A clustering approach to improve spatial representation in water-energy-food models

Abhishek Shivakumar, Thomas Alfstad and Taco Niet

1 Department of Economic and Social Affairs, United Nations, New York, United States of America
2 School of Sustainable Energy Engineering, Simon Fraser University, Surrey, BC, Canada

E-mail: abhishek.shivakumar@desa.kth.se

Keywords: water-energy-food nexus, open-source modelling, open data, sustainable development, hierarchical clustering, optimisation models

Abstract
Currently available water-energy-food (WEF) modelling frameworks to analyse cross-sectoral interactions often share one or more of the following gaps: (a) lack of integration between sectors, (b) coarse spatial representation, and (c) lack of reproducible methods of nexus assessment. In this paper, we present a novel clustering tool as an expansion to the Climate-Land-Energy-Water-Systems modelling framework used to quantify inter-sectoral linkages between water, energy, and food systems. The clustering tool uses Agglomerative Hierarchical clustering to aggregate spatial data related to the land and water sectors. Using clusters of aggregated data reconciles the need for a spatially resolved representation of the land-use and water sectors with the computational and data requirements to efficiently solve such a model. The aggregated clusters, combined together with energy system components, form an integrated resource planning structure. The modelling framework is underpinned by an open-source energy system modelling tool—OSeMOSYS—and uses publicly available data with global coverage. By doing so, the modelling framework allows for reproducible WEF nexus assessments. The approach is used to explore the inter-sectoral linkages between the energy, land-use, and water sectors of Viet Nam out to 2030. A validation of the clustering approach confirms that underlying trends actual crop yield data are preserved in the resultant clusters. Finally, changes in cultivated area of selected crops are observed and differences in levels of crop migration are identified.

1. Introduction
The past two decades have seen a resurgence in the practice of national planning. A total of 134 countries, home to almost 80% of the world’s population, had national development plans by 2018, up from 62 in 2006 (Chimhowu et al 2019). In many developing countries this renewed interest has come as a response to development challenges such as those embodied in the Millennium Development Goals and the Sustainable Development Goals (SDGs), because the comprehensive and complex nature of these goals require a comprehensive planning and monitoring framework (Chimhowu et al 2019).

Incorporating scientific and analytic evidence into decision-making processes is crucial to the success of national planning and policy processes. A systemic approach that helps identify and manage trade-offs while maximising co-benefits is needed to achieve the SDGs (Independent Group of Scientists appointed by the Secretary-General 2019). More than half the countries with national plans have embraced this view and have put emphasis on evidence-based processes, including economic modelling and scenario analysis, to develop their plans (Chimhowu et al 2019).

There is growing recognition of the need to consider interdependencies between water, energy and food sectors for effective resource planning (Miralles-Wilhelm 2016, Smajgl et al 2016, Kurian 2017, Monier et al 2018). This is reflected in the growing number of water-energy-food (WEF) models and frameworks found in the literature. These include a mix of established modelling frameworks and more
recent approaches that focus on the WEF resource nexus. Examples of the former include: IMAGE (van Vuuren et al 2015), WITCH (Bosetti et al 2015), REMIND (Luderer et al 2015), and GCAM (Calvin et al 2019); while that of the latter include: NEST (Vinca et al 2020), WHAT-IF (Payet-Burin et al 2019), AWASH (Rising 2020), K-WEFS (Purwanto et al 2021), Endo et al (2020) and Albrecht et al (2018) provide comprehensive reviews of available WEF models and frameworks. The former analyses 25 WEF reviews, while the latter systematically assesses the analytical approaches of 73 WEF studies, and then focus their analysis on 18 selected studies that represent that variety of methods used. However, this and other reviews of available WEF models highlight important gaps that exist in many currently available models and approaches (Kaddoura and El Khatib 2017, Albrecht et al 2018, Shannak et al 2018). The gaps identified include: (a) falling short of capturing interactions between water, energy, and food systems; (b) coarse spatial representation of resource systems; and (c) the lack of reproducible methods of nexus assessment.

Several nexus studies focus on dual-sector interactions e.g. water-energy, water-food, or food-energy. In particular, past approaches were typically water-centric (Smajgl et al 2016). However, in order to improve coordination of cross-sector policies, an integrated analysis that considers water, energy, and food sectors simultaneously and equally is required (Chang et al 2016, Miralles-Wilhelm 2016, Albrecht et al 2018). Such analyses need to be underpinned by similarly integrated models or frameworks. In this respect, currently available WEF models can be broadly divided into soft- and hard-linked models (Brouwer et al 2018). The former is characterised by modelling frameworks that consist of sector-specific modules (CAPRI: Blanco et al 2017; AgMIP: Lampe et al 2014, Ruane et al 2017). Cross-sectoral interactions in these models are represented by sharing of input and output data between modules. An advantage of such an approach is that the dynamics of each sector can be captured in detail. However, this approach has some drawbacks; each of the sector-specific models will prioritise the optimal configuration of that sector over others and there is scope for inconsistencies between input data and assumptions used in the different sector-specific models or modules. As mentioned above, water, energy, and food sectors need to be considered simultaneously and equally—through a fully integrated approach—to sufficiently capture their interdependencies. The approach presented here is a fully integrated (hard-linked) modelling framework.

