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 TECHNICAL REPORT

OT_CP_2016_1

Capacity building for scaling up of evidence-based best practices in agricultural production in Ethiopia

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)

J.G.B. Leenaars[1], E. Eyasu[2], H. Wösten[3], M. Ruiperez González[1], B. Kempen[1], A. Ashenafi[2], F. Brouwer[3]

1. ISRIC - World Soil Information
2. CASCAPE, National Coordination Unit
3. ALTERRA, Wageningen University & Research centre

Wageningen, 2016

This report was written to record and share results with the stakeholders of the CASCAPE project. To secure the quality of the report it was reviewed by the coordination unit and the project management team.

	Coordination	Project management
Name		
Date		
Signature		

Contents

List of Figures

List of Tables

Executive summary

The objective of this study is to produce a dataset of the major soil–landscape resources of the CASCAPE intervention woredas. The CASCAPE project operates in thirty woredas, located in six regions, therewith contributing to the Agricultural Growth Program of the Ethiopian Government in general and to the 'Soil fertility roadmap' in particular. To achieve the objective of the study, a collaborative project was established among CASCAPE project partners with scientists from ISRIC - World Soil Information[1] (ICSU World Data Centre for Soils), the CASCAPE[2] project's National Coordination Unit in partnership with universities from six regions in Ethiopia (Addis Ababa, Bahir Dar, Haramaya, Hawassa, Jimma, Mekelle) and ALTERRA[3] (Wageningen UR).

Geospatially explicit information on the major soil resources in the landscapes of Ethiopia is lacking or fragmented for much of the country and yet this information, locally observed and validated and nationally harmonised and consistent, is key for understanding the country's soils and their qualities as coherent support, complementary to soil fertility mapping as first reported for Ethiopia by Murphy (1959), to scaling up of evidence-based best practices in the agricultural growth program of the country.

The study started with a survey to identify and characterise the major soil types of the agricultural lands of a number of four kebeles selected for each of the 30 woreda. Soil profiles were georeferenced and described in the field and samples were taken, from depth intervals till beyond root zone depth of a selection of profiles, and analysed in the laboratory. The generated soil profile data were compiled in a database and the profiles were classified to soil reference groups, including qualifiers, according to the framework of the World Reference Base (WRB).

Using the soil profile data combined with spatial covariate data, the relationships between soil types and landscapes at kebele-level were statistically quantified with Random Forests modelling to produce a soil–landscape map at woreda-level. The map depicts the reference soil group, with a prefix-qualifier, predicted as most probable at given locations aggregated to polygons according to geomorphology and landscape facets.

The map was validated at woreda-level using additional soil profile data, also classified according to WRB, which were augered beyond the kebele-level. Combining áll soil profile data, the soil-landscape relationships were modelled at woreda-level to produce a final version of the map including a final round of validation. Map purity of the final raster product at a resolution of 250 m is approximately 50% and is estimated at 40% when generalised to a polygon version of the final map. The map purity is 60 and 50%, respectively, with the classification aggregated to the reference soil group.

The dataset, including soil profile data and soil-landscape resource maps, is available at: www.isric.org/projects/CASCAPE-ethiopia-woredas-soil-landscape-resources

Keywords: soil data, soil profile, soil map, soil landscape, WRB, random forests, Ethiopia, Africa

---

[1] http://www.isric.org/

[2] http://www.CASCAPE.info/

[3] http://www.wageningenur.nl/en/alterra/

Introduction

As part of the Growth and Transformation Plan (GTP), the Ethiopian Government aimed at doubling crop production during the Five Year Planning period (2011-2015). The Government of Ethiopia identified low soil fertility as one of the key challenges to meet the GTP objectives.

To address many of the soil related problems, the Ministry of Agriculture (MoA) and the Agricultural Transformation Agency (ATA) have established the project "*Ethiopian Soils Information System (EthioSIS*)" under the overall Soil Fertility Road Map of MoA. Among others, EthioSIS is set out to (i) to establish national and regional soils resource and other land resources databases; (ii) conduct surveys and soil fertility mapping to reformulate fertilizer recommendations; and (iii) develop tools for development of integrated soil fertility management technologies.

CASCAPE has entered into a collaboration agreement with the Government of Ethiopia to assist EthioSIS in various activities: (i) soil sampling and fertility mapping of 30 CASCAPE intervention woredas (through composite topsoil samples); (ii) establishment of a modern and well-functioning national soil resource database; (iii) training and capacity building in the area of soil fertility mapping (geo-statistics and soil data interpretation) and fertilizer recommendations (QUEFTS validation, fertilizer trials, etc.); and (iv) a soil characterisation study conducting detailed soil profile studies and classification of agricultural soils in all 30 CASCAPE intervention woredas.

This report summarizes findings of the fourth activity, results of which are intended to also contribute to the second activity.

Objective

The purpose of this study is to characterise and map the major soil resources within the landscapes of the CASCAPE intervention woredas and to make the data available to the project.

Data and methods used are described in Chapter 2, results in Chapter 3, and conclusions are given in in Chapter 4. The data and information produced are provided in the annexes.

Data and methods

The project was conducted through two studies, namely a soil characterisation study at kebele level and a soil mapping study at woreda level. The soil characterisation study included an exploratory survey to identify the major soil types within the landscapes and a detailed characterisation of representative soil pits to assess the soil properties and classify the major soil types. The soil profile point data collected during the soil characterisation study from selected kebeles served as input to the soil mapping study producing maps of soil types in the landscape of entire woredas.

Soil characterisation study at kebele level

The soil characterisation study consisted of two major components, namely the identification of major soil types and their characterisation.

