Biogeosciences Discussions

1   Global patterns and drivers of phosphorus pools in natural soils

2   Xianjin He[1]; Laurent Augusto[2]; Daniel S. Goll[3]; Bruno Ringeval[2]; Ying-Ping Wang[4]; Julian Helfenstein[5,]

3   [6]; Yuanyuan Huang[4]; Enqing Hou[7]

4   1 Key Laboratory of the Three Gorges Reservoir Region's Eco-Environment, Ministry of Education, Chongqing University,

5   Chongqing, China

6   2 ISPA, Bordeaux Sciences Agro, INRAE, F-33140, Villenave d'Ornon, France

7   3 Université Paris Saclay, CEA-CNRS-UVSQ, LSCE/IPSL, Gif sur Yvette, France

8   4 CSIRO Oceans and Atmosphere, Aspendale, Vic., Australia

9   5 Agroscope, 8046 Zürich, Switzerland

10   6 Soil Geography and Landscape Group, Wageningen University, 6700 AA Wageningen, The Netherlands

11   7 Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden,

12   Chinese Academy of Sciences, Guangzhou, China

13

14   *Corresponding to*: Daniel S. Goll (*dsgoll123@gmail.com*); Enqing Hou (*houeq@scbg.ac.cn*).

15

16   **Abstract.** Most phosphorus (P) in soils is unavailable for direct biological uptake as it is locked within primary or secondary

17   mineral particles, adsorbed to mineral surfaces, or immobilized inside of organic material. Deciphering the composition of

18   different P pools in soil is critical for understanding P bioavailability and its underlying dynamics. However, widely used

19   global estimates of different soil P pools are based on a dataset containing few measurements in which many regions or soil

20   types are unrepresented. This poses a major source of uncertainty in assessments that rely on these estimates to quantify soil P

21   constraints on biological activity controlling global food production and terrestrial carbon balance. To address this issue, we

22   consolidated a database of six major soil P pools containing 1857 entries from globally distributed (semi-)natural soils and 11

23   related environmental variables. The P pools (labile inorganic P (Pi), labile organic P (Po), moderately labile Pi, moderately

24   labile Po, primary mineral P, and occluded P) were measured using a sequential P fractionation method. Using the database,

25   we trained random forest regression models for each of the P pools and captured observed variation with R^2 higher than 60%.

26   We identified total soil P concentration as the most important predictor of all soil P pool concentrations, except for primary

27   mineral P concentration, which is primarily controlled by soil pH. When expressed in relative concentrations (*i.e.,* as a

28   proportion of total P), the model showed that soil pH is the most important predictor for proportions of all soil P pools, except

29   for labile Pi proportion, which is primarily controlled by soil depth. Using the trained random forest models, we predicted soil

30   P pools' distributions in natural systems at a resolution of $0.5° × 0.5°$. Our global maps of different P pools in soils as well as

31   the pools' underlying drivers can inform assessments of the role of natural P availability for ecosystem productivity, climate

32   change mitigation, and the functioning of the Earth system.

33

34

Biogeosciences

Discussions

1 Introduction

Phosphorus (P) is a key nutrient limiting plant growth across a wide range of ecosystems (Augusto et al., 2017; Elser et al., 2007; Hou et al., 2020). Soil is typically the major P source for plants in natural terrestrial ecosystems (Weihrauch and Opp, 2018). P supplied by the soil plays a vital role in determining the structures, functions, and processes in terrestrial ecosystems (Peltzer et al., 2010; Wardle et al., 2004). For example, soil P availability imposes a major constraint on plant productivity in terrestrial ecosystems worldwide (Augusto et al., 2017; Ellsworth et al., 2022; Elser et al., 2007; Hou et al., 2020; Hou et al., 2021) and affects modeled projections of terrestrial carbon cycle responses to climate change and increasing atmospheric carbon dioxide concentrations (Cunha et al., 2022; Fleischer et al., 2019; Goll et al., 2012). The size of soil P stocks is large compared to annual plant P requirements (Wang et al., 2018) and the amount of P stored in vegetation (Wang et al., 2018; Zhang et al., 2021). However, only a small proportion of soil P can be directly taken up by plants (Morel et al., 2014), with most P tightly sorbed to soil minerals, organic compounds, or organo-mineral complexes with a turnover time of centuries to millennia or longer (Helfenstein et al., 2020; Vitousek et al., 2010). Consequently, vegetation growth is often limited by P availability in ecosystems across the globe (Vitousek et al., 2010; Wardle et al., 2004). For these reasons, the investigation of P dynamics and P bioavailability in the soil requires the identification and separation of different soil P pools (Crews et al., 1995; Walker and Syers, 1976).

Our knowledge of the various pools of P occurring in the soil is based on a limited number of chronosequence studies that investigated how P is cycled during pedogenesis (Crews et al., 1995; Walker and Syers, 1976). These studies revealed that chemical weathering results in the release of P from primary minerals, after which it can be converted to organic P through biological uptake, sorbed to soil particles, or occluded within secondary minerals. The most commonly used procedures for the sequential fractionation of P in soils were developed by Hedley et al. (1982) and later modified by Tiessen and Moir (1993). In addition to forming the basis for modeling soil P dynamics, these procedures yield operationally defined pools that are used to assess soil fertility and soil development (Wang et al., 2010; Wang et al., 2022). Pools that are commonly considered are resin extractable P; 0.5 M NaCO$_3$ extractable inorganic P (Pi) and organic P (Po); 0.1 M NaOH extractable Pi and Po; 1 M HCl-extractable P; and the occluded pool, which is composed of P remaining after extraction (see section 2.1). Several studies have called the validity of sequential extractions into question, pointing out that, while it is often assumed that pools from sequential extractions contain distinct forms of P, the reality is much more complex (Condron and Newman, 2011; Gu and Margenot, 2021; Klotzbücher et al., 2019). Nevertheless, radioisotope tracer experiments show that sequentially extracted pools have distinct P exchange behaviors that result in significantly different turnover times (Buendía et al., 2010; Bünemann et al., 2004; Helfenstein et al., 2021; Helfenstein et al., 2020; Vu et al., 2010).

Numerous studies have used data from P fractionations to explore drivers of spatial differences in soil P pools from local to global scales (*e.g.*, Brucker and Spohn, 2019; Hou et al., 2018a; Yang and Post, 2011; Chen et al., 2015). Yang and Post (2011) compiled Hedley P pools data from 178 soil samples to explore P dynamics along a soil development gradient (Yang and Post, 2011). Their results generally supported the conceptual model proposed by Walker and Syers (1976): the gradual decrease of primary mineral-bound P; the continual increase and eventual dominance of occluded P; and the overall decrease of total P as pedogenesis progresses. However, this model was challenged by Yang and Post (2011), who found that labile Pi and moderately labile Pi (non-occluded P in Walker and Syers' model) formed a significant fraction of total P at every stage of pedogenic development. Augusto et al. (2017) aggregated 1684 measurements of P pools that were taken from across the globe using the Hedley fractionation method. This work found that total P content was a main factor determining the concentrations of labile Pi and organic P pools (Augusto et al., 2017). Almost concomitantly, Hou et al. (2018) used a global dataset compiled from analyses of 802 soil samples to examine climate effects on the soil P cycle and P availability and found that soil labile Pi concentration decreased with increasing mean annual temperature and precipitation (Hou et al., 2018a). Although those studies advanced our understanding of factors controlling the size of various soil P pools, their focus was largely contained to the effects of climatic factors or soil weathering stage on a few select P pools, mainly labile Pi, and organic P. Thus, we still lack

78   a comprehensive understanding of the relationships between environmental drivers and the various soil P pools at a global

79   scale.

80   Despite significant efforts to synthesize global Hedley soil P pool data, to our knowledge, only one set of global estimates

81   across natural terrestrial ecosystems exists, and this work was based on the upscaling of measurements taken from only 178

82   samples (Yang et al., 2013). These global estimates and associated maps have been used to explore global patterns of soil P

83   supply and to estimate P availability in natural and managed systems (*e.g.*, Ringeval et al., 2017; Sun et al., 2017). They have

84   also been used to calibrate or initialize a range of global P models (Wang et al., 2010; Yang et al., 2014). However, the poor

85   global coverage of the underlying data introduces significant uncertainty, potentially resulting in misinformed model

86   predictions and assessments.

87   We recently developed a new global map of soil total P concentrations and explored the underlying drivers, taking

88   advantage of improved data availability and the use of non-linear statistical modeling (He et al., 2021). Here, we constructed

89   a database of soil P pools in 1857 globally distributed (semi-)natural soils collected from 274 published studies, one order of

90   magnitude larger than the dataset used by Yang et al. (2013) (see comparison in Fig. S1). Using this database, we trained

91   random forest models to capture observed variations in Hedley P pool concentrations at the site level with two aims: (1) to

92   quantify the relative importance of different drivers of spatial variation in each soil P pool and (2) to develop global distribution

93   maps of various P pools at a spatial resolution of $0.5° \times 0.5°$ using the calibrated random forest regression model.