The spatial and temporal resolution of different nexus models varies greatly in the literature. MuSIASEM (Giampietro et al 2009, Gerber and Scheidel 2018) and the WEF Nexus Tool (Water-Energy-Food Nexus Working Group 2014, Daher and Mohtat 2015) apply to a point in time, though MuSIASEM can be applied at different spatial resolutions. Water Evaluation And Planning System (WEAP) and associated models do not include detailed spatial resolution (Sieber and SEI-US Center 2020). Prior climate, land-use, energy, and water systems (CLEWs) studies have used administrative boundaries for spatial disaggregation even though many regions crossed agro-ecological zones (Rogner et al 2009, Fischer et al 2013, Welsch et al 2014). Al-Saidi and Hussein (2021) highlights the need for spatial aspects to be re-examined and better analysed in the WEF nexus. This includes issues such as cross-regional interdependencies, trade, globalisation, relocations of production, and international collaboration. Zhang et al (2019a) finds that the spatial scope and number of sectors included in WEF studies is increasing over time but increased resolution is needed. Rasul (2014) highlights the importance of considering inter-regional connections in performing WEF analyses while Zhang et al (2018) finds that the integration of analyses across temporal and spatial boundaries and the evaluation of the combined nexus are critical. Although progress is being made, there is a need to expand the spatial resolution of existing models to address these shortcomings (Aryanpur et al 2021, Martínez-Gordón et al 2021).

Tools and methods must be accessible and reproducible but many of the tools used for WEF analysis are closed source and not openly available. The WEF Nexus Tool provides no information on model details nor how to apply the model to different jurisdictions (Water-Energy-Food Nexus Working Group 2014). WEAP requires a paid license (Sieber and SEI-US Center 2020). MuSIASEM is a framework for analysis and there is no 'tool' that supports the analyses (Giampietro et al 2009, Gerber and Scheidel 2018). A number of WEF related studies develop their own code for nexus analysis but do not publish this code open source (Zhang and Vesselinov 2017, Bieber et al 2018, Li et al 2019, Zhang et al 2019b, Sadeghi et al 2020, Yan et al 2020). NEST is open source but its econometric model details have not been made available and the tool requires a commercial solver (Parkinson et al 2018, Vinca et al 2020, Wada et al 2019). The CLEW's framework has a fully open toolset.

In this paper, we present a modelling framework that aims to address some important gaps identified in existing WEF models. We include results from an application of the framework to quantify the cross-sectoral interactions between water, energy, and food in Viet Nam. Specifically, we describe a novel modelling framework which incorporates the following features:

- Fully integrated approach, considering WEF sectors simultaneously and equally.
- Consideration of spatial aspects of WEF resources and trade-offs with computational complexities.
Flexible, scalable, and accessible, allowing for reproducible nexus assessments across different scales, geographies, and political contexts using open code and data.

The rest of the paper is organised as follows: the components of the modelling framework are described in section 2. This includes details on how the water, land-use, and energy sectors are represented. Further, a clustering tool developed to address cross-scale issues between the land-use and water is presented. Section 2 concludes with a description of the underlying modelling tool—OSeMOSYS (Howells et al 2011)—and the datasets used; section 3 first provides a description of a case study of Viet Nam, where the modelling framework is applied. An analysis of the main results from applying the framework to the selected case study is then presented. It includes a discussion of the main insights that can be derived from this application specifically for Viet Nam and more broadly for other WEF nexus assessments. Section 4 summarises the main conclusions of the study, mentioning some limitations, areas of future work, and recommendations to advance WEF modelling frameworks.

2. Methods

The methodology developed and applied in this study—presented below—aims to incorporate three main features: a fully integrated representation of WEF systems; an improved consideration of spatial aspects of WEF systems; and the development and use of a flexible, open, and accessible modelling framework.

2.1. Clustering tool to aggregate spatial data

The consideration of spatial aspects in WEF models is especially important for the water and land-use sectors. Resource dynamics such as water availability, precipitation, crop suitability, harvested area, land-use type often vary significantly by location (Shannak et al 2018, McGrane et al 2019, Liang et al 2020). In order to capture spatial differences, the water and land-use sectors are represented as sub-regional units or ‘clusters’. These sub-regional units can be derived based on, for e.g. a coarse resolution of administrative regions or detailed spatial data on agro-climatic features. While data and computational requirements in the former are relatively low, important spatial features such as soil characteristics and water availability will not be captured. In the latter case, resource dynamics can be captured at high spatial detail. However, incorporating high resolution spatial data in a long-term optimisation model will significantly increase model complexity, data requirements, and the computational burden.

In order to reconcile these two considerations, a spatial data aggregation technique is proposed and applied in this study. The spatial aggregation is based on a clustering algorithm that combines ‘cells’ of high-resolution spatial data into ‘clusters’ based on their similarity across a set of features. The specific algorithm employed here is Agglomerative Hierarchical clustering using Ward’s method (Lance and Williams 1967); in this technique, individual data points are successively combined into ‘clusters’ with the aim of minimising the variance within each cluster. The total variance—measured as the error sum of squares (ESS)—is calculated as the difference between each data point in a cluster and the cluster average for one or more selected features that the clustering tool is based on. The number of clusters is user-defined and can range between 1 (maximum ESS) and the total number of data points being clustered (ESS = 0). The improvement in accuracy of representing underlying data by increasing the number clusters is quantified in figure 8 (appendix).

The aggregation of spatial data using clustering is well-established and is employed in several geospatial analyses (Ip et al 2010, Getman et al 2015, Chen et al 2017, Siala and Mahfouz 2019, Wang et al 2019, Yue et al 2019, Cesario et al 2020). Several other clustering approaches exist—such as k-means, mean-shift, DBSCAN—each with its own advantages and disadvantages (Rodriguez et al 2019). The main reasons for selecting the hierarchical clustering are: clustering process remains the same for any number of clusters, relatively high accuracy for higher numbers of features, and avoids the need to select starting points.

