ENVIRONMENTAL RESEARCH
LETTERS

LETTER • OPEN ACCESS

Mapping social-ecological systems archetypes

To cite this article: Juan Rocha *et al* 2020 *Environ. Res. Lett.* **15** 034017

Environmental Research Letters

CrossMark

OPEN ACCESS

RECEIVED
18 October 2019

REVISED
29 November 2019

ACCEPTED FOR PUBLICATION
31 December 2019

PUBLISHED
19 February 2020

LETTER

Mapping social-ecological systems archetypes

Juan Rocha[1,2], Katja Malmborg[1], Line Gordon[1], Kate Brauman[3] and Fabrice DeClerck[1,4]

1 Stockholm Resilience Centre, Stockholm University, Kräftriket 2B, SE-10691 Stockholm, Sweden
2 Beijer Institute, Swedish Royal Academy of Sciences, Lilla Frescativägen 4A, SE-104 05 Stockholm, Sweden
3 Institute of Environment, University of Minnesota, St Paul, MN 55108, United States of America
4 Bioversity International, CGIAR, Montpellier Cedex 5, France

E-mail: juan.rocha@su.se

Keywords: social-ecological systems, archetypes, sustainable development goals

Supplementary material for this article is available online

Abstract

Achieving sustainable development goals requires targeting and monitoring sustainable solutions tailored to different social and ecological contexts. A social-ecological systems (SESs) framework was developed to help diagnose problems, identify complex interactions, and solutions tailored to each SES. Here we develop a data-driven method for upscaling the SES framework and apply it to a context where data is scarce, but also where solutions towards sustainable development are needed. The purpose of upscaling the framework is to create a tool that facilitates decision-making in data-scarce contexts. We mapped SES by applying the framework to poverty alleviation and food security issues in the Volta River basin in Ghana and Burkina Faso. We found archetypical configurations of SES in space, and discuss where agricultural innovations such as water reservoirs might have a stronger impact at increasing food availability and therefore alleviating poverty and hunger. We conclude by outlining how the method can be used in other SES comparative studies.

Zero hunger and no poverty are the first two sustainable development goals [1, 2]. Together with clean water and sanitation, they conform the most basic needs of human beings. Understanding how societies and ecosystems self-organize to meet these basic needs, is a core challenge of sustainability science [2]. Countries around the world have agreed on pursuing 17 sustainable development goals. Achieving them requires targeting and monitoring solutions that fit distinct social-ecological contexts [3]. Countries must therefore understand the diversity and dynamics of social and ecological characteristics of their territories. But data to meet these demands are not always available or even collected [4], so development of methods to quantify social-ecological contexts in data-scarce settings is imperative [5].

Nobel prize winner Elinor Ostrom advocated for embracing social-ecological complexity. Ostrom recognized that there is no universal solutions to problems of overuse of natural resources [6] and further developed a social-ecological systems' (SESs) framework hoping it would help accumulate knowledge and better understanding of what works and what does not in different SES contexts [7]. The SES framework is a nested multi-ier set of variables that has been suggested as features that characterize distinctive aspects of SES. In Ostrom's parlance a SES has 6 key subsystems: resource units (RU), resource system (RS), governance system, users (U, also actors in recent versions), interactions (I) and outcomes (O); all framed by social, economic and political settings (S) as well as by related ecosystems (ECO). Each of these subsystems has a nested second tier of variables (53 in total in the original proposal) aimed to capture key features of the first tier [7]. The framework has been typically applied to local case studies that cover relatively small areas and short periods of time (documented in two publicly available datasets: the SES Library and the SES Meta Analysis Database) [5]. Over a hundred case studies have been coded in these databases. But the scale at which they are coded makes it difficult to extrapolate their lessons to arenas relevant for policy making, or to compare them to better understand what interventions work and where [5].

In order to target development interventions, we need to be able to characterize SES at larger spatial scales and longer time horizons in data-scarce places [5, 8]. The purpose of this paper is developing a data-driven approach to upscale Ostrom's SES framework and identify a typology of SES. It is targeted to countries where available data is restricted in quality and monitoring programs may not be in place. As a working example, we studied the Volta River basin, a cross national watershed that covers roughly two thirds of Burkina Faso's and Ghana's joint territories (407 000 Km^2) [9]. The West African Sahel, where the headwaters of the Volta are located, is a highly vulnerable area where a majority of the population suffers from multi-dimensional poverty [9, 10], where rainfaill is highly variable with recurrent droughts and dry-spells [9], recurrent human and cattle diseases [9], and where population growth is leading to growing food demand [9, 11–13]. The region offers a sharp gradient in climate as well as in economic development, from the relatively rich urbanized areas in southern Ghana to northern Burkina Faso, where smallholder farming and pastoralism dominates [9, 11]. The following section outlines how we operationalize the Ostrom's SES framework to the scale of the Volta River basin and apply it using publicly available data and national statistics. Next we describe the SES archetypes found, how they change over time, and how water reservoir development explain some trends. Finally, we discuss the overall results and the applicability of our methods to other data-scarce contexts.

Clustering SES

Identifying SES archetypes from data is in essence a clustering problem, that is, a classification task of multiple elements by some measure of similarity. Identifying archetypes or systems' typologies is useful because it allows comparison between different cases with similar profiles, they reduce dimensionality, and facilitate extrapolation between cases with similar characteristics [14, 15]. Numerous methods exist to perform clustering, but before explaining the details of our choices, first we present a brief overview of what others have done when classifying SES and how our work improves previous efforts.

