

Utilising soil diagnostic features and environmental covariates to estimate nutrient content in Ethiopian soils

Gidena T. Reda^{a,b,c,*}, Gerard B.M. Heuvelink^{d,e}, David P. Wall^c, Rogier P.O. Schulte^f,
Abbadi G. Reda^b, Eyasu Elias^g, Girmay Gebresamuel^h, Rachel E. Creamer^a

^a Soil Biology Chair Group, Wageningen University and Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands

^b Tigray Agricultural Research Institute, P.O. Box: 1070, Mekelle, Ethiopia

^c Teagasc, Crops, Environment and Land Use Programme, Johnstown Castle, Co. Wexford, Ireland

^d Soil Geography and Landscape Group, Wageningen University and Research, P.O. Box 47, 6700AA, Wageningen, the Netherlands

^e ISRIC – World Soil Information, P.O. Box 353, 6700 AJ Wageningen, the Netherlands

^f Farming Systems Ecology Group, Wageningen University and Research, P.O. Box 430, 6700 AK Wageningen, the Netherlands

^g Centre for Environmental Science, College of Natural and Computational Sciences, Addis Ababa University, P.O. Box: 1176, Addis Ababa, Ethiopia

^h Mekelle University, College of Dryland Agriculture and Natural Resources, P.O. Box: 231, Mekelle, Ethiopia

ARTICLE INFO

Keywords:

Ethiopia

Diagnostic features

Environmental covariate

Nutrient content

ABSTRACT

Smallholder farmers in Ethiopia generally do not have access to soil testing services for nutrient management planning decisions; as soil analysis is too costly for most farmers. Fertiliser advice is generally accessible via blanket recommendations at a national scale. Hence, an alternative approach is needed to estimate soil nutrient content across the diverse landscapes of Ethiopia. In this study, we propose using diagnostic features to estimate soil nutrient content, which could contribute to the development of fertiliser recommendations. To achieve this the following objectives were defined: (i) to estimate soil nutrient content as influenced by soil diagnostic features; and (ii) to elucidate the influence of environmental covariates and diagnostic features on the estimation of soil nutrient levels in the Ethiopian context. Data from 550 soil profiles, distributed across Ethiopia, were collected from a range of published sources, collated and harmonised. The data were cleaned, and 496 soil profiles were prepared for modelling. To identify which diagnostic characteristics were present across these soils we applied a presence/absence scoring method to identify dominant diagnostic features. Multiple linear regression analyses were used to predict soil chemical properties from the diagnostic features and diagnostic features along with environmental covariates. The performance of the models was evaluated by applying a 10-fold cross-validation using mean error (ME), Lin's concordance correlation coefficient (LCCC), root mean square error (RMSE) and model efficiency coefficient (MEC). The MEC values for pH, TN, and CEC derived from a combination of diagnostic features and environmental covariates were 0.38, 0.33, and 0.38. The corresponding RMSE values were 0.78, 0.07 %, and 13 cmol kg⁻¹. Additionally, the LCCC values for pH, TN, and CEC were 0.62, 0.58, and 0.62, respectively. The cross-validation results for soil chemical properties showed that the model's performance improved when environmental covariates were added. Precipitation, temperature, geology and land cover were the most important environmental covariates for estimating nutrient content, along with diagnostic features of Ethiopian soils. In conclusion, the diagnostic approach offers a useful starting point for estimating soil nutrient content. However, the variation in nutrient content across the six diagnostic features was not adequately quantified, and the model's predictive performance remains insufficient for practical application at the local scale. Further expansion of the dataset is required to fully exploit the potential of these models for underpinning nutrient management decisions across Ethiopia and in other regions where access to soil test information is limited.

* Corresponding author at: Soil Biology Chair Group, Wageningen University and Research, P.O. Box 47, 6700 AA Wageningen, the Netherlands.

E-mail address: gidenatasew@gmail.com (G.T. Reda).

<https://doi.org/10.1016/j.geodrs.2025.e00962>

Received 6 September 2024; Received in revised form 25 April 2025; Accepted 26 April 2025

Available online 28 April 2025

2352-0094/© 2025 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Achieving food security was one of the Global Sustainable Development Goals (SDGs) approved by all member countries of the United Nations (UN) in 2015. The goal is to end extreme poverty and malnutrition and promote inclusive economic growth in all forms globally by 2030 (Assefa et al., 2017). However, ensuring food security in a world with a growing and more demanding population is an urgent and complex global challenge (Bhat et al., 2023). The situation is exacerbated by changing climate conditions, environmental degradation and often geo-political instability. In countries that are not part of the Organization for Economic Co-operation and Development (OECD), such as Ethiopia, food insecurity has been a serious problem for several decades (Paul et al., 2016). The Ethiopian government instigated a national food security framework in line with the goals for SDG 2 (zero hunger) in 2015 to reduce population undernourishment O'Keeffe (2016), with the goal to enhance agricultural productivity (by providing farmers with better seeds, fertilisers, irrigation systems and training on sustainable farming techniques). This was further supported by the productive safety net programme, which aims to provide a critical safety net for chronically food insecure households (Wubetie et al., 2023). However, despite the remarkable progress that has been achieved, large parts of Ethiopia continue to face limited food production due to environmental degradation, exacerbated by lack of support for farmers to implement effective nutrient and soil management practices (Bizikova et al., 2022, 2023).

One of the challenges limiting agricultural productivity in Ethiopia is suboptimal nutrient management and declining soil fertility. The country faces high rates of soil nutrient mining, which further hampers crop yield and sustainable farming (Tesfay et al., 2010; Van Beek et al., 2016). For example, in a Teff-based (Ethiopian cereal) farming system, Hailelassie et al. (2005) reported negative partial nutrient balances for N ($-25 \text{ kg N ha}^{-1} \text{ year}^{-1}$) and K ($-87 \text{ kg K ha}^{-1} \text{ year}^{-1}$) in the central highlands of Ethiopia. Abegaz et al. (2007) studied the nutrient balance of three soil types from the Atsbi catchment of Tigray and showed nutrient depletion rates of 76 kg ha^{-1} and 63 kg ha^{-1} for N and K, respectively. More widely, Hailelassie et al. (2005) calculated that the national nutrient depletion rate for N, P and K were $122 \text{ kg N ha}^{-1} \text{ year}^{-1}$, $13 \text{ kg P ha}^{-1} \text{ year}^{-1}$ and $82 \text{ kg K ha}^{-1} \text{ year}^{-1}$, respectively. This issue is further compounded by the country's diverse soil landscape, which is influenced by varying topographical and geological conditions (Berhane et al., 2013; Tura, 2017). This geological diversity has given rise to major soil groups such as Vertisols, Cambisols, Nitisols, Luvisols, Leptosols and Fluvisols, which are distributed across different regions of the country (Ali et al., 2024). This variation in soil reference groups requires a more nuanced site-specific management approach to address the nutrient depletions, as this is not currently achieved through the blanket fertiliser recommendations.

A major challenge in estimating the nutrient content of soils in Ethiopia is not the lack of spatially explicit soil data, but rather the fragmented and scattered nature of existing datasets. These data are distributed across various institutions and individuals, limiting their widespread use Ali et al. (2020) and many soil profiles lack soil chemical property data, which is needed for nutrient management strategies. To overcome this challenge, initiatives such as the Coalition of the Willing (CoW) are working to consolidate existing soil datasets into one accessible platform (Ali et al., 2024). In the absence of a locally relevant soil information system, soil fertility management decisions are mostly made based upon blanket national recommendations or by making regional specific fertilisers which include a blend of nutrients derived from the EthioSIS soil fertility map (Elias et al., 2023). These blends are recommended to farmers to replace the national blanket recommendations, but are often not suitable at a local (field) scale, as they rely on district level soil fertility conditions that do not account for local soil variability or specific farmer needs. At a national scale the soils of Ethiopia have been mapped at an exploratory scale (1:2,000,000), which does not

provide the detailed information needed for soil management decisions at the finer spatial scales of farm or field. More recently, there has been a renewed focus on the need for more local soil information. For example, Elias (2016) and Leenaars et al. (2020) described the soils of the Ethiopian highlands at a scale of 1: 250,000 in 30 high potential districts for crop production, aiming to formulate fertiliser recommendations and develop integrated soil fertility management tools. Similar efforts were made to map soil properties of Africa at 250 m resolution for the estimation of soil nutrient content (Hengl et al., 2017). However, existing maps still lack the accuracy needed for site-specific nutrient recommendations. On the other hand, laboratory analyses (or soil scanners) are inaccessible to most smallholder farmers, due to the unaffordable costs of logistics of sampling and sending samples to the labs (Gobeze and Biswas, 2023). This requires the need for an alternative tool, which would be freely available to farmers to help them predict the nutrient content of their soils, as a contribution to improving their understanding of the nutrient content of their fields and the associated fertiliser requirements.

