

# Spatial Multivariate Cluster Analysis for Defining Target Population of Environments in West Africa for Yam Breeding

Tunrayo R. Alabi, IITA, Ibadan, Nigeria

Patrick Olusanmi Adebola, IITA, Abuja, Nigeria

Asrat Asfaw, IITA, Abuja, Nigeria

David De Koeyer, IITA, Ibadan, Nigeria

Antonio Lopez-Montes, International Trade Centre (ITC), Addison House International Trade Fair Center, FAGE, Accra, Ghana

Robert Asiedu, IITA, Ibadan, Nigeria

## ABSTRACT

Yam (*Dioscorea* spp.) is a major staple crop with high agricultural and cultural significance for over 300 million people in West Africa. Despite its importance, productivity is miserably low. A better understanding of the environmental context in the region is essential to unlock the crop's potential for food security and wealth creation. The article aims to characterize the production environments into homologous mega-environments, having operational significance for breeding research. Principal component analysis (PCA) was performed separately on environmental data related to climate, soil, topography, and vegetation. Significant PCA layers were used in spatial multivariate cluster analysis. Seven clusters were identified for West Africa; four were country-specific; the rest were region-wide in extent. Clustering results are valuable inputs to optimize yam varietal selection and testing within and across the countries in West Africa. The impact of breeding research on poverty reduction and problems of market accessibility in yam production zones were highlighted.

## KEYWORDS

*Dioscorea* spp., Environmental Clusters, Mega-Environment, Target Set of Environment, Targeting, Varietal Testing, Yam Belt

## INTRODUCTION

The population of West Africa is among those in the world that are increasing most quickly and is projected to be over 500 million in 2030 (ECOWAS-SWAC/OECD, 2007). The predicted rise in the regional human population poses risks and challenges to food security. Yam (*Dioscorea* spp.) is one of the strategic crops for sustainable food production and supply in Africa and beyond as it possesses the potential to produce more food per unit of land and water than some other crops. It is a preferred staple food for over 300 million people in West Africa. It provides nutritional benefits in the form of starch, protein, vitamins, and micro-nutrients meeting the dietary demands of people in

DOI: 10.4018/IJAGR.2019070104

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

the region (Asiedu and Sartie, 2010). The crop features prominently in the social-cultural activities of the people in the region (Obidiegwu and Akpabio, 2017). Yam production in Africa, though 40% of the total for cassava, has a value which exceeds all other African staple crops and is equivalent to the total value of the top three cereal crops, maize, rice, and sorghum (FAO, 2014). Despite its high value, yam productivity in West Africa is miserably low and has remained stagnant over decades. The achieved tuber yield of less than 10 t/ha under current production practices is around 20% of its potential yield of 50 t/ha (FAO, 2014) and is further constrained by losses of about 30% in storage (Mignouna et al., 2014). The low productivity is attributed to a combination of biotic and abiotic factors mostly associated with the shortened fallow under current extensive cultivation practices.

International Institute of Tropical Agriculture (IITA) and its partners in West African countries have been implementing several research and development initiatives to unlock the potential of the yam crop for sustainable food security and wealth creation. Genetic improvement is among those initiatives that aim at translating current advances in genetics and plant breeding sciences into more resilient, productive, and end-user preferred varieties for West Africa. The genetic improvement focuses on two major yam species, white yam (*Dioscorea rotundata*) and water yam (*D. alata*). The yam genetic improvement effort in the region seeks to build the capacity for breeding to accelerate genetic gain and lead to a strong pipeline of products to meet the needs of farmers and the markets.

Better understanding of target environments is essential for a yam breeding effort that is committed to developing and identifying improved genotypes that are in some way superior with reference to plant production purposes in the region. The target environments composed of a set of farms and seasons are often highly variable and the cause of differential phenotypic expressions of plants in a crop under cultivation (Hyman et al., 2013). The connection of phenotypic expression of a crop plant (value) which depends on environment (effect) often referred as genotype  $\times$  environment interaction (GEI), is one of the major factors limiting the efficiency of plant breeding programs as it influences the nature, magnitude, and predictability of the selection response (Cooper, 1999; Ceccarelli and Grando, 2007). Although GEI poses a big challenge to breeding program efficiency it cannot be ignored but could instead be exploited. Characterizing and defining target sets of environments (TSE) for breeding and cultivar recommendation are among the strategies to exploit the space and time dimension of GEI. Environment profiling helps to strategically locate the experimental or selection sites with a greater power in predicting the breeding trials' performance in the TSE in a period extending into the future.

Several studies have shown the utility of agricultural zoning for research and development based on environmental or biophysical variables. Costantini et al., (2016) applied the approach of multivariate clustering for the study of environmental drivers for delineating wine growing zones in Italy. Garcia et al., (2014) used a combination of climate, soil, and terrain data to identify coffee growing zones with high potential in Colombia. Similarly, Muthoni et al., (2017) employed the method of cluster analysis to delineate sustainable recommendation domains for scaling maize technologies in Tanzania. Furthermore, other studies have delineated geographic space by climate and soil classification schemes as a basis for extrapolating and applying agricultural information and research to a broader spatial scale (Wood and Pardey, 1998; Padbury et al., 2002; van Waart et al., 2013). Such environmental clustering schemes have been used to detect yield variability and limiting factors for crop growth (Caldiz et al., 2002; Williams et al., 2008), to regionalize optimal crop management recommendations (Seppelt, 2000), to determine suitable locations for new crop production technologies (Geerts et al., 2006; Araya et al., 2010), and to analyze impacts of climate change on agriculture (Fischer et al., 2005).

Generally, the two approaches of matrix and cluster methodologies have been used in environmental classification studies. In a matrix classification system each variable used to delineate zones is divided into classes or class-ranges. Class cutoff values for each variable can be based on expert-opinion or frequency distributions of the variable's range of values. Zones are then formed by the matrix cells of intersecting classes (van Waart et al., 2013). This approach involves some expert

opinion in defining the range of variables and hence is subject to some limitations such as subjectivity and non-reproducibility (Williams et al., 2008). One of the pioneering examples of a matrix zonation scheme is the work of Köppen (1900) that developed a climate classification system based on multiple variables related to precipitation and temperature.

The second main approach is that of clustering, sometimes referred to as statistical stratification (Hazeu et al., 2011). It relies on multivariate statistical analyses to separate cells into a researcher-specified number of distinct zones. Cluster analysis aims at identifying natural groups in a given dataset in a manner that maximizes within-group similarity and between-group dissimilarity (Goswami et al., 2014). Ideal clusters should be compact, well-separated, and stable (Brock et al., 2008, Muthoni et al., 2017). Multivariate cluster analysis has the advantages of objectivity, explicitness, defensibility, and improved transferability (Hargrove and Hoffman, 2004; Williams et al., 2008; Moral et al., 2016).

The advances in computing power and increased availability of improved spatial environmental data from the fields of remote sensing, climate modeling, and digital soil mapping have made feasible the use of quantitative analytical approaches to characterize crop regions into similar environmental clusters (Hyman et al., 2013; Moral et al., 2016). This study sought to characterize and define the target sets of yam environments in West Africa based on biophysical environmental variables and cluster them into homologous mega-regions using the approach of quantitative spatial multivariate cluster analysis. In the study, spatial analysis was used for supporting yam improvement research, taking advantage of remote sensing data and better and improved spatial data on soils and climate. The specific objective of this study was to delineate the yam growing regions in four West African countries into homogeneous environmental clusters for yam cultivar testing and predict their performance in the final recommendation domain for wider impact with breeding products. Socioeconomic information such as gridded human population, market access and poverty were employed to investigate potential impacts of testing and disseminating selected yam cultivars within the delineated yam growing zones.

## **MATERIALS AND METHODS**

### **The Study Area**

The study area covers four West African countries, Nigeria, Bénin, Ghana, and Côte d'Ivoire. Yam growing areas of each of these countries were delineated from sub-national yam production data for between 10 and 17 years and used as the geographic extent of yam characterization with cluster analysis (Figure 1). The entire yam growing areas in each country are too large for breeding trials to be set up hence some specific target districts/regions in those countries were selected. In Ghana, 15 yam growing districts were selected and 28 districts were chosen in Bénin. In Côte d'Ivoire, nine regions were selected, and eight States were chosen in Nigeria as operational areas of the Africa Yam project. Names of districts, regions, and States selected in the four countries are shown in Table 1. Mean annual rainfall in the target area ranged from 1000 mm in the north to 2500 mm in the southern coastal belt. Occupations in the yam growing region of West Africa are mainly agrarian with over 60% of the total population of 121 million (WORLDPOP, 2016) engaged in different forms of farming activities. Farming systems in the region ranged from root and tree crops mixed in the south to mixed cereal-root crop systems in the northern part (Dixon et al., 2001). Dominant soil types in the region are Lixisols, Acrisols, Leptosols, and Luvisols (ISRIC, 2016). Soil slope is predominantly flat (< 5%) in the target region (Authors' calculation from SRTM elevation data, (Jarvis et al., 2008).

### **Environmental Data**

Owing to advances in the field of remote sensing, climate modeling, and digital soil mapping, a wealth of environmental data exists today that can be considered in characterizing yam growing regions of West Africa. In this study, 23 bioclimatic variables, 21 soil properties layers, 18 remote sensing

Figure 1. West Africa showing (a) yam production (t); (b) yam harvested area (ha) and (c) average yam yield (t/ha) within target countries

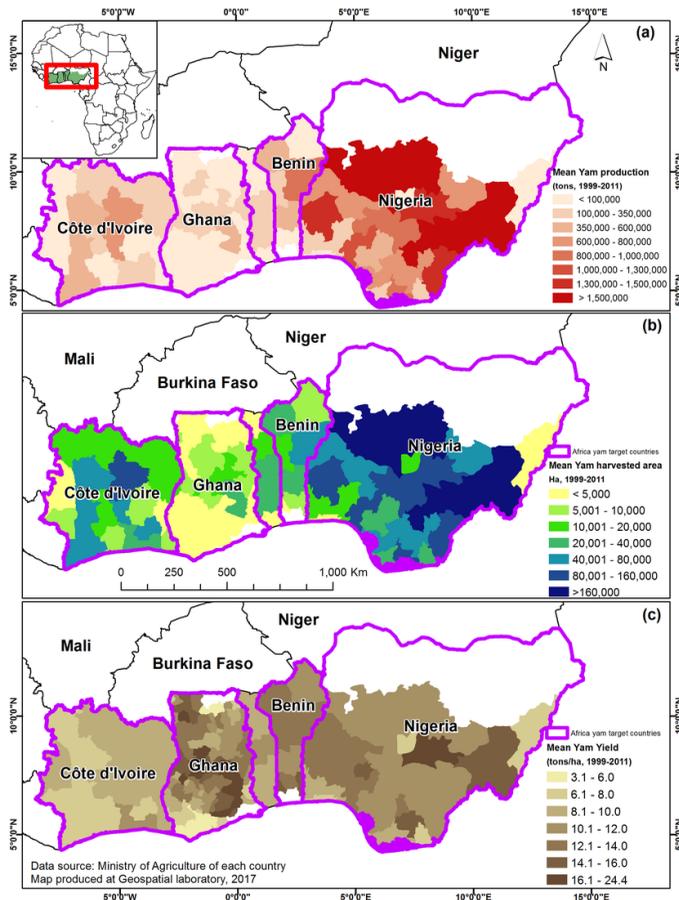


Table 1. Africa Yam project target countries and districts, states, or regions

Country	Target Districts/States/Regions
Ghana	Ejura, Sekyere West, Tachiman North, Sene East, Sene West, Nkoranza North, Sekyere East, Atebubu, Sekyere Central, Tian, Kitampo North, Nkoranza South, Kitampo South, Pru, Techiman South.
Bénin	Ketou, Savalou, Djidja, Ouesse, Bante, Glazoue, Save, Dassa, Kandi, Segbana, Kalale, Natitingou, Cobly, Boukounbe, Materi, Tanguieta, Tchaourou, Nikki, Sinede, Ndali, Perere, Bembereke, Bassila, Copargo, Pehonco, Ouake, Djougou, Kouande
Côte d'Ivoire	Bouna, Boundoukou, Korhogo, Daloa Zuenoula, Dabakala, Katiola, Ferkessadougu
Nigeria	Kwara, Federal Capital Territory, Taraba, Rivers, Ekiti, Niger, Cross River, Edo

vegetation layers, and 5 topographic variables derived from digital elevation model (DEM) data were used. In summary a total of 67 environmental layers which comprehensively define the biophysical traits of the yam growing zone of West Africa were used in defining homogeneous clusters for yam breeding. Socioeconomic variables such as total population, market access, and poverty incidence, as well as yam production data were used to estimate the potential impact of yam breeding for the region.

