Experimental Agriculture (2022), vol. 58, e27, 1–17
doi:10.1017/S0014479722000126

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Data-driven similar response units for agricultural technology targeting: An example from Ethiopia

Lulseged Tamene[1,*], Wuletawu Abera[1], Eduardo Bendito[2], Teklu Erkossa[3], Meklit Tariku[2], Habtamu Sewnet[4], Degefie Tibebe[5], Jemal Sied[6,10], Gudina Feyisa[7], Menale Wondie[8] and Kindie Tesfaye[9]

[1]International Center for Tropical Agriculture (CIAT), P.O. Box 5689, Addis Ababa, Ethiopia, [2]International Institute of Tropical Agriculture (IITA), P.O. Box 30772-00100, Nairobi, Kenya, [3]Deutsche Gesellschaft für Internationale Zusammenarbeit (GIZ), P.O. Box 100009, Addis Ababa, Ethiopia, [4]Geospatial Information Institute, P.O. Box 597, Addis Ababa, Ethiopia, [5]Water and Land Resource Center, Addis Ababa University, P.O. Box 3880, Addis Ababa, Ethiopia, [6]Ethiopian Institute of Agricultural Research, P.O. Box 2003, Addis Ababa, Ethiopia, [7]Addis Ababa University, Institute of Natural Sciences, P.O. Box 1176, Addis Ababa, Ethiopia, [8]Amhara Agricultural Research Institute (ARARI), P.O. Box 527, Bahir Dar, Ethiopia, [9]International Maize and Wheat Improvement Center (CIMMYT), P.O. Box 5689, Addis Ababa, Ethiopia and [10]Scuola Superiore Sant'Anna, Pisa, Italy
*Corresponding author. Email: lt.desta@cgiar.org

(Received 09 August 2021; revised 13 February 2022; accepted 05 March 2022)

Abstract
Ethiopia has heterogeneous topographic, climatic and socio-ecological systems. Recommendations of agricultural inputs and management practices based on coarse domains such as agro-ecological zones (AEZ) may not lead to accurate targeting, mainly due to large intra-zone variations. The lack of well-targeted recommendations may contribute to the underperformance of promising technologies. Therefore, there is a need to define units where similar environmental and biophysical features prevail, based on which specific recommendations can be made for similar response units (SRUs). We used unsupervised machine learning algorithms to identify areas of high similarity or homogeneous zones called 'SRUs' that can guide the targeting of agricultural technologies. SRUs are landscape entities defined by integrating relevant environmental covariates with the intention to identify areas of similar responses. Using environmental spatial data layers such as edaphic and ecological variables for delineation of the SRUs, we applied K- and X-means clustering techniques to generate various granular levels of zonation and define areas of high similarity. The results of the clustering were validated through expert consultation and by comparison with an existing operational AEZ map of Ethiopia. We also augmented validation of the heterogeneity of the SRUs by using field-based crop response to fertiliser application experimental data. The expert consultation highlighted that the SRUs can provide improved clustering of areas of high similarity for targeting interventions. Comparison with the AEZ map indicated that SRUs with the same number of AEZ units captured heterogeneity better with less within-cluster variability of the former. In addition, SRUs show lower within-cluster variability to optimal crop response to fertiliser application compared with AEZs with the same number of classes. This implies that the SRUs can be used for refined agricultural input and technology targeting. The work in this study also developed an operational framework that users can deploy to fetch data from the cloud and generate SRUs for their areas of interest.

Keywords: Agriculture; Clustering; Machine learning; Recommendation; Technology targeting

Introduction
The current approximately 110 million population of Ethiopia is projected to reach 180 million by 2050 (Bekele and Lakew, 2014; UNDESA, 2017). This requires an increase in huge tons of

annual cereal and meat production to meet the nutritional and food security needs of the increasing population (FAO, 2017). This situation is worsened by the fact that over 60% of the farming households in the country operate on small plots averaging less than 1 ha (e.g. CSA, 2015; Rahmato, 1994) and are fragmented with significant operational challenges (Zewdie and Tamene, 2020). Because of limited cultivable land for expansion, the efficiency of agricultural production on currently cultivated land should increase substantially to feed the growing population. This should be achieved without compromising ecological integrity and environmental sustainability. To realise this, resources should be targeted in a rational way to the production systems that have the highest potential to achieve the triple wins of poverty reduction, environmental protection and food security (Herrero *et al.*, 2014).

One of the key constraints to enhance food security and increase the overall resilience of smallholders in developing regions is a lack of informed decision making due to a combination of factors such as lack of locally relevant information, conducive institutional structure and policy issues (Aryeetey *et al.*, 2017; Covic and Hendriks, 2016; Holdsworth *et al.*, 2016; Shroff *et al.*, 2015). The major bottleneck that aggravates the impacts of these constraints is the lack of adequate data at the appropriate resolution, desired frequency, required quality and quantity to deploy data-driven knowledge-based decisions (Donatien, 2016). For example, the lack of site-specific information about the topographic settings of farms, the status of soils, weather conditions and the nutrient requirements of crops undermines effective targeting of technologies to areas where they perform better to improve productivity.

As a result of the lack of data-driven and tailored decisions, farmers are provided with blanket recommendations of technologies and management practices despite considerable differences in their farming systems in terms of environmental conditions, landscape positions, soil characteristics, crop diseases, weeds infestation and water availability. This approach ignores the need to match 'farming conditions' with technology requirements and entails the need to shift from a 'one-size-fits-all' strategy to providing tailored recommendations based on location-specific characteristics, constraints and potentials. Therefore, in regions where there are heterogeneous farming systems, classification of sites into uniform units is a crucial step to develop location-specific recommendations and thereby improve agricultural productivity and food security (Penghui *et al.* 2020; Pennock *et al.*, 1994).

Classification of landscapes into relatively homogenous units is done by creating uniform territories with regular and typical occurrence of interrelated combinations of geological composition, landforms, surface and ground waters, microclimates and soil types (Salecker *et al.*, 2019). By matching the specifications of a given development strategy with spatially referenced similar units, it is possible to delineate geographical areas where the strategy is likely to be successful and has a positive impact (Notenbaert *et al.*, 2017). This will not only enable targeting technologies to areas where they perform better but also facilitates scaling options across wide areas (Herrero *et al.*, 2014; Notenbaert *et al.*, 2013). In agriculture, the assumption is that strategies are likely to have similar response in areas that fall within the same recommendation domain. Thus, specific types of development policies, investments and livelihood options, and technologies are likely to result in a desirable effect and be adopted if they are targeted based on a recommendation domain.

The past few decades have seen widespread availability of spatial data and the advancement of robust modelling algorithms. Such developments are creating unique opportunities for optimisation of natural resource management, enhancing economic development and helping alleviate poverty on the basis of recommendation domains (Akıncı *et al.*, 2013; Elsheikh *et al.*, 2013; Freeman *et al.*, 2008; Herrero *et al.*, 2014; Hyman *et al.*, 2013; Notenbaert *et al.*, 2017, 2013; Omamo *et al.*, 2006). Different approaches have been used to create environmental units where similar processes prevail, similar recommendations can be made and similar responses can be expected. In the agricultural sector, there are several efforts to define areas of high similarity. Examples include the development of agro-ecological zones (AEZ) (e.g. FAO, 1981; Fischer and Antonie, 1994; IIASA/FAO, 2012), farming systems (e.g. Amede *et al.*, 2017;

Figure 1. Topography of parts of Ethiopia revealing complexity and diversity.

Dixon *et al.*, 2001; Rizzo *et al.*, 2013), recommendation domains (e.g. Notenbaert *et al.*, 2013; Omamo *et al.*, 2006; Tesfaye *et al.*, 2015) and topographic position (e.g. Amede *et al.*, 2020; Gerçek, 2017; Gerçek *et al.*, 2011). Recently, Muthoni *et al.* (2017) used geospatial analysis and clustering techniques to delineate relatively similar clusters for scaling improved crop varieties and good agronomic practices in Tanzania. In addition, Khoroshev (2020) developed a framework aimed at considering geographical context, matter flows and dynamic processes in developing ecological networks and identifying sites for various land use types as well as for choosing appropriate technologies.

