Advancing Methodological Consistency in CLEWs Modelling: Introducing the Model Development Infrastructure

Authors: Georgios Avgerinopoulos[1], Annie Flint Smith[1], Kavya Abeysundara[1], Mark Howells[1]

[1]Loughborough University

Abstract

Climate change mitigation, resource security, and economic development are central priorities in global policymaking. Addressing these interconnected challenges requires integrated modelling approaches that capture the interactions between climate, land, energy, and water (CLEWs) systems. While the CLEWs framework provides a structured basis for such analysis, building models from scratch remains time-consuming and error-prone. Recent standardisation initiatives have improved efficiency but remain limited in scope and flexibility.

This paper introduces the Model Development Infrastructure (MDI), a modular, Excel-based toolkit that streamlines the creation of CLEWs++ models – an extended version of the CLEWs framework that broadens sectoral coverage to capture nearly all greenhouse gas emissions. The MDI combines pre-loaded global datasets with user-supplied local information, enabling model development for any country while allowing customisation of scope and sectoral detail. Its components guide users through a structured workflow, from base-year calibration to model-ready input data optimised for the Open Source Energy Modelling System (OSeMOSYS). Key features include wide sectoral coverage, adjustable model boundaries, accessibility for non-programmers, and a modular architecture that supports maintenance and future enhancements.

Although there is space for future improvements in scalability and data integration, the MDI bridges methodological gaps in previous standardisation efforts, lowers entry barriers, improves reproducibility, and supports more comprehensive CLEWs-based policy analysis.

Keywords

Model standardisation, CLEWs, CLEWs++, MUIO, Model Development Infrastructure (MDI), OSeMOSYS, energy modelling

1. Introduction

Addressing climate change mitigation, ensuring resource security, and fostering economic development have been some of the most prominent and pressing challenges for global policymakers in recent years. Resource security encompasses not only water, food and energy systems, but also critical minerals, which has garnered increasing political attention due to geopolitical tensions over resource competition, supply chain vulnerabilities, and matters of national security (Dhir, Buisson & Kim, 2025). With the simultaneous imperative to accelerate economic growth on a global scale, these interconnected priorities require coordinated and integrated policy approaches that address their interdependencies and potential trade-offs.

The Climate, Land, Energy and Water systems (CLEWs) framework provides a structured and nexus-oriented analytical approach to managing the interactions between the resource systems underlying these policy challenges. By holistically considering the interconnections between climate, land use, energy production, and water resources, the CLEWs framework facilitates a methodological analysis of the synergies and trade-offs across different sectors. In the last fifteen years, the CLEWs framework has gained academic traction as a critical and effective tool for informing resource system management and planning (Alexander et al., 2025).

Despite advances in the conceptual understanding of applying a nexus-thinking approach to resource system management, analysing multiple system interactions is a complex task that typically relies on sophisticated mathematical models. While various modelling approaches can be employed, one common challenge persists: building a model from scratch is a time-consuming and resource-intensive endeavour. Although each model is inherently unique, certain methodological steps are transferable across applications. Repeating the process from the ground up each time increases the risk of error, leads to inefficiencies, and hinders wider and quicker adoption of CLEWs modelling approaches. Therefore, developing standardised methodologies and modelling blueprints is of paramount importance to improve the efficiency and widespread adoption of integrated resource system models.

This paper presents a new initiative aimed at addressing these shortcomings: the Model Development Infrastructure (MDI). The MDI is a structured set of tools designed for organising input data into clear, manageable components ready for model input. Its structure promotes consistency and transparency across the development of different CLEWs models, reducing the complexity of the data collection process whilst still maintaining the flexibility to build models that are tailored to specific contexts.

2. Literature Review

Over the past decade, the CLEWs framework has become increasingly prominent in academic research as a methodological approach to assessing resource system interdependencies and informing cross-sectoral policy planning. Crucially, the CLEWs framework applies an interdisciplinary, nexus thinking approach to quantitative modelling, considering how climate, land, energy and water systems are inherently interconnected (Martindale et al., 2023).