### *Bioclimatic Variables*

A comprehensive list of the 23-bioclimatic variables used in this study is presented in supplementary Table 9. These include annual rainfall, annual mean temperature, precipitation seasonality, and temperature of the warmest quarter which were obtained from Worldclim version 2 database (Fick and Hijmans, 2017). These variables were derived by interpolation of the long-term monthly values (1970-2000) of temperature and rainfall. In addition, climatic layers included the long-term value for mean sunshine hours (1983-2015). This was obtained from the recently released EUMETSAT's Satellite Application Facility on Climate Monitoring (CM SAF) monthly data described by Kothe et al., (2017). The bioclimatic variables describe the important biophysical environmental conditions for yam cultivation. Some of these parameters are described below: Annual total precipitation is the single most important climatic factor that determines yam cultivation. This represents the total water inputs to yam cultivation. Yam grows optimally in areas where annual total rainfall ranges between 1000 and 2000 mm (FAO Ecocrop, 2013). Mean annual temperature measured in degree Celsius is equally important for yam cultivation as it affects physiological traits. Yam grows best in environments where annual temperature is between 20 and 32 °C (FAO Ecocrop, 2013). Since yam cultivation can be strongly influenced by variability in the distribution of precipitation during the year, precipitation seasonality indicates the percentage of variability where larger percentages represent greater variability. Precipitation of the wettest quarter provides information on total amounts of rainfall during the wettest three months of the year which can highlight yam performance during the main growing season. Precipitation of the driest quarter is an indication of total precipitation during the driest three months of the year and can be used to relate how environmental factors may affect yam seed preservation and preparation for planting. Mean temperature of the warmest quarter is an indication of the hottest temperature conditions of the yam growing area. It provides mean temperatures during the warmest three months of the year which can be useful for examining how such environmental factors may affect germination and storage. Mean temperature of the coldest quarter represents temperatures during the coldest three months of the year and is important to capture yam response under low temperature. Apart from the basic 19 bioclimatic variables, other environmental factors were considered such as global aridity index and potential evapotranspiration (PET) which are also crucial to yam cultivation. PET is a measure of the ability of the atmosphere to remove water through evapo-transpiration (ET) processes. The FAO introduced the definition of PET as the ET of a reference crop under optimal conditions. Both global PET and the Aridity Index were produced by Trabucco and Zomer (2009) and have been downloaded from the CGIAR-CSI GeoPortal (<http://www.csi.cgiar.org>). Aridity is usually expressed as a generalized function of precipitation, temperature, and PET. An Aridity Index (UNEP, 1997) can be used to quantify precipitation availability for crops over atmospheric water demand.

### *Edaphic Variables*

Soil conditions are known to influence crop yield and performance next to climatic factors (Costantini et al., 2016). In recent times, the availability of unbiased remotely sensed environmental covariates has led to great improvements in the field of digital soil mapping thereby enhancing the availability of more detailed layers of soil properties. Gridded soil layers at a spatial resolution of 250 m were downloaded from the ISRIC-World Soil Information database (ISRIC, 2016; Hengl et al., 2017). These layers were generated using an automated mapping framework based on random forests (Hengl et al., 2017). The layers were estimated at six standard depth intervals but in this study, those depths that correspond mainly to the topsoil conditions were utilized. Generally, soil property layers were used with depths of 0-5 cm, 5-15 cm, and 0-20 cm. In some cases, averages of layers with a depth between 0-5 cm and 5-15 cm were taken and used to represent 0-15 cm soil depth. Explicit description of the 21 soil property layers used in the study is presented in supplementary Table 10. Among these are content of Soil organic carbon (SOC), Total nitrogen, Soil pH, Bulk density, Sand, silt, and clay content. Other soil parameters considered are related to the presence of micronutrients and cations such as Cation

exchange capacity (CEC), Exchangeable bases, magnesium and potassium. SOC is a proxy for soil fertility status (Hengl et al., 2015; Muthoni et al., 2017). Organic matter makes up just 2-10% of the soil's mass but it has a critical role in the physical, chemical, and biological functions of agricultural soils. SOC is a measurable component of soil organic matter which contributes to nutrient turnover and CEC, soil structure, moisture retention and availability, degradation of pollutants, greenhouse gas emissions, and soil buffering. Total nitrogen is the sum of total Kjeldahl nitrogen (ammonia, organic and reduced nitrogen) and nitrate-nitrite. It can be derived by monitoring for organic nitrogen compounds, free-ammonia, and nitrate-nitrite individually and adding the components together. It is an important indicator of fertility status. Bulk density is an indicator of soil compaction and reflects the soil's ability to function for structural support, water and solute movement, and soil aeration (Arshad et al., 1996). High bulk density is an indicator of low soil porosity and soil compaction. It may cause restrictions to root growth and poor movement of air and water through the soil. Compaction can result in shallow plant rooting and poor plant growth, influencing crop yield and reducing the vegetative cover available to protect the soil from erosion. Since yam is a tuber crop, bulk density is crucial to yield. Soil pH determines plant tolerance to acidity or alkalinity. pH levels range from 0 to 14, with 7 being neutral, below 7 acidic and above 7 alkaline. The optimal pH range for most plants is between 5.5 and 7.0; however, many plants have adapted to thrive at pH values outside this range. CEC is the total capacity of a soil to hold exchangeable cations. It is an inherent soil characteristic and so it is difficult to alter it significantly. It influences the soil's ability to hold onto essential nutrients and provides a buffer against soil acidification. Exchangeable acidity refers to the amount of acid cations, aluminum and hydrogen, occupied on the CEC. When the CEC of a soil is high, but it has a low base saturation, the soil becomes more resistant to pH changes. Exchangeable bases are commonly defined as the alkali and alkaline earth metals (principally calcium, magnesium, potassium, and sodium) attached to the clay and organic constituents of soils which can be exchanged with each other and with other positively charged ions in the soil solution. Soil electrical conductivity (ECN) is a measure of the amount of salts in soil (salinity of soil). It is an important indicator of soil health. In the soil, the ECN reading shows the level of ability the soil water has to carry an electrical current. The ECN level of soil water is a good indication of the amount of nutrients available for crops to absorb.

### ***Remote Sensing Vegetation Indices and Reflectance Bands Data***

A detailed list of 18 environmental covariates derived from remote sensing products such as Moderate Resolution Imaging Spectroradiometer (MODIS) and SPOT earth observation satellites used for the current analysis is provided in supplementary Table 11. MODIS is a key instrument aboard the Terra and Aqua satellites of NASA and SPOT is a commercial high-resolution optical imaging Earth observation satellite system of the French Space Agency. Remotely sensed vegetation layers from MODIS include normalized vegetation index (NDVI), enhanced vegetation index (EVI), mid-infrared reflectance band (MREFB7), blue reflectance band (MREFB1), gross primary productivity (GPP), and net primary productivity (NPP). Other layers from MODIS include black-sky albedo and land surface temperatures. In most cases MODIS land surface data were long-term averages of monthly composites of about 18 years (2000 to 2017). Data from SPOT earth observation satellites included the fraction of absorbed photosynthetically active radiation (SPAR) and fraction of green vegetation cover (SCOVER). These datasets were downloaded from environmental layers prepared by AfsIS (2017) for digital soil mapping in Africa (<http://africasoils.net/services/data/remote-sensing/land/>). NDVI indices result from the complex interaction between vegetation, climate, terrain, soil, and hydrology (Langella, 2008; Costantini et al., 2016). The index is a measure of vegetation greenness. Where the index is close to 0, vegetation is scarce. Where the index is close to 1, the vegetation is more abundant. NDVI varies from medium to high in yam growing regions of West Africa. The EVI is an 'optimized' vegetation index designed to enhance the vegetation signal with improved sensitivity in high biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences. EVI tends to be more sensitive

to plant canopy differences such as leaf area index (LAI), canopy structure, and plant phenology and stress than NDVI which generally responds just to the amount of chlorophyll present. Where the index is close to 0, vegetation is scarce. Where the index is close to 1, the vegetation is more abundant. GPP measures the rate at which organisms convert light energy to the chemical energy of organic compounds. Hence, the total amount of energy assimilated by plants in an ecosystem during photosynthesis (gross primary productivity) varies among environments. Highly vegetated areas produce higher levels of energy, resulting in higher GPP. Much of the energy assimilated by plants through photosynthesis is not stored as organic material but instead is used during cellular respiration. In these processes, organic compounds such as carbohydrates, proteins, and fats are broken down or oxidized to provide energy for the cell's metabolic needs. The energy not used in this process is stored in plant tissues for further use and is called net primary productivity (NPP). About 40 to 85% of GPP is not used during respiration and becomes NPP. The highest NPP in terrestrial environments occurs in swamps and marshes and tropical rainforests; the lowest occurs in deserts. The value represents the available energy remaining after photosynthetic activity in an ecosystem. Areas that are highly vegetative produce an excess of energy that represents the NPP.

### *Topographic Variables*

Topography is an important determinant of soil moisture distribution and thus drives the functioning of terrestrial ecosystems, including vegetation composition and structure (Radula et al., 2018). It affects soil erosion and therefore nutrient availability to crops. It also presents farm management complexity especially during mechanized cultivation of crops. In general, the higher the energy of the relief and the smaller the valleys and accumulation areas the harsher is the management difficulty and the greater the nutrient loss available to crops (Costantini and Barbetti, 2008). Therefore, terrain data from Shuttle Radar Topography Mission (SRTM) were obtained from SRTM 90m DEM database v4.1 produced by Jarvis et al., (2008) (<http://srtm.csi.cgiar.org>). Other topographic variables, such as slope gradient, aspect, topographic wetness index (TWI) and stream power index (STRPI), were derived from the DEM following Reuter and Nelson (2009). TWI quantifies the tendency of soil water distribution, which is affected by topography and has been frequently used to model soil moisture distribution (Zhu et al., 2014). Slope affects water erosion and nutrient leaching and thereby influences fertility of the soil. Topographic variables used in this study are listed in supplementary Table 11.

### *Socioeconomic Variables*

All agricultural activities take place within a socioeconomic context, hence information related to socioeconomics were obtained to investigate potential adoption and the impact of improved yam technologies of the AfricaYam project. Socioeconomic variables used included gridded human population 2015 data downloaded from Worldpop database (WORLDPOP, 2016). The human population layer was generated using a random forest method that estimated the number of people per pixel with national totals adjusted to match United Nation's population division estimates (Stevens et al., 2015). The human population data are a proxy for availability of markets and labor (Vanlauwe et al., 2014; Muthoni et al., 2017). Gridded spatial data representing travel time in hours from a  $1 \times 1$  km resolution pixel to the nearest town with more than 20,000 inhabitants was used as surrogate for access to market (HarvestChoice, 2015). Access to market is crucial for yam production, processing, and transportation to the point of consumption as well as access to agro-inputs such as herbicides and fertilizer. Data on estimates of the number of people living on less than \$1.25 and \$2 per day for 10 km pixel areas, based on estimates derived from combined poverty maps and survey data for the entire world, with a base year of 2005 were downloaded from HarvestChoice website (HarvestChoice, 2012). Poverty prevalence data were used to estimate the impact of the AfricaYam project to the poor living in the region. Gridded projected population data for 2050 were downloaded from Data GURU (Boke-Olén et al., 2017) and were also used to estimate the future impact of AfricaYam technologies for future conditions. To assess yam production within the yam zones of the target countries, the study

relied on spatial production allocation model (SPAM) developed by the International Food Policy Research Institute (IFPRI) for generating highly disaggregated crop-specific production data. SPAM is a spatial model to allocate crop production derived from units reporting large statistics, such as country, province, and district, to a raster grid at a spatial resolution of 5 minutes of arc normally referred to as a 10 km × 10 km pixel for simplicity. SPAM 2005 version 3 was downloaded from HarvestChoice website (You et al., 2014).

## **Data Analysis**

### *Data Preprocessing*

Environmental covariate layers were of diverse spatial resolution as can be seen from supplementary Tables 9-11. Hence, before cluster analysis, all layers were re-sampled to a spatial resolution of 250 m using bilinear transformation in ArcGIS 10.5. Bilinear transformation re-sampling method is an interpolation algorithm that determines the new value of a cell based on a weighted distance average of the four nearest input cell centers. It is useful for continuous data and will cause some smoothing of the data (ESRI, 2017). A spatial resolution of 250 m was chosen for data analysis since most of the datasets were of this resolution, especially soil layers and remote sensing vegetation layers. All layers were also projected to “Lambert Azimuthal Equal Area projection” using ArcGIS project tool.