Recent studies have focused on addressing these gaps through diagnostic approaches that integrate soil characterization, classifications and mapping. Leenaars et al. (2021) conducted a semi-detailed survey across Ethiopia, disentangling WRB soil classification into diagnostic features and mapping them using environmental covariates. This approach provides insight into soil nutrient variation and supported the formulation of site-specific fertiliser recommendation. Diagnostic features, a central component of this approach, serve as practical proxies for soil properties, offering an alternative to extensive laboratory analyses. While the diagnostic approach does not provide the detailed assessment of nutrient availability of a specific soil as laboratory analysis would, it provides a more nuanced approach compared to blanket recommendations. Building on these achievements, in this paper we propose the concept of diagnostic classes to characterise the variation in nutrient status for six diagnostic features commonly found in Ethiopia. The diagnostic approach involves the use of diagnostic features derived from the WRB classification to describe differences between contrasting soils (IUSS Working Group WRB, 2015). These diagnostic classes were then assessed for the variation in soil properties related to soil nutrient content. In this paper, we harmonised data from 550 soil profiles from across Ethiopia to: (i) estimate soil nutrient content as influenced by the diagnostic features; and (ii) elucidate the influence of environmental covariates and soil diagnostic features on the estimation of soil nutrient levels in an Ethiopian context.

2. Materials and methods

2.1. Data collation method

In total, data from 550 soil profiles were collated across the country of Ethiopia. Input data were derived from soil scientists, institutions, published papers and grey literature (Suppl. Table: SI1). While several initiatives have focused on improving soil fertility information across Ethiopia or mapping and classification of soils (Ali et al., 2024), these two types of soil information are rarely considered together. This paper recognises the variation in soil forming process, which leads to differences in genetic soil horizons, materials and properties, and focuses on the relationship between these genetic features and chemical properties. This has resulted in the compilation and harmonization of soil profile data from 550 soil profiles across Ethiopia with associated soil nutrient laboratory data (Fig. 1).

2.2. Scoring of dominant diagnostic features

The dominant diagnostic features within the WRB framework are diagnostic horizons, properties and materials. Reference soil groups (RSGs) were classified according to the WRB framework. These classified RSGs were further categorised using a presence/absence probability

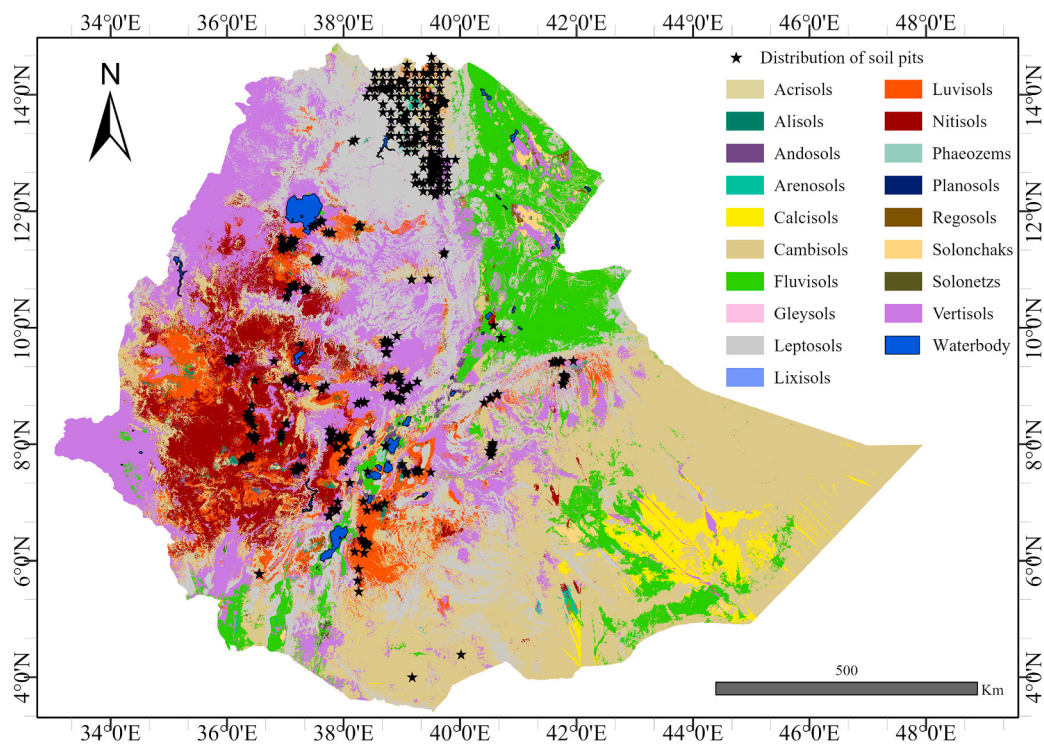


Fig. 1. Spatial distribution of the soil reference groups across Ethiopia, as mapped by Ali et al. (2024) and the distribution of 550 soil profiles collected for this study across the country. Most of soil profiles are concentrated in the highlands of Ethiopia where intensive agricultural activities are employed.

approach, grouping them based on dominant diagnostic features. This does not imply that it lacks other diagnostic features or characteristics, rather it may not contain additional diagnostic features and characteristics together with the dominant horizon. Thus, the approach provides a practical means of assessing the presence/absence of the dominant diagnostic features (Minasny et al., 2010; IUSS Working Group WRB, 2015; Simo et al., 2015; Michéli et al., 2016). These can be applied either in the field during examination and classification of a soil profile or applied where legacy data and descriptions exist from previous surveys. For Ethiopia, a comparable approach has been applied by Leenaars et al. (2021) who disentangled and mapped some 50 diagnostic features from surveyed and classified profile data. The diagnostic features are determined from specific combinations of soil properties which have developed as a result of dominant soil forming processes. These features generally manifest in the subsoil, though with a likeliness to influence the chemical properties of the surface soil. This approach considers that surface soil properties are influenced by subsurface diagnostic horizons, materials and qualifiers. For example, a vertic horizon characterized by shrink-swell clays, affects surface soil properties through a self-mulching process. The argic horizon influences the surface horizon by depleting clay and nutrient through eluviation, which can reduce soil fertility, affect water movement and alter soil structure. The surface properties of leptic soils can be strongly representative of the very shallow depth to underlying rock or parent material, which often occurs within 25 cm depth. These relationships highlight the interconnection between subsurface features and surface soil chemical characteristics. Accordingly, only soil chemical data relating to the surface are presented in this paper, as the most critical soil layer for nutrient management decisions. Presence or absence of diagnostic features was identified according to the method described by Minasny and McBratney (2007). The dominant diagnostic feature were coded ‘0’ when not present in the soil profile and ‘1’ when the diagnostic feature is a criterion or dominant for the selected soil profile. This approach was used to assign scores of dominant identifier diagnostics from all soil profiles collected. For example, we found 156 soil profiles that had vertic

diagnostic horizons, 97 of which were dominant for vertic properties and 59 not dominant yet containing vertic properties. In this paper, six commonly found diagnostic features (vertic, cambic, argic, fluvic, nitic and leptic) were coded according to the aforementioned approach (Table 1). Diagnostic features that did not fit into the six main categories were classified as ‘other’. A total of fifty-four diagnostic horizons fell into this ‘other’ category and were subsequently removed from the analysis, therefore resulting in 496 profiles for further data analysis.

2.3. Assessing the variation of soil properties across the different soil diagnostic features

For the diagnostic features presented in Table 1, summary statistics were derived by extracting diagnostic features and their associated soil properties. The diagnostic features were then assessed for variation in soil properties and in particular soil nutrient concentrations. The soil chemical properties included in this study were pH, OC (%), TN (%), P (mg kg⁻¹), and K, Ca, Mg and CEC, all expressed in cmol kg⁻¹. Details on the laboratory procedures are provided in the supplementary information (Suppl. Table: SI2). Box and whisker plots were used to reveal the minimum, maximum, median and quartiles of chemical properties for individual diagnostic features.

2.4. Multiple linear regression analysis

To estimate the soil nutrient content, we performed multiple

Table 1
Dominant and present diagnostic features.

