Developing Global Pedotransfer Functions to Estimate Available Soil Phosphorus

Yones Khaledian¹, John N. Quinton², Eric C. Brevik³*, Paulo Pereira⁴, Mojtaba Zeraatpisheh⁵

¹ – Department of Agronomy, Iowa State University, Ames, IA, USA
² – Lancaster Environment Centre, Lancaster University, Lancaster, UK
³ – Department of Natural Sciences, Dickinson State University, Dickinson, ND, USA;
³ Eric.Brevik@dickinsonstate.edu
⁴ – Environmental Management Center, Mykolas Romeris University, Ateities g. 20, LT-08303 Vilnius, Lithuania
⁵ – Department of Soil Science, College of Agriculture, Isfahan University of Technology, 84156-83111 Isfahan, Iran

* - corresponding author

Abstract

There are a large number of investigations that estimate available soil phosphorous (P), but a paucity of global data on available soil P. One significant modern challenge is developing low cost, accurate approaches to predict available soil P that are useful to scientists around the world. We conducted a global meta-analysis using data on available soil P from 738 sites, 640 in the USA and 149 in 14 other countries. Four different methods of determining available soil P, New Zealand (NZ), acid oxalate, Bray and Mehlich 3 were represented in the dataset. Inputs evaluated for inclusion in the pedotransfer functions to predict available soil P were clay (C), fine silt, (FSi) coarse silt (CSi), very fine sand (VFS), fine sand (FS), medium sand (MS), coarse sand (CS), very coarse sand (VCS), organic carbon (OC), pH, calcium (Ca), magnesium (Mg), potassium (K),
iron (Fe), aluminum (Al), and manganese (Mn). Available soil P was estimated for: 1) the entire dataset, 2) only the USA, and 3) the non-USA dataset. The best models to estimate available soil P were obtained for the NZ method (using the co-variates C, FSi, CSI, VFS, MS, CS, OC, Fe, Al, Mn, Ca, Mg, and pH) and for the acid oxalate method (using the co-variates C, FSi, Fe, Al, Mn, Ca, and Mg). Although estimation of available soil P determined with the acid oxalate method was poor for the entire dataset, good estimates were obtained for the USA and non-USA datasets separately. Models for the Bray and Mehlich 3 methods only predicted available soil P well for the non-USA dataset. Using pedotransfer function models to estimate available soil P could provide an efficient and cost effective way to estimate global distributions of a soil property that is important for a number of agricultural and environmental reasons.

Key Words: Genetic Algorithm; Soil Properties; Simulation; Statistical Models

1. Introduction

Soil phosphorus (P) is essential to living organisms and often a limiting factor for plant growth and biomass production in forests, grasslands, and croplands (Trichet et al., 2009; Smith et al., 2015). It is also a limiting element for algal growth, thus contributing to eutrophication and subsequent effects on ecosystem services associated with freshwater (Carpenter et al., 2008).

Soil P is among the most limited nutrients in the soil system because much of it is in forms that are not available to plants. This has resulted in the soil P cycle being strongly affected by human management and land use, especially in agricultural areas where P has been used intensively to fertilize the soils in industrialized agricultural systems to boost the availability of P
to crops (Bryant et al., 2006). The use of P fertilizers has resulted in large soil P surpluses in countries such as China, India, and the United States (West et al., 2014) as well as parts of Europe (Lemercier et al., 2008, Follain et al., 2009), creating a risk of ecosystem contamination. Moreover, P is indispensable in physiological and biochemical processes (Simpson et al., 2011). Hence, the availability of P controls the growth and development of crops (Wyngaard et al., 2016).

