



OPEN

A Global Database of Soil Plant Available Phosphorus

DATA DESCRIPTOR

R. W. McDowell^{1,2}✉, A. Noble¹, P. Pletnyakov¹ & P. M. Haygarth³

Soil phosphorus drives food production that is needed to feed a growing global population. However, knowledge of plant available phosphorus stocks at a global scale is poor but needed to better match phosphorus fertiliser supply to crop demand. We collated, checked, converted, and filtered a database of c. 575,000 soil samples to c. 33,000 soil samples of soil Olsen phosphorus concentrations. These data represent the most up-to-date repository of freely available data for plant available phosphorus at a global scale. We used these data to derive a model ($R^2 = 0.54$) of topsoil Olsen phosphorus concentrations that when combined with data on bulk density predicted the distribution and global stock of soil Olsen phosphorus. We expect that these data can be used to not only show where plant available P should be boosted, but also where it can be drawn down to make more efficient use of fertiliser phosphorus and to minimise likely phosphorus loss and degradation of water quality.

Background & Summary

Soil phosphorus drives food production required to feed an increasing global population that is projected to reach 10 billion people by 2050¹. It has been estimated that an additional 500 million hectares of arable land will be required to feed this increased population unless phosphorus can be either better utilised by plants or applied more efficiently². Much of this efficiency will arise from local management solutions that only apply phosphorus fertilisers where they are needed³. However, knowledge of plant-available soil phosphorus stocks is poor, globally.

Some estimates have been made of global soil total phosphorus but only considers soils in their natural state, that is without the addition of fertilisers^{4,5}. Similarly, regional estimates exist of plant available soil phosphorus stocks using measured data^{6–8}. However, global estimates of plant available soil phosphorus stocks using measured data do not exist. Instead, global stocks have been estimated using models of factors such as plant uptake, weathering and global lithology data^{9–12} or via mass balance approaches^{2,13}. It is important to know where available soil phosphorus concentrations are adequate or deficient for optimal crop growth. This knowledge enables us to better match phosphorus fertiliser supply to crop demand and to suggest where excess plant available soil phosphorus can be drawn down^{11,14}. Here we present the first global database of freely available data on plant available soil phosphorus concentrations and use these data to create a global map and calculate the global stocks of plant available soil phosphorus stocks. We chose bicarbonate-extractable Olsen phosphorus¹⁵ as the measure of plant available soil phosphorus as it is the most widely used form, globally.

Methods

Data filtering and evaluation. Data ($n = 574,375$) of available soil phosphorus were obtained from 19 regional or global databases and published studies. These were chosen for their geographic spread and representativeness of a mix of developed and developing nations and where there was a clear process in place to ensure that data were of good quality (Table 1). Prior to modelling the data to estimate global Olsen phosphorus stocks, we adopted a multi-step process (Fig. 1) to produce a globally consistent dataset. The steps comprised (1) inspecting the data and filtering it for consistent analytical methods, units, and a limit of detection (set as 2 mg kg^{-1}); (2) filtering data to remove points lacking correct geo-referencing and those falling outside an acceptable time span (from 2000–2019); (3) converting values into Olsen phosphorus concentrations via established equations (Table 2), if necessary; and (4) filtering data to remove points from depths $>20 \text{ cm}$ and eliminating any duplicate values.

Step 1 Inspect data. When examining data, we determined that the soil extraction method was recorded, and that the phosphorus extraction relied on acceptable procedures. Measurements of phosphorus based on molybdenum blue colorimetry or ion chromatography were considered comparable and acceptable. Measurements

¹AgResearch, Lincoln Science Centre, Private Bag 4749, Christchurch, 8140, New Zealand. ²Faculty of Agriculture and Life Sciences, Lincoln University, Lincoln, P O Box 84, 7647, Christchurch, New Zealand. ³Lancaster Environment Centre, Lancaster University, Lancaster, LA1 4YQ, UK. ✉e-mail: richard.mcdowell@agresearch.co.nz

Data source (short name)	Number of samples ^a	Year ^b	Conversion of filtered and evaluated data into Olsen phosphorus data ^c	Coverage	References
International Soil Reference and Information Centre (ISRIC) World Soil Information	42,026 (2,776)	2010 (2010)	2,461 converted from Bray-I P, 129 converted from Mehlich-3 P	Global	40
Global dataset of plant-available P (Hou)	215 (176)	1982–2018 (1992)	176 converted from Resin-P	Global	5
Land Use/Land Cover Area Frame Survey (LUCAS) Topsoil Survey	19,969 (15,472)	2009 (2009)		Europe	41
National Cooperative Soil Survey (NCSS)	427,238 ^d (1,759)	1956–2018 (2002)	1,105 converted from Bray-I P, 370 converted from Mehlich-3 P	US and global	42
Soil database for land surface modelling (CTSDB)	1,096 (518) ^e	1980–1996 (1996)		China	18,43
Chinese Ecosystem Research Network (CERN)	8,106 (125)	1998–2010 (2009)		China	44
World Soil Information Service (WoSIS)	56,162 (6,822)	2009–2016 (2011)	4,515 converted from Bray-I P, 1,043 converted from Mehlich-3 P	Africa	6
Soil Resources of Russia (SRR)	4,961 (145)	2010–2019 (2014)	145 converted from Kirsanov ^f	Russia	45
Australian Soil Resources Information System (ASRIS)	15,525 (4,789)	2011 (2011)	4,209 converted from Bray-I P, 12 converted from Mehlich-3 P	Australia	46
New Zealand Soil database (NZSD)	243 (106)	1995–2009 (2004)		New Zealand	47
Miscellaneous (Misc)	253 (253)	1990–2017 (2011)	65 converted from Bray-I P, 50 converted from AB-DPTA	Bangladesh, Myanmar, Nepal, Pakistan, West Africa	16,48–55
All	575,794 (32,941)	1982–2019 (2009)	19,695 required no conversion	—	—

Table 1. List of data sources used to construct the map of the estimated global soil Olsen phosphorus (P) stock. ^aThe numbers in the parentheses refer to the filtered number of data. ^bThe values in the parentheses refer to the average sampling year of the filtered data. ^cNo conversion was necessary if blank; otherwise, please refer to Table 2 for the conversion details. ^dOriginal data were pre-filtered to the top 15 cm soil layer (the 2012 data points are then further filtered). ^eData sourced from eastern provinces (Anhui, Fujian, Guangdong, Guangxi, Hainan, Hebei, Jiangsu, Jiangxi, Shandong, Sichuan, and Tianjin) and prior to 1995 are removed. ^fNo publications exist to convert Kirsanov available P⁵⁶ into Olsen phosphorus⁵⁷. We employ the weighted mean equation for Bray-I as the closest approximation of the 0.2 N HCl extractant used in the Kirsanov test.

obtained with the stannous chloride method were excluded from the database. We also inspected the data for irregularities such as different units or different detection limits. The units were restricted to mg kg⁻¹, and volumetric data (mg L⁻¹) were excluded. Where detection limits were reported (2 mg kg⁻¹), minimum values were expressed as half the detection limit (1 mg kg⁻¹). Where detection limits were not reported, we inspected each data source for repeated low concentrations and assigned detection limits equivalent to half of the values that were repeatedly reported. Values at or below the detection limit comprised < 0.1% of the final database.