Developing procedures to automatically classify landscapes into spatial entities or clusters can be essential to defining effective management units for precision farming and for scaling site-specific recommendations (Penghui *et al.*, 2020). There are, however, no standard frameworks designed to generate similar response units (SRUs) using big data and machine learning approaches. The aim of this study is, therefore, to develop an operational framework to define 'SRUs' or management zones within which similar technologies and management interventions can be recommended. We outlined a framework and approaches, with reproducible workflow and tool piloted for Ethiopia.

Approach and Methodologies

Study area

The SRU-mapping exercise is piloted in Ethiopia, which is a country with very heterogeneous landscapes (Figure 1), diverse AEZ and different farming systems. The country's elevation ranges from 116 m below sea level to over 4600 m asl and comprises more than 30 AEZs. Generally, AEZs are defined through the combination of temperature, precipitation and elevation parameters focusing on the climatic and edaphic requirements of crops and on the management systems under which the crops are grown (FAO, 1996). Agriculture is the dominant means of livelihoods in Ethiopia supporting over 80% of the population. The Ethiopian highlands (over 1500 m asl) support the majority of the population, and this is also the part of the country where crop production dominates. Still traditional farming dominates and its transformation will be needed for enhancing the quality of life and improving food security. Owing to the country's heterogeneity

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

Table 1. Key landscapes elements/variables used to drive similar response units

Variable/type	Resolution	Source
Topography (elevation, slope, aspect, wetness and topographic indexes)	90 m	https://srtm.csi.cgiar.org/ (Version 4.0)
Landscape (TPI)	270 m	https://developers.google.com/earth-engine/datasets/catalog/CSP_ERGo_1_0_Global_ALOS_mTPI
Landscape (TWI)	~500 m	https://www.hydrosheds.org
Rainfall	~5 km	https://data.chc.ucsb.edu/products/CHIRPS-2.0/ (Version 2.0)
Temperature (mean)	~5 km	https://www.worldclim.org/
Solar radiation, soil moisture, potential evapotranspiration, relative humidity	~4 km	http://www.climatologylab.org/terraclimate.html
Length of growing period	~4 km	Own analysis (derived from ???)
Soil property (sand %, clay %,, SOC, CEC, pH, total nitrogen)	250 m	www.isric.org (Version 2.0)
Vegetation index (EVI, NDVI)	1 km	https://developers.google.com/earth-engine/datasets/catalog/MODIS_006_MOD13A2

and diversity of agro-climatic and farming systems, it will be essential to ensure that appropriate options/technologies are targeted to locations with specific characteristics.

Data and data sources

Integrating key variables is crucial for the zonation of target areas into similar management units. This study uses relevant covariates (Table 1) to develop SRUs that will have similar responses to agricultural technologies such as integrated soil fertility management and climate-smart agriculture. The major factors that determine agricultural systems and productivity are topography, climate, soil and associated derivatives. These are thus the key covariates used to derive SRUs.

Topography and its derivatives (e.g. slope, aspect, wetness index and topographic index) determine landscape processes such as material flow as well as associated agricultural and landscape practices. Topography is the dominant factor in controlling the flow and accumulation of water, energy and matter in Ethiopian landscapes. It also affects the development and properties of soils and off-site environmental conditions (e.g. soil moisture, organic carbon, mineral-forming elements, etc.) in different ways. In this study, the 90 m SRTM digital elevation model was used to derive key terrain variables.

Climatic conditions such as temperature, rainfall, humidity, evaporation and their variabilities have important implications on determining the success of developed plans and interventions. Knowledge of localised climatic conditions (weather) and variability across space and time is thus critical to designing targeting options. The 5 km resolution CHIRIPS data set (Funk *et al.*, 2014) was used to represent rainfall and temperature conditions in the country; average and dekadal data sets were used to capture spatio-temporal and seasonal dynamics. Temperature was derived from 5 km WORLDCLIM data set (Fick and Hijmans, 2017). The 4 km TERRACLIME (Abatzoglou *et al.*, 2018) data were used to represent solar radiation, soil moisture and potential evapotranspiration as input variables for clustering. A 'climate derivative' called the length of the growing period (LGP), which represents overall suitability for crops and vegetation, has also been derived. It is one of the crucial components for the agricultural domain as it not only considers both rainfall and temperature dynamics but also includes other important features such as soil moisture and potential evapotranspiration.

Soil dictates the types of farming systems that can function within a defined geographical unit and the associated management options. The amount and type of input to be applied for agricultural purposes, for example, are dictated by the properties of the soil and its health. Thus, soil is the predominant organising unit related to fertiliser and agronomic advisories. Knowledge of soil type and key soil properties is essential to defining environmental conditions and their overall

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

suitability. In this study, soil texture, soil organic carbon (SOC), pH, cation exchange capacity, total nitrogen and proportion of sand and clay were used for the clustering exercise. Gridded layers for soil chemical properties with a resolution of 250 m were downloaded from the World Soil Information database (Poggio *et al.*, 2021). The weighted average value of topsoil (0–30 cm) was used as it represents the average effective rooting depth of major crops.

Vegetation indexes are useful for characterising crop health and the potential capacity of the land to sustain vegetation. The normalised difference vegetation index (NDVI) and the enhanced vegetation index (EVI) are two vegetation indexes commonly used to monitor vegetation states and processes and are included in this framework. These two variables were used in the clustering exercise of this study.

Overall, 16 variable input layers were prepared for the clustering analysis in this study (Table 1). The key data-processing steps are presented in the following sections.

Data pre-processing

Figure 2 shows the major data-processing steps used in this study. All the data sets from global and/or regional sources were clipped based on the Ethiopian boundary. Because the different data sets have varied spatial resolutions, it was necessary to adjust to a common scale. In order to maintain the details of topographic (90 m) and soil-related information (250 m), and considering that climate variables will not significantly change over short distances, all the other data sets were resampled to 1 km resolution using the weighted average resampling method.

To make a comparable (and avoid spurious) impact of variables with larger ranges of values, it was necessary to normalise data sets (Abbott, 2014). In this study, data sets from various ranges to a common range (0–1) were normalised using a min-max scaling procedure.

Variable importance was used to assess how effectively a variable can differentiate between clusters and determine whether to include it in the analysis. Fowlkes *et al.* (1988) developed a variable selection method that focuses on a reduced variable space to make the model create new clusters parsimoniously. In this study, a principal component analysis (PCA) was used to reduce the data dimension and maintain the important variables by excluding those which carry redundant information. PCA axes with eigenvalues greater than 1 were retained to ensure that only PCA axes with a significant contribution are used for further analysis (Kaiser and Rice, 1974). Further, the quality of representation of the variables was analysed using the Cos2 indicator represented in a factor map. Cos2 represents the gradient of quality to highlight the most important variables in explaining the variations retained by the principal components (Kassambara, 2016). The factor map help to visualise the cluster of correlated variables in groups (ibid.). Finally, we used moving window variance over a 10 km radius to calculate the spatial variance of each pixel for the final list of covariates.

Clustering to define SRUs

The SRU exercise targets partitioning the heterogeneous environment into similar units where similar processes prevail and similar interventions can be made. Clustering is an approach that involves classifying data points into a specific group based on the premise that data points that are in the same group/cluster would have similar properties and/or features, whereas data points in different groups would have highly dissimilar properties and/or features. The aim is to cluster areas in a manner that maximises within-group similarity with maximising between-groups dissimilarity, thus minimising the total intra-cluster variation or total within-cluster sum of squares (WSS) (Goswami *et al.*, 2014). The total WSS measures the compactness of the clustering whereby ideal clusters should be compact, well-separated and stable (Brock *et al.*, 2008). The point where the difference with the previous number of clusters flattens out represents the optimal number of clusters as determined by the elbow method (Kaufman and Rousseeuw, 1990).

