895, Oct. 2019, doi: 10.1016/J.ENPOL.2019.110895.

[6] L. Ayagapin, J. P. Praene, D. Jaggeshar, and D. Surroop, "Prospective Life Cycle Assessment: Effect of Electricity Decarbonization in Building Sector," *Energies 2021, Vol. 14, Page 3184*, vol. 14, no. 11, p. 3184, May 2021, doi: 10.3390/EN14113184.

[7] D. Surroop, P. Raghoo, and Z. M. A. Bundhoo, "Comparison of energy systems in Small Island Developing States," *Util. Policy*, vol. 54, pp. 46–54, Oct. 2018, doi: 10.1016/J.JUP.2018.07.006.

[8] D. Surroop and P. Raghoo, "Renewable energy to improve energy situation in African island states," *Renew. Sustain. Energy Rev.*, vol. 88, pp. 176–183, May 2018, doi: 10.1016/J.RSER.2018.02.024.

[9] "2050 Calculator | Faculty of Natural Sciences | Imperial College London." https://www.imperial.ac.uk/2050-calculator (accessed Feb. 23, 2022).

[10] 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.

[11] C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for Africa,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[12] C. Cannone *et al.*, “Starter Kits - Base SAND file for South America,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[13] C. Cannone *et al.*, “Starter Kits - Base SAND file for Asia,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[14] 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. .

[15] C. Cannone *et al.*, “CCG Starter Kits - Base SAND file for Africa Natural Gas Scenario,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[16] 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. .

[17] 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. .

[18] 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. .

[19] C. Cannone *et al.*, “Starter Kit - List of Countries and Regions,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[20] C. Cannone *et al.*, “Starter Kits - Technology-specific data for Base SAND file,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[21] C. Halloran *et al.*, “Starter Kits - Scripts for SAND scenarios,” *CCG Starter Kits Supporting Data and Scripts*, 2022. .

[22] “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).

[23] “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).

[24] I. Pappis *et al.*, “Energy projections for African countries ,” 2019. doi: 10.2760/678700.

[25] I. Pappis, V. Sridharan, W. Usher, and M. Howells, “KTH-dESA/jrc_temba: TEMBA 2.0 (Version v2.0.3) [Data set],” 2021. doi: http://doi.org/10.5281/zenodo.4633042.

[26] I. Renewable Energy Agency, “Planning and Prospects for Renewable Power: Eastern and Southern Africa,” 2021, Accessed: Feb. 21, 2022. [Online]. Available: www.irena.org.

[27] I. – International Renewable Energy Agency, “FUTURE OF WIND Deployment, investment, technology, grid integration and socio-economic aspects A Global Energy Transformation paper Citation About IRENA,” 2019, Accessed: Feb. 21, 2022. [Online]. Available: www.irena.org/publications.

[28] I. Renewable Energy Agency, “Planning and prospects for renewable power: WEST AFRICA 2018,” 2018, Accessed: Feb. 21, 2022. [Online]. Available: www.irena.org.

[29] E. Terpilowski-Gill, “Decarbonising the Laotian energy system Imperial College London, 2020.,” Imperial College London, 2020.

[30] “Energy Efficiency Technologies Costs - Support File for Data Manipulation Starter Data Kits | Zenodo.” https://zenodo.org/record/6206363#.YhO1xujMKUk (accessed Feb. 21, 2022).

[31] O. Okolo and H. Teng, “Analysing Nigeria’s Energy system in light of the UN’s Sustainable Development Goals,” 2017.

[32] “Data | Electricity | 2021 | ATB | NREL.” https://atb.nrel.gov/electricity/2021/data (accessed Feb. 21, 2022).

[33] “Storage Technologies Costs - Support File for Data Manipulation Starter Data Kits | Zenodo.” https://zenodo.org/record/6206365#.YhO1v-jMKUk (accessed Feb. 21, 2022).

[34] U.S. EIA, “Assumptions to the Annual Energy Outlook 2020: International Energy Module,” 2020.