In soils, P is derived mainly from weathering of the primary mineral apatite (Zhou et al., 2018) and the addition of inorganic and organic fertilizer. Total P content is high in most soils, however, only a small portion (about 13%; Sayers et al., 2008) is available for agricultural plants and microorganisms, because most P is bound strongly in soil particles and incompletely weathered material or is occluded as a secondary mineral. It is estimated that worldwide soil P content (0-50 cm) is 3.6±3 Pg in labile form, 8.6±6 Pg in organic form, 12.2±8 Pg in occluded form, and 3.2±2 Pg in secondary pools (Yang et al., 2013). The average total P in soils ranges from 200 mg/kg in older/highly weathered soils to 800 mg/kg in younger/less developed soils (Ijaz et al., 2017); the average amount of organic P in soils ranges between 30% and 65% of the total P (Condron and Tiessen, 2005). The concentration of P in the soil solution is influenced by interactions between physical, chemical and biological processes (Hinsinger, 2013; Messiga et al., 2015). P adsorption and desorption are complex processes that depend on the charge characteristics of the mineral and organic surfaces and the concentration and forms of P in the soil solution (Haygarth et al. 2005). These processes are highly pH dependent, as this alters both the characteristics of the phosphate ions in solution and the charge characteristics of the soil surfaces. In soils with high aluminum (Al) and iron (Fe) oxide contents P may be strongly bound
to mineral surfaces. However, in soils rich in carbonates (CaCO$_3$) both surface reactions between P and precipitation of P-CaCO$_3$ compounds occur (von Wandruszka, 2006) and various Ca-phosphate minerals form (Beauchemin et al., 2003), making the P unavailable to plants (Haygarth et al., 2013).

Since P is often a limiting soil factor for plant growth and excess P can have negative impacts on the environment, better ways to estimate soil P levels and spatial distribution would be useful to help scientists and stakeholders manage this important element, especially in developing nations (Borggaard et al., 2004, Krause et al., 2016). This latter point is extremely important because of the lack of resources to conduct frequent and thorough soil characterizations in these areas (Khaledian et al., 2017a, b). One option for estimating available soil P contents is to use pedotransfer functions. These consist of equations or sets of equations that allow the value of a soil property to be estimated from other properties that can be determined with simpler, faster, more efficient and/or less expensive techniques or that is already available due to previous work. Pedotransfer functions have usually been developed from multiple linear regression (MLR) models. Recently some functions have been developed using genetic algorithms (GA) to optimize the coefficients that may have been utilized in MLR (Hosseini et al., 2016; Khaledian et al., 2017a).

There have been previous attempts to develop pedotransfer functions related to available soil P, however, these have been limited in their geographic extent. Kleinman et al. (1999) developed a pedotransfer function to estimate available soil P using 59 soil samples in the USA and achieved $R^2=0.91$ for multivariate pedotransfer functions containing Al, Fe, soil organic matter, and pH as input variables. Likewise, researchers in the UK (Withers et al., 2017, $R^2$ from
5

0.64-0.87), Brazil (Valadares et al., 2017b, $R^2 = 0.73$ and 0.94), Scandinavia (Mundus et al., 2017, $R^2$ ranged from 0.46 to 0.83 using a Mitscherlich equation) and Iran (Seilsepour et al., 2008, $R^2=0.92$; Keshavarzi et al., 2016, $R^2= 0.61$ for training set and $R^2= 0.50$ for testing set; Hosseini et al., 2017, $R^2$ from 0.77-0.91) estimated available soil P based on several soil properties. However, in each of these studies the samples came from a limited geographic distribution and focused on a single available soil P test. Other studies involving pedotransfer functions for P-related issues have focused on soil phosphate adsorption capacity (Borggaard et al., 2004), available phosphate ions (Achat et al., 2011), phosphate adsorption-desorption curves (Peña and Torrent, 1990; Scheinost and Schwertmann, 1995), P mobilization (Borda et al. 2010), or remaining P (Cagliari et al., 2011). With the exception of Borggaard et al. (2004), which included a wide geographic range of samples (Canada, Denmark, Ghana, and Tanzania), these were all local (field scale) or regional (single country or a portion of a country) studies.

In this paper we aim 1) to provide the first pedotransfer functions for the global estimation of available soil P, utilising a dataset that consisted of available soil P from world-wide locations and 2) to evaluate pedotransfer functions for multiple soil P tests for the global data and national data sets from the global, USA, and non-USA locations.

