Mapping the major soil-landscape resources of the Ethiopian Highlands using random forest

Leenaars, J. G. B., Elias, E., Wosten, J. H. M., Ruiperez Gonzalez, M., & Kempen, B.

Mapping the major soil-landscape resources of the Ethiopian Highlands using random forest

J.G.B. Leenaars[a]*, E. Elias[b], J.H.M. Wösten[c], M. Ruiperez-González[a], B. Kempen[a]

[a] ISRIC - World Soil Information, PO Box 353, 6700 AJ Wageningen, The Netherlands

[b] College of Natural and Computational Sciences, Centre for Environmental Science, Addis Ababa University, Addis Ababa, Ethiopia

[c] Wageningen Environmental Research, PO Box 47, 6700 AA Wageningen, The Netherlands

* Corresponding author. Tel: +31 (0)317 483749; email: johan.leenaars@isric.org

Abstract

Geospatially explicit information of soil-landscape resources of Ethiopia is lacking or fragmented for much of the country. Recently, massive soil data were collected, however these are limited to properties related to soil fertility and valid for the topsoil only. Understanding the country's soil-landscape resources, including their qualities and constraints beyond the topsoil, remains key information for systematic and reliable scaling up of evidence-based agricultural best practices including soil fertility management recommendations. The objective of this study was to produce a coherent dataset of the major soil-landscape resources of 30 highland woredas (districts), contributing to the Agricultural Growth Program of the Government of Ethiopia. The study started with an exploratory survey to identify the major (most common) soils occurring across the landscapes followed by a full survey to assess the distribution of the identified major soils. Representative soil profiles were characterised from soil pits and classified as Reference Soil Groups (RSGs), with prefix qualifiers (PQs), according to the World Reference Base for soil resources (WRB). A large number of soil profiles was classified from auger observations. Observed soil classes at both RSG and RSG+PQ level were combined with spatial explanatory variables (covariates), representing the soil forming factors in the landscapes, and their relationships were modelled and validated by random forest. A multitude of tree models was trained using each profile for calibration in approximately two third and cross-validation in approximately one third of the models. Cross-validation showed that RSGs were predicted with a reasonable overall purity of 0.58 and RSGs+PQ were predicted with a purity of 0.48. The most relevant covariate in the models was the Geomorphology and Soils map of Ethiopia at 1: 1 M scale disaggregated into soil-landscape facets. Next models were used to predict soil classes across woredas which resulted in a 250 m resolution raster map of the most probable major soils. This raster map was generalised into a polygon map of major soil-landscape resources. The purity of this final map was estimated to be 0.54 for RSGs and 0.45 for RSGs+PQ. Soil properties relevant for agricultural interpretation, such as depth, drainage, texture, pH, CEC and organic carbon and nutrient contents, were mapped according to the RSGs depicted on the soil-landscape resources map with a RMSE/mean ratio of on average 42%. We conclude that soil expert knowledge and conventional soil-landscape survey combined with random forest modelling results in an attractive hybrid approach. The approach proves cost-effective and sufficiently accurate and can be used to inform scaling up of evidence-based agricultural best practices.

Keywords: soil survey, WRB, digital soil mapping, disaggregation, agricultural best soil fertility management practice

Short running title: Mapping major soils in Ethiopia using random forest

1 Introduction

As part of the Growth and Transformation Plan, and more specifically the Agricultural Growth Program[1], the Ethiopian Government aims at doubling crop production in the foreseeable near future. This is essential in face of a population increasing with 2.5 to 3% per year (World Bank, 2018) while, at the same time, natural resources, and particularly soil-landscape resources, are degrading at an annual rate of 3% (Berry, 2003). Soil fertility has been identified as one of the key challenges to double crop production because, while high in absolute terms (Hengl et al., 2017; Murphy, 1959), soil fertility is deficient relative to, high crop nutrient demands in Ethiopia (GYGA, 2018). However, soil fertility is declining at an estimated annual rate of some 30 kg N ha^{-1} and 15-20 kg P ha^{-1} due to unsustainable practices and soil erosion progressing at an annual rate of some 300 km^2 of arable land and 1 billion tonnes of topsoil (Berry, 2003). To address these soil related challenges, the Ministry of Agriculture and the Agricultural Transformation Agency established the Ethiopian Soils Information System (EthioSIS) implementing the Soil Fertility Road Map. Among others, EthioSIS aims to (i) establish databases comprising national and regional soil and land resources; (ii) conduct surveys and soil fertility mapping to reformulate fertiliser recommendations; and (iii) develop tools for development of integrated soil fertility management technologies. This study contributes in particular to the first EthioSIS' aim by establishing geospatial data of soil-landscape resources (soil classes occurring within the landscape). These data, together with detailed soil fertility data, are important information to assess nutrient deficiencies as the gap between crop nutrient demand and soil nutrient supply thereby addressing aims ii and iii. Information on nutrient deficiencies, coherently assessed across widely variable conditions using proven agronomic models such as WOFOST (van Diepen et al., 1989) and QUEFTS (Smaling and Janssen, 1993), is needed to formulate crop and site-specific fertiliser recommendations and key hereto are geospatial data of soil-

[1] Abbreviations: AfSP: Africa Soil Profiles database, AGP: Agricultural Growth Program, CASCAPE: capacity building for scaling up of evidence-based best practices in agricultural production in Ethiopia, EthioSIS: Ethiopia Soil Information Service, IB: in the bag, OOB: out of bag, PQ: prefix qualifier, RF: random forest, RMSE: root mean square error, RSG: Reference Soil Group, RSG+PQ: Reference Soil Group with one first prefix qualifier, SoTer: Soil and Terrain database, SRTM DEM: Shuttle Radar Topography Mission based Digital Elevation Model, WRB: world reference base for soil resources

landscape resources including its qualities and constraints. A particularly important soil quality, codetermining crop potentials and fertiliser use efficiencies and thereby suitable integrated soil fertility management packages (Sanginga and Woomer, 2009; Vanlauwe et al., 2010), is the root zone plant-available water holding capacity. This quality is constrained by rootable depth (Leenaars et al., 2018) and assessed by soil-landscape resources which are characterised over the entire profile depth. The added value of evaluating soil-landscape resources, in addition to currently available soil fertility parameters from the topsoil only, is illustrated by Figure 1 showing crop performance on two black soils of similar fertility in the topsoil but of contrasting depths. The crop on the deep black soil (Vertisol) is still growing whereas the crop on the shallow black soil (Leptosol) has already wilted due to limited soil water availability. Indeed, topsoil data alone do not suffice to assess the efficiency of crop response to fertiliser inputs which is confirmed, based on the interpretation of data generated by the Africa Soil Information Service (AfSIS) project, by Tamene et al. (2017) who recommend to use soil resource data for that purpose. Soil-landscape resources can be surveyed and mapped cost-efficiently by soil experts making low-cost field observations, minimising the number of costly soil analyses, combined with digital soil mapping techniques, minimising the number of field observations needed and replacing time and cost-intensive manual delineation of map units and manual attribution of soil classes. Maps of the major soil-landscape resources provide an adequate basis for defining site-specific fertiliser nutrient recommendations, at high resolution if combined with detailed soil fertility data, and subsequent generalisation into, manageable, spatial fertiliser product domains (Elias et al., 2019). In other words, maps of the extent and distribution in the landscape of the major soils, including their properties, reflect the very basic information to generate and communicate crop and site-specific recommendations for integrated soil fertility management and, thus, to inform scaling up of evidence-based agricultural best practices.

a) b)

<Insert Figure 1a, 1b>

The CASCAPE project (capacity building for scaling up of evidence-based best practices in agricultural production in Ethiopia) intervenes in 30 high-potential woredas (districts) in support to the AGP of the Government of Ethiopia. The present study is an initiative of the CASCAPE project and contributes particularly to EthioSIS. Specific objectives of this study are to:

- Develop a coherent geospatial dataset of the major soil-landscape resources at reconnaissance scale (approximately 1: 250,000) to inform upscaling of evidence-based agricultural best practices in 30 woredas in support to the AGP;
- Survey, map and validate the major soil resources, and associated soil properties, of 30 woredas in a cost-effective hybrid approach combining expert soil knowledge and conventional soil-landscape survey approaches with digital soil mapping techniques.

2 Methods

The study area comprised 30 high-potential woredas with a total area of 26,820 km^2 located in the Ethiopian Highlands (Figure 2) and distributed over four main administrative regions (Tigray, Amhara, Oromia and the SNNPR). The soil-landscape of the Ethiopian Highlands is distinctly different from much of the soil-landscapes in the African continent. This is particularly due to the elevation of these highlands as a result of the rising of Ethiopia over the last 75 million years and the dominant parent material of flood basalts which started to form around 30 million years ago. These basalts reach a thickness of nearly 1 km and were deposited on top of limestone underlain by sandstones and metamorphic pre-Cambrian rocks. Volcanic eruptions caused important but somewhat local deposits of ash on top of the basalts (FAO, 1984; Billy, 2015; Abbate, 2015; Elias, 2016). Much of the poorly to imperfectly drained depressions on the plateaus are covered by thick alluvium. Eventually, the Ethiopian Highlands were bisected by the Great Rift Valley and currently much of the landscape is composed of rolling to broken-hilly plateaus at an altitude of around 2000 m above sea level with little of its surface falling below 1500 m. The plateaus are deeply and steeply dissected by several major rivers causing inevitable large scale erosion. Large calderas tower above the plateaus with the summits reaching heights of up to 4550 m. The predominant climate of the Ethiopian Highlands is tropical monsoon, with important summer rains and lesser short rains in spring, with temperatures well below those at similar proximity to the equator and with alpine character at higher altitudes. Annual rainfall in the highlands is on average some 1500 mm and increases from below 800 mm in the north and east to over 2000 mm in the southwest (Elias, 2016). Potential evaporation is well below 1000 mm annually. This gives an annual rainfall surplus throughout the vast majority of the highlands (Fazzini et al., 2015) which results in the formation of deeply, but not necessarily highly, weathered soils. The variability of soil forming factors in the highlands is reflected in the variability of soils including Vertisols, Nitisols, Luvisols and important areas of Leptosols and further some Cambisols, Regosols, Alisols, Andosols, Phaeozems, Calcisols, Fluvisols and Planosols (FAO, 1984; Jones et al., 2013; Hengl et al., 2017b). Almost all these soils have very high CEC, generally exceeding 36 cmolc kg^{-1} due to high contents of clay and organic carbon (generally over 40% and 1.5%, respectively), the high activity nature of the clay and local occurrences of allophane, while the vast majority also has high base saturation indicative for fertile soils. Indeed, soil fertility in the Ethiopian Highlands is generally very high, with a large soil nutrient supply to crops, but deficient relative to the even higher crop nutrient demand. The latter is driven by potential crop transpiration, as limited by the actual availability of (soil) water, and sets the reference for any soil fertility management strategy. Simulated water-limited yield potentials, assuming sufficient nutrients, reach levels in the Ethiopian Highlands of some 6 to over 17 t ha^{-1} for maize and 4 to nearly 10 t ha^{-1} for wheat depending on the specific combination of climate and soil. Corresponding yield gaps, relative to actual yields, vary between 4 and 15 t ha^{-1} for maize and 2.5 to 7.5 t ha^{-1} for wheat (GYGA, 2018) and the overall challenge is to narrow these gaps by crop and site specific agricultural practices.

In this study we i) surveyed soil resources, ii) mapped soil resources and iii) validated the soil resources map, all in relation to the geomorphic landscape, followed by iv) mapping and validation of soil properties associated with the soil-landscape resources map, for 30 highland woredas.

2.1 Soil survey

The study started with an exploratory soil survey at kebele level (sub-districts in a woreda) to identify, characterise and classify the major soils that could be expected to occur across the woreda. Major soils were defined as soils that are most common in the surveyed area and map-able at the scale of operation. These soils can be classified at different levels of classification in accordance with the major soils of the world (Driessen and Dudal, 1989; Driessen et al., 2001). The exploratory survey at kebele level was followed by a full survey at woreda level to verify the identified and classified soils across the full woreda landscapes. The reason to start the survey with an exploratory survey at kebele level was that woredas are large areas (approximately 894 km^2 on average and varying between 168 and 2507 km^2) with important parts difficult to access whereas kebeles are about 30 times smaller (approximately 29.5 km^2 on average in a range between 4 and 94 km^2).

For each of the 30 woredas two to four kebeles were selected based on their accessibility and representativeness for agricultural land according to expert local knowledge and for woreda landscapes, resulting in a total of 112 kebeles. To guide the exploratory survey, the map of the Geomorphology and Soils of Ethiopia (FAO, 1984) was used as a base map (Figure 2) together with elevation and a slope map derived from an adjusted SRTM DEM (Vågen, 2010).

<Insert Figure 2>

Having an adequate base map is a first key step for a surveyor to build a conceptual model of soil-landscape relations (Minasny and McBratney, 2016) and design a survey scheme. The base map that we used, delineates geomorphic landscape units for the entire country at a scale of 1: 1 M and represents the soil associations distinguished at that scale. This base map is thus not spatially explicit about soil-landscape relations. However, the soil associations are described in the associated report (see Figure 3) by landscape facets that occur within each landscape unit together with the corresponding soil units and phases (FAO, 1984). These soil units are classified according to the legend of the FAO/Unesco Soil Map of the World (FAO, 1974). We used the descriptions to decide about the localisation of soil observations within each of the landscape units. The same map is also available as the Soil and Terrain (SoTer) database for north-eastern Africa (FAO, 1998) where soil units, associated with unmapped terrain components, are classified according to the revised legend (FAO, 1988).

Soil observations were made in first instance by looking at the landscape and at road cuts and surface features such as stoniness, colour and vegetation, trying to recognise the variability expected from the base map, followed by 743 auger point observations (25 per woreda). The auger points were distributed over and within the landscape units indicated by the base map attempting to cover the most important landscape facets, georeferenced using GPS and briefly described to a depth of 120 cm, unless restricted by hard rock or an impenetrable layer, according to a form prepared from the Guidelines for Soil Description (FAO, 2006). Master horizons were distinguished and described by depth interval, coarse fragments content, texture, colour, consistency, root presence and diagnostic features. The auger points were tentatively classified in the field as Reference Soil Groups (RSGs) with one prefix qualifier (PQ) according to the World Reference Base for soil resources (WRB), 2nd edition (IUSS Working Group WRB, 2007. The 3rd edition of WRB (IUSS Working Group WRB, 2015) was not yet available at the time of exploratory survey. The major soils were identified from these tentative field classifications at kebele level. These soils were assumed to be the most common in the surveyed area and map-able at targeted scale. Note that it's only after the mapping procedure that the true major, most common and map-able, soils are known.

Subsequently, representative soil profiles were selected, characterised in detail and classified. This was done for 204 profile pits (about 7 per woreda). Locations were not evenly distributed over the kebeles, but rather over the identified major soils. The master horizons with subordinate characteristics were designated to a depth of at least 180 cm (bedrock permitting) and described in detail using a form prepared for this purpose according to the Guidelines for Soil Description (FAO, 2006). Diagnostic horizons, properties and materials were identified and the profiles were preliminarily classified in the field according to WRB (IUSS Working Group WRB, 2007), specifying the RSG with its prefix and suffix qualifiers. These classifications were later validated with soil analytical data and corrected if needed.

Each horizon was sampled and analysed for selected soil properties by the Soil and Plant laboratory of the Water Works Design and Supervision Enterprise in Addis Ababa using standard laboratory procedures required for soil classification (van Reeuwijk et al., 2002): particle size distribution (sand, silt, clay content) by hydrometer method (Bouyoucos) with the fractions defined according to USDA standards (clay<0.002<silt<0.05<sand<2 mm), bulk density of the fine earth from core samples, pH-H_2O in a 1: 2.5 soil: water solution and pH-KCl in a 1: 2.5 soil: KCl (1M) solution, electric conductivity in 1: 2.5 soil: water solution, cation exchange capacity in a 1M NH4OAc solution buffered at pH-H_2O of 7 (Black), exchangeable bases with Ca and Mg by atomic absorption spectrometry and K and Na by flame photometry, organic carbon content by the method of Walkley-Black, total nitrogen by the method of Kjeldahl. Analysed for only the topsoil are available phosphorus by the method of Olsen, extractable micro nutrients (Fe, Mn, Zn, Cu) by the DTPA method and available sulphur. We used the soil analytical data to verify and possibly adjust the preliminary field classifications of the soil profiles (soil pits) that were representative for the identified major soils and to characterise these major soils, classified as RSGs and RSGs with a prefix qualifier (RSGs+PQ or soil units), in detail. The classifications of the tentatively classified auger observations were adjusted accordingly where necessary.