An Agglomerative Hierarchical clustering algorithm applied in this study is based on crop suitability data across a range of crops (e.g. maize, rice paddies, sugarcane, cassava), under multiple water supply options (rain-fed, irrigated) and input/management levels (low, intermediate, high). The definition of each category is provided in appendix. Each combination of crop, water supply, and management level is referred to as a ‘crop combination’ here. The clustering tool clusters the ‘cells’ of land in the country, region, or sub-region based on crop suitability—known as agro-climatically attainable yield—available at 5 arc-min resolution (approx. 10 km × 10 km) across a range of crop combinations. Spatial data on land cover, evapotranspiration, crop water deficit, and precipitation—also at 5 arc-min resolution—are then combined to the clusters based on their coordinates. A summary of the proposed clustering approach is shown in figure 1. The study presented in this paper considers 15 clusters. The choice of number of clusters is informed by a sensitivity analysis on the decrease in total ‘error’ (ESS) with each additional cluster. This is described in further detail in the appendix (under ‘Analysis of clustering results’).

Each resulting cluster represents cells of land with similar crop suitability—measured as
Figure 1. Summary of proposed Agglomerative Hierarchical clustering approach based on crop yields across a range of crops, water supply options, and management levels. In this example, high-resolution spatial data on land suitability (crop yield in tonnes ha$^{-1}$) for six crops aggregated into 15 clusters. As part of the clustering process, land cover and precipitation data are also aggregated accordingly.

‘agro-climatrically attainable yield’—across the geographic area of interest. The clustering process can be carried out over a range of different scales (global, regional, national, sub-national) and can be combined with other geospatial considerations such as administrative boundaries or other resource systems such as river basins and renewable energy potential. The clustering tool developed in this study is being maintained as an open-source project.

2.2. Model structure and representation of WEF sectors

The clustering process described above allows for high-resolution spatial data to be represented in a modelling framework; a linear optimisation model is used in this study. Each cluster is represented in the model as land that can be allocated to a range of uses, under a set of biophysical constraints. Each cluster has several inputs and outputs that represent the allocation of land and link it other resources, namely water and energy. The model structure of the land and water sectors is shown in figures 2 and 3. Each potential use of a land cluster is associated with different costs, water requirements, energy requirements, water outputs, and crop production. The model's objective function aims to allocate the land to different uses ‘optimally’ i.e. at the lowest overall cost over the entire modelled period under the given constraints. The total system cost being minimised is broadly divided into three categories: fixed, variable, and capital. Costs in each sector are represented as one of these three categories. For instance, fuel costs and maintenance costs for power plants are represented as variable and fixed costs, respectively. The first is dependent on the usage of a given power plant while the latter is fixed for a given power plant capacity, independent of its level of usage. In the land sector, the cost of installing a new irrigation system—to convert rainfed agriculture to irrigated—is represented as a capital cost that is incurred in the year of construction and is then available for the operational lifetime of the asset. All cost components, representing different categories and sectors, are aggregated as the total system cost in the objective function of the model. This total system cost is minimised during the optimisation process to provide a least cost ‘optimal’ solution while meeting user-defined constraints and demands.

The irrigation requirements are derived from clustered spatial data for ‘crop water deficit’ across all crop combinations. Similarly, evapotranspiration and precipitation values from the clustering are also used, the former calculated for all crop combinations. Values for run-off and groundwater recharge are estimated based on a fixed ratio between the two after considering evapotranspiration. The water inputs (precipitation and irrigation) and outputs (evapotranspiration, run-off, and groundwater recharge) are balanced across all clusters. The crop outputs from each cluster are summed up to meet an exogenous total demand.

The model presented here considers all three WEF sectors equally and simultaneously. The representation of the energy sector follows that of traditional energy system models such TIMES-MARKAL (Loulou and Labriet 2007) and MESSAGE (MESSAGE—IIASA 2020). The energy sector representation comprises a set of power plants,
electricity transmission and distribution, fuel extraction, imports and exports. The model structure of the energy sector is shown in figure 4. The figure shown focuses on the electricity sector. However, the model includes other energy vectors in each demand sector. These include: biomass in the commercial, residential, and agricultural sectors; diesel, petrol, and natural gas in the transport sector; coal, oil, and natural gas in the industry sector. Interlinkages between the water, energy, and land use systems are also represented in the model, as shown in figures 3 and 4. Groundwater recharge and surface water resulting from land based on different land uses is used as an input in the energy and water sectors for cooling powerplants and public water supply, respectively. Similarly, electricity from the energy sector is used as an input for irrigation in the agriculture sector.
Figure 4. Representation of the electricity sector in the model. The power plant categories are further disaggregated in the model by region, technology-type (e.g. steam turbine, combined cycle gas turbine), and age (existing or planned). These sub-categories are combined here for the purpose of illustration. Linkages between water, land-use, and the energy sector are considered through (1) electricity for irrigation pumping, (2) land requirements for powerplants, (3) cooling water for power plants, (4) water and land-use for biomass production.

The model is calibrated\(^5\) for the period 2016–2018 to match historical data. This includes installed powerplant capacities, energy demand by sector, land cover by land cover type, and crop area by water supply type (irrigated or rainfed).

### 2.3. Model code and data

The framework presented in this study is built using OSeMOSYS (Howells et al 2011), a widely used open source modelling tool. OSeMOSYS is a dynamic, bottom-up, multi-year system modelling tool solving a linear program to determine the least cost investment strategy required to satisfy exogenously defined resource demands. Technical, economic, and environmental implications associated with the identified least-cost systems can be easily extracted from the model results. Like other optimisation models, OSeMOSYS assumes a perfect market with perfect competition and foresight. OSeMOSYS has been applied across a range of studies at varying geographic scales to address diverse research questions (Welsch et al 2012, Taliotis et al 2016, de Moura et al 2017, Löfler et al 2017, Emodi et al 2019). OSeMOSYS is maintained by a large, active community and can be downloaded from its GitHub page\(^6\).