Identification of major soil types

Exploratory base maps were prepared to support an exploratory soil survey to assess soil variability and to identify the major soil types in the landscape. These maps served to provide prior information about which soil types to anticipate and to verify where these may occur in the landscape.

A common base map for the six regions was provided by the map compiled at 1: 1 million scale of the geomorphology and soils of Ethiopia (FAO, 1984), shown in Figure 1. The map depicts geomorphology (geology and landforms) with a legend describing the associated, spatially implicit, terrain components (facets as depicted by landscape transects and as described by slope classes). Associated to the terrain components are soil components classified according to the legend of the Soil Map of the World (FAO, 1974). The broad scale map thus represents soil associations and would need disaggregation to make the soil types spatially explicit.

Figure 1. Geomorphology and soils of Ethiopia at scale 1: 1 M (FAO, 1984)

Derived from the above study is the Ethiopian part of the Soil and Terrain database for north-eastern Africa (FAO, 1998) also at a scale of 1: 1 million. This database includes a shapefile depicting the same map as described above though with a different legend and an attribute data table. The legend and the data are harmonised according to procedures and conventions as described by FAO (1997). Annex 2 illustrates this base map together with the delineations of the CASCAPE intervention woredas projected upon Google Earth. The database provides information, for each map legend entry, about landforms, geology, land cover and alike, together with the terrain components and associated soil components and their relative proportions. Soil components are classified according the revised legend (FAO, 1990). The database version shared with the consultants also includes the classification of the dominant soil type according to WRB (FAO, 2006b) as copied from the Soil Atlas of Africa (Jones et al., 2013).

Additionally, region specific information was collected in so far as available to possibly improveme the exploratory base maps. Main sources considered here included information from the Ethiopian Mapping Authority (topographic maps at 1: 50,000 scale), Ministry of Water Resources (river basin studies), Ministry of Agriculture, the Agricultural Transformation Agency (ATA), EthioGIS (Water and Land Resources Centre), Atlas of Ethiopia (IFPRI), the six universities involved, CASCAPE offices (PRA results) and the ISRIC library.

The exploratory survey was conducted through the agricultural lands of the selected kebeles using the base maps (both delineations and content) as hypothesis to verify. Besides exploratory observations made at road cuts and through surficial features like stoniness, colour and vegetation, eight auger point observations were made per kebele (also near trial sites) resulting in a total of some 960 observations. Auger points were georeferenced using Geographic Positioning System (GPS) and described to a depth of 120 cm unless restricted by hard rock or impenetrable layer, using the field form template as prepared by ISRIC (see annex 2b) according to the guidelines for soil profile description (FAO, 2006). Auger points were tentatively classified according to WRB reference soil groups (FAO, 2006b) representing major soil types

Characterisation and classification of major soil types

Following the identification of major soil types, representative soil profiles (soil pits) were characterised in greater detail. Soil characteristics were described using the standard template prepared by ISRIC (see annex 2a) for some 6 profiles per woreda (minimally 4, maximally 8) or some 180 soil profiles in total. The locations of the profiles were not evenly distributed over the four kebeles per woreda, but over the major soil types identified per woreda.

The master horizons with subordinate characteristics were designated to a depth of at least 180 cm (bedrock permitting) and described. From each horizon, over the whole soil depth, a sample of at least 1 kg was taken (more samples were taken from thick horizons with sampled depth intervals not extending 30 cm), properly and traceably labelled, administrated, and subsequently air dried at the universities. The samples, including several (hidden) duplicate samples, were submitted to the Soil and Plant Laboratory of the Water Works and Design Supervision Enterprise in Addis Ababa for selected laboratory analyses: particle size distribution (sand, silt, clay content) by hydrometer method (Bouyoucos, 1951) with the fractions defined according to USDA ($c<0.002<si<0.05<sa<2$ mm), bulk density of the fine earth from core samples, pH H2O in 1:2.5 soil: water solution and pH KCl in 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 H2O of 7 (Black, 1965), exchangeable bases (Ca and Mg by atomic absorption spectrometry; K and Na by flame photometry), organic carbon content (Walkley and Black, 1934), total nitrogen by the method of Kjeldahl (1883), available phosphorus content by the method of Olsen (1954), available or extractable sulphur by an unspecified method and extractable micro nutrients (Fe, Mn, Zn, Cu) by the DTPA method (Lindsay and Norvell, 1978).

The soil analytical data, as sampled from the entire soil depth, were used to check the preliminary field classifications. Each soil profile was classified according to WRB (FAO, 2006b), specifying the prefix and suffix qualifiers.

Reporting of the soil characterisation study

Each of the six universities reported the methods and results of the soil characterisation studies per woreda following a detailed outline as provided by ALTERRA.

The soil profile observations (site characteristics and soil morphological, chemical and physical characteristics) were described in the reports and the data were compiled in tables in principle according to the standard template tailored from the Africa Soil Profiles database (Leenaars et al., 2014).

The soil profile data collected and compiled during the soil characterisation study at kebele level served as input to the soil mapping study at woreda level.

Soil mapping study at woreda level

Maps of the spatial extent and distribution of the major soil types, and their properties, reflect the very basic information to generate and communicate crop and site specific soil water and fertility management recommendations and thus to scale up best practices. To this effect, it was decided to prepare woreda level soil – landscape maps in a uniform way for each of the 30 CASCAPE intervention woredas.

The soil mapping study extrapolates the georeferenced soil profile point observations made at kebele level to soil-landscape maps at woreda level. Additional soil profile point observations were made beyond the kebele level to validate this base map at woreda level. A final version of the map at woreda level was produced using áll soil profile observations collected throughout the woredas. Depicted on the map is the most probable Reference Soil Group including one prefix-qualifier, as classified according to WRB (FAO, 2006), generalised from prediction grids to polygons according to geomorphology and landscape facets.