2. Materials and Methods

2.1. Data Collection

We collected data using the USA Natural Resources Conservation Services (NRCS) and National Cooperative Soil Survey (NCSS) database (http://ncsslabdatamart.sc.egov.usda.gov/) targeting the collection of data generated between 2000 and 2014. This process resulted in 738
available soil P samples, 640 in the USA and 149 in other nations; Afghanistan, Canada, Chile, China, Costa Rica, Denmark, Ghana, Guatemala, Mongolia, Nigeria, Poland, Puerto Rico, Russia and Spain (14 countries) (Figure 1).

The following information was available for each sampling point: clay (C), fine silt (FSi), coarse silt (CSi), very fine sand (VFS), fine sand (FS), medium sand (MS), coarse sand (CS), very coarse sand (VCS), organic carbon (OC), pH, calcium (Ca), magnesium (Mg), potassium (K), Fe, Al, and manganese (Mn). Soil texture was determined with sieve and pipette (NRCS analysis code 3A1a). Soil OC was measured with acid dichromate digestion and FeSO₄ titration (analysis code 6A1c). Soil pH was determined using a 1:1 water to soil ratio (analysis code 4C1a2a). Fe, Al, and Mn were determined with ammonium oxalate extraction (analysis codes 6C9b, 6G12b, and 6D5b, respectively). Ca, Mg, and K were determined as NH₄OAC extractable bases (analysis codes 6N2e, 6O2d, and 6Q2b, respectively). Four different approaches for determining available P were investigated, the New Zealand (NZ) (analysis code 6S4), acid oxalate (analysis code 4G2), Bray (analysis code 4D3a1) and Mehlich 3 (analysis code 4D6a1) methods.

Prior to data analysis, observations were grouped into categories depending on their location. Owing to the large number of potential variables being evaluated to create the pedotransfer functions, particular attention was given to determining variables that were related to available soil P. Therefore, before obtaining the pedotransfer formula, we omitted the variables that did not have a significant effect on estimation of soil P using a stepwise regression model.

2.2. Statistical Methods - Data analysis.
After selecting the variables that most influenced the prediction of available soil P, we tested the fit of linear \( y = ax + c \) and nonlinear quadratic regression \( y = ax^2 + bx + c \) formulas for 1) the entire dataset, 2) the USA, and 3) the other countries (14 countries). Following that, the formulas were optimized by genetic algorithms, and finally the fitted graphs prepared by test samples. All analysis was performed using XLSTAT publisher (Data Analysis for Microsoft Excel, Addinsoft, Paris, France)

Genetic algorithms solve complex optimization issues based on natural genetics; see Holland (1975) for details and Cropper and Comerford (2005) and Johari et al. (2010) for soil science applications. The algorithm used included three basic operators: the selection, crossover and mutation (Cieniawski et al., 1995) (Figure 2). Genetic algorithms are based on the concept of evolution by natural selection as solutions are evolved in a stochastic, iterative manner. The algorithm procedure consists of the following steps (Nelson and Odeh, 2009): 1. Start at initial time \( t=0 \); 2. Initialize a population of individuals (rules) \( P(t) \); 3. Evaluate the fitness of each individual by evaluating how well a rule predicts the distribution using a random subset of observations (training dataset) and save the best individuals in a rule archive; 4. Test against fitness criterion and terminate this rule archive if the criterion is met; otherwise 5. Increase time counter; 6. Create a new set of individuals using the rule archive and random generators; 7. Apply heuristic operators to population; 8. Go back to 3.

For all methods and datasets about 70% of the data was randomly selected to train the models and the remaining data (around 30%) was used as a validation dataset. The one exception was the data for the other countries determined with the Mehlich 3 method, which only had 12 samples. Therefore, the entire other countries Mehlich 3 dataset was used for full-cross
validation in this one instance. Finally, in order to find the influential variables in different subsets, the correlation (Pearson) of variables was considered. Significant differences were considered at a p<0.05.