The proposed modelling framework is developed around open, publicly available datasets with global coverage wherever possible. Take together with the underlying open source model, this will allow the framework to be applied in a wide range of geographical contexts, accessibility of data and assumptions, and reproducibility of nexus assessments. The clustering methodology was applied using high-resolution spatial data from the Global Agro-Ecological Zones (GAEZ) data portal (FAO and IIASA 2021) for land cover, crop suitability across a range of water supply options and management levels, annual rainfall, evapotranspiration, and crop water deficit. However, data from GAEZ is used as an example—the clustering tool can readily use alternate sources of spatial data on these factors wherever available and relevant. This categorisation is significant within the proposed modelling framework which aims to optimise long-term investment decisions in a WEF system such as

\(^5\) Model calibration in this context refers to a distinction in model results between historical data and modelled results. The modelling framework presented is exploratory rather than predictive—it is meant to be used for ‘what-if’ type analyses of rather than provide an accurate prediction of future resource systems.

\(^6\) [https://github.com/OSeMOSYS/OSeMOSYS](https://github.com/OSeMOSYS/OSeMOSYS). The specific version of the code developed for this study—modified from the original to improve the representation of land-use and water systems—is available at [https://github.com/OSeMOSYS/CLEWs/tree/master/osemosys_code](https://github.com/OSeMOSYS/CLEWs/tree/master/osemosys_code).
the cost of shifting from one management practice to obtain higher crop yields.

3. Results and discussion

In this section we first describe a real-world case study—a WEF model of Viet Nam—where the proposed methodology is applied. The main findings from this case study are then discussed.

Viet Nam’s rapid economic growth over the past two decades has been accompanied by soaring demands for energy, water, and land resources (Asian Development Bank 2015). In the WEF model for Viet Nam presented here, the energy sector components were first developed in OSeMOSYS with an outlook to the year 2030. The energy sector representation includes all powerplants in Viet Nam (Vitina et al 2017), aggregated by technology categories (e.g. coal, hydropower, natural gas, solar PV, wind), region (North, Central, South), and stage of completion (existing or planned). In total, the model is comprised of 512 ‘technologies’ with 440 ‘commodities’ connecting them. Each year of the model is divided into six representative ‘timeslices’ based on two seasons and three day-parts (base, intermediate, and peak) in each season.

The land-use and water systems are then added to the model. The first step of this process is to apply the clustering tool to aggregate high-resolution spatial data on agro-climatic features into clusters to be represented in the model. The crops selected to be represented individually in this study are cassava, coffee, maize, rice (paddy), and sugarcane. The remaining food crops are combined as vegetables and fruits. The land suitability for each crop is considered under two water supply options (irrigated and rain-fed) and three management levels (low, intermediate, high).

3.1. Comparison of crop yield values

The primary aim of the proposed clustering tool is to reconcile the need for a spatially resolved WEF model with the computational complexity of optimising such a model. The clusters resulting from the tool must therefore accurately represent trends of the actual values in the high-resolution spatial data. The WEF model utilises these clustered crop yield values to represent competition for land between different crops. Crop yields can typically be improved by investing in irrigation systems or improved cropland management or both. The extent to which crop yields can be improved through the above options varies by both crop and geographic region. Crop yield values for different crop combinations—i.e. crops under different water supply options and management levels—is shown in figure 5. These values are compared to those resulting from the proposed clustering tool.

The figure shows that the resultant clusters capture both the dominant and extreme values within the actual spatial data. Moreover, these trends are captured across all crop combinations. This can be seen by comparing the actual and clustered values in the figure on three measures: (a) range and distribution of values; (b) median values, and; (c) preservation of trends across crop combinations; the third point is especially important to capture since investment decisions in the WEF model are based on the potential improvement in yields across the crop combinations.

The yield of Cassava, for instance, can be improved from around 3 tonnes ha$^{-1}$ under rain-fed conditions and low management level to around 10 tonnes ha$^{-1}$ under irrigated conditions and high management level (both median values). A similar tripling of attainable crop yield is observed across the remaining crops. This trend in the actual crop yield data is preserved in the clustered crop yield values as well.

3.2. Land-use change

The use of high-resolution spatial data allows for other insights in addition to the national level results discussed above. For instance, it allows for an assessment of potential crop migration to minimise the total system cost over the modelled time horizon. Model results for land-use change by land-use type are shown in figure 6. It can be observed that areas under cultivation for a crop show one of the following trends: significant increase in all main areas of existing cultivation (low crop migration) or; significant increase in some main areas of existing cultivation and decrease in others (moderate crop migration) or; shift in main areas of existing cultivation from one to another.

Sugarcane, vegetables and fruits fall into the category of low crop migration; the main areas of existing cultivation for each crop—North for sugarcane and South-Central for vegetables and fruits—are further consolidated in the future. The former crop going from occupying land shares of around 10%–15% in 2020 to around 20%–35% by 2030. In the latter crop’s case, the land share in occupied clusters increases from around 15% in 2020 to almost 30% by 2030. Meanwhile, cassava, maize, and rice follow a moderate crop migration trend; the land share allocated to cassava increases in the North-Central region decreases in North and South during the same time period. Similarly, maize cultivation decreases in the South-Central region but more than doubles in parts of the North, from around 20% in 2020 to over 40% by 2030. Rice is a widely cultivated crop in Viet Nam but particularly so in the Red River Delta in the North-East and the Mekong River Delta at the Southern end of the country. The land share allocated to rice cultivation is further increased in these fertile delta regions. At the same time, some decrease in land share is observed in some parts of the North. This is likely due to competition with other crops such as sugarcane as well as other land-uses such as forests.
Figure 5. Comparison of actual and clustered crop yield values in tonnes ha$^{-1}$.