Map production

Base map and final map

A woreda level base map was produced on the basis of the georeferenced soil profile point observations reported from the kebele-level soil characterisation study. Subsequently, a final map at woreda level was produced, also using the additional georeferenced soil profile point observations.

The soil observations, compiled under a common standard together with a selection of 281 profiles added from the Africa Soil Profiles database (Leenaars et al., 2014), gives the reference soil group (RSG) and qualifiers according to WRB (FAO, 2006b) together with geo-coordinates and profile IDs. The soil classifications were related to the landscape by projecting the soil point data upon stacks of spatial covariate data representing the spatial variability of soil forming factors *climate, organisms, relief, parent material* and *time* (CLORPT), including SRTM DEM, MODIS imagery and thematic maps (see annex 7). In addition, the original Geomorphology and Soils map (FAO, 1984) was used as a covariate and was disaggregated for this purpose into spatially explicit landscape facets (terrain components defined according to slope classes) and associated spatially explicit soil types (classified according the legend of the soil map of the world (FAO, 1974)) in line with as originally reported. Soil-landscape relations were quantified statistically with Random Forests modelling. Subsequently, the Random Forests were used to predict the distribution of soil types (Reference Soil Groups (RSG) with a prefix qualifier) over the landscapes of the 30 woredas. Separately, the distribution of qualifiers was predicted.

The outcome, at this stage, are raster maps at a resolution of 250m giving the probabilities of occurrence (in %) of each of the observed prefix-RSG combinations. Moreover, the outcome includes raster maps of the probabilities of occurrence of each of the qualifiers. Qualifiers are predicted from the diagnostic horizons, properties or materials as well as from the prefixes and suffixes reported from the soil profile observations.

In the next mapping step, the gridded maps were generalised to polygon maps matching landscape features. Slope classes were assessed from the SRTM DEM, representing the different landscape facets (with associated soil types) as described by FAO (1984) from landscape transects representing the variability within the geomorphologic units. Unique combinations of the slope classes with the predicted dominant soil type (most probable Reference Soil Group including prefix) form the basis of aggregation of the prediction grids (rasters) to 'spatially homogeneous' polygons. This aggregation implies a certain level of generalization whereby isolated pixels of divergent soil-slope combinations are eliminated from the map. The resulting soil-slope polygons are subsequently related to geomorphology as mapped by FAO (1984). The imprecise delineations of this geomorphology map have been adjusted to match the delineations of the soil-slope map by applying a majority-minority rule to the intersected soil-slope polygons. The resulting map represents unique combinations of geomorphology, slope class and the dominant Reference Soil Group including prefix qualifier.

The statistics were combined with expert soil knowledge using conventional soil maps and insights into a hybrid approach for soil-landscape mapping.

Details about random forests modelling used for digital soil mapping

Tree models, including random forests, become increasingly popular for mapping soil classes (Brungard et al., 2015; Heung et al., 2016). Tree models are part of the large family of machine learning approaches. Machine learning employs computer science and statistics to uncover patterns and relationships in (large) datasets. A tree model partitions the training dataset, for example a soil sampling dataset, in subsets (Strobl et al., 2009) by means of recursive binary splitting (Figure 2). The subsets are becoming increasingly homogenous when moving down the three, i.e. the variation in the daughter nodes is smaller than the variation in the parent node. The covariates are used as partitioning variables, and each split is chosen in such a way that it maximizes the reduction of a certain accuracy measure, such as the sum of squares in case of a continuous target property or the Shannon entropy, Gini index or misclassification error in case of a categorical target property (Hastie et al., 2008). At the terminal nodes of the tree (the leafs) a simple model is fitted to the observations that are contained in the nodes. The model can be a constant, for instance the mean of the values in case of continuous data or the majority in case of categorical data.

One of the most popular tree models currently in digital soil mapping are random forest (RF) models. A random forest consists of an ensemble (a forest) of trees (Breiman, 2001; Strobl et al., 2009). Random forests combine bootstrap aggregation (bagging) and random covariate selection for partitioning, to grow a forest of trees. This means that for each tree in the forest a bootstrap sample is drawn from the calibration dataset randomly – a bootstrap sample is a subset of the full dataset drawn with replacement. Typically, the bootstrap sample contains two-thirds of the dataset. A random-forest tree is grown to the bootstrap sample. This is done as follows (Hastie et al., 2008):

1. Select m covariates at random from the full set of p covariates. m is often taken as the square root of p or $p/3$.
2. Select the best covariate/split-point among the m. This is the covariate/split-point that results in the largest reduction in error.
3. Split the node into two daughter nodes.

These steps are repeated recursively until a minimum node size is reached. The minimum node size is the minimum number of observations that must be included in a terminal node. These observations are used for the terminal node prediction (see above). Bootstrapping and growing trees is repeated many times, a random forests typically contains 500 to 1,000 trees. Each tree is used to make a prediction at the data points or at new points (for instance the nodes of a prediction grid). The predictions of the individual trees are then aggregated in a single prediction: the random-forests predictor. For continuous data, this is the average of the predictions of the individual trees. For categorical data, the prediction is based on the majority vote of the individual tree predictions (Breiman, 2001; Hastie et al., 2008).