Since the variance depends on the scale of the variables, it is customary to first standardize each variable to have mean 0 and standard deviation 1. After the standardization, the original variables with possibly different units of measurement are all in comparable units (Fodor, 2002). Hence following the method of Garcia et al., (2014) all 67 environmental variables were normalized to a scale 0 to 1 to ensure that the means of the standardized variables were 0 and the variances were equal to 1. This is known to guarantee stable convergence of weight and biases of all variables.

Derivation of topographic variables such as slope, aspect, and TWI from SRTM elevation data were performed within QGIS 3.0.1 using “SAGA GIS Topographic Wetness Index tool” (QGIS Development Team, 2018).

### *Principal Component Analysis (PCA)*

It has been observed that spatial autocorrelation and multicollinearity exist in environmental data (Shaker and Zubalsky, 2015; Muthoni et al., 2017). To address this, following the method of Costantini et al., (2016) Principal Component Analysis (PCA) was employed separately on the four types of environmental data. PCA is one of the most popular dimensionality reduction methods. It is a linear method, meaning that the transformation between the original data and the new lower dimensional representation is a linear projection. Its main purpose is dimensionality reduction, but it can also be used to explore relationships between variables. Often it is used as a preprocessing method either for data orthogonalization and eliminating redundancy caused by variable correlation or for dimensionality reduction, before employing another statistical method, such as regression or clustering (Fodor, 2002; Jolliffe et al., 2003). As principal components (PCs) are orthogonal, regression and clustering methods can proceed with data independence guaranteed (Demšar et al., 2013). Moreover, PCA as a multivariate analysis technique, allows for identifying of the variables that account for most of the total variance in data sets (Gavioli et al., 2016). PCA produces a new set of synthetic variables named principal components (PCs), which are uncorrelated among themselves (Johnson and Wichern, 2007). Hence, in each of the four sets of environmental data, PCA reduced the dimensions to a few uncorrelated latent variables represented by PC axes. Following Kaiser’s criterion (Kaiser and Rice, 1974) as utilized by Garcia et al., (2014), only PCA axes with eigenvalues greater than 1 were retained for further analysis. This ensured that only PCA axes with a significant contribution are used for further analysis. The factor loadings of variables in PCA axes indicated the strength and direction (negative or positive) of the correlation between variables and PCA axes. This information was used to infer the main contributions of each environmental layer to each significant PCA axis.

### *Multivariate Cluster Analysis*

PCA layers with eigenvalues greater than 1 resulting from the four PCA performed in the previous section were used as input into a multivariate cluster analysis using iterative self-organizing data analysis technique (ISODATA) after Metzger et al., (2012). “Iso Cluster Unsupervised Classification tool” within Multivariate toolbox of ArcGIS 10.5 Spatial Analyst extension was used to stratify the significant PCA layers from soil, climatic, vegetation, and topographic variables into environmental clusters. The “Iso Cluster tool” uses a modified iterative optimization clustering procedure, also known as the migrating means technique. The algorithm separates all cells into the user-specified number of distinct unimodal groups in the multidimensional space of the input bands (Tou and Conzalez, 1974). This type of clustering uses a process in which all samples during each iteration are assigned to existing cluster centers and new means are recalculated for every class. The process is repeated: each cell is assigned to the closest mean in multidimensional attribute space, and new means are calculated for each cluster based on the membership of cells from the iteration. Clusters are therefore homogeneous groups based on the multivariate inputs. In this case yam target zones are divided into homogeneous similarity zones according to covariates specified as inputs. Conceptually it would have been desirable to use statistical stopping rules (Gordon, 1996) to determine the optimal number of zones, reducing judgment by determining a mathematically optimal number of divisions in multivariate parameter space. Such cut-off rules balance high intra-cluster similarity and low inter-cluster similarity but such internal criteria for the quality of a clustering do not necessarily translate into good effectiveness in an application (Manning et al., 2008). Consequently, based on our knowledge of the environment, “Iso cluster tool” was assigned an initial ten clusters from which cluster aggregation was performed as in the next section.

### *Cluster Aggregation*

To provide structure and support the development of a consistent nomenclature, as well as to facilitate summarizing and reporting, it is useful to aggregate the clusters consistently to a limited set of environmental zones (Bunce et al., 2002; Leathwick et al., 2003; Metzger et al., 2012). The dendrogram tool in ArcGIS uses a hierarchical clustering algorithm and was used to aggregate the initial set of clusters to statistically dissimilar clusters. A dendrogram is a type of tree diagram showing hierarchical clustering and provides a visual way to delineate statistically dissimilar classes. The dendrogram was then used to determine the aggregation of the 10 clusters into 7 classes for the entire West Africa region.

### *Quantifying the Impact of Yam Breeding Program Using Socioeconomic Analysis*

After delineation of clusters, spatial analysis was performed to examine socioeconomic contexts and the potential impact of AfricaYam project’s improved technologies. High agricultural potential areas that are characterized by dense rural population, significant poverty levels, and high market accessibility could be of higher priority for demonstrating the impact of improved agricultural technologies (Hyman et al., 2013). Potential impact for improved yam technologies was calculated using the following five socioeconomic variables: total human population; poverty levels, projected total population in 2050, yam production, and market accessibility. Spatial analytical tools such as “overlay”, “tabulate area” and “zonal statistics as table” were employed for this within ArcGIS 10.5.

## **RESULTS AND DISCUSSION**

### **Principal Component Analysis**

Table 2 presents the results of the PCA of the four sets of environmental data: edaphic, bioclimatic, topographic, and remote sensing variables. The bulk of edaphic data was summarized by 10 principal factors, which together explained 95.4% of the variability in this data set. All 23 bioclimatic variables

**Table 2. Strongly loaded variables in the Principal Component axes (CVE: cumulative variance explained)**

<b>(a) Edaphic Variables</b>										
	<b>PCA1</b>	<b>PCA2</b>	<b>PCA3</b>	<b>PCA4</b>	<b>PCA5</b>	<b>PCA6</b>	<b>PCA7</b>	<b>PCA8</b>	<b>PCA9</b>	<b>PCA10</b>
	TEXMHT (0.57)	AWCtS (0.60)	PHIHOX (-0.62)	BDRLOG (0.93)	ALUM3S (0.82)	EMGX (0.40)	TEXMHT (0.57)	AWCtS (0.56)	BDTICM (0.46)	BDTICM (0.48)
	CLYPPT (-0.36)	AWCh1 (0.46)	SLTPPT (-0.35)		PHIHOX (0.32)	CECSOL (0.39)	SLTPPT (0.50)	BLDFIE (0.47)	NTO (-0.41)	EMGX (-0.42)
	PHIHOX (0.35)	BLDFIE (-0.41)	SNDPPT (0.35)			EXBX (0.35)	SNDPPT (-0.37)	AWCh1 (-0.42)		EXBX (-0.38)
	NTO (-0.32)	ALUM3S (0.31)				NTO (0.34)	AWCtS (-0.35)			CECSOL (0.33)
	SNDPPT (0.30)									
<b>CVE(%)</b>	36.4	58.3	69.9	76.5	82.3	86.6	90.2	92.9	94.3	95.4
<b>(b) Bioclimatic Variables</b>										
	<b>PCA1</b>	<b>PCA2</b>	<b>PCA3</b>	<b>PCA4</b>	<b>PCA5</b>	<b>PCA6</b>	<b>PCA7</b>			
	Sunhr (0.35)	BIO16 (0.44)	BIO9 (0.40)	Solarad (0.57)	Solarad (0.55)	BIO18 (0.46)	Sunhr (0.57)			
	BIO4 (0.34)	BIO13 (0.40)	BIO6 (0.38)	BIO19 (-0.47)	BIO19 (0.40)	Solarad (-0.39)	BIO4 (-0.39)			
	BIO15 (0.33)	BIO19 (0.39)	BIO11 (0.33)	BIO13 (0.32)	BIO3 (0.36)	BIO2 (0.38)				
	PET (0.33)	BIO15 (0.32)	BIO1 (0.33)		BIO4 (-0.36)	BIO7 (0.33)				
	BIO7 (0.30)		BIO10 (0.31)		PET (0.33)					
<b>CVE (%)</b>	56.0	72.1	82.7	88.7	92.5	95.2	96.7			
<b>(c) Remote Sensing Variables</b>										
	<b>PCA1</b>	<b>PCA2</b>	<b>PCA3</b>							
	SPAR (0.44)	SPAR (-0.42)	NDVI (0.65)							
	SCOVER (0.43)	SCOVER (-0.45)	EVI (0.47)							
	MWNA (-0.43)	MBNA (0.31)	MBVA (-0.30)							
	MBNA (0.40)	MBSA (0.37)								
	MBSA (0.37)	MBVA (0.32)								
		MWNA (0.31)								
<b>CVE(%)</b>	89.2	97.5	98.5							
<b>(d) Topographic Variables</b>										
	<b>PCA1</b>	<b>PCA2</b>	<b>PCA3</b>							
	Aspect (1)	TWI (0.87)	Elevation (0.9)							
		Elevation (-0.42)	TWI (0.43)							
<b>CVE(%)</b>	91.5	97.3	99.8							

were represented by seven principal factors which accounted for 96.7% of the total variability in the data set. Topographic variables were fully explained by three principal factors accounting for 99.8% of the total variability. Eighteen remote sensing variables were summarized by only three principal axes explaining about 98.5% of the total variance.

PCA1 for the edaphic data accounted for 36.4% of the variability and was mainly dominated by soil texture (TEXMHT, 0.57), clay content (CLYPPT, -0.36), and soil pH (PHIHOX, 0.35). Other significant loadings on this axis were total nitrogen (NTO), which was strongly loaded at -0.32 and sand content (SNDPPT), loaded at 0.3 (Table 2). PCA2 represented 21.9% of the variability of the soil data. It was dominated by porosity (AWCts, 0.60), available soil water capacity (AWCh1, 0.46), bulk density (BLDFIE, 0.41) and aluminum concentration (ALUM3S, 0.31). PCA3 to PCA6 respectively explained 11.6, 6.6, 5.7, and 4.3% of the variation. PCA3 was related to three factors, soil pH, silt, and sand content; out of these, only silt content never featured among the dominating variables of PCA1. PCA4 was represented by probability of occurrence of depth to bedrock (BDRLOG). PCA5 and PCA6 featured variables that were related to PCA1 and PCA2 except exchangeable Mg (EMGX), CEC of soil (CECSOL), and total exchangeable bases (EXBX). Similarly, PCA7 to PCA10 featured variables that were represented by previous PCA axes except absolute depth to bedrock (BDTICM). Exchangeable acidity (EACKCL), exchangeable aluminum (EALKCL), and electrical conductivity (ECN) were among soil properties which did not exhibit a strong relationship with any of the ten most important PCA axes, suggesting they were of less importance in defining yam growing environments.

The major edaphic factors (texture, soil pH, total nitrogen etc.) that featured prominently for the delineation of yam environments in West Africa are much as had been expected as soil characteristics have been shown to be a very important factor for yam cultivation. Texture is one of the most important properties of a soil, and it greatly affects crop production, land use, and management. Soil texture is directly related to nutrient retention and drainage capabilities (Brady and Weil 2007). Yam crop requires fertile soils rich in organic matter with good drainage and moderate water retention capacity. Yields are also observed to be higher in sandy loam and silty loam soils with higher levels of organic matter. Although yam cultivars are relatively drought-tolerant they require adequate rainfall evenly distributed throughout their growing period, particularly from 14 to 20 weeks after planting when tuber bulking occurs rapidly. According to Eruola et al., (2012), moisture variability during growth appears to be the critical factor for yam production in the humid tropics. Yam cultivars also require temperatures ranging from 25 to 30 °C and are sensitive to length of daylight periods A short daylight period promotes tuber formation and a long daylight one favors vine growth.