In many prior resource system studies, the scope has been restricted to a single sector or a limited set of sub-sectors, most often with an energy-centric focus. While these studies have yielded valuable insights, a central rationale for adopting a nexus thinking approach instead is that sector- or resource-specific methodologies, when applied in isolation, risk overlooking critical system interlinkages, synergies and trade-offs (Albrecht et al., 2018). Numerous studies have demonstrated the value of applying the CLEWs framework in practice, with it increasingly being utilised to inform national cross-sectoral policy recommendations (Alexander et al., 2025). For example, the CLEWs-based modelling analyses carried out by Kuling et al. (2022) and Weirich et al. (2013) each illustrated that analysing a resource, like energy or water, in isolation can obscure the feedback loops and trade-offs that emerge when that resource is considered in the broader integrated resource systems context. Alexander et al.'s (2025) state-of-the-art literature review of 41 peer-reviewed CLEWs studies similarly found that there are critical connections between energy and water systems that have been captured through CLEWs modelling.

Despite growing recognition of its value across diverse contexts, several persistent methodological challenges continue to limit the widespread application of the CLEWs framework. Alexander et al.'s (2025) analysis underscored how methodological inconsistencies, sectoral siloes and a lack of institutional coordination act as significant barriers to the broader uptake of CLEWs modelling. While the CLEWs framework is not tied to one specific modelling tool, the lack of a standardised method for building a CLEWs model creates inefficiencies and inconsistencies across academic research. Saedi (2016) notes that the absence of a standard workflow specifically for combining and processing data places an unnecessary burden on the modeller.

Several initiatives have attempted to streamline the model development process. Starter Data Kits (SDKs), developed for several countries, provide users with a prepared dataset that can serve as the basis for a country-specific model (Howells et al., 2023). These kits are model-independent and supplied in an easily accessible Excel format that facilitates adaptation to a range of modelling tools. Cannone et al. (2023) have also detailed clear instructions on building an SDK from scratch, allowing users to replicate the process for other countries. However, the current SDKs that have been designed are limited to the energy, focused particularly on the power sector, meaning they cannot offer the full sectoral breadth necessary for a comprehensive analysis. Given that one of the original purposes of the SDKs' creation was to evaluate greenhouse gas emissions in energy systems (Howells et al., 2023), the scope of the SDK data inherently increases the risk of neglecting cross-sectoral synergies and trade-offs in their analysis.

OSeMOSYS Global is another open-source initiative, providing an open-data model generator for creating global electricity system models (Barnes et al., 2022). One of its key strengths lies in its high flexibility for the user. Users can select the geographic region they wish to model and can adjust model specifications such as temporal resolution, including the number and distribution of time slices. This flexibility contrasts with the scope of the SDKs, which are more prescriptive in structure.

However, whilst OSeMOSYS Global provides users with the abilities to adjust specific parameter data inputs, it is designed for the power sector and therefore does not capture all sources of greenhouse gas emissions, nor does it account for the interdependencies between different resource systems. This sectoral limitation means it is best suited for electricity-focused analyses, rather than broader CLEWs assessments that require cross-sectoral datasets and modelling capacity. Furthermore, these models are often constrained by fixed sectoral structures or predefined sub-sectoral disaggregation that may not align with the specific needs of a given application. This can limit the relevance and usefulness of the resulting analysis for local decision-making, as the model structure may not reflect the most pressing issues or data realities of each context.

Other tools, such as CLEWs Global, aim to provide a ready-made foundation for building CLEWs models, particularly in the OSeMOSYS environment (Kuling et al., 2023). CLEWs Global integrates data from sources such as OSeMOSYS Global, the Global Agro-Ecological Zoning Database (GAEZ) (Saedi and Niet, 2024), CLEWSy, and Python Snakemake workflows, creating a multi-resource starting point. Nevertheless, this methodological approach assumes a relatively high level of technical competence from the research undertaker, particularly in Python programming. This high skill threshold can act as a barrier for researchers, policymakers, or planners who may have deep sectoral expertise but lack advanced coding abilities.

The existing academic literature reveals clear gaps in the current state of CLEWs model data collection and development. These include an absence of standardised workflows, a tendency towards sector-specific modelling that neglects interconnections, steep technical skill requirements, and a lack of flexibility in model structures. These limitations suggest a pressing need for tools and processes that can support the efficient, transparent and flexible development of cross-sectoral CLEWs models, enabling wider uptake and more consistent application of nexus approaches in diverse contexts. To respond to these challenges, this paper introduces the MDI as a tool for developing more consistent, flexible and accessible CLEWs models.