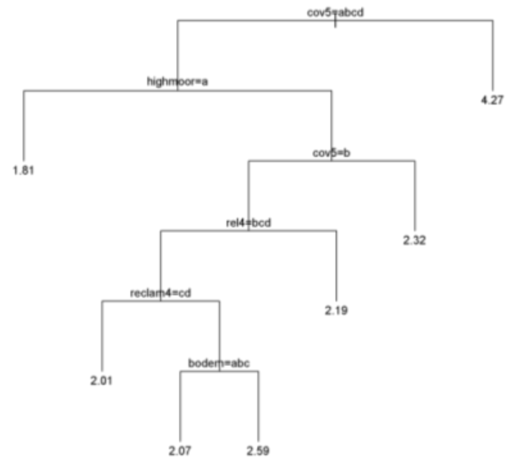

Figure 2. Example of a Random Forests tree model.

Random forests are not easy to interpret. There is no such thing as an average tree that can be visualized for interpretation for example to assess the relevance of the covariates. Despite this drawback, there are measures, the so-called 'variance importance measures' that can be used to assess the relevance of each covariate over all trees in the forest. The most advanced variance importance measure is the 'permutation accuracy importance'. Strobl et al. (2009) provide the following rationale for this measure. By randomly permutating the values of a covariate its originial association with the target variable is broken. Hence, if a permuted covariate, together with the remaining unpermuted covariates, is used to predict the target variable, then this will result in a decrease of prediction accuracy. Covariates that have the strongest association with the target variable will show the largest decrease in prediction accuracy. Thus, a reasonable measure for variable importance is the prediction accuracy before and after permuting a variable, averaged over all trees.

Map validation

Random Forests modelling includes a cross-validation procedure to assess the accuracy of the model. It sets aside a randomly selected subset (called 'bootstrapping') of soil profiles of approximately one third of the entire dataset. This subset is referred to as the 'out-of-bag' (OOB) set that are used to validate the prediction based on the model calibrated from 'in-bag' profiles. This step of calibration and validation is typcially repeated hundreds of times, here 500, with different subsets each time. The OOB profiles are used to assess the prediction accuracy of the model. This is done as follows. Once a Random-Forest tree is calibrated, it is used to predict the soil type for all OOB profiles. This is done for each tree in the forest. Because bootstrapping is done randomly, each profile will be part of the OOB set approximately one-third times the number of trees in the forest. This means that in case of this study, each profile has around 180-190 OOB predictions. The final OOB (validation) prediction is the most frequently predicted soil type. This prediction is then compared with the observed soil type. From this comparison an overall accuracy measure is computed, which is in case of soil type modelling referred to as the 'map purity' (Brus et al., 2011). Here, the map purity is defined as the percentage of soil profiles for which the predicted soil type equals the observed soil type. The results of the OOB validation can be presented in detail by purity matrices.

The map purity was assessed (step 1) for the raster base map at two levels of classification: the RSG level and the prefix-RSG combi level. With the current modelling procedure the polygon map cannot be cross-validated with independent data (such as the OOB validation for the raster maps). This means that we cannot get an objective estimate of the accuracy of the maps. Nevertheless, we checked the correspondence of the soil types according to the polygon maps with the soil types observed at the profile description locations. It should be noted, however, that this map correspondence gives a too optimistic estimate of the map purity since the same data were used for calibration and validation.

The purity of the polygon maps will likely not exceed (or is likely a fraction of) that of the raster maps from which these were derived and an estimate of the purity of the polgygon maps is given based on the fraction of the polygon map correspondence relative to the raster map correspondence.

The first Random Forests model was calibrated with kebele data only. The validation statistics are, therefore, only relevant for the base map at kebele level, but irrelevant for the parts of the base map beyond the kebele level where profile data were lacking. Additional field work was carried out to collect soil profile data beyond the kebeles to serve as an independent dataset to validate the base map at woreda-level (step 2). The data were collected by augering along transects with at least 30 auger points per woreda. The augered observations were georeferenced and classified in the field, according to WRB but also according to the local nomenclature. Few soil properties were described following the template given in annex 2b. Note that this field work is in few regions done by other consultants than those of the first phase which implies an increased chance of making different interpretations and classifications of the soils observed. An additional shortcoming is that soils cannot well be classified from (disturbed) auger observations as an adequate soil classification requires soil pit observation and soil analytical data. The auger observations thus permit to verify and confirm easily observable soil characteristics similar or comparable to soil characteristics of yet observed and classified soil profiles.

The map purity at woreda level was assessed by i) comparing the base map, both raster and polygon, with the additional soil profile data (step 2), and ii) combining the results of the cross-validation at the kebele level (of step 1) with the results of the validation with the additional data set at the woreda-level (of step 2) to get an estimate of the overall purity of the base map at woreda-level (step 3).

The final map at woreda-level (step 4) is produced by calibrating a new Random Forests model using both the data collected at kebele-level and the additional data. Again a soil-landscape polygon map was derived from the Random Forests raster predictions. The accuracy of the raster map was evaluated by the map purity based OOB cross-validation, as explained before, and the polygon map based on the correspondence of the mapped soil types with the observed soil types.

Results

Soil characterisation study

The major reference soil groups identified from the field work are, in order of importance considering number of observation points, Vertisols, Luvisols, Nitisols, Leptosols and Cambisols. These reference soil groups represent 82% of the soil profiles observed. Annex 11a gives a brief description of these major soil types, copied from Driessen et al. (2001), including main characteristics and implications for management. Other reference soil groups observed are Regosols, Fluvisols, Alisols, Planosols and Andosols, together representing 15% of the observations. The remaining 3% of the observations are classified as Arenosols, Phaeozems, Gleysols, Acrisols and Calcisols. These soil groups are also described briefly in annex 11a.

Annex 11b provides soil analytical data summarised per reference soil group considering only the first depth interval sampled from the soil profiles; note that these summary data are preliminary and not ready for being used yet.

