KU LEUVEN

FACULTY OF BIOSCIENCE ENGINEERING

Characterization of Target Population of Environments for East African Highland Banana Using a Multi-criteria Decision Methods

Vaststellen van de landgeschiktheid voor Oost-Afrikaanse Hooglandbanaan met behulp van een multi-criteria beslissingsmethode

Promoters:
Prof. Jos Van Orshoven
Department of Earth and Environmental Sciences
Division of Forest, Nature and Landscape
Bio
Prof. Rony Swennen
Department of Biosystems
Division of Crop Biotechnics

Dissertation presented in
partial fulfillment of the requirements
for the degree of Master of Science in
Master of Science in Agro- and Ecosystems Engineering

Emile Sebastian Arinaga

June 2018

Preface

I would like to thank my promotor, Prof. Jos Van Orshoven, for his guidance and patience during roadblocks in my work, my co-promotor, Prof. Rony Swennen, for his belief in my getting the job done and motivation at the right moments, my supervisor Karen Gabriels for her help in all GIS and Land Evaluation-related matters, and finally my supervisor Inge Van den Bergh for showing me what field research is like, and for helping make sense of what could sometimes seem like a text without end. Thank you all for your effort and patience!

Abstract

The East African Highland banana (EAHB) is an important staple food and cash crop in the region encompassing Uganda, Rwanda, Burundi, Tanzania, western Kenya and the eastern Democratic Republic of Congo (DRC). Its cultivation is subject to a wide variety of limitations, including soil, climate, socio-economic and pest- and disease-related factors. Breeding initiatives seeking to improve food security by developing and introducing productive, pest- and disease-resistant new EAHB hybrids can benefit from a rigorous characterization of the region in which these new hybrids are tested and released to farmers. Characterization was carried out using a Multi-criteria Decision Method (MCDM) embedded in a Land Evaluation framework and involved subdivision of the target region into fourteen sub-environments with a large degree of internal homogeneity and inter-sub-environment heterogeneity. This target region is also known as the Target Population of Environments (TPE). This methodology also involved calculation of predicted suitability and performance values, expressed as a distance to a ten-dimensional ideal point and an EAHB bunch weight respectively. Along with the predicted bunch weight, the most limiting variable in each spatial unit or pixel of the region was also calculated. Suitability was calculated using a MCDM called compromise programming (CP), while performance was calculated using Liebig's Law of the Minimum method. As input data for these calculations, data on ten soil and climate variables downloaded from open-source web platforms was used. Separate datasets containing the same soil and climate data along with actual bunch weight data were employed to set up parameters necessary for these calculations. Eventual results included maps of suitability, performance, the fourteen sub-environments and all ten soil and climate variables spanning the entire study region, as well as a more limited dataset of 64 data points spread across the banana growing regions of Uganda for which both estimated actual and predicted potential bunch weight (performance) data was available. This allowed bunch weight gaps to be calculated, whereas in the larger region, suitability predictions were compared to sub-national production figures obtained from public censuses and surveys. Suitability and performance predictions did not consistently correspond to the reality represented by both the estimated actual bunch weights and the sub-national production figures. Sub-environments also did not consistently identify the true diversity and boundaries of EAHB production systems in the region. Where inaccuracies occurred, socio-economic, pest- and disease-related and management factors were thought to be missing explanatory variables, and in several cases were identified as such in literature. However, inaccuracies were not the rule: differences in predictions and most limiting variables between sub-environments in the Ugandan banana growing regions, for example, largely reflected conclusions in available literature. The utilized land evaluation framework, which encompassed the CP and Law of the Minimum methods, was a biophysical land evaluation methodology with elements of both qualitative and quantitative land evaluation. Additional

data, especially the missing explanatory variables, is thought to have the potential to produce vastly more accurate predictions than the ones in this project.

List of Abbreviations

A	Index indicating proportion of *Musa acuminata* genome
AEZ	Agro-ecological Zoning
ALES	Automated Land Evaluation System
B	Index indicating proportion of *Musa balbisiana* genome
BW	Bunch Weight
BW_{est}	Estimated Actual Bunch Weight
BWG	Bunch Weight Gap
BW_{pred}	Predicted Potential Bunch Weight
BXW	Banana Xanthomonas Wilt
CP	Compromise Programming
DIP	Distance to the Ideal Point
DRC	Democratic Republic of Congo
EAGL	East African Great Lakes
EAHB	East African Highland Banana
ECV	Edapho-climatic variable
ET	Evapotranspiration
FSA	Farming Systems Analysis
GEI	Genotype-Environment Interaction
IITA	International Institute of Tropical Agriculture
IP	Ideal Point
IPC	Ideal Point Coordinate
ISRIC	International Soil Reference and Information Centre
LC	Land Characteristic
LGP	Length of Growing Period
LMU	Land Mapping Unit
LQ	Land Quality
LUR	Land Use Requirement
LUT	Land Use Type
MET	Multi-environment Trial
NARO	National Agricultural Research Organization
MCDM	Multi-Criteria Decision Method
RQ	Research Question
SDIP	Single-variable Distance to the Ideal Point
SG	Suitability Gap
SOC	Soil Organic Matter
SOM	Soil Organic Material
TPE	Target Population of Environments
TR4	Tropical Race 4
TSN	Total Soil Nitrogen
UBG	Ugandan Banana Growing (Region)

Table of Contents

1. Introduction

1.1 General context

Smallholder banana cultivation in the East African Great Lakes (EAGL) region provides a large part of local diets and incomes (Wairegi et al., 2009). Bananas can be cooked, eaten raw and processed into beer as well as provide fiber, construction materials and medicinal products. They are also important elements of culture (Karamura et al., 2000). A large proportion of smallholder farmers in the region grows EAHB (*Musa* spp. AAA-EA), a unique set of cultivars and varieties.

EAHBs provide a staple food for over 30 million people in East Africa and are also used as a cash crop. Most of the production is by smallholder farmers. However, the yields attained are much lower than the potential yields and have been declining in the latest decades, as has length of plantation life. By the late 1990s yields in Uganda had dropped to almost one-third of the level of yields measured in 1970 (Karamura et al., 1998). Various factors, such as low and decreasing soil fertility and limited rainfall but also increasing pests and diseases (such as nematodes and the fungus causing black leaf streak), as well as competing activities and limitations on labor contribute to this rising yield gap (Karamura, 1998; Karamura et al., 1998; Tushemereirwe et al., 2015).

A solution has been put forward by Bioversity International, a research organization that describes itself as being "*committed to attaining global food and nutrient security by safeguarding agricultural and tree biodiversity*" (Bioversity, 2016). This solution encompasses the introduction of new, resistant and productive cultivars developed by the International Institute of Tropical Agriculture (IITA) in conjunction with the Ugandan National Agricultural Research Organization (NARO). These cultivars are named 'NARITAs' after the two developing organizations and will henceforth be referred to as such (Tushemereirwe et al., 2015). The objective is that, upon large-scale adoption by smallholder farmers, the hybrids ensure more food and financial security as well as improve livelihoods across the region, in the face of a rapidly changing climate and a potentially volatile socio-economic situation. However, in order to increase chances of successful use by farmers across the extremely diverse set of environments that constitute the EAGL region, the hybrids' responses to agro-ecological conditions, as well as the factors influencing their adoption must be determined (Tushemereirwe et al., 2015).

1.2 Problem statement

The development, testing and dissemination of new varieties, such as the NARITAs, to farmers are expensive and time-consuming processes. Determining where and because of which factors EAHBs perform well could help to select optimal yield increase strategies for each (micro) region and provide a first indication as to where NARITA testing and introduction efforts are best applied. Setting up a spatial

framework by the subdivision of a target environment for breeding or extension projects into internally homogenous subunits is a crucial initial step in this process. It facilitates identification of different sub-environments, and of patterns and intercorrelations in the spatial variability of soil, climate, socio-economic and optimal production variables. Sub-environment characteristics derived at lower resolution and in a broader study (such as this Master thesis) may later be compared to more detailed socio-economic and production system baseline studies conducted in five areas in Uganda (Luweero and Mbarara districts) and Tanzania (Arusha, Kagera and Mbeya provinces). However, there are many different mechanisms for subdividing target environments into these subunits, most of which involve high-resolution data on multiple crop production-related variables (Chenu, 2015, section 2.2.4). An important challenge in this project was thus to use a subdivision method capable of identifying subunits which correctly represent real-world sub-environments, using only the datasets available (Section 3.2).

In this project, an area's optimality for EAHB production was predicted by two calculated factors representing, respectively, suitability for and performance of EAHB, the former expressed as dimensionless index values and the latter as bunch weights (BWs, unit: kg. See sections 3.3.2 and 3.3.3). The factors were computed using high-resolution geodatasets of several variables known to be limiting for EAHB production in the EAGL region. While these geodatasets were available for edapho-climatic variables (ECVs), they generally were not available for variables related to management practices, nor for socio-economic and pest and disease variables. The latter are more subject to significant variability over small spatial and short temporal scales. The fungal leaf-spot disease black leaf streak, for example, shows large differences in its development depending on the growth stage of the host banana plant, the availability of water, and whether or not farmers remove infected plant material (Fouré, 1993). This makes a single snapshot in time most likely to be inaccurate as time goes on, and social and natural changes, such as the migration of labor to other regions and the increasing unpredictability of rainy seasons, take hold. Socio-economic and management variables often show significant variability within a single village or even farm (Guy Blomme et al., 2017; Van Asten et al., 2006). Therefore, even high-resolution datasets are likely to miss significant variability among different smallholder plots.

There is thus an argument to be made for making initial suitability and performance predictions using solely ECVs, as these are generally more stable or predictable over time, and, for the EAGL region, more commonly exist in a spatially referenced, spatially continuous format, than socio-economic, management and pest and disease data. While such initial assessments are not likely to be fully accurate, they may provide valid baseline assessments. Considering socio-economic, management and pest and disease variables – henceforth collectively known as *explanatory variables* - in a separate, subsequent step has the potential to provide indications as to why ECVs alone are (not) sufficient predictors for EAHB

performance. Both the initial ECV-based assessments as well as any supplementary indications based on explanatory variables can also serve to identify priorities for further research.

1.3 Objective

To fulfil this MSc thesis research project's objective, spatial subunits representing sub-environments in the EAGL region are to be identified based on a selection of ECVs limiting for EAHB production. Subsequently, predictions of suitability and performance of smallholder EAHB cultivation in the region, calculated with the same ECVs, will be interpreted using these subunits as a spatial framework. Spatial subunits will be established using a clustering algorithm, and two different MCDMs able to incorporate and balance the influences of multiple variables at once (Estrella et al., 2014), will be employed for the suitability and performance predictions. Predicted performance, in the form of bunch weight (kg) will then be matched with estimated bunch weight data representing actual, on-farm conditions to compute *bunch weight gaps* (BWGs), the size of which is inversely proportional to the accuracy of the predicted performance. Because predicted suitability is represented by an index and can thus not be compared to any actual value expressed in the same units, its validity will only be evaluated in a rough comparison of the index values and their spatial distribution patterns with estimated actual bunch weight or other available performance data. In spatial subunits where predicted performance and/or suitability sufficiently approach reality, the most influential factors from within the model will be singled out. In subunits where predictions are inconsistent with actual production or performance, the influence of explanatory variables from outside the model will be discussed as well. Finally, the MCDMs will be briefly evaluated for their appropriateness of use in this project, and amendments to the methods used will be suggested.

1.4 Research questions

The research questions (RQs) which this thesis research project addresses are:

1. What are the edapho-climatic land use requirements (LURs) of EAHB cultivated by smallholders in Uganda?

2. What are the edapho-climatic conditions in the EAGL region, and to what extent do they meet the land use requirements of EAHB?

3. How well do performance and suitability, as predicted from a selection of the most relevant edapho-climatic LURs, approximate actual bunch weight figures in the EAHB-growing regions of Uganda? Can a similar comparison be made for the larger study region?

4. To what degree do mapped spatial distributions of ECVs and EAHB suitability values, as well as non-mapped descriptions of spatial distributions of explanatory variables, along with a supporting

framework of formal spatial subunits, correspond to the existing body of knowledge on EAHB cultivation and its various constraints?

5. Are the MCDMs as employed in this project appropriate tools for prediction of suitability or performance of EAHB production in the EAGL region, and if not, how could models be amended in similar applications in the future?

1.5 Hypothesis

EAHB is characterized by LURs that are distinct from those of lowland *Musa* spp. There is significant spatial variability in the EAGL region regarding predicted suitability for and performance of EAHB cultivation, when considering ECVs only. Many of these variables are strongly correlated, and may show similar spatial distribution trends related to the various physical, biological and anthropogenic processes that have formed the landscape, for example gradients of soil nutrient status occurring on the slopes of mountains (De Bauw et al., 2016). Spatial subunits and the associated sub-environments are characterized by the level of the ECV and explanatory variable values in that sub-environment. Predicted bunch weight figures will in many cases not correspond to actual bunch weight, and differences can be attributed to (i) inadequacy of the prediction models, (ii) empirically defined model parameters and relationships being based on insufficient data, (iii) socio-economic, management and pest and disease variables not being considered, or (iv) combinations of two or more of the previous factors. The same can be said for discrepancies between predicted suitabilities and actual performance data. Model predictions, while for a large part conforming to wider regional trends in actual EAHB performance, will be unable to explain the latter's variability at smaller (intra-village) scale. It is thought that management, which has a strong influence on soil fertility and pest and disease prevalence (which is itself strongly influenced by plant nutrient status and soil fertility), is a missing piece of the puzzle (Van Asten et al., 2004). Spatial subunits can identify several different sub-environments in the EAGL region, which, despite significant intra-subunit variability, are characterized by unique sets of average ECV and explanatory variable values. The most limiting ECVs for EAHB production are identifiable in each of the various subunits. Where an identified spatial subunit and its corresponding sub-environments do not match the boundaries of areas strongly influenced by a specific set of ECVs and explanatory variables, causes of within-subunit variability are worth looking into.

2 Literature Review

2.1 East African Highland Banana

2.1.1 EAHB: A special subgroup of *Musa*

EAHB cultivars can be divided into cooking bananas, which are edible when cooked, boiled or steamed, and beer bananas, whose flesh is bitter and inedible but can be processed into juice or beer. Cooking bananas, besides being used for subsistence, provide a source of income for farmers, who sell their excess production at local markets, whereafter it is often transported to cities to form a staple food for increasing urban populations. Banana is used as livestock fodder in mixed farming systems and is often intercropped with annual and perennial crops such as beans and coffee (Karamura et al., 1998).

Like other banana cultivar groups, the EAHB subgroup's cultivars are giant perennial herbs consisting of a rhizome, from which roots and a pseudostem made up of rolled up leaves emanate. The greatest concentrations of roots may typically be found at depths of <50 cm, and within a radius of 60 cm from the main pseudostem (Blomme et al., 2014). Propagation is typically carried out vegetatively, via side shoots (called suckers) or lab-multiplied tissue culture material. Further morphological components of a banana plant are shown in Figure 2.1.

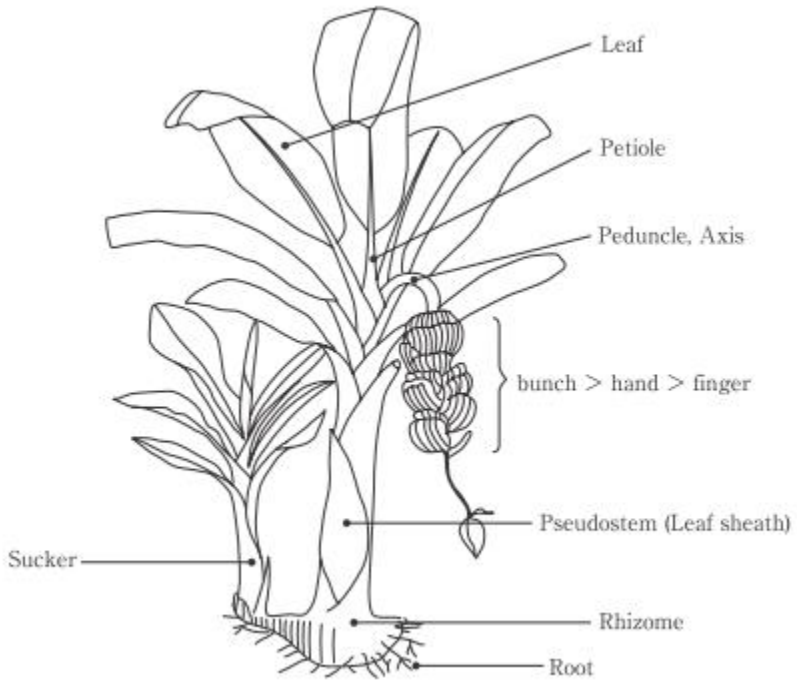

Figure 2.1: Morphology of a banana plant (IPGRI et al., 1996)

There is much speculation on when and where EAHB's ancestor was introduced to East Africa (Karamura, 1998; Komatsu et al., 2010). A commonly accepted view is that EAHB arose from a single introduction and underwent a huge population expansion through vegetative reproduction by farmers after arriving in East Africa around 2500 years ago, most likely after being introduced from Asia by way of Indian Ocean trade and/or migrations. Its current genetic diversity and sets of distinct clones are the result of consistently recurring somatic mutations and selection by farmers living in very diverse agro-ecological and socio-cultural settings, all of which happened *after* introduction into the EAGL region, which is therefore named a secondary center of banana diversity. However, several studies have demonstrated that EAHB has a narrow genetic base, and were unable to establish a consensus as to whether this limited genetic diversity could be related to the more common morphology-based classifications (Karamura, 1998; Kitavi et al., 2016; Nyine et al., 2011). Adequately mapping the diversity in this genetic base and its relationship to favorable traits such as pest, disease and drought resistance or high yield is important, especially since if a certain clone lacks resistance to a disease, this clone's genetic makes all plants of this clone type vulnerable to potentially devastating new diseases such as Banana Xanthomonas Wilt (BXW) or black leaf streak caused by the fungus *Micosphaerella fijiensis.* (Karamura et al., 2016; Kitavi et al., 2016). In the EAGL region, breeding of improved cultivars has thus focused on introducing pest and disease resistance, yield and external fruit quality traits from wild or exotic cultivars into new EAHB hybrids, such as the earlier mentioned NARITAs (section 1.1). The majority of this collection of 27 hybrids, including both beer and cooking types, resulted from integrating traits from the wild species *Calcutta 4* and selection for resistance to black leaf streak and high yield during several rounds of crossing (W. Tushemereirwe et al., 2015). Evaluation of the NARITAs has now moved on to the multi-environment testing stage, which includes both on-station trials with conditions resembling those of the local environment as much as possible, as well as farmer-managed trials being carried out to gain a detailed picture of farmers' trait preferences and local agro-ecological conditions. Farmer-managed trials thus are able to obtain feedback for the breeding process, gain insights as to which hybrid cultivar is best suited to a certain environment, and stimulate farmer adoption of hybrids right out the trials (Bioversity International, 2013). While the focus of this stage of testing is on sites in Uganda and Tanzania, where all respectively 19 and 21 hybrids are under evaluation (Bioversity International, 2013; pers. communication Inge Van den Bergh), multi-environment trials of a limited amount of NARITA cultivars have already been carried out in Rwanda (Ndayitegeye et al., 2017) and the DRC (Kamira et al., 2016).

2.1.2 Size and Distribution of Current Production

The *Musa* AAA EAHB subgroup is the most widespread in the EAGL region, appearing mostly in highland areas at altitudes of 1000-2000 m (Karamura et al., 1998). Growing areas in Uganda also

coincide with areas where the dry season is less severe (Karamura, 1998). However, EAHB's coverage of the region is far from uniform, and is very low outside aforementioned altitude ranges, and practically non-existent in the coastal areas (Karamura, 1998; Karamura et al., 1998; Figure 2.2). Other banana subgroups are grown too, including dessert or 'sweet' bananas such as those in the *Musa* AAA 'Cavendish' subgroup, found predominantly in Tanzania's coastal regions, and in the *Musa* AAA 'Gros Michel' subgroup, which enjoys some popularity in the areas around Lake Victoria, as well as other cooking types such as those in the Plantain subgroup (*Musa* AAB) in the DRC, and several exotic cultivars and introduced hybrids. In Uganda, Plantains and dessert bananas are grown on a very limited scale, comprising less than 2% and 5% of production, respectively, primarily due to high susceptibility to pests and diseases and, in the case of Plantains, to most of Uganda being too high-altitude for optimal growth (Karamura et al., 1998; Lescot, 2012). In Rwanda, Burundi, and Uganda's Central Region, a substantial proportion of banana production consists of ABB cultivars, that are used for beer brewing (Karamura et al., 1998). Beer types also reportedly take up a large proportion of total banana cropland in the DRC's South Kivu province, whereas the neighboring North Kivu province exhibits more of an even mix between dessert, cooking, beer and plantain types (Ocimati et al., 2013).

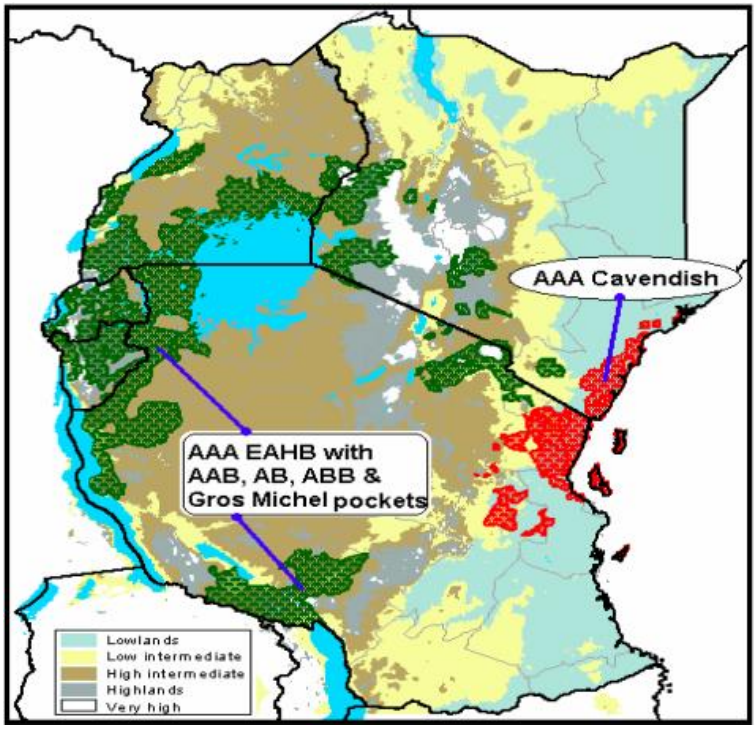

Figure 2.2: Banana Growing Areas in the EAGL region, minus the DRC (Tushemereirwe et al., 2006).

An area's predominant ethnic group or culture may play a role in determining which cultivars are grown. In Tanzania's Mbeya region, for example, Plantains form an essential part of the Nyakyusa people's

cropping system and diet, whereas the Wachagga in the Arusha/Kilimanjaro regions and the Haya in the Kagera region grow EAHB (de Langhe et al., 2001; Maruo, 2002, 2007)

Ethnicity and culture is also related to the type of agro-system employed: whereas the Haya mulch banana plantations with banana and other crop residues, the Wachagga in Kilimanjaro feed their crop residues to stall-fed cattle, applying the resulting manure to their banana plantations (Maruo, 2002, 2007). Exposure to markets and the associated consumers' preferences can lead to shifts in production, as in the DRC's North Kivu region, where local demand for cooking banana along with access to the plantain-hungry Kampala (Uganda) market has led to these types contributing to greater shares of total regional production (Dowiya et al., 2009).

Banana production systems in the region vary along with the socio-economic, cultural and agro-ecological makeup of an area, and are extremely diverse, up to the point that even within a single agro-ecological zone, they defy distinct categorization (Karamura et al., 1998). Broad categorization is often attempted: distinction is made in (Jogo et al., 2013) between the well-managed, higher-altitude (1700 m) and generally less disease-ridden EAHB systems in the Ugandan southwest, and the *Musa* ABB 'Kayinja' type beer/juice banana-dominated systems in central Uganda, which are generally less well-managed, have higher disease incidences and are located at lower (1300 m) altitudes. In the high-altitude system, intercropping is done primarily with beans and coffee, and production is divided into female-dominated subsistence gardening and male-dominated commercial cropping carried out by hired labor. In the lower-altitude system, production generates less income, is of lesser importance in the agro-system and in farmers' diets, and is mainly carried out by women (Karamura et al., 2013). Variations in these two systems can also be found throughout the rest of the EAGL region. Blomme et al. (2017) make a distinction between market-oriented production systems in southwest Uganda and subsistence systems in central Uganda, Burundi and the eastern DRC.

Production data for the various countries in the EAGL region are summarized in Table 2.1. Sources range from national statistical institutes from the countries in question (INS, 2017) to groups of experts working for non-governmental research institutes (Lescot, 2012). There are large degrees of uncertainty in many of these data, due to insufficient resources available to many of these institutes, or reasons unknown. Estimates for total annual banana production in Tanzania for the 2007-2008 agricultural year, for example, range from 1.89 Mt (NBS, 2012) to 3.82 Mt (FAO, 2013). The categories 'plantain' and 'cooking banana' are often confused, and most sources do not even make distinctions between different categories. While EAHB makes up the majority of bananas grown in Uganda, Rwanda and Burundi, this subgroup is less dominant in Tanzania, Kenya and the DRC, which boast significant production of dessert

banana and plantains. The DRC is not entirely incorporated in this project's delineation of the EAGL region, which is why data are provided on those provinces roughly falling inside the region's boundaries. Furthermore, while beer banana was not listed as a category in the Lescot (2012) data, other sources listed substantial annual beer banana production in Uganda, Rwanda and Burundi at 0.24, 0.8 and 0.84 Mt respectively (NISR, 2010; Sidi et al., 2013; UBOS, 2010).