[35] “Oil Prices - Support File for Data Manipulation Starter Data Kits | Zenodo.” https://zenodo.org/record/6206215#.YhO1yejMKUk (accessed Feb. 21, 2022).

[36] “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).

[37] 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.

[38] “Renewables.ninja.” https://www.renewables.ninja/ (accessed Feb. 21, 2022).

[39] NREL, “Annual Technology Baseline 2020 Data,” 2020.

[40] “Publications - IPCC-TFI.” https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol2.html (accessed Feb. 21, 2022).

[41] M. Brinkerink and P. Deane, “PLEXOS-World 2015,” 2020.

[42] “Global Energy Monitor.” https://www.gem.wiki/Main_Page (accessed Feb. 21, 2022).

[43] “List of Gas PowerPlants - GEO.” http://globalenergyobservatory.org/list.php?db=PowerPlants&type=Gas (accessed Feb. 21, 2022).

[44] International Energy Agency, “IEA Sankey Diagram,” 2019. .

[45] “UNSD — Energy Statistics.” https://unstats.un.org/unsd/energystats/pubs/balance/ (accessed Jun. 03, 2020).

[46] United Nations, “World Small Hydropower Development Report 2019,” 2019.

[47] IRENA, “Southern African Power Pool: Planning and Prospects for Renewable Energy,” 2013. https://www.irena.org/documentdownloads/publications/sapp.pdf (accessed Oct. 23, 2020).

[48] S. Hermann, A. Miketa, and N. Fichaux, “Estimating the Renewable Energy Potential in Africa,” Abu Dhabi, 2014.

[49] IRENA, “Analysis of Infrastructure for Renewable Power in Eastern and Southern Africa,” Abu Dhabi, 2015.

[50] I. Pappis *et al.*, “Energy projections for African countries service contract 936531.”

[51] C. Cannone *et al.*, “clicSANDMac for OSeMOSYS: a user-friendly interface for macOS users using open-source optimisation software for energy system planning,” Feb. 2022, doi: 10.21203/RS.3.RS-1384770/V1.

[52] 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.

[53] C. Cannone, “Towards evidence-based policymaking: energy modelling tools for sustainable development,” UPC Barcelona, 2020.

[54] “Blank SAND Interface for clicSAND Software | Zenodo.” https://zenodo.org/record/6203284#.YhOM4OjMJPY (accessed Feb. 21, 2022).

[55] “Scenario Constraints Excel Template for Starter Data Kits | Zenodo.” https://zenodo.org/record/6204654#.YhOcv-jMJPY (accessed Feb. 21, 2022).

[56] “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. 16, 2022).

[57] 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. .

[58] 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.

[59] “A Renewable Energy Roadmap RENEWABLE ENERGY OUTLOOK FOR ASEAN A Renewable Energy Roadmap A Renewable Energy Roadmap,” Accessed: Feb. 25, 2022. [Online]. Available: www.irena.org/publications,.

[60] L. Allington, “Asia Data Support File,” Feb. 2022, doi: 10.5281/ZENODO.6311949.

[61] IRENA and ASEAN Centre for Energy, “Renewable Energy Outlook for ASEAN,” Abu Dhabi, 2016.

[62] IRENA, “Renewable Power Generation Costs in 2019,” Abu Dhabi, 2020.

[63] Y. Li and Y. Chang, “Infrastructure Investments for Power Trade and Transmission in ASEAN+2: Costs, Benefits, Long-Term Contracts, and Prioritised Development,” 2014. Accessed: Feb. 28, 2022. [Online]. Available: http://www.adb.org/features/fast-facts-asean-infrastructure-fund.

[64] “Biogas for Domestic Cooking: Technology brief.” https://www.irena.org/publications/2017/Dec/Biogas-for-domestic-cooking-Technology-brief (accessed Feb. 28, 2022).

[65] Asia-Pacific Economic Cooperation, “APEC Energy Demand and Supply Outlook 7th Edition,” 2019.