Some of the soil profiles are also classified according to the local nomenclatures. Local nomenclatures are commonly based on landscape features and topsoil characteristics (Mulders et al., 2001) while reference soil groups are classified rather on subsoil characteristics. The main correlations between the various local soil names and the reference soil groups are shown in annex 12.

Soil data are collected for a total of 2330 profile (point) locations which corresponds with an average density of 1 profile per 11.5 km^2 for a total area of some 26,830 km^2. Added to the database is a selection of 282 profiles from the Africa Soil Profiles database, to a large extent from outside the woredas, which makes a total of 2612 soil profiles compiled.

Annex 3 gives the profile data collected from the exploratory soil survey at kebele level, including 780 soil auger observations and 221 detailed soil pit observations. The soil analytical data of the profile layers are given in annex 3b. Annex 4 gives the 1329 additional auger observations made at woreda level with relevant site data and with the soil morphologic data given per layer in annex 4b. These data, both analytical and descriptive and not limited to topsoil characteristics, permit to assess the soil fertility status of the soil as well as other important soil qualities, or constraints, such as drainage status, available soil water capacity or rootability (Leenaars et al., 2014b).

The soil profile data are georeferenced and compiled under a common standard following a tailored template prepared from the Africa Soil Profiles database (AfSP) arranging all data entries as observations and measurements by specifying the feature, property, method and value, with unit of expression, together with the lineage. Descriptive soil property data are standardised according to the data conventions of the guidelines for soil description (FAO, 2006) and numeric soil property data according to the AfSP data conventions.

The various universities used the provided template to various degrees of consistency. Consequently, the collected data could not be compiled efficiently and still needed much processing, for achieving an adequate compilation which is queryable and useable, which was done by ALTERRA/ISRIC for a selection of the most required and significant soil properties only (see annex 3). The augered data, from the additional field work, have been compiled and shared by the universities strictly following the provided template. This resulted in the fact that the data were efficiently compiled and that a wider range of property data is presented here (Annex 4).

The dataset specifies the laboratory of analysis and the laboratory methods applied as well as the units of expression applicable to the property values by a dictionary table which gives a description of the soil characteristics compiled.

The lineage of the soil profile data is specified in the dataset by reference to the unique report ID as included in the ISRIC library holding; metadata are accessible through the ISRIC library portal together (with a downloadable version of the report itself (Table 1). No lineage is defined for the additional auger data.

Table 1. Lineage of the soil profile data.

Author	Report
Kibebew Kibret, Haramaya university	library.wur.nl/WebQuery/isric/34088
Mekonnen Getahun, Bahir Dar university	library.wur.nl/WebQuery/isric/34089
Engdawork Assefa, Addis Ababa university	library.wur.nl/WebQuery/isric/34090
Amanuel Zenebe et al., Mekele university	library.wur.nl/WebQuery/isric/34091
Alemayehu Kiflu, Hawassa university	library.wur.nl/WebQuery/isric/34103
Alemayehu Regassa, Jimma university	library.wur.nl/WebQuery/isric/34104
Leenaars JGB, AJM van Oostrum and M Ruiperez Gonzalez, ISRIC	library.wur.nl/WebQuery/isric/34023

Following further validation, the soil profile data will be added to a next version of the Africa Soil Profiles database (AfSP; Leenaars et al., 2014) and subsequently imported into the World Soil Information Service (WoSIS; Ribeiro et al., 2015).

Soil mapping study

Map production

The reference soil groups observed from augers and pits are distributed throughout the landscapes as a function of the soil formation that has been taking place at the various locations which is modelled relative to spatial covariate data that mimick the soil forming factors such as parent material, climate, toposequential position, land cover and time, plus management as the sixth soil forming factor. Figure 3 illustrates the typical distribution of reference soil groups, as identified by their broad location, within the context of a landscape with more or less uniform climate and geology; it reflects differences in soil formation due to differences in toposequential position.

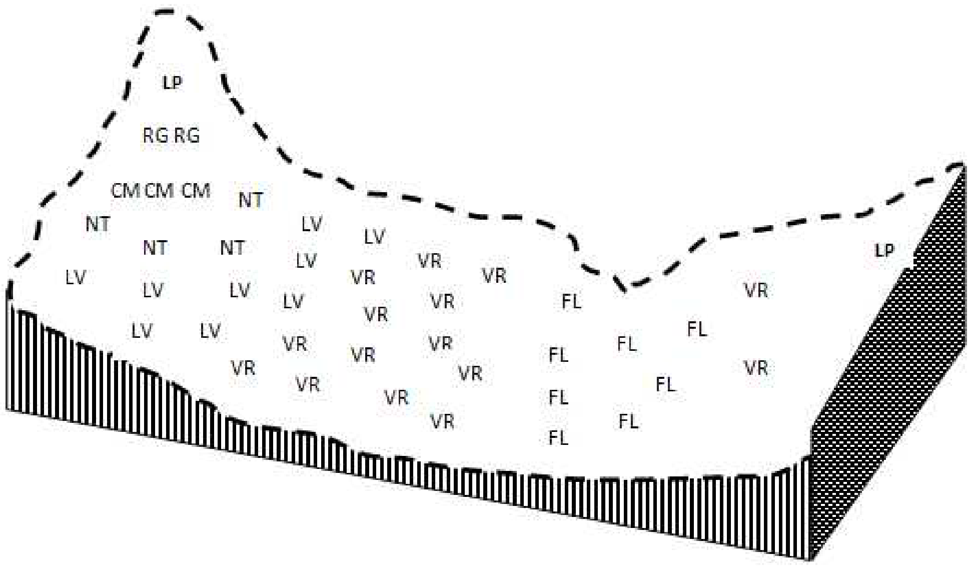

Figure 3. Example of a typical soil-landscape in Haramaya region (from Kibebew, 2014).