94

2 Material and Methods

2.1 Soil P fractionation terminology and procedure

97   The method developed by Hedley et al. (1982) and modified by Tiessen and Moir (1993) (Fig. S1) is the most commonly

98   used procedure for the sequential chemical fractionation of P in soils. This method exploits differences in solubility to separate

99   different 'forms' of P occurring in the soil. Though it cannot be used to identify specific discrete P compounds in the soil, this

100   approach has proven indispensable for the study of soil P cycling and, as such, is widely used (Condron and Newman, 2011;

101   Klotzbücher et al., 2019; Barrow et al., 2021). Here, we use the word 'pool' to indicate the concentrations quantified in each

102   step during sequential fractionation. We use the word 'proportion' to represent the size of a pool relative to total P.

103   There is disagreement about how to interpret the different pools yielded by sequential fractionation (Gu et al. 2019;

104   Barrow et al., 2021; Klotzbücher et al., 2019; Condron and Newman, 2011; Helfenstein et al., 2020). Here, we adopt a widely

105   used regime for understanding these pools: The resin Pi pool represents the soil soluble Pi pool, which is immediately

106   accessibly to plants; The HCO_3^- Pi pool can be released by ligand exchange with bicarbonate ions; This pool is available to

107   plants and persists for only short periods (*e.g.*, a growing season); Due to their functional similarity, the resin and HCO_3^- Pi

108   pools can be combined and used as an index of labile inorganic P (*i.e.,* 'available' P); The HCO_3^- Po pool represents labile Po

109   that can be utilized by plants after mineralization. The OH^- P (Pi and Po) pools indicate moderately labile P that is bound to

110   both amorphous and crystalline Al and Fe. This pool represents P that is moderately available to plants; The 1 M HCl Pi pool

111   represents primary mineral P that is bound to calcium and that can be utilized by plants after it is released by weathering; And

112   other P pools, such as residual P, are least available to plants due to their particularly low solubility.

113   To integrate data from studies that use different interpretations, we consider a set of six simplified P pools (Fig. S3):

114   labile Pi, labile Po, moderately labile Pi, moderately labile Po, primary mineral P, and occluded P. Labile Pi includes the resin

115   Pi and HCO_3^- Pi pools; labile Po and moderately labile Po are organic pools extracted by carbonate and NaOH, respectively;

116   moderately labile Pi is the NaOH Pi fraction; primary mineral P represents the 1 M HCl Pi pool; and occluded P includes any

117   remaining P (Hou et al., 2018b).

118   We collected, filtered, and processed soil P pool data (see section 2.2.) from the literature (Supplementary material).

119   First, we added all measured P pools together to calculate total soil P, unless any pools were missing. In this case, we instead

120   used the measured value of total soil P presented in that paper. Second, if phosphate was extracted using deionized water before

121   the resin P extraction step, the labile Pi pool includes both resin and aqueous P. If the extraction procedure began by using

122   sodium bicarbonate solution instead of resin P, we classified HCO_3^- Pi as labile Pi. Third, the labile Po pool and the moderately

123   labile Po pool represent the HCO_3^- Po and NaOH pools, respectively. The raw data contained other organic P pools (*e.g.,* Po

124   extracted by sonication and NaOH or by hot, concentrated HCl) which we included as part of occluded P. Fourth, if occluded

125   P was not reported, we calculated this pool's concentration by subtracting the sum of the five other pools from total P.

126   **2.2 Data source and processing**

127   We collected soil P pool data by aggregating all the publications that cited either one of two main references dedicated

128   to Hedley's method (Hedley et al., 1982; Tiessen and Moir, 1993). We included all studies that reported data from (semi-)

129   natural soils that supported primary vegetation or that had been reforested with a stand older than 10 years and no documented

130   history of P fertilization. We excluded observations taken from pot experiments, mine zones, and intertidal zones, as P pools

131   in these soils could be affected by factors different from those influencing (semi-) natural soils. Despite our best efforts, we

132   cannot rule out that our database includes data collected from soils affected by undocumented anthropogenic activities in the

133   past (*e.g.*, P fertilization occurring before reforestation), particularly in western Europe and eastern USA (*e.g.*, De Schrijver et

134   al., 2012). All data were collected at the plot scale. For data that included replicates within a plot or soil layer, average values

135   were calculated.

136   To compile our database, we first combined the two existing global databases (Augusto et al., 2017; Hou et al., 2018b).

137   Detailed information about the methods used to construct these datasets can be found in the original publications. We extracted

138   observations from these two databases by selecting only unfertilized, uncultivated, and (semi-) natural soils. This yielded 1684

139   observations from 182 studies from the dataset developed by Augusto et al. (2017) and 802 observations from 99 studies from

140   the dataset developed by Hou et al. (2018). Next, we removed 375 duplicates, after which our dataset contained 2111

141   observations from 245 studies (Figure S2). Because we use total soil P concentration as a predictor of soil P pools, we removed

142   data that did not include total soil P (calculated as the sum of P pools or measured by a separate method) or that did not identify

143   the concentration of at least one pool (*e.g.*, labile Pi, labile Po, moderately labile Pi, moderately labile Po, primary mineral P,

144   or occluded P). In this step, 816 observations were removed, resulting in a dataset that included 1295 observations from 178

145   studies.

146   Next, we added additional observations by compiling data from literature published after 2016, the final year included

147   in the database compiled by Hou et al. (2018). We used Google Scholar to search for studies published between 2016 and

148   08/08/2021 that referenced either Hedley et al. (1982) or Tiessen & Moir (1993). This search returned 701 publications citing

149   Hedley et al. (1982) and 245 citing Tiessen & Moir (1993). From this set, we selected studies that presented soil P data collected

150   using the fractionation method for (semi-)natural soils. The resulting 562 observations from 96 studies were added to our final

151   dataset, which includes a total of 1857 observations collected from 729 sites from 274 studies (Supplementary Text 1 and Fig.

152   S1).

153   In addition to soil P pool concentration and site coordinates, our dataset contains site characteristics including climate

154   variables (*i.e.*, mean annual temperature (MAT), mean annual precipitation (MAP), and potential biome), soil physiochemical

155   properties (*e.g.*, soil organic carbon concentration (SOC), soil clay and sand content, and soil pH), and elevation (Table 1).

156   Potential biome was identified using a global map of potential natural biomes (*i.e.*, the global distribution of biomes that would

157   exist in the absence of human activity) (Hengl et al., 2018). This categorization includes seven ecosystem types, including

158   tropical forest, temperate forest, boreal forest, grassland, savanna, desert, and tundra. We did not include parent material type

159   because it can be inferred from soil total P concentration and other soil properties (*e.g.*, soil texture and pH) (Augusto et al.,

160   2017; He et al., 2021). Because soil age was rarely reported, we used USDA soil order identity as a proxy for 3 age classes:

161   slightly, intermediately, and strongly weathered (Smeck, 1985; Yang et al., 2013). Among the 12 USDA soil orders, Entisols,

162   Inceptisols, Histosols, Andisols, and Gelisols are classified as slightly weathered soils. Alfisols, Mollisols, Aridisols, and

163   Vertisols are classified as intermediately weathered soils. Oxisols, Ultisols, and Spodosols are classified as strongly weathered

164   soils (Yang et al., 2013; Smeck, 1985). Given that atmospheric P inputs are small (0.3 kg P ha yr^{-1}, on average) compared to

165   soil P stocks (Mahowald et al., 2008; Wang et al., 2015) and are also highly uncertain over timescales relevant to soil

166   development, we do not consider atmospheric inputs as a predictor of P pools. As such, we did not include this information in

167   our dataset. We extracted data from each publication as available. In cases in which relevant information was not reported, we

168   extracted the missing data from gridded datasets (Table S1) based on the geographic coordinates of the study sites.

169   In random forest modelling, correlated predictors can be substituted for each other so that the importance of correlated

170   predictors will be shared, making each predictor's estimated importance smaller than its true value (Strobl et al., 2008). Thus,

171   we did not include soil total nitrogen content as it is strongly correlated with SOC ($r = +0.94$), nor did we include aridity index

172   as it is strongly correlated with MAP ($r = +0.72$). We also did not include rarely reported variables that were included in the

173   referenced studies (*e.g.*, soil extractable aluminum and iron concentrations).

2.3 Statistical modelling

175   All statistical analyses and plotting were performed in the R environment (v. 4.0.2) (R Core Team, 2018).

176   The database includes some extreme values in each P pool (Table 2). These values were likely observed in exceptional

177   geological contexts (Porder and Ramachandran, 2013) or in special soils (*e.g.*, very young volcanic soils). We included these

178   extreme values in the shared version of the dataset and in the database summary (Table 2). However, these values were

179   excluded from data used in model training, as the extremely high values could have a large influence on modeled relationships

180   between soil P pools and predictors. To this end, we only included values falling in the interval between 1% and 99%.