3. Results

To find the best regression model(s) to predict soil P by different extraction methods, six regression equations (linear and non-linear equations) were evaluated for all datasets. Table 1 gives the values of the fit indicators with the pedotransfer formulas. The NZ method had promising R² and RMSE values for all evaluated datasets (Table 1); however, a more accurate model was produced for the USA than for the entire dataset and the other locations’ data (Table 1). The formulas for linear regression in Table 1 indicate that when predicting available soil P around the globe, the importance of Al, Fe and Mn should be considered, because higher R² and RMSE values were achieved when including them. The most influential variables in the nonlinear and linear regression is Al (Table 2). In the USA, Al and Fe played influential roles in estimating available soil P, while including Mn in the regression achieved higher coefficients in the other nations’ equations. The R² values for all nonlinear models in the three datasets (R²=0.78, 0.82 and 0.84 for the entire dataset, the USA, and the other nations, respectively) were better than for the linear models, but the linear regression for the USA was also promising (R²= 0.77).

The acid oxalate method had higher R² and lower RMSE values for the linear regressions for the entire dataset and the USA locations, and for both the linear and nonlinear regressions for the other country locations. However, R² and RMSE values for the non-linear regression were not significant for the USA dataset (Table 1). Al played an important role in predicting available
soil P in almost all the formulas. The importance of Ca, Al and Mn in the entire dataset and Ca, Mg and Al in other parts of the world is clear (Table 1, 2). The nonlinear and linear models (R²=0.62 and 0.85) for other countries and the linear (R²=0.60) and nonlinear (R²=0.42) models for the USA showed promising results, but the entire dataset models did not. Linear regression models estimated available soil P better than the nonlinear regression models for the entire dataset (Table 1). Nevertheless, accuracy indicators (R² and RMSE, Table 1) showed that the input variables could not adequately predict available soil P using the entire dataset.

Results from the Bray and Mehlich 3 methods showed that apart from the USA dataset for the Mehlich 3 method and the other country locations for the Bray method, the R² and RMSE values for the other regressions were not promising (Table 1, 2).

4. Discussion

Compared to previous work, the study has considered samples from a wide geographic distribution. The fact that available P across a wide range of soil and environmental settings could be predicted using a number of relatively stable soil properties, such as sand, silt, and clay, along with other soil properties that could potentially change over shorter time periods, such as pH, extractable Ca, and total organic carbon provides the opportunity to make rapid regional and global assessments of P availability, something that is of use to both agricultural and environmental planners. Although different methods have been developed to test for available soil P, often because of the different soil conditions and properties that are encountered and different crops that are produced in various parts of the world, relationships with good predictive
ability were developed for both the NZ and acid oxalate available P methods. However, relationships for the Mehlich 3 test were poorer.

The pedotransfer functions allowed identification of the key variables controlling P availability for large scale assessments. Except for the Bray method, at least one of Al, Fe, and Mn were used in all the pedotransfer functions to predict available soil P (Table 1), indicating the importance of these variables. Available soil P is known to be closely related to the concentration of Fe and Al (Vincent et al., 2012; Seguel et al., 2015) as Al and Fe oxides are the main phosphate adsorbents in soils (Borggaard, 1990, 2002). Borggaard et al. (2004) found close relationships between amounts of adsorbed phosphate and soil contents of some forms of Al and Fe and created pedotransfer functions for predicting adsorbed phosphate from Al and Fe oxide contents that were quantified by selective extractions. The relationship between available soil P and Fe was weaker in this study than for Al, agreeing with the findings of Karlsson et al. (2008) and Turner et al. (2003).