Table 2.1: Annual production figures per country, expressed in Mt (million tons).

Country	Cooking banana	Beer banana	Dessert banana	Plantain	Total	Census Year	Source
Uganda	8.95	/	0.41	0.2	9.55	2012	Lescot, 2012
Tanzania	2.6	/	0.15	0.15	2.9	2012	Lescot, 2012
Rwanda	2.26	/	0.22	0.27	2.75	2012	Lescot, 2012
Burundi	0.21	/	0.32	1.05	1.57	2012	Lescot, 2012
DRC	/	0.74	0.83	4.97	6.54	2015	INS, 2017
DRC (North- and South-Kivu)	/	0.74	0.15	1.35	2.24	2015	
Kenya	Banana: 1.29			0.032	1.32	2016	FAO, 2013
Total (All countries)	14.02 (Kenya not incl.)	1.48 (Kenya not incl.)	1.93 (Kenya not incl.)	6.67	24.63		

2.1.3 Land Use Requirements of EAHB

Soil fertility has been declining in banana growing areas in East Africa, with as main causes increased rural population pressure and the resulting decrease in fallowing, along with rising export of cooking banana, and thus nutrients, from rural to urban areas. Soil fertility parameters have often been mentioned as important constraints to banana production (Okech et al., 2004; Taulya, 2013; Van Asten et al., 2004, 2006; Wairegi et al., 2010). Soil exchangeable K, Mg, and total soil N were found to be limiting in various areas in all countries in the EAGL region, and soil pH, soil organic matter, and soil exchangeable Ca, Zn and Cu were often mentioned as important for EAHB productivity (Nyombi et al., 2010; Wairegi et al., 2010). Temperature and rainfall are also major constraints to EAHB production, playing important roles in regulating EAHB's growth rate, time of flowering and other performance-affecting variables (McIntyre et al., 2003; Taulya et al., 2014). Climate change, in the form of ever-increasing variability of timing and intensity of rainy and dry seasons and rising temperatures, has the potential to cause substantial changes in banana growing systems worldwide and in the EAGL region (Calberto et al., 2015; Karamura et al., 1998).

Optimal and critical soil and climate variable values (= LURs) for EAHB production encountered in literature are listed in Table 2.2. These values represent a partial answer to Research Question 1 (Section 1.4); they furthermore represent a reference point for any LURs calculated using data from this project (See Section 3.3.3). When values for EAHB were not available, the corresponding values for *Musa* in general are given. Banana growth is possible between soil pH values of 4.5 and 8.5 according to De Geus (1967), but ideal ranges for EAHB are reported to lie between pH 5 and 6.5 (pers. communication Deborah Karamura). Various studies appear to confirm pH 5 as a critical minimum value (Alou et al, 2014; Sebuwufu et al., 2004; Van Asten et al., 2004). Furthermore, Alou et al. (2014) reported that bunch weights ceased to increase above pH 6.2. Soil organic matter (SOM) is reported to be limiting at a level lower than 3.0% [g SOM.g soil^{-1}.%], and total soil nitrogen (TSN) is limiting below 0.2% [g N.g soil^{-1}.%] (Odeke et al, 1999). SOM data was not available; therefore, it was necessary to convert SOM to soil organic carbon (SOC). A factor of 0.5 was used in this work to obtain a threshold level of 15 g SOC.kg soil^{-1}.

Table 2.2: Optimal and critical values for important limiting soil and climate variables for EAHB production.

Variable	Optimal and/or critical values	References
Soil pH	Optimal between pH 5-6.5 Critical values at pH 4.5 and 8.5	De Geus, 1967, pers. communication Deborah Karamura.
Soil Organic Carbon [g SOC.kg soil^{-1}]	Optimal value: 15 g.kg^{-1}	Odeke et al., 1999
Total Soil Nitrogen [g N.kg soil^{-1}]	Optimal value: 2 g.kg^{-1}	Odeke et al., 1999
Clay percentage	Critical upper value: 45%	Stover & Simmonds, 1987
Soil K^+ [cmol$_c$.kg^{-1}]	Critical lower value between 0.2-1.5 cmol$_c$.kg^{-1}	Alou et al., 2014; Odeke et al., 1999; Okech et al., 2004; Sathiamoorthy et al., 1989; Smithson et al., 2001; Van Asten et al., 2004, 2006
Soil Ca^{2+} [cmol$_c$.kg^{-1}]	Critical lower value: 3 cmol$_c$.kg^{-1}.	Van Asten et al., 2006
Soil Mg^{2+} [cmol$_c$.kg^{-1}]	Critical lower value between 0.9-1.04 cmol$_c$.kg^{-1}.	Alou et al., 2014; Gold et al., 1999; Smithson et al., 2001; Van Asten et al., 2006
K/Mg ratio	Optimal value: 0.3	Davies, 1995
Average Annual Temperature [^{0}C]	Optimal between 25-30 ^{0}C Critical values at 13^{0}C and 37^{0}C	Calberto & Staver, 2015, pers. communication Deborah Karamura
Average annual rainfall [mm.yr^{-1}]	Optimal value: >1500 mm.yr^{-1} Critical lower value: 900 mm.yr^{-1}	Calberto & Staver, 2015

EAHB bunch weight was reported to be highest for clay percentages in the range between 18.5-21.4%, and to decline at higher percentages (Alou et al., 2014). However, a positive relationship between EAHB

yield and clay percentage was reported by Wairegi et al. (2010), with a maximum around 40 %, and Stover & Simmonds (1987) listed an upper clay percentage value of 45 % as critical for banana. While doubt remains about the exact nature of the relationship, this does encourage the idea that EAHB production is highly feasible at clay percentages below 45%.

A whole range of critical minimum values for soil exchangeable K^+ have been suggested, ranging from 0.2 to 1.5 $cmol_c.kg^{-1}$ (Alou et al., 2014; Odeke et al., 1999; Okech et al., 2004; Sathiamoorthy et al., 1989; Smithson et al., 2001; Van Asten et al., 2004, 2006). The minimum soil Ca^{2+} requirement for AAA-bananas based on several international studies was listed as 3.0 $cmol_c.kg^{-1}$ soil (Van Asten et al., 2006). A strong positive relationship between Ca^{2+} and EAHB bunch weight was also documented up to soil Ca^{2+} concentrations of 2.04 $cmol_c.kg^{-1}$, as well as between soil Mg^{2+} and EAHB bunch weight when soil Mg^{2+} was between 0.36-1.04 $cmol_c.kg^{-1}$ (Alou et al., 2014). In the same study, EAHB bunch weight was far less responsive to further increases of soil Mg^{2+} content above this range, and entirely unresponsive above soil Mg^{2+} contents of 2.3 $cmol_c.kg^{-1}$. Several other publications reported critical minimum soil Mg^{2+} contents of 0.9-1 $cmol_c.kg^{-1}$ (Gold et al., 1999; Smithson et al., 2001; Van Asten et al., 2006). Soil cation ratios, such as the K/Mg and K/(K + Mg + Ca) ratios appear to influence plant resistance to weevils and nematodes (Okech et al, 2000). An optimal value of 0.3 for the K/Mg ratio was given by Davies (1995) .

The ideal temperature range for EAHB production is defined as 25-30 ºC (pers. communication, Deborah Karamura), whereas the critical values are defined for *Musa* spp. in general (Calberto & Staver, 2015; Table 2.2). Optimal and critical values for average annual precipitation are also defined for *Musa*. Banana production at precipitation rates lower than 900 $mm.yr^{-1}$ is only possible with special practices, such as irrigation or water conservation measures. Rainfed banana will suffer growth limitations at average annual rainfall rates below 1500 $mm.yr^{-1}$ (Calberto et al., 2015).

EAHB production in the EAGL region is adversely affected by a wide variety of diseases, each with its own degree of incidence and severity. One of the most ubiquitous and damaging diseases is Black Sigatoka, caused by the fungus *Mycosphaerella fijiensis* and resulting in serious yield loss due to its shriveling and drying of the banana plant's leaves. First reported in Uganda in 1989, it proceeded to colonize most of the warmer, lower-altitude areas of the EAGL region (Tushemereirwe et al., 2004). Another disease, banana Xanthomonas or bacterial wilt (BXW), caused by the *Xanthomonas campestris* pv. *musacearum* bacterium, was first reported in 2001 in limited areas in Central Uganda (Tushemereirwe et al., 2004) but has since become widespread in all six countries in the EAGL region (Ochola et al., 2014). The disease often causes 100 % yield loss and is spread to EAHB cultivars through insect vectors as well as infected farm tools and planting materials. Fusarium wilt is a disease caused by the *Fusarium*

oxysporum soil fungus. While it has rarely been reported to have an impact on EAHB plantations, mainly damaging exotic cultivars, a limited outbreak occurred among EAHB plantations in the Ugandan southwest (Blomme et al., 2013; Tushemereirwe, 1993). Other diseases with serious impact on EAHB production in the region include banana streak virus and banana bunchy top disease (Niyongere et al., 2011; Tushemereirwe et al., 2004).

Nematodes, microscopic parasitic worms that attack banana plants' roots, and weevils, insects that primarily affect the corm, are a cause of up to 85 % yield loss, and appear to have a strong effect on EAHB and plantains relative to other cultivars (Musabyimana et al., 2000). Control of nematodes through simple farmer management strategies is also difficult, especially in established plantations, as both the organisms and their effects are less visible than weevils or diseases (Kubiriba et al., 2016).

An important cause of high pest and disease pressure is the lack of effective management. Interventions such as uprooting infected mats and disinfecting farm tools are simple yet require a degree of knowledge and resources (Blomme et al., 2017). In other words, socio-economic background is an important determinant of management intensity (Karamura et al., 1998).

While the specific set of factors playing the largest role in limiting EAHB yields in a certain region has been demonstrated to vary substantially according to region or production system, research has thus far concentrated more on the effects of (and control methods for) pests and diseases than on abiotic constraints such as soil and climate variables (Wairegi et al., 2010). Additional research into how abiotic constraints influence EAHB yield and/or performance, and how abiotic, biotic and socio-economic variables influence each other, is required.

2.2 Land Evaluation

2.2.1 Biophysical and Economic Land Evaluation

Land Evaluation is defined as "*The process of assessment of land performance when used for specified purposes, involving the execution and interpretation of surveys and studies of land use, vegetation, landforms, soils, climate and other aspects of land in order to identify and make a comparison of promising kinds of land use in terms applicable to the objectives of evaluation*" (FAO, 1976). Land Evaluation is important as a method of matching a type of land use to a (subdivision of a) region with as goal the optimal productivity of resources for human use, while taking into account the prevailing conditions in the region, the available inputs and the land use type's ability to be applied sustainably in that particular region. Feeding an ever-increasing world population with a limited food production area is

also an important driving factor, as is coping with climate change, maintaining stable ecosystems across the globe and providing sustainable sources of energy (Bouma et al., 2011).

The FAO Framework for land evaluation, set up in 1976, is not a fixed and immediately implementable methodology, but rather a set of guidelines within the scope of which various methodologies, each tailored for a specific situation, may be set up (Verheye, 2002). It describes the land evaluation procedure as following a distinct set of steps, starting with the characterization of a Land Utilization Type (LUT). A LUT is defined as "*A specific land-use system with specified management methods in a defined technical and socio-economic setting, and with a specific duration"* (Rossiter et al., 1996). It does not necessarily entail the complete description of the land-use system, but rather the establishing of optimal and limiting values, also termed Land Use Requirements (LURs), for several key biophysical and socio-economic variables (Rossiter, 1996). These variables, the values of which in the study region are termed Land Qualities (LQ), are often complex and not directly measurable land properties. They are set up by integrating several directly measurable, estimated or simulated attributes of the land, termed Land Characteristics (LC). Land Qualities thus effectively bridge the gap between measurements quantifying the conditions on the ground in the study region, and one or more LUT, defined by their specific LURs. An example of a Land Quality includes the FAO's soil drainage classes, index values set up using soil texture, soil phase, slope, and soil type (Fischer et al., 2012). Subsequently, soil and climatic data is combined with crop growth specifications to divide the study region into Land Mapping Units (LMUs) (Verheye, 2002). LMUs can range in size from pixels or grid cells to entire map legend categories (Rossiter et al., 1996), and generally possess less intra-LMU variability than inter-LMU variability. Next, the LURs are 'matched' to the actual conditions in the study area, represented by LQs computed from quantitative and to varying degrees spatially explicit biophysical and socio-economic data. The goal of this 'matching' is to rate LMUs for the level at which they satisfy the LURs, which eventually leads to an overall assessment of the study region's suitability for the LUT in question.

The 1976 FAO Framework expanded upon earlier versions by extending the decision space to include not only soil variables but also climatic, socio-economic and landform variables. Also, it allowed for crop-specific assessments, as well as greater capacity to evaluate for more precisely defined LUTs, when compared to previous frameworks such as the USDA Land Capability Classification method (van Lanen, 1991). Because of this, it became necessary to rigorously define the LUT using quantitative criteria, in other words the LURs mentioned in the previous paragraph. Together, these innovations made it possible to characterize the potential of a region for LUTs of different levels of detail, ranging from a single crop type to a whole land use system with all its biophysical, socio-economic, cultural and management-related attributes, while also integrating factors from a plethora of different fields.

In theory, both biophysical and socio-economic variables were meant to be taken into account to fully characterize the production potential of the land in a study region. The FAO (FAO, 1976) broadly outlined two different strategies: a *parallel approach*, with biophysical and socio-economic assessments taking place simultaneously, and a *two-step approach*, in which biophysical land evaluation is followed by a separate socio-economic land evaluation step. However, in practice there has been little cooperation between the fields of exact sciences and economics and its more social science-oriented sub-disciplines. Biophysical land evaluation has taken precedence (FAO, 2007; Rossiter, 1995). In the more frequently employed two-step approach, socio-economic data is used either to validate the first step's findings and for initial, qualitative estimates of a LUT's applicability in a region, or it is not used at all (FAO, 1976; Verheye, 2002). Farmers and other land users have indicated that land evaluation procedures based on a narrow range of factors fail to adequately encapsulate their interests, and that information provided often forms an inadequate foundation for decision making (Jakeman et al., 2003).

Economic land evaluation follows a basic framework (Rossiter, 1995), starting with the identification of an economic measure, such as the internal rate of return or the benefit/cost ratio, which determines the economic suitability of the land. This economic measure is calculated using a combination of biophysical LC data and market information. The simplest method would be to calculate total net return from a LMU by multiplying the measured, estimated or simulated crop yield in the LMU with the crop's market price and subtracting costs. Social factors such as family size, non-farm income sources, etc. are commonly integrated in the economic land evaluation process (Rossiter, 1995).

While it is recommended to have experts from various fields working on any land evaluation project, supervised by a land evaluation specialist who can oversee the general procedure and the integration of each area of expertise into the Land Evaluation system (FAO, 2007), many land evaluation projects prioritize biophysical aspects and tend to be carried out solely by natural resource scientists. Economic land evaluation, on the other hand, cannot simply get by on either biophysical or economic expertise alone, but is defined by requiring the integration of both fields (Rossiter, 1995).

The greatest challenge in economic land evaluation lies in translating biophysical variables to the selected economic suitability measure (Rossiter, 1995). There are two approaches: using land characteristics and using land qualities. In the first approach, LCs are used in dynamic simulation models or other types of crop models to estimate crop yield and the levels of inputs required to apply a certain LUT. For example Bouraima et al. (2015) used reference evapotranspiration, soil depth, and available soil water to estimate irrigation requirements using the CROPWAT model. From the yield and input values, the total revenue and costs, and thus economic measures such as total net return, could be calculated. The second approach

uses the severity (or suitability) levels of LQs to define either yield reductions relative to a maximum attainable yield, yield delays or increased costs of remediation measures necessary to attain maximum yield despite limiting conditions (Rossiter, 1995). Economic land evaluation is well suited to convert these various manifestations of biophysical suitability into a single economic measure; yield and costs translate to total net return via summation after multiplication with market prices, and delayed yields can be added to the equation via discounting functions, in which future cash flows are calculated to have lower value in the present.

While the LQ-based method allows users to more clearly isolate the influences of individual LQs on yield, yield delay or remediation costs and the associated economic suitability measure, not every LQ has a discernible effect on yield, or can be remediated. This means that an eventual economic suitability figure will often take into account severity ratings of LQs expressed either as a decreased yield only, possibly including the discounted income from future yield, or solely as increased costs, or as a combination of both. This conclusion is also immediately relevant from an agricultural management point of view: given the biophysical and macroeconomic context, economic land evaluation can recommend optimal resource allocation strategies depending on whether land users value yield maximization, cost reduction or a more balanced alternative (Rossiter, 1995).

In traditional land evaluation, elementary land management units are typically determined using biophysical data and landscape features (FAO, 1976). However, in economic land evaluation, the fact that certain areas necessitate the same economic approach despite being biophysically heterogeneous (or different economic approaches despite being biophysically homogenous) can lead to evaluation units, and their associated suitability values and management recommendations being determined by economic rather than biophysical criteria (D. G. Rossiter, 1995). For example, it often makes more sense to assign labor costs to an entire farm than to divide them amongst different biophysical LMUs or management units, such as fields, greenhouses or soil types, within the farm. Larger regions, for example an irrigation district that requires region-wide planning of irrigation application, can also be used as evaluation units. Generally, if there is large heterogeneity within such an economically defined evaluation unit, subunits are first evaluated separately before an integrated land use management recommendation is formulated for the entire unit (Rossiter, 1995).

An advantage of economic land evaluation is that it theoretically necessitates (Rossiter, 1995) and in practice often incorporates (Johnson et al., 1991; Samranpong et al., 2009; Segerstedt et al., 2013) a sensitivity analysis, risk analysis or some other type of analysis that aims to demonstrate the probability of a certain economic scenario taking place. It thereby illustrates the impacts of the economic factors such

as price fluctuations or changes in labor availability, as well as errors in the underlying biophysical variables. These analyses frequently serve as tests of the assumptions and simplifications that economic land evaluation models are built upon, and can also shed light on the feasibility of alternative resource allocation strategies (Samranpong et al., 2009; Segerstedt et al., 2013). Furthermore, they tend to be especially highly valued by farmers, as they provide information essential to the survival of any agricultural enterprise (Samranpong et al., 2009), and arguably form an indispensable component of any land evaluation study that wishes to express effects of economic factors and attract the attention of land users.

While biophysical land evaluation is useful for establishing environmentally relevant assessments of suitability, it is often inadequate in providing results necessary for an allocation of land to LUTs that balances both long-term ecological functioning with short-term human needs. Economic land evaluation, ideally, takes into account the impacts of applying a certain LUT on both land users, determining the LUT's financial feasibility from the point of view of farmers or agricultural corporations, as well as on society as a whole (FAO, 1976). A certain flexibility in terms of formulation of results necessarily exists as well; there is not always a single 'best' LUT application strategy, as the many different actors in society often have conflicting interests and different methods of valuation. Strategies leading to community-centric improvements, such as the creation of jobs, are often weighed against those that bring about higher returns on investment for shareholders (FAO, 1976). Furthermore, while farmers' interest in the adoption of long-term, natural ecosystem-protecting measures depends on a host of factors, such as education level, availability of extension services, local socio-cultural values and the low financial risk of a measure (Smith et al., 2014; Wei et al., 2009), or the potential for profit (Greenland-Smith et al., 2016), progress in disciplines such as natural resources economics have made it possible to estimate the monetary value of services delivered by natural ecosystems (Swart et al., 2017). Economic land evaluation is thus capable of incorporating a wide range of factors, socio-economic as well as biophysical, over a wide temporal planning horizon, into an economic suitability measure, and is thereby useful in facilitating translatability of abstract, seemingly remote and unimportant biophysical factors to a more universal method of valuation that is understood by a variety of different actors (Rossiter, 1990). An example of this principle is the Automated Land Evaluation System (ALES), a computer program that enables users to build a customized biophysical land evaluation system along the lines of the 1976 FAO Framework (Rossiter, 2003; Rossiter, 1990). While not taking economic factors other than prices of inputs and outputs into consideration, ALES does offer users the chance to express biophysical suitability in economic terms, such as the net present value, gross margin or internal rate of return (Rossiter, 1995;

Rossiter, 1990). Economic limitations caused by suboptimal LC values are expressed as higher costs or lower yields.

Historically, the importance of economic land evaluation relative to biophysical land evaluation has undergone an evolution. The latter is based on LCs that tend to fluctuate relatively little, thereby giving an adequate picture of a study region's invariable characteristics (FAO, 2007). It is, however, a type of land evaluation ill-suited to portraying technological and economic change, which occurs over a shorter time scale (Johnson et al., 1996). In FAO-style land evaluation as described in the 1976 framework, unsuitable (class 'N') LMUs are subdivided into Permanently Not Suitable ('N1') and Temporarily Not Suitable ('N2') LMUs. Permanent unsuitability means that land does not meet the physical requirements for a certain LUT, and is not likely to meet them in the foreseeable future, making economic land evaluation of such LMUs unnecessary (Rossiter, 1995). Temporarily unsuitable land could be suitable based on biophysical LCs, yet would incur too many costs or would be otherwise economically unsuitable upon installation of a certain LUT (FAO, 2007). There is thus an argument to be made for a two-step procedure, using biophysical land evaluation first to give a broader, more permanent assessment and following up with economic land evaluation to determine where investments can best be made in the short term (FAO, 2007). In the FAO's Agro-Ecological Zoning (AEZ) method, for example, socio-economic considerations are indeed brought into consideration only at a later, planning- and management-related stage of the land evaluation process, after suitability assessments have already been carried out based on biophysical variables. In an example of the relevance of this approach, low-resolution land evaluation approaches in northern Portugal which considered mainly pedological factors resulted in the misclassification of productive agricultural land as better suited for forestry. This prompted the proposal of guidelines for agricultural land evaluation research projects that include integration of socio-economic factors from the household level (e.g., likelihood to emigrate, non-farm employment) to the level of national trends in the application of LUTs (de Carvalho et al., 1989).

Furthermore, in recent times the emphasis in the discipline of land evaluation has shifted. While the 1976 FAO framework recognized the need for land evaluation to recommend land use management options as favorable only if they could sustainably maximize agricultural production without depleting available resources, and historically placed a large importance on crop yield, the revised 2007 FAO framework expanded its focus to include aspects of social equity, economics and ecosystem functioning as components to be considered when assessing the sustainable suitability of a LMU (FAO, 1976; FAO, 2007; Verheye, 2002). Social equity concerns have led to a greater importance of the inclusion of all stakeholders in the process of defining LURs and the objectives of a land evaluation study. LUTs are becoming more multifunctional in their use, such as forests being used for recreation as well as purely for

timber production, and the importance of new functions is being realized, such as the water storage capacities of a grassland (Sonneveld et al., 2010). Land evaluation has, in recent decades, also become increasingly intertwined with many other disciplines. Economic, social and environmental concepts, such as water quality and carbon storage, have become more important (FAO, 2007). These phenomena all favor a land evaluation that integrates biophysical and socio-economic data in a methodology specially adapted to facilitate such integration, such as the LEFSA sequence, which attempts to combine the large-scale, biophysical characteristic-centered approach of traditional land evaluation with the farming system-scale, socio-economic factor-centered Farming Systems Analysis (FSA) methodology (Fresco et al., 1990). This framework is applied iteratively, with land evaluation supplying information on the most suitable areas to apply a certain LUT according to biophysical factors, and FSA complementing this by giving detailed feedback from the land user's perspective about the constraints and potentials facing application of said LUT (Fresco et al., 1990).

Van Lanen mentions 'integral land evaluation' as a discipline incorporating physical aspects to determine suitability of LMUs and generate management recommendations for a LUT clearly defined by societal needs and economic specifications (Smit et al., 1986; van Lanen, 1991). One integral land evaluation project in Ontario, Canada, was able to provide quantitative recommendations for the location and duration of application as well as the potential agricultural crop productivity of different LUTs across all the LMUs in the study region. These recommendations varied according to which food production target was specified and considered the effects of market forces on domestic food production needs (Smit, 1984). A project carried out in northern Australia's Herbert River District acknowledged the necessity for a land evaluation approach integrating biophysical and economic aspects over the entire process, from inputs to methodology to output (Johnson et al., 1991, 1994). It incorporated expert knowledge via an ALES-constructed system as well as crop simulation models to produce predictions of crop yield and net returns, and a risk analysis for a single LUT, namely sugarcane production, while emphasizing updatability and communicability of conclusions towards land managers (Johnson et al., 1991).

One of the causes of the imbalance between economic and biophysical land evaluation may be that research priorities, LURs and methodologies are most often directed by scientists and researchers, namely the *supply side*, rather than farmers and actual 'land users', henceforth referred to as the *demand side* (Bacic et al., 2003). In other words, the information is directed by those supplying it, rather than by those who could use it to take decisions. The MicroLEIS land evaluation decision support system, for example, was found to have been used primarily by university teaching staff. While private consultants also made up a large proportion of the total, farmers did not (De La Rosa et al., 2004). In developing countries, results of Land Evaluation studies most often pass through intermediaries such as agricultural extension

workers, adding an extra link in the communication chain (Bacic et al., 2003). Recommendations include allowing the demand side to feature more and earlier on in the decision-making and planning stages of the Land Evaluation process, so that priorities and obvious information shortages may be made clear. It is also important that communication between the supply and demand sides continues after the initial transfer of information, facilitating user-inspired adjustments in the approach (Bacic et al., 2003).