Figure 6. Land-use change by land-use type. Each map represents the share of land area (as a %) allocated to different land-uses. The top row are the values in 2020 and the bottom row are those for 2030.
3.3. Sector interactions

The WEF model provides insights on the evolution of energy, water, and land use systems over the modelled time horizon. Figure 7 below shows the main results for each sector as well as that of interlinkages between sectors. The electricity generation mix is seen to move from a system that is evenly split between natural gas, hydro and coal to one that is natural gas dominated. This is as a result of increasing use of domestic gas in Vietnam. An increasingly fossil-fuel dominated system however results in a significant increase in water demand for cooling powerplants, from 8 billion cubic meters (BCM) in 2016 to around 21 BCM in 2030—an increase of 160%. Further, it results in a dramatic increase in CO$_2$ emissions; the power sector goes from emitting 70 million tonnes of CO$_2$ in 2016 to 187 million tonnes of CO$_2$ by 2030.

The land use sector, in the meantime, shows a significant expansion of agricultural land to meet a rapidly growing population as well as to meet Vietnam's ambitious crop export targets. Total agricultural land increases from 142,000 sq. km in 2016 to 214,000 sq. km in 2030—an increase of 51%. This expansion in agricultural land comes at the cost of decreasing forests and grasslands, each declining by 50% and 44% between 2016 and 2030, respectively. Further, the increase in agricultural land results in growing water demands for irrigation, which increases from 18 BCM in 2016 to 35 BCM in 2030—an increase of 89%.

These results highlight important sectoral and intersectoral dynamics that can be analysed through a WEF model. The scenario studied was a baseline or 'business-as-usual' scenario where the main policy is a relatively un-ambitious renewable power development master plan (PDP 7). The RE target in this plan is easily reached with hydropower alone, requiring little additional investment in other renewables such as wind and solar power (each representing around 3%-4% of generation capacity by 2030). A 'renewables-led pathway' would lead to around 1.1 gigatons fewer greenhouse gas emissions by 2030. In addition, it would also significantly lower sulphur-dioxide and nitrogen-oxide emissions. In the current power development master plan (PDP7), the significant increase in emissions would have a serious impact.
Crop demands are fixed in the modelling study presented. This is indeed a simplification. However, the model itself is able to represent dynamic crop demands that respond to—among other factors—export prices, changing yields, and water availability. There is no technical limitation preventing these from being included in a future iteration of the model. The focus in this model version is on improving the spatial aspects of the proposed approach for modelling a WEF nexus. Further, the results from the study do point to a tendency of crops to respond to changing yields across Viet Nam. Therefore, studying the impact of other factors on crop selections and level of agricultural modernisation.

The total projected water demands in the model (56 bcm) can be met with the total renewable water resources (RWRs) in Viet Nam (884 bcm). However, the spatial distribution of demands and resources needs to be considered in further detail to identify potential sources of unsustainable water withdrawals. In addition, 59% of Viet Nam’s RWR is from external sources. This leaves it vulnerable to changes—both natural and man-made—in upstream water flow. This is already being seen in the increasing salination of freshwater resources downstream in places such as the Mekong River Delta. However, the representation of water dynamics such as salination is not yet included in the presented model.

The aim of such a model is to allow for more extensive scenario analyses that help identify policies and actions that can positively impact multiple sectors simultaneously. No such specific policies were included in the presented study. However, the large-scale deployment of wind power has the potential to do so. It would lead to a shift from high emission power generation capacity. This would include the additional benefit of reducing cooling water requirements for fossil fuel power plants. On the land use side, the parts of Viet Nam with the highest wind power potential are the highlands which are locations with less likelihood of competition for land with rice cultivation, the major crop of the country.

Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: 10.5281/zenodo.5348228.

Appendix

Input/management levels

GAEZ defines three generic levels of input/management: low, intermediate, and high input level (Fischer et al 2012).

Low level inputs

Under a low level of inputs (traditional management assumption), the farming system is largely subsistence based. Production is based on the use of traditional cultivars (if improved cultivars are used, they are treated in the same way as local cultivars), labour intensive techniques, and no application of nutrients, no use of chemicals for pest and disease control and minimum conservation measures.

Intermediate level inputs

Under an intermediate level of input (improved management assumption), the farming system is partly market oriented. Production for subsistence plus commercial sale is a management objective. Production is based on improved varieties, on manual labour with hand tools and/or animal traction and...
some mechanisation, is medium labour intensive, uses some fertiliser application and chemical pest disease and weed control, adequate fallows and some conservation measures.

**High level inputs**
Under a high level of input (advanced management assumption), the farming system is mainly market oriented. Commercial production is a management objective. Production is based on improved or high yielding varieties, is fully mechanised with low labour intensity and uses optimum applications of nutrients and chemical pest, disease and weed control.

**Analysis of clustering results**
The aim of any clustering process is to combine similar data points while not losing important differences within the underlying data. As mentioned in the section 2, the difference between average value of a cluster and its underlying data is measured as the error sum of squares (ESS). Aggregating all data points in a single cluster result in the maximum ESS. Conversely, ESS = 0 when the number of clusters is equal to the number of underlying data points. Each data set follows a different trend of decreasing ESS when increasing the number of clusters. This trend can be captured using an 'Elbow graph', as shown in figure 8 below. This 'Elbow graph' is for the input data set for the Viet Nam (VNM) WEF model presented in this paper.

With each additional cluster, the total ESS (measured as a % share of maximum error) decreases. The marginal gain in accuracy of representation is highest when the number of clusters is low and decreases thereafter. There is a noticeable change in the marginal gain from cluster 5 onwards as compared to clusters 1–5. This results in an 'elbow', which gives this type of figure its name.

**Computational tests**
Below is a comparison of the computational effort for three different test models, each with a different number of clusters. Table 1 also includes details on the total matrix size which indicates the size of the 'problem' to be solved.

The computations tests were carried out on a laptop with the following specifications:
- **Processor**: Intel(R) Core(TM) i7-8750H CPU@2.20 GHz, 2208 MHz, 6 core(s), 12 logical processor(s).
- **Installed physical memory (RAM)**: 16.0 GB.
- **Operating system (OS)**: Microsoft Windows 10 Home.