The idea that SES are intertwined and interdependent systems is not new: SES are human and natural coupled systems where people interact with biophysical components; they often exhibit nonlinear dynamics, reciprocal feedback loops, time lags, heterogeneity and resilience [16]. It has been suggested that complex adaptive systems, such as SES, should leave statistical signatures on social and ecological data that would allow pattern identification of typologies and make it possible to follow their spatial patterns as well as trajectories through time [17, 18]. Earlier efforts to map SES have been more general in purpose, and global in scale, such as the attempt to identify Anthromes ('human biomes') [19, 20], or general land system archetypes [21–24]. Reflecting on global consequences of land use, Foley *et al* [25] proposed a conceptual framework for bundles of ecosystem services, the idea that landscape units can be classified by the sets of goods and services that a SES co-produces, or more generally, a set of social-ecological interactions. This framework has gained empirical support [26] with studies that range from the watershed to national scales in Canada [27, 28], Sweden [29, 30], Germany [31], and South Africa [32], China [33], Alaska [34], New York city [35], and Andalusia [36]. Similar ordination methods have also been used to study regime shifts from foraging to farming societies in ancient SES [37], and early efforts on mapping SES from a vulnerability framework [38, 39].

Despite the differences in purpose, scale, resolution and datasets used, what the aforementioned studies have in common is that they attempt to map SES by combining multivariate methods of ordination and clustering algorithms to identify (i) systems' typologies and (ii) potential underlying variables of change—what explains the variability of the typologies. However, the studies do not provide guidance on how to make choices regarding optimal number of clusters or algorithmic selection (with exception of [36, 38, 39]), limiting their replicability when applied to different places or data. These choices remain idiosyncratic in the SES literature [5, 40], they depend on tacit knowledge of the researchers, and hence there has been no way of assessing best practices. Here we apply clustering techniques to SES data while explicitly addressing these limitations. To test the optimal number of clusters, 30 different indexes were compared following the sensitivity analysis described by Charrad *et al* [41], while testing the internal and stability validation of 9 different clustering techniques [42] (see methods). Each index is defined as an objective function (e.g. Silhouette, Duda, Dunn) that is maximized or minimized over the number of clusters to test. To guide us in choice of variables, we use Ostrom's SES framework [7]. We obtained publicly available datasets covering the second administrative level for Ghana and Burkina Faso (districts and provinces respectively, $N = 99$). From these, we matched any data that in a meaningful way could be used as proxies of the Ostrom variables.

Data and choice of variables

The variables and data to be used in the analysis was selected through a systematic process. First, a general review of the existing scientific and gray literature covering parts or all of the Volta basin was done using search terms such as 'food security', 'agriculture', 'farming practices' and 'development'. This generated a list of variables that, according to the literature, are important for understanding agricultural practices and increasing food availability in the study area. Second, we reviewed the available data from Burkina Faso and Ghana, including reports from national agencies and international databases. Third, both of

Table 1. Summary variables used and their equivalence with the Ostrom's SES framework.

1st tier	2nd tier	Indicators	Comments
Socio-economic and political settings (S)	S2-demographic trends	Population trend	Change in population density (most recent census / previous census)
		Inter regional migration	% people registered in a province or district who were born in different region of registration
		Intra regional migration	% people registered in a province or district who were born in another district within the same region
	S5-market incentives	Market access	Median of market access index
Resource system (RS)	RS4-human constructed facilities	Dams	One of the most common agricultural innovations in the area as insurance against drought, water source for cattle and irrigation source for crops
	RS7-predictability of system's dynamics	Variance of production	Measured in kcals, is a proxy of how stochastic crop production is related to food availability
Resource units (RU)	RU5-number of units	Cattle per Km^2	Cattle is an insurance for farmers in the area and also related to resource mobility (RU1)
		Small ruminants per capita	Small ruminants are also source of insurance
Users (U)	U1-number of users	Population density	Measured in persons per Km^2
		Ratio of farmers	% people of adult population whose main occupation is agriculture
	U2-socio-economic attributes	Ratio of children	% children age 0–14 out of total population
		Ratio of woman	% women out of total population
		Literacy	% of adult population who are literate
Related ecosystems (ECO)	ECO1-climate patterns	Aridity	Mean aridity gradient
		Mean temperature	Annual mean temperature in °C
	ECO3-flows	Soil water	Median soil water holding capacity in mm (based on soil data)
		Wet season	Number of months with precipitation >60 mm
		Slope 75	75th percentile slope in province/district
Interactions (I)	I1-harvesting levels	Kilocalories for diverse crops	Cowpea, maize, millet, rice, sorghum, soy and yam

these lists were cross-referenced with the second tier variables from the Ostrom framework. This was to identify both context-specific and generally relevant social-ecological variables that we could access data for. Finally, to achieve a good balance, the resulting list of variables and data was reviewed and correlations checked. This led to some variables being excluded due to overlaps in the information they provided (mainly done for basic demographics and climate data). This resulted in a list of 19 variables in table 1.

Since our analysis focuses on food availability we used crop production as the defining key interaction (I) of our SES characterization. It is a proxy that reflects both the capacity of the ecosystem to provide food, and also the human input (e.g. fertilizers, labor) and preferences (e.g. crop choices) necessary to co-produce the service. Crop data, both production and cultivated area, were obtained from national statistical agencies. While data does exist for 32 crops from 1993 to 2012, there are large data gaps in the time series of these data (~36%) (figure S1 available online at stacks.iop.org/ERL/15/034017/mmedia). We thus chose 7 crops with minimum missing values (3.85%), and used averages based on the last 7 years to correct for outliers (see SI for crop selection). Users (U) and their social, economic and political settings (S) were here characterized by national census statistics and their change between the years 1996 and 2006 for Burkina Faso and between 2000 and 2010 for Ghana. The ecological system (ECO) is characterized by biophysical variables from Mitchell *et al* [43] that summarizes aridity, mean temperature, precipitation and slope. Due to the short crop time series (7 years), we did not include long term variability in climate. While important, we do not expect it to have a strong statistical signature in such a short time window. The RS is a combination of variables that facilitate or inhibit crop production (our key interaction), such as the presence of water reservoirs, and the variance of food energy (kcals) produced as a proxy of predictability of system's dynamics. Predictability is found to be associated with the capacity of self-organization and the emergence of managerial rules [7]. RU were characterized by cattle per capita, since this is a source of insurance for farmers in the area [44]. All data is normalized to the range 0:1 and log transformed for distributions with heavy tails (figure S2 available online at stacks.iop.org/ERL/15/034017/mmedia). Table 1 summarizes the framework and the proxies used.