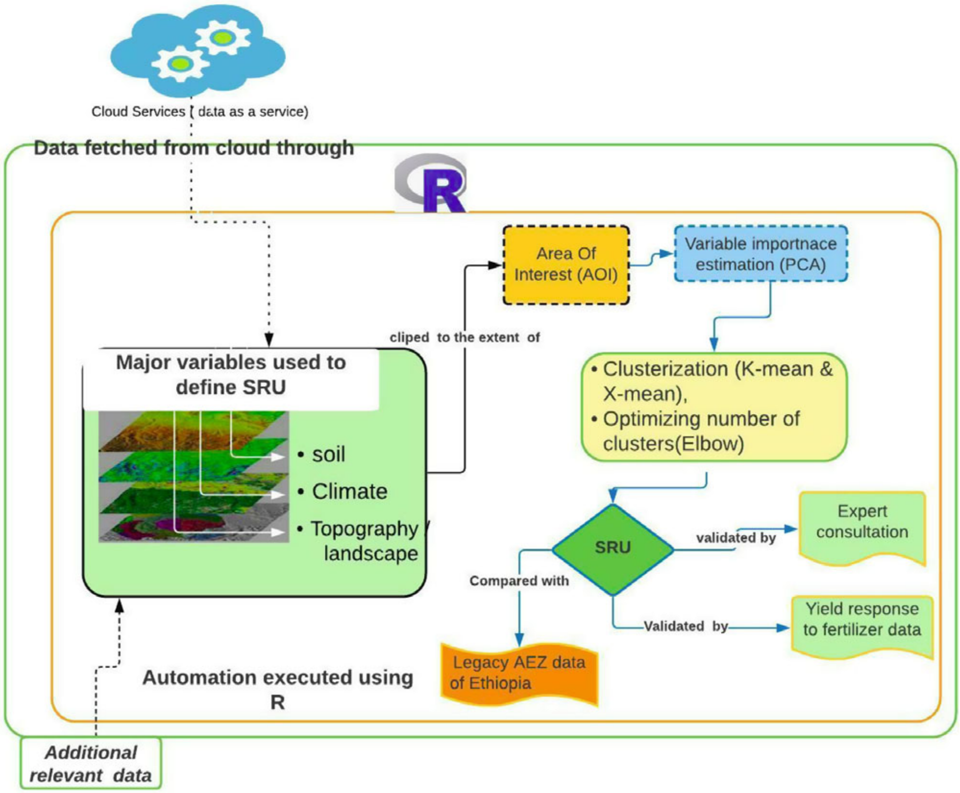

Figure 2. Flowchart showing the automation of SRU mapping.

One of the commonly used spatial clustering algorithms is the K-means (Hot and Popović-Bugarin, 2016; Jain *et al.*, 2000; Rahmani *et al.*, 2014; Shukla *et al.*, 2020). K-means is generally simple to implement and can be used with large datasets. The K-means is, however, not suited for use at a large scale due to the time it requires to give results when used for large areas and many covariates. In addition, it demands users to predefine the number of clusters to be produced. As a result, the X-means clustering method (Pelleg and Moore, 2000) from the WEKA package (Beckham *et al.*, 2016) has been developed as an extension of K-means to provide a more independent, unsupervised classification with improved computational efficiency, avoiding under parameterisation. Using X-means removes the necessity of *a pre-set number of clusters* from the user supporting the future developments of the approach developed in this study. The X-means approach is made efficient by replacing the need for K-means to compute the distance between every point to every centroid by recursive splitting of every cluster. It uses information criteria such as Bayesian information criterion (BIC) to select a model over another and whether the split at a centroid could be kept or not. The global BIC is used to define the final number of clusters. In this study, we used the X-means clustering method to group areas into similar or homogeneous zones within which similar recommendations can be made without requiring the need to predefine the number of clusters.

During the clustering exercise, all the variables (weather continues or categorical) were ingested into the classification algorithms, without any modification/creation of classes after normalisation. A moving window approach was used during classification to make the importance of the variables area/site specific.

Assessing the performances of clustering

To assess the results of the clustering algorithms, we have used qualitative and quantitative approaches. First, expert consultation has been used to gain an overall sense of the clustering results vis-à-vis expert knowledge and experiences of different geographical areas. At a one-day workshop, experts in soils, agronomy and geospatial analyses discussed the approaches used and corresponding clustering results. The maps were visualised on screen and printed in large colour prints to enable experts explore the results across the country. The second approach compared the distribution of standard deviation of optimal crop response to fertiliser application between the existing AEZ and the SRUs with the same number of classes. For this exercise, 3 commonly used AEZ maps categorised in 7 moisture belts and 15 AEZ (Hurni, 1998) and 33 zones (MoA, 2005) were selected. Corresponding SRUs with similar number of classes (7, 15 and 33) were then generated for 'one-to-one comparison" Maintaining the same level of granularity (number of classes) between the SRUs and AEZs, we compared within standard deviation between the two at a national scale. In this case, the assumption is that because the 'classes' are a result of homogeneous factors, they are expected to respond similarly to interventions. The expectation is that there will be less variation (indicated by standard deviation) in the crop response to nutrient added within similar classes than between different classes. If the standard deviation of the optimal fertiliser rate for SRU is smaller than AEZ for the same level of classes, the SRU approach of generating agriculturally homogeneous units is considered more appropriate. The optimal fertiliser recommendation used for this purpose is collated from many trial data sets (Tamene *et al.*, 2017) and analysed as outlined in Abera *et al.* (submitted).

Automation and tool development

One of the main aims of the study is to develop a scalable framework/system to delineate SRUs for different agronomic purposes. To build a generic system, the approach should primarily use globally available geospatial data sourced from the cloud with the flexibility for users to upload their own additional layers. The tool should also be designed to provide different options of clustering methods such as partitioning, distribution-based, hierarchical and fuzzy methods. The classification algorithms should also be designed to be scalable to run analysis for target areas of different spatial extent. The framework and tool developed in this study will thus enable the access of data from the cloud, and running clustering for a defined geographical area of interest using various approaches delineating similar zones. It also provides flexibility for user submit their own data to use solely for clustering and/or integrate with data derived from the cloud. The whole process is automated in an R programming environment and piloted for Ethiopia using commonly available geospatial data.

Results and Discussion

Variable selection for clustering

Figure 3 depicts the eigenvalues, variances and cumulative variances for the PCA analysis. The results show that the total cumulative variance percent of the three dimensions explained more than 80% of the variance. The first dimension explained 60% of the variance, whereas the second and third dimensions explained about 15% and 7%, respectively. Beyond the third dimension, variability lessens and the amount of new information that is carried diminishes. As a result, the first three dimensions were selected for further cluster analysis.

Figure 4 shows the quality of representation of the variables on the factor map analysed using the Cos2 indicator. Generally, well-represented variables by the principal components are positioned close to the circumference of the correlation circle, whereas the less represented ones are located close to the centre of the circle. In this case, all the variables except *slope* are well represented by the principal components. Figure 4 also shows that the distances of the variables from the origin in all covariates are high, indicating that most of these variables are useful for cluster

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

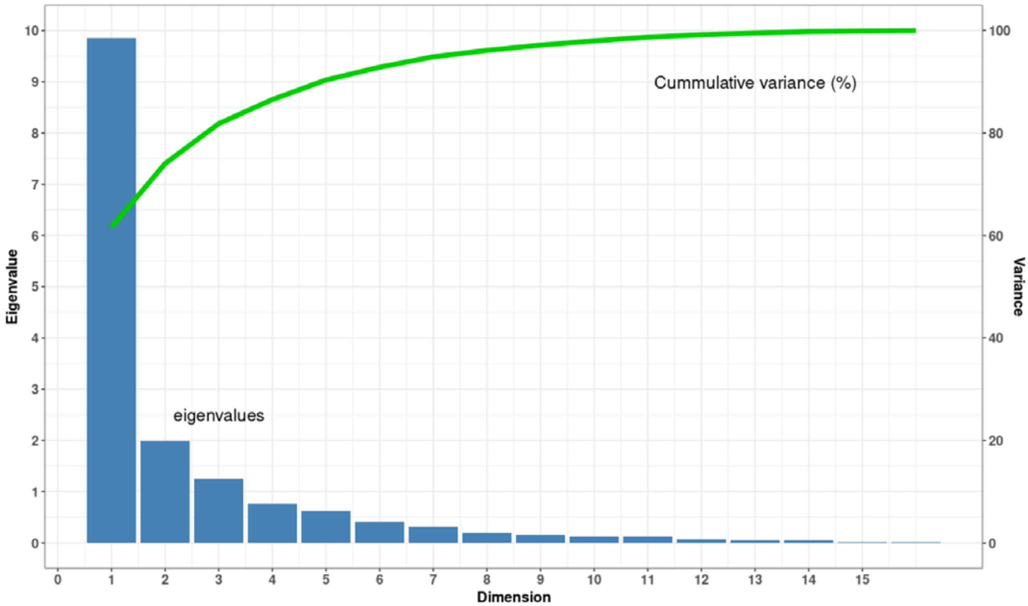

Figure 3. Eigenvalues, variance and cumulative variance of the PCA analysis.

analysis. Accordingly, most of the variables such as precipitation, total N, SOC, LGP, NDVI, EVI, PET, elevation, solar radiation and soil moisture contributed to the clustering analyses (depicted in the correlation map of Figure 4).