It can be concluded that biophysical and economic land evaluation each have distinct but complementary uses. Regarding the inclusion of both biophysical and socio-economic factors, and communicability of results with and involvement of land users, economic land evaluation is superior. However, biophysical land evaluation still has an important role to play, especially as ecosystem functions and services such as carbon storage or disciplines requiring detailed management recommendations, such as precision agriculture, grow in importance (Bouma et al., 2011). It is pivotal, however, that biophysical land evaluation be clearly situated within the context in which it is used, so that the necessary stakeholders are involved in the process and can receive any research results.

2.2.2 Qualitative and Quantitative Land Evaluation

In land evaluation, a broad division may be made between qualitative and quantitative methodologies. The 1976 FAO framework for land evaluation characterizes quantitative land suitability classification as "*a land suitability classification in which the distinctions between classes are defined in common numerical terms, usually economic, which permit objective comparison between classes relating to different classes of land use"* (FAO, 1976). These classification systems generate suitability values on a continuous scale, commonly by using formalized mathematical systems (De La Rosa et al., 2002). Qualitative methodologies, on the other hand, express suitability not on a continuous scale but in discrete albeit ordinal categories, and use systems largely defined by expert or traditional knowledge (De La Rosa et al., 2002). Due to their being defined on a continuous scale, quantitative land evaluation methods permit higher discernibility between two different LMUs for a given LUT than do qualitative methods.

In qualitative systems, expert knowledge is generally used to set up the parameters determining Land Use Types, to select LCs relevant to the suitability classification process and their limiting values, and to add any necessary specifications to the decision system used to set up the suitability classes. An example of a qualitative method is the maximum-limitation 'matching tables' method, in which the theoretical ranges of various LCs are divided up into sub-ranges, each of which defines a suitability category. In a purely qualitative method, the limits of these sub-ranges are determined by expert, traditional or empirical knowledge. LC data values in a LMU are then 'matched' with a table containing all the suitability categories. A LQ's suitability category is determined solely by its constituent LC with the lowest suitability category, and a LMU's suitability category in turn depends only on the LQ with the lowest

suitability category of all LQs relevant for determining suitability, known as the *most limiting* LQ. An alternative is the use of decision trees, in which the decision rules at the nodes as well as the trees' architecture are determined by expert knowledge (Wandahwa et al., 1996).

This expert knowledge is preferably embedded in a computer system capable of automatically applying said knowledge to determine suitability of a LMU, given the necessary inputs, and which makes it easy to adjust methodologies to fit different environments and scenarios. An example of this is the ALES land evaluation system, discussed in the preceding chapter. It enables experts to build land evaluation frameworks based on an iterative modeling process, during which decision trees are set up to guide the matching of LCs with LURs, and initial values characterizing LURs may be adjusted at any point. The resulting evaluation procedure is validated using LMUs for which extensive documentation exists regarding degree of productivity and the mechanisms contributing to it (Rossiter, 1990). While capable of providing estimates of yield and economic suitability expressed as predicted gross margins (Rossiter, 1990), these are at least partly based on experts' construction of decision trees, which by definition can result only in a finite number of branches and thus a finite number of estimated values. Land evaluation frameworks set up using ALES are therefore described as qualitative.

A definition of quantitative land evaluation was proposed by van Lanen (1991), who defined quantitative land evaluation as a system expressing suitability by assigning a value directly relatable to crop yield or another measure of productivity. Certain methods, while expressing suitability as a continuous variable, do not fulfill this criterion. An example are most parametric methods, which use expert knowledge to assign numerical scores to each LC or LQ, which are in turn mathematically processed, either via additive or multiplicative functions, to result in a productivity or suitability index value. This index value often gives an indication of suitability rather than a quantitative expression directly proportional or otherwise mathematically relatable to a measure of productivity (van Lanen, 1991).

Besides a minority consisting of expert knowledge-based parametric methods, the majority of quantitative land evaluation systems are based on the use of mechanistic, deterministic simulation models (Henny A J van Lanen, 1991). In these systems, a crop or LUT's performance is predicted based on the existing soil, landscape and management conditions and historically representative climate conditions. These conditions, represented by biophysical and socio-economic data, are input into simulation models that are built to represent the basic biological and physical processes, such as plant allocation of assimilates or soil water availability, that determine a crop or LUT's productivity or viability. Suitability is thus expressed as a figure calculated using extensive, often spatially referenced datasets processed by complex, field-

validated models. These models, while classified as quantitative under the 1976 FAO definition, do not always fall into this category under van Lanen's definition.

Such simulation models belong to the category of biophysical models, systems that allow users to predict the performance of a land use type in an area without actually implementing said LUT in that area (Rossiter, 2003). Biophysical models are not necessarily the same as land evaluation systems and demonstrate varying degrees of overlap. The 1976 FAO framework for land evaluation is described as a biophysical model which is also a land evaluation system, whereas other biophysical models are implemented as components embedded in a larger land evaluation system, such as the integration of risk analysis, a biophysical crop simulation model and expert systems described by Johnson & Cramb (1991). Recently, simulation models have also begun to play a larger role in land evaluation, as a host of factors have caused demand for quantitative evaluation to rise. These factors range from stricter, more detailed legislation defining groundwater nutrient limits to the extrapolation of consumer- and retailer-driven restrictions on food products sold in developed countries to production areas in developing countries (Bouma et al., 2011). Biophysical models are characterized by, among others, their *degree of computation*. A quantitative model has a high degree of computation, and will express suitability with greater resolution, for example as a precisely estimated crop yield figure, thereby facilitating differentiation between suitability ratings. A qualitative model has a low degree of computation and assigns LMUs or study regions to broader suitability classes, within which considerable variation could still exist, yet between which differentiation would be impossible using only the model's final results (Rossiter 2003). This is in agreement with the previously stated definitions of qualitative and quantitative land evaluation systems (De La Rosa & van Diepen, 2002; FAO, 1976), except for van Lanen's (van Lanen, 1991). Therefore, it can be stated that a land evaluation methodology employing a qualitative biophysical model will most likely be a qualitative land evaluation methodology, and likewise for methodologies containing quantitative biophysical models.

In general, it can be said that quantitative methods require a more in-depth analysis than qualitative methods, but also have the potential to yield a more detailed suitability assessment. In qualitative land evaluation, the most limiting LCs are often easy to identify, as the decision structure is very transparent. Crop yield figures calculated in quantitative methods, however, are often the result of a complicated modeling process (Johnson et al., 1994; Keating et al., 2003), in which the influences of individual LCs on the final result are often hard to discern, even though some work has gone into developing systems consisting of multiple easily replaceable modules, each representing a scientific discipline or sub-discipline (Jones et al., 2003).

In quantitative land evaluation, the final result of the suitability assessment leads to specific input recommendations, such as inorganic fertilizer requirements in $kg.ha^{-1}$, whereby the recommendation for each individual input may vary with the calculated crop yield or other suitability value. In qualitative land evaluation, fixed sets of recommended input levels are usually defined per suitability class (van Lanen, 1991).

Quantitative land evaluation has several additional advantages over its qualitative equivalent, the first of which is its ability to work with probability distributions of LQs and LCs over a temporal continuum, as opposed to being limited to LQs based on historical average values (van Lanen, 1991). Moreover, quantitative methods are based on models of universally applicable mechanistic processes and can therefore be easily adjusted to simulate environmental conditions at a finer scale, being limited only by the resolution of the input data. They can also estimate suitability of a LMU for a LUT under conditions that currently do not yet exist, especially as, with the progression of climate change and the development of new hybrid crop varieties, the effectiveness of land evaluation methods based on experience of situations in the past has been called into doubt (Bonfante et al., 2015). Qualitative methods, not being built upon the universal process simulations that form the basis of quantitative evaluations, rely more upon empirical relationships and previously encountered situations, which drastically limits the amount of potential situations that can be assessed for suitability (van Lanen, 1991).

However, quantitative methods are often more time-consuming, data-hungry and require more computing power (van Lanen et al., 1992). The two methods are seen as complementing each other, with qualitative methods appropriate for broader, large-scale assessments, which lead to identification of areas where a more detailed, quantitative evaluation would be relevant. In a project carried out in the Netherlands to assess suitability for potato cultivation, areas determined to have low potential according to an initial qualitative evaluation, such as steep slopes or areas with shallow soils, were left out of subsequent quantitative analyses. Quantitative evaluation, which included determination of probability distributions of several management-related LQs as well as crop yield simulation, focused on high-potential areas, contributing to the cost-effectiveness and time-efficiency of the project (van Lanen et al., 1992).

Many land evaluation systems incorporate elements of both qualitative and quantitative land evaluation, up to the point that such a level of interdependence and integration is reached that it becomes hard to implement the different elements separately and still obtain meaningful results. An example is the AEZ approach mentioned in the previous chapter, which provides an initial, quantitative suitability assessment for crops or LUTs in a zero-limitations scenario by using modeled estimates of climatic parameters such as reference evapotranspiration (ET_0) and length of the growing period (LGP) to calculate yield and

potential biomass. In a second, qualitative step, agro-edaphic limitations to yield are then taken into account by use of expert-determined limiting values for important constraining factors, such as soil depth, and result in 5 discrete suitability categories (Fischer et al., 2002). While these climatic and agro-edaphic assessments comprise two separate steps, each is an integral, non-negligible element of the unified AEZ framework. Another method predicted maize and sunflower yield based on a formula incorporating harvest index, respiration coefficient, adjusted leaf area index, gross biomass production and length of the growing period, along with multiplication by a soil characteristic-based correction factor (Rabati et al., 2012). The obtained predicted yield value was, however, subdivided into suitability classes according to the FAO guidelines (S1: high suitability; S2: moderate suitability; S3: critical suitability; N: not suitable), using empirically determined multiples of critical yield as the separation points between classes. Thus, while this land evaluation system would be classed as qualitative when considering system output alone, the underlying yield estimation mechanisms constitute components typical of a quantitative land evaluation process. Several other systems are heavily based on the 1976 FAO framework, yet attempt to incorporate new elements, such as standardization of the LC data to the same scale using techniques such as interpolation to pre-determined scoring categories (Kalogirou, 2002) and linear fuzzy membership functions (Nouri et al., 2017). It is common for systems to initially express suitability as a continuous variable, yet then proceed to reclassify this result into discrete suitability categories, using expert-defined category limits (Rabati et al., 2012) or other methods, such as the Jenks Natural Breaks reclassification (Nouri et al., 2017).

In general, there has been a trend of combining the strengths of both qualitative and quantitative land evaluation (De La Rosa et al., 2002), namely the former's incorporation of expert reasoning and the latter's ability to simulate dynamic natural processes, such as soil water flow and daily weather patterns, allowing the generation of certain LQs at a fine temporal scale and over a contiguous spatial area, which can then be input in expert systems or other simulation models.

It can be concluded that the majority of land evaluation studies making the distinction between qualitative and quantitative land evaluation stick to the FAO's 1976 definition (FAO, 1976), with important additions being made by (De La Rosa et al., 2002) and (van Lanen, 1991). Quantitative land evaluation methods are defined by their output, which should represent a real productivity value and is defined on a sliding scale, and by the fact that they often employ biophysical models with a high degree of computation. Qualitative land evaluation methods produce output expressed as discrete suitability categories that do not directly represent or are not convertible to a crop yield figure or other productivity value, and generally employ biophysical models with a low degree of computation that use expert knowledge to define model parameters and structure. The advent of powerful simulation models and the

wide availability of spatially referenced biophysical data have, however, catalyzed the introduction of elements of quantitative land evaluation into qualitative frameworks, which is why systems that mix elements of both qualitative and quantitative land evaluation are currently more prevalent than purely qualitative land evaluation.

2.2.3 Target Population of Environments and Land Evaluation

Most plant breeding has emphasized quantitative traits such as yield and resistance to pests and diseases, rather than adjusting to the often small-scale environmental variability in a region (Cooper et al., 1996). Such broad adaptation of certain key traits has often produced high-performing cultivars adapted to a wide range of environments, for example for maize in southern and eastern Africa (Windhausen et al., 2012). However, these cultivars may not always be well-adapted to all the specific stress patterns and small-scale environmental and management conditions within a region (Heinemann et al., 2016). The performance of a certain genotype may vary between different environments, and indeed can depend on a complex interplay of genotypic variation-induced phenology, timing and intensity of environmental stress, as well as interaction between nutrient supply and pest and disease pressure (Chenu, 2015; Chenu et al., 2009). This range of effects is described by the so-called 'genotype-environment interactions' (GEI). Two genotypes may exhibit fundamentally different phenotypic characteristics in a certain environment, while performing similarly in the breeding trial environment (Cooper et al., 1996). Strong GEI can reduce the heritability, or the proportion of phenotypic effects explainable by genotypic effects, in breeding trials, making it harder to breed for specific traits. Therefore, the objective in crop breeding is to isolate genotype effects on performance as much as possible from the more dominant environment and GEI-related effects (Chenu, 2015). To this end, multi-environment trials (METs) were introduced. These trials are run over the course of several years and take place in various distinct locations. However, selection of high-performing genotypes often happens based on METs thinly spread across a broad spatial scale and run during a limited time span. These METs may therefore fail to capture the spatial and temporal heterogeneity of environmental conditions, and thus genotype-environment interactions (Chenu, 2015). This can result in breeding programs producing mismatched varieties that, for example, perform well under optimal, low-stress conditions, yet are not viable during fluctuations of environmental variables that did not occur within the scope of the METs, or in distinct sub-environments that were not represented by any of the trial locations.

Instead of simply increasing the amount of randomly selected trial locations and the duration of the experiments (and thus the level of investment), recent publications have emphasized an approach that meticulously characterizes the possible growing environments, and uses the information gained to better direct breeding efforts. The concept of a Target Population of Environments (TPE) serves as a useful

framework for directing said breeding efforts. A TPE is defined as "*a set of environments, including spatial and temporal variability, to which improved crop varieties developed by a breeding program need to be adapted"* (Alexandre B Heinemann et al., 2016; Nyquist et al., 1991). Clustering a variety of conditions, such as climate and soil variables, into unique environment groups, can help pinpoint the traits that crop breeders need to look for in a cultivar, rather than relying on broad and often contradictory parameters.

A TPE is characterized by key classes of sub-environments, within each of which crop genotype performance is influenced by the environment in a similar manner and which contain relatively high internal homogeneity (Chenu, 2015). Once the spatial distribution and characteristics of these sub-environment classes are known, it becomes possible to breed, and eventually introduce into the field, unique new crop varieties that include adaptations to the particular limitations and stresses encountered in each portion of the area covered by the TPE. The sub-environments of a TPE typically fall into one of the following categories: *Mega-environments* and *environment types* (Chenu, 2015). The former are described as spatially explicit areas, demonstrating a high degree of internal similarity in terms of the biotic and abiotic stress types present, agricultural management requirements, and market and production characteristics (Braun et al., 1996). Environment types, conversely, are not bound to a specific location, but rather defined as a combination of temporally variable factors, such as stress pattern or ranges of climatic factors. Multiple environment types can occur at the same site, and an environment type could be present at a site one year, and absent the next.

There are three different categories of mechanisms and criteria by which areas within TPEs are assigned to a certain sub-environment: yield-based, pedo-climatic based and stress index-based TPE classification. Yield-based classifications have as advantage that they subdivide a TPE based on a criterion that is also used to measure production and is moreover the result of a complex interplay of environmental, genotypic and GEI-factors. A subdivision based on yield therefore ensures that all these factors are, to a certain degree, taken into account (Chenu, 2015). Classification may be carried out using trial site data. It is important for breeders to understand in what way their trials represent the TPE and, if applicable, its sub-environments. Therefore, characterization based on simulated crop yields, which express environmental variables over the entire spatial range of the TPE, can often correct for any incongruences caused by incomplete representation of the TPE by trial data. This, however, assumes that sufficiently high-resolution and complete data as well as a model able to integrate all relevant variables is available. Pedo-climatic classification approaches bypass yield and subdivide a TPE based directly on biophysical data (Chenu, 2015). However, yield- and pedo-climatic factor-based approaches assume that created mega-environments or environment types are permanent once they have been defined, whereas in reality

conditions can change. Therefore, stress index-based TPE classification takes into account the specific combination of site, genotype, year and management effects that influence a crop's performance (Chenu, 2015). A certain site or area could thus fall into a different sub-environment, depending on the season.

Land Evaluation and TPE methodologies are closely related: the latter may be seen as an iteration of the former. The TPE approach also strives to recommend the most suitable Land Use Type (LUT) for a certain region, albeit with a focus on matching of the optimal genotypes to the optimal environments and the identification of breeding priorities. Land Evaluation is often much broader than the TPE approach: instead of matching genotypes to a set of conditions or a geographical area, it can recommend detailed land use strategies involving management specifications, optimal combinations of species to grow at a location, and impacts on local populations and markets. In the characterization phase, however, both methodologies use spatially referenced trial, survey or simulation data as base. Both also integrate various degrees of static and dynamic modeling into their frameworks geared towards the prediction of genotype, cropping system or Land Use Type performance (Chenu, 2015; Rossiter, 2003). The TPE approach's focus on extensive characterization of the environment and the subsequent identification of any significant subdivisions could be described as a rigorous identification of LMUs in Land Evaluation. The TPE method often establishes sub-environments based on a limited amount of criteria, such as measured or simulated crop yield (Heinemann et al., 2011, 2015). While crop yield may integrate the effects of many different environmental, genotypic and GEI factors, the resulting sub-environments are often only cursorily defined in terms of the environmental and management factors contributing to yield. Land Evaluation, of which the suitability classification system is by definition based on a broad range of factors, can be used to complement the TPE approach by providing more arguments for the chosen TPE-subdivision and also by pre-testing the suitability of any newly developed specific sub-environment-targeted genotypes for their respective sub-environments.

What is proposed in this project is utilizing a MCDM embedded in a Land Evaluation framework to help establish LMUs in the EAGL region where EAHB cultivars have a high chance of successful introduction. This can also be seen as establishing classes of sub-environments within the TPE. Whether this subdivision leads to adapted breeding strategies or not is outside the scope of this work. Currently, only those subdivisions with the least limitations to EAHB growth are of interest; further work could identify low-limitation areas with respect to the NARITAs, with areas that are limiting for classic EAHB but suited to specific types of NARITA being of especial interest.

3 Materials and Methods

3.1 Study Areas

Collection and analysis of data was carried out for two different geographic levels. The larger level, also referred to in the text as the study region, consisted of the East African Great Lakes (EAGL) region. While often referred to as the area comprising Uganda, Rwanda, Burundi, part of the highlands in the eastern Democratic Republic of the Congo (DRC), North-west Tanzania and Kenya (Eledu et al., 2004; Kamira et al., 2016; Ndabamenye et al., 2013; Ndabamenye et al., 2013), exact demarcations of this region are rare. The EAGL level used in this project is based on a set of maps of the principal banana growing regions in the East African Highlands (Davies, 1995; Karamura, 1998). Areas such as Tanzania's coastal or Uganda's Northern regions, where EAHB is not commonly grown, are included to provide a continuous surface for analysis. While finding ECV geodatasets for a large area such as the EAGL geographic level was perfectly feasible, collection of actual productivity or performance data within the scope of a MSc thesis research project was constrained by time and resources. Therefore, a second geographic level, itself a smaller portion of the EAGL level, was defined. This level consisted of the Ugandan banana growing (UBG) regions, consisting of the Southwestern, Western, Central and Eastern regions of Uganda.

3.2 Data

3.2.1 ECV Data

To address the first part of Research Question 2 (see section 1.4), it was necessary to obtain data describing the edapho-climatic conditions in the study region. To this end, georeferenced soil and climate data with high spatial resolution, the result of two different past projects by soil and climate researchers (Hengl et al., 2015; Hijmans et al., 2005), were sourced from online platforms (Table 3.1). Data of nine different variables were downloaded. These variables were selected because of the importance to EAHB production as described in literature (Section 2.1.3, Table 2.2), as well as their frequently being measured in EAHB trials. Besides the nine directly downloaded variables, the additional variable 'K/Mg ratio' was calculated from the 'soil K^+' and 'soil Mg^{2+}' geodatasets. The data covered the entire EAGL geographic level, as defined in the previous section. A subset of this data was used in the analysis of the UBG geographic level. The data had, during the past projects by the aforementioned soil and climate researchers and before being downloaded for use in this project, undergone different treatments or been collated from a different set of source datasets. The soil data downloaded for this project (ISRIC, n.d.) had been produced via an automated soil mapping process, which predicted a range of soil variables using a collation of a spatially and temporally extensive soil profile observations as base data (Hengl et al.,

2015). The downloaded climatological data (Hijmans et al., 2005), on the other hand, had been generated through spatial interpolation of existing average monthly climatological data from weather stations.

Table 3.1: Basic metadata pertaining to the publically available datasets.

	Spatial Resolution	Geospatial coordinate system	Units	Variables used	Institution
WorldClim Data	30 arcseconds, approx. 0.93km	GCS_WGS_1984	°C * 10, mm	Average monthly temperature, average monthly precipitation.	Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis
ISRIC Data	250m	WGS_1984_Lambert Azimuthal_Equal_Area	Various (see Table 2.2)	pH, Soil Organic Carbon (SOC), Clay percentage, Exchangeable Ca^{2+}, Mg^{2+}, K^{+}, Total Soil N, K/Mg ratio	Africa Soil Information Service (AfSIS), ISRIC

3.2.2 Data for Model Parameter Establishment

To answer the second part of Research Question 2 (section 1.4), the degree to which EAHB's agro-ecological LURs were met in the study region had to be ascertained. While the study region's agro-ecological conditions were already described by the ISRIC and WorldClim datasets, LURs remained to be defined and also had to be translated into parameters that could be entered into both the performance and suitability models mentioned in section 1.3. To this end, data were downloaded from *AgTrials*, an online field trial data repository (CGIAR Research Program on Climate Change and Agricultural Food Security (CCAFS), 2013). This was done in accordance to the 'No Derivatives' license applicable to the data, as data were used only as input for further calculations and were thus not fundamentally adapted or exactly reproduced. The data came from trials carried out in the context of the *"Management practices and opportunities in East African highland banana (Musa spp. AAA-EA) production in Uganda"* (Wairegi, 2010). The dataset included measured bunch weight data, as well as data on the same ten ECVs contained in the WorldClim and ISRIC datasets. All data had been collected in the Western, Southwestern, Central and Eastern regions of Uganda. For more specifics on how this data was transformed into parameters for the two prediction models, readers are referred to sections 3.3.2 and 3.3.3.

3.2.3 EAHB Performance Data

3.2.3.1 Data from Previous Projects at IITA

With the WorldClim, ISRIC and AgTrials data described in the above sections, as well as the performance and suitability prediction models, it was possible to answer Research Questions 1 and 2. To

answer the question of how closely model results approximated actual EAHB production figures (Research Question 3), some of the latter needed to be obtained. To this end, data was received from a staff member at the local IITA office in Kampala, Uganda. While these data were also collected as part of the *"Management practices and opportunities in East African highland banana (Musa spp. AAA-EA) production in Uganda"* project(Wairegi, 2010), source locations were different from the AgTrials data point locations. The data were collected in 2006 from 7 different districts of Uganda, namely Bududa, Manafwa, Sironko and Mbale in the Eastern Region, Bushenyi in the Southwestern Region and Masaka and Rakai in the Central Region. The data contained observations of allometric banana plant traits used for a bunch weight estimation technique described in (Wairegi et al, 2009). This technique uses a fitted regression (Eq. 3.1) that, with reasonable trustworthiness ($R^2 = 0.73$), relates bunch weight to a set of allometric variables, namely the girth of the banana pseudostem at its base, the girth of the pseudostem at 100 cm height, the amount of hands in the bunch and the number of fingers in the outer whirl of the second-most distal hand (Figure 3.1). This relationship provided an estimate of EAHB bunch weight which has been proven to be valid independent of plant growth stage, region, cultivar and soil and foliar nutrient concentration (N, P, K, Ca and Mg). This allowed bunch weights to be estimated even though the input data for the process was collected under a wide array of different conditions and originated from many different plants. The resulting values are henceforth referred to as estimated actual bunch weight (BW_{est}) values.

$$BW_{est} = \exp[-8.908 + 0.561 \times \ln(Hands) + 0.482 \times \ln(Fingers) + 0.925 \times \ln(V_{stem})]$$

$$V_{stem} = \frac{100}{12\pi} \times \left[G_{base}^2 + G_{1m}^2 + (G_{base} \times G_{1m})\right] \qquad \text{(Eq. 3.1)}$$

with V_{stem} being the estimated pseudostem volume, G_{base} the girth of the pseudostem at base height and G_{1m} the girth of the pseudostem at a height of 1m.

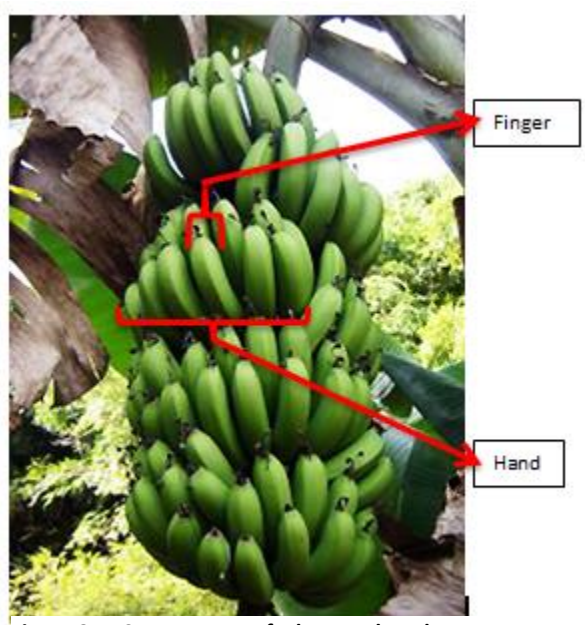

Figure 3.1: Components of a banana bunch.