The base map and final map were made available to the project, both as a raster version and a generalised polygon version. For each woreda, a print of the base polygon map was provided to the universities together with a complete legend in support of the additional field work. Annex 5 gives the full legend, with geomorphology codes explained in annex 5a and the soil-landscape legend given in annex 5b. The soil-landscape legend also provides information about the three most probable reference soil groups, with prefix qualifier, and the three most probable, independent, qualifiers including prefixes, suffixes and diagnostics. Annex 6 gives a visualisation of the final polygon map for each woreda and a summary overview of the frequencies of occurrence of reference soil groups mapped for each woreda.

Figure 4 indicates the relevance of each of the covariates, listed in annex 7, offered to the prediction model for producing the final map. The relevance is assessed from and expressed as the 'permutation accuracy importance' of each covariate.

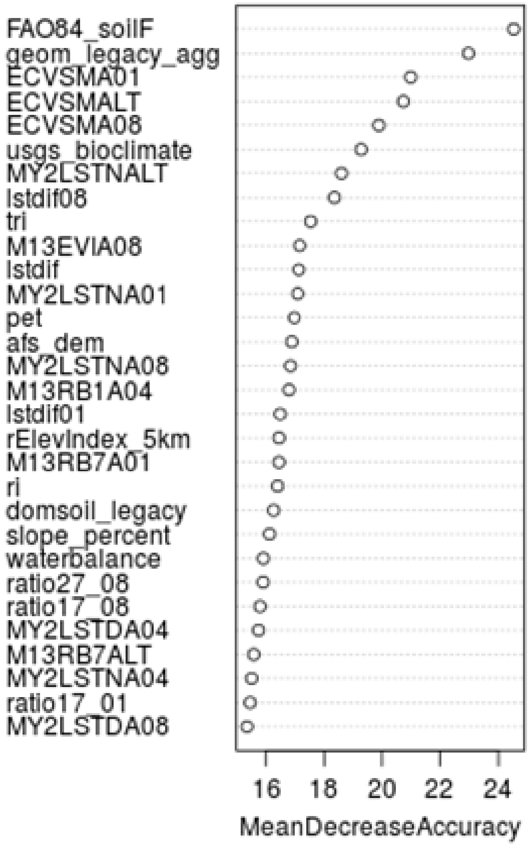

Figure 4. 'Permutation accuracy importance' of covariates in predicting the final soil map.

(See appendix 7 for abbreviations on Y-axis)

Figure 4 indicates a mean decrease of accuracy, of the final map, of 23% if the geomorphology map at 1: 1 M scale would be removed as covariate from the prediction modelling and a decrease of near 25% if the same (geomorphology) map disaggregated to soil facets (FAO84_soilF) would be removed. The map and report of the Geomorphology and Soils of Ethiopia (FAO, 1984) thus prove to be of highest relevance for predicting soil distribution, exceeding the relevance of stacks of satellite data, which coincides with the experience from many other digital soil mapping efforts that conventional legacy soil maps prove to best represent spatial variability of soil types and soil properties (Hengl et al., 2014). Annex 8 gives 'permutation accuracy importance' diagrams both for producing the final map and the base map.

Map validation

The map purity indices are reported from the Random Forests procedure and are summarised in Table 2 for the raster version of the maps. The purity of the raster base map at kebele-level (step 1) is 0.46 at the prefix-RSG level. This improves to 0.58 at the RSG level. This indicates that the prediction model captures the RSG fairly well but has difficulties in predicting the correct combination of prefix-RSG. The purity drops to 0.29/0.44 when validating with independent woreda-level data observed beyond the kebele-level (step 2). Considering all data (step 3), the purity of the base map is 0.38/0.51 at the prefix-RSG / RSG level. The final raster map (step 4) has a purity of 0.49/0.58 which exceeds that of the base map at woreda-level of step 3 but is near similar to that of the base map at kebele-level of step 1. This indicates that adding additional data to an already large dataset does not necessarily increase the overall information content in the data and neither results in better prediction accuracy at the kebele-level but does improve the information content and prediction accuracy at the woreda-level. These purities are inline with purities that are typically reported for soil class maps developed with statistical methods (Kempen et al., 2009, Kempen et al., 2012, Holmes et al., 2015, Heung et al., 2016).

The map correspondence indices are summarised in Table 2 for the polygon version of the maps. The polygon map correspondences are a fraction of the raster map correspondences and that fraction is 80-87% for the prefix-RSG combinations and 84-92% for the RSGs. Applying that fraction to the purity assessed for the raster maps gives an estimate of the purity of the final polygon maps of 0.39/0.49.

Table 2. Summary of map accuracy measures as assessed at four steps of map production.

Step	**RASTER map purity**			**POLYGON map correspondence**		
	n	Prefix-RSG	RSG	*n*	Prefix-RSG	RSG
4. Final map, all data	*2594*	0.49	0.58	*2291*	0.74	0.79
3. Base map, all data	*2594*	0.38	0.51	*2288*	0.47	0.59
2. Base map, additional data	*1328*	0.29	0.44	*1318*	0.29	0.44
1. Base map, kebele data	*1265*	0.46	0.58	*969*	0.71	0.79

(n = number of observations included, prefix = prefix qualifier, RSG = Reference Soil Group)

The map correspondence of the final polygon map is 0.74/0.79. The RSG depicted by the map corresponds with the RSG observed at similar location for 1808 out of 2291 profiles. The other 483 profiles are listed in Annex 9. The spatial distribution of the profiles of both categories is also visualised in annex 9.