Soil texture is another important factor for predicting available soil P in the pedotransfer functions (Table 1). Texture can substantially affect the chemical characteristics of soil, including the formation of stable P-Al-organic matter bonds and P leaching from soil (Negassa and Leinweber, 2009; Sugihara et al., 2012), which in turn is linked to plant available P. The influence of fine particles such as C, FSi and CSi were important in all pedotransfer functions except those for the Mehlich 3 method (Table 1). Clay content was significantly correlated to P in almost all the models, as P can be adsorbed on the surfaces of clay minerals (Shen et al., 2011; Ulén and Snäll, 2007). Clay mineralogy is also important for determining the P-sorption capability of clays. Clay minerology was not a variable investigated in this study and represents a potential direction
of future research. For the Mehlich 3 method, coarse soil texture components, such as CS, VCS, and CSi were important in the prediction equations for the USA dataset. This may be because soil texture is important in determining P leaching from soils through its influence on soil hydrology (Negassa and Leinweber, 2009).

The importance of Ca and Mg in the entire dataset and in other parts of the world for the acid oxalate method is interesting, given that the acid oxalate method is typically used in acidic soils. This could be because the acid oxalate method causes excessive solubilization of non-labile P pools, (Six et al., 2012; Valadares et al., 2017a), as e.g. Ca-P, therefore Ca shows amongst the predictor variables because Ca from Ca-P is being released into solution. Mg is probably a predictor variable given that Ca and Mg have similar chemical behavior in soils (Havlin et al., 2005).

In our study soil OC was used in the pedotransfer functions for all available soil P methods, except for the acid oxalate method. Soil OC can influence available soil P by providing binding sites (Kang et al., 2009), and the transformation of organic P strongly influences overall P bioavailability in soils (Shen et al., 2011). Yang et al. (2013) note that organic P is a major part of the global soil P pool.

It is clear that agricultural practices will have a significant influence over available P and, as we have not included them as variables in our work, will be one of the major sources of uncertainty in the study. Given that fertilizer application accounts for 42 % of total P inputs to agricultural ecosystems, or 25 % of total plant P uptake in fertilized soils (Wang et al., 2010), fertilizer inputs of P are likely to account for a meaningful amount of the variation not explained by the pedotransfer models presented here. Therefore future attempts to develop available P
pedotransfer functions may benefit from finding a way to account for fertilizer inputs. Our pedotransfer functions also did not incorporate crops, something that represents a limitation for the equations.

5. Conclusions

This research fills a clear gap in our ability to estimate available soil P at a global scale. Pedotransfer functions were developed for the common available soil P tests for global datasets. Non-linear functions provided the best predictions of available soil P for global and non-global datasets. Including fertilizer and soil management practices may improve these predictions further.

Our work has demonstrated that the development of pedotransfer functions for available soil P over large areas using routinely collected soil survey data is possible. This could help address issues such as food security by providing initial estimates of soil fertility and global environmental concerns related to available soil P.

References


near-edge structure spectroscopy and chemical fractionation. Journal of Environmental Quality 32, 1819-1819.


Figure 1. The location of sites that supplied data for this study.
Figure 2. A flowchart showing how genetic algorithms function.
Table 1. Performance indices for each model ($R^2$, RMSE, regression method and P Value) with their equations.