The dataset also contained administrative information on the location of collection, but no geographic coordinates, so locations of data were set per village using an online locator tool (Ssemulugo et al., n.d.). Where multiple farms had been surveyed in the same village, observations were averaged to arrive at single data points per village. After this process, the dataset contained 38 data points,

representing 38 villages spread over the 7 districts mentioned above.

3.2.3.2 Kabarole District Data

The bunch weight estimation data obtained from IITA did not cover some important banana-growing regions, notably Kabarole District in the Western Region. Thus, the same type of allometric data as described in the previous section was collected in several locations throughout the region, and the same type of methods were used to calculate an additional set of BW_{est} values.

To streamline the usually time-consuming process of survey taking and data recording, a mobile, tablet-based survey was set up using SMAP software, which allowed users to set up surveys tailor-made for a particular data collection purpose, and which made it possible to automatically export recorded data into an Excel file (Figure 3.2).

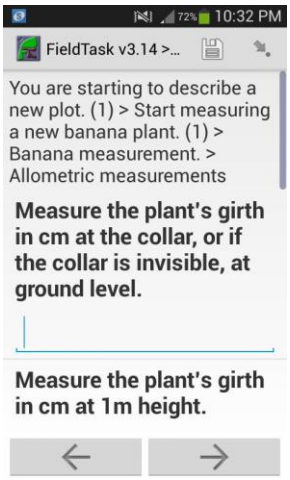

Figure 3.2: The SMAP digital survey interface.

During the fieldwork, villages were randomly selected from administrative lists of sub-counties where EAHB production was known to be high. Inside the villages, a stratified random sampling strategy using area of land owned, amount of cattle or livestock owned, and quality of building materials used in the house as stratification criteria was applied to select ten farms per village. In each farm, ten EAHB mats, which include the mother plant and all its forthcoming suckers or shoots, were randomly selected for measurement of the allometric data by walking along a zigzag trajectory through the farmers' banana producing plots.

The data collected in Kabarole contained 38 data points, from 38 different farms. These data points were combined with those received from IITA to form a new dataset, henceforth termed the Ugandan Banana Growing areas dataset, or UBG dataset. This georeferenced dataset included 76 data points, each with its own BW_{est} value.

3.2.3.3 Production Data for the EAGL Geographic Level

Datasets describing performance (in the form of BW) were, until now, all derived from the UBG geographic level. In order to fully answer RQ 3, it was necessary to evaluate how well predicted suitability and performance approximated expectations based on actual performance in the EAGL geographic level as well. To make this possible, actual production data (BW data were not available) for several countries in the EAGL geographic level was sourced from publicly available datasets (Table 3.2), and a series of maps was created displaying these data. The data on harvested area necessary for conversion of these production figures into productivity figures was sadly not available for all countries in the region. Most data were sourced from agricultural surveys carried out by national statistical agencies

or other branches of government, the exception being the HarvestChoice dataset, which was set up by an eponymous independent organization. This collection of datasets has several limitations, including the fact that data collection did not happen in the same year across the different countries, and that the production figures measured usually included substantial amounts of non-EAHB cultivars. Furthermore, sub-national data for Kenya was not available, nor was sub-province data for the DRC's Orientale province. The portions of Kenya and the DRC's Orientale province that belong within this project's delineation of the EAGL region were thus not displayed on the maps.

Table 3.2: Production datasets for the EAGL Geographic level.

Dataset name	Spatial resolution or basic spatial unit	Period of collection	Units of measurement	Reference
UCA (Ugandan Census of Agriculture)	District (Ugandan administrative unit)	2008-2009	Tons (of cooking and beer banana)	UBOS, 2010
NSCA (National Sample Census of Agriculture, Tanzania)	Region (Tanzanian administrative unit)	2007-2008	Tons (all types of banana)	NBS, 2012
NAS (National Agricultural Survey, Rwanda) 2008	District (Rwandan administrative unit)	2008	Tons (of cooking and beer banana)	NISR, 2010
ENAB (Enquete Nationale Agricole du Burundi)	Province (Burundian administrative unit)	2011-2012	Tons (of cooking and beer banana)	Sidi et al., 2013
Annuaire Statistique 2015 (DRC)	Province (DRC administrative unit)	2015	Tons (of beer banana)	INS, 2017
HarvestChoice	5-arc minute grid cell	2005	Tons (of rainfed banana)	HarvestChoice, 2015

3.3 Data Processing

3.3.1 Operations using GIS-software

For visualization of data as well as performing calculations of the Predicted Potential Bunch Weight (BW_{pred}) and the Distance to the Ideal Point (DIP), using the performance and suitability models mentioned in section 1.3, ArcMap software version 10.3 was used. ISRIC and WorldClim data on the ten ECVs was imported, converted to a raster format, clipped with the boundaries of the East African Great Lakes region (as previously defined) and subsequently resampled and snapped so that the raster data layers for the ten ECVs all overlaid each other perfectly and all had the same 1 km x 1 km resolutionSubsequently, the georeferenced ECV data from the ISRIC and WorldClim datasets were

processed into new raster data layers portraying the DIP and BW_{pred} values for the entire EAGL geographic level. These reworked datasets together were named the *EAGL dataset*, used only to refer to the data contained in the ten ECV layers, and thus not to be confused with the EAGL geographic level in which it is located. Furthermore, a raster layer containing data on the most limiting variable according to the performance model was also produced (see section 3.3.2). The 76 data points in the UBG dataset (located in the UBG geographic level), did not yet contain data on the 10 ECVs specified in section 3.2.1. However, due to location data being available for each data point, it was possible to match data points with EAGL dataset pixels (and their corresponding ECV values) in which the data points were located. An overview of all datasets as they were used in data processing is given in Table 3.4. An outline of what the different datasets were used for is given in Figure 3.3. Details on these operations are given in sections 3.3.2 and 3.3.3.

Table 3.4: Overview of main datasets. Legend for acronyms in figure: ECV = Edapho-climatic variable, BW_{est} = Estimated actual bunch weight, BW_{pred} = Predicted potential bunch weight, DIP = Distance to the Ideal Point.

Dataset	Observations	Variables
AgTrials	179 data points	10 ECVs, BW_{est}
Ugandan Banana Growing	76 data points (64 used in analysis)	10 ECVs, BW_{est}, BW_{pred}, DIP.
East African Great Lakes	1 489 151 pixels	10 ECVs, BW_{pred} (for part of dataset only; see 3.3.1), DIP.

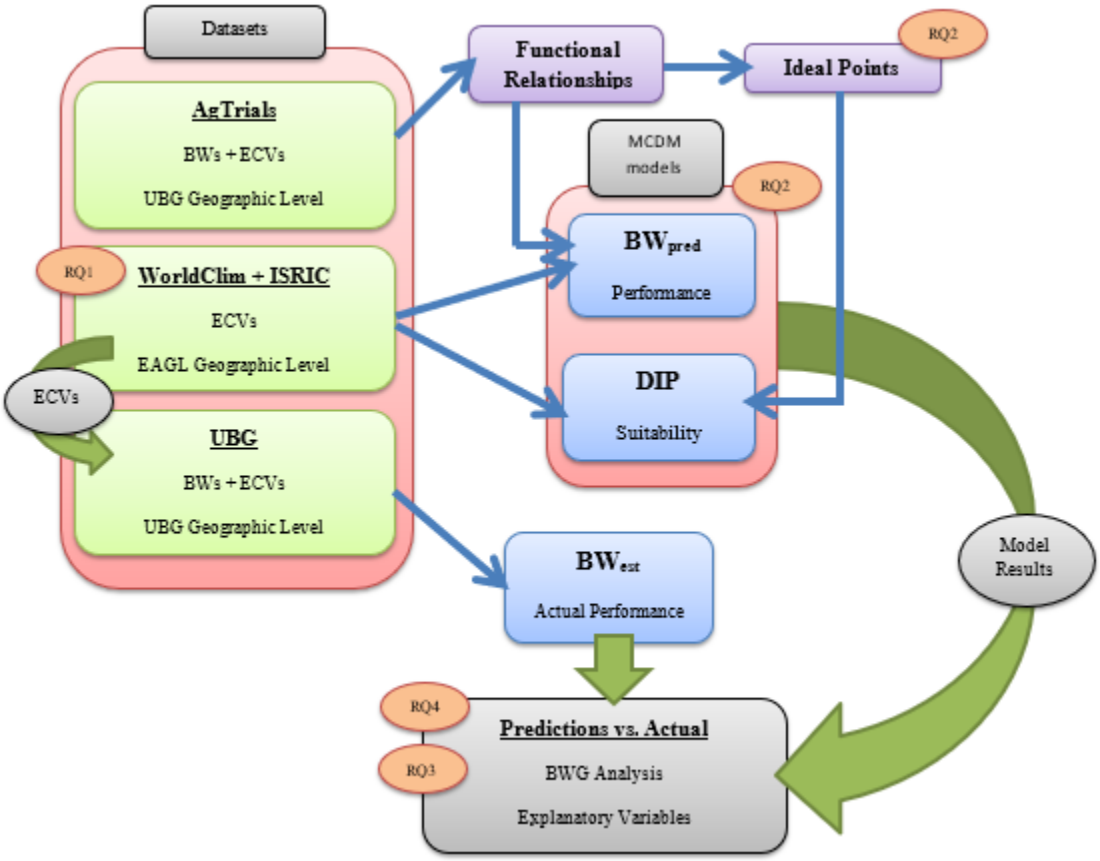

Figure 3.3: Overview of the different datasets and what they are used for, as well as where Research Questions (RQ) 1-4 are answered. Legend for acronyms in figure: BW = Bunch Weight, ECV = Edapho-climatic variable, UBG = Ugandan Banana Growing, RQ = Research Question, MCDM = Multi-Criteria Decision Method, BW_{pred} = Predicted potential bunch weight, BW_{est} = Estimated actual bunch weight, DIP = Distance to the Ideal Point, BWG = Bunch weight gap.

3.3.2 Functional Relationships and Predicted Potential Bunch Weight

In order to execute the performance model mentioned earlier, data from the AgTrials database were used to set up boundary lines (also termed functional relationships) between the 10 ECVs and the bunch weights measured in Uganda's Western, Central and Eastern Regions. The purpose of these relationships was to predict bunch weight based on a single ECV. First, the dataset was divided into 10 two-dimensional sub-datasets, one being bunch weight and the other being the 10 edapho-climatic variables. Secondly, the Mahalanobis distance, a measure expressing the multivariate distance between a point and a distribution, was used to remove outliers. The data points for which this distance was greater than 10.8276 (equivalent to $p = 0.001$ and $df = 1$ in the Chi-squared table) were removed from their respective two-dimensional sub-datasets. Thereafter, to set up the boundary lines, the BOLIDES system was used as

described by (Schnug et al., 1996). This involved recalculating each data point's measured bunch weight value according to a piecewise constant upper boundary step function for data point clouds of each variable, and subsequently fitting a 4th-degree polynomial to the data points (Figure 3.4). All functional relationships were set up according to this method, except for the bunch weight-SOC relationship, which was amended from a 4th-degree polynomial to a 2nd-degree polynomial due to the relationship assuming unrealistic (negative) values for bunch weight.

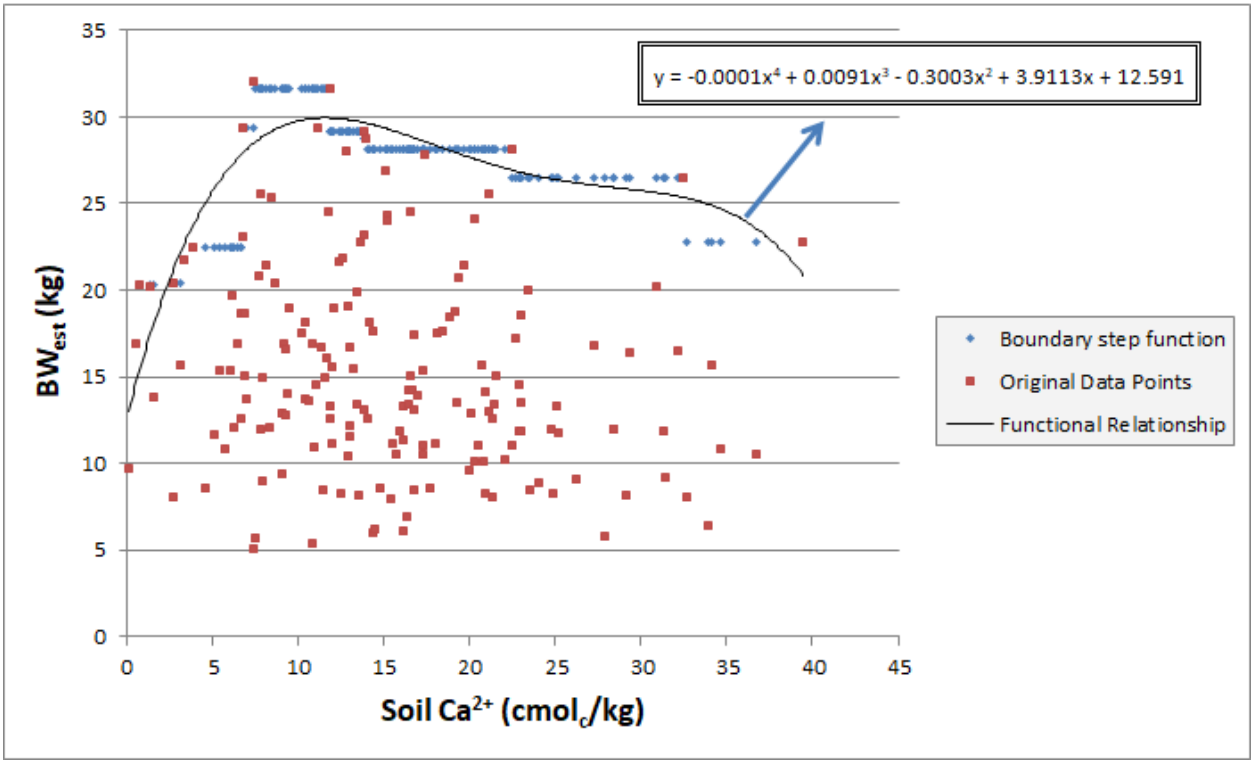

Figure 3.4: Illustration of the process of setting up a boundary line. Red points are original data points, blue points represent their new coordinates after recalculation with the boundary step function.

While the AgTrials dataset contained points from various regions in Uganda, and there was a precedent for working with region-specific functional relationships (Wairegi et al., 2010), there were not enough data points from each region to compute region-specific functional relationships. Therefore, functional relationships were set up using all of the dataset's 179 data points and were assumed to be identical for all EAHB producing regions of Uganda (Table 3.5). They were also assumed valid for the non-Ugandan areas of the EAGL geographic level.

Table 3.5: Functional relationships between bunch weight and 10 edapho-climatic variables, set up using data from the AgTrials dataset.

Variable [Unit]	Relationship
pH [-]	$y = 0.2224 \times x^4 - 4.645 \times x^3 + 31.089 \times x^2 - 64.056 \times x + 11.397$
Clay [%]	$y = -0.00007 \times x^4 + 0.0088 \times x^3 - 0.3971 \times x^2 + 7.6769 \times x - 22.86$
Soil Ca^{2+} [cmolc.kg^{-1}]	$y = -0.0001 \times x^4 + 0.0091 \times x^3 - 0.3003 \times x^2 + 3.9113 \times x + 12.591$
Soil Mg^{2+} [cmolc.kg^{-1}]	$y = -0.0016 \times x^4 + 0.0474 \times x^3 - 0.867 \times x^2 + 7.1209 \times x + 11.925$
Soil Organic Carbon (SOC) [g SOC.kg^{-1}]	$y = -0.1387 \times x^2 + 4.8553 \times x - 10.493$
Total Soil Nitrogen [g N.kg soil^{-1}]	$y = -3.2727 \times x^4 + 24.038 \times x^3 - 67.609 \times x^2 + 85.951 \times x - 9.977$
Average Annual Rainfall [mm]	$y = 0.00000000007 \times x^4 + 0.000000008 \times x^3 - 0.0001 \times x^2 + 0.2141 \times x - 90.469$
K/Mg ratio [-]	$y = -441.57 \times x^4 + 987.02 \times x^3 - 766.49 \times x^2 + 241.82 \times x + 3.7232$
Average Annual Temperature [°C]	$y = -0.6165 \times x^4 + 50.315 \times x^3 - 1540.8 \times x^2 + 20985 \times x - 107236$
K^+ [cmolc.kg^{-1}]	$y = 0.0031 \times x^4 + 0.1546 \times x^3 - 2.2488 \times x^2 + 6.8069 \times x + 24.239$

Subsequently, the functional relationships were used to compute a series of predicted potential bunch weights (BW_{pred} values), one per edapho-climatic variable (ECV), which meant ten ECV-specific BW_{pred} values per pixel in the EAGL dataset. Of these ten BW_{pred} values, the smallest value was taken as the final BW_{pred} in an application of Liebig's Law of the Minimum (Eq. 3.2).

A proportion of data points in the UBG dataset as well as pixels in the EAGL dataset had one or more ECV values outside the AgTrials dataset's ECV ranges, where the functional relationships were not defined. These points and pixels were left out of this analysis. This resulted in 64 remaining data points and only 235,846 pixels (or 15.8 % of the total) in the UBG and EAGL datasets respectively for which the BW_{pred} was calculated and considered valid.

$$BW_{pred,i} = \min\{f_K(x_{K,i}),\ f_{Mg}(x_{Mg,i}), f_{Ca}(x_{Ca,i}), f_{pH}(x_{pH,i}), f_{SOC}(x_{SOC,i}), f_{TSN}(x_{TSN,i}),\ f_{K/Mg}(x_{K/Mg,i}), f_{Clay}(x_{Clay,i}), f_{prec}(x_{prec,i}), f_{temp}(x_{temp,i})\} \quad \text{(Eq. 3.2)}$$

with $BW_{pred,i}$ being the predicted potential bunch weight in pixel i of the EAGL dataset, f_K the functional relationship for soil K, and $x_{K,i}$ the value for soil K in point i.

The AgTrials dataset provided data for soil organic matter content (SOM, %). This contrasted with the ECV dataset, whose ISRIC source data provided soil organic carbon content (SOC, %), but no SOM. In order for a functional relationship for SOC to be set up using points from the AgTrials dataset, so that, in turn, BW_{pred} could be computed for SOC, the SOM values in the AgTrials data had to be converted to SOC values Traditionally the Van Bemmelen factor (0.58) has been used. However, a recent publication cast doubt on the relevance of this factor, stating it was based on an original measurement of the carbon content of humic acid, not SOM in general (Pribyl, 2010). While coming with the caveat that a single factor can never be representative of the full scope of variation in SOM carbon content, a revised factor of 0.5 was proposed and has been used in this work.

Besides the Law of the Minimum-calculated BW_{pred} value, the particular ECV contributing to the final, minimum BW_{pred} value was recorded, and is henceforth termed the "most limiting variable". Also, Bunch Weight Gaps (BWGs) were calculated by subtracting BW_{est} values from the corresponding BW_{pred} values. These BWG values can be seen as a partial answer to Research Question 3 (see section 1.4): a large absolute value for BWG indicates that the Law of the Minimum model's prediction does not approximate actual bunch weights in the geographic level, as represented by the BW_{est} values, very well, while small absolute values for BWG mean the model's prediction was accurate.

3.3.3 Compromise Programming to Calculate Distance to the Ideal Point

In order to carry out a suitability prediction while taking into account the ten ECVs, a MCDM called CP was used (Estrella et al., 2014). This measure typically involves calculation of a continuous function that defines distance to an ideal point (DIP) according to the formula below (Eq. 3.3.).

$$Distance\ to\ ideal\ point\ for\ cell\ i\ and\ variable\ j$$

$$= \sqrt{[\left(\sum_{j=1}^{n}\left(\frac{ideal\ point\ value(j) - data\ point\ value(i)}{\max(j) - \min(j)}\right)^2\right)]} \qquad \text{(Eq. 3.3)}$$

The ideal point (IP) was defined as a point in a ten-dimensional space, with the ideal point's coordinates (IPCs) determined by the maxima of the functional relationships in Table 3.5. This could be seen as a more quantitative alternative to the literature-derived LURs in Table 2.2 and is the definitive way in which Research Question 1 (See Section 1.4) is answered. However, there were a number of complications in this process. Where the functional relationships exhibited more of a plateau or maximal range than a single maximum point, this entire range was taken as the *ideal range*, within which distance to the ideal point coordinate in question was equal to zero. Further complications included the fact that while the original formula called for the use of the difference between an independent variable's (in this case an ECV's) ideal and anti-ideal points in the denominator, almost all of the ideal points were (except in the case of precipitation) located in the middle of the variable's range, making it hard to determine a

single anti-ideal point. This meant that a point with a x-coordinate of ($x_{Ideal\ Point} - a$) was the same distance removed from the IPC as a point at ($x_{Ideal\ Point} + a$), with an arbitrary value. Thus, quotient terms with denominator values given by an ECV's complete range were used (Table 3.6). While this diminished the eventual DIP value to values generally between 0-0.5 (instead of 0-1.0 as in the original formula), taking measures such as assigning a separate quotient term to data points above and below the ideal point coordinate would give the points from the smaller range (and smaller denominator) a greater weight than those from the broader range. Therefore, ranges were set up only by removing Mahalanobis outliers from complete ECV ranges in the EAGL region, determined from the ISRIC and WorldClim datasets. The variables *K/Mg ratio*, *Clay percentage* and *Average annual precipitation* had ideal ranges, not points. In these cases, numerators in quotient terms were defined as distances to the edge of the ideal range when points were located outside the range and were equal to zero when points were located inside the ideal range. Therefore, these three ECVs were the only ones for which separate lower and upper quotient terms were used. The DIP was calculated for all pixels in the EAGL dataset, and thus also for the UBG dataset.

Table 3.6: Ideal point coordinates and quotient terms for each ECV. Lower range terms are for the data points with ECV values lower than the IPC, and upper range terms for those above the IPC.

Variables	Ideal Point Coordinate (IPC)	Lower range quotient term	Upper Range quotient term
K/Mg	Ideal range from 0.303-0.828	(0.308 - x) / 0.921 (When ECV value below ideal range)	(x – 0.828) / 0.921 (When ECV value above ideal range)
Average annual precipitation [mm.yr^{-1}]	Ideal range: all values > 1506.39	(1506.39 - x) / 1726	0
Average Annual Temperature [^{0}C]	21.105	(21.105 - x)/ 14.1	
Mg^{2+} [cmolc.kg^{-1}]	7.035	(7.035 - x) / 8.005	
K^{+} [cmolc.kg^{-1}]	1.908	(1.908 - x) / 2.082	
Ca^{2+} [cmolc.kg^{-1}]	11.55	(11.55 - x) / 25.52	
TSN [gN.kg soil^{-1}]	1.753	(1.753 - x) / 3.54	
pH	5.94553	(5.94553 - x) / 3.6	
Clay percentage [%]	Ideal range from 22.017 - 43.906	(22.017 - x) / 25.111 (When ECV value below ideal range)	(x – 43.906) / 25.111 (When ECV value above ideal range)

SOC [g SOC.kg soil^{-1}]	17.503	(17.503 - x) / 41

It was hypothesized that certain ECVs would be consistently closer to their ideal points than others. To determine for which variables this was the case, the DIP was recalculated for each ECV individually according to Eq. (3.4), resulting in ten DIP values calculated in one-dimensional space, one for each ECV. Different quotient terms were applied for variables with an ideal range instead of an ideal point, in the same manner as in the original calculation of the DIP (section 3.3.3), and calculations were carried out for both the EAGL and UBG datasets.

$$SDIP\ for\ variable\ j\ in\ data\ point\ i$$

$$= \sqrt{\left(\frac{ideal\ point\ value(j) - data\ point\ value(i)}{\max(j) - \min(j)}\right)^2} \qquad \text{(Eq. 3.4)}$$

In data analysis, SDIP values were also often used instead of ECVs, due to their essentially representing the same data while also incorporating information in IPCs.

3.3.5 Cluster Analysis

A large region like the EAGL geographic level is characterized by an extremely heterogeneous set of environments. To better understand spatial patterns of predicted performance, suitability, gaps, limiting variables and ECVs, it was deemed useful to divide the EAGL level into multiple, characteristic units. To set up these units, a formalized statistical algorithm was needed. Cluster analysis is a technique that defines multivariate clusters of data points exhibiting a certain degree of similarity with (or proximity in multivariate space to) each other. Each cluster also shows dissimilarity (or multivariate distance) to any other clusters. This fits the description of LMUs in section 2.2, namely that inter-unit variability is significantly larger than intra-unit variability. Cluster analysis was therefore deemed a suitable technique and was thus carried out on all four sets of variables mentioned in the paragraph above, for the EAGL region using RStudio software (RStudio, 2015).