Step 2 Correct for space and time. We determined whether data points were correctly geo-referenced and occurred within an acceptable time span. To increase the likelihood that data points were correctly geo-referenced we excluded any data that were incorrectly reported or located in aquatic systems, glaciers, or permanent snowpacks. To generate an acceptable and consistent time span we restricted our data to the period from 2000 to 2019, except for three datasets relating to areas with unchanged land use. The first dataset was a global metanalysis dataset of soils under native land use (largely forestland)⁵ with a mean sampling year of 1992. As these samples were obtained from natural land uses, they were not expected to be influenced by anthropogenic phosphorus inputs. The second dataset involved 25 sites in Sahel and West African countries sampled in 1990¹⁶. Despite increases in green vegetation, land use intensification in these areas was very limited¹⁷. We therefore considered these soils to be representative of current practices. The third dataset included 17,920 values from the Second National Soil Farm Survey of China¹⁸ collected between 1980 and 1996. Major changes occurred in both the land use and land use intensity in eastern China, but not in western China, from 1986 to 2010^{19–21}. To account for likely changes in the soil Olsen phosphorus concentration, we excluded data from the second survey prior to 1995 along with data pertaining to eastern provinces (Anhui, Fujian, Guangdong, Guangxi, Hainan, Hebei, Jiangsu, Jiangxi, Shandong, Sichuan, and Tianjin). We retained data originating from the remaining provinces where no land use change or intensification was noted^{22–24}.

Step 3 Convert the data. We converted all data into soil Olsen phosphorus concentration data using regression equations suitable for Bray-I phosphorus, Resin-phosphorus, Kirsanov-phosphorus, AB-DPTA-phosphorus and for calcareous and non-calcareous soils²⁵ considering Mehlich-3 phosphorus (Table 2). The slopes and coefficients of these equations were weighted according to the number of data points in each dataset. We note that pH can strongly affect these conversions especially if soil tests are used inappropriately; for example, using the Bray-I

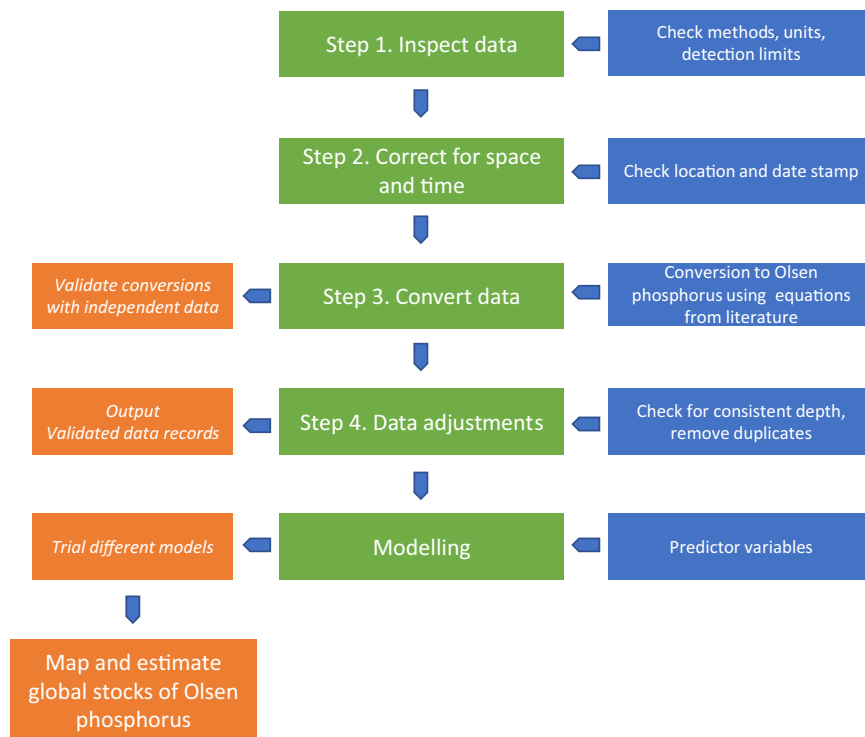


Fig. 1 Flowchart of the steps involved in filtering, evaluation, and modelling of soil Olsen phosphorus data. Note that the blue and orange boxes are sub tasks associated with each step and resulting outputs, respectively.

Soil test/ pH range	Soil characteristics	Number of samples	Jurisdiction	Slope	Intercept	R ²	Reference
Conversion from Mehlich-3^a							
4.3–6.8	Non-calcareous	91	US	0.33	6.9	0.88	⁵⁸
5.3–8.2	Non-calcareous	~180	US	0.47	1.2	0.79	⁵⁹
	Calcareous	~60	US	0.45	1.8	0.81	⁵⁹
3.8–8.6	Non-calcareous	66	Italy	0.70	−0.6	0.71	⁶⁰
	Calcareous	54	Italy	0.37	0.4	0.73	⁶⁰
Weighted mean							
	Non-calcareous	337		0.47	2.4	0.80	
	Calcareous	114		0.41	1.1	0.77	
Conversion from Bray-I							
5.3–7.45	Non-calcareous	~180	US	0.43	2.4	0.68	⁵⁹
6.0–8.0	Non-calcareous	165	US	0.68	3.9	0.95	⁶¹
4.3–6.8	Non-calcareous	91	US	0.30	2.9	0.85	⁵⁸
Weighted mean		436		0.49	3.1	0.82	
Conversion from AB-DPTA							
6.7–9.1	^b	35	India	1.81	4.1	0.50	⁶²
Conversion from Resin							
5.1–7.7	Non-calcareous	59	US	0.71	0.1	0.91	⁶³
	Calcareous	30	US			0.94	⁶⁴

Table 2. Regression equations to convert Mehlich-3, Bray-I and Resin phosphorus into Olsen phosphorus for calcareous and non-calcareous soils. To convert the data into Olsen phosphorus data, we use slope and intercept values calculated as weighted mean values for the calcareous and non-calcareous FAO soil types²⁵. ^aWe consider the conversion equation of Zbiral and Némec⁶⁵ to be too geographically limited (Czech Republic only) and too variable to be included. ^bUnknown. Soil characteristics were not listed, and no site-specific coordinates were available to independently check them.

phosphorus test will dissolve calcium phosphates that are sparingly available to plants. We excluded Bray-I P data from soils \geq pH 7 from our database. Moreover, while pH was found to have a minimal effect on conversions of Mehlich-3 phosphorus to Olsen phosphorus²⁶, nevertheless we used separate equations for calcareous