Of the seven principal components that accounted for 96.7% of the total variability attributable to the original 23 climatic variables, 56% was explained by PCA1 (Table 2). Annual sunshine hours (Sunhr, 0.35), temperature seasonality (BIO4, 0.34), precipitation seasonality (BIO15, 0.33), potential evapo-transpiration (PET, 0.33), and temperature annual range (BIO7, 0.3) were the dominant contributors to the variance explained by PCA1. PCA2 represented about 16.1% of the variability and the axis was strongly related to precipitation and its derivatives. It was dominated by precipitation of wettest quarter (BIO16), precipitation of wettest month (BIO13), precipitation of coldest quarter (BIO19), and precipitation seasonality (BIO15). Interestingly, PCA3 which accounted for about 10% of the total variation was clearly dominated by temperature related variables (mean temperature of driest quarter (BIO9), minimum temperature of coldest month (BIO6), mean temperature of coldest quarter (BIO11), annual mean temperature (BIO1), and mean temperature of warmest quarter (BIO10)). PCA4 accounting for 6% of the variability was also dominated mainly by precipitation variables (BIO19 and BIO13) except annual solar radiation. PCA5 to PCA7 accounted for a total of 8% of the variability and were related to variables that had featured in the previous axes except isothermality (BIO3), precipitation of warmest quarter (BIO18), and mean temperature diurnal range (BIO2). Numerous authors have observed that rainfall and temperature are very crucial to the growth and development of crops (Bannayan et al., 2011, Milošević et al., 2015, Nouri et al., 2017).

This confirms the indicated relative importance of climatic parameters related to the amount and distribution of precipitation as well as temperature observed in this study.

PCA1 to PCA3 explained over 98% of the variability attributable to 18 original remote sensing layers. PCA1 accounted for about 89% of the total variation and was dominated by five remote sensing variables composed mainly of layers from SPOT and MODIS bands. Fraction of absorbed photosynthetically active radiation (SPAR) and fraction of green vegetation cover (SCOVER) from SPOT earth observation satellite are among the five. Others are average white-sky NIR albedo (MWNA), average black-sky NIR albedo (MBNA), and average black-sky SWIR albedo (MBSA) from MODIS satellite. PCA2 explained about 8% of variability and included all dominating variables in PCA1 except an additional albedo band (average black-sky VIS albedo (MBVA)) from MODIS satellite. Vegetation indices (NDVI and EVI) dominated the last significant PCA3. Other remote sensing attributes such as gross primary productivity (GPP), net primary productivity (NPP), land surface temperature for both day and night (LSTD and LSTN), and MODIS reflectance bands 1, 2, 3, and 7 (MREFB1 to MREFB7) did not exhibit a strong relationship with PCA1 to PCA3 suggesting that they are of less importance in characterizing the yam growing environment of West Africa. However, the importance of spectral vegetation indices such as NDVI and EVI, as indicator of crop performance is well documented in literature. For instance, Gusso et al., (2014) found good correlation between EVI and crop yield in Brazil while Kogan et al., (2012) concluded that NDVI is a cumulative indicator of crop growth in Kansas, USA. These observations from other researchers confirm the relative importance of satellite derived vegetation related products in delineating yam growing environment observed in this study.

The first three components (PCA1 to PCA3) explained about 99.8% of the original five topographic data (Table 2). PCA1 accounted for about 92% of the variability and was dominated by aspect while PCA2 represented about 5% and was dominated by topographic wetness index (TWI) and elevation. PCA3 accounted for about 2.5% of the total variation and was still dominated by elevation and TWI. Soil slope did not appear with strong loadings in the three most important PCA axes probably because soil slope had been captured by terrain aspect which featured significantly in PCA1. Aspect and slope are known to be related as can be seen from its description. Aspect describes the direction in which a slope faces and relates to the degree of solar exposure and it is known to affect soil properties significantly. Begum et al., (2010) found that soils of the north-facing slope had higher SOC content, moisture, faunal abundance and diversity. They also concluded that variations due to topographic aspect induced varied microclimates and caused differences in soil temperature and soil fertility. Moreover, Geroy et al., (2011) observed that soil porosity, soil organic matter and silt content were affected by soil aspect and each contributed to greater soil water retention. The crop also usually performs well at low and medium elevation up to 900 masl making elevation an important factor contributing to the delineation of a yam environment as observed in this study.

## Cluster Analysis

### *Environmental Clustering Analysis for West Africa*

Hierarchical clustering algorithm was employed whereby sum of squared Euclidean distance is minimized as proposed by Ward (1963). The result of this is presented in the dendrogram (Figure 2). Seven distinct clusters were identified for the entire region using the cutoff of  $R^2=0.7$  (70% dissimilarity). These are presented as seven yam mega-environments (Figure 3 and Table 3), hereby referred to as target set of environments (TSE1 to TSE7). TSE1 occupied about 8.1 million ha (6.7%) and occurred uniquely in the humid forest ecology of Nigeria while TSE2 (11 million ha) featured mainly at the coastal part of Ghana and Côte d'Ivoire with some fractional appearance in the western part of coastal Nigeria. TSE3, TSE4, and TSE6 respectively occupied land areas of about 23.4, 32.3, and 26.4 million ha and featured across the entire region, suggesting they might be useful in the dissemination of advanced yam varieties in West Africa.

Incidentally, the three mega-environments are responsible for about 70% of the current yam production in the study area (Table 3). Hence, selection and cultivar testing performed in the three



Figure 3. West Africa showing results of characterization of yam growing area into seven mega-environments

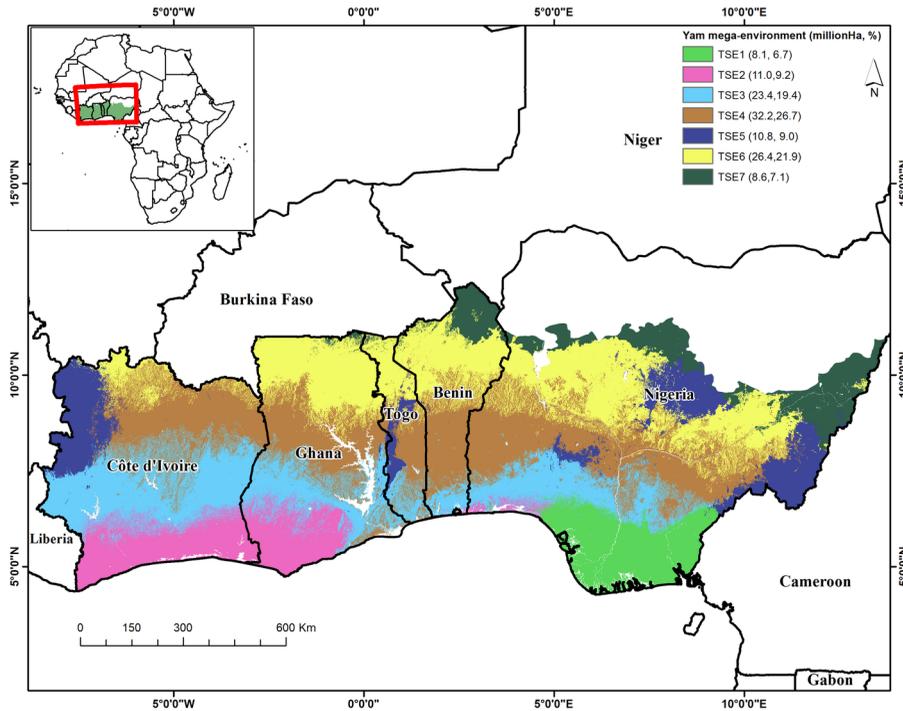


Table 3. Yam environmental clusters by area and percentage for West Africa

Country	TSE1	TSE2	TSE3	TSE4	TSE5	TSE6	TSE7	Total (Country)
<b>Area (million Ha) Under Environmental Cluster</b>								
Bénin	0.00	0.03	0.80	3.71	0.02	5.03	1.85	11.44
Côte d'Ivoire	0.00	6.27	11.02	8.41	4.20	1.87	0.00	31.77
Ghana	0.00	4.33	5.33	6.40	0.09	6.63	0.17	22.94
Nigeria	8.12	0.42	6.23	13.68	6.53	12.89	6.57	54.43
Grand Total	8.12	11.04	23.37	32.19	10.84	26.43	8.59	120.59
<b>Percentage (%) of Area Under Environmental Cluster</b>								
Bénin	0.0	0.3	7.0	32.4	0.1	44.0	16.2	9.5
Côte d'Ivoire	0.0	19.7	34.7	26.5	13.2	5.9	0.0	26.3
Ghana	0.0	18.9	23.2	27.9	0.4	28.9	0.7	19.0
Nigeria	14.9	0.8	11.4	25.1	12.0	23.7	12.1	45.1
Grand Total	6.7	9.2	19.4	26.7	9.0	21.9	7.1	100.0
Yam production (million tons)	6.5	2.1	10.5	11.8	2.2	9.6	2.4	45.2
% Yam production	14.3	4.6	23.2	26.2	4.9	21.2	5.4	100.0

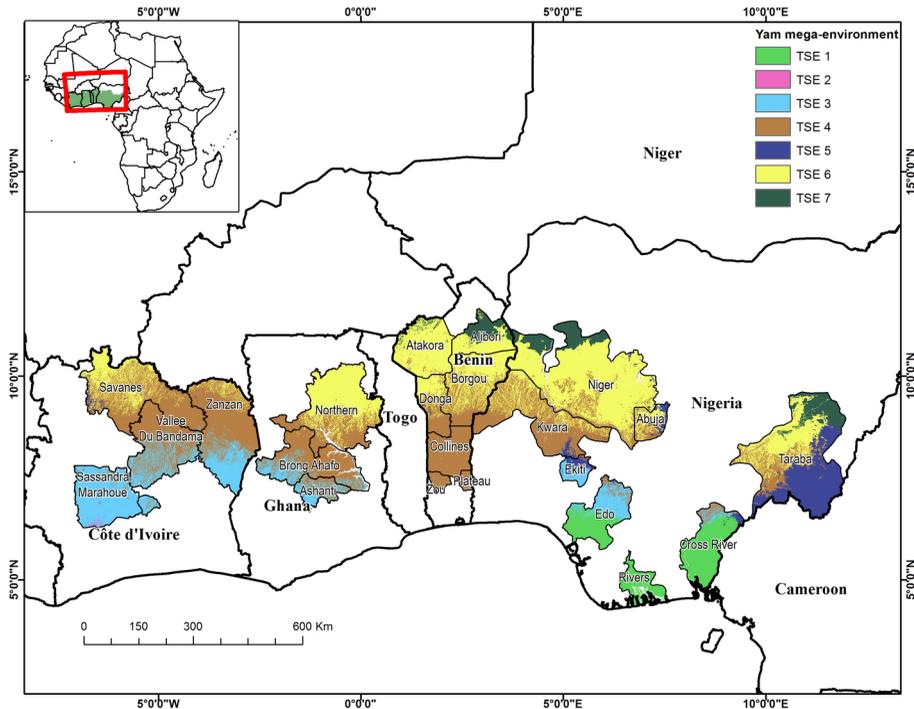
Table 4. Yam environmental clusters by area and percentage for Nigeria

State	TSE1	TSE2	TSE3	TSE4	TSE5	TSE6	TSE7	Total (State)
	Area (*000 Ha) Under Environmental Cluster							
Cross River	1,679	0.04	157	156	70	-	-	2,061
Edo	996	2.81	804	141	13	-	-	1,956
Ekiti	-	-	319	30	174	-	-	523
FCT	-	-	0.03	229	127	378	-	734
Kwara	-	-	9.72	2,512	94	919	-	3,534
Niger	-	-	4.69	704	37	5,324	874	6,943
Rivers	739	0.02	-	-	-	-	-	739
Taraba	0.01	-	1.93	930	2,454	1,803	806	5,994
Target yam area	3,414	2.87	1,296	4,701	2,968	8,424	1,679	22,485
Country yam area	8,116	415	6,225	13,677	6,534	12,893	6,567	54,427
State	Percentage (%) of Area Under Environmental Cluster							
Cross River	81.5	0.0	7.6	7.5	3.4	-	-	9.2
Edo	50.9	0.1	41.1	7.2	0.7	-	-	8.7
Ekiti	-	-	61.0	5.7	33.3	-	-	2.3
FCT	-	-	0.0	31.2	17.3	51.5	-	3.3
Kwara	-	-	0.3	71.1	2.6	26.0	-	15.7
Niger	-	-	0.1	10.1	0.5	76.7	12.6	30.9
Rivers	100.0	0.0	-	-	-	-	-	3.3
Taraba	0.0	-	0.0	15.5	40.9	30.1	13.4	26.7
Target yam area	15.2	0.0	5.8	20.9	13.2	37.5	7.5	100.0
Country yam area	14.9	0.8	11.4	25.1	12.0	23.7	12.1	100.0
Yam production (million tons)	6.48	0.23	7.06	8.63	2.08	6.98	2.37	33.82
% Yam production	19.1	0.7	20.9	25.5	6.1	20.6	7.0	100.0

### Environmental Clustering Analysis for Nigeria

Results of multivariate clustering analysis for Nigeria are shown in Table 4 and Figure 4. The yam production areas are clustered into six mega-environments or TSE covering the seven States with a total land area of about 22 million ha. TSE 1 is obviously a cluster of environments within the forest ecology as it extended from Rivers, Edo, and Cross River States and covered about 15% of the total yam growing belt of Nigeria. TSE 3 is the next cluster with significant occurrence and occupied about 1.3 million ha representing about 5.8% of the total target area. TSE 4 is another important cluster, being the second largest occupied in terms of area (4.7 million ha) after TSE 6 within the study area. It covered a major part of Kwara and extends throughout the Middle Belt to the southern part of Taraba. TSE 4 covered a total of 13.7 million ha in the yam growing environment and accounts for about 25.5% of the current yam production. This implies that about 26% of the total number of variety testing trials in Nigeria should be conducted in this cluster distributed across States according to their respective proportion. TSE 6 occupied the largest area of about 8.4 million ha within the target States and represents about 12.8 million ha of the entire yam growing area of the country. Breeding trials carried out in this environment will be representative of about 24% of the total yam area in the

Figure 4. West Africa showing results of clustering analysis within the AfricaYam project's' target regions/districts



country. Current yam production within this TSE is equally high at about 7 million t (21%). TSE 7 is situated in the northernmost part of Niger and Taraba covering about 1.7 million ha for both States. TSE 5 represented about 13.2% of the target area and occurs mainly in the highlands of Taraba with some patches in the northern part of Ekiti. This environment will be ideal for testing yam varieties for adaptation to high altitude ecologies with cool temperatures.