<table>
<thead>
<tr>
<th>Methods</th>
<th>Locations</th>
<th>Training Set</th>
<th>Validation Set</th>
<th>Range</th>
<th>Regression method</th>
<th>$R^2$</th>
<th>RMSE</th>
<th>P-value</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ</td>
<td>The Whole World</td>
<td>129</td>
<td>55</td>
<td>96.0</td>
<td>Linear</td>
<td>0.69</td>
<td>14.59</td>
<td>0.00</td>
<td>NZ = 66.73+3.83E-02<em>CSI-0.73</em>VFS-0.11<em>CS+8.93</em>Fe+37.30<em>Al+1.58E-02</em>MN+0.16<em>Ca-4.29E-02</em>Mg-7.44*pH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-Linear</td>
<td>0.78</td>
<td>13.42</td>
<td>0.00</td>
<td>NZ = 61.86+1.55<em>CSI-2.31</em>VFS-37.26<em>Fe+91.39</em>Al-0.40<em>Ca+0.36</em>Mg-7.34<em>pH-3.67E-02</em>CSI^2+5.46E-02<em>VFS^2+17.11</em>Fe^2-23.05<em>Al^2+1.78E-02</em>Mn^2+5.92E-03<em>Ca^2-3.50E-03</em>Mg^2+0.15*pH^2</td>
</tr>
<tr>
<td></td>
<td>The USA</td>
<td>95</td>
<td>40</td>
<td>96.0</td>
<td>Linear</td>
<td>0.77</td>
<td>12.84</td>
<td>0.00</td>
<td>NZ = 20.64+0.13<em>C-0.67</em>VFS+0.83<em>MS-0.68</em>CS+0.37<em>OC+15.14</em>Fe+39.66*Al</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-Linear</td>
<td>0.82</td>
<td>11.48</td>
<td>0.00</td>
<td>NZ = 44.13+6.26E-02<em>C-4.48</em>VFS+1.12<em>MS-1.29</em>CS+2.02<em>OC-3.91</em>Fe+75.10<em>Al-3.54E-03</em>C^2+0.12<em>VFS^2-2.24E-02</em>MS^2+2.77E-02<em>CS^2-3.86E-02</em>OC^2+3.96<em>Fe^2-19.45</em>Al^2</td>
</tr>
<tr>
<td></td>
<td>The Other Locations</td>
<td>34</td>
<td>15</td>
<td>92.0</td>
<td>Linear</td>
<td>0.79</td>
<td>13.36</td>
<td>0.00</td>
<td>NZ = 96.69+6.41E-02<em>C+0.62</em>FSi+0.13<em>MS+1.63E-02</em>MN+12.91*pH</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-Linear</td>
<td>0.84</td>
<td>14.05</td>
<td>0.00</td>
<td>NZ = -5.45+1.12<em>C-1.37</em>FSi+0.43<em>MS+0.03</em>Mn+18.80<em>pH-9.20E-03</em>C^2+3.04E-02<em>FSi^2-1.14E-02</em>MS^2-1.81E-06<em>Mn^2-2.53</em>pH^2</td>
</tr>
<tr>
<td>Acid Oxalate</td>
<td>The Whole World</td>
<td>236</td>
<td>102</td>
<td>1195.6</td>
<td>Linear</td>
<td>0.40</td>
<td>165.50</td>
<td>0.00</td>
<td>Acid Oxal = 263.56-3.43<em>C-4.77</em>FSi+119.24<em>Fe+220.28</em>Al+5.07E-02*Mn</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-Linear</td>
<td>0.39</td>
<td>185.30</td>
<td>0.00</td>
<td>Acid Oxal = 59.74-2.18<em>C+5.03</em>FSi+198.87<em>Fe+568.73</em>Al+1.86E-02<em>Mn+1.09E-02</em>C^2-5.38E-02<em>FSi^2-3.84</em>Fe^2-232.38<em>Al^2+5.79E-06</em>Mn^2</td>
</tr>
<tr>
<td></td>
<td>The USA</td>
<td>214</td>
<td>91</td>
<td>5825.8</td>
<td>Linear</td>
<td>0.60</td>
<td>561.64</td>
<td>0.92</td>
<td>Acid Oxal = -96.20-5.12<em>Al+6.21</em>FSi+914.60<em>Al+0.23</em>Mn+0.77*Ca</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Non-Linear</td>
<td>0.42</td>
<td>296.78</td>
<td>0.92</td>
<td>Acid Oxal = -50.03-6.32<em>C+5.52</em>FSi+1063.51<em>Al+0.42</em>Mn+ ...</td>
</tr>
<tr>
<td>Location</td>
<td>Bray</td>
<td>Mehlich</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>----------</td>
<td>------</td>
<td>---------</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Indiana, Iowa, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin and Wyoming</td>
<td>Linear 0.62</td>
<td>Linear 0.23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Other Locations</td>
<td>28 12 705.1</td>
<td>83 35 86.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Acid Oxal = 39.19-4.80<em>C+1175.29</em>Al+8.33<em>Ca-4.08</em>Mg</td>
<td>Acid Oxal = 12.09-32.35<em>Al-0.41</em>Ca+24.58*K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Chile, China, Costa Rica, Denmark, Ghana, Guatemala, Poland and Puerto Rico</td>
<td>Non-Linear 0.85</td>
<td>Non-Linear 0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Whole World</td>
<td>Linear 0.27</td>