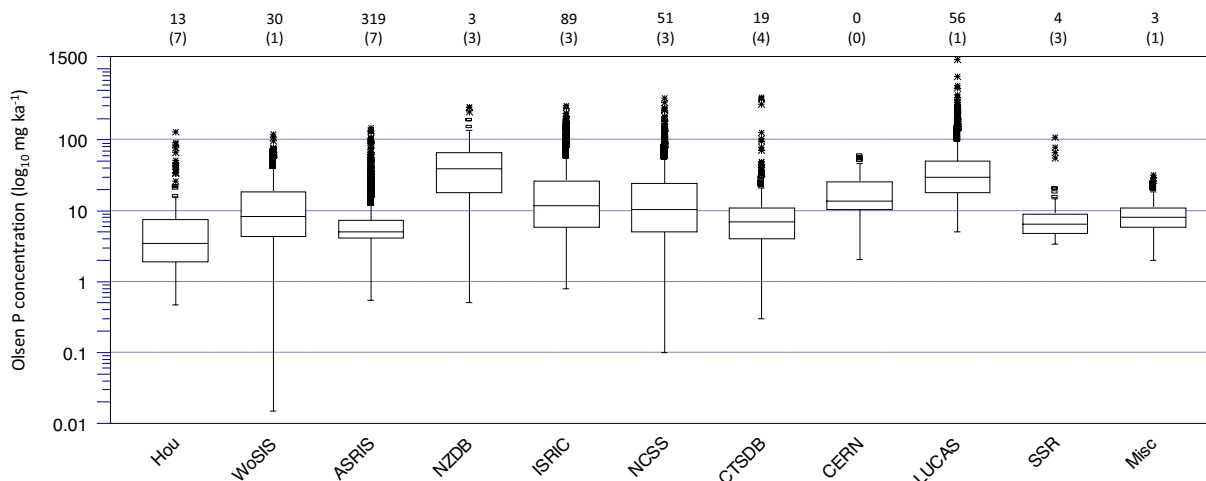


Fig. 2 Box plots showing the 25th, 50th and 75th percentiles (top, middle and bottom of each box), the upper and lower fences (the 75th and 25th percentiles plus and minus 1.5 times the interquartile range, respectively) and minor (>75th percentile but <upper fence) and major (>upper fence) Olsen P concentrations for each database. The values at the top indicate the number (and percentage in parentheses) of major outliers in each database.

and non-calcareous soils (Table 2). The proportions of the total sites converted from Bray-I, Resin, Kirsanov, AB-DPTA and Mehlich-3 phosphorus at this stage were 50.7%, 0.4%, 0.6%, 0.1% and 4.8%, respectively, but after step 4, the proportions changed to 37.4%, 0.5%, 0.4%, 0.2%, and 4.7%, respectively. Nearly 57% of all the samples required no conversion (please refer to the filtering and conversion tabs in Final_Filtered_Raw_OlsenP_Plus_Predictors.xls or Steps_1_to_4.csv²⁷).

Step 4 Adjustment to a consistent sampling depth and removal of duplicate values. Sites sampled at multiple depths were averaged to the top 20 cm, considering the proportion of a given sample within the top 20 cm and any variance in the bulk density at a certain depth²⁵ ($n = 11,756$). For instance, if a sample was collected at depths from 15–25 cm, the sample influenced the mean value only within the 0–20 cm depth interval by a quarter (assuming all the soil samples exhibited the same bulk density). We did not make any adjustment for stratification of Olsen phosphorus concentrations in the deeper soil sample. However, much less stratification of Olsen phosphorus occurs with depth owing to strong sorption of phosphorus by the topsoil²⁸. Where there were multiple concentrations for the same coordinates, we adopted the mean value ($n = 176$). Deeper samples and any duplicate values at a specific site and date were removed ($n = 15,791$).

Our final global dataset contained 33,102 values distributed across 89 countries, with a mean concentration of 26 mg kg^{-1} . Over our sampling period, the mean sampling year was 2009 (Table 1). The percentage of major outliers (calculated as 1.5 times the interquartile range plus the upper fence of each database) varied from zero to seven (Fig. 2). However, when examining the whole database, the percentage of major outliers was <1%. We therefore did not remove outliers from the final database.

Modelling. The filtered data ($n = 32,941$) were paired with predictor variables obtained from a wide variety of sources (Table 3). These predictor variables were chosen due to their high likelihood of influencing soil Olsen phosphorus stocks and included catchment characteristics, hydrological and climate parameters, land use, population, and ecological classifications^{6,29}. We extracted data for each predictor variable from the sources outlined in Tables 3, 4 at a resolution of 1 km^2 , resulting in 933,120,000 points per variable considering the global land mass.

Prior to statistical analysis, log-transformed Olsen phosphorus concentrations were confirmed as approximately normally distributed with the Shapiro-Wilk test. A range of models was trialled to predict Olsen phosphorus concentrations. However, to minimise the likelihood that models were being overfitted we conducted a principal components analysis on 17 variables that were likely autocorrelated, being produced on a monthly timestep (e.g., EVI, NDVI, precipitation, mean temperature, mean maximum temperature, and mean minimum temperature). These components explained 96.9% of the variance in the set of variables and were all highly significant ($P < 0.001$) in the first model tried (a simple linear model) and so were included in all our models. Following the simple linear model, we developed a mixed effects model, then a random forest model, followed by generalised additive model (GAM) fitted with the *mgcv*³⁰ procedure in R. Although the random forest model developed explained most of the most variance in the data, the computational requirements were too high for it to be applied on a global scale. We chose to implement the generalised additive model to predict log Olsen phosphorus concentrations, globally (Table 5).

During modelling we used 70% of the data to train the models, while the remaining 30% was reserved to evaluate model performance. However, after finding little difference in predictive power between models using 70% or all the data, we chose to create the final model based on all the data.

It was not possible to predict values for the countries not included in the training data (representing 27.8% of global area). However, through the modelling process, country (geopolitical boundary) was an important