Full purity matrices are given in annex 10, for steps 1, 2 and 4, comparing the RSG as observed from the ‘out-of-bag’ profiles with the RSG as predicted on the map.

Major soil types summarised

The final map of soil-landscape resources, generalised to polygons, is summarised per woreda in annex 6a and is summarised for all 30 CASCAPE intervention woredas together in Table 3 by giving the frequencies of occurrence (%) of the major reference soil groups. The frequencies (%) as counted from profile point observations are also given together with the frequencies as counted from predictions at similar point locations.

The relative importance varies, depending on the entry point for assessing the relative frequencies. Where the order of importance, considering number of observations, was in a former paragraph reported as Vertisols, Luvisols, Nitisols, Leptosols and Cambisols, representing 82% of the soil profiles observed, the order considering the area mapped is Nitisols, Vertisols and Leptosols, representing 83% of the area mapped, followed by Luvisols and Planosols.

These figures suggest that Luvisols are underrepresented on the maps, possibly due to the fact that these brownish soils may take an intermediate landscape position between the relatively poorly drained blackish vertisols and relatively well drained reddish nitisols, resulting in the latter two soils being more strongly expressed, relative to the landscape position, and overrepresented at the cost of the Luvisols. The relative representation of leptosols largely exceeds the relative number of observations which is well explainable by the fact that leptosols occupy the highest and steepest least accessible landscape positions. Cambisols and also Regosols and Fluvisols were observed frequently (15.5%) but are strongly underrepresented on the map with only 2%. This may well be due to the fact that these soils are young and 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. The Fluvisols are likely merged with the Vertisols, in the lower landscape positions, and the regosols with the leptosols. The Cambisols can be expected within close distance with most other reference groups.

Table 3. Frequencies of occurrence (%) of reference soil groups in the intervention woredas

RSG	Area mapped	Observations counted	Predictions counted
Nitisols	30.4	17.5	21.6
Vertisols	26.5	21.7	25.9
Leptosols	26.2	13.5	13.9
Luvisols	10.5	21.4	20.3
Planosols	1.8	2.3	2.0
Alisols	1.5	3.0	3.2
Regosols	1.1	4.0	2.3
Cambisols	0.9	8.1	4.9
Andosols	0.5	2.0	1.8
Arenosols	0.4	1.3	1.5
Fluvisols	0	3.2	1.3
Phaeozems	0	1.1	0.6
Acrisols	0	0.3	0.4
Calcisols	0	0.2	0.1
Gleysols	0	0.4	0.2
Lakes	0.1	0.0	0.0
	100%	100%	100%

Conclusions

The distribution of soils in the woreda landscapes has been inventoried using geomorphology as a directive and augering as a rapid approach to verify variability within the landscape. This variability was captured by the identification of the major soil resources which were characterised in detail and classified according to the World Reference Base. Next, a map was produced of the soil-landscape resources at woreda level using the soil profile observations collected at kebele level. This base map proved reasonably accurate, with a map purity of 58%, but the predictive capacity of the soil-landscape model proved less adequate to predict soil variability beyond the kebele level, as shown by the a map purity of 44% when comparing the map with additionally augered soil observations. Using áll soil profile observations collected at the woreda level, a final map was produced which represents larger soil variability though with a map purity of again 58%. Generalised to polygons, this map purity is 49%.

The map purity is reasonable and justifies a comparable approach to extrapolate soil-landscape data and findings to other woredas. However, as indicated, map purity strongly depends on the quantity, quality and reliability of the in-field profile classifications, which should be enhanced.

Reference soil groups, and qualifiers, may be assessed from a detailed field observation complemented with the necessary soil analytical data. Alternatively, it is difficult to classify a soil correctly from an auger observation which is 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.

Some auger observations were also classified according to the local (farmers) nomenclature. When done in consultation with a local land user, this may prove to be a more reliable entry to in-field soil classification whereby within-class variability can be assessed from augering till greater depth. Local soil names, plus possibly within-class variability identified from deep augering, may serve as the basis for detailed ‘scientific’ soil characterisation and classification according to WRB and for mapping the variability of soil-landscape resources. The advantage of such approach is the enhanced reliability of the in-field classification of soil observations and thus of the derived soil maps as well as the enhanced communicability and relevance for farmers and extension workers, without jeopardising the relevance and communication with the scientific community.

Outlook

By collecting and processing additional soil profile data, using a consistent approach as described in this report, reliable soil-landscape resource maps can be produced efficiently for additional woredas in Ethiopia. The results of this study, and of other well documented studies, may be used as input for the Ethiopian Soil Information Service (EthioSIS) resp. World Soil Information Service (WoSIS); once digitised and quality-assessed such data may serve soil mapping efforts in support to scaling up of best agricultural practices in Ethiopia and beyond. For this, the soil property data, of morphological, physical and chemical nature, need further interpretation relative to land use requirements and measured crop response to agricultural practices managing soil water and soil nutrients.

Acknowledgements

Funding for the CASCAPE project was provided by the Directorate-General for International Cooperation (DGIS) of the Netherlands Ministry of Foreign Affairs through the Netherlands Embassy in Ethiopia.

Acknowledgements are due to Arie van Kekem who initiated this study on soil-landscape resources within the CASCAPE project and contracted ISRIC – World Soil Information (JL) for expert advice.

Special thanks are due to 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. Koos Dijkshoorn, formerly of ISRIC, provided guidance as a consultant during the field work in two regions.