The optimum number of clusters and SRUs

The K-means clustering result in Figure 5a shows the output of the WSS for the computation of 1–150 clusters. This visual representation is typically referred to as the 'elbow method' and allows the user to identify an appropriate number of clusters for an unsupervised classification exercise (Chiang and Mirkin, 2010). Though there is no straightforward 'rule' to determine which number of clusters can best perform, the general recommendation is to consider the position where the elbow tends to plateau compared with the other number of clusters (Jain *et al.*, 2000; Khan and Mohamudally, 2020). In Figure 5a, it is possible to identify the position(s) where the WSS of the clusters tends to flatten out. On the basis of the distinct elbows, three example clusters (with 9, 30 and 53 classes) can be distinguished in this study (Figure 5a). The corresponding spatial SRUs for the above three classes are shown in Figure 5b.

The level of homogeneity of clusters needed can vary with the specific applications, and each of the maps shown in Figure 5b can be recommended for different applications. When considering macro-granular partitioning, it is possible to use fewer classes compared with applying for detailed process understanding and recommendation. In this example, SRU 9 (Figure 5b) can be used for applications/advisories that do not require detailed site specificity, such as identifying farming systems where detailed studies can be conducted. In this case, for instance, the lowlands like most parts of Somalia and the Afar region represent one unit each. The south-western highland forested area came out to be another separate unit (Figure 5b). When the number of clusters increases with the desire to obtain more detailed and homogeneous areas, SRUs with 30 and 53 units can be more applicable. In these clusters, we can identify units where similar climatic and major soil types occur and where agricultural practices and interventions can be prioritised. SRUs with 53 clusters can be used for detailed recommendations at high spatial resolution such as sub-catchments and lower area coverages.

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

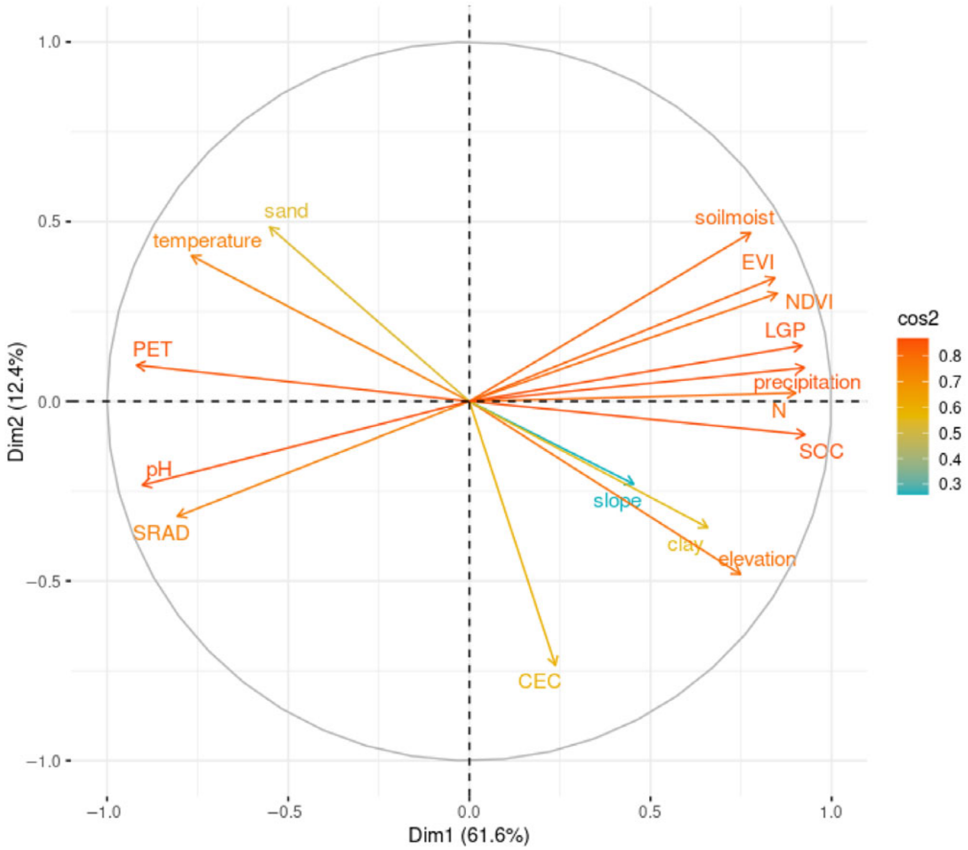

Figure 4. The representation quality of variables in the correlation plot.

Because we are operating at a national scale with complex topographic and climatic conditions, it was not possible to define an 'optimal' number of clusters using the K-means approach. This means the K-means approach requires predefined information about the number of clusters required. We thus employed the X-means clustering approach. Figure 6 shows the optimal number of clusters determined using the X-means clustering methods for Ethiopia. In this case, the algorithm resulted in 37 SRUs, after which it was not possible to split the existing units further. The SRU 37 is derived based on unsupervised classifier approach with no requirement to predefine the number of clusters because the X-means algorithm optimises the number of clusters to minimise WSS without the need to produce redundant clustering. This means that 37 clusters represent best approximation to partition Ethiopia into SRUs based on the covariates employed in the study. These units can be considered domains where targeted recommendations can be made considering specific and/or a combination of environmental variables present in those areas.

In principle, each SRU can be attributed using various environmental variables and users can obtain SRU properties for further scrutiny. However, owing to the heterogeneity of Ethiopia's landscape, specifically the highlands, the role of different attributes in creating the SRUs is complex (Figure 6). In order to facilitate interpretation, spatially aggregated mean value for all the covariates at SRU level (associated with Figure 6) is provided in Appendix I (Table A1). Users can compile the legend following the major environmental characteristics of each unit using the table provided with the statistics of each covariate in the clusters (Appendix I, Table A1).

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

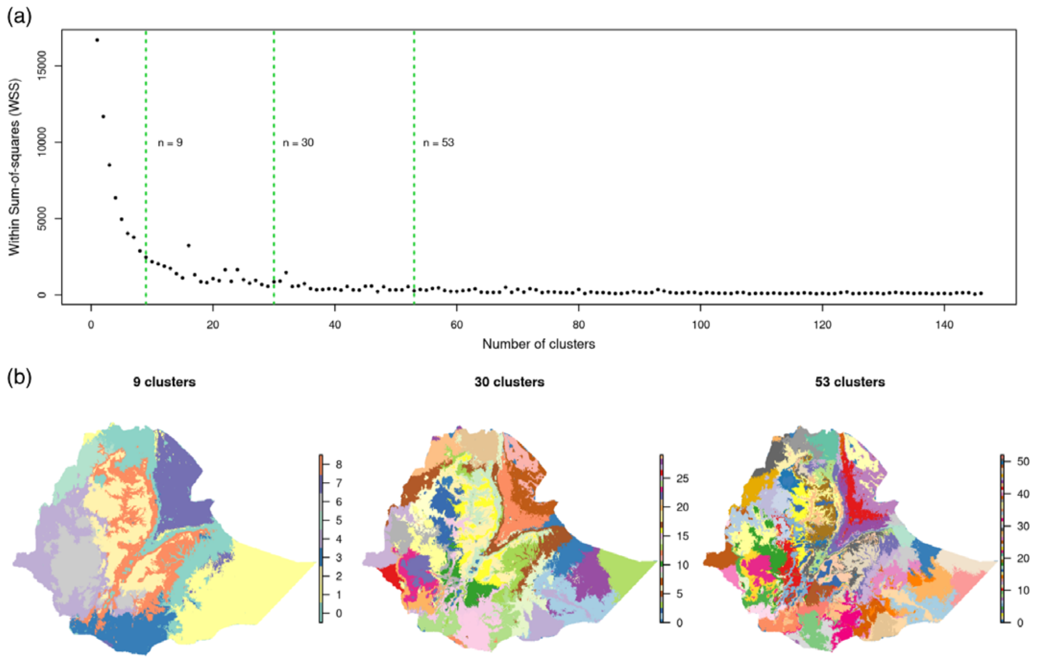

Figure 5. (a) Number of clusters using the elbow method in K-means clustering and (b) examples of clusters (SRUs) with three different number of classes.