Following the exploratory survey at kebele level a full survey was conducted to verify the distribution and extent of the major soils across the entire woredas. Soil profile observations were collected by additional augering beyond the kebele level along transects defined based on the base map. For this purpose, we updated the base map by disaggregating it into spatially explicit soil units and by replacing soil units classified according to the legend of the soil map of the world (FAO, 1974) by soil units (RSGs+PQ) identified during the exploratory survey. The landscape units depicted on the map were first disaggregated into spatially explicit landscape facets which are distinguished by FAO (1984) as part of landscape transects that represent the soil-landscape (soil association) within each landscape unit based on relative position and slope class (see Figure 3). Relative positions could not be distinguished based on explicit criteria, whereas slope classes could, and we used relative elevation as a proxy. We assessed relative elevation in a 5 km radius and slope from the hydrologically adjusted SRTM DEM (Vågen, 2010), which we resampled from 90 m to 250 m, and then mapped classes of relative elevation (low, medium, high) and slope, whereby the slope classes were defined according to the current standards of the Guidelines for soil description (FAO, 2006). We then allocated soil units, and associated phases, to each of the resulting landscape facets as described by FAO (1984). This latter step was slightly time intensive and required soil expert knowledge due to the interpretation still needed for relating the relative elevation classes with the relative positions, and soils, distinguished by FAO (1984) and for defining soils for landscape facets that are steeper or less steep than as distinguished by FAO (1984). Soil units, classified according to the legend of the FAO/Unesco Soil Map of the World (FAO, 1974), were replaced by soil units, classified according to WRB (IUSS Working Group WRB, 2007), by extrapolating the georeferenced classifications collected during the exploratory survey from the kebele level extent to the woreda level extent. This extrapolation was done using a digital soil mapping approach, exactly similar to the approach described in section 2.3, whereby the disaggregated map mentioned above served as a covariate. Besides extrapolating RSGs and RSGs+PQ, we also extrapolated the qualifiers, including diagnostic qualifiers, independently. We then produced an updated base map with a legend indicating the landscape unit, the landscape facet, the 3 most probable RSGs+PQ and the 3 most probable qualifiers.

<Insert Figure 3>

In total 1329 additional auger points were made across the 30 woredas (about 45 per woreda), which were georeferenced, briefly described using a field form prepared from the Guidelines for soil description (FAO, 2006) and classified as RSG+PQ with reference to the earlier identified and classified major soils. Tentative field classifications were made, despite the fact that soils cannot be properly classified from (disturbed) auger observations because proper soil classification requires detailed soil pit observations and soil analytical data. Nevertheless, auger observations permitted to assess easily observable soil characteristics as a basis to verify diagnostic features and confirm the soil classes and qualifiers anticipated from the updated base map. Soil profiles were also classified according to local nomenclatures. This permits to correlate soils classified according to WRB with local soil names and more importantly to effectively communicate with farmers about soil and soil-related management. This match is not perfect because local soil names are commonly based on landscape features and topsoil characteristics (Mulders et al., 2001) whereas RSGs and RSGs+PQ are classified rather based on subsoil diagnostics.

Eventually, a total of 2276 soil profile point observations was realised during the survey including 204 pit observations and 2072 auger observations (743 augerings at kebele level and another 1329 at woreda level). Added to these were 282 reliably classified profiles from the Africa Soil Profiles (AfSP) database (Leenaars et al., 2014; 2014b), to a large extent from outside the woredas but located in the highlands, and 37 virtual profiles representing Leptosols randomly located in undersurveyed, inaccessibly steep, parts of the landscapes where the orginal base map indicates Lithosols, making a total of 2595 profiles. See Figure 4 for the location of soil profile observations in 4 woredas including the 4 x 4 kebeles selected for the exploratory survey.

<Insert Figure 4>

The soil profile data were compiled under a common standard, ready for use, following a tailor made template prepared from an extended version of the AfSP database (Leenaars et al., 2014). This database arranges all data entries as observations and measurements by specifying the soil object (defining the soil in 4D), property, method and value, including the unit of measure or the dictionary associated with the value, together with the lineage. (Also see Batjes et al., 2017). Numeric soil property data were standardised according to data conventions of AfSP and descriptive soil property data according to the conventions of the Guidelines for soil description (FAO, 2006) and, where lacking, according to AfSP conventions copied from SoTer conventions (van Engelen and Wen, 1995).

2.2 Soil modelling and mapping

The spatial distribution of major soils across the landscapes at woreda level was modelled and mapped with random forest. Random forest modelling is a popular machine learning technique for digital soil mapping (Grimm et al., 2008; Hengl et al., 2015; Heung et al., 2014) in line with the future direction of soil mapping as described by Minasny and McBratney (2016) and Brevik et al. (2016). We developed two random forest models: one to model and map RSGs and one for RSGs+PQ (soil units or RSGs with one first prefix qualifier). These models were trained with the classifications of the collected 2595 georeferenced soil profile observations projected onto a stack of 139 spatial explanatory layers, the covariates. These covariates represent the spatial variability within the landscape of the soil forming factors (*climate, organisms, relief, parent material* and *time*, as well as *management* explicitly mentioned here as the only factor directly influenced by humans) that are drivers of soil spatial variation. The covariates included terrain variables derived from the SRTM DEM

(Vågen, 2010), climatic variables from WorldClim (Hijmans et al., 2005) and MODIS imagery including derived land cover and vegetation index maps (AfSIS, 2015) as well as thematic maps such as the original and disaggregated version of the Geomorphology and Soils map of Ethiopia (FAO, 1984) and each of the covariates was resampled to a spatial resolution of 250 m.

Random forest is a so-called ensemble method, meaning that a model is composed of a multitude (a forest) of decision tree models (Breiman, 2001; Strobl et al., 2009). We trained forests with 500 tree models. Each tree model randomly splits the input dataset (n = 2595) in a calibration and a validation set. This split is made randomly for each model. The calibration set contains (by default) approximately 63% of the input dataset and the validation set 37%, which implies that each soil profile point observation served calibration 315 times and validation 185 times. The validation set is typically referred to as the 'out-of-bag' dataset (OOB). The calibration set, which could be called the 'in-the-bag' (IB) dataset, is used to construct the tree model. The algorithm does this by splitting the calibration set into two nodes (subsets) in such a way that these two nodes are as homogenous as possible with respect to the observed soil classes that are contained in each node. Each of the two nodes is then further split into two new and even more homogenous nodes. This splitting process continues until a stopping criterion is met. Splitting is done on basis of covariate values. For each split, the random forest algorithm randomly selects a subset of covariates that it will evaluate (by default equal to the square root of the number of covariates in case of a categorical target variable such as a soil class). From this subset, in our study consisting of 11 or 12 covariates, the optimal covariate is selected in such a way that a maximum purity is achieved in the two new nodes.

Random forest allows to assess the relevance of each covariate over all trees in the forest. We used the 'permutation accuracy importance' measure for this purpose (Strobl et al., 2009). Basically it measures the decrease in prediction accuracy when the values of one covariate are randomly permuted (hereby breaking the association with the target variable as if the covariate is excluded from the model) and then adding this permuted covariate together with the unpermuted covariates to the model to predict the target variable (Strobl et al., 2009). The decrease in accuracy is large for important covariates that correlated strongly with the target variable and small for less important covariates that are less strongly correlated. This reduced the risk of model overfitting.

The two trained random forest models were subsequently applied to the covariate stack to predict (map) RSGs and RSGs+PQ across the 30 woredas at a spatial resolution of 250 m (429,120 pixels). For each pixel in the map, the algorithm produces one prediction for each tree model in the forest. In our case, with two forests, each pixel received 500 predictions of RSG and 500 of RSG+PQ. (We summarised these predictions in soil class specific probability maps). The final prediction for a pixel is then the most frequently predicted RSG and RSG+PQ and the outcome at this stage are two raster maps of major soils predicted as RSGs and as RSGs+PQ.

In a next step, these two raster maps were generalised to two polygon maps, because raster maps are too granular to support the envisaged final purpose of the soil maps. Unique combinations of the predicted major soils (RSGs and RSGs+PQ) with landscape facets formed the basis of the aggregation of individual pixels to 'spatially homogeneous' polygons. This aggregation implied a certain level of generalisation whereby isolated pixels (or two connected pixels) of divergent soil-facet combinations were eliminated from the map. The resulting soil-facet polygons were subsequently related to the coarse scaled polygons of the map of geomorphic landscape units (FAO, 1984). The imprecise delineation of these landscape units was adjusted to match the delineations of the soil-facet map by applying a majority-minority rule to the intersected soil-facet polygons. The two resulting final maps represent unique combinations for each of the woredas of the geomorphic landscape unit, landscape facet and major soil, with the major soil classified as RSG and as RSG+PQ. The legend further indicates the 2nd and 3rd most probable RSGs and RSGs+PQ.

2.3 Soil map validation

The accuracy of the raster maps was assessed through cross-validation with independent data. This cross-validation is an integral part of the random forest algorithm which makes use of the OOB dataset for this purpose. Remember that for each tree model, 37% of the input data was randomly selected and set aside (put 'out-of-bag) and that these OOB data are not used to build (calibrate) the model. Instead, these OOB data serve as a test sample to assess the prediction accuracy of that particular model. Once a tree model is built using the calibration data, it is used to predict the target variable (soil class) at the locations of each of the OOB observations. In this way, each profile observation of our input dataset served approximately 185 times as an OOB observation (37% of 500 trees) and received 185 independent, OOB, predictions. The final prediction at the location of a given OOB observation is then taken, in case of a categorical target variable as in this study, as the most frequently predicted class. This gave us an independent prediction of the RSG and RSG+PQ at the location of each of the 2595 profile observation in the input dataset. These independent predictions were then compared to the observed RSGs and RSGs+PQ and accuracy measures were computed from this comparison.

We followed Brus et al. (2011) for the selection of accuracy measures for categorical soil maps. We computed the overall purity as a measure for the overall accuracy, here defined as the fraction (0 - 1) of the input data for which the OOB predicted class equals the observed class. The overall purity is composed of the map unit purity and class representation also referred to as the user's accuracy and the producer's accuracy, respectively. The map unit purity is defined as the fraction of OOB predictions of a particular class that is similar to the observed class (e.g. the fraction of predicted Luvisols that is observed as Luvisols). The class representation is defined as the fraction of an observed class that is correctly OOB predicted (e.g. the fraction of observed Luvisols that is predicted as Luvisols).

The polygon maps generalised from the raster maps could not be independently validated with the OOB predictions from the random forest procedure. This means that no objective measure of the accuracy of the polygon maps could be provided. Instead of evaluating the overall purity, we evaluated the 'overall correspondence' or naive accuracy by comparing soil classes observed at all the profile point locations with soil classes depicted on the final polygon maps. The overall correspondence was computed from this comparison as the fraction (0 - 1) for which the observed soil class corresponded with the depicted soil class. All profile point locations were used for this purpose, however excluding those profiles that were added from the AfSP database or that were virtual and not located within the woredas (n = 2291), which implies that this overall correspondence gives an optimistic estimate of the map accuracy since the map is not independent from the training data.

We estimated the overall purity of the final polygon maps, which were generalised from the raster maps, assuming that the ratio (fraction) of the overall purity (OP) relative to the overall correspondence (OC) is similar for polygon maps (pOP/pOC) and raster maps (rOP/rOC). For this, we assessed the overall correspondence of the raster map (rOC) and estimated the overall purity of the polygon map (pOP) from pOP = pOC x rOP/rOC.

2.4 Mapping of derived soil properties and map validation

Soil classes are regarded as carriers of soil information which we illustrated by producing maps for a selection of soil properties relevant for agricultural interpretation. Selected were depth to bedrock and drainage class (observed from the soil profile as a whole) and textural fractions, bulk density, pH, CEC, exchangeable bases, organic carbon and nutrient contents (measured from samples taken from individual soil profile horizons or depth intervals). We attributed these properties to the RSGs using the available soil profiles data and specified, for each property and each RSG, the mean, median, minimum, maximum and standard deviation. Each soil property was mapped by attributing its mean value to the RSGs depicted on the polygon map of major soil-landscape resources.

Data preparation prior to attribution included (i) conversion of censored values for depth to bedrock (e.g. > 100 cm) to absolute values for depth to bedrock (explained hereafter), (ii) conversion of drainage classes (V, P, I, M, W, S, E) to ordinal values for drainage (1, 2, 3, 4, 5, 6, 7) and (iii) conversion, by fitting mass preserving splines (Bishop et al., 1999), of soil analytical data originally measured and compiled for variable depth intervals to soil analytical data expressed according to two standard depth intervals (0 - 30 cm and 30 - 100 cm).

Censored data on depth to bedrock are highly valuable but not unambiguously interpreted. We translated censored values to absolute values by comparing censored values reported for depth to bedrock with the class values reported for rootable depth (classes according to FAO, 2006). The average was taken from the censored depth value and the bottom depth value of the rootable depth class in case the censored depth corresponded with the rootable depth class (e.g. > 70 cm corresponding with the class of 50 - 100 cm results in an estimated absolute value of 85 cm). The censored depth value was excluded from the attribution in case that the censored depth exceeded the bottom depth indicated by the rootable depth class (e.g. > 70 cm compared to the class of 30 - 50 cm). Also excluded were censored depth values for profiles that lacked a reported class value for rootable depth.

The accuracy of soil property maps was assessed by comparing the values of the soil property map with values of the soil property observed or measured at corresponding profile point locations. Assessed were the mean error, root mean square error and the relative root mean square error (in %) relative to the average value reported per RSG. We also assessed the Pearson correlation coefficient (r) between mapped and observed property values as well as the overall purity of the drainage class map.

3 Results

The results are reported in detail by Leenaars et al. (2016) and are available[2] under a Creative Commons license (CC-BY 4.0). Detailed results have also been included in the work of Elias (2016).

3.1 Soil survey

Soil profile data collected during the survey were compiled according to conventions of the AfSP database (Leenaars et al., 2014). The profiles (2595) were georeferenced and classified as RSG+PQ according to WRB and the majority also according to local nomenclatures. The match between RSGs and local soil names (53) is not reported here and proved not univocal, not in the least because the local names were in different languages. Table 1 lists the RSGs which were identified from soil profile observations together with a selection of soil properties observed or measured for each RSG. Depth of soil and drainage class were not reported for Acrisols. Soil analytical data were not available for Alisols, Andosols and Arenosols since, apparently, no representative soil profile had been sampled for analysis, and described in detail, for any of these RSGs. Apparently, these RSGs were identified and classified directly from auger observations based on the expert knowledge of the experienced surveyors, e.g. for Alisols considering parent material (Trachyte, Rhyolite), elevation (> 2,300 masl.), high annual rainfall, presence of acid tolerant crops, colour (dark reddish brown) and presence of an Argic horizon with much Manganese concretions (Abebe, 2007).

[2] www.isric.org/projects/capacity-building-scaling-evidence-based-best-practices-agricultural-production-ethiopia

<Insert Table 1>

Table 1 shows that RSGs do not differ significantly between each other for most of the soil properties except for specific RSGs which are different for specific properties. Leptosols, Regosols and Calcisols are relatively shallow, Gleysols, Phaeozems, Planosols and Vertisols are relatively less well drained and Acrisols, Gleysols, Nitisols and Planosols have relative low base saturation. Clay content, CEC, exchangeable bases, organic carbon and nutrient contents are generally high to very high for all RSGs. Sulphur content seems low but this may be due to the laboratory method used and exchangeable K also seems quite low.

RSGs identified from field work are, in order of importance considering the number of observation points, Vertisols, Luvisols, Nitisols, Leptosols and Cambisols (Table 2). These RSGs represent 82% of the soil profiles observed. Other RSGs observed are Regosols, Fluvisols, Alisols, Planosols and Andosols, together representing 15% of the observations, and the remaining 3% of the observations are classified as Arenosols, Phaeozems, Gleysols, Acrisols and Calcisols.

The original 1: 1 M version of the base map used for the exploratory survey at kebele level appeared valuable for guiding surveyors across the major landscapes and detect the soils occurring at kebele level. The exploratory survey was insufficient though to detect the variability in soils at woreda level and Alisols, Andosols and Arenosols were not identified. The updated 1: 250,000 version of the base map prepared for the full survey at woreda level did not depict these unidentified soils but proved effective after all in guiding the surveyors across a wider variability in spatially explicit soil-landscape resources. This resulted in the detection of additional soil-landscape combinations and the identification of additional soils. The updated base map was independently validated using profile observations collected during the full survey beyond the kebele level and had an overall purity of only 0.44. The same base map had an overall purity at kebele level of 0.58 when independently validated using the OOB procedure and the profile observations collected at kebele level. A major weak point of the base map, both the original and updated version, was the delineation of the landscape units which is imprecise.