Dominant environmental clusters varied by State (Table 4). For instance, the dominant environments in Cross River were represented in TSE 1 which accounted for over 80% of that State while the same cluster was completely dominant in Rivers. Similarly, TSE 3 was prominent in Edo and Ekiti, respectively covering over 41 and 61% of the States. TSE 4 was most important in FCT, Abuja, and Kwara. Niger exhibited the prominent environmental clusters TSE 6 and TSE 7 while Taraba was the only State where TSE 4 -TSE 7 were predominant clusters. These results could be used to strategically locate and conduct yam trial sites/selection environments and resource allocation for each target State in the country. They are equally useful in ensuring that experiments are conducted in representative locations within the TSEs which will facilitate scaling out yam breeding results for wider impact. Estimates of total yam production were lowest in TSE 2 at about 0.23 million t and highest at TSE 4 (8.6 million t, Table 4).

#### Environmental Clustering Analysis for Bénin

The results of analysis of environmental characterization of yam growing area of Bénin are shown in Figure 4 and Table 5. The yam target set of environments comprises the seven *départements* of Alibori, Atakora, Borgou, Collines, Donga, Plateau, and Zou with a total area of about 8.8 million ha. The most extensive environmental cluster was TSE 6 with an area of about 4.9 million ha (55%) followed by TSE 4 covering about 37% of the whole target area. Within the yam target area TSE 1- TSE 3 and TSE 5 were non-existent while only TSE 4 and TSE 6 were the two most prominent

Table 5. Yam environmental clusters by area and percentage for Bénin

Département	TSE1	TSE2	TSE3	TSE4	TSE5	TSE6	TSE7	Total (Département)
	Area ('000 ha) Under Environmental Cluster							
Alibori	-	-	-	-	-	761	555	1,316
Atakora	-	-	-	30	6.38	1,915	112	2,064
Borgou	-	-	0.04	901	0.99	1,658	-	2,560
Collines	-	-	5.96	1,395	0.06	0.06	-	1,401
Donga	-	-	0.58	573	9.86	525	-	1,109
Plateau	-	-	23.81	156	-	-	-	180
Zou	-	-	7.58	211	-	-	-	219
Target yam area	-	-	37.98	3,267	17.28	4,859	668	8,848
Country yam area	0.05	29.83	801	3,711	16.62	5,032	1,852	11,443
Département	Percentage (%) of Area Under Environmental Cluster							
Alibori	-	-	-	-	-	57.8	42.2	14.9
Atakora	-	-	-	1.5	0.3	92.8	5.4	23.3
Borgou	-	-	0.0	35.2	0.0	64.8	-	28.9
Collines	-	-	0.4	99.6	0.0	0.0	-	15.8
Donga	-	-	0.1	51.7	0.9	47.3	-	12.5
Plateau	-	-	13.2	86.8	-	-	-	2.0
Zou	-	-	3.5	96.5	-	-	-	2.5
Target yam area	-	-	0.4	36.9	0.2	54.9	7.5	100.0
Country yam area	0.0	0.3	7.0	32.4	0.1	44.0	16.2	100.0
Yam production (tons)	-	-	0.04	0.84	0.00	1.24	0.06	2.19
% Yam production	-	-	2.0	38.4	0.1	56.6	3.0	100.0

environmental clusters covering about 92% of the area. TSE 7 occurred at the northern part of the target zone with an area of 0.6 million ha and represents about 16% of the yam producing belt of the country. The important yam environment clusters varied by *départements*. For example, only TSE 6 and TSE 7 occurred in Alibori while TSE 4 and TSE 6 were predominant in Borgou and Donga. TSE 4 was the dominant cluster in Zou, Collines, and Plateau with respective percentages of occurrence at 96.5, 99.6, and 86.5.

### Environmental Clustering Analysis for Côte d'Ivoire

The results of yam environment characterization using spatial multivariate analysis in Côte d'Ivoire are shown in Figure 4 and Table 6. Two dominant clusters that emerged are TSE 3 and TSE 4, representing respectively 33% and 53% of the target yam area of the current breeding operation in the country. Although TSE 2 and TSE 5 occurred in insignificant proportions within the target study area, they are representative of about one-third (33%) of the entire yam growing area in the country (Table 6). Further examination of environmental cluster distribution among the regions in the target area depicted that only four regions - Savanes, Vallée du Bandama, Zanzan, and Sassandra Maraho - had significant representation. Two environmental clusters, TSE 2 and TSE 4, were predominant in Zanzan and Savanes regions while Sassandra Maraho accounted for the significant size of cluster TSE 3 (94%). TSE 2 which was situated mainly in the southern parts of Ghana and Côte d'Ivoire

Table 6. Yam environmental clusters by area and percentage for Côte d'Ivoire

Regions	TSE1	TSE2	TSE3	TSE4	TSE5	TSE6	TSE7	Total (Regions)
	Area (*000 ha) Under Environmental Cluster							
Savanes	-	-	5.0	2,422.1	81.3	1,476.6	1.4	3,986
Vallée Du Bandam	-	-	939.3	1,870.1	0.3	6.6	-	2,816
Zanzan	-	-	1,051.9	2,519.6	3.0	229.4	-	3,804
District Autonom	-	-	171.8	32.0	-	-	-	204
Sassandra Maraho	-	33.4	2,219.5	103.2	4.7	-	-	2,361
Target Yam Area	-	33.4	4,387.5	6,946.8	89.3	1,712.6	1.4	13,171
Country yam area	1.2	6,272.2	11,016.0	8,407.3	4,203.8	1,869.3	2.1	31,772
	Percentage (%) of Area Under Environmental Cluster							
Savanes	-	-	0.1	60.8	2.0	37.0	0.0	30.3
Vallée Du Bandam	-	-	33.4	66.4	0.0	0.2	-	21.4
Zanzan	-	-	27.7	66.2	0.1	6.0	-	28.9
District Autonom	-	-	84.3	15.7	-	-	-	1.5
Sassandra Maraho	-	1.4	94.0	4.4	0.2	-	-	17.9
Target yam area	-	0.3	33.3	52.7	0.7	13.0	0.0	100.0
Country yam area	0.0	19.7	34.7	26.5	13.2	5.9	0.0	100.0
Total yam production (million tons)	-	1.7	2.4	0.8	0.1	0.2	-	5.2
% Yam production	-	32.8	46.6	15.1	2.5	3.0	-	100.0

seemed not well captured within the current target area for breeding operations and was represented by only 0.3% of the current yam breeding research targets. TSE 2 is an important environment for yam cultivation as 33% of total yam production in Côte d' Ivoire takes place within this cluster, indicating under-representation and gaps when study area selection was made. Our result of the yam environment clustering could serve as a useful tool to strategically represent and locate the testing sites for yam breeding operations for the country.

### *Environmental Clustering Analysis for Ghana*

Environmental cluster analysis for Ghana showed that about 7.1 million ha are within the current target areas of the breeding program (Table 7). Out of the seven clusters only three had significant proportions within the current target yam area of the breeding program. These are TSE 3, TSE 4, and TSE 6 with 15, 51, and 34%, of occurrence respectively. Although TSE 3 covered an area of 1.01 million ha within the target study area, it is representative of 5.3 million ha in the entire yam growing region. This implies that breeding trials conducted in TSE 3 will be representative of a larger area within Ghana's yam growing ecologies as well as the West African region. Similarly, experiments conducted within TSE 4 (Table 7) will be representing a larger area of over 6.4 million ha in the yam growing region. More detailed assessment of the distribution of environmental clusters among regions in the target area revealed that about 89% of clusters occurred within Brong Ahafo region (35%) and Northern region (53.6%), leaving only 11% for Ashanti. In Ashanti and Brong Ahafo, only TSE 3 and TSE 4 occurred in significant proportions while TSE 6 featured prominently in Northern region. Although TSE 2 occupied a significant area of about 4.3 million ha (18.9%) of the entire yam growing area, it was not well represented in the project target area, probably owing to its low

Table 7. Yam environmental clusters by area and percentage for Ghana

Region	TSE1	TSE2	TSE3	TSE4	TSE5	TSE6	TSE7	Total (Regions)
	Area ('000 Ha) Under Environmental Cluster							
Ashanti	-	3.1	461	333	1.2	-	-	799
Brong Ahafo	-	-	617	1,859	0.1	13.3	-	2,489
Northern	-	-	1.2	1,386	3.1	2,411	-	3,801
Yam target area	-	3.1	1,079	3,578	4.4	2,424	-	7,089
Country yam area	0.1	4,327	5,332	6,396	88.6	6,634	167	22,944
	Percentage (%) of Area Under Environmental Cluster							
Ashanti	-	0.4	57.7	41.7	0.1	-	-	11.3
Brong Ahafo	-	-	24.8	74.7	0.0	0.5	-	35.1
Northern	-	-	0.0	36.5	0.1	63.4	-	53.6
Yam target area	-	0.0	15.2	50.5	0.1	34.2	-	100.0
Country yam area	0.0	18.9	23.2	27.9	0.4	28.9	0.7	100.0
Yam production (million t)	-	0.15	0.96	1.59	0.01	1.20	0.01	3.92
Yam production (%)	-	3.9	24.4	40.5	0.3	30.7	0.2	

importance in yam production. TSE 2 accounted for only about 3.9% of the total yam production in the country. This is not unexpected as it is predominantly an area where cultivation of cocoa and other tree crops is predominant.

### Socioeconomic Analysis of Yam Growing Zones of Target Countries

The results of socioeconomic analysis of the yam producing areas of the target countries are presented in Table 8. A quick overview of these results shows that about 181 million people currently live in the yam zones with about 68% living below poverty levels at \$2 a day. The projected population of the region for 2050 is 366 million. Yam production in the entire region is about 45 million t with over 33 million t (75%) produced in Nigeria. Among the target countries, the development indicator of poverty varied substantially with Bénin having the worst poverty level of 86% and Côte d'Ivoire the lowest, at about 45%. Nigeria ranked second for levels of poverty in the yam producing regions (79%) while Ghana ranked third (57%). Poverty levels varied slightly within yam growing clusters of the target countries. For instance, TSE 5 and TSE 7 of Bénin displayed the highest level of poverty incidence at about 99% while TSE 2 which exhibited the lowest was 72%. In Ghana, poverty level was highest at 78 and 82% in the two northernmost clusters (TSE 6 and TSE 7) while it became low at the southernmost clusters of environments, TSE 2 to TSE 4. Poverty incidence in these countries revealed an increasing gradient from the south to the north. This is probably due to decreasing agricultural potential from the coast to the hinterland as agriculture is the mainstay of the populace in the region. Moreover, this trend follows the annual rainfall gradient which decreases as latitude increases, and consequently production potential equally declines since most agricultural activities are rain-fed.