181   We used random forest regression models (Breiman, 2001) to predict global patterns of distribution for individual soil P

182   pools. All models included the same 11 predictors: MAT, MAP, potential biome, total P, soil depth, SOC, soil clay and sand

183   content, soil pH, elevation, and soil weathering stage (Table S1). The random forest analysis accounts for interactions and

184   nonlinear relationships between predictors and is appropriate for handling the multicollinearity problem in the multivariate

185   regression (Delgado-Baquerizo et al., 2017). The fit for each tree was determined by randomly selecting test cases. In order to

186   compare the relative importance of different variables for predicting the size of a pool, we performed random forest regression

187   analysis using the R package *caret* by applying the embedded R package *randomForest* version 3.1 (Liaw and Wiener, 2002)

188   with an automated *mtry* parameter. Five-fold cross-validation was performed using the R package *caret* (v. 6.0-86) (Kuhn,

189   2020) to evaluate model performance. The mean decrease in accuracy (%IncMSE) was used to evaluate the relative importance

190   of each variable as a predictor of a soil P pool. The mean decrease in accuracy plot shows how the accuracy of the fitted model

191   declines with the exclusion of a predictor. The greater the decline in accuracy, the more important the variable is for prediction.

192   In this study, the importance measure was calculated for each tree and averaged across the forest (500 trees). Our model found

193   that all 11 variables are important for predicting pool concentrations; thus, all were used as predictors as we developed the

194   global distribution map. Partial dependence plots showed the marginal effect of each continuous predictor on soil P pool

195   concentration. We used the *partial_dependence* function in the R package *edarf* version 1.1.1 (Jones and Linder, 2016) to

196   calculate the partial dependence of the response on an arbitrary dimensional set of continuous predictors from a fitted random

197   forest model.

198   Finally, we used the built forest regression models for each of the soil P pools for global prediction by using the predict

199   function in the *ranger* package with globally gridded datasets for all 11 variables. The *predict* function in the *ranger* package

200   (Wright and Ziegler, 2017) can compute the standard error of a predicted value. To estimate standard errors based on out-of-

201   bag predictions, we used the infinitesimal jackknife for bagging approach (Wager et al., 2014). Prior to use for global predictions,

202 driver variables were re-gridded to the same spatial resolution of $0.5° \times 0.5°$ (the original resolution of each predictor can be
203 found in Table S1). We did not mask croplands or other areas heavily influenced by human activity (*e.g.*, urban areas), so pool
204 concentrations predicted for these regions should be interpreted as the natural state prior to anthropogenic activity.

205   Because we trained models to predict P pool concentrations and proportions using the same 11 variables, we had two
206 options for developing global maps of P pool proportions: (1) dividing a pool's concentration by total P (He et al., 2021), or
207 (2) using our trained model. The resulting maps (Fig. S6) are highly correlated, with Pearson correlation coefficients from 0.61
208 to 0.98. Model accuracy was higher for predicted concentrations than it was for predicted proportions. (Fig. 2 & 3). Therefore,
209 we developed our map using the model to predict P pool concentration, after which these predictions were used with total soil
210 P concentrations to calculate P pool proportions (He et al., 2021) rather than predicting them using random forest models.

211   Soil depth was used as a predictor, allowing models to predict soil P pool concentration for any given depth (Hengl et al.,
212 2017). The partial dependence plot indicated that soil P pool concentration decreased approximately linearly with soil depth
213 in the top 30 cm and that there was no apparent trend with depth in the subsoil (30-100 cm) (Fig. S2-7); As such, we calculated
214 global soil P pool concentration at 0 cm, 10 cm, 20 cm, 30cm, and 100 cm. Averages for a depth interval (*e.g.*, 0-30 cm or 0-
215 100 cm) can be derived by calculating the weighted average of the predictions within that interval (Hengl et al., 2017).

216 **3 Results**

217 **3.1 Characters of P pools in natural soils across the world**

218   Our soil P pool database includes 1857 measurements from 729 geographically distinct sites and covers 6 continents, all
219 major biomes, and all 12 USDA soil orders in terrestrial ecosystems (Fig. 1). The database includes pool concentrations
220 measured in samples collected from the topsoil (0-30 cm) to a depth of 450 cm, with 83% of the measurements taken from the
221 topsoil. In the database, both concentrations and proportions of all P pools were highly skewed to the right (Fig. S4 & S5).

222   The largest pool among the six pools considered is occluded P, with a global mean concentration of 260.5 mg kg^{-1} and a
223 mean proportion of 41.9% (Table 2). Primary mineral P has a global mean concentration of 106.8 mg kg^{-1} and a mean
224 proportion of 19.0%. The labile and moderately labile Po pools have global mean concentrations of 31.1 mg kg^{-1}and 120.3 mg
225 kg^{-1}, respectively, and mean proportions of 5.9% and 19.5%, respectively. Moderately labile Pi has a global mean concentration
226 of 58.7 mg kg^{-1} and a mean proportion of 9.3%. Labile Pi represents the smallest proportion of total P with a global mean
227 concentration of 37.1 mg kg^{-1} and a mean proportion of 5.9% (Table 2).

228 **3.2 Model performance of different P pools in soils**

229   The random forest regression models explained 62%, 64%, 60%, 83%, 76%, and 82% of the variance in the concentrations
230 of labile Pi, labile Po, moderately labile Pi, moderately labile Po, primary P, and occluded P, respectively (Fig. 2). Using the
231 importance measure (%IncMSE), we identified the most important predictor for concentrations of soil labile Pi, labile Po,
232 moderately labile Pi, moderately labile Po, and occluded P as total P concentration, while the most important predictor for soil
233 primary P was the soil pH (Fig. 2). The random forest regression models explained 48%, 58%, 52%, 64%, 80%, and 58% of
234 the variance in proportions of labile Pi, labile Po, moderately labile Pi, moderately labile Po, primary P, and occluded P,
235 respectively (Fig. 3). Based on the importance measure, the most important predictor for proportions of soil labile Po,
236 moderately labile Pi, moderately labile Po, primary P, and occluded P was soil pH, while the most important predictor for labile
237 Pi proportion was soil depth (Fig. 3).

238 **3.3 Global patterns of P pools in natural soils**

239　　　　The maps we developed indicate that soil P pool concentrations showed substantial differences between biomes (Fig. 4).

240　　Lower P pools concentrations were found in warm and/or humid biomes (*e.g.*, tropical forest and savanna), while higher P

241　　pool concentrations were found in northern cold biomes (*e.g.*, tundra and boreal forest) (Fig. 4C). Estimated subsoil P pool

242　　concentrations showed similar patterns to those identified in the topsoil. The spatial patterns of pool proportions were different

243　　from those of pool concentrations across biomes (Fig. 7 & 8). For example, variation in the proportion of P held in the labile

244　　Pi pool was relatively small compared to variation observed in pool concentrations; moreover, the proportion of occluded P

245　　tended to increase in the transition from tundra and boreal forest to tropical forest and savanna (Fig. 4D). It should be noted

246　　that the mapped predictions of P pool concentrations across biomes are not consistent with the measured data (Fig. S4), which

247　　indicate that total soil P in tropical forests is higher than in any other biome. This result suggests a sampling bias due to

248　　overrepresentation of high total soil P sites in the tropical forest data.

249　　　　Our global predictions revealed that average values across all P pools were higher in slightly weathered soils compared

250　　to those in strongly weathered soils (Fig. 4A), reflecting the strong effect of soil development on soil P depletion. While

251　　occluded P proportion increased with soil development, the proportions of labile and moderately labile P (Pi and Po) were

252　　fairly independent of soil weathering stage (Fig. 4B).

253　　　　There are significant differences between our predictions and those made by Yang et al. (2013) in both the magnitude and

254　　the spatial patterns associated with most P pool concentrations. The two global estimates were only weakly to moderately

255　　correlated (Pearson correlation coefficients between 0.09 and 0.38) (Fig. S8). Yang et al.'s predictions are lower than ours for

256　　organic P, moderately labile Pi, primary mineral P, and occluded P concentrations (Table S5). Although average values for

257　　labile Pi concentrations estimated by Yang *et al.* were close to ours, they were only weakly correlated with each other (Pearson

258　　correlation coefficient of 0.09).

259　　　　Partial dependence plots (Fig. 5) and the results of Pearson correlation analysis (Table 3) revealed that total P

260　　concentration was significantly and positively correlated with concentrations for all six pools. SOC was significantly and

261　　negatively correlated with primary mineral P concentration, but positively correlated with the other five pool concentrations.

262　　MAT and MAP were significantly and negatively correlated with concentrations of all soil P pools. Soil pH was significantly

263　　and positively correlated with primary mineral P concentration, but significantly and negatively correlated with concentrations

264　　of the other five P pools. The results of Pearson correlation analysis also indicated that P pool concentrations were well

265　　correlated with each other, with the exception of primary mineral P; this pool was negatively correlated with labile Po and not

266　　correlated with moderately labile Po concentration.