<td>Linear 0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray1 = 22.52+1.71E-03<em>C+7.90E-02</em>FSi+0.12<em>CSi+0.68</em>FS+0.72<em>CS+2.19</em>OC-5.10*pH</td>
<td>Bray1 = 12.09-32.35<em>Al-0.41</em>Ca+24.58*K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Canada, Ghana, Nigeria, Poland, Costa Rica and Guatemala</td>
<td>Non-Linear 0.31</td>
<td>Non-Linear 0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Other Locations</td>
<td>Linear 0.64</td>
<td>Linear 0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray1 = 131.27-0.70<em>C-0.29</em>FSi-1.27<em>CSi+1.27</em>VCS+2.40<em>OC+3.96</em>Mg-17.63*pH</td>
<td>Bray1 = -8.60+0.77<em>CSI+1.63</em>CS-1.49<em>VCS-1.15</em>OC-2.14<em>Mg+21.96</em>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The USA</td>
<td>Linear 0.64</td>
<td>Linear 0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray1 = 42.02-5.78*pH</td>
<td>Bray1 = 124.16-35.31<em>pH+2.56</em>pH^2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Alaska, Delaware, Hawaii, Illinois, Louisiana, Montana, North Carolina, Oklahoma, Pennsylvania, South Dakota, Tennessee, Vermont and Virginia</td>
<td>Non-Linear 0.28</td>
<td>Non-Linear 0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray1 = 124.16-35.31<em>pH+2.56</em>pH^2</td>
<td>Bray1 = 124.16-35.31<em>pH+2.56</em>pH^2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The Whole World</td>
<td>Linear 0.23</td>
<td>Linear 0.63</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray1 = 779.56-2.02<em>C+4.11</em>FSi-3.51<em>CSI+2.48</em>VCS+25.30<em>OC-21.12</em>Mg-247.63<em>pH+1.37E-02</em>C^2-9.95E-02<em>FSi+2.0</em>CSI+2-3.76E-02<em>VCS</em>2-5.64<em>OC</em>2+0.75<em>Mg</em>2+19.69*pH^2</td>
<td>Bray1 = -8.60+0.77<em>CSI+1.63</em>CS-1.49<em>VCS-1.15</em>OC-2.14<em>Mg+21.96</em>K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>The USA</td>
<td>Non-Linear 0.44</td>
<td>Non-Linear 0.44</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray1 = 12.09-32.35<em>Al-0.41</em>Ca+24.58*K</td>
<td>Bray1 = 12.09-32.35<em>Al-0.41</em>Ca+24.58*K</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Notes:**
- The formulae for Acid Oxal, Bray1, and Mehlich are shown in the table.
- The USA and Other Locations are categorized based on their geographical regions.
- The values for Bray1 are given in the form of linear and non-linear functions, along with their coefficients and R² values.
- The geographical regions are listed with their respective states, countries, and areas.
<table>
<thead>
<tr>
<th>Location</th>
<th>Number of Locations</th>
<th>Mean</th>
<th>Non-Linear</th>
<th>Linear</th>
<th>Non-Linear Equation</th>
<th>Linear Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>California, Connecticut, Delaware, Iowa, Kansas, Michigan, Mississippi, Nevada, New Mexico, North Dakota, Oregon, Vermont, West Virginia and Wyoming</td>
<td>12</td>
<td>76.6</td>
<td>0.62</td>
<td>13.29</td>
<td>0.044</td>
<td>Mehlich = -0.17-3.77E-02<em>CSi+1.46</em>CS-1.89<em>VCS+0.18</em>OC-3.82<em>Mg+40.47</em>K+0.02<em>CSi^2+2.48E-02</em>CS^2-1.58E-02<em>VCS^2-0.07</em>OC^2+4.30E-02<em>Mg^2-7.169</em>K^2</td>
</tr>
<tr>
<td>The Other Locations</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Costa Rica and Poland</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note: clay (C), fine silt (FSi), coarse silt (CSi), very fine sand (VFS), fine sand (FS), medium sand (MS), coarse sand (CS), very coarse sand (VCS), total organic carbon (OC), extractable iron (Fe), extractable aluminum (Al), extractable manganese (Mn), extractable calcium (Ca), extractable magnesium (Mg) and extractable potassium (K)
Table 2. Correlation (Pearson) between soil co-variates and extraction methods for different datasets. Significant differences are indicated as p < 0.05* and p <0.01**. ns, not significant