Clustering results are highly variable depending on the clustering algorithm used and the clustering validity measure (also known as a stopping rule) employed. Hierarchical methods define a series of cluster solutions by agglomerating data points into clusters one by one, or by dividing an initial single cluster containing all points into smaller clusters. These methods proved to be too computationally intensive to run with RStudio when carried out on a dataset as large as the EAGL dataset. Non-hierarchical k-means clustering, which requires fixing the amount of clusters prior to carrying out the analysis, requires less computing power, and while sometimes outperformed by other types of non-hierarchical clustering (Beynon et al., 2012), or performing badly when data is skewed, remains one of

the commonly used, easy to implement and historically effective clustering algorithms (Jain, 2010; Liu et al., 2010). Therefore, k-means clustering was carried out on the EAGL dataset's ECV and SDIP values, with the prespecified number of clusters varying from two to the maximum amount possible with the available computing power.

To select the optimal number of clusters for each set of variables, and thus to arrive at a clustering solution that most adequately represented real-life subdivisions in the data, Pseudo-F values (also known as *Calinski-Harabasz* index values) were used as a validity measure. Identifying a single validity measure that adequately represents quality of clustering is often difficult: measures are often biased towards one or another type of algorithm, are sensitive to effects of noise or different densities of data, and have been tested only on unrealistically strongly clustered datasets (Lamirel, 2016). The Pseudo-F value or *Calinski-Harabasz* index, while generally performing well, is unsuited to higher levels of noise in clustered data (Guerra et al., 2012; Liu et al., 2010; Milligan et al., 1985). Although other indexes, such as the *S_Dbw* and *Gamma* measures, demonstrated stable performances across a range of different categories, datasets and clustering algorithms and can thus compensate for this deficiency (Guerra et al., 2012; Liu et al., 2010), the Pseudo-F value was selected for its ease of computation and previous record of good performance relative to other validity measures. (Milligan et al., 1985). The clustering solution with the largest Pseudo-F value was assumed to have the smallest within-cluster sum of squares and the largest between-cluster sum of squares, making it optimal.

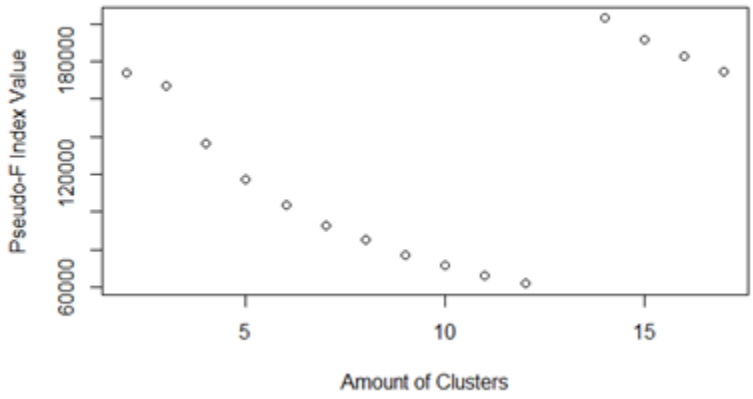

Figure 3.5: Pseudo-F index values for different clustering solutions using single-variable distance to the ideal point (SDIP) values.

This method indicated the 14-cluster solution obtained from the EAGL dataset's SDIP values as optimal (Figure 3.5). All of this solution's 14 clusters contained >1% of pixels. Clustering carried on ECV values did not result in higher Pseudo-F index values than clustering carried on SDIP values, and the most

optimal clustering solution provided information only on a limited number of clusters (4). While providing some insights as to the broader trends, the ECV values-based clustering solutions were disregarded in the remainder of the analysis in favor of the 14-cluster SDIP values-based solution, which with its 14 relevant clusters was able to provide more detailed information on the various sub-regions of the EAGL region. Further results of cluster analysis are given in section 4.4.

3.4 Data Analysis

In this project, data analysis largely revolved around locating and quantifying bunch weight gaps in the UBG dataset, in order to find an answer to RQs 3 and a partial answer to RQ4. A strictly quantitative analysis of gaps was indeed only possible with the UBG dataset, due to the lack in this project, of high-spatial resolution actual bunch weight or other performance data, and thus of quantitative bunch weight or other gaps for the EAGL geographic level. Additionally, performance predictions, represented by BW_{pred} values, could only be computed in a small part of the larger geographic level. However, some actual production data on a coarser scale were available for large portions of the EAGL level (section 3.2.2.3), as was the DIP (representing suitability). This made identification and explanation of gaps possible in the EAGL level, too, albeit not in a quantitative manner, and based more on visual observation of maps. Gaps analyzed in the UBG dataset were still termed BWGs, while the less strictly quantitative gaps in the EAGL dataset were named suitability gaps (SGs). A conceptualization of this two-pronged approach towards gap analysis is depicted in Figure 3.7.

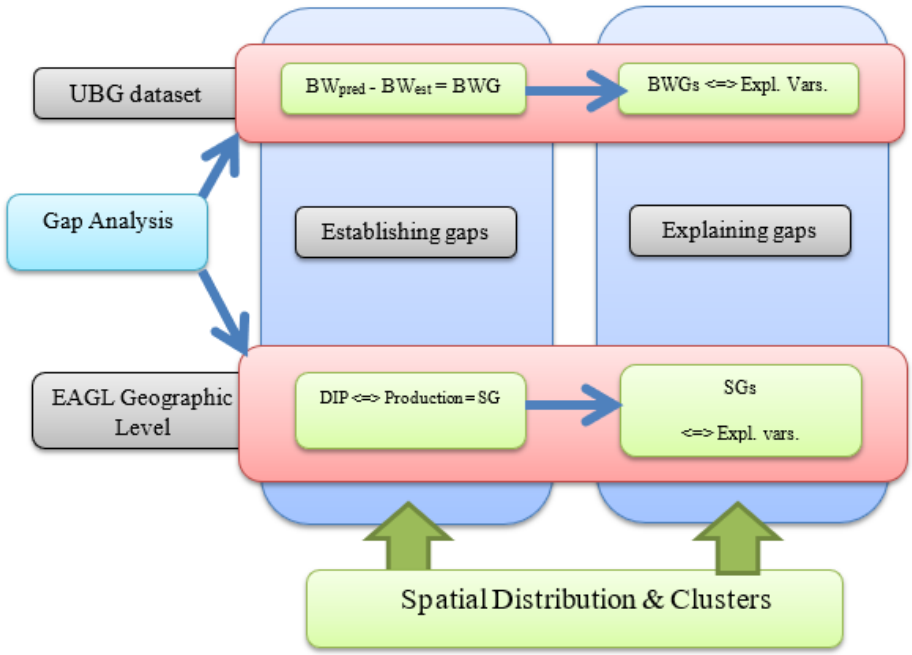

Figure 3.7: Conceptualization of data analysis, with different steps and datasets used. Legend for acronyms in figure: ECV = Edapho-climatic variable, UBG = Ugandan Banana Growing, BW_{pred} = Predicted potential bunch weight, BW_{est} = Estimated actual bunch weight, DIP = Distance to the Ideal Point, BWG = Bunch weight gap, Expl. vars. = Explanatory variables, SG = Suitability Gap.

In section 1.5, it was hypothesized that the size and spatial distribution of BWGs (and, by extension, the SGs in the EAGL geographic level) could be related either to explanatory variables or inadequacies in the performance and suitability models used. Several situations were, in fact, possible (Table 3.7). In one of these, gaps were non-existent, in which case the models adequately described reality, and ECVs were the only variables determining actual production and suitability. Another possibility was for gaps to be present, and for the ECVs' influence on actual suitability and performance, while not telling the entire story, to be correctly represented by the models. In this case, the remaining proportions of gaps were explained by the explanatory variables. A complication, however, was posed by the possibility that models insufficiently portrayed the ECVs' real effects. The very existence of gaps, not to mention which variables explained them, was made uncertain by this possibility. However, quantifying the degree to which models were accurate was outside the scope of this data analysis. This project's general assumption, during data analysis, remained that the models were correct. Any issues surrounding the inadequacy of models were considered in the Discussion, as were effects of explanatory variables.

Table 3.7: Three possible situations that could result from analysis of bunch weight and SGs. ECV stands for 'Edapho-climatic variable'.

Gaps	Variables explaining gaps	Models able to correctly convey influence of ECVs?
Non-existent	None	Yes
Present	Explanatory variables	Yes
Present/Non-existent?	Explanatory variables/ECVs/both?	No

3.4.1 Gap Analysis

3.4.2.1 Bunch Weight Gaps in the UBG dataset

In the UBG dataset, it was possible to quantitatively analyze BWGs. Data points were divided into different groups according to cluster (= sub-environment) membership. This allowed the occurrence of significant differences in population mean BWGs between the different groups, as well as the presence or absence of BWGs inside each group, to be tested. This division also facilitated the making of conclusions on the spatial distribution of different (combinations of) variables influencing the size of the BWGs. When significant inter-cluster or -limiting variable differences in population mean BWGs existed, the significance of population mean differences between the same pair of data point groups was tested for DIP and all ECVs, SDIPs and explanatory variables. Cluster- or limiting variable-derived groups were further characterized by the presence or absence of BWGs, determined by testing for significant differences between population means of BW_{est} and BW_{pred} within each group. In the absence of gaps, it was assumed that the Law of the Minimum model was an adequate predictor for bunch weight (see first situation, Table 3.7).

3.4.2.2 SGs in the EAGL Geographic Level

Due to the EAGL dataset's enormous size (1,489,151 pixels) and the often-low resolution or spatially irregular unit division of production data in the dataset, analysis of gaps at the EAGL geographic level was often based solely on visual analysis of maps. The only variables boasting spatial resolutions of 1x1 km and for which data were available across the entire geographic level were the ECVs, SDIPs, DIP and cluster membership. First, areas with high and low SGs were established. Secondly, any overlaps or spatial congruencies between these regions and ECVs, SDIPs, explanatory variables and clusters were identified, along with any obvious spatial congruencies *between* these variables.

3.4.2.3 Differences of population means

To determine the significance of inter-group differences in population means, or for the presence of BWGs, formal statistical tests were used. Depending on whether the data exhibited normality and heteroscedasticity or not, different tests were carried out (Figure 3.6). Statistical tests were carried out in pairs: an initial test with H_0: "There is no significant difference between the population means from which the two samples were taken", which established the presence or absence of a significant difference between the means of at least one pair of samples, and a subsequent post-hoc test that established precisely which the sample pairs in question were. Samples in this case refer to cluster- or limiting variable-derived groups in the UBG dataset. The tests are all suited to data taken in unpaired or independent samples.

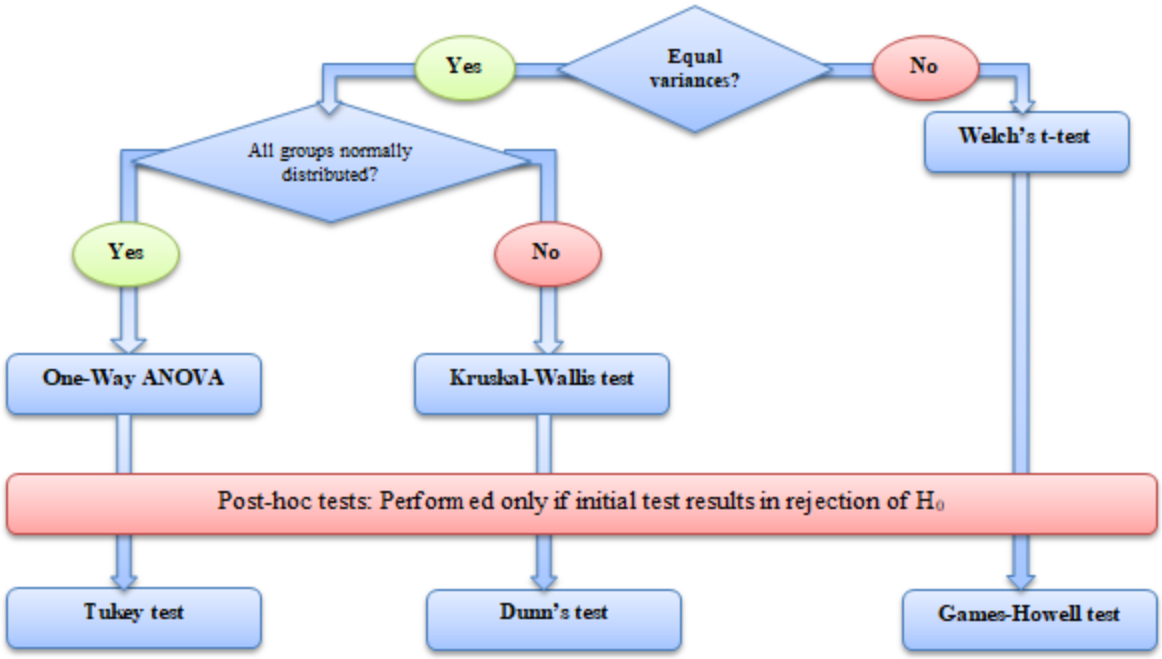

Figure 3.6: Decision framework for application of statistical tests to determine differences between population means of samples taken from different sets of data points.

Besides testing for differences *between* groups, testing for significance of difference between population means of BW_{est} and BW_{pred} was also an important part of the analysis. This involved paired samples, meaning that the values of BW_{est} and BW_{pred} between which significances of differences of population means was being tested belonged to the same group or dataset. Thus, a set of tests for paired (or dependent) samples was used, once again involving testing for normality and heteroscedasticity. The correct test for each situation was selected using the decision framework in Figure 3.7.

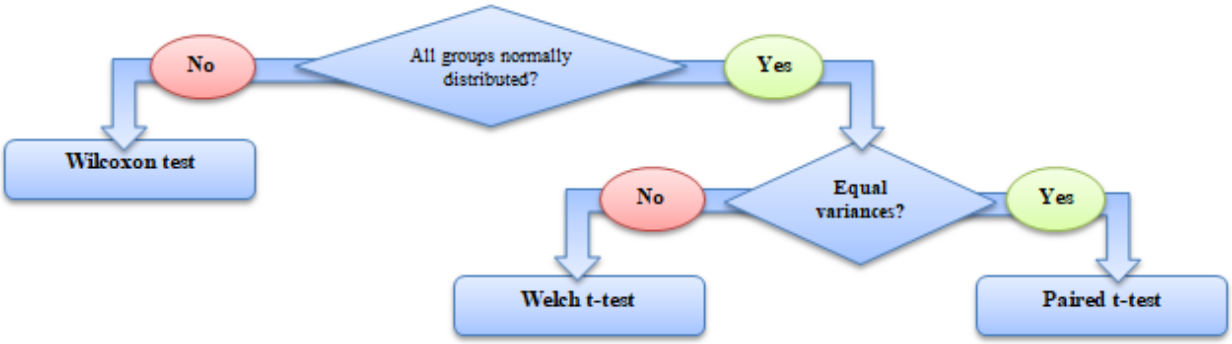

Figure 3.7: Decision framework for application of statistical tests to determine differences between population means of two samples taken from the same set of data points.

4 Results and Discussion

4.1 Production Data in the EAGL Geographic Level

Production figures for various countries in the EAGL region are given in Figure 4.1 (NBS, 2012; NISR, 2010; Sidi et al., 2013; UBOS, 2010). For Uganda, Rwanda and Burundi, figures represent cooking and beer banana production per administrative unit, which consists largely of EAHB. In Tanzania, there is no distinction made between banana types, with figures representing total banana production. The majority of production in the Arusha, Kilimanjaro, Kagera and Mbeya regions can be assumed to be EAHB (Byabachwezi et al., 2005; Maruo, 2002). In the DRC, production figures for beer banana (the only other categories given were cooking banana and plantain, which aren't represented within the EAHB subgroup) were available per province, yet only the North- and South-Kivu provinces fall entirely inside the EAGL region as delineated for this project. Only the easternmost part of the Orientale province, for which separate production figures were unavailable, fell inside the region. Total annual production of beer banana in 2015 was 479,414 tons in North-Kivu and 262,585 tons in South-Kivu. Production figures dated from different years (See section 3.2.3.3). No sub-national statistics on banana production of any kind were available at time of writing for Kenya. However, to fill in these gaps, and to provide sub-province or -region-scale data, the HarvestChoice dataset was used to create the map in Figure 4.2.

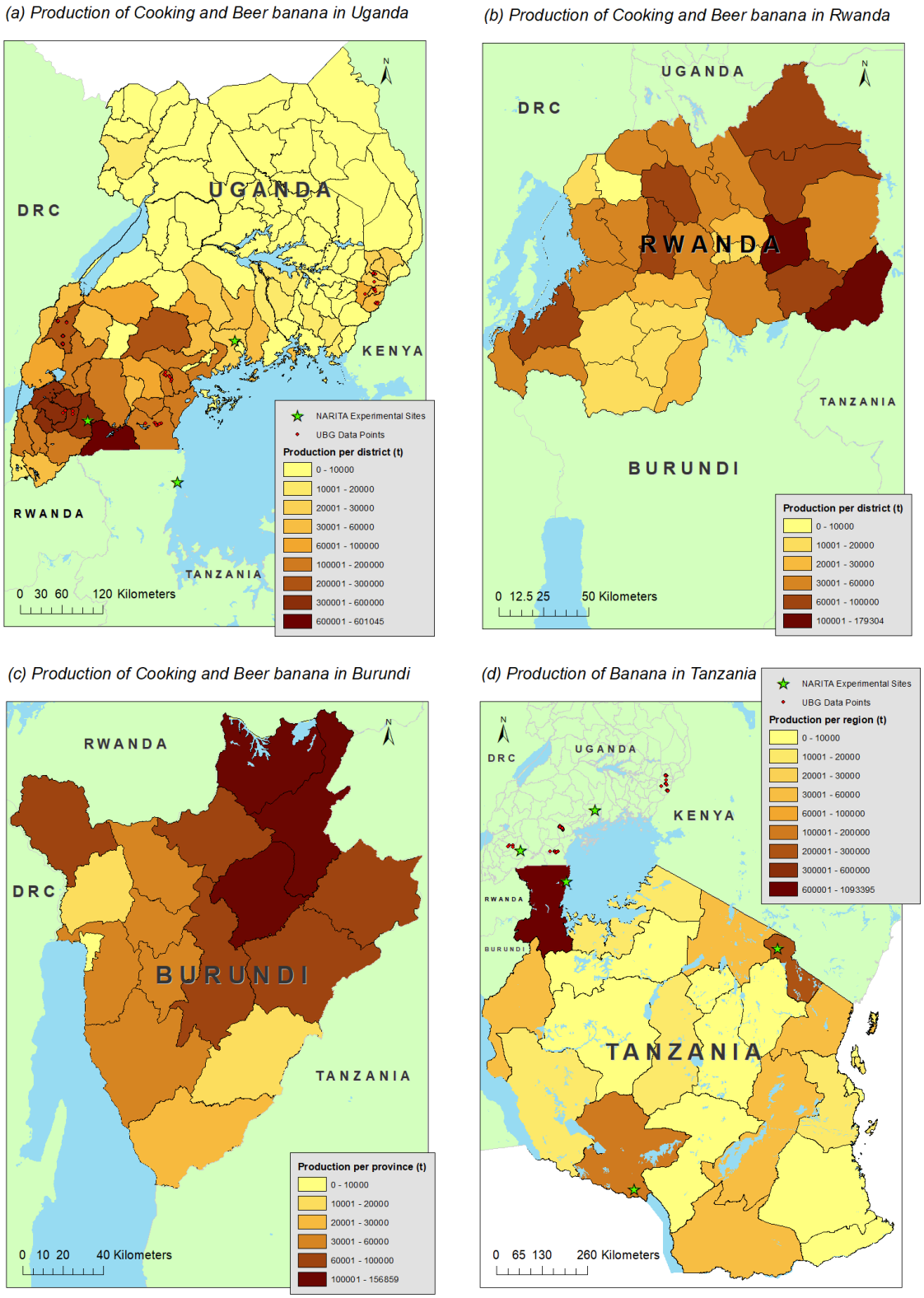

Figure 4.1: Production figures per administrative unit for the countries of (clockwise from top left) Uganda, Rwanda, Tanzania and Burundi.

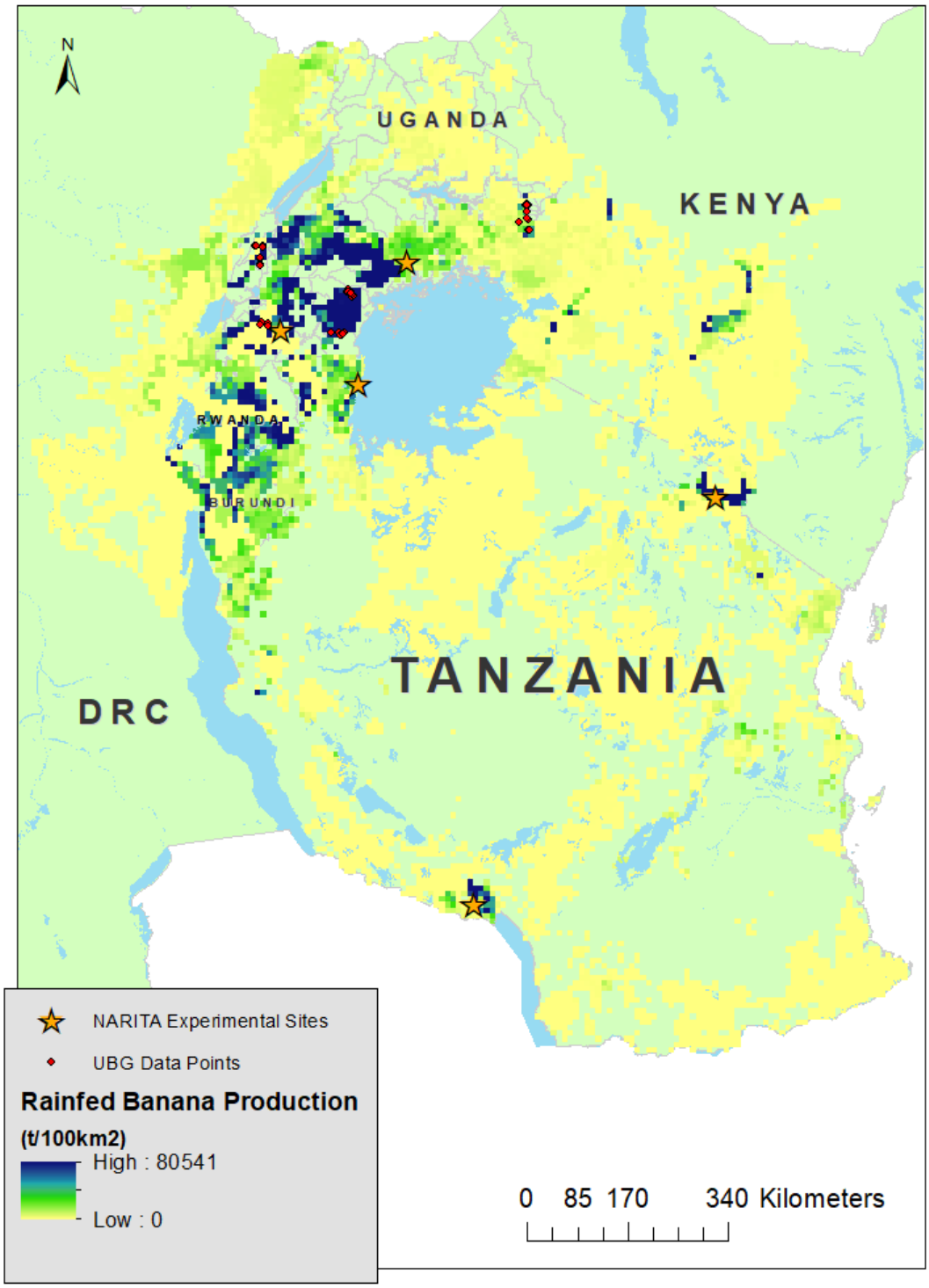

Figure 4.2: Production data for Rainfed banana across the EAGL geographic level.

4.2 Distribution of SDIP and DIP at the EAGL Geographic Level

As an answer to the first part of RQ2 (Section 1.4), the edapho-climatic conditions in the EAGL region as retrieved from the WorldClim and ISRIC datasets are represented by Figure 4.3, in which the 10 ECVs mentioned in Section 3.2.1 are displayed. The second part of RQ2, namely how well EAHB's LURs were fulfilled in the region, is addressed by the DIP and its single-variable sub-terms (SDIPs, see Figures 4.4 and 4.5). Larger DIP values in a certain pixel mean that conditions in this pixel are less able to meet the LURs (specified by all 10 IPCs). SDIPs are more specific: for example, larger $SDIP_{Mg}$ values in a specific pixel mean nothing more than a greater difference between that pixel's soil Mg^{2+} value and the IPC for soil Mg^{2+} (7.035 $cmolc.kg^{-1}$).