**Comparison of clustering with counterfactual (no clustering)**
In this section, results of a counterfactual (no clustering) are compared with that of 15 clusters. The series of figures below show crop yield (normalised) for a range of crop combinations. Each line in a figure
Table 1. Comparison of computational effort required to solve models with different numbers of clusters.

|                     | 7 clusters | 15 clusters | 30 clusters |
|---------------------|------------|-------------|-------------|
| Number of rows      | 372 207    | 398 127     | 440 907     |
| Number of columns   | 302 183    | 348 767     | 426 407     |
| Matrix generation   | 19 s       | 24 s        | 40 s        |
| Solution            | 201 s      | 614 s       | 932 s       |

Figure 9. Crop yield (normalised) for a range of crop combinations in the counterfactual case (no clustering).

represents crop suitability for a cell of land included in that cluster. For the case of no clustering, the lines represent all cells of land in the country.

No clustering
In the counterfactual case (with no clustering), all data is aggregated at the national level. It is clear from the figure below that crop suitability varies significantly between cells of land for different crops. These important differences would therefore not be captured in a model that uses un-clustered data as an input.

15 clusters (as used in the model presented in this paper)
Below is a series of figures (figures 10–24) showing the results of the clustering process. It is clear from the figures—when compared to the counterfactual case—that the cells of land with similar crop suitability across different crop combinations are clustered together. This is highlighted by the fact that the lines that were scattered more widely in the counterfactual case (figure 9), they are bunched more closely together when clustered (into 15 clusters in this case).
Figure 10. Crop yield (normalised) for a range of crop combinations in cluster 1 out of a total of 15 clusters.

Figure 11. Crop yield (normalised) for a range of crop combinations in cluster 2 out of a total of 15 clusters.
Figure 12. Crop yield (normalised) for a range of crop combinations in cluster 3 out of a total of 15 clusters.

Figure 13. Crop yield (normalised) for a range of crop combinations in cluster 4 out of a total of 15 clusters.
Figure 14. Crop yield (normalised) for a range of crop combinations in cluster 5 out of a total of 15 clusters.

Figure 15. Crop yield (normalised) for a range of crop combinations in cluster 6 out of a total of 15 clusters.
Figure 16. Crop yield (normalised) for a range of crop combinations in cluster 7 out of a total of 15 clusters.

Figure 17. Crop yield (normalised) for a range of crop combinations in cluster 8 out of a total of 15 clusters.
Figure 18. Crop yield (normalised) for a range of crop combinations in cluster 9 out of a total of 15 clusters.

Figure 19. Crop yield (normalised) for a range of crop combinations in cluster 10 out of a total of 15 clusters.
Figure 20. Crop yield (normalised) for a range of crop combinations in cluster 11 out of a total of 15 clusters.

Figure 21. Crop yield (normalised) for a range of crop combinations in cluster 12 out of a total of 15 clusters.
Figure 22. Crop yield (normalised) for a range of crop combinations in cluster 13 out of a total of 15 clusters.

Figure 23. Crop yield (normalised) for a range of crop combinations in cluster 14 out of a total of 15 clusters.
Figure 24. Crop yield (normalised) for a range of crop combinations in cluster 15 out of a total of 15 clusters.

ORCID iDs

Abhishek Shivakumar  https://orcid.org/0000-0002-2535-4134
Taco Niet  https://orcid.org/0000-0003-0266-2705