267　　　　Partial dependence plots (Fig. 6) and Pearson correlation analysis (Table 3) revealed that soil pH was positively correlated

268　　with the primary mineral P proportion and negatively correlated with the other five P pool proportions. Soil labile Po,

269　　moderately labile Pi, and moderately labile Po proportions decreased substantially with an increase in MAT, while the occluded

270　　P proportion increased with MAT. Soil labile Po, moderately labile Pi, and moderately labile Po proportions increased

271　　substantially with increasing total P concentration, while the soil labile Pi and occluded P proportions decreased substantially

272　　with total P concentration.

273

274　　4 Discussion

275　　4.1 Improved mapping of different P pools in global natural soils

276　　　　We trained random forest regression models using 11 variables to predict six soil P pools at different depths in

277　　(semi-)natural terrestrial ecosystems, resulting in significant improvements over earlier estimates (Yang et al., 2013). First, we

278　　used a new global map of total P concentrations in natural soils (He et al., 2021) as a predictor. Because total P is an important

279   predictor and is highly correlated with all other P pools (Fig. 2 & 7), a higher quality map of total soil P provides much

280   improvement to the global estimates of P pool concentrations in soils (He et al., 2021). Second, Yang *et al.* (2013) used a

281   limited number (n=178) of measurements of Hedley P pools across soils. Our database represents a nearly ten-fold increase,

282   which can better represent the heterogeneous conditions on Earth. Third, Yang *et al.* (2013) estimated P pool concentrations

283   using total soil P concentrations, global soil order maps, and average proportions of various P pools for different soil orders.

284   In our study, we found that soil orders were less informative than other environmental predictors. By including more predictors

285   (*e.g.,* SOC, climate, and soil pH) our model offers significant improvements for capturing the variation observed in soil P

286   composition across the globe.

287   Differences between our estimates of different P pools and those presented by Yang *et al.* (2013) have significant

288   implications for soil P availability to vegetation. The averages and median values of Yang *et al.*'s predicted soil organic P,

289   moderately labile Pi, and occluded P concentrations were substantially lower than our estimates. Evidence suggests that soil

290   organic P and moderately labile Pi remain bioavailable on timescales of days to months (Helfenstein et al., 2020; Augusto et

291   al., 2017; Maharjan et al., 2018), while occluded P is bioavailable on the order of years to millennia (Hou et al., 2016; Wang

292   et al., 2007). Thus, soil P availability might be larger than previously assumed in assessments based on estimates by Yang *et*

293   *al.* (2013) (e.g., Sun et al. 2017).

4.2 Major drivers of different P pools in natural soils

295   Our results indicate that global variation in soil P pools is jointly controlled by total P concentration, soil pH, soil

296   development, climatic factors, and soil depth. Given that our models explain > 48% of the variance observed in P pools, our

297   results (see Figures 2 and 3) suggest that edaphic properties and climatic factors play significant roles in the size and

298   composition of different soil P pools globally. We discuss the effects of three major factors in more detail below.

299   **<u>Effects of total soil P concentration on P pools</u>**

300   We found that total soil P concentration was a prominent predictor of most soil P pools at the global scale and that total

301   P was positively correlated with all P pool concentrations and Po pool proportions. This is consistent with findings at local

302   (Turner and Blackwell, 2013) and global (Augusto et al., 2017; Hou et al., 2018; Harrison, 1987) scales. Total soil P is

303   influenced by multiple soil forming factors (*e.g.,* parent material P concentration, climate, soil organic carbon content, and soil

304   texture) (He et al., 2021). Thus, total soil P provides an integrated measure of factors that regulate the size of the P pools.

305   Moreover, this result is consistent with the emerging idea of substrate-based P cycling in natural ecosystems (Lang et al., 2017;

306   Lang et al., 2016): Soils with high total P content are usually also associated with a large primary mineral P pool. At these P-

307   rich sites, plant and microbial communities tend to promote P release from primary minerals, with subsequent biological and

308   abiotic transformations resulting in high concentrations in all other P pools (Lang et al., 2016; He et al., 2021) and higher

309   proportions of organic P (Hou et al., 2018c). In contrast, at P-poor sites, plant and microbial communities are more reliant on

310   P recycling systems that promote the mineralization of Po by soil microbes (Achat et al., 2009; Marklein and Houlton, 2012)

311   and the mobilization of moderately labile Pi or even occluded P (Augusto et al., 2017) to sustain the P supply. Therefore, soil

312   P pool concentrations are expected to co-vary with total soil P concentration.

313   Interestingly, our predictions indicate that labile Pi concentrations are not primarily controlled by soil P supply (*i.e.,* total

314   soil P) on a global scale, but by biological processes such as plant uptake, microbial uptake, immobilization, and mineralization

315   (Yang et al., 2013; Cross and Schlesinger, 1995; Weihrauch and Opp, 2018; Hou et al., 2016). Some work suggests that in

316   strongly weathered soils (with limited P stocks), mineralization of Po could be a major source of labile Pi (Vitousek, 1984;

317   Achat et al., 2009). In addition, fluctuating redox conditions in highly weathered soils can cause the release of labile Pi from

318   the moderately labile P pool through the reduction of Fe^{3+} minerals (Chacon et al., 2006; Liptzin and Silver, 2009). This is

319   consistent with other studies that found that the labile Pi pool is quite stable across weathering stages (Cross and Schlesinger,

320   1995; Yang and Post, 2011; Zhang et al., 2005).

321

Effects of soil pH on P pools

323   Consistent with previous studies (Hou et al., 2018c; Kruse et al., 2015; Oburger et al., 2011; Barrow et al., 2020), our
324   results indicate that soil pH is an important predictor of P pool concentrations and proportions in natural soils globally. The
325   relative importance of pH is unsurprising, since the sequential fractionation procedure is based on dissolving a soil sample in
326   solutions of varying acidity/alkalinity. However, the observed pH effects also support the existing mechanistic understanding
327   of the various pools. The strong positive correlation of primary P and soil pH is expected because 1) the primary P pool is
328   composed mainly of calcium phosphate/apatite, which is highly soluble at low pH but becomes less soluble with increasing
329   pH and 2) soil pH declines with soil weathering intensity (Delgado-Baquerizo et al., 2020) (*e.g.*, the highest values of soil pH
330   are usually found in dry regions where chemical weathering rates are limited by water availability (Slessarev et al., 2016)).
331   Both factors affect the transformation of primary mineral P to other forms.

332   Soil pH shows important but negative influences on the proportions of other soil P pools (*i.e.*, proportions of labile Pi and
333   Po, moderately labile Pi and Po, and occluded P). There are several possible explanations for these relationships. First, low
334   soil pH (< 5.0) inhibits soil microbial activities and the extracellular activity of phosphatase enzymes (Aciego Pietri and
335   Brookes, 2008; Eivazi and Tabatabai, 1977; Xu et al., 2017). Thus, in acidic soils, more organic P (*i.e.*, labile Po) may
336   accumulate than in neutral soils. Second, decreasing soil pH is associated with the accumulation of Fe and Al oxides, which
337   leads to enhanced adsorption of P (*i.e.*, labile Pi and labile Po). Third, pH tends to decrease as soil weathering advances and
338   base cations are progressively washed out (Slessarev et al. 2016). As soils weather, occluded P accumulates, which explains
339   why the occluded pool decreases with increasing pH. Fourth, increasing soil pH is associated with enhanced adsorption of
340   dissolved Pi to Ca and Mg, reducing the amount of labile Pi available for plants and soil microorganisms (Fink et al., 2016;
341   Gerke, 2015). This could explain the negative relationship between soil pH and the labile Pi proportion identified in this study.

342

Effects of soil development on P pools

344   The variation of P pools across weathering stages predicted by our model partially supports Walker and Syers' (1976)
345   theory based on soil chronosequences, and is consistent with more nuanced models of soil P evolution that consider variation
346   in tectonic uplift and geological composition (Buendía et al., 2010). While our results are consistent with expectations from
347   Walker and Syers' theory about the increase in the proportion of occluded P that occurs at the expense of primary and organic
348   P during soil development, they do not support Walker and Syers' ideas regarding the evolution of the labile Pi and moderately
349   labile Pi pools. The evolution of occluded P is commonly explained by the increase of Al and Fe oxide minerals and the
350   decrease of soil pH; In addition to being fixed onto Fe and Al oxides, P that is released from primary minerals or mineralized
351   from organic matter can be occluded by adsorbing to mineral surfaces (Crews et al., 1995; Selmants and Hart, 2010). This
352   mechanistic understanding is in line with our findings about the dependence of soil P pools on soil pH. However, our results
353   disagree with the prediction by Walker and Syers' (1976) model in that labile Pi and moderately labile Pi (non-occluded P in
354   Walker and Syers' model) formed significant fractions of total P throughout all soil orders across weathering stages. This could
355   be due to the coarse classification of weathering stages in our study, which may be insufficient to characterize the end members
356   of the range. This explanation is supported by the small proportion of 1 M HCl P in the slightly weathered soil and the moderate
357   amounts of P remaining in strongly weathered soils. The inconsistency may also be due to the fact that Walker and Syers'
358   model was based on a system with negligible tectonic uplift, as argued by Buendía et al. (2010). In Buendía *et al.*'s model, the
359   primary P pool is replenished by the uplift effect, a prediction that is supported by our finding of considerable primary mineral
360   P pools occurring across soil weathering stages.