Diagnostic features	vertic	cambic	argic	nitic	fluvic	Leptic
Dominant (D)	97	107	114	70	23	85
Present but not dominant (P)	59	7	40	0	13	65
Dominant and Present (D + P)	156	114	154	70	36	150

regression analyses to evaluate the relationship between soil properties (dependent variables) and diagnostic features, as well as diagnostic features combined with environmental covariates (independent variables). Initially, we conducted multiple regression analyses using only the diagnostic features as covariates to assess their ability to explain variation in soil chemical properties and estimate nutrient content. Subsequently, we performed additional multiple regression analyses incorporating environmental covariates along with the diagnostic features, to evaluate their combined contribution to explaining soil nutrient content. Around fifteen environmental covariates related to soil forming factors were considered (Regmi and Rasmussen, 2018). These environmental covariates were collected from ISRIC-World Soil Information, organised for Ethiopia at 250 m resolution from various sources (Suppl. Table: SI3). In each step, the environmental covariates are evaluated for addition to or subtraction from the set of covariates based on the step Akaike Information Criterion (stepAIC) (Zhang, 2016). The subtraction step iteratively eliminates the least informative covariates from the model. Accordingly, seven environmental covariates were finally selected based on their significance and ability to explain the variation in soil chemical properties, which included terrain slope, elevation, topographic position index, geology, land cover, mean annual temperature and precipitation. These selected covariates were considered as environmental indicators that influence the spatial distribution of soil chemical properties in the study area along with the diagnostic features. In multiple regression models, categorical variables for diagnostic features and environmental covariates were dummy-coded, with one category set as the reference class (intercept). This baseline allowed the effects of other categories to be interpreted relative to the reference class or intercept. The reference classes we used were cambic for diagnostic feature, basic volcanic for geology, natural vegetation for land cover, humid for precipitation, sloping for slope terrain, middle slope for topographic position index and cold for temperature. Among the dataset OC, TN, P and K content did not show a normal distribution hence, prior to regression modelling we applied a natural logarithm transformation method to better meet the normal distribution assumptions (Suppl. Figure: SI1). Subsequently, most of the observations were symmetrically distributed around the mean, except logTN, which was still somewhat skewed to the right. After analysing the log-transformed data, the results were back-transformed to their original scale for presentation. However, the data transformation step was not applied for the summary statistics, which were calculated on the raw data.

2.5. Validation and model evaluation

Prediction performance of the multiple linear regression model was evaluated using a 10-fold cross-validation procedure, using R software. The original dataset is randomly divided into ten equally sized subsets. The model is trained on nine of these subsets and tested on the remaining subset, ensuring that no data from the same profile appears in both training and test datasets. This procedure is done ten times iterations, each time withholding a different subset for testing. Performance metrics are computed over all test sets combined so that these metrics are effectively based on all observations. Four metrics were used to verify the performance of MLR models: the mean error (ME), root mean square error (RMSE), model efficiency coefficient (MEC) and Lin's concordance correlation coefficient (LCCC). These metrics were computed as follows.

$$ME = \frac{1}{n} \sum_{i=1}^n (P_i - O_i) \quad (1)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2} \quad (2)$$

$$MEC = 1 - \frac{\sum_{i=1}^n (P_i - O_i)^2}{\sum_{i=1}^n (O_i - \bar{O})^2} \quad (3)$$

$$LCCC = \frac{2r\sigma_o\sigma_p}{\sigma_o^2 + \sigma_p^2 + (\bar{O} - \bar{P})^2} \quad (4)$$

where P_i and O_i are the predicted and observed soil chemical properties; n is the number of observations; $\bar{P}$ and $\bar{O}$ are the means for the predicted and observed soil nutrient content; σ_o^2 and σ_p^2 are the variances of observed and predicted values; and r is the Pearson correlation coefficient between the predicted and observed values. ME indicates possible bias, RMSE measures the quality of the predictions, MEC measures the predictive performance of the model, and LCCC measures the degree to which the predicted and observed values follow the 1:1 line.

2.6. Software and statistical packages used in this study

All statistical analyses and modelling were conducted in R software version 4.3.1 (R Core Team, 2023). Point data were overlain on the environmental covariate maps to extract the associated covariate values for each soil profile location creating a spatial data frame using the "sp" R package. Data were assessed for regression assumptions using the "car" R package. In this study, the argument stepAIC function was assigned to a combination of forward and backward elimination in the "Mass" package (Zhang, 2016). The relative importance of each environmental covariate in MLR model was assessed by the "relweights" function in the "car" package assesses the relative importance of covariates in a regression model and calculates these relative weights by evaluating how each covariate contributes to the model's explanatory power across all possible sub models (Kabacoff, 2011). The "hydroGOF" and "relaimpo" R statistical packages were used to measure the predictions of a statistical model align with the observed data and compute cross-validation metrics, respectively.

3. Results

3.1. Variation of soil chemical properties across the diagnostic features

Table 1 shows the number of profiles where the dominant diagnostic features were classified (i.e. designated as a diagnostic feature) and where features were recognised as diagnostic properties, but this did not result in a diagnostic horizon designation (present but not dominant properties).

Fig. 2 displays boxplots of soil properties across diagnostic features. These boxplots provide a visual summary of the distribution of each of the soil properties associated with each diagnostic feature. The central tendency of the data is highlighted by the median, while the inter-quartile range indicates the variability within each class.

3.2. Diagnostic features as predictors of soil chemical properties

Multiple linear regression coefficients in Table 2 show that there were significant variations in soil pH, organic carbon (OC), total nitrogen (TN), available phosphorous (P), exchangeable Mg and K and cation exchange capacity (CEC) contents among the diagnostic features. In this analysis, the cambic horizon was used as the reference class. The lowest average pH value was found in nitic and argic horizons (5.9), which had significantly lower pH values compared to the cambic horizon. Similarly, the highest average contents of OC (3.2 %), TN (1.2 %) and CEC (41 cmol kg⁻¹) were found in both the nitic and argic horizons with significantly higher values compared to the cambic horizon, whereas the lowest values for these three parameters were obtained in soils with a leptic diagnostic horizon. Total nitrogen and CEC were also significantly higher in soils with vertic properties compared to a cambic horizon. In

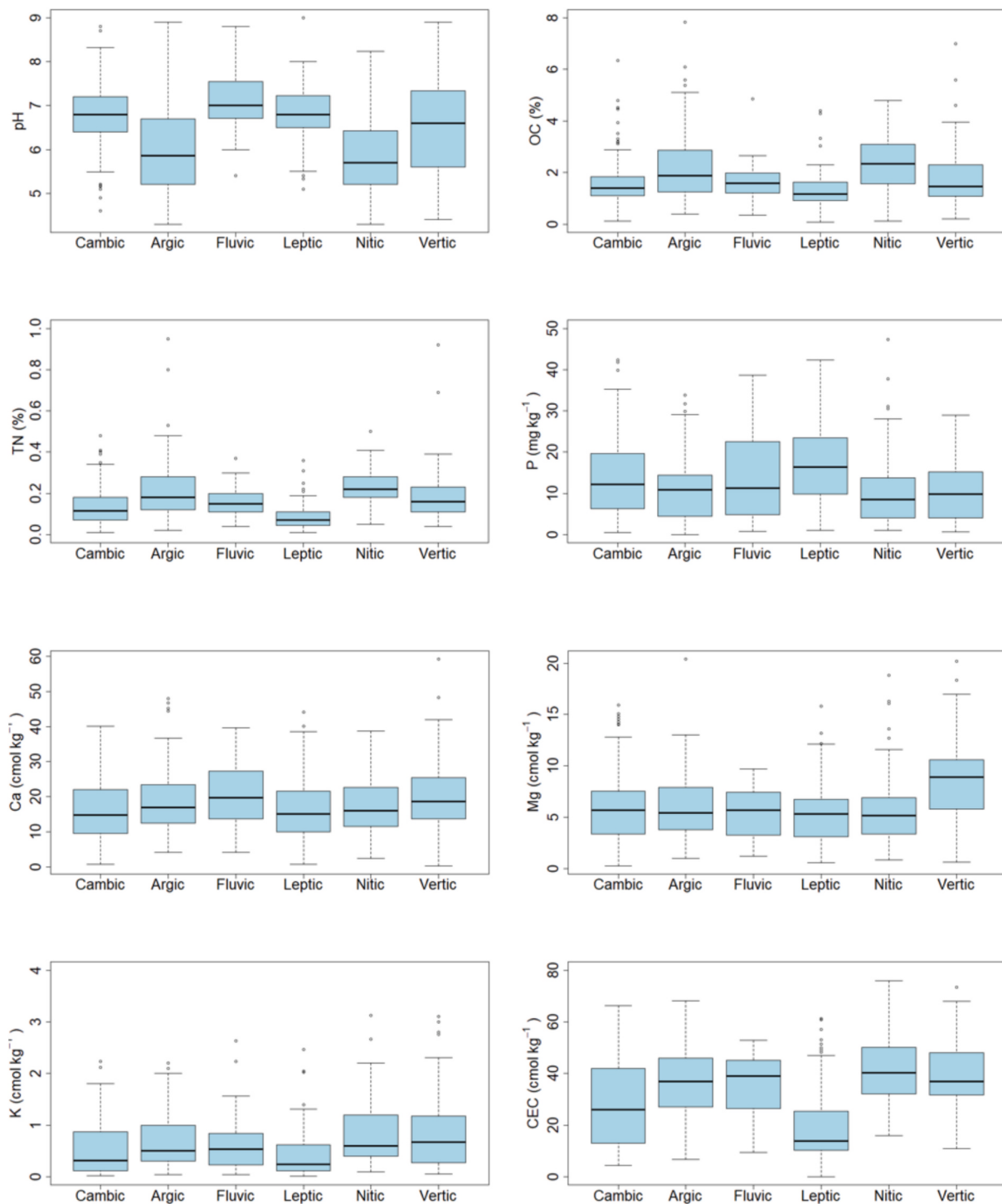


Fig. 2. Comparison among diagnostic features against the soil chemical properties with data sets of (vertic = 97; argic = 114; cambic = 107; leptic = 85; nitic = 70 and fluvic = 23). The dark lines represent the median value, box represents the interquartile range, tails represent minimum and maximum values, and the outliers are represented by dots for the selected soil chemical properties (pH, OC, TN, P, Ca, Mg, K and CEC).