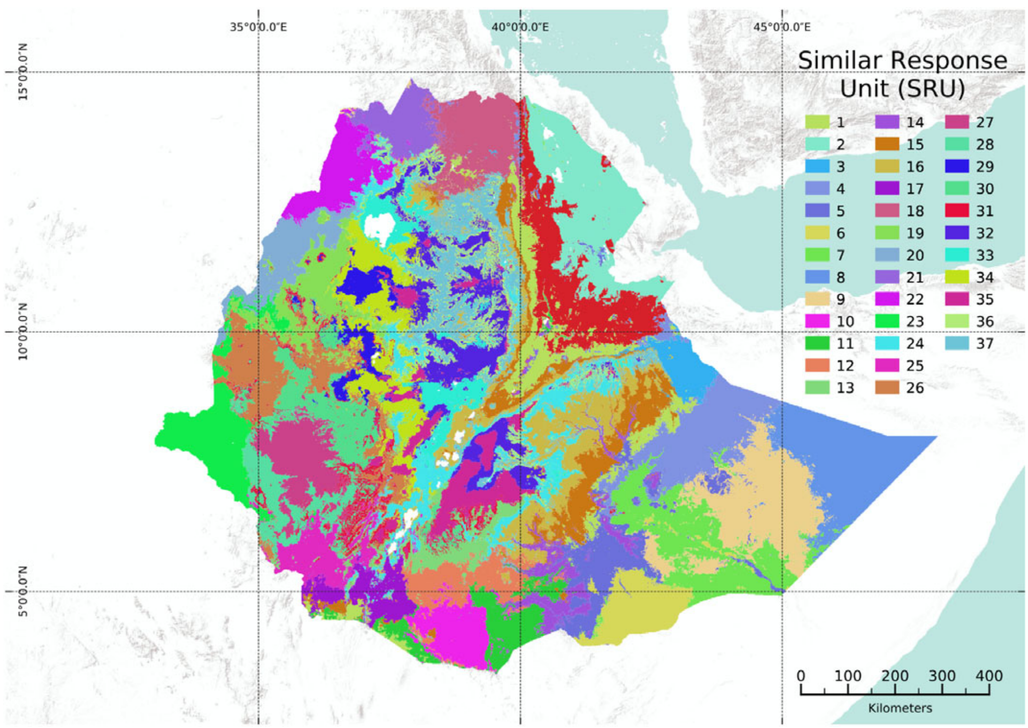

Figure 6. The pattern and spatial distribution of optimal SRUs based on X-means clustering algorithm.

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

Performance of the clustering and validity of SRU maps

The outputs shown in Figures 5 and 6 were presented to national experts who have come from different parts of the country and are familiar with the farming systems. The idea was to discuss the 'concept of SRUs' and assess the results. The participants appreciated the need to develop an automated system that can enable deriving clusters that can be used for targeting interventions. They also stressed the generality of the existing AEZs to repressing the heterogeneity of the farming systems and challenge to ascribe targeted advisories. After checking the various clusters, it was agreed that SRU with 9 clusters is too general while SRU with 53 clusters is too detailed and complex to comprehend visually. Considering this, the participants indicated the relevance and applicability of clusters 30 and 37 to facilitate agricultural decision making. These maps have captured heterogeneity well, while at the same time the number of clusters is optimal to manage for operational purposes. This, among others, is because the refined zonation can be used to provide detailed and location-specific advisories that will not be possible at generalised levels. The intermediate level of classification can also overcome the difficulty to develop and prescribe advisories at plot level (too detailed). However, the experts suggested the need to validate the results properly using quantitate and field data.

The other assessment was based on comparing the distribution of standard deviation of optimal crop response to fertiliser application between the existing AEZ and the SRUs with the same number of clusters (Figure 7). The result shows that for the same level of 7 and 15 clusters, the approach used in this study produced a lower standard deviation value than the AEZs for all of Ethiopia. This suggests that the use of SRUs for targeting fertiliser recommendation is preferable as it contains more homogeneous units in terms of response to fertiliser application; however, this changes when the number of clusters increases. The standard deviation of response to fertiliser application becomes similar when we compare AEZ and SRU clusters of 33 units. This basically suggests the advantage of a larger number of clusters that can capture heterogeneity better for agricultural technology targeting and recommendation. This was also corroborated by the experts' opinions. Future analysis will use field data related to integrated soil fertility management to assess the optimal number of clusters that can capture variability across space and scale.

Generally, it is noted that the 'classification approach' is essential for planning and targeting in situations where broader agro-ecological and farming system approaches are not plausible and, at the same time, plot-level interventions/advisories cannot be developed at the current state of data availability, especially in developing countries.

Clustering tool and reproducibility

Over the past few decades, widespread availability of spatial data and the advancement of robust modelling algorithms have increased. Such developments are creating unique opportunities that help answer targeted questions and prioritisations related to optimisation of natural resource management as well as steps to be taken for economic development and facilitation of poverty alleviation on the basis of recommendation domains (Akıncı *et al.*, 2013; Elsheikh *et al.*, 2013; Hyman *et al.*, 2013). Jasiewicz *et al.* (2014) introduced landscape similarity mapping using a numerical measure that assesses affinity between different landscapes based on the similarity between the patterns of their constituent landform elements. Automating the operationalisation of such techniques using similar and standard data sets can enable standardisation of the approaches and comparison of associated results. A study by Muthoni *et al.* (2017) used geospatial analysis and clustering techniques to delineate relatively similar clusters for scaling improved crop varieties and good agronomic practices. This approach added value to the previous ones by providing options to compare different clustering approaches.

In this study, we moved one step further in order to allow users access to relevant data from the cloud and/or add their own data and choose a specific geographical area of interest to run clustering. We developed a generic framework that supports creating a scalable system to be used to

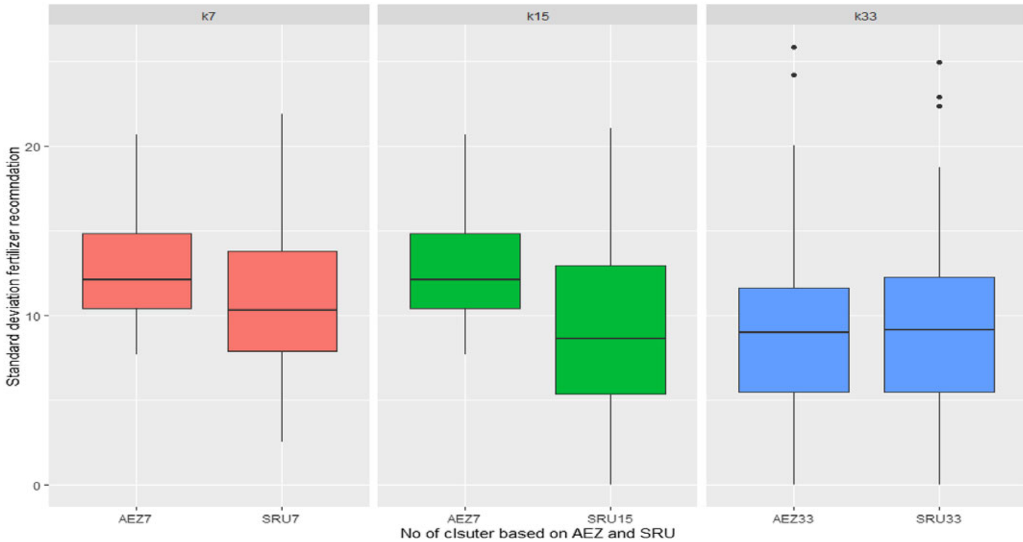

Figure 7. Standard deviation of recommended nitrogen fertiliser according to the agro-ecological zones and classification approach used in this study.

delineate SRUs for different purposes. The data analytics component provides options for several clustering methods (e.g. partitioning, distribution-based, hierarchical and fuzzy). The classification algorithm is also designed such that it can be scalable to run analysis for different extents but maintaining a standard procedure. The functionality of the system thus enables accessing data from the cloud, running clustering for a defined geographical area of interest and automatically performing clustering for different agronomic purposes.