3. Model Development Infrastructure (MDI)

3.1 MDI Overview

The MDI generates and organises input data intended for use across a range of modelling software tools. While its outputs can be adapted for use across multiple platforms and tools, the data structure is optimised for integration with the Model User Interface for OSeMOSYS (MUIO). The main features of the MDI, outlined below, are specifically designed to overcome several of the limitations observed in previous CLEWs modelling efforts:

- **Broader scope**: Earlier initiatives have primarily focused on the power sector or, at best, selected areas within the energy system. In contrast, the MDI enables the development of comprehensive CLEWs models and also incorporates waste and industrial processes; elements not previously included in standard CLEWs models. This extended coverage of nearly 100% of greenhouse gas (GHG) emissions, is referred to as CLEWs++.

- **Adjustable scope**: While the MDI allows for the construction of a detailed CLEWs++ model across multiple sectors, it also supports flexibility in model scope. Users may exclude certain sectors (e.g. such as waste) if not relevant to their analysis. Similarly, sectoral detail can be customised. For instance, the industrial sector can be disaggregated into sub-sectors; if a user opts not to model aluminium production explicitly, it can be grouped under a general 'other industry' category.

- **Ease of use**: The MDI is based on a series of Excel files and requires no coding skills. Users populate these files in sequence, following clear, step-by-step instructions. This design minimises the likelihood of errors and lowers the barrier to entry for non-programmers.

- **Modular and maintainable**: Instead of relying on a single master file, the MDI consists of multiple interlinked files. This modularity ensures that updates or improvements to one component do not necessarily disrupt the entire workflow, enhancing both usability and long-term maintainability.

3.2 Technical Description

The MDI is distributed as a single folder containing all the files required to develop a CLEWs++ model. The folder is publicly available and can be accessed on Zenodo (Avgerinopoulos et al., 2025). To begin, the user should download and unzip the folder, then consult the included instructions.

The MDI contains a certain amount of pre-loaded data, enabling the user to construct a model for any country with full CLEWs++ scope. Users are, however, encouraged to supplement this with additional, locally sourced data to improve model accuracy and robustness. The MDI represents an ongoing effort, with continuous improvements and updates aimed at enhancing its scope, functionality, and usability.

The MDI is organised into a set of files and tools, each performing a specific role in the model development workflow. Used in sequence, these components guide the user from initial data collection through to the preparation of fully structured model input data ready for implementation in MUIO.

Table 1 summarises each component, its format, and its primary function. Detailed descriptions are provided in the following subsections.

Table 1. Summary of the components of the Model Development Infrastructure (MDI), including file format and primary function.

Component	Format	Purpose
00 – Instructions	Word	Explains the purpose of the MDI, how to use it, and troubleshooting advice.
00 – Technology Database	Excel	Stores techno-economic data for model technologies; can be used beyond the MDI.
01_Tool – Calibration	Excel	Uses historical data to establish base-year conditions and activity levels by technology.

02_Tool – Technoeconomics	Excel	Converts Technology Database inputs into modelling-ready parameters, including interpolation.
03_Tool – Demand Projections	Excel	Projects future demand based on base-year values and growth rates.
04_Tool – Residual Capacity Power	Excel	Builds vintage capacity profiles for power and heat plants.
05_Tool – Year Split and Demand Profiles	Excel	Defines time-slice shares and demand distribution profiles.
06_Tool – VRE Capacity Factors	Excel	Calculates wind and solar capacity factors for each time slice.
07_Tool – Constraints	Excel	Allows the user to apply predefined or custom modelling constraints.
08_Tool – Residual Capacity Others	Excel	Calculates vintage capacity profiles for other capital-cost technologies.
09_File – Data Structuring for MUIO	Excel	Organises outputs from all tools into MUIO-ready format.
Profiles	Folder	Contains country-specific demand profiles and renewable capacity factors from Renewables.ninja.

To provide further clarity on the capabilities and intended applications of each component, the subsequent sections present an expanded discussion of the tools summarised in Table 1.

00 – Instructions (Word document)

This is the first file the user should consult. It outlines the purpose of the MDI, provides guidance on its intended use, and offers troubleshooting advice.