Variable	Unit	Year measured	Source
Mean monthly precipitation	mm	2010	⁶⁶
Mean monthly temperature	°C	2010	⁶⁶
Mean monthly maximum temperature	°C	2010	⁶⁶
Mean monthly minimum temperature	°C	2010	⁶⁶
Bioclimatic index 1 = Annual mean temperature	°C	2009	⁶⁶
Bioclimatic index 2 = Mean diurnal range (mean monthly temperature (maximum temperature - minimum temperature))	°C	2009	⁶⁶
Bioclimatic index 3 = Isothermal conditions (BIO2/BIO7) (× 100)	—	2009	⁶⁶
Bioclimatic index 4 = Temperature seasonality (standard deviation × 100)	—	2009	⁶⁶
Bioclimatic index 5 = Maximum temperature of the warmest month	°C	2009	⁶⁶
Bioclimatic index 6 = Minimum temperature of the coldest month	°C	2009	⁶⁶
Bioclimatic index 7 = Annual temperature range (BIO5-BIO6)	°C	2009	⁶⁶
Bioclimatic index 8 = Mean temperature of the wettest quarter	°C	2009	⁶⁶
Bioclimatic index 9 = Mean temperature of the driest quarter	°C	2009	⁶⁶
Bioclimatic index 10 = Mean temperature of the warmest quarter	°C	2009	⁶⁶
Bioclimatic index 11 = Mean temperature of the coldest quarter	°C	2009	⁶⁶
Bioclimatic index 12 = Annual precipitation	mm	2009	⁶⁶
Bioclimatic index 13 = Precipitation of the wettest month	mm	2009	⁶⁶
Bioclimatic index 14 = Precipitation of the driest month	mm	2009	⁶⁶
Bioclimatic index 15 = Precipitation seasonality (coefficient of variation)	—	2009	⁶⁶
Bioclimatic index 16 = Precipitation of the wettest quarter	mm	2009	⁶⁶
Bioclimatic index 17 = Precipitation of the driest quarter	mm	2009	⁶⁶
Bioclimatic index 18 = Precipitation of the warmest quarter	mm	2009	⁶⁶
Bioclimatic index 19 = Precipitation of the coldest quarter	mm	2009	⁶⁶

Table 3. Climatic variables and the units, years and sources of the variables used to predict the Olsen phosphorus concentration.

Variable	Unit	Year measured	Source
Catchment area	km ²		⁶⁷
Enhanced vegetation index (EVI, monthly)	EVI at a 1-km ² resolution	2010	⁶⁸
Normalised difference vegetation index (NDVI, monthly)	NDVI at a 1-km ² resolution	2010	⁶⁸
Mean annual runoff	mm	2014	⁶⁹
Slope	%		⁶⁷
Altitude	m above sea level		⁶⁷
Development status	Yes/no		⁷⁰
Population density	Persons/km ²	2000	⁷¹
Subsoil group number retrieved from SoilGrids	—	2016	²⁵
Percentage of cropland	%	2000	²⁴
Percentage of bare land	%	2015	²³
Percentage of grassland or shrubland if not forestland, cropland or bare land	%	2015	^{22–24}
Biome	—	2016	⁷²
Country	—	2016	⁷⁰

Table 4. Biophysical and geographic variables and the units, years and sources of the variables used to predict the Olsen phosphorus concentration.

predictor (see: R_model_outputs.docx²⁷). To predict Olsen P concentrations for countries with no data, we randomly sampled 5% of each country in the dataset and renamed the country for those observations as “other” before rerunning the model. Thus, the “other” countries represented a weighted average of the countries present in the training data. This procedure may have biased the predictions for the “other” countries, as the model would be weighted towards countries with more training data, which may not be representative of those countries not represented in the training data. Users should be aware of this modelling fix and are advised to consult Country_counts.csv to judge the number of data points for each country.

Once models were run, predicted concentrations were back-transformed and corrected for the retransformation bias with the smearing estimate method³¹:

Model approach	R squared (%)	AIC	NSE
Simple linear model	51	77345	0.204
Mixed-effect model	48	78317	0.214
Generalised additive model	54	75458	0.172
Random forest	68	– ¹	0.051

Table 5. Approaches and performance metrics (Akaike Information Criterion, AIC; Nash Sutcliffe Efficiency, NSE) for each of the models tested. ¹AIC is not available for the Random Forest model as it has no log likelihood.

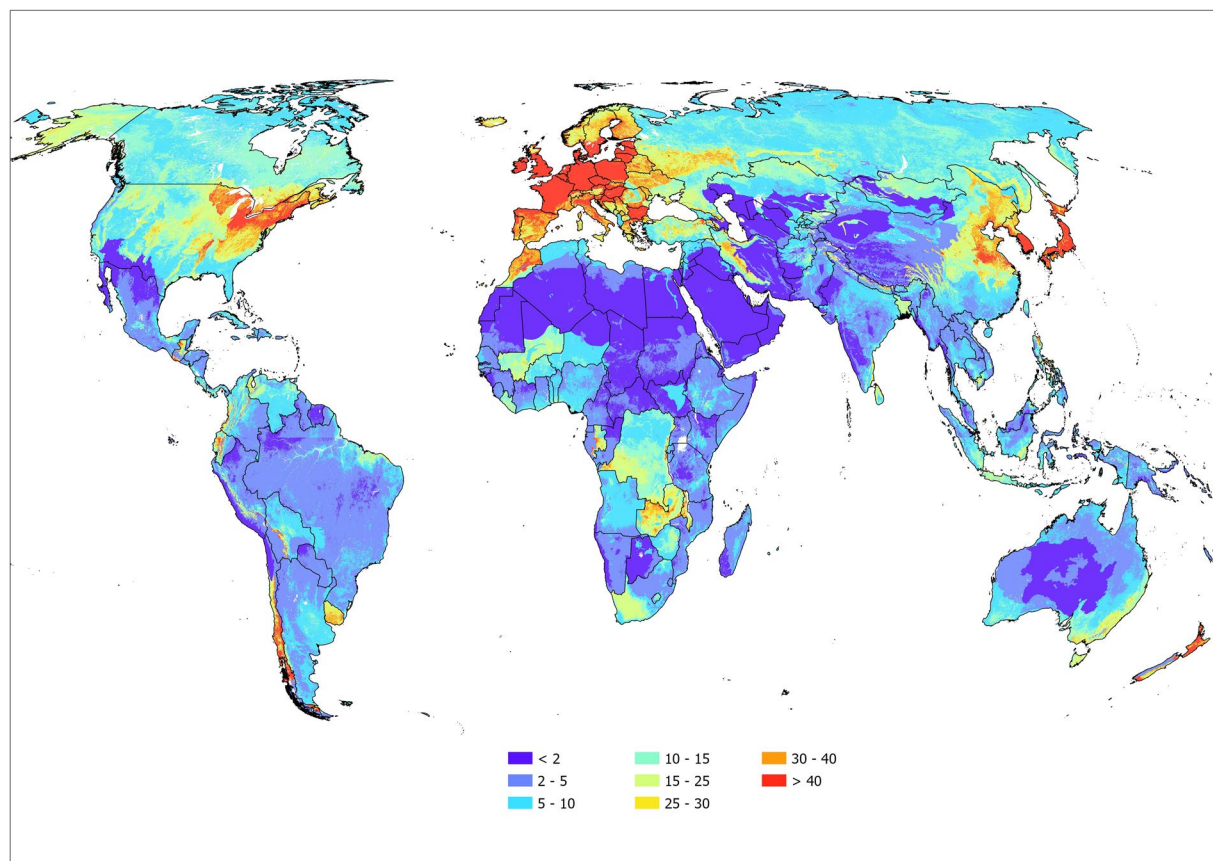


Fig. 3 Global topsoil Olsen phosphorus concentration (mg kg^{-1}). The mapped land parcels are plotted at a resolution of 1-km^2 and were calculated from a database containing ca. 575,000 soil samples of freely available data with a wide geographic coverage. An interactive version of this map, allowing users to discover predicted concentrations at selected points is available at: <https://world-olsen.agr.nz/>.

$$S = \frac{1}{n} \sum_{i=1}^n e^{\hat{\varepsilon}_i}$$

where ε_i denotes the residuals of the regression models. The correction factor (S) is applied over the whole range of predictions, as it is assumed that the residuals are homoscedastic.