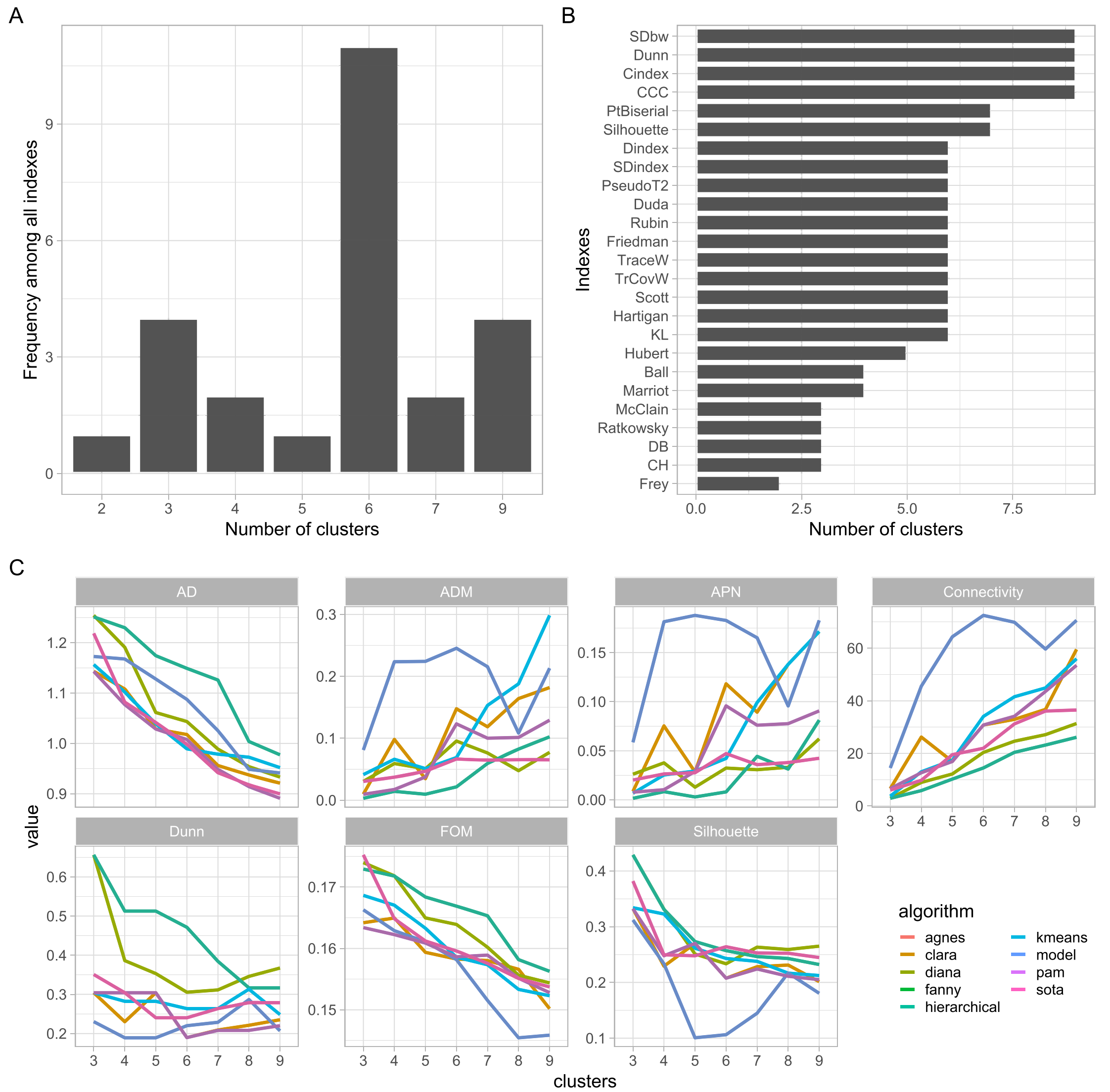

Figure 1. Algorithm selection and number of clusters. A sensitivity analysis to choose adequate number of groups shows that six clusters is the partition favored (A) by 11 out of 30 indexes tested (B). Hierarchical clustering is the best performing algorithm as evaluated through seven measures of performance. For mathematical formulae and indication on how to read each index or performance measure, we refer the reader to [28, 29].

Results

The Volta river basin is composed of distinct sets of SESs. The clustering search (figure 1(A)) identifies an optimal number of six archetypes suggested by 11 out of 30 indexes, followed by three clusters (four indexes), and nine clusters (four indexes). The test for internal and stability validation of nine different clustering algorithms [42] suggests that hierarchical clustering is the best performing technique (figure 1(C)). Henceforth subsequent results use hierarchical clustering and six clusters to characterize SES in the Volta basin. Figure S3 further compares the cluster partitions of the basin with the top performing algorithms, showing a relatively good agreement between the different techniques and a strong north–south gradient regardless of the algorithm used.