3.2 Soil models and maps

Of the 139 covariates, the Geomorphology and Soils map of Ethiopia (FAO, 1984) disaggregated to soil-landscape facets (FAO84_soilF) proved to be the most relevant for modelling of the soil variability across the woreda landscapes, exceeding the relevance of stacks of satellite data. The permutation index in Figure 5 shows a mean decrease in accuracy of near 25% if this map would be removed (permuted) as a covariate in the random forest modelling of RSGs. This covariate is particularly important to explain soil variability occurring within geomorphic landscapes at the, relatively, short distance scale of the topo-sequence. The accuracy decreases with another 23% if the Geomorphology and Soils map itself (geo_legacy_agg) would be removed. This covariate is particularly important to explain long distance soil variability occurring at the scale of the country and this study.

<Insert Figure 5>

Polygonised, the RSGs mapped show an order of importance at the locations of observation which is near similar to that of the RSGs observed. Vertisols, Nitisols, Luvisols, Leptosols and Cambisols are predicted for 87% of the points (Table 2) followed by Alisols, Regosols, Planosols and Andosols (9%). Arenosols, Fluvisols, Phaeozems, Acrisols, Gleysols and Calcisols are predicted at 4% of the points only. The order of importance considering the area mapped is dominated by Nitisols, Vertisols and Leptosols, representing 83% of the area mapped, followed by Luvisols (10%) and, near negligible, small areas (0.4 to 1.8% each) of Planosols, Alisols, Regosols, Cambisols, Andosols and Arenosols (Table 2). Negligible is the area mapped as Acrisols, Calcisols, Fluvisols, Gleysols or Phaeozems, each occupying less than a half permille (< 0.05%) of the total area. Apparently, the predictions in terms of area differ considerably from the predictions at point locations. These figures suggest that Nitisols, and to a lesser extent Vertisols and Leptosols, seem to be overrepresented on the map at the cost of Luvisols, Cambisols and most of the other RSGs. The apparent underrepresentation of Luvisols is possibly due to the fact that Luvisols genetically, morphologically and laterally grade towards Vertisols and Nitisols and these soils are not always easily distinguished. The relative position of each of these soil classes along the topo-sequence is not unambiguously clear with Vertisols - Nitisols - Luvisols observed, from relatively low to high position, in one landscape and Vertisols - Luvisols - Nitisols in another and doesn’t necessarily coincide with schematic sequences as proposed by Driessen et al. (2001). This is probably due to variations in parent material along the topo-sequence which disturb the pedological sequence. The large representation of Leptosols on the map exceeds the relative number of observations which is well explained by the fact that Leptosols occupy the highest and steepest, least accessible landscape positions. Cambisols, Regosols and Fluvisols were observed frequently but are strongly underrepresented on the map. This may well be due to the fact that these soils are developed in relatively young parent materials and this soil-landscape relationship is apparently inadequately modelled due to spatial covariates which do not adequately reflect such young landscape areas with unclear characteristics. The Fluvisols are likely confused with Vertisols in the lower landscape positions and the Regosols with Leptosols in higher landscape positions. Cambisols can be expected within close distance to most other Reference Soil Groups. Figure 6 shows the final map of major soil-landscape resources for Omonada woreda as an example of the map for 30 woredas.

<Insert Figure 6>

The RSGs mapped for each woreda vary considerable between the different woredas, and regions, throughout the country, as is shown by Table 2 which summarises the relative map areas (%) occupied by each of the RSGs for each of the woredas.

<Insert Table 2>

3.3 Soil map validation

Accuracy statistics of the soil maps are presented in Table 3. Reported are the overall purity for the raster maps and deducted estimates of the overall purity for the polygon maps. The so-called overall correspondence is given for both the raster and polygon versions of the maps. A distinction is made for RSGs and RSGs+PQ.

<Insert Table 3>

The overall purity for RSGs is 0.58 and drops to 0.49 for RSGs+PQ. These accuracies are very reasonable and well in line with accuracies that are typically reported for soil class maps developed with statistical methods (Grinand et al., 2008; Kempen et al., 2009; Kempen et al., 2012b; Holmes et al., 2015, Heung et al., 2016) but also for soil class maps developed with conventional methods (Kempen et al., 2012). The overall correspondence is 0.84 for RSGs and 0.81 for RSGs+PQ. Evidently, the correspondence largely exceeds the purity. The polygon map, generalised from the raster map, shows a correspondence of 0.79 for RSGs and 0.74 for RSGs+PQ. This suggests that the accuracy of the polygon map is some 94% of that of the raster map for RSGs and 91% for RSGs+PQ. The purity estimated for the polygon map, as deducted from preceding figures, is 0.54 for RSGs and 0.45 for RSGs+PQ. Altogether, the random forest model clearly captured the RSGs fairly well but had some difficulties in capturing the larger variability associated with RSGs+PQ.

Table 4 gives the full OOB validation matrix indicating the map unit purities as well as the class representations of observed versus predicted RSGs. Calcisols, Alisols and Andosols are the map units showing highest map unit purity followed by Luvisols, Phaeozems and Vertisols. The map units for Calcisols, Alisols and Andosols were very small in size though. Alisols and Vertisols are the observed classes that are best represented by the map followed by Nitisols, Arenosols, Luvisols, Andosols and Leptosols.

<Insert Table 4>

3.4 Soil property map and validation

The soil properties summarised in Table 1 for each RSG were mapped through the RSGs depicted on the soil-landscape resources map. Figure 7 illustrates this for drainage class, pH-H_2O, extractable zinc and available P. At the national scale, drainage, pH and available P seem to be correlated and to increase from the southwest to the north and east whereas extractable zinc shows a contrary trend.

a) b)

c) d)

<Insert Figure 7a, 7b, 7c, 7d>

The overall purity of the drainage class map is 0.64 which is considered very adequate. Drainage class was, even though an ordinal variable, further included with the numeric variables. The accuracy statistics of the 24 soil property maps are reported in Table 5, both for the maps as a whole and each RSG separately. The accuracy of the maps of depth and drainage could not be assessed for Acrisols, because these properties were not recorded for Acrisols, and neither for Calcisols and Gleysols because the area on the map depicting Calcisols and Gleysols does not correspond with any point observation on depth and drainage. The accuracy of the soil analytical property maps could not be assessed for Acrisols, Andosols, Alisols, Arenosols, Gleysols and Regosols for similar reasons. Andosols, Alisols and Arenosols were not sampled and Acrisols, Gleysols and Regosols were mapped in areas which do not correspond with any sampled point observation.

The accuracy of the property maps is altogether quite reasonable and varies among the different soil properties and the different RSGs. Relative RMSE (relative to the average property value reported for each RSG) is on average 42% which is considered an acceptable small error but varies between 9 and 166%. Soil properties particularly accurately mapped by RSGs (in terms of relative RMSE) are drainage, clay, bulk density, pH and CEC. Particularly inaccurate are depth, exchangeable K and extractable Zn. The apparent inaccuracy of the depth map is due to the large relative RMSE reported for depth in the area mapped as Leptosols. The depth map further appears accurate though. Relative good Pearson correlations, between mapped and observed soil property values, are shown for drainage, depth, pH and sum of exchangeable bases and particularly bad correlations for exchangeable K and especially extractable S. The correlation is low though when considering all selected properties with an average coefficient of 0.38. Bias or mean error is generally very small.

<Insert Table 5>

4 Discussion and conclusions

Expert soil knowledge, used in a conventional soil-landscape survey approach, was combined with random forest, currently one of most popular digital soil mapping techniques, into a hybrid approach to produce a geospatial dataset of the major soil-landscape resources of 30 Ethiopian Highland woredas at an accuracy which is adequate and in a manner which is time and cost effective. This dataset includes tables with newly collected soil property data for 2276 georeferenced soil profiles classified according to WRB, compiled together with data and classifications for a few hundred profiles added from the Africa Soil Profiles database, and maps of the major soil-landscape resources (maps of RSGs, with associated soil properties, and of RSGs+PQ depicted in combination with major landscapes and landscape facets). This dataset can be included in the Ethiopian Soil Information Service (EthioSIS), and also in the World Soil Information Service (WoSIS)[3], in support to scaling up of agricultural best practices in Ethiopia, and beyond.

The final map of major soil-landscape resources is a polygon map, generalised from raster, depicting RSGs according to geomorphic landscape features at a targeted scale of 1: 250,000 (apparent scale of approximately 1: 150,000). The map accuracy, expressed by the overall purity, was estimated to be 0.54. This estimate was deducted from the, independently assessed, overall purity of 0.58 of the raster version of the map and the, so-called, overall correspondence of 0.79 of the polygon map and 0.84 of the raster map. Major soil-landscape resources were also defined and mapped as RSGs+PQ and the overall purity of both the polygon and raster map was slightly lower with 0.45 and 0.49, respectively. Apparently, higher precision comes at the cost of lower accuracy and we expect that the accuracy would become very low if we had adhered to the rules for creating map legends, suggested by IUSS Working Group WRB (2015), implying for our targeted scale to use the first 3 applicable principal qualifiers (RSGs+3PQ). Besides the overall purity, evaluating the predicted most probable soil class, we could had assessed the entropy as an additional measure of accuracy by evaluating the relative probabilities of the most and less probable soil classes. We considered this too much detail and not necessarily more informative, not in the least because less probable soils are, by definition, minor soils. The overall purity is basically a measure of the accuracy of the model predictions at independent OOB cross validation points. This cross validation is an integral part of the random forest modelling procedure which we consider a major advantage of the procedure. Alternatively, additional survey would be needed for soil map validation, with a probability sampling design providing unbiased estimates of the accuracy expressed as the areal fraction of the map

[3] www.isric.org/explore/wosis

correctly predicted (Brus et al., 2011), however, additional survey is time and cost intense. Soil map production using the random forest procedure is also very time and cost effective compared to alternative more conventional approaches, such as the Soil and Terrain (SoTer) approach, which requires manual or automated (e.g. Dobos et al., 2005) delineation of landscape features at the appropriate scale and manual attribution of soil classes. Summarising, overall purities for the RSG maps are very reasonable and in line with purities that are typically reported for soil class maps developed with either statistical or conventional methods, while the latter are substantially more time- and cost intense. Some 4290 to 17,160 profile point observations would have been required to map the 30 woredas at a targeted scale of 1: 250,000 using a conventional approach with 1 to 4 observations per cm^2 on the paper map (11,920 to 47,680 profile observations at the apparent scale of 1: 150,000) whereas we used only 2595 profile observations.

Evidently, the accuracy of the map depends much on the adequacy of the survey data and covariates. As an example, we used a legacy soil-landscape map, disaggregated into spatially explicit soil-landscape facets, as covariate in the random forest modelling procedure and this covariate contributed very much to the accuracy of the final map. This result highlights the importance of including readily available information on soils and landscapes in the covariates which coincides with the experience from other digital soil mapping efforts that conventional legacy soil maps, of adequate quality, prove to best represent spatial variability of soil classes and properties compared to other covariates (Hengl et al., 2014; Kempen et al., 2019). In fact, and more generally speaking, legacy soil maps can be effectively updated and further improved using DSM techniques possibly enhancing scale, precision and accuracy, as illustrated by e.g. Kempen et al. (2009; 2015), Zeraatpisheh et al. (2019), Moller et al. (2019), and updating of the classification to new standards comparable to the approach applied in this study. The importance of adequate survey data is illustrated by the example wherein we assessed the extrapolative capacity of a random forest model trained with data collected during the explorative survey at kebele level (947 profiles) to predict RSGs at woreda level. The prediction accuracy at kebele level was 0.58 but extrapolated to the woreda level, as validated by observations made during the full survey beyond the kebele level, reduced to only 0.44. The survey data at kebele level were apparently inadequate to inform random forest predictions at woreda level. It is therefore recommended to conduct the exploratory survey directly at the targeted level (scale and extent) and to use a legacy soil-landscape map, disaggregated into spatially explicit soil-landscape facets, as a base map from the start. For preparing the base map we recommend not to use random forest to replace legacy soil units by soil units classified according to WRB, like we did, and rather maintain the legacy soil units. Map accuracy depends not only on the modelling approach, the covariates, the base map and the quantity and distribution of the survey data but also on the quality and reliability of the survey data (the classifications) itself. The latter is the most critical and possibly weakest point in the entire workflow and requires expert knowledge and ample experience. Reference Soil Groups, and prefix and suffix qualifiers, may be assessed from diagnostic features, largely reflecting soil forming processes, interpreted and diagnosed on the basis of detailed field observations complemented with the necessary soil analytical data. Alternatively, it is difficult to classify soils correctly from auger observations which were meant to determine or confirm in-field variability of soil characteristics possibly similar or comparable to characteristics yet observed in detail from yet classified soils. For instance, it is not obvious to distinguish Vertisols (Chromic) from Vertic Cambisols or Vertic Luvisols and Nitic Luvisols from Luvic Nitisols as all these soils may be reddish-brown to brownish-red in colour, deep, clayey and sticky throughout. This difficulty combined with the fact that the majority of observations was made by auger, distinguishing only 1 prefix qualifier, certainly had an important impact on the map accuracy. This may also explain why the inclusion of prefix qualifiers, even though enhancing precision, resulted in larger errors and hence lower map accuracy. The accuracy would likely be enhanced by higher level grouping of soil observations with comparable field characteristics but this would negatively impact precision. In-field classification according to an easier classification system may result in more reliable data (classes) which would likely contribute to a higher accuracy. The majority of auger observations was also classified according to local (farmers) nomenclatures. When done in consultation with a local land user, this may prove a reliable entry to in-field soil classification. Local soil names, plus possible within-name variability identified from augering till greater depth, may serve as a relative easy basis for surveying, and mapping, of soil resources whereby the local soil names can be characterised in detail and classified according to WRB. Another advantage of such approach would be the enhanced communicability towards farmers and extension workers without jeopardising the relevance and communication with the scientific community. We assessed, but not reported, the correlation between RSGs and local soil names which proved incomplete and not straightforward.

A soil class represents a collection of soil properties that are interrelated with each other as the result of similar soil formation beyond the topsoil translating into a collection of soil qualities. These qualities define the functions that a soil can provide, e.g. for supporting agricultural productivity or ecosystem services, and the management practices required to enhance those functions. Soil class maps represent the spatial variability of these collections at the scale of operation. This is illustrated by a number of soil property maps representing the mean values associated with each RSG. The accuracy of the soil property maps seems reasonable but is different for different properties. The inaccuracy, here expressed by the relative RMSE (relative to the mean), is on average 42% and varies between 9% for bulk density and 166% for extractable zinc. Different RSGs prove not very different in terms of the selected soil properties except for specific properties that are typical or diagnostic for specific RSGs. For example, soil nutrient parameters may be similar for different RSGs but other soil properties such as drainage, depth, sum of exchangeable bases or pH are different for different RSGs and show relative good correlations between observed and mapped property values. These differences determine differences in aeration, rootability, acidity and fertility and therewith differences in crop response to fertiliser nutrients. Trivial but key for producing soil property maps from soil class maps is that a representative soil profile is described in detail and sampled for laboratory analysis for each of the identified soil classes (RSGs). Evidently, soil properties can also be mapped directly using DSM techniques, instead of using the soil classes as carriers, but such approach requires sufficient point data on the properties of interest, including analytical point data. In this study however we collected only a small number of soil samples and instead a large number of soil classifications, from auger observations, which allowed us to map soil properties at limited laboratory costs.