The back-transformed predictions of Olsen phosphorus concentrations in topsoil were projected globally in ArcGIS. Raster grids were created at a spatial resolution of 0.025 degrees (ca. 1 km^2 near the equator), which corresponds to the coarsest grid cells associated with the input data, as listed in Tables 3 and 4.

Post-processing adjustments. Our preliminary modelling established that the biome and development status of a given country were important factors influencing the projection of Olsen phosphorus concentration in that country (see: *R_model_outputs.docx*²⁷). However, most of the data used to generate our global model were derived from developed regions and productive biomes. To determine if large areas were being modelled with a paucity of data, we split the database into biomes and whether a data point pertained to a developing or underdeveloped country. After inspecting the data, we found that five biomes suffered from a paucity of data ($n < 100$): deserts (within the desert and xeric shrubland biome); flooded grasslands and savannas and mangroves (developed); tropical and subtropical dry broadleaf forests, and montane grasslands and shrublands (underdeveloped).

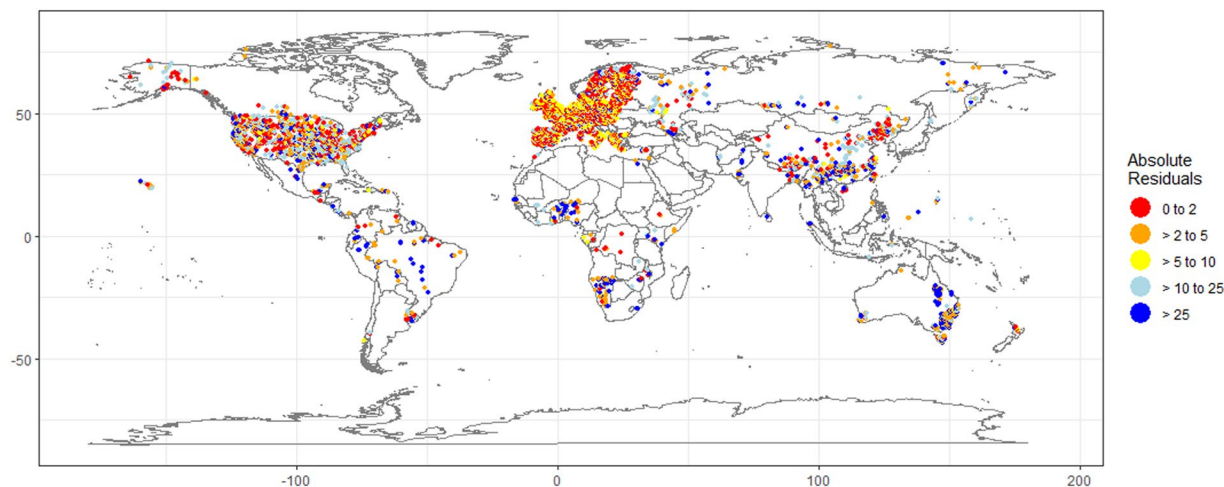


Fig. 4 Map of the residuals for each data point calculated as the difference between GAM predictions and the original value and classed the percentage difference into five classes: 0–2, 2.1–5, 5.1–10, 10.1–25 and >25%.

These biomes represented 12.6%, 0.9%, <0.1%, 3.7% and 0.5%, respectively, of the global land area, but are all largely unproductive. As previous studies have identified lower but stable soil phosphorus concentrations in unproductive biomes than in productive biomes, we used literature data to replace our modelled estimates of soil Olsen phosphorus. These biomes were assigned Olsen phosphorus concentrations (mg kg^{-1}) of 2.0³², 5.4³³, 3.5³⁴, 3.1³⁵ and concentrations between 1 and 3 depending on slope and elevation³⁶ (see: Final_Filtered_Raw_OlsenP_Plus_Predictors.xlsx; post modelling processing tab or Steps_1_to_4.csv²⁷).

Few data were available for South Africa. However, a prior spatial model of the mean soil available phosphorus (Bray-I phosphorus) in South Africa was available at the provincial level³⁷. This model was generated from >10,000 data points and performed better (for South Africa) than our model ($r^2 = 0.68$ cf. 0.54). Hence, we converted the modelled South African Bray-I phosphorus concentrations at a provincial level into Olsen phosphorus concentrations and applied it instead of ours.

Calculation and soil Olsen phosphorus stocks. To predict soil Olsen phosphorus stocks, the predicted concentration data (Fig. 3) were multiplied by bulk density data²⁵. Predicted Olsen concentration and bulk density data were assumed to cover 1-km² land parcels with a topsoil thickness of 20 cm. The mass in each pixel was calculated in kilotons. The predicted global stock (across 136 M km² of land) is estimated to be 318,618 kt ($\pm 21,985$ kt), while continental stocks are estimated to be: 47,847 ($\pm 3,301$), 86,474 ($\pm 4,483$), 84,401 ($\pm 7,279$), 60,517 ($\pm 4,176$), 13,374 (± 951), and 26,005 kt ($\pm 1,795$ kt), for Africa, Asia, Europe, North America, Oceania, and South America, respectively. Variation in stocks were calculated as the coefficient of variation using $\hat{c}V_{raw} = \sqrt{e^{s_{in}^2} - 1}$ for each estimate in the dataset³⁸ and the “metrumrg” package in R³⁹ (see also R_code_output.docx). The mean coefficient of variation was 0.069 or 6.9%. The stocks and area calculated for each continent (and country) are given in Stats_by_Continent.xlsx. The mean stock for countries was 1356 kt, ranging from <1 for small Caribbean Island nations to 39267 kt for the US.

On average the percentage difference between the predicted and observed data (i.e., residuals) was 14.9%. We classed the percentage differences into 0–2, 2.1–5, 5.1–10, 10.1–25 and >25%. The percentage of our predictions that were in each class was 14, 19, 20, 30 and 18%, respectively (see Final_Filtered_Raw_OlsenP_Plus_Predictors.xlsx Residuals by continent tab or Residuals_by_continent.csv²⁷). A map of the percentage differences is given in Fig. 4.

Data Records

The data and code used in modelling and outputs are available in Figshare²⁷. A list of the data files and outputs is available in Supplementary Table 1.