[66] Argus, “Argus Biomass Markets Weekly Biomass Market News and Analysis Issue 20-47,” 2020.

[67] G. N. P. de Moura, L. F. L. Legey, and M. Howells, “A Brazilian perspective of power systems integration using OSeMOSYS SAMBA – South America Model Base – and the bargaining power of neighbouring countries: A cooperative games approach,” *Energy Policy*, vol. 115, pp. 470–485, Apr. 2018, doi: 10.1016/j.enpol.2018.01.045.

[68] “Africa Clean Energy Corridor.” https://irena.org/cleanenergycorridors/Africa-Clean-Energy-Corridor (accessed Feb. 28, 2022).

[69] P. Howes, J. Bates, A. Brown, R. Diaz-Chavez, S. Christie, and A. Bayley, “Global Biomass Markets Final Report,” 2018.

[70] “Countries ranked by Electric power transmission and distribution losses (% of output) - Asia.” https://www.indexmundi.com/facts/indicators/EG.ELC.LOSS.ZS/rankings/asia (accessed Feb. 28, 2022).

[71] McKinsey, “McKinsey Refinery Reference Desk,” 2020. .

[72] “ALL CLEAN PVT LT | Clean Cooking Alliance.” https://cleancooking.org/sector-directory/all-clean-pvt-lt/ (accessed Feb. 28, 2022).

[73] OLADE, “Energy Outlook of Latin America and the Caribbean 2019,” 2019.

[74] “Open Energy Data Initiative (OEDI).” https://data.openei.org/ (accessed Feb. 28, 2022).

Boxes

Boxes 1-4 are in the supplementary files section.

Figures

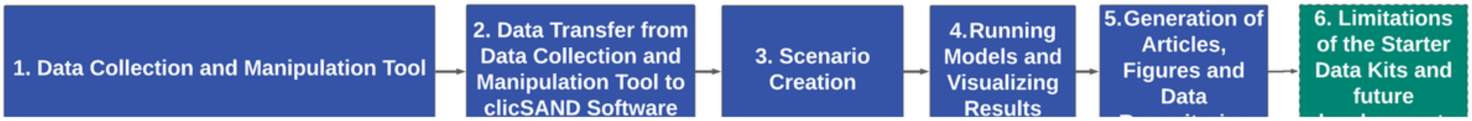

Figure 1

Description of the complete process to develop a Starter Data Kit for any country. The **Boxes** include detailed instructions on a particular action that needs to be performed. The paper reads well as a step-by-step guide.

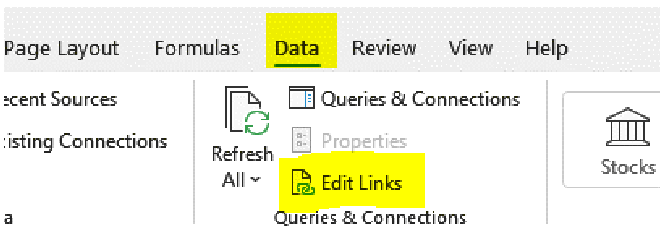

Figure 2

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

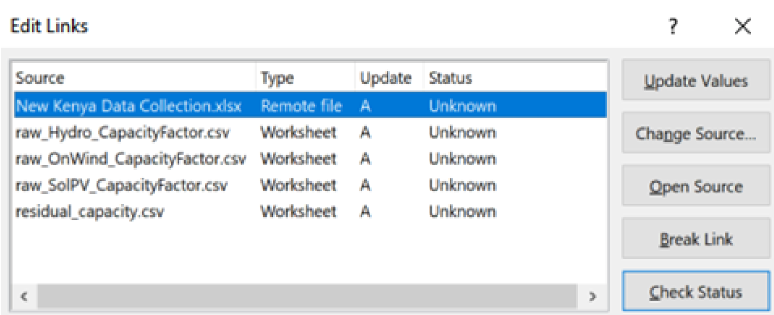

Figure 3

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

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FinalMethodsXZenodoSupplementaryFiles.docx
- Boxes.docx