Following the analogy of bundles of ecosystem services [25, 27], we also map the sets of SES variables that covary in space using the archetypes found by the clustering analysis (figure 2). SES archetypes follow a north–south gradient, with SES in the north (cluster 1 and 2) characterized by arid environments. In these clusters, the kilocalories produced come primarily from cowpea, millet and sorghum, with relatively higher values from rice and maize in southern Burkina Faso (cluster 2). Though clusters 1 and 2 are similar, cluster 2 has in average higher cattle per km^2, both higher intra- and inter- regional migration, higher literacy rates, faster population growth, but lower ratio of women and children (figures S4–S5). Cluster 3 concentrates higher kilocalorie production with relatively high production of all crops analyzed except for maize. Maize reaches its production peak in the south—in clusters 4–5—also characterized by the production of yam and rice, lower cattle per km^2, and fewer small ruminants per capita. The highest urbanization and literacy occurs in clusters 4 and 5. External migration

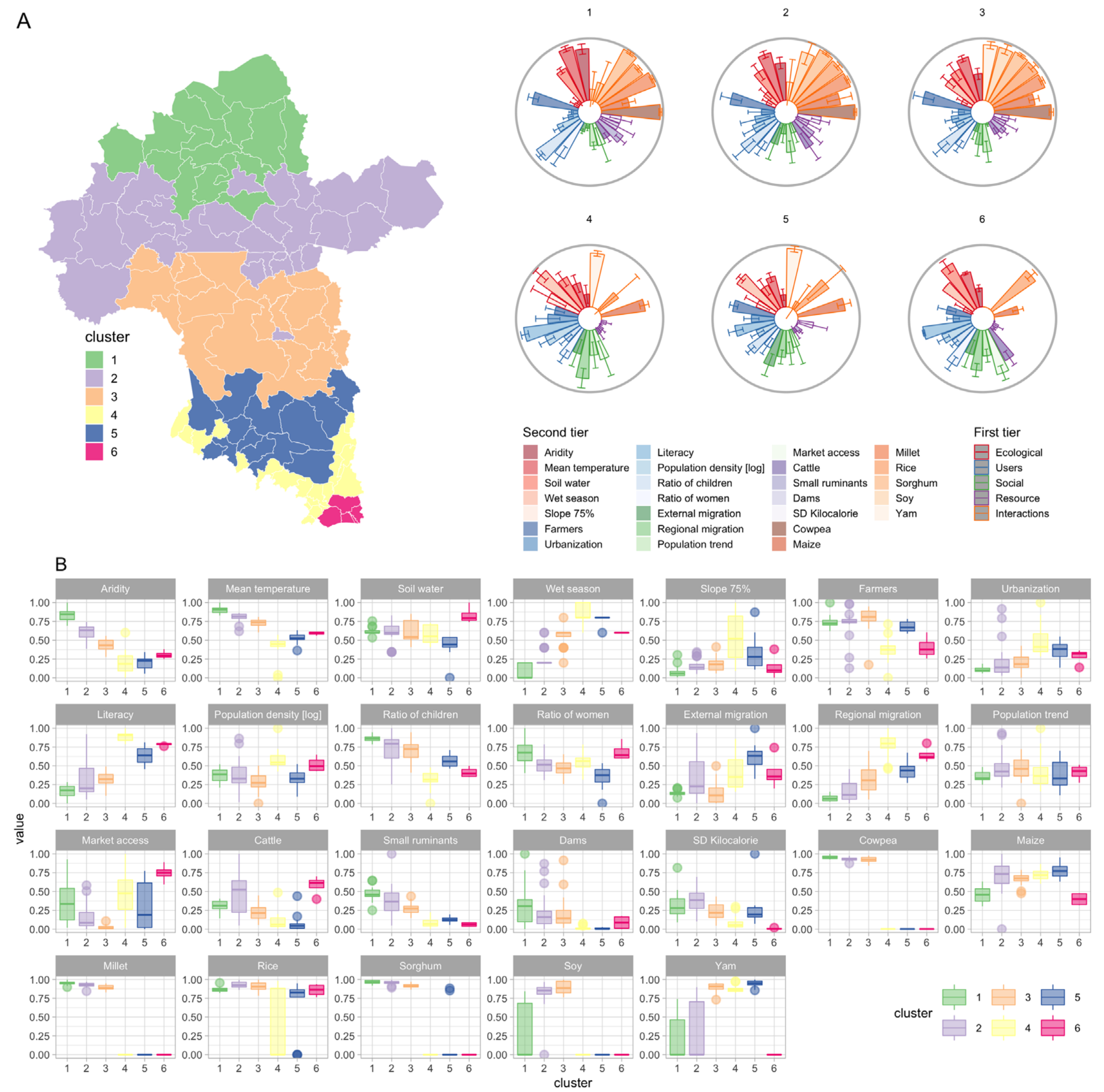

Figure 2. Bundles of SES variables. The SES framework was operationalized by analyzing datasets that are used as proxies for Ostrom's suggested variables at spatial units that correspond to the second administrative level (A). The clusters correspond to the best partition obtained with Hierarchical clustering (figure S3). For each cluster there is a flower diagram of average normalized units for each second tier variable. The outer gray circle shows the unit circle, if the mean is equal to one the bar would be as high as the circle's perimeter. The base of the bars shows the zero values. The height of the bar exhibits the variable mean for each cluster, while the error bars show standard deviations. Note that for some variables the error bars go above 1 or below zero, indicating that the mean is not necessarily the best summary statistic for all variables. (B) Shows the distribution of each variable and allows for cluster comparison within variable; figure S4 complements this figure by showing the intra cluster variation while figure S5 shows the statistical differences between clusters per variable.

is low in cluster 3 but increases again further south in clusters 4–6. Cattle per km^2 and market access reach their highest in cluster 6. The south is also dominated by longer wet seasons and higher soil water content. Note that while water reservoir density is higher in cluster 1, the variance of kilocalorie (our proxy for predictability) is high in clusters 1, 2, 3 and 5. Figures S4 and S5 further explore significant differences between clusters for each second tier variable.