Technical Validation

Validating conversions to Olsen phosphorus. The conversions from Mehlich-3 P or Bray-I P to Olsen phosphorus were validated against the National Cooperative Soil Survey database, which contains observations of both Olsen phosphorus and Mehlich-3 P or Bray-I P for 97 samples. With the use of equations for either Bray-I P (Olsen phosphorus = $0.49 \times \text{Bray-I P} + 3.1$) or Mehlich-3 P (Olsen phosphorus = $0.47 \times \text{Mehlich-3 P} + 2.4$ for non-calcareous soils and Olsen phosphorus = $0.41 \times \text{Mehlich-3 P} + 1.1$ for calcareous soils), we predicted Olsen phosphorus concentrations and compared these estimates to measured Olsen phosphorus concentrations. The regression outputs ($P < 0.001$) indicate that the slope between the measured and predicted values approaches 1 (0.998 for Bray-I P and 0.928 for Mehlich-3 P; Fig. 5), suggesting that the equations are suitable for general use as a conversion tool.

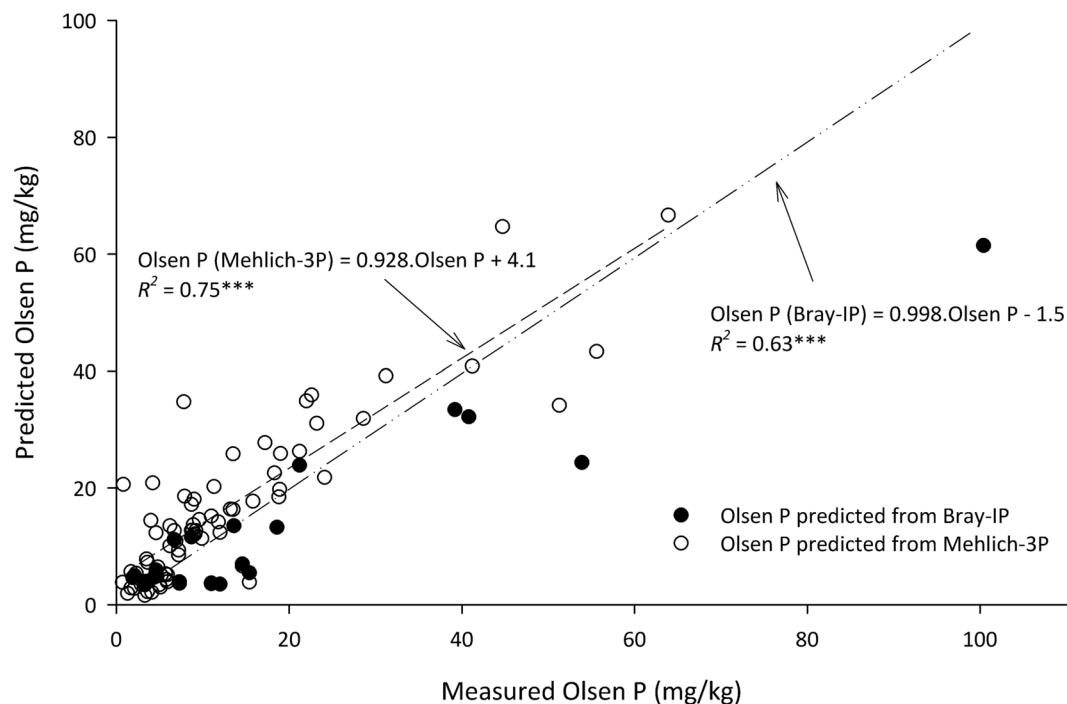


Fig. 5 Validation of Olsen phosphorus (P) predictions via the equations for Bray-IP and Mehlich-3P in Table 2 and independently sourced data from the NCSS. In addition to a significant fit ($P < 0.001$), and slope approaching 1, the Nash Sutcliffe Efficiency was >0.7 for each regression.

Validating soil Olsen phosphorus stocks. We compared our estimate of the topsoil Olsen phosphorus stock in sub-Saharan Africa to the previously modelled and published phosphorus stock in Sub-Saharan Africa. These published stocks were expressed as Mehlich-3 phosphorus⁶, so we converted Mehlich-3 phosphorus stocks to Olsen phosphorus stocks for 1-km² parcels of calcareous and non-calcareous soils using the equations provided in Table 2. Excluding the Saharan Desert²³, our modelled estimate was 36,875 kt of Olsen phosphorus for the 0–20 cm depth. After converting the published stock of Mehlich-3 phosphorus (estimated for the 0–30 cm depth) into Olsen phosphorus by the equations in Table 2, the Olsen phosphorus stock was 28,890 kt; our estimate was 27% greater, but 1% greater if the stock for the forested land was also removed.

Usage Notes

These data and the estimated global distribution of soil Olsen phosphorus stocks can be used to estimate where soil Olsen phosphorus is deficient or more than required for optimal crop growth. This can guide more efficient use of fertiliser stocks and can also indicate the potential for phosphorus loss from land to water, for example via erosion, which can impair water quality through eutrophication¹². However, it should be noted that such assessments are best done at a continental scale or at most a country or basin scale owing to the paucity of data in some regions, leading to high variability in the modelled stocks. It is advised that work requiring soil Olsen phosphorus stocks for policy at smaller scales therefore be supported by more localised sampling.

Code availability

The following code and outputs are available on-line²⁷:

- The R code and outputs of the efficacy and performance of the models employed to estimate the global Olsen phosphorus concentration from the predictor variables.
- Python code describing the filtering and post-processing steps involved in the use and analysis of the raw data and predictor variables to estimate global soil Olsen phosphorus concentration values and stocks.

Received: 6 September 2022; Accepted: 14 February 2023;

Published online: 07 March 2023

References

1. Lee, R. The Outlook for Population Growth. *Science* **333**, 569–573 (2011).
2. Mogollón, J. M. *et al.* More efficient phosphorus use can avoid cropland expansion. *Nature Food* **2**, 509–518 (2021).
3. Haygarth, P. M. & Rufino, M. C. Local solutions to global phosphorus imbalances. *Nature Food* **2**, 459–460 (2021).
4. He, X. *et al.* Global patterns and drivers of soil total phosphorus concentration. *Earth Syst. Sci. Data Discuss.* **2021**, 1–21 (2021).
5. Hou, E., Tan, X., Heenan, M. & Wen, D. A global dataset of plant available and unavailable phosphorus in natural soils derived by Hedley method. *Scientific Data* **5**, 180166 (2018).
6. Hengl, T. *et al.* Soil nutrient maps of Sub-Saharan Africa: assessment of soil nutrient content at 250 m spatial resolution using machine learning. *Nutrient Cycling in Agroecosystems* **109**, 77–102 (2017).