References

Al-Saidi M and Hussein H 2021 The water-energy-food nexus and COVID-19: towards a systematization of impacts and responses Sci. Total Environ. 779 146529
Albrecht T R, Crootof A and Scott C A 2018 The water-energy-food nexus: a systematic review of methods for nexus assessment Environ. Res. Lett. 13 043002
Aryanpur V, O’Gallachoir B, Dai H, Chen W and Glynn J 2021 A review of spatial resolution and regionalisation in national-scale energy systems optimisation models Energy Strategy Rev. 37 100702
Asian Development Bank 2015 Viet Nam: energy sector assessment, strategy, and road map Asian Development Bank
Bieber N, Ker J H, Wang X, Triantafyllidis C, van Dam K H, Koppelaar R H E M and Shah N 2018 Sustainable planning of the energy-water-food nexus using decision making tools Energy Policy 113 584–607
Blanco M, Ramos F, van Doorslaer B, Martinez P, Fumagalli D, Ceglar A and Fernández F J 2017 Climate change impacts on EU agriculture: a regionalized perspective taking into account market-driven adjustments Agric. Syst. 156 52–66
Bosetti V, Marangoni G, Borgonovo E, Díaz Amador I, Barron R, McJean H C, Politis S and Friley P 2015 Sensitivity to energy technology costs: a multi-model comparison analysis Energy Policy 80 244–63
Brouwer F, Agerinopoulos G, Fazekas D, Laspidou C, Mercure J-F, Pollitt H, Ramos E P and Howells M 2018 Energy modelling and the nexus concept Energy Strategy Rev. 19 1–6
Calvin K et al 2019 GCAM v5.1: representing the linkages between energy, water, land, climate, and economic systems Geosci. Model Dev. 12 677–98
Cesario E, Vinci A and Zhu X 2020 Hierarchical clustering of spatial urban data Numerical Computation: Theory and Algorithms, Lecture Notes in Computer Science ed Y D Sergeyev and D E Kvasov (Cham: Springer International Publishing) pp 223–31
Chang Y, Li G, Yao Y, Zhang L and Yu C 2016 Quantifying the water-energy-food nexus: current status and trends Energies 9 65
Chen H, Yang C and Xu X 2017 Clustering vehicle temporal and spatial travel behavior using license plate recognition data J. Adv. Transp. 2017 1738085
Chimhouw A O, Hulme D and Munro L T 2019 The ‘new’ national development planning and global development goals: processes and partnerships World Dev. 120 76–89
Daher B T and Mohitar R H 2015 Water–energy–food (WEF) Nexus Tool 2.0: guiding integrative resource planning and decision-making Water Int. 40 748–71
de Moura P, Loureiro Legey G N, Balderrama L F and Howells M G P 2017 South America power integration, Bolivian electricity export potential and bargaining power: an OSeMOSYS SAMBA approach Energy Strategy Rev. 17 27–36
Emodi N V, Chaiechi T and Alam Beg A B M R 2019 Are emission reduction policies effective under climate change conditions? A backcasting and exploratory scenario approach using the LEAP-OSeMOSYS model Appl. Energy 236 1183–217
Endo A et al 2020 Dynamics of water–energy–food nexus methodology, methods, and tools Curr. Opin. Environ. Sci. Health 13 46–60 Environmental Monitoring Assessment: water-energy-food nexus FAO, IIASA 2021 GAEZ v4 data portal (available at: https://gaez.fao.org/) (Accessed 30 August 2021)
Fischer G, Hiznyk E, van Velthuizen H and Hermann S 2013 Climate, Land, Energy & Water Strategies: A Case Study of Mauritius: Agro-ecological Assessment: Methodology and Results (Laxenburg, Austria: IIASA)
Fischer G, Nachtgeraede F O, Prieler S, Teixeira E, Töth G, van Verelst V H and Wilberg D L 2012 Global agro-ecological zones: model documentation
Gerber J-F and Scheidt A 2018 In search of substantive economics: comparing today’s two major socio-metabolic approaches to the economy—MUEFA and MuSIASEM EcoL. Econ. 144 186–94
Getman D, Lopez A, Mai T and Dyson M 2013 Methodology for clustering high-resolution spatiotemporal solar resource data (No. NREL/TP-6A20-63148, 1225896) (National Renewable Energy Laboratory) (https://doi.org/10.2172/1225896)
Giampietro M, Mayumi K and Ramos-Martín J 2009 Multi-scale integrated analysis of societal and ecosystem metabolism (MuSIASEM): theoretical concepts and basic rationale Energy WEC. 2006 313–22
Howells M et al 2011 OSeMOSYS: the open source energy modeling system Energy Policy 39 9850–70
Independent Group of Scientists appointed by the Secretary-General 2019 Global sustainable development report 2019: the future is now—science for achieving sustainable development (United Nations, New York)
Ip A, Fong S and Liu E 2010 Optimization for allocating BEV recharging stations in urban areas by using hierarchical clustering 2010 6th Int. Conf. on Advanced Information Management and Service (IMS). Presented at the 2010 6th Int. Conf. on Advanced Information Management and Service (IMS) pp 460–5
Kaddoura S and El Khatib S 2017 Review of water-energy-food Nexus tools to improve the Nexus modelling approach for integrated policy making Environ. Sci. Policy 77 114–21
Kurian M 2017 The water-energy-food nexus: trade-offs, thresholds and transdisciplinary approaches to sustainable development Environ. Sci. Policy 68 97–106
Lampe M et al 2014 Why do global long-term scenarios for agriculture differ? An overview of the AgMIP global economic model intercomparison Agric. Econ. 45 3–20
Lance G N and Williams W T 1967 A general theory of classificatory sorting strategies: II. Clustering systems Comput. J. 10 271–7
Li M, Singh V P, Fu Q, Liu D, Li T and Zhou Y 2019 Optimization of agricultural water–food–energy nexus in a random environment: an integrated modelling approach Stoch. Environ. Risk Assess. 35 3–19
Liang Y et al 2020 Quantifying direct and indirect spatial food–water–energy (FWE) nexus in China Environ. Sci. Technol. 54 9791–803
Löffler K, Hainsch K, Burandt T, Oei P-Y, Kemfert C and von Li M, Singh V P, Fu Q, Liu D, Li T and Zhou Y 2019 Optimization of agricultural water–food–energy nexus in a random environment: an integrated modelling approach Stoch. Environ. Risk Assess. 35 3–19
Liang Y et al 2020 Quantifying direct and indirect spatial food–water–energy (FWE) nexus in China Environ. Sci. Technol. 54 9791–803
Löffler K, Hainsch K, Burandt T, Oei P-Y, Kemfert C and von Hirschhausen C 2017 Designing a model for the global energy system—GeneYS-Mod: an application of the open-source energy modeling system (OSeMOSYS) Energies 10 1468
Loulou R and Labriet M 2007 ETAPS-TIAM: the TIMES integrated assessment model Part I: model structure Comput. Manage. Sci. 5 7–50