soils with fluvic features CEC was also significantly higher compared to a cambic horizon. Some of the exchangeable bases varied significantly among the diagnostic features. Exchangeable Mg concentrations were found to be significantly higher in soils with vertic properties compared to a cambic horizon. Similarly, nitic, vertic and argic were obtained to be statistically significant for K when compared to soils characterized by a cambic horizon. The diagnostic features demonstrated varying degree of explanatory power for different soil chemical properties, as measured by the MEC. For all soil chemical attributes, the diagnostic features

accounted for less than 10 % of the variance, only those with a MEC > 10 % presented as an example. The model efficiency coefficient from the multiple linear regression analysis showed that the diagnostic features when considered in isolation, accounted for only 15 % of the variability across the soil profiles for pH, 17 % for TN and 18 % for CEC (Table 4). These values indicates that the diagnostic features captured a moderate proportion of the variability in these three properties. For the remaining soil chemical properties, the explanatory power was more limited.

Table 2

Results of multiple linear regression coefficients for prediction models of the diagnostic features (estimates with SE in parenthesis*).

diag. feat.	pH	OC (%)	TN (%)	P (mg kg ⁻¹)	Ca (cmol kg ⁻¹)	Mg (cmol kg ⁻¹)	K (cmol kg ⁻¹)	CEC (cmol kg ⁻¹)
Intercept	6.7 (0.08)	2.5 (1.03)	1.13 (1.00)	12 (1.07)	16 (0.92)	5.9 (0.33)	1.4 (1.03)	28 (1.41)
Argic	−0.75*** (0.12)	0.58*** (0.04)	0.07*** (0.01)	−2.2* (0.13)	2.2 (1.28)	0.026 (0.46)	0.15* (0.05)	7.4*** (1.95)
Fluvic	0.38 (0.20)	0.07 (0.07)	0.03 (0.02)	−1.6 (0.23)	3.4 (2.19)	−0.47 (0.79)	0.18 (0.07)	6.6* (3.34)
Leptic	0.10 (0.13)	−0.26* (0.04)	−0.31** (0.01)	2.8 (0.13)	−0.21 (1.30)	−0.56 (0.50)	−0.06 (0.04)	−7.4*** (2.11)
Nitic	−0.76*** (0.14)	0.69*** (0.14)	0.08*** (0.03)	−2.9** (0.14)	1.9 (1.46)	−0.032 (0.53)	0.31*** (0.05)	13*** (2.23)
Vertic	−0.19 (0.12)	0.13 (0.15)	0.05** (0.01)	−2.2 (0.13)	2.6 (1.33)	2.6*** (0.48)	0.31*** (0.04)	10*** (2.04)

*Estimate (SE) = estimated coefficient and standard error; Diagnostic features (diag. feat.): Highlighting distinctive features of soil profile; *, ** and *** indicates significance at the 5 %, 1 % and 0.1 % probability level, respectively.

3.3. Diagnostic features and environmental covariates as predictors of soil chemical properties

The model efficiency coefficient from the multiple linear regression analysis showed that the diagnostic features when considered in isolation was limited. Further multiple regression analyses were used to ascertain if other environmental covariates could improve the explained variance. This further analysis showed some improvement in the

model's accuracy for predicting the soil chemical properties (Table 3). The environmental covariates: precipitation, diagnostic features and land cover together explained 38 % of the total variability in pH. Temperature did not significantly contribute to variance in pH. In the model prediction, argic and nitic diagnostic horizons had significantly lower pH compared to the cambic horizon. Similarly, among the precipitation categories, soils in moist, sub- moist and sub-humid regions had significantly higher pH values compared to humid regions. For TN,

Table 3

Results of multiple linear regression coefficients for prediction models of diagnostic features and environmental covariates (estimates with SE in parenthesis*).

Environmental covariates		Model coefficient of soil chemical properties							
		pH	OC	TN	P	Ca	Mg	K	CEC
			%	%	mg kg ⁻¹	cmol kg ⁻¹			
D_h	Intercept	5.2 (0.48)	4.9 (1.18)	1.2 (1.06)	3.7 (1.77)	10 (4.99)	4.1 (1.76)	1.6 (1.07)	45 (7.76)
	Argic	−0.36** (0.11)	0.48* (0.05)	0.05** (0.01)	−0.33 (0.22)	2.10 (1.31)	−0.53 (0.45)	ns	3.3 (1.93)
	Fluvic	0.28 (0.18)	0.31 (0.08)	0.04 (0.02)	−0.39 (0.38)	3.60 (2.16)	−0.03 (0.74)	ns	6.6* (3.13)
	Leptic	0.08 (0.11)	−0.36 (0.04)	−0.03** (0.01)	0.64* (0.22)	0.87 (1.37)	−0.16 (0.47)	ns	−4.5* (2.02)
	Nitic	−0.37** (0.12)	0.42 (0.06)	0.05** (0.01)	−0.62 (0.26)	1.50 (1.52)	−0.95 (0.52)	ns	7.3** (2.26)
Precip	Vertic	−0.07 (0.11)	1.00 (0.05)	0.04* (0.02)	−0.42 (0.22)	2.30 (1.34)	2.3*** (0.45)	ns	6.4** (1.95)
	Moist	1.2*** (0.10)	−1.1*** (0.04)	−0.06*** (0.01)	1.9*** (0.19)	0.28 (1.17)	−0.73 (0.41)	−0.16* (0.04)	−3.7* (1.74)
	Semi_arid	0.93 (0.57)	2.27 (0.34)	0.01 (0.06)	6.4 (1.78)	0.99 (6.80)	−0.07 (2.31)	−0.06 (0.31)	−0.03 (9.91)
	Sub_humid	0.75*** (0.09)	−0.60** (0.04)	−0.04*** (0.01)	2.0*** (0.17)	2.60* (1.04)	0.55 (0.35)	−0.03 (0.03)	−1.8 (1.53)
	Sub_moist	1.10*** (0.16)	−1.15** (0.07)	−0.08*** (0.02)	2.0* (0.33)	−1.31 (2.01)	−2.2** (0.69)	−0.29** (0.07)	−11.1** (3.00)
Geo	Carbonate	ns	1.0 (0.04)	−0.03 (0.01)	0.41 (0.23)	−1.72 (1.46)	−1.11* (0.50)	−0.16* (0.06)	−7.8*** (2.14)
	Metamorphic	ns	−0.60* (0.06)	−0.06*** (0.02)	2.7** (0.26)	−3.13 (1.60)	−1.01 (0.55)	−0.31*** (0.05)	−8.6*** (2.37)
	Siliciclastic sediment	ns	−0.19 (0.05)	−0.03* (0.01)	1.9** (0.21)	0.51 (1.34)	0.65 (0.45)	−0.12 (0.05)	−0.43 (1.98)
	Unconsol. sediment	ns	0.15 (0.08)	0.01 (0.02)	0.89 (0.35)	3.90 (2.01)	0.33 (0.68)	0.20 (0.08)	−0.29 (3.03)
	Crop land	0.23 (0.14)	−0.23 (0.07)	−0.4 (0.02)	0.15 (0.28)	−1.71 (1.73)	−0.63 (0.58)	0.08 (0.07)	−3.1 (2.56)
LC	Grass land	0.97** (0.30)	0.05 (0.16)	0.01 (0.03)	−0.47 (0.74)	3.72 (3.65)	2.31 (1.24)	0.24 (0.45)	4.2 (5.32)
	Herbaceous cover	0.32 (0.30)	0.21 (0.16)	−0.01 (0.03)	−1.2 (0.74)	−3.30 (3.54)	−3.8** (1.21)	0.28 (0.45)	−5.1 (5.13)
	Tree cover	1.10*** (0.27)	−1.2** (0.14)	−0.09** (0.02)	4.0* (0.65)	7.91* (3.03)	1.42 (1.04)	0.05 (0.11)	8.6* (4.47)
	Cool	0.35 (0.47)	−1.5** (0.22)	ns	12.1* (1.73)	9.53 (4.76)	2.80 (1.67)	ns	−6.2 (7.29)
	Tepid	0.63 (0.46)	−1.2 (0.22)	ns	6.7 (1.73)	6.61 (4.71)	1.71 (1.65)	ns	−11 (7.22)
Tem	Warm	0.39 (0.47)	−1.5* (0.22)	ns	2.8 (1.78)	4.51 (4.84)	0.53* (1.70)	ns	−16 (7.41)
	Flat	ns	1.9* (0.14)	ns	ns	ns	ns	0.33 (0.13)	2.3 (5.39)
	Gently slopping	ns	1.0 (0.04)	ns	ns	ns	ns	−0.02 (0.03)	−0.82 (1.68)
	Mod. steep	ns	0.15 (0.05)	ns	ns	ns	ns	−0.12 (0.04)	−6.3** (2.09)
	Nearly level	ns	0.49 (0.06)	ns	ns	ns	ns	0.07 (0.05)	0.41 (2.32)
Slope	Steep	ns	−0.05 (0.10)	ns	ns	ns	ns	−0.12 (0.09)	−9.21* (3.53)
	Strongly slopping	ns	0.42 (0.06)	ns	ns	ns	ns	−0.09 (0.05)	−4.21 (2.16)
	Vgently slopping	ns	0.21 (0.06)	ns	ns	ns	ns	0.03 (0.05)	−0.67 (2.31)
	Flat	ns	ns	ns	ns	ns	2.23 (1.05)	ns	9.91* (4.68)
	Lower slope	ns	ns	ns	ns	ns	4.2*** (1.11)	ns	8.52 (4.87)
TPI	Ridge	ns	ns	ns	ns	ns	1.22 (0.94)	ns	3.90 (4.06)
	Upper slope	ns	ns	ns	ns	ns	0.42 (1.27)	ns	2.41 (5.44)
	Valley	ns	ns	ns	ns	ns	0.73 (0.93)	ns	2.90 (4.05)