361

Effects of soil depth on P pools

363   We found that soil P pools varied significantly with soil depth. Total soil P in topsoil was higher than in subsoil due to
364   biological uplift, which was reported by previous studies (Jobbágy and Jackson, 2001; Porder and Chadwick, 2009). The labile
365   and moderately labile P pool concentrations (in both inorganic and organic pools) showed a similar trend. In contrast, the
366   primary P and occluded P concentrations in topsoil were lower than in the subsoil. This can be explained by the fact that topsoil
367   tends to be more weathered and developed than the subsoil (Achat et al., 2012; Chen et al., 2021).
368

369   **4.3 Limitations and prediction uncertainty**

370   In our database, some regions were underrepresented (*e.g.*, northern Canada, middle and northern Asia, and inner Africa),
371   which may result in low accuracy of the predicted values in those regions. In the tropics, high P soils were overrepresented
372   and accuracy of predicted values in tropical regions may be quite low. Our database contains four times as many observations
373   from surface mineral soils (0-30cm) than it does from soils deeper than 30 cm. As such, the predicted concentrations of different
374   P pools for deep soils may suffer from larger uncertainties. Finally, large portions of variation remain unexplained, especially
375   variation in soil labile Pi concentrations and proportions (40% and 52% unexplained, respectively), indicating that other factors
376   that were not accounted for play a role. These factors may include microbial processes, Fe and Al oxide concentrations, plant
377   community composition, atmospheric deposition, and soil erosion (Kruse et al., 2015; Achat et al., 2016). These limitations
378   highlight the need for additional measurements, particularly from underrepresented regions and the subsoil as well as
379   measurements of closely associated variables, especially those related to labile Pi.
380

381   **5 Conclusion**

382   Here, we compiled the largest database to date of different soil P pools. Using machine learning modelling, we quantified
383   the relative importance of multiple predictors for estimating different soil P pools and estimated these pools at the global scale.
384   Our results indicated that the global concentrations of soil labile Pi, labile Po, moderately labile Pi, moderately labile Po, and
385   occluded P could be predicted mainly by the total soil P concentration, while primary P concentration was mainly predicted by
386   soil pH and total soil P concentration. For predicting proportions of total P, soil pH was the most important predictor for all P
387   pools at the global scale, with the exception of labile Pi proportions, for which soil depth was the main driver. In addition, our
388   results also revealed significant effects of climate and other edaphic factors on spatial variation in P pools. We concluded that
389   edaphic properties and climatic factors were significant predictors of soil P pools, including concentration and proportion of
390   total P. These findings represent a significant step towards improving understanding of global variations in different soil P
391   pools. Our global maps of predictions of different P pools will be important to efforts to improve global scale biogeochemical
392   models of the P cycle.
393
394

398   **Author contributions**
399   D.S.G., Y.W. and E.H designed this study. L.A., E.H, and X.H. collected the data. X.H., E.H., L.A., D.S.G., B.R., Y.W., J.H.,
400   and Y.H. discussed analyzing methods. X.H. conducted the analysis and drafted the manuscript. All authors discussed the
401   results and contributed to the manuscript.

Biogeosciences

Discussions

Open Access

EGU

409 **Competing interests**
410 The authors declare that they have no conflict of interest.
411
412
413
414
415 **References**

416 Achat, D.L., Augusto, L., Bakker, M.R., Gallet-Budynek, A. and Morel, C.: Microbial processes controlling P availability in
417 forest spodosols as affected by soil depth and soil properties. Soil Biology and Biochemistry, 44, 39-48,
418 https://doi.org/10.1016/j.soilbio.2011.09.007, 2012.
419 Achat, D.L., Bakker, M.R. and Morel, C.: Process-Based Assessment of Phosphorus Availability in a Low Phosphorus Sorbing
420 Forest Soil using Isotopic Dilution Methods. Soil Sci. Soc. Am. J., 73, 2131-2142, https://doi.org/10.2136/sssaj2009.0009,
421 2009.
422 Achat, D.L., Pousse, N., Nicolas, M., Brédoire, F. and Augusto, L.: Soil properties controlling inorganic phosphorus
423 availability: general results from a national forest network and a global compilation of the literature. Biogeochemistry,
424 127, 255-272, https://doi.org/10.1007/s10533-015-0178-0, 2016.
425 Aciego Pietri, J.C. and Brookes, P.C.: Relationships between soil pH and microbial properties in a UK arable soil. Soil Biology
426 and Biochemistry, 40, 1856-1861, https://doi.org/10.1016/j.soilbio.2008.03.020, 2008.
427 Augusto, L., Achat, D.L., Jonard, M., Vidal, D. and Ringeval, B.: Soil parent material—A major driver of plant nutrient
428 limitations in terrestrial ecosystems. Glob. Change Biol., 23, 3808-3824, https://doi.org/10.1111/gcb.13691, 2017.
429 Barrow, N.J., Debnath, A. and Sen, A.: Measurement of the effects of pH on phosphate availability. Plant Soil, 454, 217-224,
430 https://doi.org/10.1007/s11104-020-04647-5, 2020.
431 Barrow, N.J., Sen, A., Roy, N. and Debnath, A.: The soil phosphate fractionation fallacy. Plant Soil, 459, 1-11,
432 https://doi.org/10.1007/s11104-020-04476-6, 2021.
433 Breiman, L.: Random Forests. Mach. Learn., 45, 5-32, https://doi.org/10.1023/A:1010933404324, 2001.
434 Brucker, E. and Spohn, M.: Formation of soil phosphorus fractions along a climate and vegetation gradient in the Coastal
435 Cordillera of Chile. Catena, 180, 203-211, https://doi.org/10.1016/j.catena.2019.04.022, 2019.
436 Buendía, C., Kleidon, A. and Porporato, A.: The role of tectonic uplift, climate, and vegetation in the long-term terrestrial
437 phosphorous cycle. Biogeosciences, 7, 2025-2038, https://doi.org/10.5194/bg-7-2025-2010, 2010.
438 Bünemann, E.K., Steinebrunner, F., Smithson, P.C., Frossard, E. and Oberson, A.: Phosphorus Dynamics in a Highly
439 Weathered Soil as Revealed by Isotopic Labeling Techniques. Soil Sci. Soc. Am. J., 68, 1645-1655,
440 https://doi.org/10.2136/sssaj2004.1645, 2004.
441 Chacon, N., Flores, S. and Gonzalez, A.: Implications of iron solubilization on soil phosphorus release in seasonally flooded
442 forests of the lower Orinoco River, Venezuela. Soil Biology and Biochemistry, 38, 1494-1499,
443 https://doi.org/10.1016/j.soilbio.2005.10.018, 2006.
444 Chen, C.R., Hou, E.Q., Condron, L.M., Bacon, G., Esfandbod, M., Olley, J. and Turner, B.L.: Soil phosphorus fractionation
445 and nutrient dynamics along the Cooloola coastal dune chronosequence, southern Queensland, Australia. Geoderma, 257-

446   258, 4-13, https://doi.org/10.1016/j.geoderma.2015.04.027, 2015.

447 Chen, X., Feng, J., Ding, Z., Tang, M. and Zhu, B.: Changes in soil total, microbial and enzymatic C-N-P contents and

448   stoichiometry with depth and latitude in forest ecosystems. Sci. Total Environ., 151583,

449   https://doi.org/10.1016/j.scitotenv.2021.151583, 2021.

450 Condron, L.M. and Newman, S.: Revisiting the fundamentals of phosphorus fractionation of sediments and soils. J. Soils

451   Sediments, 11, 830-840, https://doi.org/10.1007/s11368-011-0363-2, 2011.

452 Crews, T.E., Kitayama, K., Fownes, J.H., Riley, R.H., Herbert, D.A., Mueller-Dombois, D. and Vitousek, P.M.: Changes in

453   Soil Phosphorus Fractions and Ecosystem Dynamics across a Long Chronosequence in Hawaii. Ecology, 76, 1407-1424,

454   https://doi.org/10.2307/1938144, 1995.

455 Cross, A.F. and Schlesinger, W.H.: A literature review and evaluation of the Hedley fractionation; applications to the

456   biogeochemical cycle of soil phosphorus in natural ecosystems. Geoderma, 64, 197-214, https://doi.org/10.1016/0016-

457   7061(94)00023-4, 1995.