The whole process is automated in an R programming environment and piloted for Ethiopia using globally available geospatial data. The system being put in place is generic and is being expanded to create a clustering analytic platform. The platform gives reproducible results which allow users to interactively choose data sources, use expert knowledge, experiment and compare the result of several algorithms and be flexible in order to work at any spatial scale and resolution to provide SRUs for interventions for sub-Saharan African countries. The tool generated is available for the public in an open repository (https://github.com/EiA2030/validation), including the workflows generated in this study.

Conclusion and Future Research Direction

Matching agricultural operations and inputs to the crop requirement, as is the case with precision agriculture, requires understanding the within-field variation of underlying biophysical factors. Because real-time monitoring and tailoring farm management to field-level variation are not possible, an alternative option is classifying the target area into homogenous units. This exercise becomes increasingly important to enhance agricultural productivity under optimised resource use for areas with fragmented farming systems and heterogeneous landscapes, as is the case in Ethiopia. To benefit from such advances, the current effort is made to create a scalable and operational tool that can harvest relevant data from the cloud and enable users to partition areas into uniform zones. The tool and its generic workflow have been piloted for Ethiopia.

The workflow and automated system demonstrated in this study can be used to create homogeneous landscape units using different clustering algorithms. The system is flexible to allow users to either refine or run targeted zoning as more relevant data are made available. The study serves to demonstrate the possibility of aggregating SRUs in a standardised way, ensuring transferability

https://doi.org/10.1017/S0014479722000126 Published online by Cambridge University Press

to other regions and settings. Although the algorithms used in this study are standard packages available in commonly used statistical software, the power of the system presented here is the practical convenience of offering an integrated solution whereby users could readily source the relevant data for different geographies, run the data analytics and obtain reproducible results including submission of the study area with defined territory.

The approach in this study targeted the use of global coverage data (or available at the scale of interest) that can be harvested from the cloud and harmonised for integrated analysis. As a result, some important data/variables that are not commonly available or that have questionable accuracy have not been used. For instance, geomorphology is an important factor that relates soil types with topographic catena and is among the major components of soil-forming factors. However, geomorphology data were not used in this study because we doubted the quality of the existing data a national scale. Such addititional data and improved analityics can improve the results.

The approach employed in this study demonstrates its potential to zoning spatial geographies into uniform clusters. Next steps will focus on validation the SRUs using 'ground information' and fine-tune their applicability. In addition, an attempt will be made to develop SRUs that will be specific to target defined issues such as climate-smart agriculture and other agronomic practices. In addition, functionalities will be incorporated to provide more options to the user (e.g. masking out non-agricultural areas) and assessing performances in an automated manner.

Acknowledgements. We thank the Supporting Soil Health Initiatives (SSHI) of GIZ-Ethiopia project for its support to gather data and conduct analysis. The Excellence in Agronomy (EiA) team provided great input and advnaced the analytics. The Africa RISING project funded by United States Agency for International Development and Feed the Future; and the Accelerating Impacts of CGIAR Climate Research in Africa (AICCRA) projects have also supported staff time of some of the authors. In addition, we recognise the support from the Water, Land and Ecosystems programme of the CGIAR as well as the European Union–International Fund for Agricultural Development project under the CGIAR Research Program on Climate Change, Agriculture and Food Security. We are also indebted to the great support of the 'coalition of the willing' members who were instrumental in generating various ideas and supporting analytics including providing data sets.

Funding Support. This work was supported, in whole or in part, by the Bill & Melinda Gates Foundation [INV-005460]. Under the grant conditions of the Foundation, a Creative Commons Attribution 4.0 Generic License has already been assigned to the Author Accepted Manuscript version that might arise from this submission.

Data Availability Statement. The data used to produce this article is available and accessible.

Conflict of Interest. We declare that there is no conflict of interest associated with all co-authors.

References

Abatzoglou J.T., Dobrowski S.Z., Parks S.A. and Hegewisch K.C. (2018). TerraClimate, a high-resolution global dataset of monthly climate and climatic water balance from 1958–2015. *Scientific Data* **5**, 1–12.

Abbott D. (2014). *Applied Predictive Analytics: Principles and Techniques for the Professional Data Analyst.* John Wiley & Sons. 10475 Crosspoint Boulevard.

Akıncı H., Özalp A.Y. and Turgut B. (2013). Agricultural land use suitability analysis using GIS and AHP technique. *Computers and Electronics in Agriculture* **97**, 71–82.

Amede T., Auricht C., Boffa J.-M., Dixon J., Mallawaarachchi T., Rukuni M. and Teklewold Deneke T. (2017). *A farming system framework for investment planning and priority setting in Ethiopia.* ACIAR Technical Reports Series No. 90. Canberra: Australian Centre for International Agricultural Research, 52pp.

Amede T., Gashaw T., Legesse G., Tamene L., Mekonen K., Thorne P. and Schultz S. (2020). Landscape positions dictating crop fertilizer responses in wheat-based farming systems of East African Highlands. *Renewable Agriculture and Food Systems* 1–13. doi: 10.1017/S1742170519000504

Aryeetey R., Holdsworth M., Taljaard C., Hounkpatin W.A., Colecraft E., Lachat C., Nago E., Hailu T., Kolsteren P. and Verstraeten R. (2017). Evidence-informed decision making for nutrition: African experiences and way forward. *Proceedings of the Nutrition Society* **76**, 589–596.

Beckham C., Hall M. and Frank E. (2016). WekaPyScript: classification, regression, and filter schemes for WEKA implemented in Python. *Journal of Open Research Software*, **4**(1), p. e33. http://doi.org/10.5334/jors.108

Bekele A. and Lakew Y. (2014). Projecting Ethiopian Demographics from 2012–2050 Using the Spectrum Suite of Models. Policy Brief, Ethiopian Public Health Association. http://www.healthpolicyproject.com/pubs/724_PROJECTINGETHIOPIAN.pdf