*Estimate (SE) = estimated coefficient and standard error; ns = Not selected for the model to the respective parameter; Precip = Precipitation; D_h = Diagnostic features; LC = land cover; geo = Geology; Temp = Temperature; TPI = Topographic position index; *, ** and *** indicates significance at the 5 %, 1 % and 0.1 % probability level, respectively.

the inclusion of environmental covariates (precipitation, land cover, geology and temperature) alongside the diagnostic features (D.h) explained around 33 % of the total variability (Table 4). Furthermore, the regression model revealed that the argic, nitic and vertic diagnostic horizons had significantly higher TN levels compared to cambic, while the leptic horizon was associated with lower TN content. Among geological classes, soils based on metamorphic materials exhibited significantly lower TN levels compared to basic volcanic parent materials. Precipitation also had a significant impact on TN, with climatic conditions of moist, sub-humid and sub-moist conditions associated with lower TN levels compared to humid conditions. Land cover type such as crop land, grass land and herbaceous cover did not significantly affect TN content. However, the presence of trees was strongly related with lower TN levels compared to natural vegetation class. Both diagnostic horizons and environmental covariates (precipitation, land cover, slope gradient, slope position) explained around 38 % of the total variability in CEC across the 496 profiles. Among the soil diagnostic features, fluvic, nitic and vertic classes had significantly higher CEC compared to cambic horizons, suggesting that such soils are more capable of retaining essential nutrients i.e. nutrient cations. Conversely, leptic horizons were linked with the lower CEC. In terms of geological classes, both metamorphic and siliciclastic sediment parent materials resulted in a significant decrease in CEC compared to soils developed from basic volcanic parent materials, indicating that these parent materials were associated with soils with lower nutrient retention and exchange capacity. The presence of tree cover resulted in significantly higher CEC when compared to natural vegetation. The slope of the land affects CEC, with different slope conditions showing varied impacts. Moderately steep and steep slopes were associated with significantly lower CEC compared to middle slope. Whereas no significant difference was found for flat and nearly level or gentle slopes, compared to the middle slope. The topographic position index showed that flat areas had significantly higher CEC, indicating better conditions for nutrient retention compared to middle slope. Sites with warm temperatures had significantly lower CEC compared to regions with cold temperatures. This suggests that soils in warmer climates have a reduced capacity to retain and exchange cations. Precipitation levels also impacted CEC, with moist and sub-moist conditions significantly decreasing CEC, when compared to humid condition.

3.4. Evaluation of model performance

Table 4 presents the 10-fold cross-validation statistics used to assess the predictive performance of the models. The model metrics ranged from 0.02 to 0.38 for the MEC, 0.03 to 14.4 for RMSE and 0.27 to 0.62 for LCCC. The predication models for TN and Mg show relatively low RMSE values and good LCCC scores, indicating accurate and reliable predictions. Models for pH and CEC have moderate RMSE and LCCC values, suggesting reasonably good predictions. However, the models for P, Ca and K have higher RMSE values and lower MEC, indicating that these models may require further improvement. The zero ME across the all the metrics suggest that the models were unbiased.

Table 4

Model performance of multiple linear regression through cross-validation metrics.

Covariates	Metrics	pH	OC	TN	P	Ca	Mg	K	CEC
Diagnostic features	ME	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RMSE*	0.90	0.33	0.08	0.77	9.19	3.44	0.31	14.4
	MEC	0.15	0.10	0.17	0.05	0.02	0.09	0.07	0.18
	LCCC	0.40	0.37	0.44	0.24	0.13	0.30	0.27	0.44
Diagnostic features and environmental covariates	ME	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	RMSE*	0.78	0.30	0.07	0.72	0.08	0.03	0.30	13.00
	MEC	0.38	0.29	0.33	0.18	0.14	0.27	0.16	0.38
	LCCC	0.62	0.54	0.58	0.43	0.38	0.52	0.40	0.62

* RMSE is expressed in the same unit as the dependent variable (soil chemical properties).

3.5. Relative importance of covariates

The importance of the environmental covariates and diagnostic co-variables for each dependent variable (soil chemical properties) are shown in Fig. 3. Precipitation and diagnostic horizon had the highest effect on the majority of soil chemical properties assessed, in particular pH, OC, TN, P and Ca. In the final model prediction, of TN, variation across sites was explained by the co-variables; precipitation, diagnostic horizons, temperature, geology and land cover. From these significant environmental covariates that explained TN, the most influential predictors were precipitation (contributing about 35 %), diagnostic horizons (contributing about 28 %) and geology (contributing about 20 %). Precipitation, diagnostic horizon, geology were also important predictors describing the variability in P, however, temperature and land cover had a greater influence, and the effect of precipitation was less pronounced than on TN. For pH and OC, precipitation was the most dominant descriptor. For Mg and K, the effect of precipitation was less dominant, with the diagnostic horizon and the geology as the most significant covariates.

4. Discussion

The current fertiliser recommendation system in Ethiopia primarily relies on blanket advice at the national level, where nitrogen fertiliser in the form of urea is recommended, while at the district/regional level, blended fertilisers are recommended, based on the EthioSIS soil fertility map (Elias et al., 2019; Dargie et al., 2022). However, at community level the agricultural landscapes within these regions exhibit high variability in soil properties, due to factors such as management practices, parent materials and the landscape characteristics (Sileshi, 2022). This variability underscores the need for more localised site-specific fertiliser management to maximise the efficiency of the applied fertiliser (Reda et al., 2019; Cheng et al., 2023). Currently blend fertilisers are applied throughout the country. However, their use as a generalised solution often leads to nutrient imbalance, with some (micro) nutrients being applied insufficiently to meet the crop demand, while other are over-applied where they are not needed, reducing the capacity to reach optimal yields (Elias et al., 2019). While high resolution soil fertility maps may offer a relatively easy and practical alternative for guiding fertiliser recommendations, their use would still benefit from integration with localised soil observations such as key diagnostic features and environmental covariates that originally underpinned these maps. Furthermore, although infrared proximal soil sensors are currently under trial to estimate soil nutrient status, but are not yet effective for application at the local scale (Gobeze and Biswas, 2024; Najdenko et al., 2024). To explore other alternatives, this paper assessed whether it was possible to predict the nutrient status for macro-nutrient concentrations (using existing soil chemical properties) from 496 described profiles across Ethiopia, utilising the main descriptors of soil diagnostic features and environmental covariates. If feasible this would provide advisors with a baseline assessment from which farmers could receive more nuanced advice on fertilisation based on the major soil diagnostic features and the environmental characteristics.