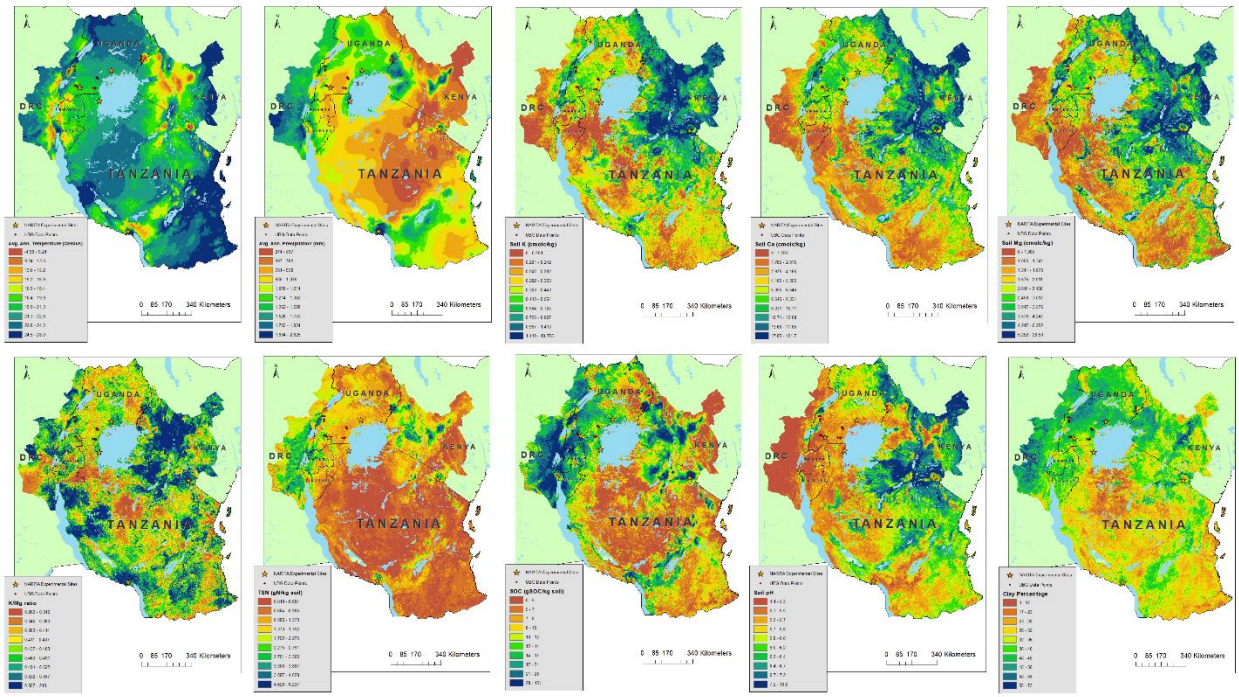

Figure 4.3: ECVs across the EAGL region. Clockwise from top left: Average annual temperature, average annual precipitation, soil K^+, soil Ca^{2+}, soil Mg^{2+}, Clay percentage, soil pH, soil organic carbon (SOC), total soil nitrogen (TSN) and the K/Mg ratio. For more detailed maps: See Annex, section 7.1.

A large area such as the EAGL region contains a multitude of factors influencing the location and variation trends of the ECVs. Some of these factors can be discerned through visually surveying the maps of the ECVs and SDIPs in Figures 4.3 and 4.4. Foremost among them are topographic features, such as Mount Elgon along the Ugandan-Kenyan border, Mt. Kilimanjaro in Tanzania and the highlands in the southwest of Kenya, or the lower elevations of the Rift valley in northern and western Uganda. Soil characteristics such as TSN, SOC, soil Ca^{2+} and K^+ all demonstrate concentrations of high values where elevations are greatest, as well as often showing steep gradients on the slopes or foothills. Proximity to water bodies appears to be related to high average annual precipitation, as does the presence of mountain

ranges, although rain shadow effects are also prominent. ECVs in some places coincide with vegetation elements, such as the presence of forest cover. This is most obvious in the case of the DRC's areas under tropical rainforest, which boast high SOC, clay and precipitation, as well as low soil pH (Figure 4.3).

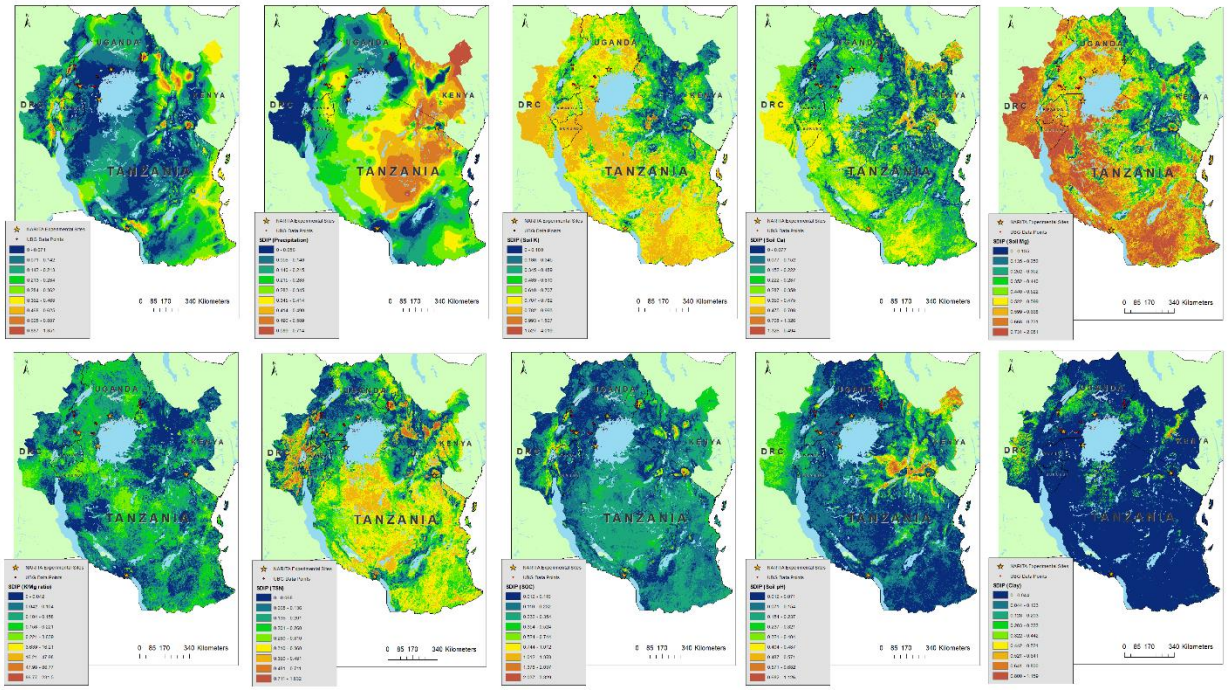

Figure 4.4: Single-variable Distances to the Ideal Point (SDIPs) across the EAGL region. Clockwise from top left: the SDIPs for Average annual temperature, average annual precipitation, soil K^+, soil Ca^{2+}, soil Mg^{2+}, Clay percentage, soil pH, soil organic carbon (SOC), total soil nitrogen (TSN) and the K/Mg ratio. For more detailed maps: See Annex, section 7.2.

Areas with the lowest DIPs (Figure 4.5) coincide with the EAHB-producing areas in Figure 2.2 and/or the high-production zones in Figure 4.1 in the areas around Mt. Elgon in eastern Uganda and western Kenya, the area of central Kenya roughly between Nairobi and Mt. Kenya, the area surrounding Mts. Kilimanjaro and Meru in northeastern Tanzania, parts of the Albertine Rift valley in western Uganda and the Rusizi river valley on the border between Burundi and the DRC. However, the majority of Uganda's banana growing areas along the Lake Victoria shore and in the southwest, along with the Rwandan northeast, are ranked only in the intermediate ranges of DIPs, while the Tanzanian Kagera and Mbeya regions and Usambara mountains and Kenya's western lakeshore areas consist of mixes of limited areas of high suitability interspersed with larger areas of moderate to low suitability (or high DIP). The majority of the Rwandan and Burundian areas of high production as well as the more western parts of the DRC's North- and South-Kivu provinces have high DIP values and thus low suitability.

The **Luweero** site and the surrounding area in Central Uganda are characterized by an extremely low SDIP for (and thus high suitability regarding) temperature, relatively low SDIPs for SOC, soil pH, TSN

and K/Mg ratio, intermediate values for soil Ca^{2+} and precipitation, and high SDIPs for soil K^{+} and Mg^{2+}. The area around the **Mbarara** site in southwestern Uganda is quite similar, with higher SDIPs for TSN and precipitation and a lower SDIP for Clay percentage the only notable differences. Furthermore, the site in the Tanzanian **Kagera** region and surrounding areas on Lake Victoria's western shore have intermediate to high suitability regarding all ECVs except for soil K^{+} and Mg^{2+}. The test site in the **Kilimanjaro** region (northeast Tanzania) is located in an area of strong gradients centered on Mt. Kilimanjaro. Gradients are also not always monotonous: along a transect from the test site to the summit, the SDIPs for Ca^{2+}, precipitation, TSN, SOC and soil pH reach at least one local minimum or maximum before rising or falling again. $SDIP_{Mg}$ and $SDIP_{K}$ decrease monotonously from test site to summit, while $SDIP_{Temp}$ increases. Only clay percentage and K/Mg ratio do not vary according to these concentric gradients, being roughly uniformly suitable in a wide region around Mt. Kilimanjaro. Finally, the experimental site in the **Mbeya** region, located in southern Tanzania has low SDIP values (high single-variable suitability) regarding precipitation, temperature and all soil variables except for soil K^{+} and Ca^{2+} (intermediate) and soil Mg^{2+} (high).

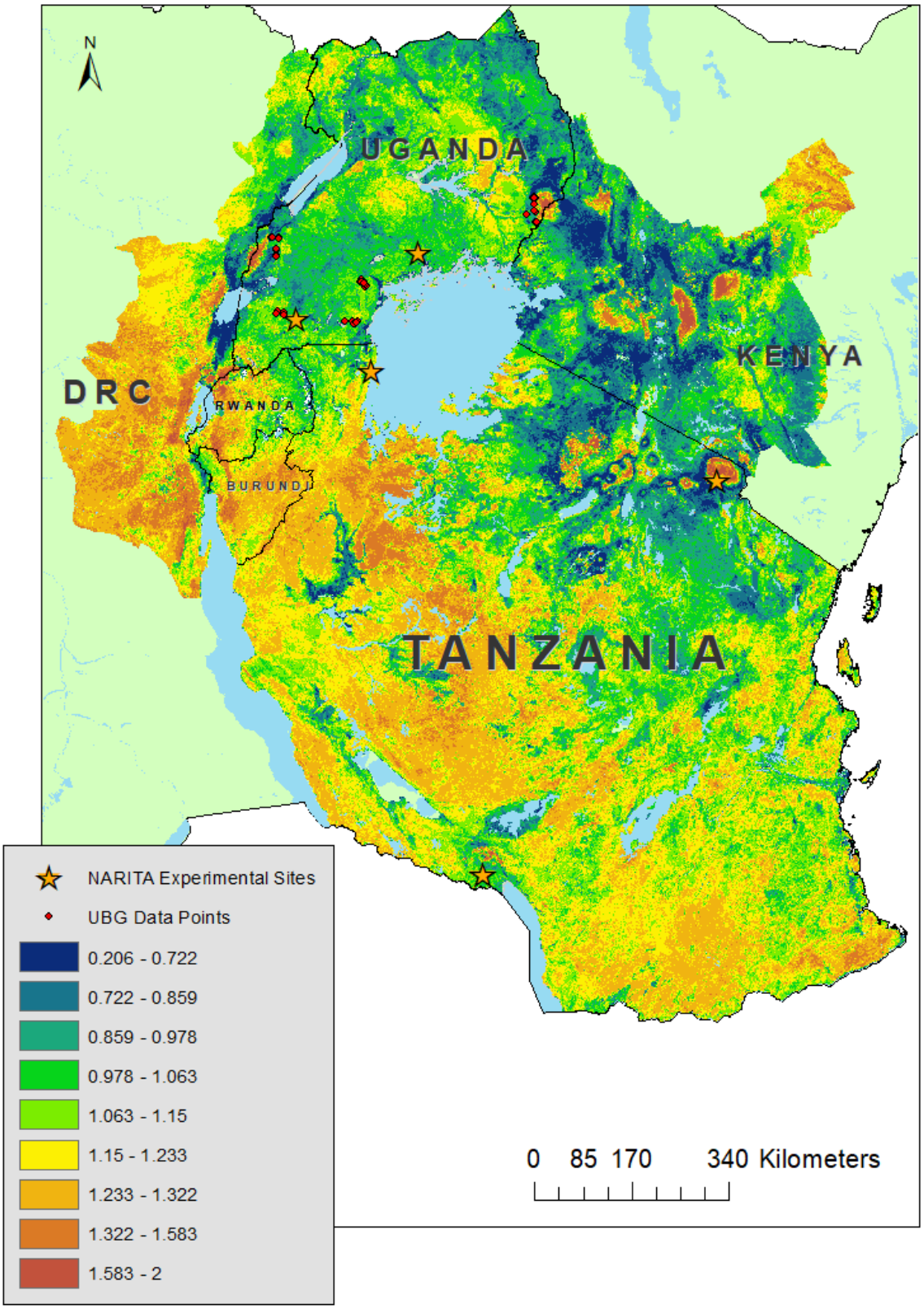

Figure 4.5: Distance to the Ideal Point (DIP) values across the EAGL geographic level.

4.3 Most Limiting Variables for EAHB Production

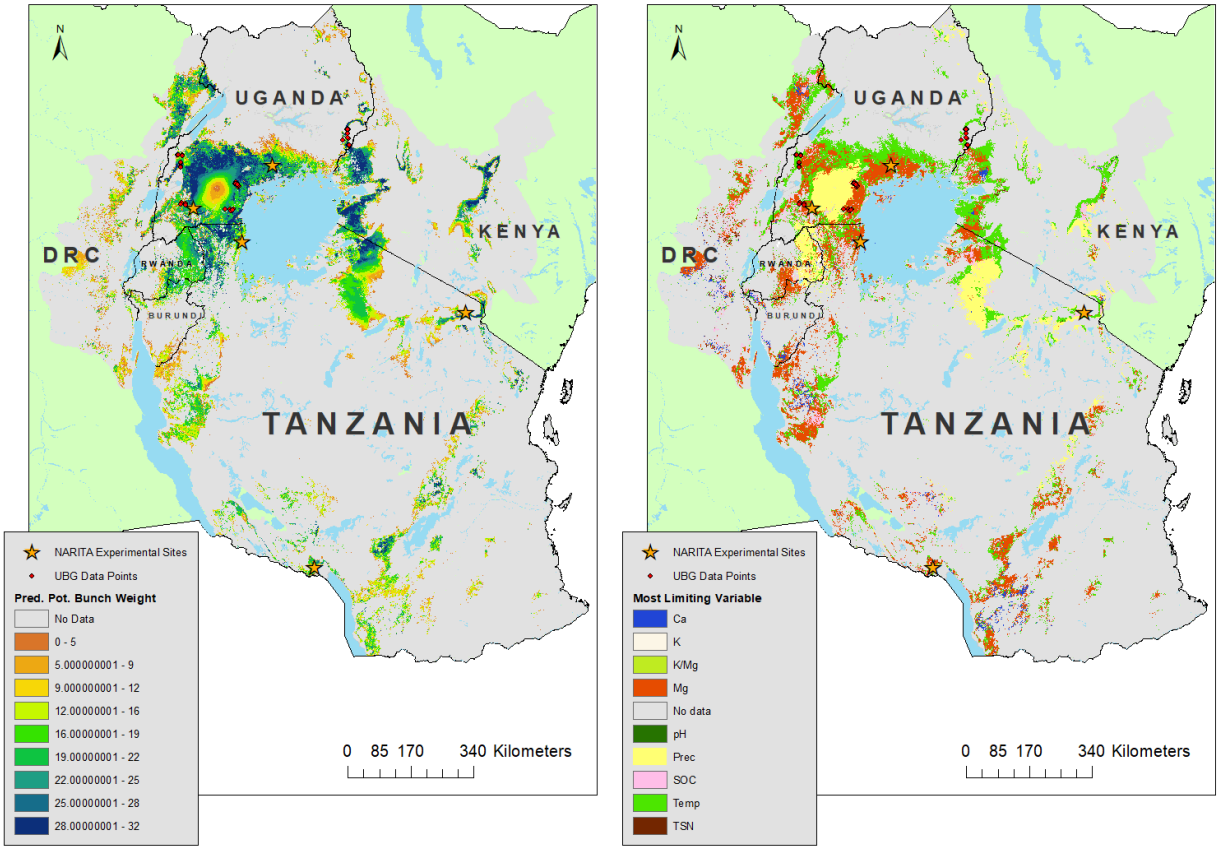

Figure 4.6: Predicted potential bunch weight data (in kg, left map) and most limiting variables across the EAGL geographic level (right map). For more detailed maps: See Annex, section 7.3.

In each data point where the Law of the Minimum method was applied to calculate BW_{pred}, the most limiting variable was also computed. While the abundance of 'No data' values for these two variables across the EAGL geographic level means it was difficult to determine whether trends are outliers or not, notable phenomena can still be identified, such as the low-BW_{pred}, precipitation-limited zone in southwestern Uganda along with higher-BW_{pred}, soil Mg^{2+}- and temperature-limited zones scattered across the region. The extent of areas in which there is data for these variables is in itself interesting: pixels containing at least one ECV value falling outside that ECV's range in the AgTrials dataset were labeled 'No data', meaning those pixels that *do* have valid BW_{pred} and most limiting variable values are similar to the AgTrials dataset as far as the 10 ECVs are concerned (see Sections 3.2.2 and 3.3.2). Interestingly, the extent of these data-containing pixels in many areas roughly mirrors the EAHB-producing areas in Figure 2.2. All five NARITA experimental sites also fall within areas where data-containing pixels are concentrated.

As far as experimental sites are concerned, the Luweero, Mbarara, Kagera and Mbeya sites and their surrounding areas are most limited by soil Mg^{2+}, which coincides with their high $SDIP_{Mg}$ values. The Ugandan Lake Victoria shore appears to be most limited by soil Mg^{2+}, while about 50 km further inland, this is replaced by temperature. The experimental site in the Kilimanjaro region lies in an area of high variability of ECVs and SDIPs, and the same can be said of most limiting variables: there are roughly concentric bands of temperature-, precipitation-, soil Mg^{2+}- and SOC-limited areas around the summit of Mt. Kilimanjaro.

4.4 Sub-environments of the Target Population of Environments

Median values of all SDIPs except that of clay percentage varied widely across the 14 clusters into which the EAGL region was divided (Figure 4.7). A map of these clusters is displayed in Figure 4.9. Each cluster had its own unique set of median SDIP values, for example cluster 6, which appeared mostly at very high altitudes (Figure 4.9) as found near the summits of Mts. Kilimanjaro, Elgon and Kenya, and had generally very high SDIP values (low suitability) for temperature, TSN and SOC, along with relatively very suitable values for K/Mg ratio, pH and precipitation. Looking at the boxplot of the DIP values per cluster (Figure 4.8), it could be seen that these characteristics translated into a cluster that, while very variable, in general contained some of the least suitable (highest DIP) environments for EAHB production when taking into account the 10 ECVs only. Cluster 7, which had the lowest median DIP, was situated mostly in the Kenyan highlands and the adjacent region in northeastern Tanzania. Notably, clusters were often arrayed in concentric circles around high-altitude areas occupied by cluster 6, with clusters 13, 12 and 7 frequently appearing in that same order, as altitude decreased. Cluster 7 had generally below average (compared to the other clusters' median SDIPs) median SDIPs for K^+, Mg^{2+}, Ca^{2+}, K/Mg ratio, SOC and TSN, but an average median $SDIP_{Temp}$ and an above average median $SDIP_{Prec}$. The principal cooking and beer banana producing areas in Uganda and Rwanda according to Figure 4.1, as well as the western Kenyan rainfed banana growing areas along Lake Victoria's western shore (Figure 4.2) belong to cluster 4, which has, compared to the other clusters, relatively low medians (and thus is relatively suitable) for all SDIPs except those for soil K^+, Mg^{2+} and the K/Mg ratio (Figure 4.7). Southwestern Rwanda and large parts of the Burundian interior (excluding the heights of the Nile-Congo watershed divide, which run in a rough V-shape parallel to the country's western and southeastern borders), which are also important beer and cooking banana producing regions (Figure 4.1), belong to cluster 5. Large parts of the rainfed banana-producing areas in the eastern DRC (Figure 4.2) also belong to cluster 5. This cluster is characterized by relatively low SDIPs for temperature, TSN, SOC and precipitation and relatively high SDIPs for soil K^+, Mg^{2+} and Ca^{2+}. Its median SDIPs for soil K^+, Mg^{2+} and Ca^{2+} and temperature are slightly larger than cluster 4's, while its median SDIP for temperature is slightly smaller. This translates into cluster 5 having a slightly higher DIP or being slightly less suitable for EAHB

production than cluster 4 (Figure 4.8). The banana-producing areas in central Kenya (Figure 4.2) as well as the banana growing areas around Mt. Kilimanjaro are primarily located in clusters 12 and 13. Cluster 13 has relatively high SDIP values for soil K^+ and Mg^{2+}, SOC, TSN and temperature, but a relatively low SDIP for precipitation, somewhat resembling cluster 6 (Figure 4.7). It also has the second-highest median DIP (Figure 4.8). Cluster 12 has relatively low SDIPs for pH, SOC, TSN and soil Ca^{2+}, along with average values for temperature, precipitation, soil K^+ and Mg^{2+}, resulting in the second-lowest median DIP of all 14 clusters. Banana production in these central Kenyan and northeastern Tanzanian highland areas thus appears to be grown in both very suitable and very unsuitable environments. Notably, banana production in cluster 7 is quite low, despite this cluster's high suitability. According to the HarvestChoice map in Figure 4.2, banana production is also generally absent in the unsuitable 'mountaintop' environments of cluster 6, except perhaps in a small area in the Mbeya region in Tanzania (Figure 4.9).

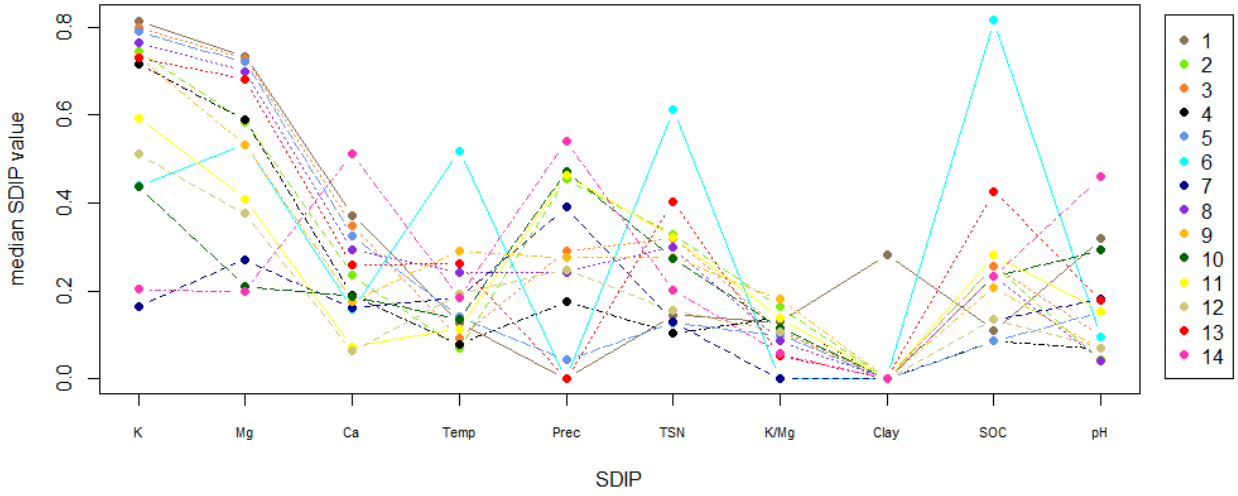

Figure 4.7: **Median SDIP values per cluster for the SDIP clustering solution.**

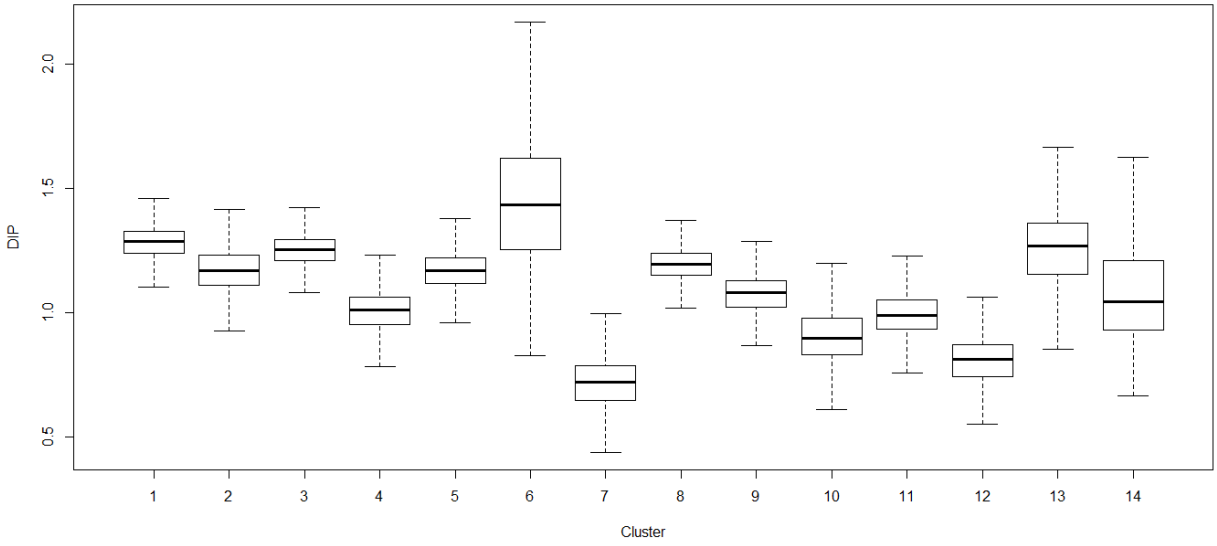

Figure 4.8: Distributions of DIP values in each of the SDIP clustering solution's 14 clusters.