Total yam production computed from SPAM 2005 Version 3 (You et al., 2014) varied substantially among the project countries with Nigeria leading with 33.8 million t owing mainly to larger areas of cultivation. Total production in the other three target countries was about 11 million t with Côte d'Ivoire producing 5.2 million t, Ghana 3.9 million t, and Bénin the least, 2.2 million t (Table 8). Average yam yield in the target area was about 9.3 t/ha. Although yam production varied significantly

**Table 8. Socioeconomic characteristics of the target clusters of yam production environment in the four West African countries**

	Total Population	No. of Poor People Living Below \$2/Day	Poor People Living Below \$2/Day (%)	Projected Population in 2050	Total Yam Production (t)	Mean Yam Yield (t/ha)	Market Access (Mean Hour of Travel to Towns > 20,000 Inhabitants)
<b>Benin</b>							
TSE 1	-	-	-	-	-	-	-
TSE 2	304,830	218,716	71.75	1,042,579	-	-	0.9
TSE 3	3,535,249	2,686,059	75.98	4,956,610	43,032	5.8	2.3
TSE 4	3,030,291	2,451,580	80.90	5,159,076	840,994	8.3	2.9
TSE 5	4,603	4,557	99.00	12,137	1,838	12.2	1.9
TSE 6	2,398,018	2,190,876	91.36	3,924,065	1,239,354	11.5	3.5
TSE 7	674,227	667,049	98.94	1,042,276	64,677	10.6	5.8
Bénin Total/Mean	9,947,219	8,218,838	86.32	16,136,743	2,189,895	8.4	2.9
<b>Côte d'Ivoire</b>							
TSE 1	57	-	-	5,575	-	-	-
TSE 2	8,599,905	3,533,243	41.08	17,855,217	1,710,063	8.8	5.3
TSE 3	8,837,875	3,565,882	40.35	15,839,021	2,430,641	8.0	3.4
TSE 4	2,837,158	1,060,552	37.38	4,400,581	786,318	7.3	3.1
TSE 5	982,751	438,125	44.58	1,893,518	131,838	6.8	3.2
TSE 6	1,047,842	320,780	30.61	1,292,984	157,889	6.6	2.7
TSE 7	917	673	73.38	6,623	-	-	-
Côte d'Ivoire Total/Mean	22,306,506	8,919,255	44.56	41,293,519	5,216,748	7.7	3.6
<b>Ghana</b>							
TSE 1	-	-	-	-	-	-	-
TSE 2	6,387,837	2,378,594	37.24	11,264,383	153,520	9.4	2.7
TSE 3	10,938,677	4,311,761	39.42	17,980,772	955,316	7.5	3.1
TSE 4	4,741,111	1,813,659	38.25	6,238,857	1,589,787	8.6	3.8
TSE 5	34,996	22,568	64.49	66,957	10,370	10.0	3.4
TSE 6	3,634,132	2,983,489	82.10	6,032,148	1,204,019	9.8	3.1
TSE 7	478,093	373,996	78.23	378,305	8,851	9.3	2.5
Ghana Total/Mean	26,214,847	11,884,067	56.62	41,961,422	3,921,863	8.9	3.2
<b>Nigeria</b>							
TSE 1	36,050,893	27,790,799	77.09	79,558,130	6,475,538	10.0	3.8
TSE 2	1,618,627	1,320,723	81.60	7,455,192	233,886	7.0	3.4
TSE 3	29,251,122	21,785,165	74.48	65,615,067	7,055,282	10.9	2.5
TSE 4	26,770,892	20,060,252	74.93	50,498,141	8,630,461	11.7	3.6
TSE 5	7,230,050	5,569,116	77.03	17,601,654	2,078,062	9.3	5.0
TSE 6	12,456,182	9,836,577	78.97	23,641,143	6,981,449	11.7	3.6
TSE 7	9,483,191	8,450,674	89.11	21,956,086	2,368,779	9.7	3.6
Nigeria Total/Mean	122,860,957	94,813,306	79.03	266,325,414	33,823,457	10.5	3.6
Target countries / Total /mean	181,329,528	123,835,466	68	365,717,097	45,151,963	9.3	3.5

among the countries and within yam growing clusters, yield varied only slightly within the target area. Average yield was highest in Nigeria at 10.5 t/ha, followed by Ghana (9.0 t/ha) while it was lowest in Côte d'Ivoire at 7.7 t/ha. From this analysis, average yam yield of 9.3 t/ha in the region is below the attainable yield of 50 t/ha (FAO, 2014). Hence the AfricaYam project seeks to bridge the yield gap and produce varieties suitable for sustainable intensification and higher productivity.

Access to markets for inputs and outlets for production is very important for agricultural development. Poor access can minimize farmers' productivity and profitability (Ulimwengu et al., 2009). Results of market access analysis using travel time to settlements with populations above 20,000 showed that average travel time in the entire region was 3.5 hours. According to HarvestChoice (2015) those places with a travel time less than 2 hour are classified as having good market access; and those with 2-4 hours of travel have medium access; those with travel time longer than 4 hours have poor access. It is clear from Table 8 that accessibility was generally medium in the whole target area with some variations among countries and clusters. Yam production environment TSE 5 of Benin Republic was found to have the best market access - about 1.9 hours - while the production domain with the worst access - a mean travel time of about 5.3 hours - was found in TSE 2 in Côte d'Ivoire where about 33% of the yam is being produced. Accessibility was probably worst due to high forest ecology of TSE 2. The problem of market accessibility appeared more pronounced in production environment TSE 5 in Nigeria with a mean travel time to markets of about 5 hours. This may be probably due to the high-altitude environment of TSE 5. Our results suggest that poor market accessibility in the forest zone and high-altitude environments may affect the effectiveness and impact of the AfricaYam project's breeding program.

## CONCLUSION AND RECOMMENDATION

Most often trial sites for varietal testing are chosen based on convenience and ease of access. Cluster maps provide unbiased guides for site selection for varietal testing to optimally represent the target set of production environments. Results of our analysis suggest that the current breeding target areas of the yam improvement programs are truly representative of environmental characteristics in over 90% of the yam growing regions of Nigeria and Bénin but not well represented in Ghana and Côte d'Ivoire. This analysis enables us to discover where gaps exist in breeding programs.

This yam production environment mapping also further highlighted the presence of other potential areas less targeted by current breeding efforts in the four West African countries where yam variety development or testing could help match the varieties onto the target environments through scaling up and out the research findings for wider impact. Moreover, the clustering results can be used to optimize varietal testing and selection programs within and across the target countries. The result of the socioeconomic analysis showed possible impact on poverty reduction in the target area and suggested problems of access to market in the production zones.

## ACKNOWLEDGMENT

This work is an output of the AfricaYam Project, funded by the International Institute of Tropical Agriculture (IITA), Ibadan, Nigeria, through a research grant (OPP1052998) received from the Bill and Melinda Gates Foundation. The authors gratefully acknowledge the support from both organizations. The contributions of the IITA yam breeding team, and the AfricaYam partners in CNRA, Côte d'Ivoire, UAC, Benin, NRCRI, Nigeria, EBSU, Nigeria, SARI and CSRI, Ghana for data collection and analyses are highly appreciated.

## REFERENCES

- Africa Soil Information Service (AfsIS). (2017). AfsIS MODIS Collection: Land Surface Temperature, December 2017, version 6 Release. Palisades, NY: Center for International Earth Science Information Network (CIESIN), Columbia University. Retrieved from <ftp://africagrids.net/1000m/MYD11A2/Version006/>
- Araya, A., Keesstra, S. D., & Stroosnijder, L. (2010). A new agro-climatic classification for crop suitability zoning in northern semi-arid Ethiopia. *Agricultural and Forest Meteorology*, *150*(7-8), 1057–1064. doi:10.1016/j.agrformet.2010.04.003
- Arshad, M. A., Lowery, B., & Grossman, B. (1996). Physical tests for monitoring soil quality. In J. W. Doran & A. J. Jones (Eds.), *Methods for Assessing Soil Quality*. Madison, WI: ACSESS Digital Library.
- Asiedu, R., & Sartie, A. (2010). Crops that feed the world 1. Yams: Yams for income and food security. *Food Security*, *2*(4), 305–315. doi:10.1007/s12571-010-0085-0
- Bannayan, M., Lotfabadi, S. S., Sanjani, S., Mohamadian, A., & Aghaalikhani, M. (2011). Effects of precipitation and temperature on crop production variability in northeast Iran. *International Journal of Biometeorology*, *55*(3), 387–401. doi:10.1007/s00484-010-0348-7 PMID:20706741
- Begum, F., Bajracharya, R. M., Sharma, S., & Sitaula, B. K. (2010). Influence of slope aspect on soil physico-chemical and biological properties in the mid-hills of central Nepal. *International Journal of Sustainable Development and World Ecology*, *17*(5), 438–443. doi:10.1080/13504509.2010.499034
- Boke-Olén, N., Abdi, A. M., Hall, O., & Lehsten, V. (2017). High-resolution African population projections from radiative forcing and socio-economic models, 2000 to 2100. *Scientific Data*, *4*, 160130. doi:10.1038/sdata.2016.130 PMID:28094785
- Brady, N. C., & Weil, R. R. (2007). *The Nature and Properties of Soils* (14th ed.). Upper Saddle River, NJ: Prentice Hall.
- Brock, G., Pihur, V., Datta, S., & Datta, S. (2008). CValid: An R Package for Cluster Validation. *Journal of Statistical Software*, *25*(4), 1–22. doi:10.18637/jss.v025.i04
- Bunce, R. G. H., Carey, P. D., Elena-Roselló, R., Orr, J., Watkins, J. W., & Fuller, R. (2002). A comparison of different biogeographical classifications of Europe, Great Britain and Spain. *Journal of Environmental Management*, *65*(2), 121–134. doi:10.1006/jema.2002.0533 PMID:12197075
- Caldiz, D. O., Haverkort, A. J., & Struik, P. C. (2002). Analysis of a complex crop production system in interdependent agro-ecological zones: A methodological approach for potatoes in Argentina. *Agricultural Systems*, *73*(3), 297–311. doi:10.1016/S0308-521X(01)00085-3
- Ceccarelli, S., & Grando, S. (2007). Decentralized-Participatory Plant Breeding: An Example of Demand Driven Research. *Euphytica*, *155*(3), 349–360. doi:10.1007/s10681-006-9336-8
- Conrad, O., Bechtel, B., Bock, M., Dietrich, H., Fischer, E., Gerlitz, L., & Böhner, J. et al. (2015). System for Automated Geoscientific Analyses (SAGA) v. 2.1.4. *Geoscientific Model Development*, *8*(7), 1991–2007. doi:10.5194/gmd-8-1991-2015
- Cooper, M. (1999). Concepts and strategies for plant adaptation research in rainfed lowland rice. *Field Crops Research*, *64*(1-2), 13–34. doi:10.1016/S0378-4290(99)00048-9
- Costantini, E. A. C., & Barbetti, R. (2008). Environmental and visual impact analysis of viticulture and olive tree cultivation in the province of Siena (Italy). *European Journal of Agronomy*, *28*(3), 412–426. doi:10.1016/j.eja.2007.11.012
- Costantini, E. A. C., Lorenzetti, R., & Malorgio, G. (2016). A multivariate approach for the study of environmental drivers of wine economic structure. *Land Use Policy*, *57*, 53–63. doi:10.1016/j.landusepol.2016.05.015
- Demšar, U., Harris, P., Brunson, C., Fotheringham, A. S., & McLoone, S. (2013). Principal Component Analysis on Spatial Data: An Overview. *Annals of the Association of American Geographers*. doi:10.1080/0045608.2012.689236