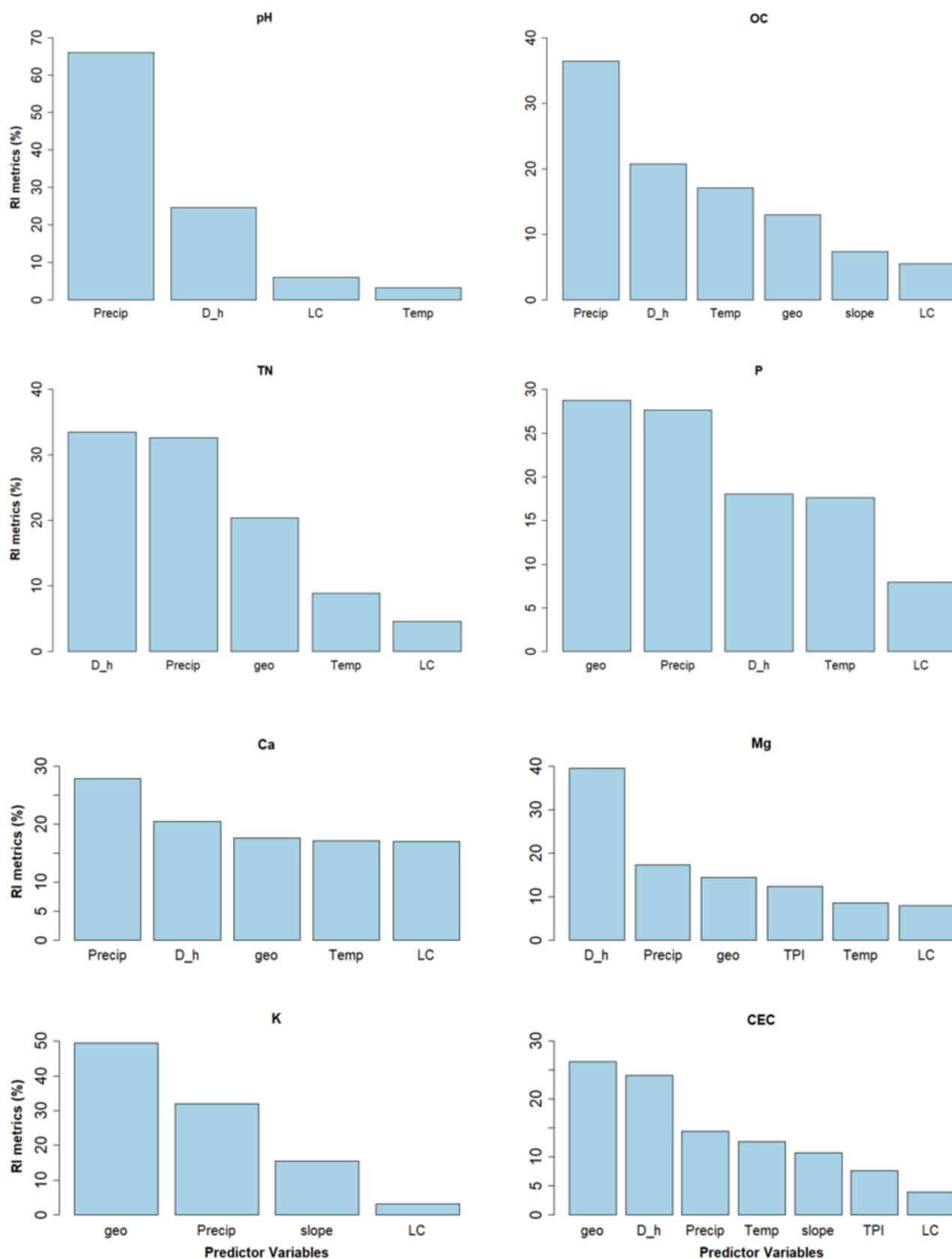


Fig. 3. Bar plots describing the relative importance metrics of environmental variables generated using the R statistical package for the prediction models of pH, OC, TN, P, Ca, Mg, K and CEC. D_h, diagnostic horizons; Precip, mean annual precipitation; Temp, mean annual temperature; geo, geology; LC, land cover; TPI, topographic position index; slope, terrain slope.

4.1. Utilising diagnostic features to estimate soil nutrient content

The multiple linear regression analyses highlight the role of diagnostic features in estimating nutrient content at local level. While the diagnostic features reflect important soil forming processes, their ability to predict chemical properties remains insufficient (Table 4). However, to demonstrate how the diagnostic approach can be used to estimate soil nutrient content, we derived estimated values for TN, pH and CEC by comparing each of the diagnostic features to the cambic horizon, which serves as a reference class. For TN, the baseline content in the cambic horizon was 1.13 ± 0.08 %. Soils with a vertic horizon were associated with a 0.05 % unit increase in TN, leading to an estimated TN content of 1.18 ± 0.12 %. Similarly, soils with argic, fluvic, leptic and nitic features were estimated to have TN contents of 1.2 ± 0.12 %, 1.1 ± 0.20 %, 0.82 ± 0.13 % and 1.2 ± 0.14 %, respectively. The differences in TN content among the diagnostic features can be attributed to variations in their pedogenic processes, organic matter content and drainage conditions. Argic and nitic horizons exhibited higher TN levels due to their ability to accumulate organic matter as a result of weathering processes and relatively stable structural characteristics that support nitrogen retention. Similar findings have been reported in soils with well-developed horizons (Smith et al., 2015). In contrast the leptic class had the lowest TN content, which can be explained by their shallow nature and limited organic matter accumulation. Soils with argic and nitic horizons exhibited significantly lower pH values compared to cambic horizon (6.7 ± 0.08), with an estimated pH of 5.9 ± 0.12 for argic and 5.9 ± 0.14 for nitic. This can be described by the characteristics of argic and nitic horizons, which are rich in clay and influenced by processes such as illuviated clay accumulation and weathering of primary minerals increase soil acidity, respectively. In contrast, fluvic and leptic soils exhibited a higher pH (7.1 ± 0.20 and 6.8 ± 0.13 , respectively), which could be attributed to the relatively young and less weathered nature of fluvic materials and the parent materials underling the leptic class. The pH soils with a dominant vertic horizon (6.5 ± 0.12) did not differ significantly from those with a cambic horizon, likely due to clay swelling and shrinkage processes that affect vertic horizons. For CEC, the baseline content in the cambic horizon was 28 ± 1.41 cmol kg⁻¹. Vertic, nitic and argic horizons showed significant differences compared to cambic horizon, with increases of 10 cmol kg⁻¹ for vertic, 13 cmol kg⁻¹ for nitic and 7.4 cmol kg⁻¹ for argic. This resulted in estimated CEC values of 38 ± 2.04 cmol kg⁻¹, 41 ± 2.23 cmol kg⁻¹ and 35 ± 1.95 cmol kg⁻¹, respectively. The leptic soil class had lower CEC, estimated at 21 ± 2.11 cmol kg⁻¹. Cation exchange capacity has been widely reported to be influenced by soil texture, organic matter content and the presence of clay minerals Solly et al. (2020), which could explain the variation observed across different horizons in this study. On the other hand, the analysis using environmental covariates alone resulted in relatively modest model efficiencies compared to the diagnostic feature, and lower than those obtained from the combined model of diagnostic features and environmental covariates (Suppl. Table: SI4).

4.2. Utilising diagnostic features and environmental covariates to estimate soil nutrient content

The utilisation of multiple regression analyses of diagnostic features and environmental covariates to estimate nutrient content associated with the 496 site locations provided some nuance, in understanding the main factors that affected nutrient estimation. The reference class coefficient can be either an increase or decrease, depending on how the covariates interact with the diagnostic features. This interaction can lead to a recalibration of the baseline reference class values as the model accounts for the influence of broader environmental factors. For the majority of soil chemical properties, precipitation classes and soil diagnostic features were the most important predictors. The inclusion of environmental covariates along with diagnostic features for TN, the baseline content in the cambic horizon was 1.2 ± 1.06 %. For argic and

nitic horizons, resulting in an estimated TN content of 1.3 ± 0.01 %. Similarly, vertic and fluvic soils had estimated TN content of 1.2 ± 0.02 %, while leptic soils were estimated at 1.1 ± 0.01 %. Environmental factors, including climate and vegetation further modulated TN content. The baseline value in pH was estimated 5.2. To estimate a pH value for vertic and argic horizons in semi-arid climate for the arable land, can be estimated 6.3 and 6.0, respectively. While fluvic and leptic soils approximately estimated around 6.6 and 6.4. The inclusion of environmental covariates and diagnostic features also revealed significant differences in CEC compared to the cambic horizon (45 cmol kg⁻¹). The estimated CEC for vertic, argic and nitic horizons in semi-arid climate arable soils with calcium carbonate content, can be estimated 40.5 cmol kg⁻¹, 37.4 cmol kg⁻¹ and 41.4 cmol kg⁻¹, respectively. Fluvic and leptic soils in similar conditions approximately estimated around 40.6 cmol kg⁻¹ and 29.6 cmol kg⁻¹, respectively. This increment in CEC, could be due to increased organic matter decomposition and mineral weathering processes (Solly et al., 2020). The estimated values of pH, TN and CEC based solely on diagnostic features were relatively higher compared to those obtained from the combined model. This is likely because diagnostic features reflect the intrinsic characteristics of the soil. When environmental covariates such as climate, topography and land use were added to the model, they introduce site-specific variability and constraints that can either increase or reduce the estimated values, depending on the local conditions.