7. Ballabio, C. *et al.* Mapping LUCAS topsoil chemical properties at European scale using Gaussian process regression. *Geoderma* **355**, 113912 (2019).
8. Tóth, G., Guicharnaud, R.-A., Tóth, B. & Hermann, T. Phosphorus levels in croplands of the European Union with implications for P fertilizer use. *European Journal of Agronomy* **55**, 42–52 (2014).
9. Yang, X., Post, W. M., Thornton, P. E. & Jain, A. The distribution of soil phosphorus for global biogeochemical modeling. *Biogeosciences* **10**, 2525–2537 (2013).
10. Ringeval, B. *et al.* Phosphorus in agricultural soils: drivers of its distribution at the global scale. *Global Change Biol.* **23**, 3418–3432 (2017).
11. Sattari, S. Z., Bouwman, A. F., Giller, K. E. & van Ittersum, M. K. Residual soil phosphorus as the missing piece in the global phosphorus crisis puzzle. *Proceedings of the National Academy of Sciences* **109**, 6348–6353 (2012).
12. Alewell, C. *et al.* Global phosphorus shortage will be aggravated by soil erosion. *Nature Communications* **11**, 4546 (2020).
13. Zhang, J. *et al.* Spatiotemporal dynamics of soil phosphorus and crop uptake in global cropland during the 20th century. *Biogeosciences* **14**, 2055–2068 (2017).
14. Ros, M. B. H. *et al.* Towards optimal use of phosphorus fertiliser. *Scientific Reports* **10**, 17804 (2020).
15. Olsen, S. R., Cole, C. V., Watanbe, F. S. & Dean, L. A. in *United States Department of Agriculture. Circular No. 939* 19 (U.S. Dept. of Agriculture, Washington, D.C., 1954).
16. Manu, A., Bationo, A. & Geiger, S. C. Fertility status of selected millet producing soils of West Africa with emphasis on phosphorus. *Soil Sci.* **152**, 315–320 (1991).
17. Olsson, L., Eklundh, L. & Ardö, J. A recent greening of the Sahel—trends, patterns and potential causes. *J. Arid Environ.* **63**, 556–566 (2005).
18. Ministry of Agriculture of the People's Republic of China. *The second national soil survey farm fertility database (1980–1996)*, <http://soil.geodata.cn/data/datadetails.html?dataguid=159793232590769&docId=24> (2019).
19. Deng, X., Huang, J., Rozelle, S. & Uchida, E. Cultivated land conversion and potential agricultural productivity in China. *Land Use Policy* **23**, 372–384 (2006).
20. Fan, Y. *et al.* Entropies of the Chinese Land Use/Cover Change from 1990 to 2010 at a County Level. *Entropy* **19**, e19020051 (2017).
21. Bai, Z. *et al.* China's livestock transition: Driving forces, impacts, and consequences. *Science Advances* **4**, eaar8534 (2018).
22. Hansen, M. C. *et al.* High-Resolution Global Maps of 21st-Century Forest Cover Change. *Science* **342**, 850–853 (2013).
23. Di Miceli, C., Carroll, M., Sohlberg, R., Kim, D. & Townshend, J. MOD44B MODIS/Terra Vegetation Continuous Fields Yearly L3 Global 250m SIN Grid V006. NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MOD44B.006> (2015).
24. U.S. Department of the Interior & NASA. *Global Food Security-support Analysis Data 30 meter (GFSAD30) Cropland Extent*, <https://lpdac.usgs.gov/news/release-of-gfsad-30-meter-cropland-extent-products/> (2017).
25. Hengl, T. *et al.* SoilGrids250m: Global gridded soil information based on machine learning. *PLOS ONE* **12**, e0169748 (2017).
26. Steinfurth, K., Hirte, J., Morel, C. & Buczko, U. Conversion equations between Olsen-P and other methods used to assess plant available soil phosphorus in Europe – A review. *Geoderma* **401**, 115339 (2021).
27. McDowell, R. W., Pletnyakov, P., Noble, A. & Haygarth, P. M. A Global Database of Soil Plant Available Phosphorus. *figshare* <https://doi.org/10.6084/m9.figshare.14241854> (2023).
28. Sharpley, A. N. Depth of surface soil-runoff interaction as affected by rainfall, soil slope, and management. *Soil Sci. Soc. Am. J.* **49**, 1010–1015 (1985).
29. Stutter, M. I. *et al.* Land use and soil factors affecting accumulation of phosphorus species in temperate soils. *Geoderma* **257**, 29–39 (2015).
30. Wood, S. N. *Generalized Additive Models: An Introduction with R*. 2nd Edition edn, 496 (CRC Press, 2017).
31. Duan, N. Smearing Estimate: A Nonparametric Retransformation Method. *Journal of the American Statistical Association* **78**, 605–610 (1983).
32. Issaka, R. N., Masunaga, T., Kosaki, T. & Wakatsuki, T. Soils of inland valleys of West Africa. *Soil Sci. Plant Nutr.* **42**, 71–80 (1996).
33. Lavado, R. S. & Jorge, O. S. & Patricia, N. H. Impact of Grazing on Soil Nutrients in a Pampean Grassland. *Journal of Range Management* **49**, 452–457 (1996).
34. Boto, K. G. & Wellington, J. T. Soil Characteristics and Nutrient Status in a Northern Australian Mangrove Forest. *Estuaries* **7**, 61–69 (1984).
35. Liu, X. *et al.* Available Phosphorus in Forest Soil Increases with Soil Nitrogen but Not Total Phosphorus: Evidence from Subtropical Forests and a Pot Experiment. *PLOS ONE* **9**, e88070 (2014).
36. Bhandari, J. & Zhang, Y. Effect of altitude and soil properties on biomass and plant richness in the grasslands of Tibet, China, and Manang District, Nepal. *Ecosphere* **10**, e02915 (2019).
37. Khuthadzo, M. Phosphorus status. ARC-ISCW, Department of Agriculture, Land Reform and Rural Development <http://daffarcgis.nda.agric.za/portal/home/item.html?id=f62859757f4249a8bc71292083b18a2d> (2003).
38. Koopmans, L. H., Owen, D. B. & Rosenblatt, J. I. Confidence intervals for the coefficient of variation for the normal and log normal distributions. *Biometrika* **51**, 25–32 (1964).
39. Bergsma, T. T. *et al.* Facilitating pharmacometric workflow with the metrumrg package for R. *Comput. Methods Programs Biomed.* **109**, 77–85 (2013).
40. Batjes, N. H. Overview of soil phosphorus data from a large international soil database. 56 (Wageningen and ISRIC - World Soil Information, Wageningen, The Netherlands, 2011).
41. Toth, G., Jones, A. & Montanarella, L. LUCAS Topsoil Survey methodology, data and results. **141** (Publications Office of the European Union, Luxembourg, 2013).
42. National Cooperative Soil Survey. National Cooperative Soil Survey Characterization Database. (2020).
43. Shangguan, W. *et al.* A China data set of soil properties for land surface modeling. *Journal of Advances in Modeling Earth Systems* **5**, 212–224 (2013).
44. Chinese Ecosystem Research Network 1998–2010. *Soil nutrient data of major farmland ecosystems in China (1990–2006)* <http://soil.geodata.cn/data/datadetails.html?dataguid=118007709206684&docId=31> (2019).
45. Ministry of Agriculture of the Russian Federation. *United State Register of Soil Resources of Russia Version 1.0*, <http://egrpr.esoil.ru/> (2019).
46. McKenzie, N. J., Jacquier, D. W., Maschmedt, D. J., Griffin, E. A. & Brough, D. M. The Australian Soil Resource Information System (ASRIS) Technical Specifications. Revised Version 1.6, June 2012. (CSIRO, Canberra, Australia, 2012).
47. Taylor, M. D., Kim, N. D., Hill, R. B. & Chapman, R. A review of soil quality indicators and five key issues after 12 yr soil quality monitoring in the Waikato region. *Soil Use. Manage.* **26**, 212–224 (2010).
48. Talpur, N. A. *et al.* Soil fertility mapping of chilli growing areas of Taluka Kunri, Sindh, Pakistan. *Sindh University Research Journal (Science Sries)* **48**, 547–552 (2016).
49. Buerkert, A., Bationo, A. & Piepho, H.-P. Efficient phosphorus application strategies for increased crop production in sub-Saharan West Africa. *Field Crops Res.* **72**, 1–15 (2001).
50. Tsado, P. A., Osunde, O. A., Igwe, C. A., Adeboye, M. K. A. & Lawal, B. A. Phosphorus sorption characteristics of some selected soil of the Nigerian Guinea Savanna. *International Journal of AgriScience* **2**, 613–618 (2012).
51. Panday, D., Maharjan, B., Chalise, D., Shrestha, R. K. & Twanabasu, B. Digital soil mapping in the Bara district of Nepal using kriging tool in ArcGIS. *PLoS one* **13**, e0206350–e0206350 (2018).