Accurate soil-landscape resources maps can be produced efficiently for additional woredas in Ethiopia. It is recommended to apply a similar approach for future surveys under similar circumstances using the following steps: 1) prepare a base map for the woreda which indicates hypothetical soil classes according to landscape facets disaggregated from the Geomorphology and Soils map of Ethiopia. Distinguish and map landscape facets based on slope class and relative elevation (low-mid-high) within 5 km distance, 2) define the minimum number of observations required for the targeted scale and prepare a survey plan following the base map, 3) conduct an exploratory survey following the base map throughout the woreda, thus not starting at the kebele level, to identify major soil classes, 4) characterise and classify soil profiles, from detailed soil pit descriptions, which are representative for the identified soil classes and define diagnostic characteristics which are recognisable from an auger, 5) continue with the survey throughout the woreda following the same base map to verify the distribution of the identified soil classes, explicitly linking the classifications of the observation points with the classifications of the soil pits and where needed making additional soil pit observations for additionally identified soil classes, 6) confirm and where needed correct the field-assessed classifications using soil analytical results first for soil pits and then the related auger observations, 7) compile the soil data strictly according to the standards of the targeted database template, 8) prepare spatial covariates representative for soil forming factors including management, 9) model and independently validate soil-landscape relations using DSM (random forest) with georeferenced soil classifications as input and the base map as one of the covariates. Consider producing maps of the diagnostic qualifiers to add them to the covariate stack for predicting RSGs, 10) apply the validated model to produce a map at high resolution and, if needed, generalise the map into a polygon map following geomorphic landscape features, 11) map soil properties according to the soil-landscape resources map. Additional survey can be conducted for additional map validation if considered necessary and time and budget permitting.

Acknowledgements

This study was conducted under the CASCAPE project funded by the Directorate-General for International Cooperation (DGIS) of the Netherlands Ministry of Foreign Affairs through the Netherlands Embassy in Ethiopia.

Special thanks are due to the consultants from the participating Ethiopian universities for carrying out most of the field work and reporting; Kibebew Kibret of Haramaya University, Mekonnen Getahun of Bahir Dar University, Engdawork Assefa of Addis Ababa University, Amanuel Zenebe et al. of Mekele University, Alemayehu Kiflu of Hawassa University and Alemayehu Regassa of Jimma University. An important contribution was made by Ashenafi Ali who synthesised the soil data at national level. Koos Dijkshoorn provided guidance as a consultant during the field work in two regions together with Arie van Kekem and the lead author guiding the field work each in two other regions. Much thanks are due to Arie van Kekem who initiated the CASCAPE project and this very study.

References

Abbate E., P. Bruni, and M. Sagri, 2015. Geology of Ethiopia: a review and geomorphological perspectives. In: Billi (ed.), 2015. Landscapes and Landforms of Ethiopia, World Geomorphological Landscapes: 117-136. Springer, Dordrecht.

Abebe M., 2007. Nature and management of acid soils in Ethiopia. Haramaya University, Dire Dawa. 82 p.

AfSIS, 2015. Africa Soil Information Service project. Covariates for land and climate developed from remotely sensed data. Earth institute, Columbia University, New York. Available URL: http://africasoils.net/services/data/remote-sensing; ftp://africagrids.net (accessed on: 2015).

Alemayehu Kiflu, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas in Southern Nations Nationalities People Regional State. CASCAPE report. Hawassa University, Ethiopia. 511 p. (library.wur.nl/WebQuery/isric/2259398).

Alemayehu Regassa, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas in Western Oromia. CASCAPE report. Jemma University, Ethiopia. 117 p. (library.wur.nl/WebQuery/isric/2259397).

Amanuel Zenebe, Girmay Gebresamuel, and Atkilt Girma, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas in Southern Tigray, Ethiopia. CASCAPE report. Mekelle University, Ethiopia. 284 p. (library.wur.nl/WebQuery/isric/2259410).

Batjes N.H., E. Ribeiro, A. van Oostrum, J.G.B. Leenaars, T. Hengl, and J. Mendes de Jesus, 2017. WoSIS: providing standardised soil profile data for the world. Earth Syst. Sci. Data 9: 1-14. (doi.org/10.5194/essd-9-1-2017).

Berry L., 2003. Land degradation in Ethiopia: its extent and impact. In: L. Berry, J. Olson, and D. Campbell, 2003. Assessing the extent, cost and impact of land degradation at the national level: findings and lessons learned from seven pilot case studies. Technical report. Commissioned by Global Mechanism with support from the World Bank.

Billi P., 2015. Geomorphological landscapes of Ethiopia. In: Billi (ed.), 2015. Landscapes and Landforms of Ethiopia, World Geomorphological Landscapes: 3-32. Springer, Dordrecht.

Bishop T.F.A., A.B. McBratney, and G.M. Laslett, 1999. Modelling soil attribute depth functions with equal-area quadratic smoothing splines. Geoderma 91 (1-2): 27-45. Elsevier.

Blum W.E.H., P. Schad, and S. Nortcliff, 2018. Essentials of soil science. Soil formation, functions, use and classification (World Reference Base, WRB). Borntraeger Science Publishers, Stuttgart. 171 p.

Breiman L., 2001. Random forest. Machine Learning 45 (1): 5-32. Kluwer. (doi.org/10.1023/A:1010933404324).

Brevik E.C., C. Calzolari, B.A. Miller, P. Pereira, C. Kabala, A. Baumgarten, and A. Jordán, 2016. Soil mapping, classification, and pedologic modeling: History and future directions. Geoderma 264(B): 256-274. Elsevier.

Brungard C.W., J.L. Boettinger, M.C. Duniway, S.A. Wills, and T.C. Edwards Jr, 2015. Machine learning for predicting soil classes in three semi-arid landscapes. Geoderma 239–240(0): 68-83. Elsevier.

Brus D.J., B. Kempen, and G.B.M. Heuvelink, 2011. Sampling for validation of digital soil maps. European Journal of Soil Science 62(3): 394-407.

Dobos, E., J. Daroussin, and L. Montanarella. 2005. An SRTM-based procedure to delineate SOTER Terrain Units on 1:1 and 1:5 million scales. EUR 21571 EN. Office for Official Publications of the European Communities, Luxembourg. 55 p.

Driessen P.M., J. Deckers, O. Spaargaren, and F. Nachtergaele (eds.), 2001. Lecture notes on the major soils of the world. World soil resources reports No. 94. FAO, Rome. (www.fao.org/3/a-y1899e.pdf).

Driessen P.M., and R. Dudal (eds.), 1989. Lecture notes on the, geography, formation, properties and use of the, major soils of the world. Agricultural University Wageningen, KU Leuven. 296 p.

Engdawork Assefa, 2015. Characterisation and classification of the major agricultural soils in CASCAPE intervention woredas in the central highlands of Oromia region, Ethiopia. CASCAPE report. Addis Ababa University, Ethiopia. (library.wur.nl/WebQuery/isric/2259411).

Elias E., P.F. Okoth, and E.M.A. Smaling, 2019. Explaining bread wheat (Triticum aestivum) yield differences by soil properties and fertilizer rates in the highlands of Ethiopia. Geoderma 339: 126-133. Elsevier.

Elias E., 2016. Soils of the Ethiopian Highlands: Geomorphology and Properties. CASCAPE Project, ALTERRA, Wageningen UR, the Netherlands. (library.wur.nl/WebQuery/isric/2259099).

FAO, 2006. Guidelines for soil description. FAO, Rome.

FAO, 1998. The Soil and Terrain Database for north-eastern Africa. Crop production systems zones of the IGAD sub region. Land and water digital media series 2. FAO, Rome.

FAO, 1988. FAO/Unesco Soil Map of the World, revised legend. World Resources Report 60. FAO, Rome. Reprinted, with corrections, as Technical Paper 20. ISRIC, Wageningen, 1989. (library.wur.nl/WebQuery/isric/2264662).

FAO, 1984. Assistance to land-use planning, Ethiopia. Geomorphology and Soils. Based on the work of Henricksen B.L., S. Ross, S. Tilimo and H.Y. Wijntje-Bruggeman. UNDP, FAO. Addis Ababa. (library.wur.nl/WebQuery/isric/2257365). Monograph & 8 map sheets.

FAO, 1974. FAO/Unesco Soil Map of the World, 1: 5,000,000. Vol. I: legend. Unesco, Paris. 65 p. (library.wur.nl/WebQuery/isric/2258200).

Fazzini M., C. Bisci, and P. Billi, 2015. The Climate of Ethiopia. In: P. Billi (ed.), 2015. Landscapes and Landforms of Ethiopia, World Geomorphological Landscapes: 65-87. Springer, Dordrecht. (doi.org/10.1007/978-94-017-8026-1_3).

Grimm R., T. Behrens, M. Märker, and H. Elsenbeer, 2008. Soil organic carbon concentrations and stocks on Barro Colorado Island - Digital soil mapping using Random forest analysis. Geoderma 146 (1-2): 102-113. Elsevier.

Grinand C., D. Arrouays, B. Laroche, and M.P. Martin, 2008. Extrapolating regional soil landscapes from an existing soil map: Sampling intensity, validation procedures, and integration of spatial context. Geoderma 143: 180–190. Elsevier.

GYGA, 2018. Global Yield Gap and Water Productivity Atlas. Available URL: www.yieldgap.org (accessed on: December 30th, 2018).

Hastie T., R. Tibshirani, and J. Friedman, 2008. The Elements of Statistical Learning. Data Mining, inference and prediction. Second edition. Springer, New York.

Hengl T., J.G.B. Leenaars, K.D. Shepherd, M.G. Walsh, G.B.M. Heuvelink, et al., 2017. Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. Nutrient Cycling in Agroecosystems 109: 77-102. Springer, Netherlands. (doi.org/10.1007/s10705-017-9870-x).

Hengl T., J. Mendes de Jesus, G.B.M. Heuvelink, M. Ruiperez Gonzalez, .., J.G.B. Leenaars, et al., 2017b. SoilGrids250m: Global gridded soil information based on machine learning. PLoS ONE 12(2). (doi.org/10.1371/journal.pone.0169748).

Hengl T., G.B.M. Heuvelink, B. Kempen, J.G.B. Leenaars, M.G. Walsh, K.D. Shepherd, A. Sila, R.A. MacMillan, J. Mendes de Jesus, L. Tamene, and J.E. Tondoh, 2015. Mapping Soil Properties of Africa at 250 m Resolution: Random forest Significantly Improve Current Predictions. PLoS ONE 10(6). (doi.org/10.1371/journal.pone.0125814).

Heung B., H.C. Ho, J. Zhang, A. Knudby, C.E. Bulmer, M.G. Schmidt, 2016. An overview and comparison of machine-learning techniques for classification purposes in digital soil mapping. Geoderma 265: 62-77. Elsevier.

Heung B., C.E. Bulmer, and M.G. Schmidt, 2014. Predictive soil parent material mapping at a regional-scale: A Random forest approach. Geoderma 214–215: 141-154. Elsevier.

Hijmans, R.J., S.E. Cameron, J.L. Parra, P.G. Jones and A. Jarvis, 2005. Very high resolution interpolated climate surfaces for global land areas. International Journal of Climatology 25: 1965-1978. (doi.org/10.1002/joc.1276).

Holmes, K.W., E.A. Griffi, and N.P. Odgers, 2015. Large-area spatial disaggregation of a mosaic of conventional soil maps: evaluation over Western Australia. Soil Research 53(8): 865-880.

IUSS Working Group WRB, 2015. World reference base for soil resources 2014, update 2015. International soil classification system for naming soils and creating legends for soil maps. World soil resources reports No. 106. FAO, Rome.

IUSS Working Group WRB, 2007. World reference base for soil resources 2006, first update 2007. A framework for international classification, correlation and communication. World soil resources reports No. 103. FAO, Rome.

Jones A., H. Breuning-Madsen, M. Brossard, A. Dampha, J. Deckers, O. Dewitte, T. Gallali, S. Hallett, R. Jones, M. Kilasara, P. Le Roux, E. Michéli, L. Montanarella, O. Spaargaren, L. Thiombiano, E. Van Ranst, M. Yemefack, and R. Zougmore (eds.), 2013. Soil atlas of Africa. European Commission, Publications Office of the European Union, Luxembourg. (doi.org/10.2788/52319).

Kempen B., S. Dalsgaard, A.K. Kaaya, N. Chamuya, M. Ruipérez-González, A. Pekkarinen, and M.G. Walsh, 2019. Mapping topsoil organic carbon concentrations and stocks for Tanzania. Geoderma 337: 164-180. Elsevier.

Kempen B., D.J. Brus, and F. de Vries, 2015. Operationalizing digital soil mapping for nationwide updating of the 1:50,000 soil map of the Netherlands. Geoderma 241-242: 313-329. Elsevier.

Kempen B., D.J. Brus, J.J. Stoorvogel, G.B.M. Heuvelink, and F. de Vries, 2012. Efficiency Comparison of Conventional and Digital Soil Mapping for Updating Soil Maps. SSSAJ 76 (6): 2097-2115. (doi.org/10.2136/sssaj2011.0424).

Kempen B., D.J. Brus, and G.B.M. Heuvelink, 2012b. Soil type mapping using the generalised linear geostatistical model: A case study in a Dutch cultivated peatland. Geoderma 189-190: 540-553. Elsevier.

Kempen B., D.J. Brus, G.B.M. Heuvelink, and J.J. Stoorvogel, 2009. Updating the 1:50,000 Dutch soil map using legacy soil data: A multinomial logistic regression approach. Geoderma 151(3-4): 311-326. Elsevier.

Kibebew Kibret, 2014. Characterisation of agricultural soils in CASCAPE intervention woredas in Eastern region. CASCAPE report. Haramaya University, Ethiopia. 263 p. (library.wur.nl/WebQuery/isric/2259414).

Leenaars J.G.B., L. Claessens, G.B.M. Heuvelink, T. Hengl, M. Ruiperez González, L.G.J. van Bussel, N. Guilpart, H. Yang, and K.G. Cassman, 2018. Mapping rootable depth and root zone plant-available water holding capacity of the soil of sub-Saharan Africa. Geoderma 324: 18-36. Elsevier.

Leenaars J.G.B., M. Ruiperez González, and B. Kempen, 2018b. Extrapolation of fertilizer nutrient recommendations for major food crops in West Africa; a proof of concept (with dataset). Project report (draft) for IFDC, USAID - West Africa Fertilizer Program, Accra. ISRIC - World Soil Information, Wageningen. Available URL: www.isric.org/projects/taking-fertilizer-recommendations-scale-major-crops-west-africa.

Leenaars J.G.B., E. Elias, H. Wösten, M. Ruiperez González, B. Kempen, A. Ashenafi, and F. Brouwer, 2016. Major soil-landscape resources of the CASCAPE intervention woredas, Ethiopia. Soil information in support to scaling up of evidence-based best practices in agricultural production (with dataset). CASCAPE Project, ALTERRA, Wageningen UR, the Netherlands. (library.wur.nl/WebQuery/isric/2259349).

Leenaars J.G.B., A.J.M. van Oostrum, and M. Ruiperez Gonzalez, 2014. Africa Soil Profiles Database, Version 1.2. A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa (with dataset). ISRIC Report 2014/01. Africa Soil Information Service (AfSIS) project and ISRIC - World Soil Information, Wageningen. (library.wur.nl/WebQuery/isric/2259472).

Leenaars J.G.B., B. Kempen, A.J.M. van Oostrum, and N.H. Batjes, 2014b. Africa Soil Profiles Database: A compilation of georeferenced and standardised legacy soil profile data for Sub-Saharan Africa. In: Arrouays et al. (eds.), 2014. GlobalSoilMap: 51-57. Taylor & Francis group, London.

Mekonnen Getahun, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas of Amhara region. CASCAPE report. Bahir Dar University, Ethiopia. 89 p. (library.wur.nl/WebQuery/isric/2259413).

Minasny B., and A.B. McBratney, 2016. Digital soil mapping: A brief history and some lessons. Geoderma 264(B): 301-311. Elsevier.

Møller A.B., B. Malone, N.P. Odgers, A. Beucher, B.V. Iversen, M.H. Greve, and B. Minasny, 2019. Improved disaggregation of conventional soil maps. Geoderma 341: 148-160. Elsevier.

Mulders M., J.G.B. Leenaars, and A. Belemviré, 2001. Soil resources in Sahelian villages. In: L. Stroosnijder and T. van Rheenen (eds.). Agro-silvo-pastoral land use in Sahelian villages. Advances in geo-ecology 33: 101-130. IUSS, Catena Verlag.

Murphy H.F., 1959. A report on the fertility status of some soils of Ethiopia. Imperial Ethiopian college of Agriculture and mechanical arts. Addis Ababa, Ethiopia. (library.wur.nl/WebQuery/isric/2265435).

Sanginga N., and P.L. Woomer (eds.), 2009. Integrated soil fertility management in Africa: principles, practices and development process. TSBF-CIAT, Nairobi. 263 p.

Smaling E.M.A., and B.H. Janssen, 1993. Calibration of QUEFTS: a model predicting nutrient uptake and yields from chemical soil fertility indices. Geoderma 59: 21-44. Elsevier.

Strobl C., J. Malley, and G. Tutz, 2009. An Introduction to Recursive Partitioning: Rationale, Application, and Characteristics of Classification and Regression Trees, Bagging, and Random forest. Psychological Methods 14(4): 323-348.