4.3. Model evaluation performance

Overall, the performance metrics indicated weak to moderate predictive accuracy, with low mean errors across all chemical properties. Among variables assessed pH, TN, and CEC showed relatively better model performance. The MEC values for TN, pH, and CEC using diagnostic features were 0.17, 0.15, and 0.18, and the corresponding RMSE values were 0.90, 0.08 %, and 14.4 cmol kg⁻¹, respectively. Additionally, the LCCC values for pH, TN, and CEC were 0.40, 0.44, and 0.62, respectively. While environmental covariates were included alongside the diagnostic features, the model accuracy improved, with MEC values increasing to 0.38, 0.33, and 0.38, and RMSE values decreasing to 0.78, 0.07 %, and 13 cmol kg⁻¹ for pH, TN, and CEC, respectively. The LCCC values also improved to 0.52, 0.58, and 0.74, for pH, TN and CEC, respectively. The model that incorporates both diagnostic features and environmental covariates performed better than the model based solely on diagnostic features or environmental covariates. However, despite this improvement, the overall performance of the model remains relatively poor. These findings suggest that a significant portion of soil property variability remains challenging to predict using diagnostic features and the selected environmental covariates. This limited improvements in prediction accuracy can be attributed to several factors. First, the legacy soil profiles used in the model were inadequately distributed across the country, offering a restricted range of soils and diagnostic features. Additionally, these datasets were derived from multiple sources, often containing measurement and positional errors, introducing additional variability where the models were not parameterized to account for. Another source of uncertainty lies in the presence-absence probabilities used to derive the dominant diagnostic features, which further limited the predictive reliability of the models. Further, we assumed that a soil profile can have only one diagnostic feature (for example, if it is argic, then it can not be nitic or fluvic or another feature), though this is not always the case. These limitations underscore that the initial soil nutrient level estimations should be regarded as merely the first step in an ongoing process of refinement. While this approach may not offer the precision of laboratory soil tests, it may bridge the gap between blanket recommendations (often not fit for the context in which they are applied) and tailored solutions, allowing for more site-specific interventions.

4.4. Limitation of the study

In this study, a clear regional clustering pattern is observed in the datasets, which is more than likely influenced by the different main agricultural systems in the studied area. Areas with fewer soil profiles, possibly due to the lower intensity of agricultural practices compared to areas where crop cultivation or mixed farming systems are more prevalent. In these areas soil data collection is more common. The application of this approach to farmers and even extension agents may face challenges in accurately identifying diagnostic features, specifically without specialized training. Another limitation is the lack of a centralized repository for soil data, making it challenging to utilize soil information effectively. While there are some existing systems for data storage, they often fail to integrate both soil diagnostic and soil properties, leaving gaps in the overall understanding of the soil conditions. Recent initiatives have started to address these challenges; however, these efforts are still developing, and there is a pressing need to refine and harmonise the data collection practices applied across Ethiopia. Hence, incorporating soil diagnostics into future sampling campaigns is vital for advancing soil research and management in Ethiopia. This approach will lead to more precise data analyses and more effective solutions for soil management related challenges. Therefore, further expansion of the dataset is required.

5. Conclusion

The diagnostic approach offers a useful starting point for estimating soil nutrient content. However, the variation in nutrient content across the six diagnostic features was not adequately quantified, and the model's predictive performance remains insufficient for practical application at the local scale. These findings highlight the need for ongoing methodological refinements to better capture the complexities of soil nutrient assessments. Future studies should adopt more advanced modelling techniques, incorporating environmental covariates, as these have shown potential to improve model performance when used alongside diagnostic features. Furthermore, combining this approach with existing soil maps could enhance both spatial accuracy and practical relevance. Expanding the dataset and including multiple diagnostic features per profile will also be crucial for strengthening the robustness of the predictions.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We would like to thank ISRIC – World Soil Information, especially Johan. G. B. Leenaars and Maria Ruiperez Gonzalez for providing us soil profile data and the environmental covariates. Prof. Mitiku Haile and other individuals who provide us with soil profile data are also appreciated.

Appendix A. Supplementary data

Supplementary information to this article can be found online at <https://doi.org/10.1016/j.geodrs.2025.e00962>.

Data availability

Data will be made available on request.

References

- Abegaz, A., Keulen, H. van, Haile, M., Oosting, S.J., 2007. Nutrient dynamics on smallholder farms in Teghane, northern highlands of Ethiopia. In: *Advances in Integrated Soil Fertility Management in sub-Saharan Africa: Challenges and Opportunities*. Springer Netherlands, Dordrecht, pp. 365–378. https://doi.org/10.1007/978-1-4020-5760-1_34. LK - <https://wur.on.worldcat.org/oclc/5665005193>.
- Ali, A., Luelsseged, T., Teklu, E., 2020. Identifying, cataloguing, and mapping soil and agronomic data in Ethiopia. CIAT Publication no. 506. Addis Ababa (Ethiopia): International Center for Tropical Agriculture (CIAT), p. 42.
- Ali, A., Erkossa, T., Gudeta, K., Abera, W., Mesfin, E., Mekete, T., Haile, M., Haile, W., Abegaz, A., Tafesse, D., Belay, G., Getahun, M., Beyene, S., Assen, M., Regassa, A., Selassie, Y.G., Schulz, S., Tamene, L., Elias, E., Abebe, A., et al., 2024. Reference soil groups map of Ethiopia based on legacy data and machine learning-technique: EthioSoilGrids 1.0. *SOIL* 10 (1), 189–209. <https://doi.org/10.5194/soil-10-189-2024>.
- Assefa, Y., Damme, W. Van, Williams, O.D., Hill, P.S., 2017. Successes and challenges of the millennium development goals in Ethiopia : lessons for the sustainable development goals. *BMJ Glob. Health* 1–7. <https://doi.org/10.1136/bmjgh-2017-000318>.
- Berhane, G., Martens, K., Al, N., Walraevens, K., 2013. Water leakage investigation of micro-dam reservoirs in Mesozoic sedimentary sequences in northern Ethiopia. *J. Afr. Earth Sci.* 79, 98–110. <https://doi.org/10.1016/j.jafrearsci.2012.10.004>.
- Bhat, S.A., Hussain, I., Huang, N.-F., 2023. Soil suitability classification for crop selection in precision agriculture using GBRT-based hybrid DNN surrogate models. *Eco Inform* 75 (April), 102109. <https://doi.org/10.1016/j.ecoinf.2023.102109>.
- Bizikova, L., De Brauw, A., Rose, M.E., Laborde, D., Motsumi, K., Murphy, M., Parent, M., Picard, F., Smaller, C., 2022. IISD.org ii Achieving Sustainable Food Systems in a Global Crisis: Ethiopia International Institute for Sustainable Development Achieving Sustainable Food Systems in a Global Crisis: Ethiopia Ceres2030 Deep Dives into the Nexus of Food Systems, Climate Change, and Diets Achieving Sustainable Food Systems in a Global Crisis: Ethiopia. www.iisd.org.
- Bizikova, L., De Brauw, A., Rose, M.E., Laborde, D., Motsumi, K., Murphy, M., Parent, M., Picard, F., Smaller, C., 2023. IISD.org ii Achieving Sustainable Food Systems in a Global Crisis: Summary Report International Institute for Sustainable Development Achieving Sustainable Food Systems in a Global Crisis: Summary Report Achieving Sustainable Food Systems in a Global Crisis: Summary Report. www.iisd.org.
- Cheng, L., Wang, L., Wang, X., Ou, Y., Liu, H., Hou, X., 2023. The various effect of cow manure compost on the degradation of imazethapyr in different soil types. *Chemosphere* 337 (March), 139325. <https://doi.org/10.1016/j.chemosphere.2023.139325>.
- Dargie, S., Girma, T., Chibsa, T., Kassa, S., Boke, S., Abera, A., Haileselassie, B., Addisie, S., Amsalu, S., Haileselassie, M., Soboka, S., Abera, W., Weldeamayay, S.G., 2022. Balanced fertilization increases wheat yield response on different soils and agroecological zones in Ethiopia. *Exp. Agric.* 58, e23. <https://doi.org/10.1017/S0014479722000151>.
- Elias, E., 2016. Soils of the Ethiopian Highlands: Geomorphology and Properties. CASCAPE Project, ALTERA, Wageningen University and Research Centre (Wageningen UR). The Netherlands, p. 385.
- Elias, E., Okoth, P.F., Smaling, E.M.A., 2019. Explaining bread wheat (Triticum aestivum) yield differences by soil properties and fertilizer rates in the highlands of Ethiopia. *Geoderma* 339, 126–133. <https://doi.org/10.1016/j.geoderma.2018.12.020>.
- Elias, E., Okoth, P.F., Stoorvogel, J.J., Berecha, G., Berecha, G., Mellisse, B.T., Mekuriaw, A., Gebresamuel, G., Selassie, Y.G., Biratu, G.K., Smaling, E.M.A., 2023. Cereal yields in Ethiopia relate to soil properties and N and P fertilizers. *Nutr. Cycl. Agroecosyst.* 126 (2), 279–292. <https://doi.org/10.1007/s10705-023-10291-z>.
- Gobeze, T.B., Biswas, A., 2023. The need for streamlining precision agriculture data in Africa. *Precis. Agric.* 24 (1), 375–383. <https://doi.org/10.1007/s11119-022-09928-w>.
- Gobeze, T.B., Biswas, A., 2024. Preserving soil data privacy with SoilPrint: a unique soil identification system for soil data sharing. *Geoderma* 442, 116795. <https://doi.org/10.1016/j.geoderma.2024.116795>.
- Haileselassie, A., Priess, J., Veldkamp, E., Teketay, D., Lesschen, J.P., 2005. Assessment of soil nutrient depletion and its spatial variability on smallholders' mixed farming systems in Ethiopia using partial versus full nutrient balances. *Agric. Ecosyst. Environ.* 108, 1–16. <https://doi.org/10.1016/j.agee.2004.12.010>.
- Hengl, T., Leenaars, J.G.B., Shepherd, K.D., Walsh, M.G., Heuvelink, G.B.M., Mamo, T., Tilahun, H., Berkhout, E., Cooper, M., Fegraus, E., Wheeler, I., Kwabena, N.A., 2017. Soil nutrient maps of sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutr. Cycl. Agroecosyst.* 109 (1), 77–102. <https://doi.org/10.1007/s10705-017-9870-x>.
- 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.
- Kabacoff, R.I., 2011. In: Stirling, S. (Ed.), *R in Action: Data Analysis and Graphics with R*. Manning Publications Co.
- Leenaars, J.G.B., Elias, E., Wösten, J.H.M., Ruiperez-González, M., Kempen, B., 2020. Mapping the major soil-landscape resources of the Ethiopian highlands using random forest. *Geoderma* 361 (March), 114067. <https://doi.org/10.1016/j.geoderma.2019.114067>.
- Leenaars, J.G.B., Ali, A., González, M., Ruiperez, Kempen, B., Mantel, S., Kiflu, G., 2021. Semi-detailed survey and mapping of WRB soil resources in Ethiopia using different pathways to optimize between accuracy and precision.