00 – Technology Database (Excel file)

This file is intentionally not referred to as a *tool*, unlike the subsequent Excel files, as it is intended for use beyond the MDI framework. As its name suggests, the Technology Database stores all techno-economic data required for the technologies included in a model. Collecting this data is a critical step in the modelling process and must be carried out by the user. Once gathered, the data should be entered into the Technology Database in accordance with the guidelines provided. Where time or resources are limited, the user may rely on the default global values already included for most key technologies.

01_Tool – Calibration (Excel file)

This tool forms the backbone of the MDI. Its purpose is to provide the initial conditions for a reference year, using historical data. It is pre-populated with energy balances and power generation by source for all countries. Building on the energy balances, the tool generates the necessary data to describe each sub-sector included in the MDI. While the energy balances are the most critical input, additional user-sourced data can improve the quality of the calibration. If such data is unavailable, default values can be applied.

For example, the energy balances provide data on total gas consumption in the residential sector for each country, but do not indicate the split between heating and cooking. In such cases, the user can either obtain specific data, use default values, or apply reasonable assumptions. The outputs of this tool include:

- **Activity levels by technology** (e.g. heat output from gas boilers; kilometres travelled by petrol cars)

- **Total final service or commodity demand by case** (e.g. total heat demand in the residential sector; total passenger-kilometres by all car types)

- **Ratios useful for subsequent modelling stages** (e.g. electricity-to-heat ratio for non-specified industry)

02_Tool – Technoeconomics (Excel file)

This tool draws primarily on data from the **Technology Database** and reformats it into a structure suitable for modelling. For example, while the **Technology Database** provides capital cost data for milestone years only (e.g. 2030 and 2040), modelling requires values for all intermediate

https://doi.org/10.33774/coe-2025-vst77-v2 Content not peer-reviewed by Cambridge University Press.

years. **02_Tool- Technoeconomics** addresses this by applying linear interpolation to generate the necessary annual data points.

The tool focuses on compiling all key techno-economic parameters required to model a technology in OSeMOSYS. Most of the necessary data are sourced directly from the **Technology Database**, meaning that only minimal additional inputs are needed at this stage. The main supplementary input concerns **emission intensities**, for which users may either provide their own values or utilise the default data included in the tool.

03_Tool - Demand Projections (Excel file)

This tool uses the base-year demand values calculated in **01_Tool – Calibration** as the starting point. For each future year, it projects demand by multiplying the base-year value by a specified growth rate, thereby producing annual demand estimates for each demand category.

The tool is pre-populated with growth rates for most demand types and countries. These default values are derived from a simplified techno-economic analysis that uses GDP and population trends as key drivers. Users are encouraged to revise these growth rates where more accurate or context-specific data are available.

An additional function of this tool is to allow the user to define whether each demand is categorised as **Accumulated Annual** or **Specified Annual Demand**. Accumulated Annual Demand is the total yearly demand for a resource that must be met by the end of the year, without restrictions on when during the year it is delivered. Specified Annual Demand is a yearly demand that must be satisfied in each time slice, with the total distributed according to a defined temporal profile (e.g., electricity load curves). This distinction is essential in later modelling stages.

04_Tool - Residual Capacity Power (Excel file)

This tool uses the commissioning date and installed capacity of each type of (combined heat and) power plant to aggregate this data and generate a vintage capacity profile for each technology type. This enables the model to represent the remaining operational capacity of existing plants over time in a realistic manner.

Where such commissioning data are unavailable, this tool may be omitted. In such cases, the vintage capacity profile for power plants can instead be calculated using the same method applied to most end-use technologies (see **08_Tool – Residual Capacity Others**).

05_Tool - Year Split & Load Profiles (Excel file)

This tool calculates two key input parameters for the model.

1. **Year Split** – This represents the share of each time slice within a year. At the time of writing, the tool provides a fixed year split for 30 predefined time slices. Future versions are expected to allow users to define custom time slices and year splits.

2. **Demand Profiles** – These profiles apply to the demands identified as *Specified Annual Demand* in the previous tool (**03_Tool – Demand Projections**). A demand profile expresses the distribution of demand occurrence throughout the year as a share of the total. To generate these profiles, the tool requires both the year split and an hourly demand profile for each demand type. Hourly profiles for a selection of key demands are preloaded in the MDI and can be accessed in the *Profiles* folder.