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

Vågen T.G., 2010. Africa Soil Information Service: Hydrologically Corrected/Adjusted SRTM DEM (AfrHySRTM). International Center for Tropical Agriculture - Tropical Soil Biology and Fertility Institute (CIAT-TSBF), World Agroforestry Centre (ICRAF), Center for International Earth Science Information Network (CIESIN), Columbia University. Available URL: https://cmr.earthdata.nasa.gov/search/concepts/C1214155420-SCIOPS

van Diepen C.A., J. Wolf, H. van Keulen, and C. Rappoldt, 1989. WOFOST: a simulation model of crop production. Soil Use Management 5(1): 16–24.

van Engelen V.W.P., and T.T. Wen (eds.), 1993. Global and National Soils and Terrain Digital Databases (SOTER): Procedures Manual. UNEP-ISSS-ISRIC-FAO. ISRIC - World Soil Information, Wageningen. (Also published as World Soil Resources Report 74, Rev.1).

van Reeuwijk L.P. (ed.), 2002. Procedures for Soil Analysis. Technical paper 9, sixth edition. ISRIC - World Soil Information, Wageningen, the Netherlands.

Vanlauwe B., A. Bationo, J. Chianu, K.E. Giller, R. Merckx, U. Mokwunye, P. Pypers, K. Shepherd, E. Smaling, P.L. Woomer, and N. Sanginga, 2010. Integrated soil fertility management. Operational definition and consequences for implementation and dissemination. 19th World Congress of Soil Science, Brisbane, Australia. Outlook Agriculture 39(1): 17-24.

World Bank, 2018. Databank. World development indicators. Available URL: data.worldbank.org (accessed on: January 2nd, 2019).

Wösten J.H.M., J. Bouma, and G.H. Stoffelsen, 1985. Use of soil survey data for regional soil water simulation models. Soil Sci. Soc. Am. J. 49: 1238-1244.

Zeraatpisheh M., S. Ayoubi, C.W. Brungard, and P. Finke, 2019. Disaggregating and updating a legacy soil map using DSMART, fuzzy *c*-means and *k*-means clustering algorithms in Central Iran. Geoderma 340: 249-258. Elsevier.

List of figures

Figure 1

Contrasting crop performance on two soils with similar topsoil (fertility) and contrasting depths (water): a) deep black soil, with on the background b) shallow black soil.

Figure 2

Geomorphology and Soils map of Ethiopia (FAO, 1984; in thin black polygons) projected upon elevation (colours) serving as base map to guide the exploratory survey in 30 Ethiopian Highland woredas (bold black polygons).

Figure 3

Example of the disaggregation of the Geomorphology and Soils map of Ethiopia (FAO, 1984) according to the landscape facets and soil units (FAO, 1974) defined for a soil association (Rp^{4}_{v}) in a given landscape unit (Rp: undulating to rolling residual plateaus).

Figure 4

Illustration of soil profile point locations in 4 out of the 30 woredas (black polygons), including 4 selected kebeles per woreda (visualised by a yellow colour). Black dots: soil pits of exploratory survey, open dots: augers of exploratory survey, blue dots: augers of full survey, red dots: virtual profiles.

Figure 5

Random forest covariate importance (permutation accuracy importance) for modelling RSGs.

Figure 6

Polygon map of the major soil-landscape resources in Omonada woreda at the level of RSGs (RSGs indicated by colour, major landscapes by bold black polygons and landscape facets by thin black polygons). Rgv: Major river gorges, canyons and escarpments including very steep side slopes and plateau terraces, Rh2v: high to mountainous relief hills, Rm2v: moderate to high relief hills, Rp4v: undulating to rolling plateaus.

Figure 7 a, b, c, d (left up, right up, left low, right low)

Soil property maps for a) drainage class, b) pH-H_2O, c) extractable Zn, d) available P. Colour range (red-orange-yellow-green-light blue-dark blue) indicates low to high property values.

Figure 1 a & b

Figure 2

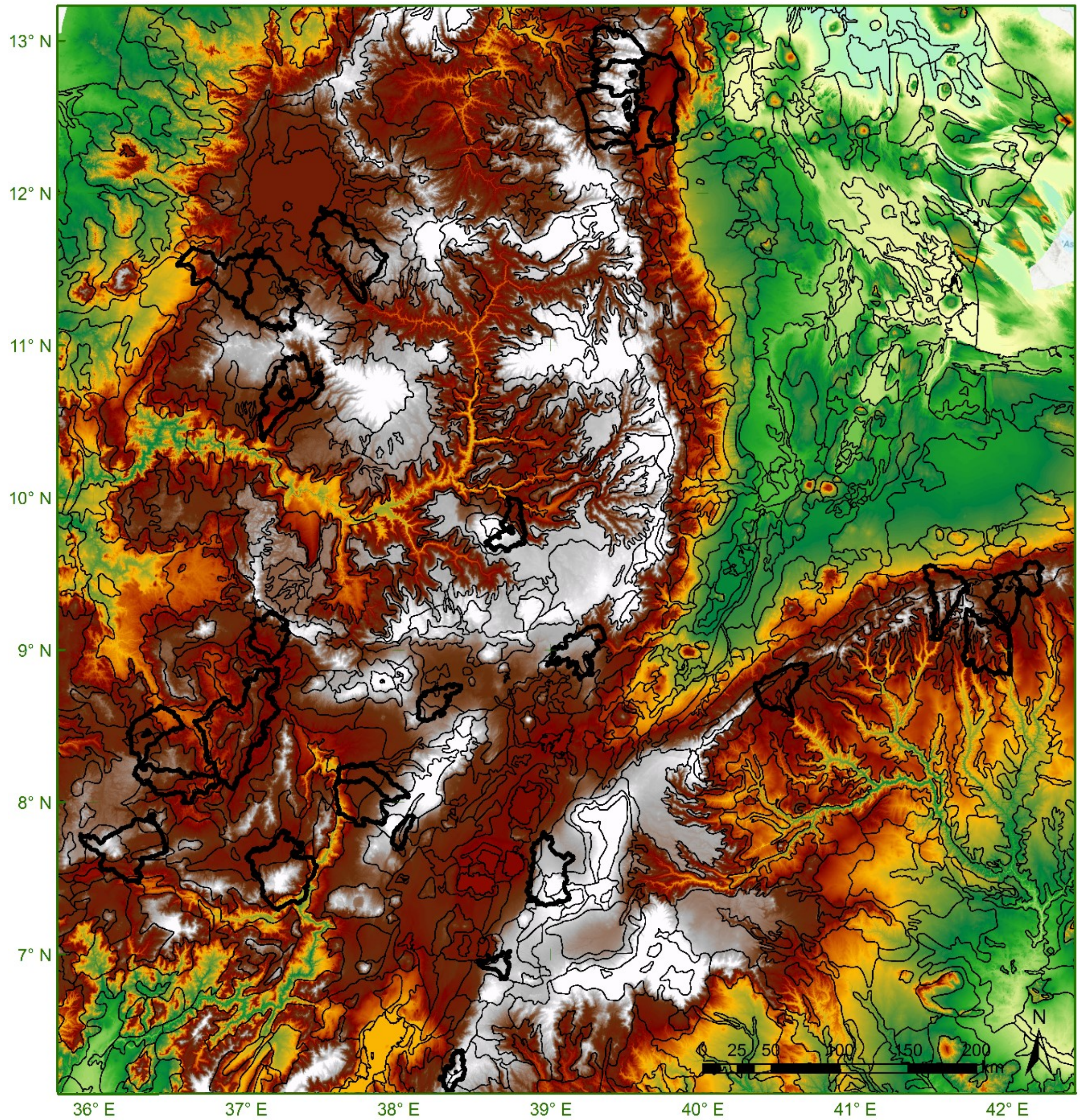
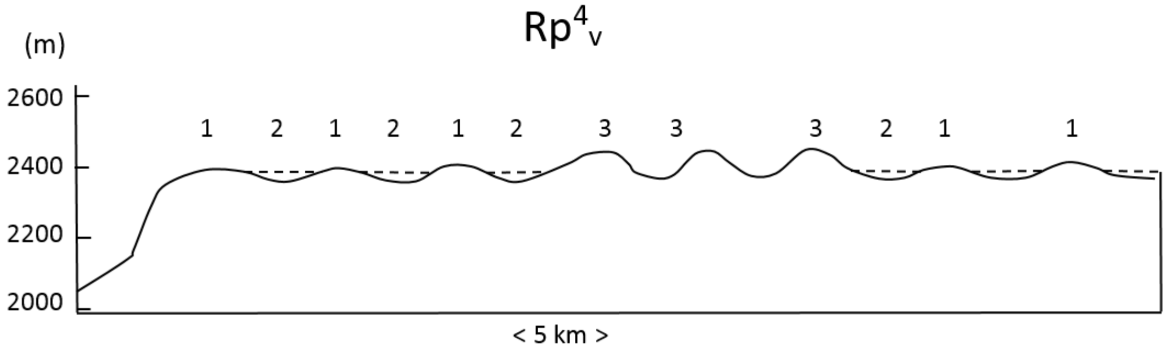

13° N
12° N
11° N
10° N
9° N
8° N
7° N
36° E
37° E
38° E
39° E
40° E
41° E
42° E
0 25 50 100 150 200
Km
N

Figure 3

Rp^4_v



REPORTED SOIL ASSOCIATION				
Number	1	2	3	-
Facet	undulating plateaux	depressions deficient drainage	rolling plateaux	moderately steep hills
Area (%)	40	40	20	-
Slope (%)	2 - 8	0 - 2	8 - 16	-
Soils Rp^4_v	Chromic Vertisols	Eutric Fluvisols	Dystric Nitosols	-
DISAGGREGATED SOIL ASSOCIATION				
Slope (%) & rel. elevation	2 - 5 5 - 10 mid	0 - 2	10 - 15 5 - 10 high	15 - 30
Soils Rp^4_v	Chromic Vertisols	Eutric Fluvisols	Dystric Nitosols	Dystric Cambisols

Figure 4

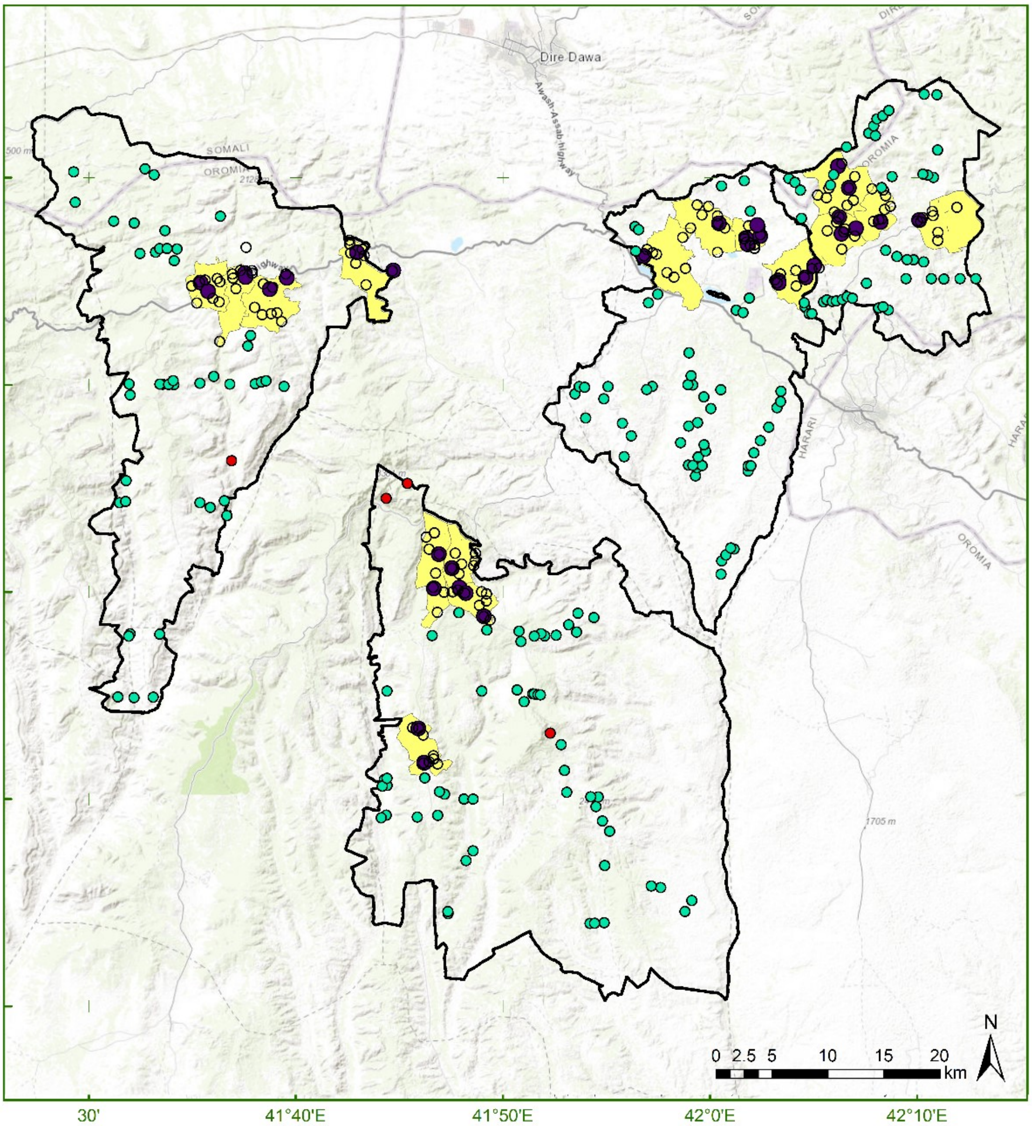

Dire Dawa
SOMALI
OROMIA
HARARI
OROMIA
30'
9°20'N
9°10'N
9°0'N
8°50'N
30'
41°40'E
41°50'E
42°0'E
42°10'E
0 2.5 5 10 15 20
km
N