- Michéli, E., Láng, V., Owens, P.R., Mcbratney, A., Hempel, J., 2016. Testing the pedometric evaluation of taxonomic units on soil taxonomy — a step in advancing towards a universal soil classification system. *Geoderma* 264, 340–349. <https://doi.org/10.1016/j.geoderma.2015.09.008>.
- Minasny, B., Mcbratney, A.B., 2007. Incorporating taxonomic distance into spatial prediction and digital mapping of soil classes. *Geoderma* 142, 285–293. <https://doi.org/10.1016/j.geoderma.2007.08.022>.
- Minasny, B., Mcbratney, A.B., Hartemink, A.E., 2010. Global pedodiversity, taxonomic distance, and the World Reference Base. *Geoderma* 155 (3–4), 132–139. <https://doi.org/10.1016/j.geoderma.2009.04.024>.
- Najdenko, E., Lorenz, F., Dittert, K., Ols, H.-W., 2024. Rapid in-field soil analysis of plant-available nutrients and pH for precision agriculture—a review. *Precis. Agric.* 25 (6), 3189–3218. <https://doi.org/10.1007/s11119-024-10181-6>.
- O’Keeffe, P., 2016. The role of Ethiopia’s public universities in achieving the United Nations sustainable development goals. *Int. Rev. Educ.* 62 (6), 791–813. <https://doi.org/10.1007/s11159-016-9599-9>.
- Paul, C.J., Weinthal, E.S., Bellemare, M.F., Jeuland, M.A., 2016. Social capital, trust, and adaptation to climate change: Evidence from rural Ethiopia. *Glob. Environ. Chang.* 36, 124–138. <https://doi.org/10.1016/j.gloenvcha.2015.12.003>.
- R Core Team, 2023. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Reda, G.T., Dargie, S., Hailu, B., Welde, K., Eyasu, G., 2019. Calibration of phosphorus fertilizer for barley (*Hordeum vulgare* L.) on cambisols of Endamekhoni, Northern Ethiopia. *Cogent. Food Agric.* 5 (1), 1–10. <https://doi.org/10.1080/23311932.2019.1670564>.
- Regmi, N.R., Rasmussen, C., 2018. Predictive mapping of soil-landscape relationships in the arid Southwest United States. *Catena* 165 (February), 473–486. <https://doi.org/10.1016/j.catena.2018.02.031>.
- Sileshi, G.W., 2022. Dose-response models to guide site-specific nutrient management and lessons for fertiliser trial design in sub-Saharan Africa. *Exp. Agric.* 58, e2. <https://doi.org/10.1017/S0014479721000193>.
- Simo, I., Schulte, R.P.O., Corstanje, R., Hannam, J.A., Creamer, R.E., 2015. Validating digital soil maps using soil taxonomic distance : a case study of Ireland. *GEODRS* 5, 188–197. <https://doi.org/10.1016/j.geodrs.2015.07.002>.
- Smith, P., Cotrufo, M.F., Rumpel, C., Paustian, K., Kuikman, P.J., Elliott, J.A., McDowell, R., Griffiths, R.I., Asakawa, S., Bustamante, M., House, J.I., Sobocká, J., Harper, R., Pan, G., West, P.C., Gerber, J.S., Clark, J.M., Adhya, T., Scholes, R.J., Scholes, M.C., 2015. Biogeochemical cycles and biodiversity as key drivers of ecosystem services provided by soils. *SOIL* 1 (2), 665–685. <https://doi.org/10.5194/soil-1-665-2015>.
- Solly, E.F., Weber, V., Zimmermann, S., Walthert, L., Hagedorn, F., Schmidt, M.W.I., 2020. A critical evaluation of the relationship between the effective cation exchange capacity and soil organic carbon content in Swiss Forest soils. *Front. Forests Global Change* 3. <https://doi.org/10.3389/ffgc.2020.00098>.
- Tesfay, A., Cornelis, W.M., Nysse, J., Govaerts, B., Gebregziabher, T., Oicha, T., 2010. Impact of conservation agriculture on runoff, soil loss and crop yield on a Vertisol in the northern Ethiopian highlands. In: 19th World Congress of Soil Science, Soil Solutions for a Changing World, August, 2005–2008.
- Tura, T.T., 2017. Holy Hills role in biodiversity conservation and mountain management : implication of multi-purpose Ethiopian orthodox Tewahido Church in Sustainable Mountain Conservation. *J. Biodivers. Endanger. Species* 5 (1), 179. <https://doi.org/10.4172/2332-2543.1000179>.
- Van Beek, C.L., Elias, E., Heesmans, H., Yihene, G.S., Tsegaye, A., Feyisa, H., Tolla, M., Melmuye, Y., Gebremeskel, Y., Menist, S., 2016. Soil nutrient balances under diverse agro-ecological settings in Ethiopia. *Nutr. Cycl. Agroecosyst.* 106 (3), 257–274. <https://doi.org/10.1007/s10705-016-9803-0>.
- Wubetie, H.T., Zewotir, T., Mitku, A.A., Dessie, Z.G., 2023. Household food insecurity levels in Ethiopia: quantile regression approach. *Front. Public Health* 11. <https://doi.org/10.3389/fpubh.2023.1173360>.
- Zhang, Z., 2016. Variable selection with stepwise and best subset approaches. *Ann. Transl. Med.* 4 (7). <https://doi.org/10.21037/atm.2016.03.35>.