Brock G., Pihur V., Datta S. and Datta S. (2008). clValid: an R package for cluster validation. *Journal of Statistical Software* **25**, 1–22.
Chiang M.M.T. and Mirkin B. (2010). Intelligent choice of the number of clusters in k-means clustering: an experimental study with different cluster spreads. *Journal of Classification* **27**, 3–40.
Covic N. and Hendriks S.L. (2016). Achieving a nutrition revolution for Africa: the road to healthier diets and optimal nutrition. *ReSAKSS Annual Trends and Outlook Report 2015*. International Food Policy Research Institute (IFPRI). http://dx.doi.org/10.2499/9780896295933
CSA Central Statistical Agency. (2015). *Agricultural Sample Survey 2014/2015: Report on Land Utilization (Private Peasant Holdings, Meher Season)*. CSA, Addis Ababa: Ethiopia.
Dixon J.A., Gibbon D.P. and Gulliver A. (2001). *Farming Systems and Poverty: Improving Farmers' Livelihoods in a Changing World*. Rome, Italy: Food & Agriculture Org.
Donatien B. (2016). Poor data affects Africa's ability to make the right policy decisions. *The Conversation*. https://theconversation.com/poor-data-affects-africas-ability-to-make-the-right-policy-decisions-64064
Elsheikh R., Shariff A.R.B.M., Amiri F., Ahmad N.B., Balasundram S.K. and Soom M.A.M. (2013). Agriculture land suitability evaluator (ALSE): a decision and planning support tool for tropical and subtropical crops. *Computers and Electronics in Agriculture* **93**, 98–110.
FAO (1981). Report on the Agro-Ecological Zones Project 1978–81. Vol. 1, Methodology and results for Africa; Vol. 2, Results for Southwest Asia; Vol. 3, Methodology and results for South and Central America; Vol. 4, Results for Southeast Asia. [FAO] World Soil Resources Report, 48/1,-4, AGL-FAO, Rome.
FAO (1996). Agro-ecological Zoning. Guidelines. FAO Soils Bulletin 73, 78 pp. Soil Resources, Management and Conservation Service, FAO Land and Water Development Division. Food and Agriculture Organization of the United Nations, Rome, Italy. ISBN 92-5-103890-2
FAO (2017). The future of food and agriculture–trends and challenges. *Annual Report* **296**.
Fick S.E. and Hijmans R.J. (2017). WorldClim 2: new 1km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302–4315.
Fischer G.W. and Antonie J. (1994). Agro-Ecological Land Resources Assessment for Agricultural Development Planning; A Case Study of Kenya: Making Land Use Choices for District Planning. World Soil Resources Reports 71/9, FAO/IIASA, Land and Water Development Division, International Institute for Applied Systems Analysis, Rome.
Fowlkes E.B., Gnanadesikan R. and Kettenring J.R. (1988). Variable selection in clustering. *Journal of Classification* **5**, 205–228.
Freeman A., Noenbaert A., Herrero M., Thornton P.K. and Wood S. (2008). Strategies and priorities for integrated agriculture for development in the SSA CP. *Presentation on the Selection of Pilot Learning Sites for the Sub Saharan Africa Challenge Program (SSA CP), Accra, Ghana.*
Funk C.C., Peterson P.J., Landsfeld M.F., Pedreros D.H., Verdin J.P., Rowland J.D., Romero B.E., Husak G.J., Michaelsen J.C. and Verdin A.P. (2014). A quasi-global precipitation time series for drought monitoring. *US Geological Survey Data Series* **832**, 1–12.
Gerçek D. (2017). A conceptual model for Delineating Land Management Units (LMUs) using geographical object-based image analysis. *ISPRS International Journal of Geo-Information* **6**, no. 6: 170. https://doi.org/10.3390/ijgi6060170
Gerçek D., Toprak V. and Strobl J. (2011). Object-Based classification of land forms based on their local geometry and geomorphometric context. *International Journal of Geographical Information Science* **25** (2011), pp. 1011–1023. https://doi.org/10.1080/13658816.2011.558845
Goswami R., Chatterjee S., and Prasad B. (2014). Farm types and their economic characterization in complex agro-ecosystems for informed extension intervention: Study from Coastal West Bengal, India. *Agricultural Food and Economics.* **2**, 1–24. https://doi.org/10.1186/s40100-014-0005-2
Herrero M., Notenbaert A., Thornton P., Pfeifer C., Silvestri S., Omolo A. and Quiros C. (2014). *A Framework for Targeting and Scaling-Out Interventions in Agricultural Systems*. CCAFS Working Paper No. 62. Copenhagen, Denmark: CGIAR Research Program on Climate Change, Agriculture and Food Security (CCAFS).
Holdsworth M., Aryeetey R.N., Jerling J., Taljaard C., Nago E., Colecraft E., Lachat C., Kolsteren P., Hailu T. and Verstraeten R. (2016). The challenges, opportunities, and lessons learned in evidence-informed decision making in Africa. In Covic N. and Hendriks SL. (eds), *Achieving a Nutrition Revolution for Africa: The Road to Healthier Diets and Optimal Nutrition*. International Food Policy Research Institute (IFPRI), pp. 115–130.
Hot E. and Popović-Bugarin V. (2015). Soil data clustering by using K-means and fuzzy K-means algorithm. *Telfor Journal*, **8**(1): 56–61.
Hurni H. (1998). Agroecological Belts of Ethiopia: Explanatory notes on three maps at a scale of 1:1,000,000. Research Report, Soil Conservation Research Program, Addis Ababa.
Hyman G., Hodson D. and Jones P. (2013). Spatial analysis to support geographic targeting of genotypes to environments. *Frontiers in Physiology* **4**, 1–13.
IIASA/FAO (2012). Global Agro-ecological Zones (GAEZ v3.0): Model documentation. IIASA, Laxenburg, Austria and FAO, Rome, Italy, pp. 1–179.

Jain A.K., Duin R.P.W. and Mao J. (2000). Statistical pattern recognition: a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**, 4–37.

Jasiewicz J., Netzel P. and Stepinski T.F. (2014). Landscape similarity, retrieval, and machine mapping of physiographic units. *Geomorphology* **221**, 104–112.

Kaiser H.F. and Rice J. (1974). Little jiffy, mark IV. *Educational and Psychological Measurement* **34**, 111–117.

Kassambara A. (2016). Practical guide to principal component methods in R: unsupervised machine learning ((Multivariate Analysis Book 1). https://www.datanovia.com/en/product/practical-guide-to-cluster-analysis-in-r/

Kaufman L. and Rousseeuw P.J. (2009). *Finding Groups in Data: An Introduction to Cluster Analysis* (Vol. **344**). New York: John Wiley & Sons. ISBN: 0-471-87876-6

Khan D.M. and Mohamudally N. (2010). A multiagent system (MAS) for the generation of initial centroids for k-means clustering data mining algorithm based on actual sample data points. In *The 2nd International Conference on Software Engineering and Data Mining*. IEEE, pp. 495–500.

Khoroshev A. (2020). Landscape-ecological approach to spatial planning as a tool to minimize socio-ecological conflicts: case study of Agrolandscape in the Taiga Zone of Russia. *Land* **9**, 192. https://doi.org/10.3390/land9060192

MoARD (Ministry of Agriculture and Rural Development). (2005). *Major Agro-ecological Zones of Ethiopia*. Addis Ababa, Ethiopia: Forestry, Land Use and Soil Conservation Department.

Muthoni F.K., Guo Z., Bekunda M., Sseguya H., Kizito F., Baijukya F. and Hoeschle-Zeledon I. (2017). Sustainable recommendation domains for scaling agricultural technologies in Tanzania. *Land Use Policy* **66**, 34–48.

Notenbaert A., Herrero M., De Groote H., You L., Gonzalez-Estrada E. and Blummel M. (2013). Identifying recommendation domains for targeting dual-purpose maize-based interventions in crop-livestock systems in East Africa. *Land Use Policy* **30**, 834–846.

Notenbaert A., Pfeifer C., Silvestri S. and Herrero M. (2017). Targeting, out-scaling and prioritising climate-smart interventions in agricultural systems: lessons from applying a generic framework to the livestock sector in sub-Saharan Africa. *Agricultural Systems* **151**, 153–162.

Omamo S.W., Diao X., Wood S., Chamberlin J., You L., Benin S., Wood-Sichra U. and Tatwangire A. (2006). *Strategic Priorities for Agricultural Development in Eastern and Central Africa*. Research Report. Washington, DC: International Food Policy Research Institute (IFPRI). doi: 10.2499/9780896291584RR150

Pelleg D. and Moore A.W. (2000). X-means: Extending k-means with efficient estimation of the number of clusters. *In Proceedings of the 17th International Conference on Machine Learning* **1**, 727–734.

Penghui J., Manchun L. and Liang C. (2020). Dynamic response of agricultural productivity to landscape structure changes and its policy implications of Chinese farmland conservation. *Resources, Conservation and Recycling* **156**, 104724–104743.

Pennock D.J., Anderson D.W. and De Jong E. (1994). Landscape-scale changes in indicators of soil quality due to cultivation in Saskatchewan, Canada. *Geoderma* **64**, 1–19.

Poggio L., de Sousa L.M., Batjes N.H., Heuvelink G., Kempen B., Ribeiro E. and Rossiter D. (2021). SoilGrids 2.0: producing soil information for the globe with quantified spatial uncertainty. *Soil* **7**, 217–240.

Rahmani M.K.I., Pal N. and Arora K. (2014). Clustering of image data using K-means and fuzzy K-means. *International Journal of Advanced Computer Science and Applications* **5**, 160–163.

Rahmato D. (1994). Land tenure and land policy in Ethiopia after the Derg. Land Tenure project, institute of development research. *Addis Ababa University* **8**, 1–20.

Rizzo D., Marraccini E., Lardon S., Rapey H., Debolini M., Benoît M. and Thenail C. (2013) on Geografisk Tidsskrift-Danish *Journal of Geography* **113**(2): 71–86.

Salecker J., Dislich C., Wiegand K., Meyer K.M. and Péer G. (2019). EFForTS-LGraf: a landscape generator for creating smallholder-driven land-use mosaics. *PloS One*. doi: 10.1371/journal.pone.0222949

Shroff Z., Aulakh B., Gilson L., Agyepong I.A., El-Jardali F. and Ghaffar A. (2015). Incorporating research evidence into decision-making processes: researcher and decision-maker perceptions from five low-and middle-income countries. *Health Research Policy and Systems* **13**, 1–14.

Shukla R.K., Agrawal J., Sharma S., Chaudhari N.S. and Shukla K.K. eds. (2020). *Social Networking and Computational Intelligence: Proceedings of SCI-2018 (Vol. 100)*. Singapore Pte Ltd: Springer Nature.