Figure 5

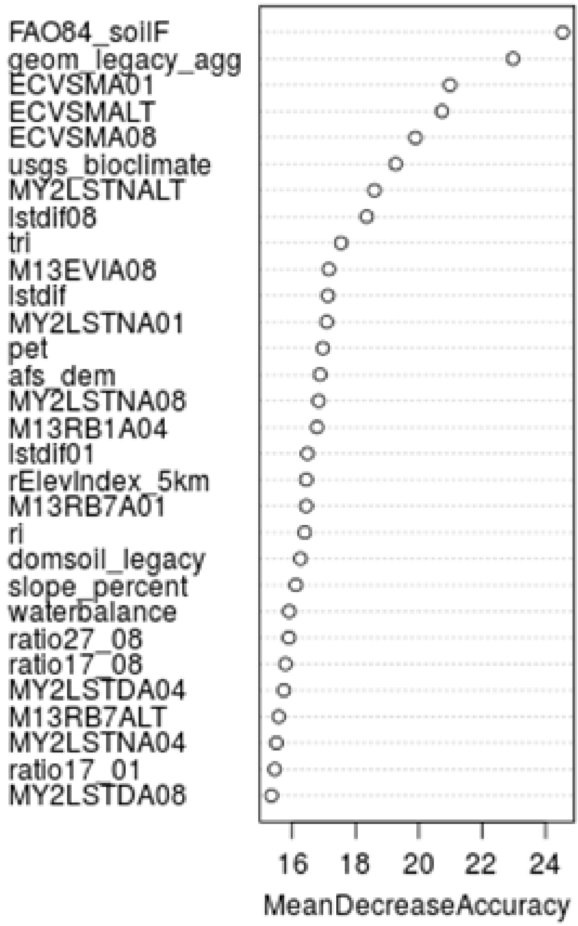
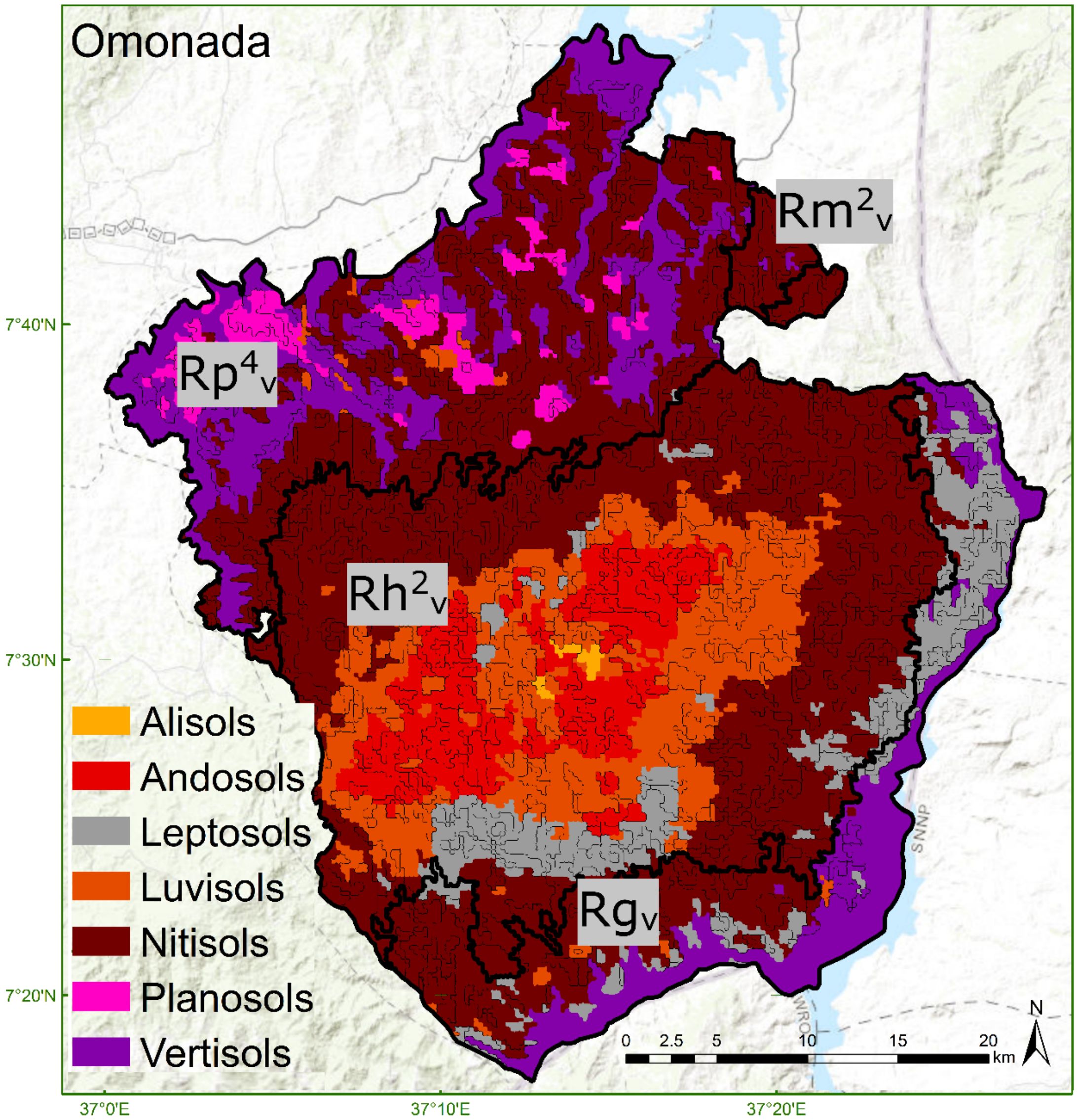
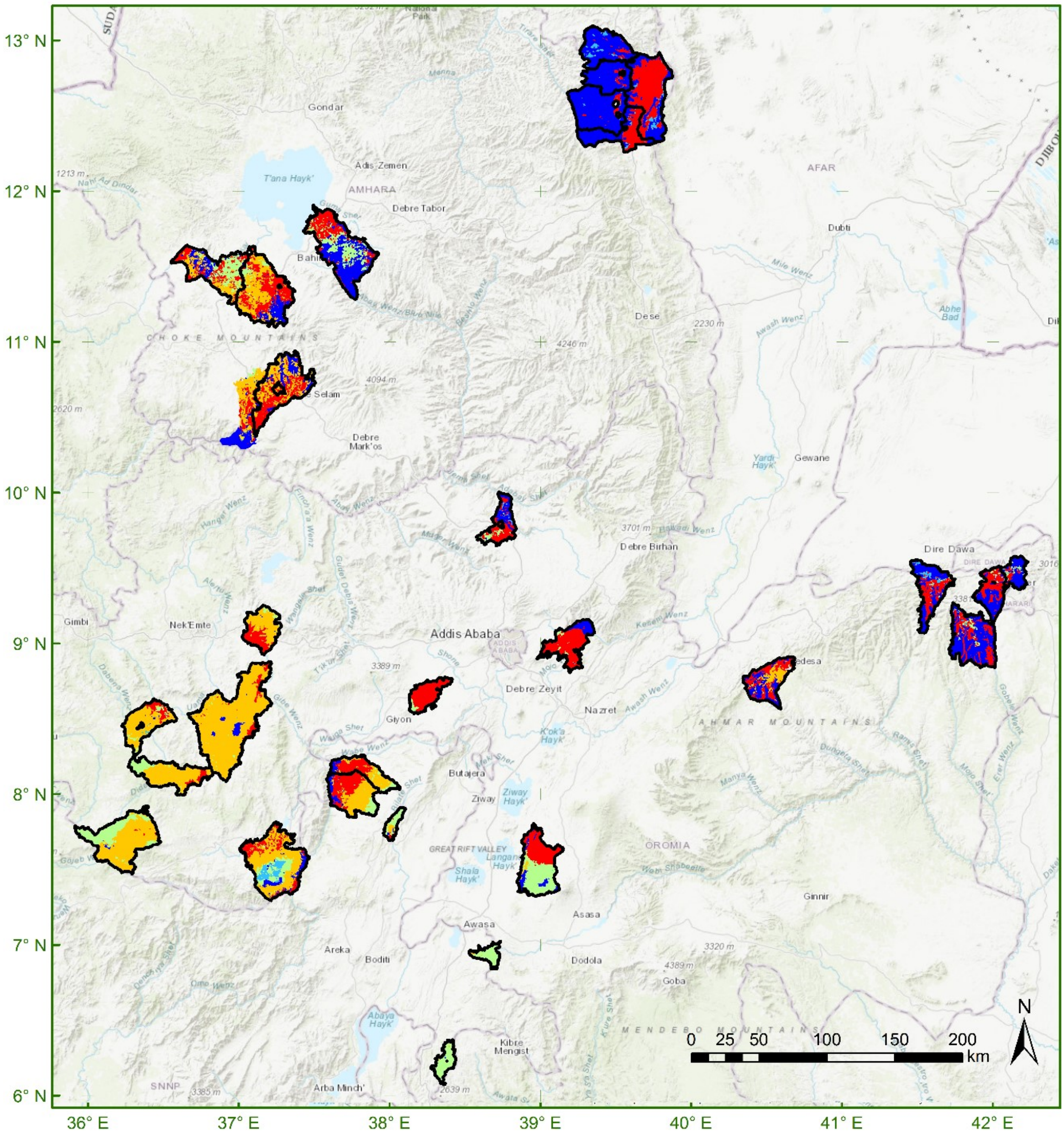

FAO84_soilF
geom_legacy_agg
ECVSMA01
ECVSMALT
ECVSMA08
usgs_bioclimate
MY2LSTNALT
lstdif08
tri
M13EVIA08
lstdif
MY2LSTNA01
pet
afs_dem
MY2LSTNA08
M13RB1A04
lstdif01
rElevIndex_5km
M13RB7A01
ri
domsoil_legacy
slope_percent
waterbalance
ratio27_08
ratio17_08
MY2LSTDA04
M13RB7ALT
MY2LSTNA04
ratio17_01
MY2LSTDA08
16
18
20
22
24
MeanDecreaseAccuracy