<table>
<thead>
<tr>
<th>Methods</th>
<th>Locations</th>
<th>C</th>
<th>FSi</th>
<th>CSI</th>
<th>VFS</th>
<th>FS</th>
<th>MS</th>
<th>CS</th>
<th>VCS</th>
<th>OC</th>
<th>PH</th>
<th>Ca</th>
<th>Mg</th>
<th>K</th>
<th>Fe</th>
<th>Al</th>
<th>Mn</th>
</tr>
</thead>
<tbody>
<tr>
<td>NZ</td>
<td>The Whole World</td>
<td>-</td>
<td>-</td>
<td>0.04*</td>
<td>-0.06*</td>
<td>-</td>
<td>-</td>
<td>-0.08*</td>
<td>-</td>
<td>-</td>
<td>-0.39**</td>
<td>-0.18**</td>
<td>-0.22**</td>
<td>-</td>
<td>0.61**</td>
<td>0.72**</td>
<td>0.30**</td>
</tr>
<tr>
<td></td>
<td>The USA</td>
<td>-0.07**</td>
<td>-</td>
<td>-0.07*</td>
<td>0.03**</td>
<td>0.03**</td>
<td>-</td>
<td>0.36**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.68**</td>
<td>0.81**</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Other Locations</td>
<td>0.35**</td>
<td>0.60**</td>
<td>-</td>
<td>-</td>
<td>-0.35**</td>
<td>-</td>
<td>-</td>
<td>-0.45**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td></td>
<td>-</td>
<td></td>
<td></td>
<td>0.61**</td>
</tr>
<tr>
<td>Acid Oxalate</td>
<td>The Whole World</td>
<td>-0.12*</td>
<td>0.27**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.47**</td>
<td>0.46**</td>
<td>0.20**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The USA</td>
<td>-0.03**</td>
<td>0.22**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.08**</td>
<td>-</td>
<td>-</td>
<td>0.57**</td>
<td>0.43**</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>The Other Locations</td>
<td>-0.23**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.21**</td>
<td>0.05**</td>
<td>-</td>
<td>0.66**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bray</td>
<td>The Whole World</td>
<td>-0.22**</td>
<td>-0.23**</td>
<td>-0.27**</td>
<td>-</td>
<td>0.36**</td>
<td>-</td>
<td>0.17**</td>
<td>0.13**</td>
<td>-0.20**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>The USA</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.34**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>The Other Locations</td>
<td>-0.42**</td>
<td>-0.31**</td>
<td>-0.28**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.09**</td>
<td>0.20**</td>
<td>-0.22**</td>
<td>-</td>
<td>-0.09**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Mehlich</td>
<td>The Whole World</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.28**</td>
<td>0.43**</td>
<td>-</td>
<td>-0.11**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>The USA</td>
<td>-</td>
<td>-</td>
<td>0.26**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.06**</td>
<td>-0.03**</td>
<td>-0.03**</td>
<td>-</td>
<td>0.13**</td>
<td>0.56**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>The Other Locations</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-0.60**</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

Note: clay (C), fine silt (FSi), coarse silt (CSI), very fine sand (VFS), fine sand (FS), medium sand (MS), coarse sand (CS), very coarse sand (VCS), total organic carbon (OC), extractable iron (Fe),
extractable aluminum (Al), extractable manganese (Mn), extractable calcium (Ca), extractable magnesium (Mg) and extractable potassium (K)