Tamene L., Amede T., Kihara J., Tibebe D. and Schulz S. (eds.).(2017). A review of soil fertility management and crop response to fertilizer application in Ethiopia:towards development of site- and context-specific fertilizer recommendation. CIAT Publication No. 443. International Center for Tropical Agriculture (CIAT), Addis Ababa, Ethiopia. 86 p. Available at: http://hdl.handle.net/10568/82996

Tesfaye K., Jaleta M., Jena P. and Mutenje M. (2015). Identifying potential recommendation domains for conservation agriculture in Ethiopia, Kenya, and Malawi. *Environmental management* **55**, 330–346.

UNDESA (2017). World Population Prospects: The 2017 Revision, DVD Edition., United Nations, Department of Economic and Social Affairs, Population Division. Visited on 10/10.2021: https://www.un.org/development/desa/publications/world-population-prospects-the-2017-revision.html

Zewdie Y. and Tamene L. (2020). Toward an inclusive and evidence-based approach to farmland consolidation in Ethiopia. *Addis Ababa: International Center for Tropical Agriculture*, 30. https://cgspace.cgiar.org/handle/10568/110149

Appendix A

Table A1. The mean value for each environmental attribute associated with each cluster mapped in Figure 6

Cluster	Temp	Precip	SolRad	PET	Elev	Slope	Moist	SOC	TN	pH	CEC	Sand	Clay	EVI	NDVI	LGP
1	25	52.6	23.6	15.8	921.9	1.40	0.76	20.54	13.71	7.73	32.92	39.54	27.64	0.16	0.28	9.35
2	28.7	20.1	23.9	17.86	327.3	1.76	0.02	15.48	14.60	7.87	24.05	46.81	21.09	0.06	0.11	2.12
3	20.2	63.4	24.0	15.16	1594.6	1.27	0.29	22.99	18.68	8.08	31.34	36.74	36.36	0.20	0.32	11.54
4	24.7	38.1	24.7	16.7	1021.63	0.92	0.44	21.90	13.62	7.95	27.68	42.97	30.12	0.14	0.23	5.67
5	26.5	35.3	24.3	16.5	632.44	1.98	1.09	26.08	17.32	7.48	25.00	36.99	31.25	0.22	0.38	5.38
6	28.7	29.0	24.23	18.80	320.42	0.57	0.52	25.00	15.38	7.69	19.57	35.95	32.91	0.15	0.24	4.09
7	27.95	31.37	24.95	17.74	429.01	0.76	0.65	23.51	16.63	7.87	24.49	35.07	29.90	0.15	0.23	4.70
8	25.8	20.4	25.46	18.48	591.46	0.26	0.06	19.51	16.42	7.95	26.65	36.64	29.59	0.17	0.28	3.73
9	26.3	33.52	25.35	17.25	624.08	0.65	0.57	22.64	18.17	7.98	27.46	33.08	31.37	0.18	0.30	4.24
10	20.34	57.85	22.09	13.67	1274.36	1.30	2.60	24.37	13.18	6.82	25.79	57.64	23.30	0.25	0.44	9.46
11	23.08	45.57	22.68	14.46	967.32	1.12	1.46	22.85	13.22	7.29	24.82	47.36	29.81	0.25	0.44	9.96
12	20.48	61.07	22.03	12.82	1397.61	3.09	4.49	30.54	20.01	6.50	26.24	48.33	31.16	0.29	0.49	11.93
13	19.94	96.83	21.80	12.09	1678.60	5.14	9.06	39.35	36.59	6.20	29.53	36.90	35.62	0.40	0.64	17.03
14	24.08	45.39	23.59	14.96	925.88	4.62	1.35	28.43	18.71	7.07	25.49	39.90	32.67	0.26	0.45	9.05
15	23.33	61.59	23.57	14.43	1183.34	2.13	1.90	26.21	18.42	7.52	35.34	37.74	31.64	0.23	0.41	11.53
16	21.60	78.06	23.31	13.51	1434.70	2.85	2.67	30.89	22.73	7.29	37.52	32.37	36.34	0.26	0.45	14.53
17	24.66	47.52	22.15	13.37	893.37	1.73	2.16	28.33	22.09	7.27	46.46	35.97	37.26	0.33	0.55	11.75
18	20.57	65.88	25.45	14.06	1846.23	6.37	2.65	28.88	26.46	7.43	32.46	41.69	27.71	0.17	0.29	9.16
19	23.33	143.33	22.18	13.72	1131.77	3.60	10.21	31.39	29.58	5.95	27.73	31.80	36.26	0.34	0.57	17.76
20	26.36	113.11	22.93	16.02	768.77	1.82	8.49	24.01	24.02	6.35	30.83	33.90	36.03	0.30	0.51	16.47
21	25.39	75.13	25.93	16.23	1165.23	3.50	2.99	25.94	22.38	7.38	39.80	35.34	32.89	0.24	0.39	12.80
22	26.96	87.49	25.13	17.18	840.97	2.12	3.72	23.18	19.37	6.75	49.44	26.85	40.72	0.27	0.46	14.87
23	26.40	108.10	20.66	13.61	549.67	0.93	9.15	26.31	26.54	6.23	21.92	40.68	30.33	0.38	0.64	19.81
24	20.40	96.49	22.71	12.63	1634.88	4.62	4.37	35.37	30.31	6.80	36.92	30.26	38.67	0.33	0.56	17.54
25	25.45	80.06	21.54	13.08	745.34	2.17	10.55	34.38	35.43	6.47	35.54	34.61	38.70	0.42	0.67	18.00
26	21.36	137.85	21.37	12.09	1435.03	3.30	11.01	34.24	35.21	5.91	26.46	35.32	32.41	0.36	0.60	19.95
27	18.51	177.76	20.22	10.88	1921.04	5.26	21.17	46.96	48.84	5.56	24.95	30.36	38.38	0.47	0.72	25.53
28	23.22	127.77	20.53	12.07	1006.33	4.51	16.66	38.86	42.08	5.93	24.20	34.40	37.76	0.47	0.74	23.49
29	16.59	169.78	21.15	11.24	2262.16	4.51	8.25	44.54	42.77	5.69	30.57	22.12	44.78	0.33	0.55	19.03
30	19.28	171.65	20.94	11.44	1827.97	3.77	13.76	41.62	42.58	5.74	27.33	28.82	36.85	0.41	0.64	21.60
31	19.44	144.21	21.17	11.56	1802.93	11.68	12.14	42.46	46.34	5.89	27.28	32.77	36.25	0.41	0.64	21.93
32	14.54	110.75	21.89	10.90	2658.30	3.99	3.46	39.20	34.63	6.42	40.77	23.99	39.92	0.26	0.45	15.84
33	18.47	103.77	22.54	12.08	2038.83	2.46	3.67	32.38	26.14	6.79	42.25	22.50	42.87	0.23	0.40	14.52

(*Continued*)

Table A1. (*Continued*)

Cluster	Temp	Precip	SolRad	PET	Elev	Slope	Moist	SOC	TN	pH	CEC	Sand	Clay	EVI	NDVI	LGP
34	18.12	135.08	21.60	11.74	2009.31	3.22	6.82	38.82	32.87	5.93	32.32	23.88	41.76	0.30	0.51	18.06
35	13.16	123.79	21.30	10.26	2861.61	5.66	6.85	46.60	47.24	6.09	29.67	32.52	32.75	0.36	0.56	21.12
36	16.86	102.51	22.35	11.73	2286.14	13.27	3.33	37.56	36.19	6.61	38.99	27.85	37.90	0.26	0.47	14.94
37	19.92	87.48	22.59	12.93	1792.64	7.65	2.37	32.91	28.89	7.03	36.26	33.80	33.49	0.23	0.41	13.38

Temp: annual mean temperature; Precip: annual mean precipitation in mm; SolRad: net solar radiation in W/m^2; PET: potential evapotranspiration in mm; elev: elevation in metre; slope: slope in %; moist: soil moisture in mm; SOC: soil organic carbon in g/kg; TN: total nitrogen in g/kg; pH: soil pH; cation exchange capacity in cmol/kg; sand: sand in %; clay: clay in %; EVI: enhanced vegetation index; NDVI: net difference vegetation index; LGP: length of the growing period.

Cite this article: Tamene L, Abera W, Bendito E, Erkossa T, Tariku M, Sewnet H, Tibebe D, Sied J, Feyisa G, Wondie M, and Tesfaye K. Data-driven similar response units for agricultural technology targeting: An example from Ethiopia. *Experimental Agriculture*. https://doi.org/10.1017/S0014479722000126