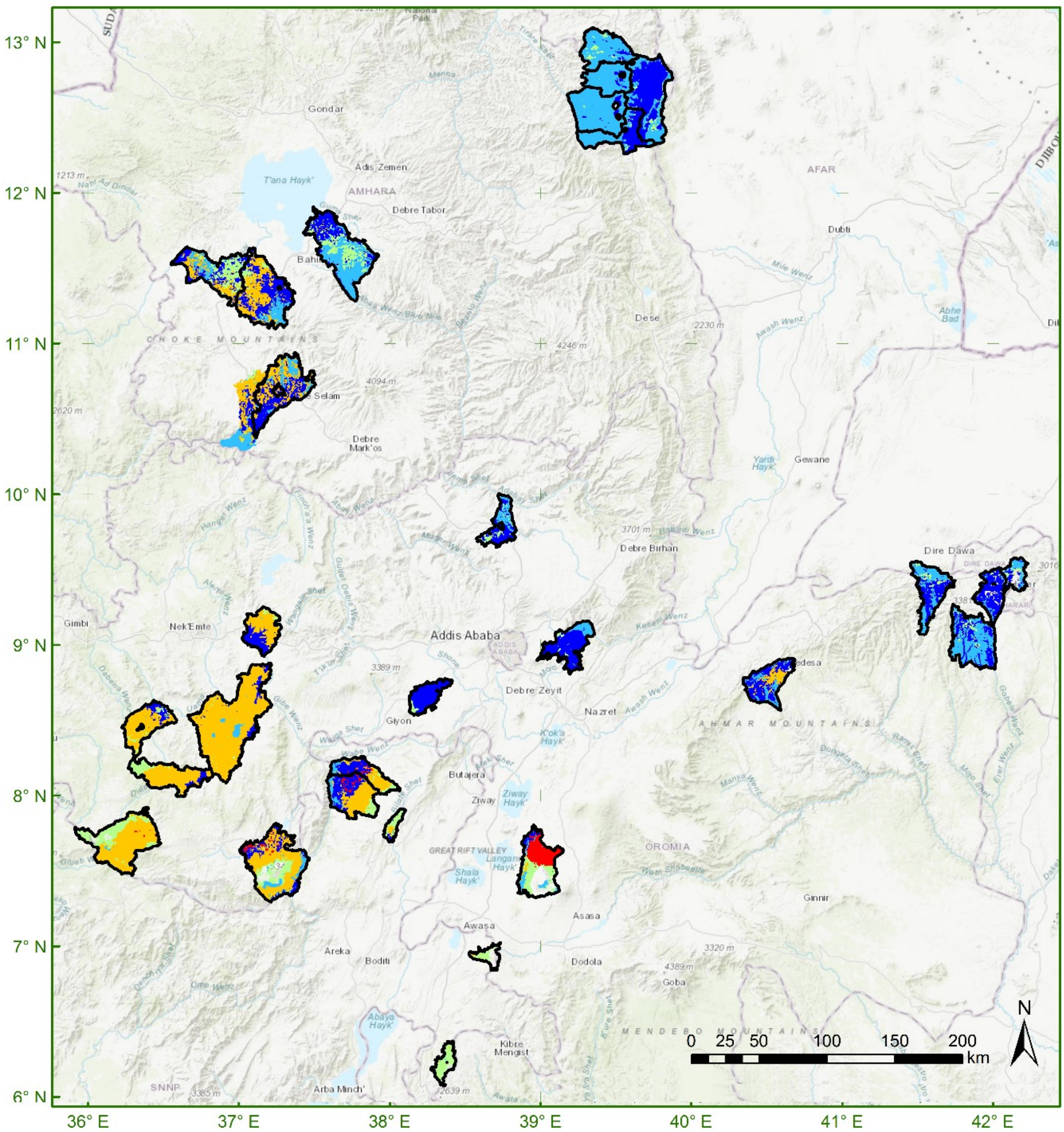
Figure 6

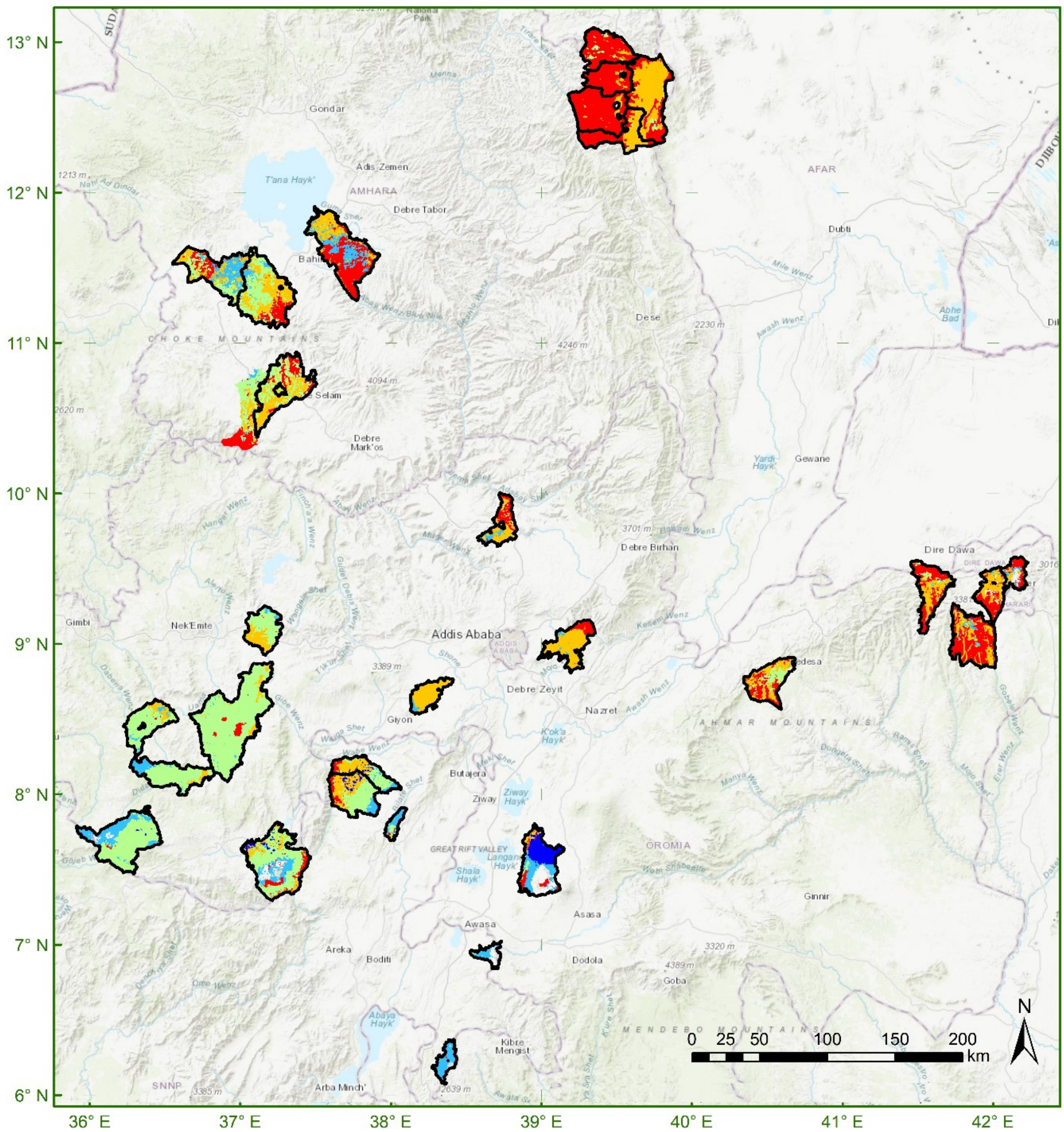


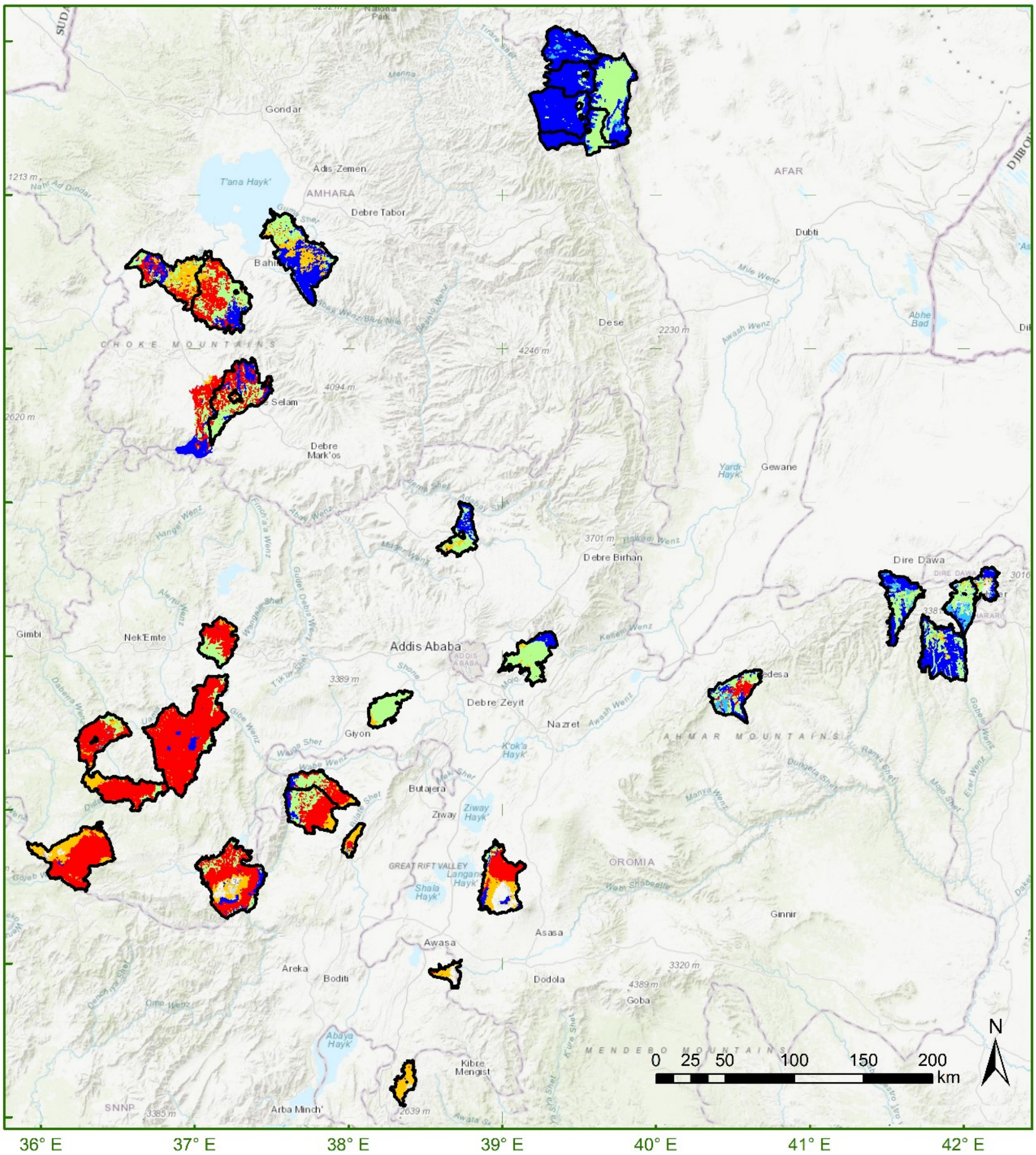
Omonada
Rm2v
Rp4v
Rh2v
Rgv
7°40'N
7°30'N
7°20'N
37°0'E
37°10'E
37°20'E
Alisols
Andosols
Leptosols
Luvisols
Nitisols
Planosols
Vertisols
0 2.5 5 10 15 20 km
N

Figure 7 a, b, c & d




13° N
12° N
11° N
10° N
9° N
8° N
7° N
6° N
36° E
37° E
38° E
39° E
40° E
41° E
42° E
Gondar
Adis Zemen
T'ana Hayk'
AMHARA
Debre Tabor
AFAR
Dubti
Dese
CHOKE MOUNTAINS
Debre Mark'os
Gewane
Debre Birhan
Dire Dawa
Addis Ababa
Gimbi
Nek'Emte
Debre Zeyit
Nazret
Giyon
AHMAR MOUNTAINS
Butajera
Ziway
GREAT RIFT VALLEY
OROMIA
Asasa
Ginnir
Awasa
Areka
Boditi
Dodola
Goba
MENDEBO MOUNTAINS
Kibre Mengist
Arba Minch'
SNNP
0 25 50 100 150 200
km
N


13° N
12° N
11° N
10° N
9° N
8° N
7° N
6° N
36° E
37° E
38° E
39° E
40° E
41° E
42° E
Gondar
Adis Zemen
T'ana Hayk'
AMHARA
Debre Tabor
AFAR
Dubti
CHOKE MOUNTAINS
Dese
Debre Mark'os
Gewane
Debre Birhan
Dire Dawa
Gimbi
Nek'Emte
Addis Ababa
Debre Zeyit
Nazret
Giyon
AHMAR MOUNTAINS
Butajera
Ziway
GREAT RIFT VALLEY
OROMIA
Asasa
Ginnir
Awasa
Areka
Boditi
Dodola
Goba
MENDEBO MOUNTAINS
Kibre Mengist
Arba Minch'
SNNP
0 25 50 100 150 200
km
N


13° N
12° N
11° N
10° N
9° N
8° N
7° N
6° N
36° E
37° E
38° E
39° E
40° E
41° E
42° E
Gondar
Adis Zemen
T'ana Hayk'
AMHARA
Debre Tabor
AFAR
Dubti
Dese
CHOKE MOUNTAINS
Debre Mark'os
Gewane
Debre Birhan
Dire Dawa
Gimbi
Nek'emte
Addis Ababa
Debre Zeyit
Nazret
Giyon
AHMAR MOUNTAINS
Butajera
Ziway
GREAT RIFT VALLEY
OROMIA
Ginnir
Awasa
Asasa
Areka
Boditi
Dodola
Goba
MENDEBO MOUNTAINS
Kibre Mengist
Arba Minch'
SNNP
0 25 50 100 150 200
km
N

List of tables

Table 1

Property	**RSG**	**Acrisols**	**Alisols**	**Andosols**	**Arenosols**	**Calcisols**	**Cambisols**	**Fluvisols**	**Gleysols**	**Leptosols**	**Luvisols**	**Nitisols**	**Phaeozems**	**Planosols**	**Regosols**	**Vertisols**	**All**
Soil profile																	
Depth	AVG	-	115	98	94	39	95	101	108	26	117	110	118	119	36	113	98
(cm)	SD	-	21	18	21	8	13	22	14	17	26	24	19	12	16	29	48
	n	-	*60*	*13*	*25*	-	*41*	*2*	-	*264*	*266*	*258*	*2*	*33*	*48*	*277*	*1289*
Drainage	AVG	-	4.9	5.0	6.9	-	5.1	5.2	3.3	6.4	4.9	4.7	3.3	3.0	6.6	3.1	4.9
(class)	SD	-	0.6	0.0	0.3	-	0.9	1.1	1.9	0.7	0.6	0.7	0.8	0.4	0.6	1.1	1.4
	n	-	*62*	*15*	*25*	-	*41*	*2*	-	*248*	*265*	*261*	*2*	*32*	*50*	*276*	*1279*
Soil depth interval		(0-30 cm)															
Clay	AVG	57	-	-	-	57	43	34	45	48	47	57	49	42	38	55	50
(g/100g)	SD	3	-	-	-	-	10	8	6	9	12	12	6	6	5	10	12
BD	AVG	1.04	-	-	-	1.32	1.26	1.24	1.08	1.21	1.13	1.15	1.23	1.04	1.29	1.22	1.17
(kg/dm3)	SD	0.01	-	-	-	-	0.13	0.18	0.06	0.08	0.10	0.09	0.10	0.02	0.04	0.12	0.12
pH-H_2O	AVG	4.3	-	-	-	7.8	6.4	7.2	4.8	6.7	5.6	5.4	6.6	5.0	7.2	6.8	6.0
(-)	SD	0.1	-	-	-	-	1.0	1.1	0.4	1.0	0.8	0.4	1.1	0.5	0.8	0.9	1.0
CEC	AVG	36.9	-	-	-	49.4	39.7	37.5	39.7	52.2	39.9	42.2	46.3	31.4	33.9	49.1	42.8
(cmolc/kg)	SD	2.4	-	-	-	-	6.4	4.0	7.5	6.6	8.5	6.4	4.5	4.9	6.4	9.1	8.8
ExBases	AVG	17.3	-	-	-	45.2	31.6	29.7	16.3	46.9	29.0	25.8	38.1	16.3	22.9	41.2	31.8
(cmolc/kg)	SD	0.0	-	-	-	-	8.9	3.0	2.7	6.6	9.9	6.1	8.7	3.4	5.6	9.8	11.1
BSat	AVG	47	-	-	-	91	80	79	41	90	73	61	82	52	68	84	74
(%)	SD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OrgC	AVG	19.1	-	-	-	11.3	13.9	16.8	23.4	16.9	21.9	21.6	16.4	16.6	12.7	16.5	19.6
(g/kg)	SD	0.3	-	-	-	-	4.0	9.4	14.6	7.5	8.8	7.6	7.0	7.7	1.2	7.0	9.5
TN	AVG	2.24	-	-	-	1.21	1.60	1.67	2.51	1.79	2.07	1.90	1.47	1.78	1.39	1.69	1.90

(g/kg)	SD	0.10	-	-	-	-	0.47	0.74	1.52	0.78	0.72	0.61	0.43	0.81	0.23	0.59	0.77
Av P	AVG	10.6	-	-	-	22.5	14.2	18.0	8.4	16.9	11.4	8.5	12.8	6.1	13.8	13.3	12.1
(mg/kg)	SD	1.9	-	-	-	-	4.5	6.6	1.6	3.5	6.2	5.7	7.4	3.1	1.0	7.3	6.6
Exch K	AVG	0.32	-	-	-	0.89	0.66	0.44	0.27	1.32	0.77	0.76	1.22	0.32	0.29	0.56	0.71
(cmolc/kg)	SD	0.04	-	-	-	-	0.63	0.36	0.02	0.51	0.53	0.60	0.25	0.11	0.19	0.32	0.52
	n	-	-	-	-	*1*	*16*	*4*	-	*4*	*48*	*56*	*1*	*2*	-	*71*	*203*
Extr S	AVG	1.37	-	-	-	1.06	0.88	1.05	0.72	1.03	1.14	0.94	0.90	0.41	0.65	0.88	1.00
(mg/kg)	SD	0.73	-	-	-	-	0.40	0.46	0.68	0.36	0.60	0.42	0.42	0.06	0.03	0.28	0.48
	n	-	-	-	-	*1*	*16*	*4*	-	*4*	*47*	*55*	*1*	*2*	-	*71*	*201*
Extr Zn	AVG	4.4	-	-	-	7.3	4.1	0.9	8.3	0.5	7.0	4.4	6.4	8.3	0.7	2.1	4.5
(mg/kg)	SD	5.0	-	-	-	-	7.4	0.8	5.7	0.1	8.3	5.9	11.3	2.3	0.8	4.7	6.7
	n	-	-	-	-	*1*	*15*	*4*	-	*4*	*46*	*56*	*1*	*2*	-	*67*	*196*
<u>Soil depth interval</u>		<u>(30-100 cm)</u>															
Clay	AVG	60	-	-	-	52	44	40	52	-	54	64	61	59	34	59	55
(g/100g)	SD	2	-	-	-	-	10	5	14	-	13	13	8	1	3	11	14
BD	ME	1.05				1.30	1.25	1.26	1.19	-	1.13	1.12	1.20	1.19	1.19	1.24	1.17
(kg/dm3)	SD	0.01				-	0.10	0.17	0.04	-	0.09	0.08	0.04	0.03	0.07	0.13	0.11
pH-H_2O	AVG	4.5	-	-	-	7.9	6.6	7.4	5.1	-	5.5	5.3	6.8	5.3	7.5	7.2	6.1
(-)	SD	0.4	-	-	-	-	1.1	1.0	0.5	-	1.0	0.5	1.1	0.5	0.9	0.9	1.2
CEC	AVG	35.0	-	-	-	43.2	39.4	36.7	38.0	-	40.9	39.4	50.4	48.2	36.0	50.6	42.7
(cmolc/kg)	SD	2.4	-	-	-	-	7.5	3.8	6.8	-	11.5	6.8	5.8	7.8	8.6	10.7	10.6
ExBases	AVG	18.9	-	-	-	41.3	30.1	30.5	16.2	-	29.4	23.1	41.4	28.5	28.1	44.8	31.9
(cmolc/kg)	SD	1.2	-	-	-	-	10.2	4.4	2.6	-	12.8	8.0	9.5	8.5	13.2	11.5	13.5
BSat	AVG	54	-	-	-	96	76	83	43	-	72	59	82	59	78	89	75
(%)	SD	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
OrgC	AVG	19.1	-	-	-	11.3	13.9	16.8	23.4	16.9	21.9	21.6	16.4	16.6	12.7	16.5	19.6
(g/kg)	SD	0.3	-	-	-	-	4.0	9.4	14.6	7.5	8.8	7.6	7.0	7.7	1.2	7.0	9.5
TN	AVG	1.07	-	-	-	0.40	0.91	0.83	0.59	-	1.21	0.99	0.94	0.68	0.70	1.18	1.09
(g/kg)	SD	0.10	-	-	-	-	0.34	0.28	0.31	-	0.51	0.40	0.14	0.02	0.04	0.56	0.56
Av P	AVG	-	-	-	-	-	8.0	-	-	-	9.8	13.8	-	-	14.5	18.1	11.8
(mg/kg)	SD	-	-	-	-	-	-	-	-	-	3.2	-	-	-	-	7.2	5.0

Exch K	AVG	0.23	-	-	-	0.93	0.56	0.55	0.27	-	0.60	0.66	1.01	1.09	0.20	0.52	0.60
(cmolc/kg)	SD	0.05	-	-	-	-	0.64	0.51	0.12	-	0.38	0.70	0.33	0.74	0.14	0.40	0.51
	n	-	-	-	-	*1*	*16*	*4*	-	*0*	*48*	*56*	*1*	*1*	-	*69*	*196*
Extr S	AVG	-	-	-	-	-	1.06	-	-	-	0.97	1.19	-	-	0.63	0.99	0.97
(mg/kg)	SD	-	-	-	-	-	-	-	-	-	0.42	-	-	-	-	0.39	0.37
Extr Zn	AVG	-	-	-	-	-	8.1	-	-	-	5.6	4.7	-	-	0.1	1.8	4.7
(mg/kg)	SD	-	-	-	-	-	-	-	-	-	3.3	-	-	-	-	1.8	3.4
	n	-	-	-	-	*0*	*0*	*0*	-	*0*	*8*	*3*	*0*	*0*	-	*4*	*15*

Table 2

All 30 woredas		**RSG**	**Acrisols**	**Alisols**	**Andosols**	**Arenosols**	**Calcisols**	**Cambisols**	**Fluvisols**	**Gleysols**	**Leptosols**	**Luvisols**	**Nitisols**	**Phaeozems**	**Planosols**	**Regosols**	**Vertisols**	**Lake**	**Sum (%)**
All. Observations counted (%)		->	0.3	3.0	2.0	1.3	0.2	8.1	3.2	0.4	13.5	21.4	17.5	1.1	2.3	4.0	21.7	0.0	100
All. Predictions counted (%)		->	0.4	3.2	1.8	1.5	0.1	4.9	1.3	0.2	13.9	20.3	21.6	0.6	2.0	2.3	25.9	0.0	100
All. Area mapped (%)		->	0.0	1.5	0.5	0.4	0.0	0.9	0.0	0.0	26.2	10.5	30.4	0.0	1.8	1.1	26.5	0.1	100
University	**Woreda**																		
Mekele	Alamata		-	-	-	-	-	2.3	0.2	-	55.8	-	-	-	-	-	42.0	-	100
Mekele	Ambalage		-	-	-	-	-	8.4	-	-	86.1	-	-	-	-	-	5.4	-	100
Mekele	Endamehone		-	-	-	-	-	0.3	-	-	93.4	-	-	-	-	-	6.5	-	100
Mekele	Ofla		-	-	-	-	-	0.5	0.2	-	93.7	-	-	-	-	-	5.6	0.2	100
Mekele	Rya Azebo		-	-	-	-	-	3.8	0.0	-	36.0	-	-	-	-	-	60.3	-	100
Bahir Dar	Bure		-	-	-	-	-	-	-	-	30.3	3.5	44.7	-	-	-	21.6	-	100
Bahir Dar	Dera		-	-	-	-	-	0.0	0.1	-	45.4	23.0	0.8	-	-	-	30.6	0.3	100
Bahir Dar	Jebitenan		-	-	-	-	-	-	-	-	19.2	0.2	39.6	-	-	-	41.1	-	100
Bahir Dar	Mecha		-	-	-	-	-	-	-	-	15.0	6.3	39.2	-	-	-	39.8	-	100
Bahir Dar	South Achefer		-	-	-	-	-	-	-	-	15.0	29.0	26.4	-	-	-	29.8	-	100
Addis Ababa	Bako Tibe		-	-	-	-	-	-	-	-	0.2	8.0	56.8	-	-	0.1	34.9	-	100
Addis Ababa	Bedcho		-	-	-	-	-	-	-	-	0.1	7.9	0.3	-	-	-	91.8	-	100
Addis Ababa	Gimbichu		-	-	-	-	-	-	-	-	18.5	4.4	-	0.0	-	-	77.1	-	100
Addis Ababa	Girar Jarso		-	-	-	-	-	2.9	-	-	37.5	12.9	0.2	-	-	-	46.6	-	100
Addis Ababa	Munesa		-	23.5	-	-	-	-	0.0	-	9.3	23.4	3.7	0.1	33.1	-	7.0	-	100
Jimma	Bedele Zuriya		0.1	-	-	-	-	0.9	-	-	0.1	8.0	75.6	-	-	0.5	15	-	100