52. Tin Maung, A. *Developing sustainable soil fertility in southern Shan State of Myanmar: a thesis presented in partial fulfilment of the requirements for the degree of Doctor of Philosophy in Soil Science at Massey University, Palmerston North, New Zealand* Doctor of Philosophy (Ph.D.) thesis, Massey University, (2001).
53. Guppy, C., Win, S. S., Win, T., Thant, K. M. & Phyo, K. N. in *Myanmar soil fertility and fertilizer management* 101–107 (International Fertilizer Development Center and the Division of Soil Science, Water Utilization and Agricultural Engineering, Department of Agricultural Research, Ministry of Agriculture, Livestock and Irrigation, Yezin, Nay Pyi Taw, Myanmar, 2017).
54. Khan, S., Khan, Q. U., Khan, M. J., Khatak, S. G. & Khan, A. A. Soil phosphate sorption characteristics of selected calcareous soil series of Southern Punjab, Pakistan. *Journal of Environmental Science and Management* **21-1**, 1–7 (2018).
55. Shil, N., Saleque, M. A., Islam & Jahiruddin, M. Soil fertility status of some of the intensive crop growing areas under major agroecological zones of Bangladesh. *Bangladesh Journal of Agricultural Research* **41**, 735–757 (2016).
56. Ginzburg, K. E. in *Agrochemical methods for soil research* (ed A. V. Sokolov) 106–191 (Nauka, 1975).
57. Ulen, B., Djodjic, F., Buciene, A. & Masauskiene, A. in *Ecosystem Health and Sustainable Agriculture 1: Sustainable Agriculture* (ed C. Jakobsson) Ch. 9, 82–101 (The Baltic University Programme, 2009).
58. Wolf, A. M. & Baker, D. E. Comparisons of soil test phosphorus by Olsen, Bray I, Mehlich I and Mehlich III methods. *Communications in Soil Science & Plant Analysis* **16**, 467–484 (1985).
59. Mallarino, A. P. in *Proceedings of the twenty-fifth North Central extension-industry soil fertility conference* (ed G. Rehm) 96–101 (Potash & Phosphate Institute, Manhattan, KS 1995).
60. Buondonno, A., Coppola, E., Felleca, D. & Violante, P. Comparing tests for soil fertility: 1. Conversion equations between Olsen and Mehlich 3 as phosphorus extractants for 120 soils of south Italy. *Commun. Soil Sci. Plant Anal.* **23**, 699–716 (1992).
61. Malo, D. D. & Gelderman, R. H. Portable soil test laboratory results compared to standard soil test values. *Communications in Soil Science & Plant Analysis* **15**, 909–927 (1984).
62. Malathi, P. & Stalin, P. Evaluation of AB - DTPA extractant for multnutrients extraction in soils. *International Journal of Current Microbiology and Applied Sciences* **7**, 1192–1205 (2018).
63. Mallarino, A. P. & Atia, A. M. Correlation of a Resin Membrane Soil Phosphorus Test with Corn Yield and Routine Soil Tests. *Soil Sci. Soc. Am. J.* **69**, 266–272 (2005).
64. Nesse, P., Grava, J. & Bloom, P. R. Correlation of several tests for phosphorus with resin extractable phosphorus for 30 alkaline soils. *Commun. Soil Sci. Plant Anal.* **19**, 675–689 (1988).
65. Zbíral, J. & Némec, P. Comparison of Mehlich 2, Mehlich 3, CAL, Egner, Olsen and 0.01 M CaCl₂ extractants for determination of phosphorus in soils. *Commun. Soil Sci. Plant Anal.* **33**, 3405–3417 (2002).
66. Fick, S. E. & Hijmans, R. J. WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology* **37**, 4302–4315 (2017).
67. United States Department of the Interior - United States Geological Survey. HydroSHEDS. (U.S. Dept. of the Interior, U. S. Geological Survey, Washington D.C., 2008).
68. Didan, K. MOD13A3 MODIS/Terra vegetation Indices Monthly L3 Global 1km SIN Grid V006. NASA EOSDIS Land Processes DAAC <https://doi.org/10.5067/MODIS/MOD13A3.006> (2015).
69. Ghiggi, G., Humphrey, V., Seneviratne, S. I. & Gudmundsson, L. GRUN: an observation-based global gridded runoff dataset from 1902 to 2014. *Earth Syst. Sci. Data* **11**, 1655–1674 (2019).
70. The World Bank. *Table 3.2 World Development Indicators: Agricultural Inputs*, <http://wdi.worldbank.org/table/3.2> (2015).
71. Ryan, J. *et al.* in *Advances in Agronomy* Vol. **114** (ed D. L., Sparks) 91–153 (Academic Press, 2012).
72. Dinerstein, E. *et al.* An Ecoregion-Based Approach to Protecting Half the Terrestrial Realm. *Bioscience* **67**, 534–545 (2017).

Acknowledgements

The authors thank those individuals who supplied the data in published reports and journal articles or stored data in open repositories. These individuals are cited in the Supplementary Data files. R.W.M., P.P. and A.N. thank the New Zealand Our Land and Water National Science Challenge for the support provided while generating the data and writing this manuscript (Ministry for Business, Innovation and Employment contract C10X01507).

Author contributions

R.W.M. and P.M.H. conceived the idea and wrote the manuscript. P.P. conducted the geographic analysis. A.N. and R.W.M. developed and performed the statistical analysis. R.W.M. sourced and collated the raw data.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41597-023-02022-4>.

Correspondence and requests for materials should be addressed to R.W.M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023