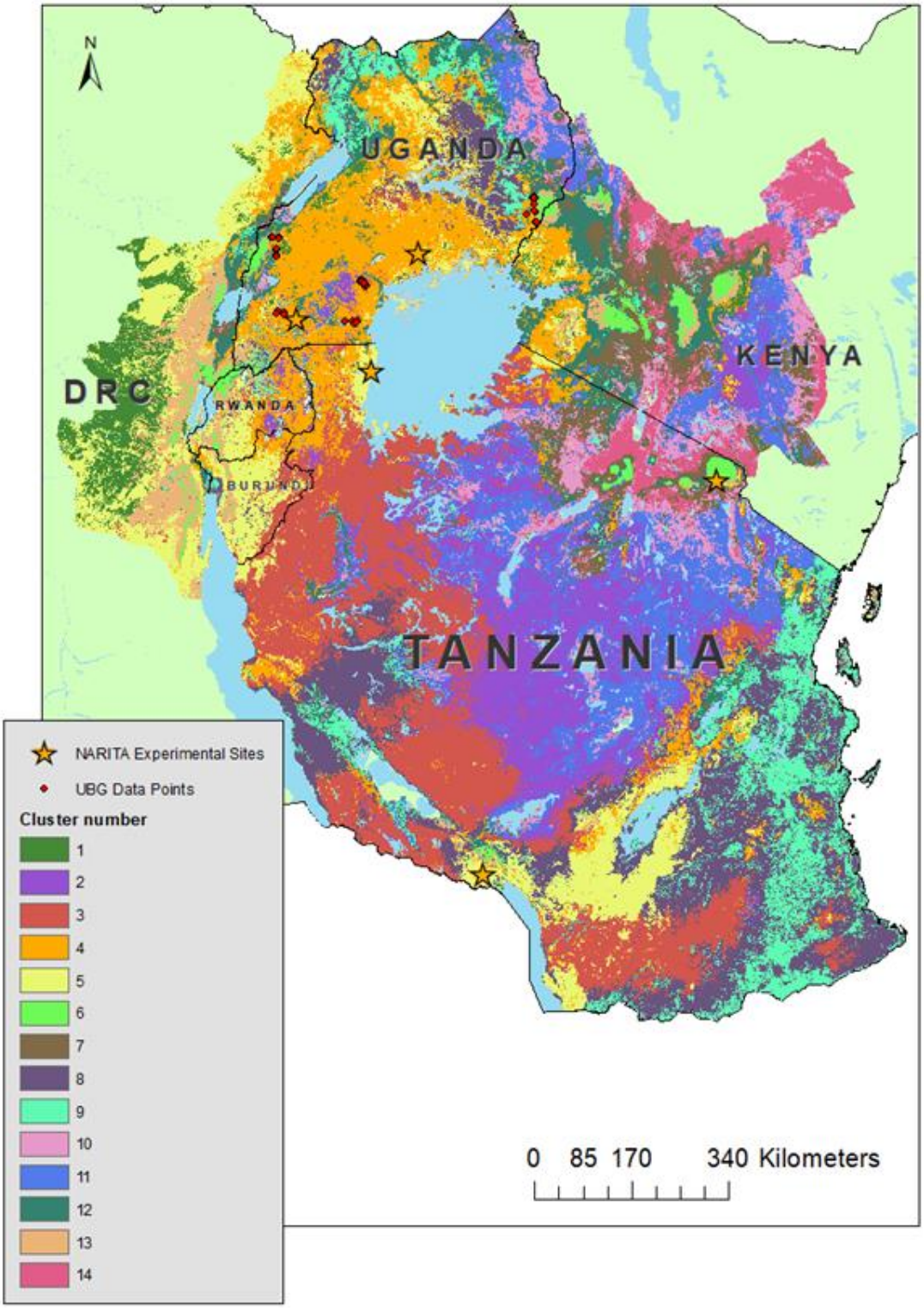

Figure 4.9: Spatial distribution of the 14 different clusters across the EAGL geographic level.

4.5 Performance Gaps and Sub-environments in the UBG Geographic Level

Of the UBG dataset's 64 data points, 42.2 % belonged to cluster 12 and 54.8 % to cluster 4. BW_{est} was significantly ($p < 0.05$) higher than BW_{pred} in cluster 12, whereas BW_{est} was lower than BW_{pred} in cluster 4. Cluster 4 did have a significantly larger mean BWG than cluster 12; the average sample BWGs for the two clusters were 2.2 kg and -5.89 kg. This could be interpreted as cluster 4 underperforming and cluster 12 outperforming expectations (which are defined by BW_{pred} and thus the 10 ECVs and their ideal point coordinate values). The data points belonging to cluster 4 in the UBG dataset have significantly higher SDIPs for soil Mg^{2+}, SOC, K/Mg ratio and precipitation than those belonging to cluster 12, but a lower SDIP for temperature (Figure 4.10).

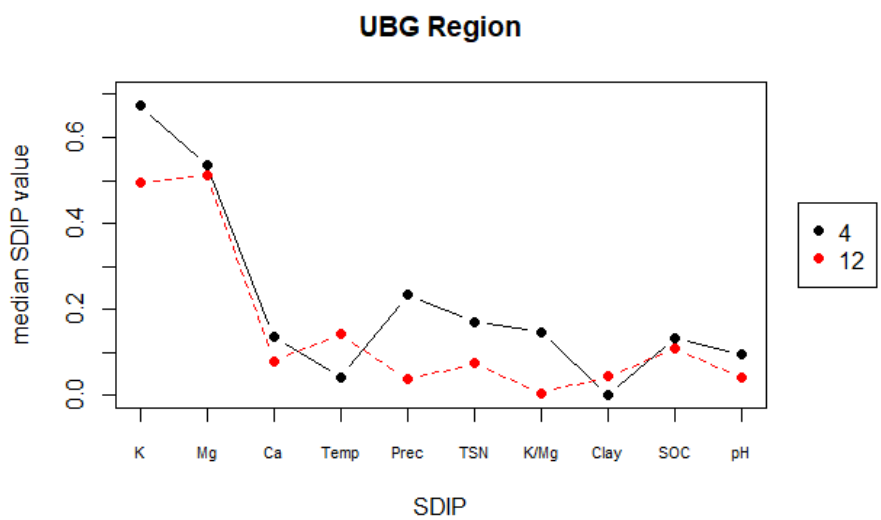

Figure 4.10: Median SDIP values for all ten SDIP types in clusters 4 and 12 in the UBG dataset. Are these lines making any sense?

In the UBG dataset, the only limiting variables present are temperature, precipitation, SOC, K/Mg ratio and soil Mg^{2+}. In cluster 12, temperature dominates as limiting factor, accounting for 96.3% of points (Figure 4.11). This is in accordance with this cluster's slightly higher SDIP for temperature. Cluster 4 contains a more varied array of most limiting variables. Across the entire UBG dataset, average annual temperature is most frequently the most limiting factor (51.6% of data points), followed by soil Mg^{2+} (21.0%), K/Mg ratio (16.1%), SOC (8.1%) and average annual precipitation (3.2%). Based on these results, it can be said that cluster 12 is mainly temperature-limited, while cluster 4's limitations stem primarily from soil variables.

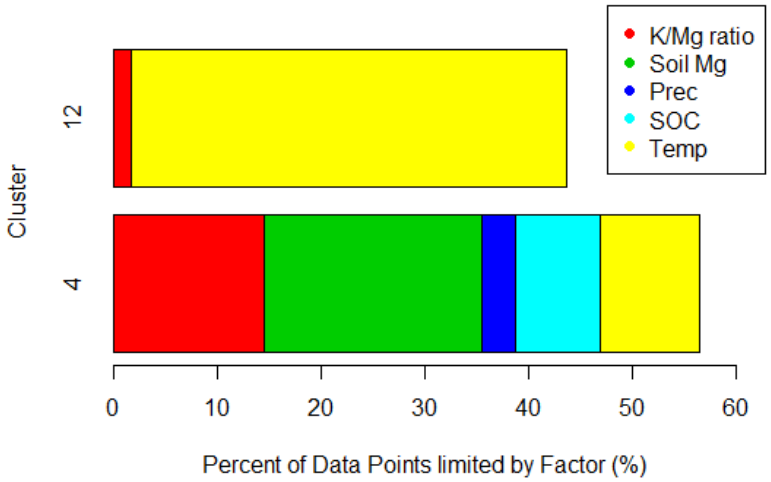

Figure 4.11: Limiting factors present in the two major clusters in the UBG dataset.

Most UBG data points belonging to cluster 12 are situated in the higher-altitude zones around Mt. Elgon and in Kabarole district. Management recommendations for data points in cluster 4 involve soil fertility improvement, through techniques such as manuring and mulching in the absence of any inorganic fertilizer. Mulch serves to increase SOC content of the soil, introduces nutrients and retains moisture, making it an interesting solution not only in those data points where soil variables are most limiting, but also in those points where precipitation is limiting. Mulching comes with the caveat that it might contribute to increased weevil activity, and must therefore be coupled with anti-weevil management techniques, such as trapping (McIntyre et al., 2003). Mulching and manuring is often prohibitively expensive and depends on the local availability of mulch or manure. Furthermore, Wairegi et al. (2010) mentions that a variable could be 'most limiting' in a certain area due to its own far from optimal values, or due to the elevated optimality of values for all other variables. Tackling one problem could very well lead to another, for example the removal of mulch to deny weevils a favorable environment could lead to a resurgence of weeds. Therefore, it is important to look at solutions for a whole array of limiting variables, not only that which is *most* limiting.

While many papers mention decreasing *inherent* soil fertility as a cause of yield declines across the region, the majority of EAHB, except for the extremely fertile volcanic soils found in parts of Rwanda, Burundi, Tanzania and the DRC, is cultivated on inherently nutrient-poor ferralsol- and acrisol-type soils. These soils depend largely on mineralization of soil organic matter for their nutrients, suggesting that management practices such as mulching or manuring could be greater determinants of variability of soil fertility and yield than inherent soil fertility (Van Asten et al., 2006). Pest and disease incidence and severity is often affected by a complex interplay of climatic, soil, socio-economic and management-

related variables. Higher-altitude data points such as those in cluster 12, for example, are less likely to experience the same intensity of a disease like black leaf streak than data points in the lowlands (Erima et al., 2017; Van Asten et al., 2004). In another example, the *Xanthomonas campestris* bacterium generally survives longer in moist soil, yet will generally not survive longer than 35 days due primarily to predation and competition effects of other micro-organisms and nematodes, whose incidence is itself positively affected by high soil moisture contents (Mwebaze et al., 2006). Higher levels of soil K^+, Ca^{2+} and nitrogen reduced disease incidence and lengthened the bacterium's dormant period in controlled in-vitro experiments, while increasing the levels of these nutrients in plant tissues (Atim et al, 2013).

One difference that the clusters fail to accurately portray is the much higher performance of EAHB in southwestern Uganda than in central Uganda (Gold et al., 1999; Karamura et al., 1998). Both of these areas are dominated by cluster 4, except for a limited area of southwestern Uganda where cluster 2 (a cluster with a *higher* median DIP and less suitability than cluster 4) makes an appearance. The difference between these areas is therefore thought to depend more on socio-economic variables, such as a more motivated labor force, a more market-oriented production system and a greater degree of investment in expensive pest and disease control methods (Blomme et al., 2017; Gold et al., 1999; Jogo et al., 2013). More market-oriented farmers can use their funds to purchase additional labor and the necessary materials for implementation of disease control frameworks. They are thus more likely to successfully adopt these frameworks and less likely to suffer large losses caused by disease. In the central region as opposed to the southwest, banana cultivation is carried out primarily by women, which also has a negative impact on adoption of better management practices, as information disseminated via agricultural extension and farmer field schools often reaches only men (Jogo et al., 2013). Farmers in the higher-altitude areas in the southwest have also been shown to be more likely to carry out agro-ecological intensification measures such as mulching, trapping insects, fallowing and introducing pests' natural predators (E. B. Karamura et al., 2013).

An issue affecting both DIPs/SDIPs and BW_{pred} is the fact that the bunch weight and ECV data from the AgTrials dataset, which was used to set up functional relationships and, by extension, the IPCs, consists of a mix of 'pure' on-farm data and demonstration plot data collected exclusively in the southern half of Uganda (Wairegi et al., 2010). The ideal functional relationship is one which portrays BW as limited *only* by the ECV in question. Non-ideal management by farmers may have adversely affected BWs of a majority of mats, resulting in depressed functional relationships. It may also have directly affected certain soil variables, leading to data points being distributed across different levels of these soil variables than they might be under conditions of ideal management. Areas with ECV conditions different to those in southern Uganda might have the same result. This can cause eventual functional relationships to be

completely different, as evidenced by the clay percentage relationships in two other studies employing functional relationships (Alou et al., 2014; Wairegi et al., 2010). The choice of fitting function is also important: inverse parabolic relationships in this project contrast starkly with the asymptotic ones in other works (Alou et al., 2014; Wairegi et al., 2010). It remains to be seen whether different base data and fitting functions would lead to entirely different results. However, when comparing IP values in Table 3.6 to literature-derived ideal and critical values and ranges in Table 2.2, there is rough similarity for the ideal values of soil pH, SOC, TSN, temperature and soil K^+, that the functional relationship-derived ideal range for clay percentage falls just below the critical point listed in Table 2.2 and that the calculated ideal range for precipitation is almost exactly the same as the literature-derived range. Only the computed IP values for Soil Mg^{2+} and soil Ca^{2+} are the only ECVs for which the computed IP values are very different from literature values. Exaggeratedly high IP values for these two variables could have resulted in elevated SDIPs and DIPs for pixels which would normally have fallen close to the IP. However, it remains possible that these calculated IP values represent EAHB's true preference regarding levels of soil Mg^{2+} and soil Ca^{2+} in the region.

4.6 Situation of Methods in a Land Evaluation and TPE framework

As mentioned in Section 2.2.3, there are similarities between Land Evaluation and TPE methodologies. The sub-environments set up through cluster analysis are simply another form of LMUs, and both methodologies are known to use ECVs to both subdivide and evaluate a study area. In case of the TPE methodology, this refers to the pedo-climatic method of setting up sub-environments (Chenu, 2015, see Section 2.2.3).

This project and the CP and Law of the Minimum methods employed in it correspond more to the definition of biophysical land evaluation than that of economic land evaluation. To compute DIP and BW_{pred} values, only soil and climate data (the ECVs) were used, with socio-economic variables only being considered in a later, somewhat cursory step. This mirrors Land Evaluation's two-step approach (Section 2.2.1). There is an absence of any type of economic metrics (such as internal rate of return or benefit/cost ratio), and economic variables play no role in the subdivision of the study region. This does not mean that integrating certain aspects of economic land evaluation wouldn't be a good idea. Sensitivity analyses and increased communication with farmers, phenomena often linked to economic rather than biophysical land evaluation (Jakeman et al., 2003; Samranpong et al., 2009), could both have been used to set up weights for use in the CP method. Feedback from the demand side (farmers, see Section 2.2.1) in carrying out land evaluation is important (Bacic et al., 2003), which is exactly where the farmer-managed trials in the NARITA breeding process come in.

Determining whether this project's overall methodology belongs mainly to a quantitative or qualitative land evaluation framework is less clear-cut. While IPCs and other parameters used in the CP and Law of the Minimum methods are determined from data, and not from expert knowledge, implying a quantitative land evaluation framework (De La Rosa et al., 2002, section 2.2.2). This position is reinforced by the fact that both methods' outputs are continuous figures, not discrete categories (De La Rosa et al., 2002). However, the CP method produces a dimensionless index not directly relatable to EAHB productivity, which according to van Lanen (1991) means it is a qualitative land evaluation method. Furthermore, both methods' decision structures are quite transparent, with BW_{pred} able to be matched with the most limiting variable in the same location, and DIP able to be taken apart into its constituent SDIPs. To conclude, it can be said that methods used in this project retain elements of both quantitative and qualitative land evaluation.

5 Conclusion

This project had two main subjects: carrying out a MCDM embedded in a Land Evaluation framework, and situating the methods employed in this framework (i), in order to identify and describe different sub-environments in the EAGL region regarding their suitability for EAHB production by smallholder farmers, and their characteristic ECV values (ii).

The CP and Law of the minimum methods used only edapho-climatic data to generate suitability and performance predictions, respectively, and didn't take into account or generate any economic variables. Therefore, while economic land evaluation would have been an interesting addition to this project, due to its translatability to terms accessible to farmers and other 'demand side' stakeholders, as well as its capacity to take into account temporal variability and uncertain situations, this project was limited to biophysical land evaluation. The picture is less clear when it comes to qualitative and quantitative land evaluation frameworks, with the CP and Law of the minimum methods each containing elements of either type of framework. This is, however, quite common in modern land evaluation, which strives to find balance between overly complex quantitative methodologies dependent on computation-intensive biophysical models and qualitative methodologies that fail to provide a sufficient degree of detail or flexibility under highly variable conditions.

This project's second main subject was subdivided into two geographic levels. In the EAGL geographic level, spanning the entire EAGL region, extensive and high-resolution datasets for ten soil and climate variables selected for their known limiting influence on EAHB performance were computed into performance and suitability values using the two methods mentioned above. Only the suitability values, expressed as DIPs, could be calculated for the entire geographic level. Areas of high EAHB or banana production volume often coincided with areas of high suitability, with notable exceptions being the banana producing regions of central and southwestern Uganda, northeastern Rwanda, Kenya's western lakeshore and Tanzania's Mbeya, Kagera and Usambara mountains regions. It can thus be said that DIP values based on ECVs only either do not give an entirely accurate picture of suitability for EAHB production, or that cultivation is happening in areas despite those areas' low suitability. NARITA experimental sites are generally found in areas of intermediate to high suitability.

The performance predictions, represented as BW_{pred} values, were only calculable in those areas bearing similarity to the locations of the source data for the functional relationships used to calculate the BW_{pred} values. A map of most limiting variables for the region identified average annual precipitation, average annual temperature and soil Mg^{2+} as the most important limiting variables in the region for which BW_{pred} was calculated. Low BW_{pred} values appeared to coincide with precipitation-limited areas.

Spatial subunits computed using a clustering algorithm, represented fourteen different sub-environments, each characterized by a unique combination of median SDIP values. Sub-environments, while corresponding to large-scale landscape and vegetation features such as lakes, mountain ranges and rainforests, did not always adequately portray the diversity of banana-producing systems in the EAGL region. In some cases, banana production was equally high in adjacent sub-environments with drastically different suitabilities and ECVs (central Kenya, northeastern Tanzanian highlands), whereas in other cases, distinct production systems with completely different productivities were located in the same sub-environment (central and southwestern Uganda).

In the UBG geographic level, most data points belonged to either a lower-altitude sub-environment in central and southwestern Uganda, where EAHB production was limited mainly by SOC, soil Mg^{2+} and the K/Mg ratio, or a higher-altitude sub-environment in the Mt. Elgon and Kabarole areas where production was mainly limited by temperature. BWGs and DIPs in the higher-altitude sub-environment were significantly lower than those in the lower-altitude sub-environment.

One of this project's main flaws was the focus on ECVs. Suitability and performance predictions were often inconsistent with actual performance and production data, and the important role ascribed in literature to socio-economic, pest and disease and management factors in determining the type and productivity of EAHB production system only confirms the fact that soil and climate variables only give part of the total picture. Suggestions for amelioration of this project's methods include obtaining more data on the missing (explanatory) variables, expansion of the CP and Law of the Minimum methods with these additional variables, determination of weights for use in CP and the setting up of functional relationships using base data from other EAHB-producing areas in the EAGL region.

6 References

Alou, I. N., Van Asten, P. J. A., & Tenywa, M. M. (2014). Biophysical and Crop Management Gradients Limiting Yields of East African Highland Banana (Musa spp. AAA-EA) within farms in low input Cropping Systems, *4*(3), 27–44.

Atim, M., Beed, F., Tusiime, G., Tripathi, L., & van Asten, P. (2013). High Potassium, Calcium, and Nitrogen Application Reduce Susceptibility to Banana Xanthomonas Wilt Caused by *Xanthomonas campestris* pv. *musacearum*. *Plant Disease*, *97*(1), 123–130. https://doi.org/10.1094/PDIS-07-12-0646-RE

Bacic, I. L. Z., Rossiter, D. G., & Bregt, A. K. (2003). planners and decision-makers : a case study in Santa, *1992*, 12–18. https://doi.org/10.1079/SUM2002154

Bagamba, F. (2007). *Market access and agricultural production: the case of banana production in Uganda*. https://doi.org/10.1080/13545701.2013.855320

Beynon, M. J., Heffernan, M., & Mcdermott, A. M. (2012). Psychological Contracts and Job Satisfaction: Clustering Analysis Using Evidential c-means and Comparison with other Techniques. *Intelligent Systems in Accounting, Finance and Management.*, *19*, 247–273. https://doi.org/10.1002/isaf

Bioversity. (2016). No Title. Retrieved from http://www.bioversityinternational.org/

Bioversity International. (2013). *Improvement of Banana for Smallholder Farmers: Work Package 4 – Regional Testing*.

Blomme, G., Jacobsen, K., Ocimati, W., Beed, F., Ntamwira, J., Sivirihauma, C., … Karamura, E. (2014). Fine-tuning banana Xanthomonas wilt control options over the past decade in East and Central Africa. *European Journal of Plant Pathology*, (March). https://doi.org/10.1007/s10658-014-0402-0

Blomme, G., Ocimati, W., Sivirihauma, C., Vutseme, L., Mariamu, B., Kamira, M., … Ntamwira, J. (2017). A control package revolving around the removal of single diseased banana stems is effective for the restoration of Xanthomonas wilt infected fields. *European Journal of Plant Pathology*, *149*(2), 385–400. https://doi.org/10.1007/s10658-017-1189-6

Blomme, G., Ploetz, R., Jones, D., De Langhe, E., Price, N., Gold, C., … Buddenhagen, I. (2013). A historical overview of the appearance and spread of Musa pests and pathogens on the African continent: Highlighting the importance of clean Musa planting materials and quarantine measures. *Annals of Applied Biology*, *162*(1), 4–26. https://doi.org/10.1111/aab.12002

Bonfante, A., & Bouma, J. (2015). The role of soil series in quantitative land evaluation when expressing effects of climate change and crop breeding on future land use. *Geoderma*, *259–260*, 187–195. https://doi.org/10.1016/j.geoderma.2016.05.018

Bouma, J., Stoorvogel, J. J., & Sonneveld, M. P. W. (2011). Land Evaluation for Landscape Units. In *Handbook of Soil Sciences* (pp. 1–22). CRC Press. https://doi.org/doi:10.1201/b11267-40

Bouraima, A. K., Weihua, Z., & Chaofu, W. (2015). Irrigation water requirements of rice using Cropwat model in Northern Benin. *International Journal of Agricultural and Biological Engineering*, *8*(2), 58–64. https://doi.org/10.3965/j.ijabe.20150802.1290

Braun, H.-J., Rajaram, S., & Ginkel, M. (1996). CIMMYT's approach to breeding for wide adaptation. *Euphytica*, *92*(1–2), 175–183. https://doi.org/10.1007/BF00022843

Byabachwezi, M., Mkulila, S., & Karamura, D. (2005). *Tanzania Musa Expedition 2005*.

Calberto, G., & Staver, C. (2015). An assessment of global banana production and suitability under climate change scenarios.

CGIAR Research Program on Climate Change and Agricultural Food Security (CCAFS). (2013). AgTrials: Global Agricultural Trial Repository and Database. Retrieved from http://www.agtrials.org/

Chenu, K. (2015). Characterizing the Crop Environment - nature, significance and applications. In V. Sadras & D. Calderini (Eds.), *Crop Physiology. Applications for Genetic Improvement and Agronomy* (pp. 321–348). Academic Press. https://doi.org/10.1016/B978-0-12-417104-6/00021-2

Chenu, K., Chapman, S. C., Tardieu, F., McLean, G., Welcker, C., & Hammer, G. L. (2009). Simulating the yield impacts of organ-level quantitative trait loci associated with drought response in maize: A "gene-to-phenotype" modeling approach. *Genetics*, *183*(4), 1507–1523. https://doi.org/10.1534/genetics.109.105429

Cooper, M., & Byth, D. E. (1996). Understanding plant adaptation to achieve systematic applied crop improvement–a fundamental challenge. In *Plant adaptation and crop improvement* (pp. 5–23). CABI Wallingford, UK.

Davies, G. (1995). Banana and plantain in the East African highlands. In S. Gowen (Ed.), *Bananas and Plantains*. Springer, Dordrecht. https://doi.org/https://doi.org/10.1007/978-94-011-0737-2_18

De Bauw, P., Van Asten, P., Jassogne, L., & Merckx, R. (2016). Agriculture , Ecosystems and Environment Soil fertility gradients and production constraints for coffee and banana on volcanic mountain slopes in the East African Rift : A case study of Mt . Elgon, *231*, 166–175.

de Carvalho, A., Coelho Barros, V., Fragata, A., & da Rosa Pires, A. (1989). Land Resources, Agricultural Change and Economic Development: Methodological Guidelines for a Proposed Research Project. In *Methods and socio-economic criteria for the analysis and prevision of land use and land evaluation* (pp. 189–196).

De Geus, J. G. (1967). *Fertilizer Guide for Tropical and Subtropical Farming*. Centre d'Étude de l'Azote, Zurich.