06_Tool - Renewables CFs (Excel file)

This tool calculates the capacity factor for wind and solar technologies for each time slice. Similar to **05_Tool – Year Split and Demand Profiles**, it uses the predefined year split in combination with hourly capacity factor data to produce the final time-slice-specific values.

Pre-loaded capacity factor profiles, disaggregated by country and by technology (wind and solar), are included in the MDI and can be accessed in the *Profiles* folder.

07_Tool - Constraints (Excel file)

This tool assists the user in defining constraints that influence the behaviour of the model. A set of predefined constraints is included; however, it is at the user's discretion to determine which of these are relevant to their case study and to specify their parameter values accordingly.

In addition to the default options, the user may introduce custom constraints to address specific modelling requirements or policy considerations.

08_Tool - Residual Capacity Others (Excel file)

This tool calculates the vintage capacity profile for all technologies that have an associated capital cost. For technologies without capital costs, residual capacity is not necessary.

To generate these profiles, the user first runs a scenario in which no residual capacity is specified. This allows the model to determine the capacity required in the base year, given the predefined technology activity. A linear decline in capacity is then assumed, based on the operational lifetime of the relevant technology.

Model Input Data (Excel file)

This file is not a tool in the conventional sense, as it performs no calculations. Instead, it compiles the outputs from all preceding tools and organises the data into a structure that is ready for direct use in building a model within MUIO.

Profiles (folder)

This folder contains one Excel file for each country. Each file includes profiles for key demands as well as capacity factor data for wind and solar technologies. All data are sourced from Renewables Ninja (Renewables.Ninja, n.d.).

4. Discussion and Conclusions

The MDI addresses key methodological limitations in previous CLEWs modelling efforts by offering a broader sectoral scope, adjustable model boundaries, and a modular design that can help expedite the data collection and calibration process. By combining pre-loaded global datasets with the ability to integrate local information, the MDI strikes a balance between accessibility and customisability, enabling model development for any country without extensive coding expertise.

Compared with earlier initiatives that focused primarily on power or selected energy sub-sectors, the CLEWs++ scope within the MDI allows for the collection of data, and thus modelling, of numerous sectors. This has extended benefits to previous CLEWs and sector-specific models as it provides more comprehensive greenhouse gas coverage, including waste and industrial process emissions, and allows one to potentially explore more interconnections between several sectors beyond just climate, energy, land, and water. The modular architecture not only facilitates targeted updates but also reduces the risk of disrupting the entire workflow when introducing improvements. This approach supports both reproducibility and long-term maintainability – two factors often cited as challenges in complex system modelling.

Since this is the first version of the MDI, some limitations remain, but these do not diminish its contribution to addressing key methodological challenges such as standardisation, reproducibility, and accessibility. For instance, the reliance on Excel can constrain automation of very large-scale applications, as multiple interlinked files may occasionally strain system capacity. Likewise, the pre-loaded datasets are designed to lower entry barriers and support rapid model construction, but they are best viewed as a starting point that should be complemented with robust, locally validated data. Finally, while the MDI is built with flexibility in mind, its current functionality is most strongly aligned with the OSeMOSYS platform, which may delay adoption by users of other modelling environments.

Future development priorities include expanding the library of default datasets, enabling user-defined time slices, improving integration with other modelling environments, and enhancing automation for large-scale or multi-region applications by using macros or Power Query. These improvements will help to further lower entry barriers and broaden the MDI's applicability in both research and policy contexts.

In conclusion, the MDI provides a practical, structured approach to building CLEWs++ models that can accelerate the adoption of integrated systems analysis in climate, resource, and development planning. By reducing the time, effort, and risk associated with starting from scratch, it supports more consistent, transparent, and reproducible modelling workflows, ultimately strengthening the analytical basis for informed decision-making.

5. References

Albrecht, T., Crootof, A. & Scott, C. (2018) The Water-Energy-Food Nexus: A systematic review of methods for nexus assessment. *Environmental Research Letters.* 13 10.1088/1748-9326/aaa9c6.