Jimma	Dedesa	-	-	-	-	-	0.1	-	-	-	18.7	72.7	-	-	0.0	8.5	-	100
Jimma	Gera	-	3.2	-	-	-	0.0	-	0.0	0.5	29.3	66.6	-	0.5	-	-	-	100
Jimma	Limu Seka	-	-	-	-	-	0.0	-	-	2.6	1.0	89.4	-	-	-	7.1	-	100
Jimma	Omonada	-	0.2	8.2	-	-	-	-	-	8.3	13.5	50.0	-	3.1	-	16.8	-	100
Haramaya	Girawa	-	-	-	0.1	0.1	0.8	-	-	65.2	4.3	0.2	-	-	0.6	28.8	-	100
Haramaya	Habro	-	-	-	-	-	0.1	-	-	27.8	1.2	20.0	-	-	10.6	39.3	1.3	100
Haramaya	Haromaya	-	-	-	7.0	-	0.9	0.0	-	14.5	2.3	0.2	-	-	28.1	45.9	1.2	100
Haramaya	Kombolicha	-	-	-	23.3	-	1.0	0.1	-	43.9	10.5	0.7	-	-	5.1	15.3	-	100
Haramaya	Meta	-	-	-	-	-	7.7	-	-	50.9	2.9	0.0	-	-	4.2	34.3	-	100
Hawasa	Bule	-	8.2	-	-	-	0.4	-	-	0.5	87.9	2.9	-	-	-	-	-	100
Hawasa	Cheha	-	-	-	-	-	-	-	-	9.6	11.0	40.1	-	4	-	35.2	-	100
Hawasa	Enemor Ener	-	-	-	-	-	-	-	-	10.0	9.9	44.9	-	4.6	-	30.7	-	100
Hawasa	Malga	-	42.9	-	-	-	-	-	-	-	45.7	11.1	-	-	-	-	-	100
Hawasa	Misirak Azerenet Ber	-	-	-	-	-	-	-	-	1.9	66.1	18.2	-	2.6	-	11.2	-	100

Table 3

	RASTER			POLYGON		
	n	**RSG**	**RSG+PQ**	***n***	**RSG**	**RSG+PQ**
Overall purity	2594	0.58	0.49	*2291*	0.54	0.45
Overall correspondence	2294	0.84	0.81	*2291*	0.79	0.74

Table 4

Map units / Profile classes	Acrisols	Alisols	Andosols	Arenosols	Calcisols	Cambisols	Fluvisols	Gleysols	Leptosols	Luvisols	Nitisols	Phaeozems	Planosols	Regosols	Vertisols	Map unit purity	TOTAL
Acrisols	2	0	0	0	0	6	0	0	0	3	0	0	0	0	0	0.18	11
Alisols	0	61	0	0	0	0	1	3	3	13	1	0	0	0	0	0.74	82
Andosols	0	0	33	0	0	2	1	0	6	2	0	0	0	1	2	0.7	47
Arenosols	0	0	0	22	0	1	1	0	3	3	2	0	0	5	2	0.56	39
Calcisols	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	1	2
Cambisols	5	0	3	0	0	61	5	0	15	9	6	2	0	3	17	0.48	126
Fluvisols	0	0	0	0	0	5	14	0	0	0	7	1	0	0	8	0.4	35
Gleysols	0	0	0	0	0	0	0	0	0	0	1	0	3	0	0	0	4
Leptosols	0	1	7	2	2	47	6	0	205	21	14	0	0	25	30	0.57	360
Luvisols	2	12	1	3	0	14	4	1	31	344	58	2	4	12	39	0.65	527
Nitisols	0	4	0	3	0	17	11	2	44	87	314	1	8	17	52	0.56	560
Phaeozems	0	0	0	0	0	4	1	0	0	1	0	10	0	0	0	0.62	16
Planosols	0	0	0	0	0	1	1	4	2	5	3	1	27	0	9	0.51	53
Regosols	0	0	1	0	0	3	1	0	11	3	6	0	1	25	9	0.42	60
Vertisols	0	1	8	3	0	48	38	1	31	63	41	11	17	16	394	0.59	672
Class repr.	0.22	0.77	0.62	0.67	0.5	0.29	0.17	0	0.58	0.62	0.69	0.36	0.45	0.24	0.7	**0.58**	-
TOTAL	9	79	53	33	4	209	84	11	351	554	453	28	60	104	562	-	2594

Table 5

Property	RSG	Acrisols	Alisols	Andosols	Arenosols	Calcisols	Cambisols	Fluvisols	Gleysols	Leptosols	Luvisols	Nitisols	Phaeozems	Planosols	Regosols	Vertisols	All
<u>Soil profile</u>																	
Depth	ME	-	-0.6	-4.1	-4.3	-	-3.1	24.1	-	15.9	-1.7	-8.4	6.9	0.0	4.6	-12.5	-1.5
(cm)	RMSE	-	21	16	28	-	35	24	-	39	31	36	7	21	25	42	36
r = 0.69	rRMSE	-	18	16	30	-	36	24	-	150	27	28	6	18	69	32	74
	n	-	*60*	*13*	*25*	-	*41*	*2*	-	*264*	*266*	*258*	*2*	*33*	*48*	*277*	*1289*
Drainage	ME	-	0.0	0.0	-0.1	-	0.0	1.6	-	-0.3	0.0	0.1	-0.3	0.2	0.0	0.4	0.1
(class)	RMSE	-	0.7	0.0	0.5	-	0.8	1.6	-	1.0	0.8	0.9	0.3	0.7	0.6	1.5	1.0
r = 0.72	rRMSE	-	14	0	7	-	15	31	-	16	15	19	9	24	9	48	26
	n	-	*62*	*15*	*25*	-	*41*	*2*	-	*248*	*265*	*261*	*2*	*32*	*50*	*276*	*1279*
<u>Soil depth interval</u>		<u>(0-30 cm)</u>															
Clay	ME	-	-	-	-	0.00	0.56	-0.79	-	-2.97	-0.89	-1.10	-6.50	-1.64	-	-3.41	-1.78
(g/100g)	RMSE	-	-	-	-	0.0	9.8	8.2	-	7.6	11.0	12.2	6.5	4.1	-	11.5	11.2
r = 0.39	rRMSE	-	-	-	-	0	23	24	-	16	23	21	13	10	-	21	22
BD	ME	-	-	-	-	0.00	-0.01	-0.05	-	0.04	0.01	-0.02	-0.09	0.02	-	0.00	-0.01
(kg/dm3)	RMSE	-	-	-	-	0.00	0.15	0.20	-	0.07	0.10	0.12	0.09	0.04	-	0.09	0.11
r = 0.39	rRMSE	-	-	-	-	0	12	16	-	6	9	11	7	4	-	7	9
pH-H_2O	ME	-	-	-	-	0.00	-0.26	-0.07	-	-0.42	-0.12	-0.02	-0.34	0.57	-	-0.08	-0.09
(-)	RMSE	-	-	-	-	0.0	1.2	1.2	-	1.1	0.7	0.6	0.3	0.6	-	0.9	0.8
r = 0.61	rRMSE	-	-	-	-	0	19	16	-	17	13	11	5	12	-	14	13
CEC	ME	-	-	-	-	0.00	-1.66	-0.51	-	-6.91	-1.70	0.54	-0.11	8.48	-	-2.38	-1.28
(cmolc/kg)	RMSE	-	-	-	-	0.0	4.7	4.5	-	8.2	8.0	8.3	0.1	9.9	-	8.9	8.1
r = 0.39	rRMSE	-	-	-	-	0	12	12	-	16	20	20	0	32	-	18	19
ExBases	ME	-	-	-	-	0.00	-3.47	-0.88	-	-9.28	-1.17	-0.87	0.22	12.33	-	-1.85	-1.52

(cmolc/kg)	RMSE	-	-	-	-	0.0	8.3	3.4	-	13.0	9.1	7.6	0.2	14.9	-	10.4	9.2
r = 0.56	rRMSE	-	-	-	-	0	26	12	-	28	31	29	1	92	-	25	29
OrgC	ME	-	-	-	-	0.00	1.07	-1.05	-	0.11	0.66	0.24	8.41	0.66	-	-0.44	0.18
(g/kg)	RMSE	-	-	-	-	0.0	4.0	6.6	-	5.7	9.1	8.0	8.4	4.6	-	6.8	7.5
R = 0.38	rRMSE	-	-	-	-	0	29	39	-	34	42	37	51	28	-	41	39
TN	ME	-	-	-	-	0.00	0.16	-0.09	-	0.15	0.05	0.12	0.58	-0.14	-	-0.09	0.03
(g/kg)	RMSE	-	-	-	-	0.00	0.48	0.46	-	0.59	0.76	0.72	0.58	0.36	-	0.55	0.65
r = 0.30	rRMSE	-	-	-	-	0	30	28	-	33	36	38	40	20	-	33	35
Av P	ME	-	-	-	-	0.00	0.35	-1.67	-	-3.22	-0.88	-0.40	0.29	3.12	-	1.26	0.09
(mg/kg)	RMSE	-	-	-	-	0.00	4.66	5.90	-	7.14	6.23	4.74	0.29	5.65	-	6.62	5.87
r = 0.42	rRMSE	-	-	-	-	0	33	33	-	42	55	56	2	92	-	50	52
Exch K	ME	-	-	-	-	0.00	-0.08	-0.13	-	-0.68	-0.07	0.05	-0.02	0.39	-	0.08	0.01
(cmolc/kg)	RMSE	-	-	-	-	0.00	0.54	0.23	-	0.89	0.55	0.63	0.02	0.62	-	0.38	0.53
r = 0.10	rRMSE	-	-	-	-	0	83	52	-	67	71	83	1	193	-	69	76
	n	-	-	-	-	*1*	*16*	*4*	-	*4*	*48*	*56*	*1*	*2*	-	*71*	*203*
Extr S	ME	-	-	-	-	0.00	0.14	-0.22	-	-0.15	0.02	-0.03	0.28	0.25	-	0.03	0.01
(mg/kg)	RMSE	-	-	-	-	0.00	0.43	0.26	-	0.33	0.45	0.62	0.28	0.41	-	0.35	0.46
r = 0.20	rRMSE	-	-	-	-	0	49	25	-	32	39	65	31	100	-	40	49
	n	-	-	-	-	*1*	*16*	*4*	-	*4*	*47*	*55*	*1*	*2*	-	*71*	*201*
Extr Zn	ME	-	-	-	-	0.00	-0.82	-0.12	-	0.01	0.68	1.46	-5.09	-4.47	-	0.14	0.49
(mg/kg)	RMSE	-	-	-	-	0.00	5.93	0.75	-	0.27	9.14	6.54	5.09	5.50	-	4.54	6.48
r = 0.31	rRMSE	-	-	-	-	0	145	81	-	52	131	149	79	66	-	211	166
	n	-	-	-	-	*1*	*15*	*4*	-	*4*	*46*	*56*	*1*	*2*	-	*67*	*196*
Soil depth interval		(30-100 cm)															
Clay	ME	-	-	-	-	0.00	1.00	1.35	-	-	-0.96	-1.56	-11.9	-1.53	-	-3.05	-1.72
(g/100g)	RMSE	-	-	-	-	0.0	10.1	5.8	-	-	12.3	13.6	11.9	1.5	-	13.4	12.8
r = 0.42	rRMSE	-	-	-	-	0	23	14	-	-	23	21	20	3	-	23	22
BD	ME	-	-	-	-	0.00	-0.01	-0.01	-	-	0.01	0.01	0.01	-0.01	-	-0.03	-0.01
(kg/dm3)	RMSE	-	-	-	-	0.00	0.11	0.19	-	-	0.09	0.13	0.01	0.01	-	0.10	0.11
r = 0.39	rRMSE	-	-	-	-	0	9	15	-	-	8	12	1	1	-	8	9
pH-H_2O	ME	-	-	-	-	0.00	-0.35	-0.04	-	-	-0.19	-0.02	-0.15	0.58	-	-0.10	-0.11

(-)	RMSE	-	-	-	-	0.0	1.2	0.9	-	-	0.9	0.6	0.2	0.6	-	0.9	0.9
r = 0.70	rRMSE	-	-	-	-	0	19	12	-	-	16	11	2	11	-	13	14
CEC	ME	-	-	-	-	0.00	-2.72	1.32	-	-	-3.21	1.99	-1.49	7.69	-	-2.26	-1.18
(cmolc/kg)	RMSE	-	-	-	-	0.0	4.9	4.3	-	-	10.0	9.8	1.5	7.7	-	10.7	9.7
r = 0.41	rRMSE	-	-	-	-	0	13	12	-	-	25	25	3	16	-	21	22
ExBases	ME	-	-	-	-	0.00	-3.27	2.35	-	-	-2.26	-0.11	5.96	8.24	-	-2.94	-1.77
(cmolc/kg)	RMSE	-	-	-	-	0.0	7.9	4.4	-	-	10.4	10.0	6.0	8.2	-	12.7	10.8
r = 0.60	rRMSE	-	-	-	-	0	26	14	-	-	35	43	14	29	-	28	34
OrgC	ME	-	-	-	-	0.00	1.15	1.08	-	-	1.00	-0.45	-2.68	0.10	-	-1.05	-0.15
(g/kg)	RMSE	-	-	-	-	0.0	3.1	2.4	-	-	5.1	4.3	2.7	0.1	-	4.4	4.4
r = 0.27	rRMSE	-	-	-	-	0	37	29	-	-	43	44	29	2	-	40	41
TN	ME	-	-	-	-	0.00	0.09	0.10	-	-	0.09	-0.02	-0.08	-0.02	-	-0.13	-0.02
(g/kg)	RMSE	-	-	-	-	0.00	0.33	0.20	-	-	0.53	0.44	0.08	0.02	-	0.49	0.47
r = 0.23	rRMSE	-	-	-	-	0	36	24	-	-	44	44	9	3	-	42	42
	n	-	-	-	-	*1*	*16*	*4*	-	*0*	*48*	*56*	*1*	*1*	-	*69*	*196*
Av P	ME	-	-	-	-	-	-	-	-	-	0.43	-2.06	-	-	-	-3.81	-1.20
(mg/kg)	RMSE	-	-	-	-	-	-	-	-	-	1.90	3.37	-	-	-	9.17	5.16
r = 0.34	rRMSE	-	-	-	-	-	-	-	-	-	19	24	-	-	-	51	32
	n	-	-	-	-	*0*	*0*	*0*	-	*0*	*8*	*3*	*0*	*0*	-	*4*	*15*
Exch K	ME	-	-	-	-	0.00	-0.10	-0.12	-	-	-0.10	0.09	0.09	-0.11	-	0.03	0.00
(cmolc/kg)	RMSE	-	-	-	-	0.00	0.54	0.42	-	-	0.35	0.72	0.09	0.11	-	0.38	0.51
r = 0.18	rRMSE	-	-	-	-	0	96	77	-	-	59	109	9	10	-	73	83
	n	-	-	-	-	*1*	*16*	*4*	-	*0*	*48*	*56*	*1*	*1*	-	*69*	*196*
Extr S	ME	-	-	-	-	-	-	-	-	-	0.04	-0.30	-	-	-	0.05	-0.02
(mg/kg)	RMSE	-	-	-	-	-	-	-	-	-	0.40	0.45	-	-	-	0.29	0.38
r = -0.14	rRMSE	-	-	-	-	-	-	-	-	-	41	38	-	-	-	30	38
Extr Zn	ME	-	-	-	-	-	-	-	-	-	0.30	2.17	-	-	-	-0.18	0.55
(mg/kg)	RMSE	-	-	-	-	-	-	-	-	-	2.99	2.66	-	-	-	1.36	2.58
r = 0.60	rRMSE	-	-	-	-	-	-	-	-	-	53	57	-	-	-	74	60
	n	-	-	-	-	*0*	*0*	*0*	-	*0*	*8*	*3*	*0*	*0*	-	*4*	*15*
All	rRMSE	-	16	8	18	0	37	28	-	39	37	41	16	36	39	42	42