The potential relationships between Ostrom's framework components were further investigated by applying vector fitting to non-metric multi-dimensional scaling (figure 3). We use this approach to better understand what explains the variability of the archetypes found. The ordination method applied to each set of variables reveals that the clustering (figure 2(B)) is highly driven by the interactions (crops) category (figure 3 top row). Clusters 1–2 tend to produce similar crops and rely heavily on cattle and small ruminants; they are also where dams are more abundant. A similar ordination on the ecological variables (ECO) supports the idea that water reservoirs and cattle have been highly correlated to places with high aridity, and occur in places where the crop portfolio is characterized by cowpea and sorghum (cluster 1), or rice and soy (clusters 2–3). The predictability of the RS, measured as the variance of kilocalorie production over 7 years of data, shows that clusters 2–3 are the areas where crop productivity is more unpredictable, with potential implications for food availability. They are also areas with

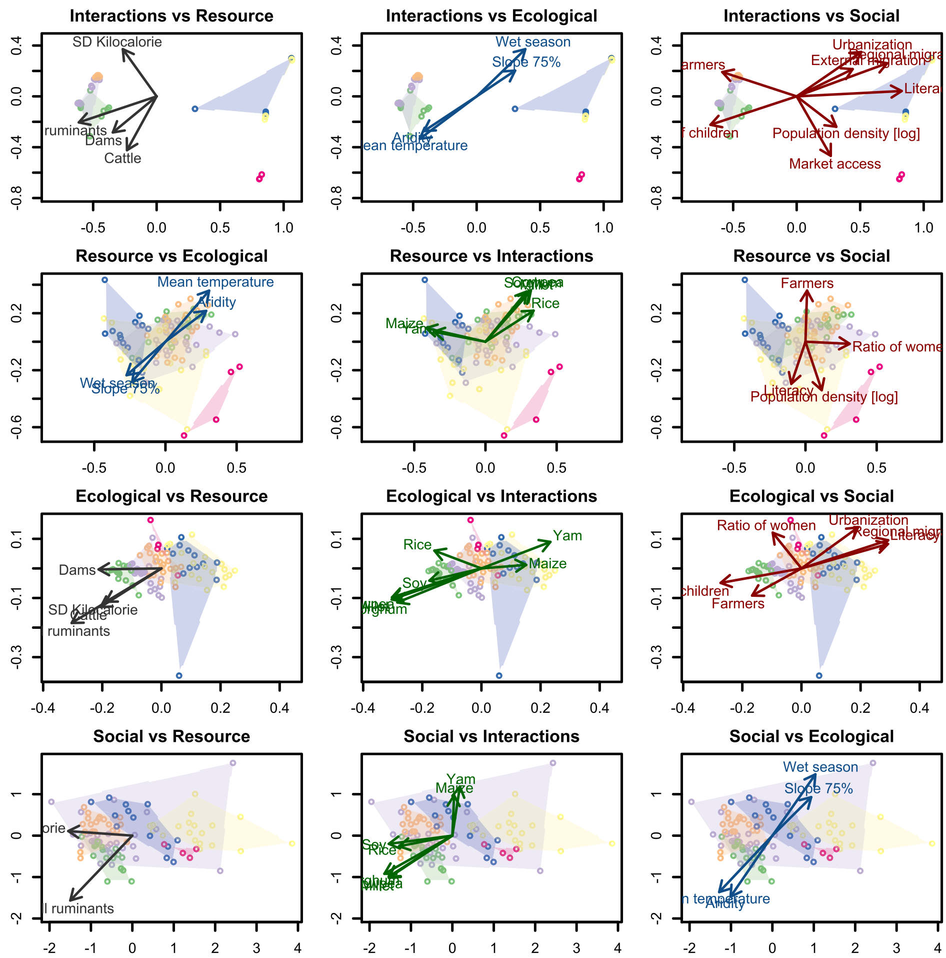

Figure 3. Relationships between SES framework components. Each subsystem in the Ostrom's framework is ordered with non-metric multi-dimensional scaling and vectors are fitted for all other variables that significanlty ($p < 0.001$) explain the variation of the ordination. Each plot title shows the dataset in which the ordination was applied versus the dataset used for the vector fitting. The colors of points and contours correspond to the archetypes found in figure 1.

higher densities of farmers and children. These relationships across the different components of the Ostrom's framework holds when looking at years where complete crop data is available (figure S6, also see figure S1 for data gaps). It suggests that the results are robust across time despite limited longitudinal data. Our results also agree with previous characterizations of the Volta basin based solely on crop data from 1990s [9], regional agricultural characterizations of the Sub-Saharan Africa [39], and finer socio-economic survey data in two of the districs we studied [12].

Discussion

This paper outlines a data-driven routine for operationalising Ostrom's SES framework and mapping SES archetypes. The method can be executed using exclusively publicly available data and open access software [41, 42], making suitable for replication in other data limited settings beyond our test case in the Volta river basin. By applying clustering with a sensitivity analysis routine to the Volta river basin case, we have shown how the method performs in a setting with restricted data quality and still renders useful insights. We have found that the Volta basin can be best described by six SES archetypes strongly characterized by their crop productivity profiles but also by social variables such as urbanization, literacy, and migration.

Our results also suggest that the construction of water reservoirs can improve food availability in places in the basin historically exposed to high variability in food supply such as northern parts of Burkina Faso (cluster 1, figure 3). In addition, our analysis suggests that additional reservoirs would likely have a strong impact on the SES represented by clusters 2 and 3, because in these regions social and ecological conditions are similar to cluster 1 and food production predictability is low (based on a high standard deviation of kilocalorie production). Ostrom found that predictability on resource dynamics enables the capacity to self-organize and the emergence of rules for managing resources [7]. The clusters identified help us understand where these two conditions meet: high crop variability and potential for water reservoirs to buffer variation allowing people to better manage their resources. Our results allow us to identify patterns of SES where the abundance of reservoirs does correlate with production of certain crops. Based on this pattern we can speculate where an additional dam is likely to have a similar effect, or more importantly, where an additional reservoir is not likely to have an impact on food production. These patterns are relevant for SES theory development because it help us understand what works and where [40], or to circumvent the problem of context dependence that Ostrom describes. However, these results cannot be interpreted as causal effects. The public datasets used do not provide extensive time series, instrumental variables, or randomized control trials required to test for causality. In addition, other factors that are expected to play a role on the production of crops are not controlled for due to lack of data at the appropriate spatial scale, such as the use of fertilizers and pesticides. Further tests for causality require longitudinal data on water reservoirs (when were they built, water storage capacity, location), complete longitudinal data on water and irrigation intensive crops such as vegetables, and proxies for fertilizers and pesticides use at the second-level administrative level. Our results only allow us to approximate food production from the perspective of average municipalities or districts. Fine-scale attributes of individuals or households such as access to food or reliance are not observable from the scale of our study.