458 Cunha, H.F.V., Andersen, K.M., Lugli, L.F., Santana, F.D., Aleixo, I.F., Moraes, A.M., Garcia, S., Di Ponzio, R., Mendoza,

459   E.O., Brum, B., Rosa, J.S., Cordeiro, A.L., Portela, B.T.T., Ribeiro, G., Coelho, S.D., de Souza, S.T., Silva, L.S., Antonieto,

460   F., Pires, M., Salomão, A.C., Miron, A.C., de Assis, R.L., Domingues, T.F., Aragão, L.E.O.C., Meir, P., Camargo, J.L.,

461   Manzi, A.O., Nagy, L., Mercado, L.M., Hartley, I.P. and Quesada, C.A.: Direct evidence for phosphorus limitation on

462   Amazon forest productivity. Nature, 608, 558-562, https://doi.org/10.1038/s41586-022-05085-2, 2022.

463 De Schrijver, A., Vesterdal, L., Hansen, K., De Frenne, P., Augusto, L., Achat, D.L., Staelens, J., Baeten, L., De Keersmaeker,

464   L., De Neve, S. and Verheyen, K.: Four decades of post-agricultural forest development have caused major redistributions

465   of soil phosphorus fractions. Oecologia, 169, 221-234, https://doi.org/10.1007/s00442-011-2185-8, 2012.

466 Delgado-Baquerizo, M., Eldridge, D.J., Maestre, F.T., Karunaratne, S.B., Trivedi, P., Reich, P.B. and Singh, B.K.: Climate

467   legacies drive global soil carbon stocks in terrestrial ecosystems. Sci. Adv., 3, e1602008,

468   https://doi.org/10.1126/sciadv.1602008, 2017.

469 Delgado-Baquerizo, M., Reich, P.B., Bardgett, R.D., Eldridge, D.J., Lambers, H., Wardle, D.A., Reed, S.C., Plaza, C., Png,

470   G.K., Neuhauser, S., Berhe, A.A., Hart, S.C., Hu, H., He, J., Bastida, F., Abades, S., Alfaro, F.D., Cutler, N.A., Gallardo,

471   A., García-Velázquez, L., Hayes, P.E., Hseu, Z., Pérez, C.A., Santos, F., Siebe, C., Trivedi, P., Sullivan, B.W., Weber-

472   Grullon, L., Williams, M.A. and Fierer, N.: The influence of soil age on ecosystem structure and function across biomes.

473   Nat. Commun., 11, 4721-4721, https://doi.org/10.1038/s41467-020-18451-3, 2020.

474 Eivazi, F. and Tabatabai, M.A.: Phosphatases in soils. Soil Biology and Biochemistry, 9, 167-172,

475   https://doi.org/10.1016/0038-0717(77)90070-0, 1977.

476 Ellsworth, D.S., Crous, K.Y., De Kauwe, M.G., Verryckt, L.T., Goll, D., Zaehle, S., Bloomfield, K.J., Ciais, P., Cernusak,

477   L.A., Domingues, T.F., Dusenge, M.E., Garcia, S., Guerrieri, R., Ishida, F.Y., Janssens, I.A., Kenzo, T., Ichie, T., Medlyn,

478   B.E., Meir, P., Norby, R.J., Reich, P.B., Rowland, L., Santiago, L.S., Sun, Y., Uddling, J., Walker, A.P., Weerasinghe,

479   K.W.L.K., van de Weg, M.J., Zhang, Y., Zhang, J. and Wright, I.J.: Convergence in phosphorus constraints to

480   photosynthesis in forests around the world. Nat. Commun., 13, 5005, https://doi.org/10.1038/s41467-022-32545-0, 2022.

481 Elser, J.J., Bracken, M.E.S., Cleland, E.E., Gruner, D.S., Harpole, W.S., Hillebrand, H., Ngai, J.T., Seabloom, E.W., Shurin,

482   J.B. and Smith, J.E.: Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and

483   terrestrial ecosystems. Ecol. Lett., 10, 1135-1142, https://doi.org/10.1111/j.1461-0248.2007.01113.x, 2007.

484 Fink, J.R., Inda, A.V., Bavaresco, J., Barrón, V., Torrent, J. and Bayer, C.: Adsorption and desorption of phosphorus in

485   subtropical soils as affected by management system and mineralogy. Soil and Tillage Research, 155, 62-68,

486   https://doi.org/10.1016/j.still.2015.07.017, 2016.

487 Fleischer, K., Rammig, A., Kauwe, D.M.G., Walker, A.P., Domingues, T.F., Fuchslueger, L., Garcia, S., Goll, D.S., Grandis,

488   A., Jiang, M., Haverd, V., Hofhansl, F., Holm, J.A., Kruijt, B., Leung, F., Medlyn, B.E., Mercado, L.M., Norby, R.J., Pak,

489   B., Randow, V.C., Quesada, C.A., Schaap, K.J., Valverde-Barrantes, O.J., Wang, Y.P., Yang, X., Zaehle, S., Zhu, Q.,
490   Lapola, D.M. and Oak Ridge National Lab. ORNL, O.R.T.U.: Amazon forest response to CO2 fertilization dependent on
491   plant phosphorus acquisition. Nat. Geosci., 12, 736-741, https://doi.org/10.1038/s41561-019-0404-9, 2019.

492   Gerke, J.: The acquisition of phosphate by higher plants: Effect of carboxylate release by the roots. A critical review. J. Plant
493   Nutr. Soil Sci., 178, 351-364, https://doi.org/10.1002/jpln.201400590, 2015.

494   Goll, D.S., Brovkin, V., Parida, B.R., Reick, C.H., Kattge, J., Reich, P.B., van Bodegom, P.M. and Niinemets, Ü.: Nutrient
495   limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling.
496   Biogeosciences, 9, 3547-3569, https://doi.org/10.5194/bg-9-3547-2012, 2012.

497   Gu, C. and Margenot, A.J.: Navigating limitations and opportunities of soil phosphorus fractionation. Plant Soil, 459, 13-17,
498   https://doi.org/10.1007/s11104-020-04552-x, 2021.

499   Harrison, A.F., Soil Organic Phosphorus. A Review of World Literature. CAB International, Wallingford.,1987.

500   He, X., Augusto, L., Goll, D.S., Ringeval, B., Wang, Y., Helfenstein, J., Huang, Y., Yu, K., Wang, Z., Yang, Y. and Hou, E.:
501   Global patterns and drivers of soil total phosphorus concentration. Earth Syst. Sci. Data, 13, 5831-5846,
502   https://doi.org/10.5194/essd-13-5831-2021, 2021.

503   Hedley, M.J., Stewart, J.W.B. and Chauhan, B.S.: Changes in Inorganic and Organic Soil Phosphorus Fractions Induced by
504   Cultivation   Practices   and   by   Laboratory   Incubations.   Soil   Sci.   Soc.   Am.   J.,   46,   970-976,
505   https://doi.org/10.2136/sssaj1982.03615995004600050017x, 1982.

506   Helfenstein, J., Frossard, E., Pistocchi, C., Chadwick, O., Vitousek, P. and Tamburini, F.: Soil Phosphorus Exchange as
507   Affected by Drying-Rewetting of Three Soils From a Hawaiian Climatic Gradient. Frontiers in Soil Science, 1, 738464,
508   https://doi.org/10.3389/fsoil.2021.738464, 2021.

509   Helfenstein, J., Pistocchi, C., Oberson, A., Tamburini, F., Goll, D.S. and Frossard, E.: Estimates of mean residence times of
510   phosphorus   in   commonly   considered   inorganic   soil   phosphorus   pools.   Biogeosciences,   17,   441-454,
511   https://doi.org/10.5194/bg-17-441-2020, 2020.

512   Hengl, T., Mendes, D.J.J., Heuvelink, G.B., Ruiperez, G.M., Kilibarda, M., Blagotic, A., Shangguan, W., Wright, M.N., Geng,
513   X., Bauer-Marschallinger, B., Guevara, M.A., Vargas, R., MacMillan, R.A., Batjes, N.H., Leenaars, J.G., Ribeiro, E.,
514   Wheeler, I., Mantel, S. and Kempen, B.: SoilGrids250m: Global gridded soil information based on machine learning. Plos
515   One, 12, e0169748, https://doi.org/10.1371/journal.pone.0169748, 2017.

516   Hengl, T., Walsh, M.G., Sanderman, J., Wheeler, I., Harrison, S.P. and Prentice, I.C.: Global mapping of potential natural
517   vegetation: an   assessment   of   machine   learning   algorithms   for   estimating   land   potential.   Peerj,   6,   e5457,
518   https://doi.org/10.7717/peerj.5457, 2018.