- Dixon, J., Gulliver, A., & Gibbon, D. (2001). *Farming Systems and Poverty: Improving Farmers' Livelihoods in a Changing World*. Rome and Washington DC, USA: FAO and World Bank.
- FAO Ecocrop. (2013). Ecocrop database. FAO, Rome, Italy. Retrieved from <http://ecocrop.fao.org/ecocrop/>
- ECOWAS-SWAC/OECD. (2007). The Atlas on Regional Integration in West Africa. Population series. Demographic trends. 16p <http://www.oecd.org/swac/publications/39802965.pdf>
- Eruola, A. O., Bello, N. J., Ufoegbune, G. C., & Makinde, A. A. (2012). Application of Rainfall-Potential Evapotranspiration Model for Determining Optimum Planting Date of Yam (*Dioscorea rotundata*) in a Tropical Wet-and-Dry Climate. *International Journal of Plant Research*, 2(2), 36–40. doi:10.5923/j.plant.20120202.07
- ESRI. (2017). *Copyright 1999-2017. ArcGIS 10.5 Computer software*. Redlands, CA: ESRI.
- FAO. (2014). Food and Agriculture Organization of the United Nations. Retrieve dfrom [www.fao.org/statistics/en/](http://www.fao.org/statistics/en/) FAO, Rome. Italy
- Fick, S.E. &R.J. Hijmans, (2017). Worldclim 2: New 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*.
- Fischer, G., Shah, M., Tubiello, F. N., & Van Velhuizen, H. (2005). Socio-economic and climate change impacts on agriculture: An integrated assessment, 1990–2080. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, 360, 2067–2083. doi:10.1098/rstb.2005.1744 PMID:16433094
- Fodor, I. K. (2002). *A survey of dimension reduction techniques. LLNL Technical Report*. Livermore, CA: Lawrence Livermore National Laboratory. Retrieved from <http://www.llnl.gov/CASC/sapphire/pubs/148494.pdf> doi:10.2172/15002155
- García, L. J. C., Posada-Suárez, H., & Läderach, P. (2014). Recommendations for the regionalizing of coffee cultivation in Colombia: A methodological proposal based on agro-climatic indices. *PLoS One*, 9(12), e113510. doi:10.1371/journal.pone.0113510 PMID:25436456
- Gavioli, A., de Souza, E. G., Bazzi, C. L., Guedes, L. P. C., & Schenatto, K. (2016). Optimization of management zone delineation by using spatial principal components. *Computers and Electronics in Agriculture*, 127, 302–310. doi:10.1016/j.compag.2016.06.029
- Geerts, S., Raes, D., Garcia, M., Del Castillo, C., & Buytaert, W. (2006). Agro-climatic suitability mapping for crop production in the Bolivian Altiplano: A case study for quinoa. *Agricultural and Forest Meteorology*, 139(3-4), 399–412. doi:10.1016/j.agrformet.2006.08.018
- Geroy, I. J., Gribb, M. M., Marshall, H. P., Chandler, D. G., Benner, S. G., & Mcnamara, J. P. (2011). Aspect influences on soil water retention and storage (2011). *Hydrological Processes*, 25(25), 3836–3842. doi:10.1002/hyp.8281
- Gordon, A. (1996). *Null models in cluster validation*. In W. Gaul & D. Pfeifer (Eds.), *From data to knowledge* (pp. 32–44). New York: Springer. doi:10.1007/978-3-642-79999-0\_3
- Goswami, R., Chatterjee, S., & Prasad, B. (2014). Farm types and their economic characterization in complex agro-ecosystems for informed extension intervention: Study from coastal West Bengal India. *Agricultural and Food Economics*, 2(1), 1–24. doi:10.1186/s40100-014-0005-2
- Gusso, A., Ducati, J., Veronez, M. R., Sommer, V., & da Silveira, L. G. Jr. (2014). Monitoring heat waves and their impacts on summer crop development in southern Brazil. *Agricultural Sciences*, 5(04), 353–364. doi:10.4236/as.2014.54037
- Hargrove, W. W., & Hoffman, F. M. (2004). Potential of multivariate quantitative methods for delineation and visualization of ecoregions. *Environmental Management*, 34(1), S39–S60. doi:10.1007/s00267-003-1084-0 PMID:15883870
- Hartkamp, A. D., White, J. W., Aguilar, R. A., Banziger, M., Srinivasan, G., Granados, G., & Crossa, J. (2000). Maize Production Environments Revisited: a GIS-Based Approach. CIMMYT & NRG, Mexico, D.F.
- HarvestChoice. (2012). “Sub-national Poverty and Extreme Poverty Prevalence.” International Food Policy Research Institute, Washington, DC., and University of Minnesota, St. Paul, MN. Available online at <http://harvestchoice.org/node/4751>

- HarvestChoice. (2015). Travel time to nearest town over 20K (mean, hours, 2000). International Food Policy Research Institute, Washington, DC., and University of Minnesota, St. Paul, MN. Retrieved from [http://harvestchoice.org/data/tt\\_20k](http://harvestchoice.org/data/tt_20k)
- Hazeu, G. W., Metzger, M. J., Múcher, C. A., Perez-Soba, M., Renetzeder, C., & Andersen, E. (2011). European environmental stratifications and typologies: An overview. *Agriculture, Ecosystems & Environment*, 142(1-2), 29–39. doi:10.1016/j.agee.2010.01.009
- Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Walsh, M. G., Shepherd, K. D., & Tondoh, J. E. et al. (2015). Mapping soil properties of Africa at 250 m resolution: Random forests significantly improve current predictions. *PLoS One*, 10(6), e0125814. doi:10.1371/journal.pone.0125814 PMID:26110833
- Hengl, T., Mendes de Jesus, J., Heuvelink, G. B. M., Ruiperez Gonzalez, M., Kilibarda, M., Blagotić, A., & Kempen, B. et al. (2017). SoilGrids250m: Global gridded soil information based on machine learning. *PLoS One*, 12(2), e0169748. doi:10.1371/journal.pone.0169748 PMID:28207752
- Hyman, G., Hodson, D., & Jones, P. (2013). Spatial analysis to support geographic targeting of genotypes to environments. *Frontiers in Physiology*, 4, 40. doi:10.3389/fphys.2013.00040 PMID:23515351
- ISRIC. (2016). Soil property maps of Africa at 250 m. Retrieved from <http://www.isric.org/data/AfSoilGrids250m>
- Jarvis, A., Reuter, H. I., Nelson, A., & Guevara, E. (2008). Hole-filled seamless SRTM data V4, International Centre for Tropical Agriculture (CIAT). Retrieved from <http://srtm.csi.cgiar.org>
- Johnson, R. A., & Wichern, D. W. (2007). *Applied Multivariate Statistical Analysis* (6th ed.). New Jersey: Pearson.
- Jolliffe, I. T., Trendafilov, N. T., & Uddin, M. (2003). A modified principal component technique based on the LASSO. *Journal of Computational and Graphical Statistics*, 12(3), 531–547. doi:10.1198/1061860032148
- Kaiser, H. F., & Rice, J. (1974). Little jiffy: Mark iv. *Educational and Psychological Measurement*, 34(1), 111–117. doi:10.1177/001316447403400115
- Kogan, F. N., Salazar, L., & Roytman, L. (2012). Forecasting crop production using satellite-based vegetation health indices in Kansas, USA. *International Journal of Remote Sensing*, 33(9), 2798–2814. doi:10.1080/01431161.2011.621464
- Köppen, W., (1900). Versuch einer Klassifikation der Klimate, vorzugsweise nach ihren Beziehungen zur Pflanzenwelt. *Geographie Zeitschrift*, 6, 657–679, 593–611.
- Kothe, S., Pfeifroth, U., Cremer, R., Trentmann, J., & Hollmann, R. (2017). A Satellite-Based Sunshine Duration Climate Data Record for Europe and Africa. *Remote Sensing*, 9(5), 429. doi:10.3390/rs9050429
- Langella, G. (2008). *Spatial analysis of pedological and environmental features by means of digital soil mapping* [PhD Thesis]. Università degli Studi di Napoli Federico II, Italy.
- Leathwick, J. R., Overton, J. M., & McLeod, M. (2003). An environmental domain classification of New Zealand and its use as a tool for biodiversity management. *Conservation Biology*, 17(6), 1612–1623. doi:10.1111/j.1523-1739.2003.00469.x
- Manning, C. D., Prabhakar, R., & Schutze, H. (2008). *Introduction to information retrieval*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511809071
- Metzger, M. J., Bunce, R. G. H., Jongman, R. H. G., Sayre, R., Trabucco, A., & Zomer, R. (2012). A high-resolution bioclimate map of the world: A unifying framework for global biodiversity research and monitoring. *Global Ecology and Biogeography*. doi:10.1111/geb.12022
- Mignouna, B. D., Abdoulaye, T., Alene, A. D., Asiedu, R., & Manyong, V. M. (2014). *Characterization of Yam-growing Households in the Project Areas of Nigeria*. International Institute of Tropical Agriculture. Ibadan, Nigeria: IITA.
- Milošević, D. D., Savić, S. M., Stojanović, V., & Popov-Raljić, J. (2015). Effects of precipitation and temperatures on crop yield variability in Vojvodina (Serbia). *Italian Journal of Agrometeorology*, 3, 35–46.

- Moral, F. J., Rebollo, F. J., Paniagua, L. L., & García-Martín, A. (2016). A GIS-based multivariate clustering for characterization and ecoregion mapping from a viticultural perspective. *Spanish Journal of Agricultural Research*, 14(3), e0206. doi:10.5424/sjar/2016143-9323
- Muthoni, F. K., Zhe, G., Mateete, B., Sseguyac, H., Kizitod, F., Baijukyae, F., & Hoeschle-Zeledon, I. (2017). Sustainable recommendation domains for scaling agricultural technologies in Tanzania. *Land Use Policy*, 66, 34–48. doi:10.1016/j.landusepol.2017.04.028
- Notenbaert, A., Herrero, M. D. E., Grootte, H., You, L., Gonzalez-Estrada, E., & Blummel, M. (2013). Identifying recommendation domains for targeting dual-purpose maize-based interventions in crop-livestock systems in East Africa. *Land Use Policy*, 30(1), 834–846. doi:10.1016/j.landusepol.2012.06.016
- Nouri, M., Homae, M., & Bannayan, M. (2017, September). Climate variability impacts on rainfed cereal yields in west and northwest Iran. *International Journal of Biometeorology*, 61(9), 1571–1583. Epub2017Apr18. doi:10.1007/s00484-017-1336-y PMID:28421270
- Obidiegwu, J.E., Akpabio, E.M., (2017). The geography of yam cultivation in southern Nigeria: Exploring its social meanings and cultural functions, *Journal of Ethnic Foods*. doi:10.1016/j.jef.2017.02.004
- Padbury, G., Waltman, S., Caprio, J., Coen, G., McGinn, S., Mortensen, D., & Sinclair, R. et al. (2002). Agroecosystems and land resources of the northern Great Plains. *Agronomy Journal*, 94, 251–261.
- QGIS Development Team. (2018). QGIS Geographic Information System. Open Source Geospatial Foundation Project. Retrieved from <http://qgis.osgeo.org>
- Radula, M. W., Szymuraa, T. H., & Szymurab, M. (2018). Topographic wetness index explains soil moisture better than bioindication with Ellenberg's indicator values. *Ecological Indicators*, 85(February), 172–179. doi:10.1016/j.ecolind.2017.10.011
- Reuter, H. I., & Nelson, A. (2009). Geomorphometry in ESRI packages. In T. Hengl & H. I. Reuter (Eds.), *Geomorphometry: Concepts, software and Applications* (pp. 269–292). Amsterdam, Netherlands: Elsevier. doi:10.1016/S0166-2481(08)00011-1
- Seppelt, R. (2000). Regionalised optimum control problems for agroecosystem management. *Ecological Modelling*, 131(2-3), 121–132. doi:10.1016/S0304-3800(00)00270-2
- Shaker, R. R., & Zubalsky, S. (2015). Examining patterns of sustainability across Europe: A multivariate and spatial assessment of 25 composite indices. *International Journal of Sustainable Development and World Ecology*, 22(1), 1–13. doi:10.1080/13504509.2015.1055524
- Stevens, F. R., Gaughan, A. E., Linard, C., & Tatem, A. J. (2015). Disaggregating census data for population mapping using random forests with remotely-sensed and ancillary data. *PLoS One*, 10(2), e0107042. doi:10.1371/journal.pone.0107042 PMID:25689585
- Tesfaye, K., Jaleta, M., Jena, P., & Mutenje, M. (2015). Identifying potential recommendation domains for conservation agriculture in Ethiopia, Kenya, and Malawi. *Environmental Management*, 55(2), 330–346. doi:10.1007/s00267-014-0386-8 PMID:25331642
- Tou, J. T., & Conzalez, R. C. (1974). *Pattern recognition principles*. Reading, MA: Addison-Wesley.
- Trabucco, A., & Zomer, R. J. (2009). Global Potential Evapo-Transpiration (Global-PET) and Global Aridity Index (Global-Aridity) Geo-Database. In *CGIAR Consortium for Spatial Information*. Retrieved from <http://www.csi.cgiar.org>
- Ulimwengu, J., Funes, J., Headey, D., & You, L. (2009). Paving the Way for Development? The Impact of Transport Infrastructure on Agricultural Production and Poverty Reduction in the Democratic Republic of Congo. *IFPRI Discussion Paper 00944 December 2009*. International Food Policy Research Institute (IFPRI).
- UNEP (United Nations Environment Programme). (1997). *World atlas of desertification 2ED*. London: UNEP.
- van Waart, J., Kersebaum, K. C., Peng, S., Milner, M., & Cassman, K. G. (2013). A protocol for estimating crop yield potential at regional to national scales. *Field Crops Research*, 143, 34–43. doi:10.1016/j.fcr.2012.11.018

Vanlauwe, B., Coyne, D., Gockowski, J., Hauser, S., Huising, J., Masso, C., & Van Asten, P. et al. (2014). Sustainable intensification and the African smallholder farmer. *Current Opinion in Environmental Sustainability*, 8, 15–22. doi:10.1016/j.cosust.2014.06.001

Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236–244. doi:10.1080/01621459.1963.10500845

Williams, C. L., Hargrove, W. W., Liebman, M., & James, D. E. (2008). Agro-ecoregionalization of Iowa using multivariate geographical clustering. *Agriculture, Ecosystems & Environment*, 123(1-3), 161–174. doi:10.1016/j.agee.2007.06.006

Wood, S., & Pardey, P. G. (1998). Agroecological aspects of evaluating agricultural R and D. *Agricultural Systems*, 57(1), 13–41. doi:10.1016/S0308-521X(97)00057-7

WORLDPOP. (2016). WorldPop: High resolution age structured population distribution maps. Retrieved from <http://www.worldpop.org.uk/>

You, L., Wood-Sichra, U., Fritz, S., Guo, Z., See, L., & Koo, J. (2014). Spatial Production Allocation Model (SPAM) 2005 Beta Version. Retrieved from <https://harvestchoice.org/tools/crop-production-spam-0>

Yu, B., & Guo, Z. (2015). *Measurement of agricultural productivity in Africa South of the Sahara: a spatial typology application*. Washington, DC: International Food Policy Research Institute.