Cited and used literature

Africa Soil Information Service (AfSIS), 2013. AfSIS Climate Collection: WorldClim, 2013 release. Center for International Earth Science Information Network (CIESIN), Columbia University, New York. (ftp://africagrids.net/1000m/WorldClim/)

Africa Soil Information Service (AfSIS), website (africasoils.net/).

Alemayehu Kiflu, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas in Southern Nations Nationalities People Regional State. CASCAPE report. Hawassa university, Ethiopia. (library.wur.nl/WebQuery/isric/34103).

Alemayehu Regassa, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas in Western Oromia. CASCAPE report. Jimma university, Ethiopia. (library.wur.nl/WebQuery/isric/34104).

Amanuel Zenebe, Girmay Gebresamuel and Atkilt Girma, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas in Southern Tigray, Ethiopia. CASCAPE report. Mekele university, Ethiopia. (library.wur.nl/WebQuery/isric/34091).

Batjes, N.H., 2016. Harmonised soil property values for broad-scale modelling (WISE30sec) with estimates of global soil carbon stocks. Geoderma 269, 61-68 (with data set, available at www.isric.org).

Breiman, L., 2001. Random forests. Machine Learning 45(1), 5-32.

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

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

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. Wageningen University, ITC, Catholic University Leuven, ISRIC, FAO, Rome. (ftp://ftp.fao.org/agl/agll/docs/wsrr94e.pdf).

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/34090).

FAO, 1984. Assistance to land-use planning, Ethiopia. Geomorphology and soils. FAO, UNDP, Addis Ababa. (library.wur.nl/WebQuery/isric/5322, includes 8 map sheets). Based on Henricksen et. al.

FAO, 1997. Soils of East Africa, 1: 1 million scale. User's guide version 1.0. FAO, Rome. (library.wur.nl/WebQuery/isric/27879

FAO, 1998. The Soil and Terrain Database for northeastern Africa. Land and water digital media series 2. FAO, Rome.

FAO, 2006. Guidelines for soil description. FAO, Rome, Italy. 97 p.

FAO, 2006b. World reference base for soil resources 2006. A framework for international classification, correlation and communication. World soil resources reports No. 103. IUSS, ISRIC, FAO, Rome. 128 p.

Hastie T., R. Tibshirani and J. Friedman, 2008. The Elements of Statistical Learning. Data Mining, Inference, and Prediction. Second Edition. Springer Series in Statistics. Springer, New York.

Hengl T., J. Mendes de Jesus, R.A. MacMillan, N.H. Batjes, G.B.M. Heuvelink, E. Ribeiro, A. Samuel-Rosa, B. Kempen, J.G.B. Leenaars, M.G. Walsh and M. Ruiperez Gonzalez, 2014. SoilGrids1km - global soil information based on automated mapping. PLoS ONE 9(8). (http://journals.plos.org/plosone/article?id= 10.1371/journal.pone.0105992).

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

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

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, R. Zougmore (eds.), 2013. Soil atlas of Africa. European Commission, Publications Office of the European Union, Luxembourg. 176 p. ISBN 978-92-79-26715-4. DOI: 10.2788/52319. (eusoils.jrc.ec.europa.eu/library/maps/africa_atlas/data.html).

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.

Kempen B., D.J. Brus, G.B.M. Heuvelink, 2012. Soil type mapping using the generalised linear geostatistical model: A case study in a Dutch cultivated peatland. Geoderma 189-190, 540-553. doi: 10.1016/j.geoderma.2012.05.028

Kibebew Kibret, 2014. Characterisation of agricultural soils in CASCAPE intervention woredas in Eastern region. CASCAPE report. Haramaya university, Ethiopia. (library.wur.nl/WebQuery/isric/34088).

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. (www.isric.org/data/africa-soil-profiles-database-version-01-2).

Leenaars J.G.B., T. Hengl, M. Ruiperez Gonzalez, J. Mendes de Jesus, G.M.B. Heuvelink, J. Wolf, L. van Bussel, L. Claessens, H. Yang and K.G. Cassman, 2015. Root zone plant-available water holding capacity of the Sub-Saharan Africa soil, version 1.0. Gridded functional soil information (dataset RZ-PAWHC SSA v. 1.0). ISRIC report 2015/02. Collaboration project of Africa Soil Information Service (AfSIS) and Global Yield Gap and Water Productivity Atlas (GYGA). ISRIC - World Soil Information, Wageningen.

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. IUSS. Catena verlag, 101-130.

Mekonnen Getahun, 2015. Characterisation of agricultural soils in CASCAPE intervention woredas of Amhara region. CASCAPE report. Bahir Dar university, Ethiopia. (library.wur.nl/WebQuery/isric/34089).

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. (library.wur.nl/WebQuery/isric/261).

Ribeiro E., N.H. Batjes, J.G.B. Leenaars, A.J.M. van Oostrum and J. Mendes de Jesus, 2015. Towards the standardization and harmonization of world soil data: procedures manual ISRIC World Soil Information Service (WoSIS version 2.0). ISRIC Report 2015/03. ISRIC – World Soil Information, Wageningen. (available at: www.isric.org/data/wosis)

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 Forests. Psychological Methods 14(4), 323-348.

Trabucco A., and R.J. Zomer, 2009. Global Aridity Index (Global-Aridity) and Global Potential Evapo-Transpiration (Global-PET) Geospatial Database. CGIAR Consortium for Spatial Information. (www.cgiar-csi.org).

USGS Rocky Mountain Geographic Science Center, 2009. African surficial lithology. (rmgsc.cr.usgs.gov/ecosystems/).

USGS Rocky Mountain Geographic Science Center, 2009. Africa Land Surface Forms. (rmgsc.cr.usgs.gov/ecosystems/).