Alexander, K. A., Tan, N., Gardumi, F., Plazas-Niño, F. A., Kuling, K., Ramos, E. & Martindale, L., 2025. Analysing the development of the climate, land, energy, and water systems (CLEWs) modelling framework: a state-of-the-art review. *Environmental Research* (preprint), published 05 August. Available at: https://doi.org/10.1088/2752-5295/adf504 [Accessed 14 August 2025].

Avgerinopoulos, G., Flint Smith, A., Abeysundara, M. K. & Howells, M., 2025. Model Development Infrastructure. Zenodo. Available at: https://doi.org/10.5281/zenodo.16875355 [Accessed 14 August 2025].

Barnes, T., Shivakumar, A., Brinkerink, M. *et al.* OSeMOSYS Global, an open-source, open data global electricity system model generator. *Sci Data* **9**, 623 (2022). Available at: https://doi.org/10.1038/s41597-022-01737-0.

Cannone, C., Allington, L., Cervantes Barron, K., Charbonnier, F., Zachau Walker, M., Halloran, C., Yeganyan, R., Tan, N., Cullen, J. M., Harrison, J., To, L. S. & Howells, M. (2023) Designing a zero-order energy transition model: How to create a new Starter Data Kit. *MethodsX.* 10 102120. 10.1016/j.mex.2023.102120.

Dhir, S., Buisson, E. & Kim, T.-Y., 2025. Growing geopolitical tensions underscore the need for stronger action on critical minerals security. IEA Commentaries, 9 February. Available at: https://www.iea.org/commentaries/growing-geopolitical-tensions-underscore-the-need-for-stronger-action-on-critical-minerals-security [Accessed 14 August 2025].

Howells, M., Quiros-Tortos, J., Merven, B., Arderne, C., Gardumi, F., Laskurain-Iturbe, I., Moksnes, N., Mentis, D., Taliotis, C., Ramos, E., Burandt, T., Ouedraogo, N., Usher, W., Broad, O., Pappis, I. and Conradie, B. (2023) ‘Energy starter data kits for evidence-based policy making: An open-source, replicable and flexible approach for national energy planning’, *Energy Strategy Reviews*, 47, p. 101092. Available at: https://doi.org/10.1016/j.esr.2023.101092.

Kuling, K., Saedi, Y., Barnes, T., Sunder Rajan, A., and Niet, T. (2023) ‘CLEWs Global: An open source, open data Climate, Land, Energy, and Water systems model generator’. Presented at the 41st International Energy Workshop, June 2023, Colorado, USA. Available at: https://summit.sfu.ca/item/36641 [Accessed 21 August 2025].

Kuling, C., Huppmann, D., Welsch, M., Martinez, G., Brinkerink, M., Broad, O., Kober, T. and Rogelj, J. (2022) ‘Climate, Land, Energy and Water systems (CLEWs) modelling to support national development planning: A review of applications and opportunities’, *Environmental Science & Policy*, 136, pp. 1–13. Available at: https://doi.org/10.1016/j.envsci.2022.06.005.

Martindale, L., Cannone, C., Niet, T., Hodgkins, R., Alexander, K. & Howells, M. (2023) Empowering Tomorrow's Problem Solvers: Nexus Thinking and CLEWs Modelling as a Pedagogical Approach to Wicked Problems. *Energies.* 16 (14), https://doi.org/10.3390/en16145539.

Renewables.ninja, n.d. *Renewables.ninja*. Available at: https://www.renewables.ninja/ [Accessed 21 August 2025].

Saedi, Y. (2016) 'Enhancing Open Source CLEWs Models with High-Resolution Land and Water Data', *Simon Fraser University Summit Institutional Repository*. Available at: https://summit.sfu.ca/item/38005.

Saedi, Y. & Niet, T., 2024. *Climate, Land, Energy and Water systems: a Global Model*. GitHub. Available at: https://github.com/OSeMOSYS/CLEWs_GAEZ [Accessed 21 August 2025].

Weirich, F., Howells, M., Lyth, J., Bazilian, M., Hermann, S., Welsch, M., Segerström, R., Rogner, H.-H. and Rogner, H. (2013) 'Towards an integrated climate, land-use, energy and water (CLEW) framework: A case study of Mauritius', *Natural Resources Forum*, 37(4), pp. 255–268. Available at: https://doi.org/10.1111/1477-8947.12019.