Our work extends previous efforts for mapping SES [27–32, 38, 39, 45] in that it considers a broader range of both social and ecological variables outlined by Ostrom. Our work contributes, to our knowledge, one of the few attempts to upscale Ostrom's framework to a multinational scale that matches the scale of the resource flow dynamics: the basin (see [46]). The unit of analysis was second-level administrative units ($N = 99$ in the Volta basin) that clustered into six SES archetypes. Previous efforts have relied heavily in qualitative data, which restricts the analysis to smaller sample sizes (i.e. see work by [45], $N = 12$; or [46], $N = 26$). While our approach looses some of the richness provided by case studies, it allows us to take advantage of data over a larger spatial scale and few observations across time to draw comparisons among diverse places.

Our study also contributes to a growing literature characterizing systems archetypes in SES [14, 15, 40]. The concept of archetype, first introduced by Forrester [47], refers to canonical structures or causal building blocks to many dynamical systems and their managerial problems. Forrester believed that there should be at most 20 archetypes that capture 90% of policy issues that most managers encounter [47]. Typically depicted as causal diagrams, archetypes represent recurrent feedback structures reduced to 4 problem archetypes and their solutions [48–50]. These building blocks have inspired the identification of resilience surrogates in SES [51], shared causes [52] and potential interactions of regime shifts in SES [53], as well as archetypical configurations of land systems' change [54]. The structural building blocks approach to archetypes is complemented by its empirical counterpart [55], focused on the identification of common empirical typologies that can result from recurrent causal structures. Our archetype mapping falls within this second tradition, similar to previous data-driven efforts on land use archetypes [21–24] and anthromes [19, 20]. Our contribution differs, however, from previous land use studies in including a broader set of social-ecological variables inspired by Ostroms SES framework but at the cost of resolution. Data driven identification of land use archetypes typically used raster data at 1 arc degree or 1 Km^2 grids [22, 24]. Our coarser resolution enabled us to include a larger variety of social-ecological variables at a scale where management strategies are usually decided. Our data-driven approach would benefit, however, of the integration of a building blocks perspective by including relevant processes such as large-scale land acquisitions, public policies, trade, or certification schemes, as shown in other structure oriented studies [54, 56].

Our approach is limited by data availability. For example, we have not included any variable in the Ostrom SES framework that describes the governance of the system (G), or the use of fertilizers or tractors (RS). Although indicators of governance and agricultural modernization do exist at the national scale (e.g. governance indicators, World Bank database, only available for Ghana; fertilizers FAO database), they cannot provide insight at the scale of this study, limiting our conclusions. This suggests that including governance indicators such as how often people share food, what is the structure of the social networks, or efficiency of local institutions at managing existing water infrastructure in national monitoring programs such as census or national surveys would markedly improve the ability to characterize SES. If and when these type of data become available for the Volta basin or elsewhere, they can easily be incorporated into the SES analysis here proposed.

Our work also improves replicability and reproducibility compared to previous efforts of mapping SES [27–32]. Previous work has relied on only one clustering technique and strong context dependent knowledge to make subjective decisions about the number of clusters to fit and the clustering technique to apply. While this is a valid approach, it limits scalability and reproducibility because choices made for one place may not be appropriate in other places. Here we have used an updated routine with a sensibility analysis that helps the researcher to make such choices guided by the patterns already contained in the data. While this type of machine learning approach can never replace the richness of local knowledge, it does facilitates the practical application of the method in absence of in-depth qualitative data (e.g. lack of coverage), and in settings where field work validation is restricted (e.g. war zones). This approach can thus complement and guide where qualitative research efforts could be most effectively deployed. In addition, though we cannot claim causality, the patterns here presented can be useful for policy making or identifying priority areas for future investments. Central to sustainability science problems is distinguishing where and when solutions works and are transferable [7]. Here we applied an approach to mapping SES that can help distinguishing context dependent from generalizable solutions in data-scarce contexts.

Conclusion

Advancing theories on sustainability science requires articulating existing SES frameworks to generalizable and replicable analysis of large-scale systems. Achieving the sustainable development goals depends on distinguishing where a sustainable solution is context dependent or where it could be generalized to different SES arenas. Here we have advanced an approach to identify SES by updating clustering routines with a sensitivity analysis that allows us to reinterpret a binational dataset in the Volta basin. We identified where and under which conditions an agricultural innovation such as water reservoirs is likely to influence the food security in one of the most arid areas of the world. These patterns can inform policy decisions. Identifying patterns of variables in space and time that characterize different SES is key for further developing theories of sustainability, testing when interventions work, and mapping how nations progress towards sustainable development goals. The methods here outlined are generalizable to other developing country settings, and we hope they will help rigorously test under which conditions the Ostrom's SES framework can have policy relevant implications.