519   Hou, E., Chen, C., Kuang, Y., Zhang, Y., Heenan, M. and Wen, D.: A structural equation model analysis of phosphorus
520   transformations   in   global   unfertilized   and   uncultivated   soils.   Glob.   Biogeochem.   Cycle,   30,   1300-1309,
521   https://doi.org/10.1002/2016GB005371, 2016.

522   Hou, E., Chen, C., Luo, Y., Zhou, G., Kuang, Y., Zhang, Y., Heenan, M., Lu, X. and Wen, D.: Effects of climate on soil
523   phosphorus   cycle   and   availability   in   natural   terrestrial   ecosystems.   Glob.   Change   Biol.,   24,   3344-3356,
524   https://doi.org/10.1111/gcb.14093, 2018a.

525   Hou, E., Luo, Y., Kuang, Y., Chen, C., Lu, X., Jiang, L., Luo, X. and Wen, D.: Global meta-analysis shows pervasive
526   phosphorus   limitation   of   aboveground   plant   production   in   natural   terrestrial   ecosystems.   Nat.   Commun.,   11,   637,
527   https://doi.org/10.1038/s41467-020-14492-w, 2020.

528   Hou, E., Tan, X., Heenan, M. and Wen, D.: A global dataset of plant available and unavailable phosphorus in natural soils
529   derived by Hedley method. Sci. Data, 5, 180166, https://doi.org/10.1038/sdata.2018.166, 2018b.

530   Hou, E., Wen, D., Jiang, L., Luo, X., Kuang, Y., Lu, X., Chen, C., Allen, K.T., He, X., Huang, X. and Luo, Y.: Latitudinal
531   patterns of terrestrial phosphorus limitation over the globe. Ecol. Lett., 24, 1420-1431, https://doi.org/10.1111/ele.13761,

532   2021.

533   Hou, E., Wen, D., Kuang, Y., Cong, J., Chen, C., He, X., Heenan, M., Lu, H. and Zhang, Y.: Soil pH predominantly controls
534   the forms of organic phosphorus in topsoils under natural broadleaved forests along a 2500 km latitudinal gradient.
535   Geoderma, 315, 65-74, https://doi.org/10.1016/j.geoderma.2017.11.041, 2018c.

536   Jobbágy, E.G. and Jackson, R.B.: The Distribution of Soil Nutrients with Depth: Global Patterns and the Imprint of Plants.
537   Biogeochemistry, 53, 51-77, https://doi.org/10.1023/A:1010760720215, 2001.

538   Jones, Z.M. and Linder, F.J.: edarf: Exploratory Data Analysis using Random Forests. The Journal of Open Source Software,
539   1, 92, https://doi.org/10.21105/joss.00092, 2016.

540   Klotzbücher, A., Kaiser, K., Klotzbücher, T., Wolff, M. and Mikutta, R.: Testing mechanisms underlying the Hedley sequential
541   phosphorus extraction of soils. J. Plant Nutr. Soil Sci., 182, 570-577, https://doi.org/10.1002/jpln.201800652, 2019.

542   Kruse, J., Abraham, M., Amelung, W., Baum, C., Bol, R., Kühn, O., Lewandowski, H., Niederberger, J., Oelmann, Y., Rüger,
543   C., Santner, J., Siebers, M., Siebers, N., Spohn, M., Vestergren, J., Vogts, A. and Leinweber, P.: Innovative methods in
544   soil phosphorus research: A review. J. Plant Nutr. Soil Sci., 178, 43-88, https://doi.org/10.1002/jpln.201400327, 2015.

545   Kuhn,   M.:   caret: Classification   and   Regression   Training.   R   package   version   6.0-86.,
546   https://doi.org/https://github.com/topepo/caret/, 2020.

547   Lang, F., Bauhus, J., Frossard, E., George, E., Kaiser, K., Kaupenjohann, M., Krüger, J., Matzner, E., Polle, A., Prietzel, J.,
548   Rennenberg, H. and Wellbrock, N.: Phosphorus in forest ecosystems: New insights from an ecosystem nutrition
549   perspective. J. Plant Nutr. Soil Sci., 179, 129-135, https://doi.org/10.1002/jpln.201500541, 2016.

550   Lang, F., Krüger, J., Amelung, W., Willbold, S., Frossard, E., Bünemann, E.K., Bauhus, J., Nitschke, R., Kandeler, E., Marhan,
551   S., Schulz, S., Bergkemper, F., Schloter, M., Luster, J., Guggisberg, F., Kaiser, K., Mikutta, R., Guggenberger, G., Polle,
552   A., Pena, R., Prietzel, J., Rodionov, A., Talkner, U., Meesenburg, H., von Wilpert, K., Hölscher, A., Dietrich, H.P. and
553   Chmara, I.: Soil phosphorus supply controls P nutrition strategies of beech forest ecosystems in Central Europe.
554   Biogeochemistry, 136, 5-29, https://doi.org/10.1007/s10533-017-0375-0, 2017.

555   Liaw, A. and Wiener, M.: Classification and Regression by randomForest. R News, 2, 18-22,2002.

556   Liptzin, D. and Silver, W.L.: Effects of carbon additions on iron reduction and phosphorus availability in a humid tropical
557   forest soil. Soil Biology and Biochemistry, 41, 1696-1702, https://doi.org/10.1016/j.soilbio.2009.05.013, 2009.

558   Maharjan, M., Maranguit, D. and Kuzyakov, Y.: Phosphorus fractions in subtropical soils depending on land use. Eur. J. Soil
559   Biol., 87, 17-24, https://doi.org/10.1016/j.ejsobi.2018.04.002, 2018.

560   Mahowald, N., Jickells, T.D., Baker, A.R., Artaxo, P., Benitez-Nelson, C.R., Bergametti, G., Bond, T.C., Chen, Y., Cohen,
561   D.D., Herut, B., Kubilay, N., Losno, R., Luo, C., Maenhaut, W., McGee, K.A., Okin, G.S., Siefert, R.L. and Tsukuda, S.:
562   Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts.
563   Glob. Biogeochem. Cycle, 22, GB4026, https://doi.org/10.1029/2008GB003240, 2008.

564   Marklein, A.R. and Houlton, B.Z.: Nitrogen inputs accelerate phosphorus cycling rates across a wide variety of terrestrial
565   ecosystems. New Phytol., 193, 696-704, https://doi.org/10.1111/j.1469-8137.2011.03967.x, 2012.

566   Morel, C., Ziadi, N., Messiga, A., Bélanger, G., Denoroy, P., Jeangros, B., Jouany, C., Fardeau, J., Mollier, A., Parent, L.,
567   Proix, N., Rabeharisoa, L. and Sinaj, S.: Modeling of phosphorus dynamics in contrasting agroecosystems using long-term
568   field experiments. Can. J. Soil Sci., 94, 377-387, https://doi.org/10.4141/cjss2013-024, 2014.

569   Oburger, E., Jones, D.L. and Wenzel, W.W.: Phosphorus saturation and pH differentially regulate the efficiency of organic
570   acid anion-mediated P solubilization mechanisms in soil. Plant Soil, 341, 363-382, https://doi.org/10.1007/s11104-010-
571   0650-5, 2011.

572   Peltzer, D.A., Wardle, D.A., Allison, V.J., Baisden, W.T., Bardgett, R.D., Chadwick, O.A., Condron, L.M., Parfitt, R.L.,
573   Porder, S., Richardson, S.J., Turner, B.L., Vitousek, P.M., Walker, J. and Walker, L.R.: Understanding ecosystem
574   retrogression. Ecol. Monogr., 80, 509-529, https://doi.org/10.1890/09-1552.1, 2010.

14

575   Porder, S. and Chadwick, O.A.: Climate and Soil-Age Constraints on Nutrient Uplift and Retention by Plants. Ecology, 90,
576   623-636, https://doi.org/10.1890/07-1739.1, 2009.

577   Porder, S. and Ramachandran, S.: The phosphorus concentration of common rocks—a potential driver of ecosystem P status.
578   Plant Soil, 367, 41-55, https://doi.org/10.1007/s11104-012-1490-2, 2013.

579   R Core Team: R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna,
580   Austria, https://doi.org/https://www.R-project.org/, 2018.

581   Ringeval, B., Augusto, L., Monod, H., van Apeldoorn, D., Bouwman, L., Yang, X., Achat, D.L., Chini, L.P., Van Oost, K.,
582   Guenet, B., Wang, R., Decharme, B., Nesme, T. and Pellerin, S.: Phosphorus in agricultural soils: drivers of its distribution
583   at the global scale. Glob. Change Biol., 23, 3418-3432, https://doi.org/10.1111/gcb.13618, 2017.

584   Selmants, P.C. and Hart, S.C.: Phosphorus and soil development: Does the Walker and Syers model apply to semiarid
585   ecosystems? Ecology, 91, 474-484, https://doi.org/10.1890/09-0243.1, 2010.