Ludger G et al 2015 Description of the REMIND model (version 1.6) (SSRN Scholarly Paper No. ID 2697070) (Rochester, NY: Social Science Research Network) (https://doi.org/10.2139/ssrn.2697070)
Martinez-Gordón R, Morales-España G, Sijin J and Fiaai A P C 2021 A review of the role of spatial resolution in energy systems modelling: lessons learned and applicability to the North Sea region Renew. Sustain. Energy Rev. 141 110857
McGrane S J et al 2019 Scaling the nexus: towards integrated frameworks for analysing water, energy and food Geogr. J. 185 419–31
MESSAGE—Iiasa 2020 (available at: https://iiasa.ac.at/web/home/research/modelsData/MESSAGE/MESSAGE.en.html) (Accessed 26 May 2020)
Miralles-Willhelm F 2016 Development and application of integrative modeling tools in support of food-energy-water nexus planning—a research agenda J. Environ. Stud. Sci. 6 3–10
Monier E et al 2018 Toward a consistent modeling framework to assess multi-sectoral climate impacts Nat. Commun. 9 1–8
Parkinson S C, Makowski M, Krey V, Sedraoui K, Almasoud A H and Djilali N 2018 A multi-criteria model analysis framework for assessing integrated water-energy system transformation pathways Appl. Energy 210 477–86
Pausitan K et al 2019 Quantifying carbon for agricultural soil management: from the current status toward a global soil information system Carbon Manage. 10 567–87
Payet-Burin R, Kromann M, Pereira-Cardenal S, Strzapk M and Bauer-Gottwein P 2019 WHAT-IE: an open-source decision support tool for water infrastructure investment planning within the water–energy–food–climate nexus HydroL. Earth Syst. Sci. 23 4129–52
Purwanto A, Suśnik J, Suryadi F X and de Fraiture C 2021 Quantitative simulation of the water-energy-food (WEF) security nexus in a local planning context in Indonesia Sustain. Prod. Consum. 25 198–216
Rasul G 2014 Food, water, and energy security in South Asia: a nexus perspective from the Hindu Kush Himalayan region Environ. Sci. Policy 39 35–48
Rising J 2020 Decision-making and integrated assessment models of the water-energy-food nexus Water Secur. 9 100056
Rodriguez M Z, Comin C H, Casanova D, Bruno O M, Amancio D R, Costa L, Da F and Rodrigues F A 2019 Clustering algorithms: a comparative approach PLoS One 14 e0212536
Rogner H-H, Young C, Herman S, Welsch M, Howells M, Dercon G, Nguyen M, Fischer G and Veld H 2009 Seeking CLEWS—climate, land, energy and water strategies—a pilot case study in Mauritius UN Commission on Sustainable Development (CSD)
Ruane A C et al 2017 An AgMIP framework for improved agricultural representation in integrated assessment models Environ. Res. Lett. 12 125003
Sadeghi S H, Shariﬁ Moghadam E, Delavar M and Zarghami M 2020 Application of water-energy-food nexus approach for designating optimal agricultural management pattern at a watershed scale Agric. Water Manage. 233 106071
Shannak S, Mabrey D and Vittorio M 2018 Moving from theory to practice in the water–energy–food nexus: an evaluation of existing models and frameworks Water Energy Nexus 1 17–25
Siala K and Mahfouz M Y 2019 Impact of the choice of regions on energy system models Energy Strategy Rev. 25 75–85
Sieber J and SEI-US Center 2020 WEAP (water, energy and agriculture) (available at: www.weap21.org) ( Accessed 30 August 2021)
Sloat L I, Davis S J, Gerber J S, Moore F C, Ray D K, West P C and Mueller N D 2020 Climate adaptation by crop migration Nat. Commun. 11 1243
Smaijl A, Ward J and Pluschke L 2016 The Water–food–energy Nexus—realising a new paradigm J. Hydrol. 533 533–40
Sridharan V, Shivakumar A, Niet T, Ramos E P and Howells M 2020 Land, energy and water resource management and its impact on GHG emissions, electricity supply and food production—insights from a Ugandan case study Environ. Res. Commun. 2 085003
Taliotis C, Shivakumar A, Ramos E, Howells M, Mentis D, Sridharan V, Broad O and Mofo L 2016 An indicative analysis of investment opportunities in the African electricity supply sector—using TEMBA (the electricity system model base for Africa) Energy Sustain. Dev. 31 50–66
van Vuuren D P et al 2015 Pathways to achieve a set of ambitious global sustainability objectives by 2050: explorations using the IMAGE integrated assessment model Technol. Forecast. Soc. Change 98 303–23
Vinca A et al 2020 The nexus solutions tool (NEST): an open platform for optimizing multi-scale energy-water-land system transformations Geosci. Model. Dev. 13 1095–21
Vitiša A, Dupont N and Togeby M 2017 Renewable energy scenarios for Vietnam: technical report
Wada Y et al 2019 Co-designing INDUS WATER-ENERGY-LAND FUTURES One Earth 1 183–94
Wang Z, Han Q and de Vries B 2019 Land use/land cover and accessibility: implications of the correlations for land use and transport planning Appl. Spat. Anal. Policy 12 923–40
Water-Energy-Food Nexus Working Group 2014 WEF Nexus Tool 2.0 (available at: www.wefnuxustool.org/login.php) (Accessed 13 June 2020)
Welsch M et al 2014 Adding value with CLEWS—modelling the energy system and its interdependencies for Mauritius Appl. Energy 113 1434–45
Welsch M, Howells M, Bazilian M, DeCarolis J F, Hermann S and Rogner H H 2012 Modelling elements of smart grids—enhancing the OSeMOSYS (open source energy modelling system) code Energy 46 337–50
Yan X, Fang L and Mu L 2020 How does the water-energy-food nexus work in developing countries? An empirical study of China Sci. Total Environ. 716 134791
Yue H, Guan Q, Pan Y, Chen L, Lv J and Yao Y 2019 Detecting clusters over intercity transportation networks using K-shortest paths and hierarchical clustering: a case study of mainland China Int. J. Geogr. Inf. Sci. 33 1082–105
Zhang C, Chen X, Li Y, Ding W and Fu G 2018 Water-energy-food nexus: concepts, questions and methodologies J. Clean. Prod. 195 623–39
Zhang P, Zhang L, Chang Y, Xu M, Hao Y, Liang S, Liu G, Yang Z and Wang C 2019a Food-energy-water (FEW) nexus for urban sustainability: a comprehensive review Resour. Conserv. Recyle 142 215–24
Zhang T Y, Dong I Y, Feng X M and Tan Q 2019b Optimization model for enhancing water-use efficiency from the perspective of water-energy-food nexus IOP Conf. Ser. Earth Environ. Sci. 344 012130
Zhang X and Vesselinov V V 2017 Integrated modeling approach for optimal management of water, energy and food security nexus Adv. Water Resour. 101 1–10