Methods

The optimal number of clusters was tested by comparing 30 different indexes and their performance following a sensitivity analysis. In general, each index requires maximization or minimization of some measure, that helps the researcher decide the optimal number of clusters in the data. A description and interpretation of each index can be found in Charrad *et al* [41]. We further test the internal validation and stability validation of 9 different clustering techniques: hierarchical clustering, self-organizing maps, k-means, partitioning around medoids *pam*, divisible hierarchical algorithm *diana*, a sampling based clustering *clara*, a fuzzy clustering *fanny*, self-organizing trees *sota*, and model-based algorithm [42]. While the first technique offers a robust estimation of the number of clusters in the data, the second helps choosing an optimal clustering algorithm. Results are typically presented as non-metric multi-dimensional ordinations and their spatial distribution in maps of the Volta basin. Alternatively, principal component ordination results are shown in SI. The maximum dissimilarity distance was used to maximize the distance between components, while the Ward aggregation method was used to minimize the total within-cluster variance [41]. For visualizations we used a less restrictive Manhattan distance to ensure convergence. We further investigate the interdependences between Ostrom's nested variables, by reiterating the ordination on a set of variables of interest (e.g. interactions [I]) and performing vector fitting with the remaining variable sets (e.g. resource (RU and RS), social (U and S), ecological (ECO)) (figure 2). The same procedure was performed over time for years with complete data to test how robust are our estimates over time (figure S6).

Acknowledgments

We appreciate feedback on previous versions of the manuscript by Megan Meacham, Ingo Fetzer, and Miina Porkka. Ilja Lange and Esther Klingenberg provided assistance with data cleaning and literature review. LG and FD contributed the initial idea; all authors designed the research, KM cleaned the data, JR wrote the code and ran the analysis with guidance of KM, LG, KB and FD. All authors discussed the results and wrote the paper. The authors declare no conflict of interest. All authors were supported by grant

Data availability

The data that support the finding of this study are openly available at DOI:10.6084/m9.figshare.6709247 and code at https://github.com/juanrocha/TAI-Volta.

ORCID iDs

Juan Rocha https://orcid.org/0000-0003-2322-5459
Katja Malmborg https://orcid.org/0000-0003-1660-4514
Line Gordon https://orcid.org/0000-0002-3520-4340
Kate Brauman https://orcid.org/0000-0002-8099-285X
Fabrice DeClerck https://orcid.org/0000-0002-3631-8745

References

[1] Assembly U 2015 *Transforming Our World: The 2030 Agenda for Sustainable Development (New York, United Nations)* (http://scholar.google.com/scholar?q=related:ojKXWYYcKVYJ:scholar.google.com/&hl=en&num=20&as_sdt=0,5)
[2] Lubchenco J, Barner A K, Cerny-Chipman E B and Reimer J N 2015 Sustainability rooted in science *Nat. Geosci.* **8** 741–5
[3] Griggs D *et al* 2013 Policy: sustainable development goals for people and planet *Nature* **495** 305–7
[4] Reyers B, Stafford Smith M, Erb K-H, Scholes R J and Selomane O 2017 Essential variables help to focus sustainable development goals monitoring *Curr. Opin. Environ. Sust.* **26–27** 97–105
[5] Partelow S 2018 A review of the social-ecological systems framework: applications, methods, modifications, and challenges *Ecology and Society* **23** 36
[6] Ostrom E 2007 A diagnostic approach for going beyond panaceas *Proc. Natl Acad. Sci. USA* **104** 15181–7