De La Rosa, D., Mayol, F., Diaz-Pereira, E., Fernandez, M., & De La Rosa, D. J. (2004). A land evaluation decision support system (MicroLEIS DSS) for agricultural soil protection With special reference to the Mediterranean region, *19*, 929–942. https://doi.org/10.1016/j.envsoft.2003.10.006

De La Rosa, D., & van Diepen, C. A. (2002). Qualitative and Quantitative Land Evaluations. In *Encyclopedia of Life SUpport Systems (EOLSS-UNESCO)* (pp. 3–12). Eolss Publishers, Oxford, UK.

de Langhe, E., Karamura, D., & Mbwana, A. (2001). *Tanzania Musa Expedition 2001*.

Dowiya, N. B., Rweyemamu, C. L., & Maerere, A. P. (2009). Banana (Musa spp. Colla) cropping systems, production constraints and cultivar preferences in eastern Democratic Republic of Congo. *Journal of Animal and Plant Sciences (JAPS)*, *4*(2), 341–356.

Eledu, C. A., Karamura, E. B., & Tushemereirwe, W. K. (2004). Agroecological distribution of banana systems in the Great Lakes Region. *African Crop Science Journal*, *12*(1), 33–42.

Erima, R., Kubiriba, J., Komutunga, E., Nowakunda, K., Namanya, P., Seruga, R., … Tushemereirwe, W. K. (2017). Banana pests and diseases spread to higher altitudes due to increasing temperature over the last 20 years. *African Journal of Environmental Science and Technology*, *11*(12), 601–608. https://doi.org/http://dx.doi.org/10.5897/AJEST2016.2173

Estrella, R., Cattrysse, D., & Van Orshoven, J. (2014). Comparison of Three Ideal Point-Based Multi-Criteria Decision Methods for Afforestation Planning, 3222–3240. https://doi.org/10.3390/f5123222

Fao. (1976). *A framework for land evaluation. FAO soils bulletin n.32*. https://doi.org/M-51

FAO. (2007). *Land Evaluation: Towards a Revised Framework*. *FAO Land and Water Discussion Papers*.

FAO. (2013). FAO Statistics Division. Retrieved April 15, 2018, from http://faostat3.fao.org/

Fischer, G., Nachtergaele, F. O., Prieler, S., Teixeira, E., Toth, G., van Velthuizen, H., … Wiberg, D. (2012). Global Agro-ecological Zones (GAEZ): Model Documentation, 1–179.

Fischer, G., Velthuizen, H. Van, Shah, M., & Nachtergaele, F. (2002). *Global Agro-ecological Assessment for Agriculture in the 21st Century : Methodology and Results*. *Analysis* (Vol. RR-02-02). https://doi.org/http://www.iiasa.ac.at/Admin/PUB/Documents/RR-02-002.pdf

Fouré, E. (1993). *Etude de la dynamique saisonnière de la cercosporiose noire sur plantains et de ses répercussions sur les possibilités de lutte raisonnée en milieu paysan.* Douala (CMR): CRBP.

Fresco, L., Huizing, H., van Keulen, H., Luning, H., & Schipper, R. (1990). *Land evaluation and farming systems analysis for land use planning*. Rome: Rome : FAO,.

Gold, C. S., Karamura, E. B., Kiggundu, A., Bagamba, F., Abera, A. M. K., Gold, C. S., … Bagamba, F. (1999). Geographic shifts in the highland cooking banana (Musa spp ., group AAA-EA) production in Uganda. *International Journal of Sustainable Development and World Ecology*, *6*(1), 45–59. https://doi.org/10.1080/13504509.1999.9728471

Greenland-Smith, S., Brazner, J., & Sherren, K. (2016). Farmer perceptions of wetlands and waterbodies: Using social metrics as an alternative to ecosystem service valuation. *Ecological Economics*, *126*, 58–69. https://doi.org/10.1016/j.ecolecon.2016.04.002

Guerra, L., Robles, V., Bielsa, C., & Larrañaga, P. (2012). A comparison of clustering quality indices using outliers and noise. *Intelligent Data Analysis*, *16*, 703–715. https://doi.org/10.3233/IDA-2012-0545

HarvestChoice. (2015). Banana Harvested Production (t, 2005). Retrieved from http://harvestchoice.org/data/bana_p

Heinemann, A. B., Barrios-perez, C., Ramirez-villegas, J., Arango-londoño, D., Bonilla-findji, O., Medeiros, J. C., & Jarvis, A. (2015). Variation and impact of drought-stress patterns across upland rice target population of environments in Brazil. *Journal of Experimental Botany*, *66*(12), 3625–3638. https://doi.org/10.1093/jxb/erv126

Heinemann, A. B., Ramirez-villegas, J., Lívio, T., Souza, P. O., Dirceu, A., Geraldo, J., … Jarvis, A. (2016). Agricultural and Forest Meteorology Drought impact on rainfed common bean production areas in Brazil. *Agricultural and Forest Metereology*, *225*, 57–74. https://doi.org/10.1016/j.agrformet.2016.05.010

Heinemann, A. B., & Sentelhas, P. C. (2011). Environmental group identification for upland rice production in central Brazil. *Scientia Agrícola*, *68 (5)*(October), 540–547.

Hengl, T., Heuvelink, G. B. M., Kempen, B., Leenaars, J. G. B., Tamene, L., & Tondoh, J. E. (2015). Mapping Soil Properties of Africa at 250 m Resolution : Random Forests Significantly Improve Current Predictions, 1–26. https://doi.org/10.1371/journal.pone.0125814

Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, G., & Jarvis, A. (2005). Very High Resolution Interpolated Climate Surfaces for Global Land Areas, *1978*, 1965–1978. https://doi.org/10.1002/joc.1276

Huber, D. M., & Haneklaus, S. (2007). Managing nutrition to control plant disease. *Landbauforschung Volkenrode*, *57*(4), 313–322.

INS. (2017). *Annuaire Statistique 2015*.

IPGRI, INIBAP, & CIRAD. (1996). *Descriptors for Banana (Musa spp.)*.

ISRIC. (n.d.). SoilGrids - global gridded soil information. Retrieved from ftp://ftp.soilgrids.org/data/recent/

Jain, A. K. (2010). Data clustering : 50 years beyond K-means. *Pattern Recognition Letters*, *31*(8), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

Jakeman, A. J., & Letcher, R. A. (2003). Integrated assessment and modelling: Features, principles and examples for catchment management. *Environmental Modelling and Software*, *18*(6), 491–501. https://doi.org/10.1016/S1364-8152(03)00024-0

Jogo, W., Karamura, E., Tinzaara, W., Kubiriba, J., & Rietveld, A. (2013). Determinants of Farm-Level Adoption of Cultural Practices for Banana Xanthomonas Wilt Control in Uganda. *Journal of Agricultural Science*, *5*(7), 70–81. https://doi.org/10.5539/jas.v5n7p70

Johnson, A. K. L., & Cramb, R. A. (1991). Development of a simulation based land evaluation system using crop modelling, expert systems and risk analysis. *Soil Use and Management*, *7*(4), 239–246.

https://doi.org/10.1111/j.1475-2743.1991.tb00881.x

Johnson, A. K. L., & Cramb, R. A. (1996). Integrated land evaluation to generate risk-efficient land-use options in a coastal catchment. *Agricultural Systems*, *50*(3), 287–305. https://doi.org/10.1016/0308-521X(94)00072-Y

Johnson, A. K. L., Cramb, R. A., & Wegener, M. K. (1994). The use of crop yield prediction as a tool for land evaluation studies in Northern Australia. *Agricultural Systems*, *46*(1), 93–111. https://doi.org/10.1016/0308-521X(94)90171-B

Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., … Ritchie, J. T. (2003). *The DSSAT cropping system model*. *European Journal of Agronomy* (Vol. 18). https://doi.org/10.1016/S1161-0301(02)00107-7

Kalogirou, S. (2002). Expert systems and GIS : an application of land suitability evaluation, *26*, 89–112.

Kamira, M., Ntamwira, J., Sivirihauma, C., Ocimati, W., van Asten, P., Vutseme, L., & Blomme, G. (2016). Agronomic performance of local and introduced plantains, dessert, cooking and beer bananas (Musa spp.) across different altitude and soil conditions in eastern Democratic Republic of Congo. *African Journal of Agricultural Research*, *11*(43), 4313–4332. https://doi.org/10.5897/AJAR2016.11424

Karamura, D. (1998). *Numerical Taxonomic Studies of the East African Highland Bananas (Musa AAA-East Africa) in Uganda*.

Karamura, D., Kitavi, M., Nyine, M., Ochola, D., Ocimati, W., Muhangi, S., & Talengera, D. (2016). Genotyping the local banana landrace groups of East Africa, 67–74. https://doi.org/10.17660/ActaHortic.2016.1114.9

Karamura, E. B., Jogo, W., Rietveld, A., Ochola, D., Staver, C., & Tinzaara, W. (2013). Effectiveness of Agro-Ecological Intensification Practices in Managing Pests in Smallholder Banana Systems in East and Central Africa.

Karamura, E., Frison, E., Karamura, D. A., & Sharrock, S. (1998). Banana production systems in eastern and southern Africa. In *Bananas and Food Security. Proceedings of an International Symposium held in Douala, Cameroon* (pp. 401–412).

Karamura, E., Frison, E., Karamura, D. A., & Sharrock, S. (2000). Banana production systems in eastern and southern Africa, 401–412.

Keating, B. A., Carberry, P. S., Hammer, G. L., Probert, M. E., Robertson, M. J., Holzworth, D., … Smith, C. J. (2003). An overview of APSIM, a model designed for farming systems simulation. *European Journal of Agronomy*, *18*(3–4), 267–288. https://doi.org/10.1016/S1161-0301(02)00108-9

Kitavi, M., Downing, T., Lorenzen, J., Karamura, D., Onyango, M., Nyine, M., … Spillane, C. (2016). The triploid East African Highland Banana (EAHB) genepool is genetically uniform arising from a single ancestral clone that underwent population expansion by vegetative propagation. *Theoretical and Applied Genetics*, *129*(3), 547–561. https://doi.org/10.1007/s00122-015-2647-1

Komatsu, K., Sato, Y., Kitanishi, K., Suzuki, K., & Muranga, F. I. (2010). *Cooking banana in Africa*.

Kubiriba, J., Ssali, R. T., Barekye, A., Akankwasa, K., Tushemereirwe, W. K., & Batte, M. (2016). The performance of East African highland bananas released in farmers ' fields and the need for their further improvement, 231–238. https://doi.org/10.17660/ActaHortic.2016.1114.32

Lamirel, J.-C. (2016). Reliable Clustering Indexes. *LNAI*, (9799), 103–114. https://doi.org/10.1007/978-3-319-42007-3

Lescot, T. (2012). 2010 Banana producion and trade. *Fruitrop*, (200), 74–78.

Liu, Y., Li, Z., Xiong, H., Gao, X., & Wu, J. (2010). Understanding of Internal Clustering Validation Measures. In *20120 IEEE International Conference on Data Mining*. https://doi.org/10.1109/ICDM.2010.35

Maruo, S. (2002). Differentiation of Subsistence Farming Patterns Among the Haya Banana Growers in Northwestern Tanzania. *African Study Monographs*, *23*(December), 147–175. Retrieved from http://jambo.africa.kyoto-u.ac.jp/kiroku/asm_normal/abstracts/pdf/23-4/p147-175.pdf

Maruo, S. (2007). Development of the Plantain-based Culture of the Nyakyusa of Southern Tanzania. *Change*, (March), 21–38.

McIntyre, B. D., Gold, C. S., Ssali, H., & Riha, S. J. (2003). Effects of mulch location on banana weevil, soil and plant nutrients, soil water and biomass in banana fields. *Biol Fertil Soils*, 74–79. https://doi.org/10.1007/s00374-003-0681-7

Milligan, G. W., & Cooper, M. C. (1985). An Examination of Procedures for Determining the Number of Clusters in a Data Set. *Psychometrika*, *50*(2), 159–179.

Musabyimana, T., Seshu Reddy, K. V., Ngode, L., & Prasad, J. S. (2000). Evaluation of Banana Cultivars

for Resistance To the Banana Weevil Cosmopolites Sordidus and Nematode Complex in Western Kenya. In K. Craenen (Ed.), *International Symposium for Banana and Plantain for Africa* (pp. 137–143). Acta Horticulturae.

Mwebaze, J., Tusiime, G., Tushemereirwe, W., & Kubiriba, J. (2006). The survival of Xanthomonas Campestris PV.musacearum in soil and plant debris. *African Crop Science Journal*, *14*(2), 121–127. https://doi.org/10.4314/acsj.v14i2.46166

NBS. (2012). *National Sample Census of Agriculture 2007/2008. Volume II: Crop Sector - National Report* (Vol. II). Retrieved from http://www.agriculture.go.tz/publications/english docs/Crops National Report (2008).pdf%5Cnhttp://catalog.ihsn.org/index.php/catalog/3794%5Cnhttp://catalog.ihsn.org/index.php/catalog/3794/download/52493

Ndabamenye, T., Van Asten, P. J. A., Blomme, G., Vanlauwe, B., Swennen, R., Annandale, J. G., & Barnard, R. O. (2013). Scientia Horticulturae Ecological characteristics and cultivar influence optimal plant density of East African highland bananas (Musa spp ., AAA-EA) in low input cropping systems. *Scientia Horticulturae*, *150*, 299–311. https://doi.org/10.1016/j.scienta.2012.11.037

Ndabamenye, T., Vanlauwe, B., Van Asten, P. J. A., Blomme, G., Swennen, R., Uzayisenga, B., … Barnard, R. O. (2013). Influence of plant density on variability of soil fertility and nutrient budgets in low input East African highland banana (Musa spp . AAA-EA) cropping systems, 187–202. https://doi.org/10.1007/s10705-013-9557-x

Ndayitegeye, O., Blomme, G., Ocimati, W., Crichton, R., Ntamwira, J., Muchunguzi, P., … Gaidashova, S. (2017). Multi-locational evaluation of cooking banana cultivars NARITA 4 hybrid and Mpologoma in Rwanda. *African Journal of Agricultural Research*, *12*(41), 3068–3071. https://doi.org/10.5897/AJAR2017.12608

NISR. (2010). *Rwanda National Agriculture Survey 2008.* Kigali, Rwanda.

Niyongere, C., Ateka, E., Losenge, T., Blomme, G., & Lepoint, P. (2011). Screening Musa genotypes for banana bunchy top disease resistance in Burundi. *Acta Horticulturae*, *897*, 439–448. https://doi.org/10.17660/ActaHortic.2011.897.60

Nouri, H., Mason, R. J., & Moradi, N. (2017). Land suitability evaluation for changing spatial organization in Urmia County towards conservation of Urmia Lake. *Applied Geography*, *81*, 1–12.

https://doi.org/10.1016/j.apgeog.2017.02.006

Nyine, M., & Pillay, M. (2011). The Effect of Banana Breeding on the Diversity of East African Highland Banana (Musa , AAA), 225–230.

Nyombi, K., Asten, P. J. A. Van, Corbeels, M., Taulya, G., Leffelaar, P. A., & Giller, K. E. (2010). Field Crops Research Mineral fertilizer response and nutrient use efficiencies of East African highland banana (Musa spp ., AAA-EAHB , cv . Kisansa). *Field Crops Research*, *117*(1), 38–50. https://doi.org/10.1016/j.fcr.2010.01.011

Nyquist, W. E., & Baker, R. J. (1991). Estimation of heritability and prediction of selection response in plant populations. *Critical Reviews in Plant Sciences*, *10*(3), 235–322. https://doi.org/10.1080/07352689109382313

Ochola, D., Jogo, W., Odongo, M., Tinzaara, W., Onyango, M., & Karamura, E. (2014). Household dynamics influencing effective eradication of Xanthomonas wilt in smallholder banana systems in Ugunja division-Kenya. *African Journal of Agricultural Research*, *9*(26), 2031–2040. https://doi.org/10.5897/AJAR2014.8758

Ocimati, W., Karamura, D., Rutikanga, A., Sivirihauma, C., Ndungo, V., Adheka, J., … Blomme, G. (2013). Musa Germplasm Diversity Status across a Wide Range of Agro-ecological Zones in Rwanda, Burundi and Eastern Democratic Republic of Congo. In *Banana Systems in the Humid Highlands of Sub-Saharan Africa* (pp. 8–21). CAB International.

Odeke, M., Rubaihayo, P. R., & Osiru, D. S. O. (1999). Effect of spacing, stage and method of desuckering on bunch size and yield of banana cultivar Kibuzi(AAA-EA). *African Crop Science Journal*, *7*(4), 349–353. Retrieved from http://www.bioline.org.br/request?cs99025

Okech, S. H., Gold, C. S., Speijer, P. R., Ssali, R. T., & Karamura, E. B. (2000). Relationships between soil fertility, banana weevil and nematodes in hte East African highland cooking banana in Ntungamo, Southwestern Uganda. *Acta Horticulturae*, *19*(2), 505–514.

Okech, S. H., Gold, C. S., & Ssali, H. (2004). Effects of potassium deficiency , drought and weevils on banana yield and economic performance in Mbarara , Uganda, (1996), 511–519.

Pribyl, D. W. (2010). Geoderma A critical review of the conventional SOC to SOM conversion factor. *Geoderma*, *156*(3–4), 75–83. https://doi.org/10.1016/j.geoderma.2010.02.003

Rabati, a. P., Jafarzadeh, a. a., Shahbazi, F., Rezapour, S., & Momtaz, H. R. (2012). Qualitative and

quantitative land-suitability evaluation for sunflower and maize in the north-west of Iran. *Archives of Agronomy and Soil Science*, *58*(11), 1229–1242. https://doi.org/10.1080/03650340.2011.579597

Rossiter, D. G. (1990). ALES: a framework for land evaluation using a microcomputer. *Soil Use and Management*, *6*(1), 7–20. https://doi.org/10.1111/j.1475-2743.1990.tb00790.x

Rossiter, D. G. (1995). Economic land evaluation: why and how. *Soil Use and Management*, *11*(3), 132–140. https://doi.org/10.1111/j.1475-2743.1995.tb00511.x

Rossiter, D. G. (1996). A theoretical framework for land evaluation, *72*, 165–190.

Rossiter, D. G. (2003). Biophysical Models in Land Evaluation. In *Encyclopedia of Life Support Systems (EOLSS) Section 1.5 "Land Use and Land Cover"* (Vol. II, pp. 1–16).

RStudio, T. (2015). RStudio: Integrated Development for R. Boston, MA: RStudio, Inc. Retrieved from http://www.rstudio.com/

Samranpong, C., Ekasingh, B., & Ekasingh, M. (2009). Economic land evaluation for agricultural resource management in Northern Thailand. *Environmental Modelling & Software*, *24*(12), 1381–1390. https://doi.org/10.1016/j.envsoft.2009.07.004

Sathiamoorthy, S.; Jeyabaskaran, K. J. (1989). Potassium Management of Banana, (1988), 499–516.

Schnug, E., Heym, J., & Achwan, F. (1996). Establishing critical values for soil and plant analysis by means of the Boundary Line Development System (BOLIDES) Aus dem Institut für Pflanzenernährung und Bodenkunde. *Communications in Soil Science and Plant Analysis*, *27*(May), 2739–2748. https://doi.org/10.1080/00103629609369736

Sebuwufu, G., Rubaihayo, P. R., & Blomme, G. (2004). Sebuwufu et al, 2004.pdf. *African Crop Science Journal*, *12*(1), 85–93.

Segerstedt, A., & Bobert, J. (2013). Revising the potential of large-scale Jatropha oil production in Tanzania: An economic land evaluation assessment. *Energy Policy*, *57*, 491–505. https://doi.org/10.1016/j.enpol.2013.02.023

Sidi, T., & Yameogo, A. E. (2013). *Enquete Nationale Agricole du Burundi 2011-2012*.

Smit, B., Brklacich, M., Dumanski, J., MacDonald, K. B., & Miller, M. H. (1986). Integral land evaluation and its application to policy. *Soil Survey and Land Evaluation*, *6*(3), 99–113.

Smith, H. F., & Sullivan, C. A. (2014). Ecosystem services within agricultural landscapes-Farmers' perceptions. *Ecological Economics*, *98*, 72–80. https://doi.org/10.1016/j.ecolecon.2013.12.008

Smithson, P. C., Mcintyre, B. D., Gold, C. S., Ssali, H., & Kashaija, I. N. (2001). Nitrogen and potassium fertilizer vs . nematode and weevil effects on yield and foliar nutrient status of banana in Uganda, 239–250.

Sonneveld, M. P. W., Hack-ten Broeke, M. J. D., van Diepen, C. A., & Boogaard, H. L. (2010). Thirty years of systematic land evaluation in the Netherlands. *Geoderma*, *156*(3–4), 84–92. https://doi.org/10.1016/j.geoderma.2010.02.023

Ssemulugo, M., & Rich, J. (n.d.). Schooling Uganda. Retrieved from https://schooling.ug/

Stover, R. H., & Simmonds, N. W. (1987). *Bananas* (3rd ed.). Harlow, Essex, England: Longman Scientific & Technical.

Swart, J. A. A., & Zevenberg, J. (2017). Utilitarian and nonutilitarian valuation of natural resources: a game-theoretical approach. *Restoration Ecology*, 1–10. https://doi.org/10.1111/rec.12504

Taulya, G. (2013). East African highland bananas (Musa spp. AAA-EA) "worry" more about potassium deficiency than drought stress. *Field Crops Research*, *151*, 45–55. https://doi.org/10.1016/j.fcr.2013.07.010

Taulya, G., van Asten, P. J. A., Leffelaar, P. A., & Giller, K. E. (2014). Phenological development of East African highland banana involves trade-offs between physiological age and chronological age. *European Journal of Agronomy*, *60*, 41–53. https://doi.org/10.1016/j.eja.2014.07.006

Tushemereirwe, W., Batte, M., Nyine, M., Tumuhimbise, R., Barekye, A., Ssali, T., … Swennen, R. (2015). *Performance of Narita Banana Hybrids in the Preliminary Yield Trial , Uganda*.

Tushemereirwe, W. K. (1993). Fusarium Wilt of Banana in Uganda. In *International symposium on recent developments in banana cultivation technology* (pp. 240–246).

Tushemereirwe, W. K., Bagamba, F., Katungi, E., Kikulwe, E. M., Karamura, D., Edmeades, S., & Smale, M. (2006). *Uganda Country Report: A baseline assessment of banana production and management practices in Uganda*.

Tushemereirwe, W. K., Kangira, A., Kubiriba, J., Nakyanzi, M., & Gold, C. S. (2004). Diseases Threatening Banana Biodiversity in Uganda. *African Crop Science Journal*, *12*(1), 19–26.

UBOS. (2010). *Uganda Census of Agriculture: Crop Area and Production Report* (Vol. 4). Retrieved from www.ubos.org

Van Asten, P. J. A., Gold, C. S., Okech, S. H., Tushemereirwe, W. K., Gaidashova, S. V., & De Waele, D. (2004). Soil quality problems in East African banana systems and their relation with other yield loss factors, *13*(2), 20–25.

Van Asten, P. J. A., Gold, C. S., Wendt, J., Waele, D. De, Ssali, H., & Tushmereirwe, W. K. (2006). The contribution of soil quality to yield and its relationship with other factors in Uganda, *1950*(December 2003), 8–9.

van Lanen, H. A. J. (1991). *Qualitative and quantitative physical land evaluation: An operational approach*.

van Lanen, H. A. J., Hacktenbroeke, M. J. D., Bouma, J., & Degroot, W. J. M. (1992). A Mixed Qualitative Quantitative Physical Land Evaluation Methodology. *Geoderma*, *55*, English Article.

Verheye, W. (2002). The FAO Guidelines for Land Evaluation. In *Encyclopedia of Life SUpport Systems (EOLSS-UNESCO)* (Vol. II, pp. 76–100).

Wairegi, L. W. I. (2010). *Management practices and opportunities in East African highland banana (*. Wageningen University.

Wairegi, L. W. I., Asten, P. J. A. Van, Tenywa, M., & Bekunda, M. (2009). Scientia Horticulturae Quantifying bunch weights of the East African Highland bananas (Musa spp . AAA-EA) using non-destructive field observations, *121*, 63–72. https://doi.org/10.1016/j.scienta.2009.01.005

Wairegi, L. W. I., Van Asten, P. J. A., Tenywa, M., & Bekunda, M. (2009). Scientia Horticulturae Quantifying bunch weights of the East African Highland bananas (Musa spp . AAA-EA) using non-destructive field observations, *121*, 63–72. https://doi.org/10.1016/j.scienta.2009.01.005

Wairegi, L. W. I., Van Asten, P. J. A., Tenywa, M. M., & Bekunda, M. A. (2010). Field Crops Research Abiotic constraints override biotic constraints in East African highland banana systems. *Field Crops Research*, *117*(1), 146–153. https://doi.org/10.1016/j.fcr.2010.02.010

Wandahwa, P., & Van Ranst, E. (1996). Qualitative land suitability assessment for pyrethrum cultivation in west Kenya based upon computer-captured expert knowledge and GIS. *Agriculture, Ecosystems and Environment*, *56*(3), 187–202. https://doi.org/10.1016/0167-8809(95)00641-9

Wei, Y. P., Chen, D., White, R. E., Willett, I. R., Edis, R., & Langford, J. (2009). Farmers' Perception of Environmental Degradation and their Adoption of Improved Management Practices in Alxa, China. *Land Degradation and Development*, *20*(April), 336–346. https://doi.org/10.1002/ldr

Windhausen, V. S., Wagener, S., Magorokosho, C., Makumbi, D., Vivek, B., Piepho, H., … Wagener, S. (2012). Strategies to Subdivide a Target Population of Environments : Results from the CIMMYT-Led Maize Hybrid Testing Programs in Africa. *Crop Science*, *52*(September-October), 2143–2152. https://doi.org/10.2135/cropsci2012.02.0125