Zhu, H. D., Shi, Z. H., Fang, N. F., Wu, G. L., Guo, Z. L., & Zhang, Y. (2014). Soil moisture response to environmental factors following precipitation events in a small catchment. *Catena*, 120, 73–78. doi:10.1016/j.catena.2014.04.003

## APPENDIX

### Supplementary Tables

Table 9. 23 Bioclimatic variables and their description

Variable	Description	Original Resolution (m)	Source
solarRad	Annual Solar radiation (kJ m <sup>-2</sup> day <sup>-1</sup> )	1000	Fick & Hijmans, 2017
Aridity Index	Aridity index(fraction)	1000	Trabucco & Zomer, 2009
PET	Potential evapotranspiration (mm)	1000	Trabucco & Zomer, 2009
Sunhr	Annual Sunshine hour (hour)	5000	Kothe et al., 2017
BIO1	Annual Mean Temperature (°C)	1000	Fick & Hijmans, 2017
BIO2	Mean Diurnal Range (Mean of monthly (max temp - min temp)) (°C)	1000	Fick & Hijmans, 2017
BIO3	Isothermality (BIO2/BIO7) (* 100) (fraction)	1000	Fick & Hijmans, 2017
BIO4	Temperature Seasonality (standard deviation *100) (°C)	1000	Fick & Hijmans, 2017
BIO5	Max Temperature of Warmest Month (°C)	1000	Fick & Hijmans, 2017
BIO6	Min Temperature of Coldest Month (°C)	1000	Fick & Hijmans, 2017
BIO7	Temperature Annual Range (BIO5-BIO6) (°C)	1000	Fick & Hijmans, 2017
BIO8	Mean Temperature of Wettest Quarter(°C)	1000	Fick & Hijmans, 2017
BIO9	Mean Temperature of Driest Quarter (°C)	1000	Fick & Hijmans, 2017
BIO10	Mean Temperature of Warmest Quarter (°C)	1000	Fick & Hijmans, 2017
BIO11	Mean Temperature of Coldest Quarter (°C)	1000	Fick & Hijmans, 2017
BIO12	Annual Precipitation (mm)	1000	Fick & Hijmans, 2017
BIO13	Precipitation of Wettest Month (mm)	1000	Fick & Hijmans, 2017
BIO14	Precipitation of Driest Month (mm)	1000	Fick & Hijmans, 2017
BIO15	Precipitation Seasonality (Coefficient of Variation) (fraction)	1000	Fick & Hijmans, 2017
BIO16	Precipitation of Wettest Quarter (mm)	1000	Fick & Hijmans, 2017
BIO17	Precipitation of Driest Quarter (mm)	1000	Fick & Hijmans, 2017
BIO18	Precipitation of Warmest Quarter (mm)	1000	Fick & Hijmans, 2017
BIO19	Precipitation of Coldest Quarter (mm)	1000	Fick & Hijmans, 2017

Table 10. 21 Edaphic variables and description

Variable	Description	Resolution (m)	Source
ALUM3S	Aluminum concentration (Mehlich 3) 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
AWCtS	Porosity (volumetric fraction) based on PTF 0-20 cm	250	ISRIC, 2016, Hengl et al., 2017
EACKCL	Exchangeable acidity (KCl) 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
EALKCL	Exchangeable Aluminum (KCl) for 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
ECN	Electrical conductivity for 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
EMGX	Exchangeable Mg for 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
EXBX	Exchangeable bases total 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
EXKX	Exchangeable K for 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
NTO	Total nitrogen for 0-20 cm depth	250	ISRIC, 2016, Hengl et al., 2017
AWCh1	Available soil water capacity (volumetric fraction) 0-15cm	250	ISRIC, 2016, Hengl et al., 2017
BDRICM	Depth to bedrock (R horizon) up to 200 cm	250	ISRIC, 2016, Hengl et al., 2017
BDRLOG	Probability of occurrence (0-100%) of R horizon	250	ISRIC, 2016, Hengl et al., 2017
BDTICM	Absolute depth to bedrock (in cm)	250	ISRIC, 2016, Hengl et al., 2017
BLDFIE	Bulk density (fine earth) in kg / cubic-meter at depth 0-15 cm	250	ISRIC, 2016, Hengl et al., 2017
CECSOL	Cation exchange capacity of soil in cmolc/kg (0-15cm)	250	ISRIC, 2016, Hengl et al., 2017
CLYPPT	Clay content (0-2 micro meter) mass fraction in % (0-15 cm)	250	ISRIC, 2016, Hengl et al., 2017
ORCDRC	Soil organic carbon content in g per kg (0-15 cm)	250	ISRIC, 2016, Hengl et al., 2017
PHIHOX	Soil pH x 10 in H <sub>2</sub> O (0-15cm)	250	ISRIC, 2016, Hengl et al., 2017
SLTPPT	Silt content (2-50 micro meter) mass fraction in %	250	ISRIC, 2016, Hengl et al., 2017
SNDPPT	Sand content (50-2000 micro meter) mass fraction in %	250	ISRIC, 2016, Hengl et al., 2017
TEXMHT	Texture class (USDA system)	250	ISRIC, 2016, Hengl et al., 2017

Table 11. 23 Remote sensing vegetation layers and terrain variables

Variable	Description	Resolution (m)	Source
SPAR	SPOT long-term average (1998–2014) fraction of absorbed photosynthetically active radiation	1000	AfSIS, 2017
SCOVER	SPOT long-term average (1998–2014) fraction of Green Vegetation Cover	1000	AfSIS, 2017
EVI	MODIS long-term (2000–2017) average enhanced vegetation index	250	AfSIS, 2017
NDVI	MODIS long-term (2000–2017) average normalized difference vegetation index	250	AfSIS, 2017
MREFB1	MODIS long-term (2000–2017) average blue reflectance	250	AfSIS, 2017
MREFB2	MODIS long-term (2000–2017) average green reflectance	250	AfSIS, 2017
MREFB3	MODIS long-term (2000–2017) average red reflectance	250	AfSIS, 2017
MREFB7	MODIS long-term (2000–2017) average Mid infra-red reflectance	250	AfSIS, 2017
GPP	MODIS long-term average (2000-2010) Gross Primary Productivity	1000	AfSIS, 2017
NPP	MODIS long-term average (2000-2010) Net Primary Productivity	1000	AfSIS, 2017
LAI	MODIS long-term average (2000-2010) Leaf Area Index	1000	AfSIS, 2017
MPAR	MODIS long-term average (2000–2016) fraction of absorbed photosynthetically active radiation	1000	AfSIS, 2017
MBNA	MODIS long-term (2000–2016) average black-sky NIR albedo	500	AfSIS, 2017
MBSA	MODIS long-term (2000–2016) average black-sky SWIR albedo	500	AfSIS, 2017
MBVA	MODIS long-term (2000–2016) average black-sky VIS albedo	500	AfSIS, 2017
MWNA	MODIS long-term (2000–2016) average white-sky NIR albedo	500	AfSIS, 2017
LSTD	MODIS long-term 2002-2017 average day land surface temperature	1000	AfSIS, 2017
LSTN	MODIS long-term 2002-2017 average night land surface temperature	1000	AfSIS, 2017
Elevation	Shuttle Radar Topography Mission (SRTM) DEM	90	Jarvis et al., 2008
TWI	Topographic Wetness Index	90	SAGA GIS*
STRPI	Stream Power Index	90	SAGA GIS*
slope	Soil Slope	90	SAGA GIS*
Aspect	Soil Aspect	90	SAGA GIS*

\*SAGA GIS is by Conrad et al., 2015

*Tunrayo Alabi is currently the GIS Support services Manager at International Institute of Tropical Agriculture (IITA) where he has worked for about 23 years. He provides support on GIS and remote sensing application to agricultural research in sub-Saharan Africa. He obtained a B.Sc Electrical/Electronic Engineering with First Class from the Department of Electrical Engineering, University of Ibadan, Nigeria (1989) and an M.Sc in Geographic Information Systems (GIS) in the Department of Geography of the same University (1999). He was trained for 2 months at Environmental Systems Research Institute (ESRI) in USA, Redlands, California in the application of GIS in the year 2004 and was a Visiting Research Scholar at the University of New Hampshire, Durham, the USA on the application of remote sensing to agricultural research in 2007. His area of research is the application of earth observation to food security, climate change studies, Geospatial capacity building, geostatistical and land suitability analysis, spatial targeting of agricultural technology in Africa.*

*Patrick Adebola is a Plant Breeder/Geneticist. He joined IITA-Abuja, Nigeria as Project Leader of the Africa Yam Project. Dr Adebola oversees and coordinates the project activities in the four target countries. He obtained his BSc, MSc and PhD degrees from the University of Ilorin, Nigeria in 1987, 1990 and 2003 respectively. He also obtained a master's degree in Business Leadership (MBL) from the Graduate School of Business Leadership, University of South Africa in 2014. Dr. Adebola has over 20 years' experience in plant breeding research and community engagement in West and Southern Africa and had managed and supervised several international projects and networks. Before joining the IITA, Dr. Adebola works for the Africa Rice Center as the Deputy Director General for the Central Agricultural Research Institute (CARI), Liberia in 2016. He was a Research Team Manager and Head of the Plant Breeding Division at the Agricultural Research Council, ARC-VOPI, South Africa (2005 – 2016). He also previously works for the Cocoa Research Institute of Nigeria as a research scientist and rose to the position of Assistant Chief Research Officer and Program Manager Biotechnology (1993-2004). He was a postdoctoral fellow (2004-2005) in the department of Botany, University of Fort Hare, South Africa and a recipient of the Rothamsted International Postdoctoral Fellowship hosted by the Food and Environment Research Agency (FERA), Sand Hutton, York, United Kingdom during 2008-2010.*

*Asrat A. Asfar is a plant breeder and geneticist by training. Worked as plant breeder for more than 15 years in national and international research systems. Developed improved varieties for release and populations for genetic studies in many crops ranging from grain legumes to root and tuber crops.*

*Antonio Lopez Montes is an Agronomist and plant breeder and geneticist by training. Worked as a plant breeder for more than 26 years in national and international research systems. Developed 21 improved and released varieties of Cassava and of yam. He also developed the basic crosses for the Sequenciation of the genome of the White yam, interspecific crosses of yam and populations for genetic studies in Yam.*

*Robert Asiedu is the Director, Research for Development, for Western Africa at the International Institute of Tropical Agriculture (IITA). He also leads the Institute's work on Biotechnology and Crop Improvement. He joined IITA in 1989 to work as a breeder of root and tuber crops, after a Postdoctoral Research Fellowship at the International Maize and Wheat Improvement Center (CIMMYT), Mexico where he worked on the transfer of useful traits to wheat from its wild relatives. He holds a PhD degree in Agricultural Science from the University of Adelaide, Australia and a BSc Honors degree from the Kwame Nkrumah University of Science and Technology, Kumasi, Ghana.*