[7] Ostrom E 2009 A general framework for analyzing sustainability of social-ecological systems *Science* **325** 419–22
[8] Cox M 2014 Understanding large social-ecological systems: introducing the SESMAD project *Int. J. Commons* **8** 265–76
[9] Lemoalle J and de Condappa D 2010 Farming systems and food production in the Volta Basin *Water Int.* **35** 655–80
[10] Malik K 2014 *Human Development Report 2014* United Nations Development Program (www.undp.org) (https://doi.org/978-92-1-056659-9)
[11] Institut National de la Statistique et de la Demographie 2009 *Redensement general de la population et de l'habitation (RGPH) de 2006. Analyse des resultats definitifs* (Burkina Faso: Ministere de Economie et des finances)
[12] Douxchamps S *et al* 2015 Linking agricultural adaptation strategies, food security and vulnerability: evidence from West Africa *Reg. Environ. Change* **16** 1305–17
[13] Ofosu E A, van der Zaag P, van de Giesen N C and Odai S N 2010 Productivity of irrigation technologies in the White Volta basin *J. Phys. Chem. EAR* **35** 706–16
[14] Eisenack K *et al* 2019 Design and quality criteria for archetype analysis *Ecol. Soc.* **24** 6
[15] Oberlack C *et al* 2019 Archetype analysis in sustainability research: meanings, motivations, and evidence-based policy making *Ecol. Soc.* **24** 26
[16] Liu J *et al* 2007 Complexity of coupled human and natural systems *Science* **317** 1513–6
[17] Holland J H 2012 *Signals and Boundaries* (Cambridge, MA: MIT Press)
[18] Levin S A 2000 *Fragile Dominion* (Cambridge, MA: Perseus Publishing)
[19] Ellis E C, Goldewijk K K, Siebert S, Lightman D and Ramankutty N 2010 Anthropogenic transformation of the biomes, 1700 to 2000 *Glob. Ecol. Biogeogr.* **19** 589–606
[20] Ellis Ramankutty 2008 Putting people in the map: anthropogenic biomes of the world *Front. Ecol.* **6** 523–523
[21] Surendran Nair S, Preston B L, King A W and Mei R 2016 Using landscape typologies to model socioecological systems: application to agriculture of the United States Gulf Coast *Environ. Modelling Softw.* **79** 85–95
[22] Václavı'k T, Lautenbach S, Kuemmerle T and Seppelt R 2013 Mapping global land system archetypes *Glob. Environ. Change* **23** 1637–47
[23] Ropero R F, Aguilera P A and Rumı' R 2015 Analysis of the socioecological structure and dynamics of the territory using a hybrid Bayesian network classifier *Ecol. Modelling* **311** 73–87
[24] Levers C *et al* 2015 Archetypical patterns and trajectories of land systems in Europe *Reg. Environ. Change* **40** 214
[25] Foley J A *et al* 2005 Global consequences of land use *Science* **309** 570–4
[26] Spake R, Lasseur R, Crouzat E and Bullockl J 2017 Unpacking ecosystem service bundles: Towards predictive mapping of synergies and trade-offs between ecosystem services *Global Environmental Change* **47** 37–50
[27] Raudsepp-Hearne C, Peterson G D and Bennett E M 2010 Ecosystem service bundles for analyzing tradeoffs in diverse landscapes *Proc. Natl Acad. Sci. USA* **107** 5242–7
[28] Renard D, Rhemtulla J M and Bennett E M 2015 Historical dynamics in ecosystem service bundles *Proc. Natl Acad. Sci. USA* **112** 201502565
[29] Meacham M, Queiroz C, Norström A and Peterson G 2016 Social-ecological drivers of multiple ecosystem services: What variables explain patterns of ecosystem services across the Norrström drainage basin? *Ecology and Society* **21** art14
[30] Queiroz C *et al* 2015 Mapping bundles of ecosystem services reveals distinct types of multifunctionality within a Swedish landscape *Ambio* **44** 89–101
[31] Rabe S-E, Koellner T, Marzelli S, Schumacher P and Grêt-Regamey A 2016 National ecosystem services mapping at multiple scales The German exemplar *Ecol. Indicators* **70** 357–72
[32] Hamann M, Biggs R and Reyers B 2015 Mapping socialecological systems: identifying 'green-loop' and 'red-loop' dynamics based on characteristic bundles of ecosystem service use *Glob. Environ. Change* **34** 218–26
[33] Su C *et al* 2011 Ecosystem management based on ecosystem services and human activities: a case study in the Yanhe watershed *Sust. Sci.* **7** 17–32
[34] Alessa L N, Kliskey A A and Brown G 2008 Socialecological hotspots mapping: a spatial approach for identifying coupled socialecological space *Landscape Urban Plan.* **85** 27–39
[35] McPhearson T, Kremer P and Hamstead Z A 2013 Mapping ecosystem services in New York City: applying a socialecological approach in urban vacant land *Ecosyst. Serv.* **5** 11–26
[36] Martín-López B *et al* 2017 Delineating boundaries of social-ecological systems for landscape planning: a comprehensive spatial approach *Land Use Policy* **66** 90–104
[37] Ullah I I T, Kuijt I and Freeman J 2015 Toward a theory of punctuated subsistence change *Proc. Natl Acad. Sci. USA* **112** 9579–84
[38] Kok M *et al* 2015 A new method for analysing socio-ecological patterns of vulnerability *Reg. Environ. Change* **16** 229–43
[39] Sietz D, Ordoñez J C, Kok M and Janssen P 2017 Nested archetypes of vulnerability in African drylands: where lies potential for sustainable agricultural intensification? *Environ. Res. Lett.* **12** 095006
[40] Sietz D *et al* 2019 Archetype analysis in sustainability research: methodological portfolio and analytical frontiers *Ecology and Society* **24** 34
[41] Charrad M, Ghazzali N, Boiteau V and Niknafs A 2014 NbClust: an RPackage for determining the relevant number of clusters in a data set *J. Stat. Softw.* 61 (http://jstatsoft.org/v61/i06/)
[42] Brock G, Pihur V, Datta S and Datta S 2008 clValid: an R package for Cluster validation *J. Stat. Softw.* **25** 1–22
[43] Mitchell T D and Jones P D 2005 An improved method of constructing a database of monthly climate observations and associated high-resolution grids *Int. J. Climatol.* **25** 693–712
[44] Frelat R *et al* 2016 Drivers of household food availability in sub-Saharan Africa based on big data from small farms *Proc. Natl Acad. Sci. USA* **113** 458–63
[45] Leslie H M *et al* 2015 Operationalizing the social-ecological systems framework to assess sustainability *Proc. Natl Acad. Sci. USA* **112** 5979–84
[46] Oberlack C and Eisenack K 2018 Archetypical barriers to adapting water governance in river basins to climate change *J. Inst. Econ.* **14** 527–55
[47] Lane D 1998 Can we have confidence in generic structures? *J. Oper. Res. Soc.* **49** 936–47
[48] Wolstenholme E F and Wolstenholme E 2003 Towards the definition and use of a core set of archetypal structures in system dynamics *Syst. Dyn. Rev.* **19** 7–26
[49] Meadows D H and Wright D 2009 *Thinking in Systems: A Primer* (White River Junction, VT: Chelsea Green Publishing)
[50] Senge P M 1990 *The Fifth Discipline: The Art and Practice of the Learning Organization* (New York: Doubleday)
[51] Bennett E M, Cumming G S and Peterson G D 2005 A systems model approach to determining resilience surrogates for case studies *Ecosystems* **8** 945–57
[52] Rocha J C, Peterson G D and Biggs R 2015 Regime shifts in the anthropocene: drivers, risks, and resilience *PLoS One* **10** e0134639
[53] Rocha J C, Peterson G, Bodin O and Levin S 2018 Cascading regime shifts within and across scales *Science* **362** 1379–83
[54] Rocha J *et al* 2019 Toward understanding the dynamics of land change in Latin America: potential utility of a resilience approach for building archetypes of land-systems change *Ecol. Soc.* **24** art17
[55] Stone E and Cutler A 1996 Introduction to archetypal analysis of spatio-temporal dynamics *Physica* D **96** 110–31
[56] Oberlack C, Tejada L, Messerli P, Rist S and Giger M 2016 Sustainable livelihoods in the global land rush? Archetypes of livelihood vulnerability and sustainability potentials *Glob. Environ. Change* **41** 153–71