586   Slessarev, E.W., Lin, Y., Bingham, N.L., Johnson, J.E., Dai, Y., Schimel, J.P. and Chadwick, O.A.: Water balance creates a
587   threshold in soil pH at the global scale. Nature, 540, 567-569, https://doi.org/10.1038/nature20139, 2016.

588   Smeck, N.E.: Phosphorus dynamics in soils and landscapes. Geoderma, 36, 185-199, https://doi.org/10.1016/0016-
589   7061(85)90001-1, 1985.

590   Strobl, C., Boulesteix, A., Kneib, T., Augustin, T. and Zeileis, A.: Conditional variable importance for random forests. Bmc
591   Bioinformatics, 9, 307, https://doi.org/10.1186/1471-2105-9-307, 2008.

592   Sun, Y., Peng, S., Goll, D.S., Ciais, P., Guenet, B., Guimberteau, M., Hinsinger, P., Janssens, I.A., Peñuelas, J., Piao, S.,
593   Poulter, B., Violette, A., Yang, X., Yin, Y. and Zeng, H.: Diagnosing phosphorus limitations in natural terrestrial
594   ecosystems in carbon cycle models. Earth's Future, 5, 730-749, https://doi.org/10.1002/2016EF000472, 2017.

595   Tiessen, H. and Moir, J., Characterization of available P by sequential extraction, in Soil Sampling and Methods of Analysis,
596   edited by M. R. Carter, pp. 75–86, Lewis Publishers., In,1993.

597   Turner, B.L. and Blackwell, M.S.A.: Isolating the influence of pH on the amounts and forms of soil organic phosphorus. Eur.
598   J. Soil Sci., 64, 249-259, https://doi.org/10.1111/ejss.12026, 2013.

599   Vitousek, P.M.: Litterfall, Nutrient Cycling, and Nutrient Limitation in Tropical Forests. Ecology, 65, 285-298,
600   https://doi.org/10.2307/1939481, 1984.

601   Vitousek, P.M., Porder, S., Houlton, B.Z. and Chadwick, O.A.: Terrestrial phosphorus limitation: mechanisms, implications,
602   and nitrogen–phosphorus interactions. Ecol. Appl., 20, 5-15, https://doi.org/10.1890/08-0127.1, 2010.

603   Vu, D.T., Tang, C. and Armstrong, R.D.: Transformations and availability of phosphorus in three contrasting soil types from
604   native and farming systems: A study using fractionation and isotopic labeling techniques. J. Soils Sediments, 10, 18-29,
605   https://doi.org/10.1007/s11368-009-0068-y, 2010.

606   Wager, S., Hastie, T. and Efron, B.: Confidence Intervals for Random Forests: The Jackknife and the Infinitesimal Jackknife.
607   J. Mach. Learn. Res., 15, 1625-1651,2014.

608   Walker, T.W. and Syers, J.K.: The fate of phosphorus during pedogenesis. Geoderma, 15, 1-19, https://doi.org/10.1016/0016-
609   7061(76)90066-5, 1976.

610   Wang, R., Balkanski, Y., Boucher, O., Ciais, P., Peñuelas, J. and Tao, S.: Significant contribution of combustion-related
611   emissions to the atmospheric phosphorus budget. Nat. Geosci., 8, 48-54, https://doi.org/10.1038/ngeo2324, 2015.

612   Wang, Y., Ciais, P., Goll, D., Huang, Y., Luo, Y., Wang, Y., Bloom, A.A., Broquet, G., Hartmann, J., Peng, S., Penuelas, J.,
613   Piao, S., Sardans, J., Stocker, B.D., Wang, R., Zaehle, S. and Zechmeister-Boltenstern, S.: GOLUM-CNP v1.0: a data-
614   driven modeling of carbon, nitrogen and phosphorus cycles in major terrestrial biomes. Geosci. Model Dev., 11, 3903-
615   3928, https://doi.org/10.5194/gmd-11-3903-2018, 2018.

616   Wang, Y.P., Houlton, B.Z. and Field, C.B.: A model of biogeochemical cycles of carbon, nitrogen, and phosphorus including

617   symbiotic nitrogen fixation and phosphatase production. Glob. Biogeochem. Cycle, 21, GB1018,
618   https://doi.org/10.1029/2006GB002797, 2007.

619 Wang, Y.P., Huang, Y., Augusto, L., Goll, D.S., Helfenstein, J. and Hou, E.: Toward a Global Model for Soil Inorganic
620   Phosphorus Dynamics: Dependence of Exchange Kinetics and Soil Bioavailability on Soil Physicochemical Properties.
621   Glob. Biogeochem. Cycle, 36, e2021GB007061, https://doi.org/10.1029/2021GB007061, 2022.

622 Wang, Y.P., Law, R.M. and Pak, B.: A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere.
623   Biogeosciences, 7, 2261-2282, https://doi.org/10.5194/bg-7-2261-2010, 2010.

624 Wardle, D.A., Walker, L.R., Bardgett, R.D. and Sveriges, L.: Ecosystem Properties and Forest Decline in Contrasting Long-
625   Term Chronosequences. Science, 305, 509-513, https://doi.org/10.1126/science.1098778, 2004.

626 Weihrauch, C. and Opp, C.: Ecologically relevant phosphorus pools in soils and their dynamics: The story so far. Geoderma,
627   325, 183-194, https://doi.org/10.1016/j.geoderma.2018.02.047, 2018.

628 Wright, M.N. and Ziegler, A.: ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R. J.
629   Stat. Softw., 77, 1-17, https://doi.org/10.18637/jss.v077.i01, 2017.

630 Xu, X., Schimel, J.P., Janssens, I.A., Song, X., Song, C., Yu, G., Sinsabaugh, R.L., Tang, D., Zhang, X. and Thornton, P.E.:
631   Global pattern and controls of soil microbial metabolic quotient. Ecol. Monogr., 87, 429-441,
632   https://doi.org/10.1002/ecm.1258, 2017.

633 Yang, X. and Post, W.M.: Phosphorus transformations as a function of pedogenesis: A synthesis of soil phosphorus data using
634   Hedley fractionation method. Biogeosciences, 8, 2907-2916, https://doi.org/10.5194/bg-8-2907-2011, 2011.

635 Yang, X., Post, W.M., Thornton, P.E. and Jain, A.: The distribution of soil phosphorus for global biogeochemical modeling.
636   Biogeosciences, 10, 2525-2537, https://doi.org/10.5194/bg-10-2525-2013, 2013.

637 Yang, X., Thornton, P.E., Ricciuto, D.M. and Post, W.M.: The role of phosphorus dynamics in tropical forests – a modeling
638   study using CLM-CNP. Biogeosciences, 11, 1667-1681, https://doi.org/10.5194/bg-11-1667-2014, 2014.

639 Zhang, C., Tian, H., Liu, J., Wang, S., Liu, M., Pan, S. and Shi, X.: Pools and distributions of soil phosphorus in China. Glob.
640   Biogeochem. Cycle, 19, GB1020, https://doi.org/10.1029/2004GB002296, 2005.

641 Zhang, Y., Guo, Y., Tang, Z., Feng, Y., Zhu, X., Xu, W., Bai, Y., Zhou, G., Xie, Z. and Fang, J.: Patterns of nitrogen and
642   phosphorus pools in terrestrial ecosystems in China. Earth Syst. Sci. Data, 13, 5337-5351, https://doi.org/10.5194/essd-
643   13-5337-2021, 2021.

644
645
646
647

648 **Table 1. Summary of training data used to predict soil P pool concentrations.** P10 and P90 indicate percentile rank of 10%

649 and 90%, respectively. Proportions from literature (PFL) and proportions from gridded maps (PFGM) indicate proportions of

650 measurements from the literature and extracted from global gridded maps, respectively.

Group	Variables	Unit	Min	P10	Median	P90	Max	PFL[*]	PFGM[#]
Climate	MAT	°C	-12	1.1	12.8	25.7	30.0	96%	4%
	MAP	mm yr^{-1}	10	414	970	2750	5180	96%	4%
Soil property	Total P	mg kg^{-1}	4.8	114.0	455.5	1107.9	14973.6	100%	0%
	SOC	g kg^{-1}	<0.1	4.8	24.4	130	545.2	87%	13%
	Soil pH	unitless	3.0	4.2	5.7	8.1	10.5	92%	8%
	Soil clay	g kg^{-1}	<0.1	70.0	195.5	410.7	945.5	52%	48%
	Soil sand	g kg^{-1}	<0.1	164.9	420.0	757.6	982.0	49%	51%
	Depth	cm	0.5	4.2	10.0	47.5	450.0	100%	0%
	Soil order	unitless	12 USDA soil orders					80%	20%
Vegetation	Biome	unitless	8 major biomes					0%	100%
Topography	Elevation	m	-2	37	616	3015	4813	85%	15%

651 MAT: Mean annual temperature; MAP: Mean annual precipitation; SOC: Soil organic carbon.
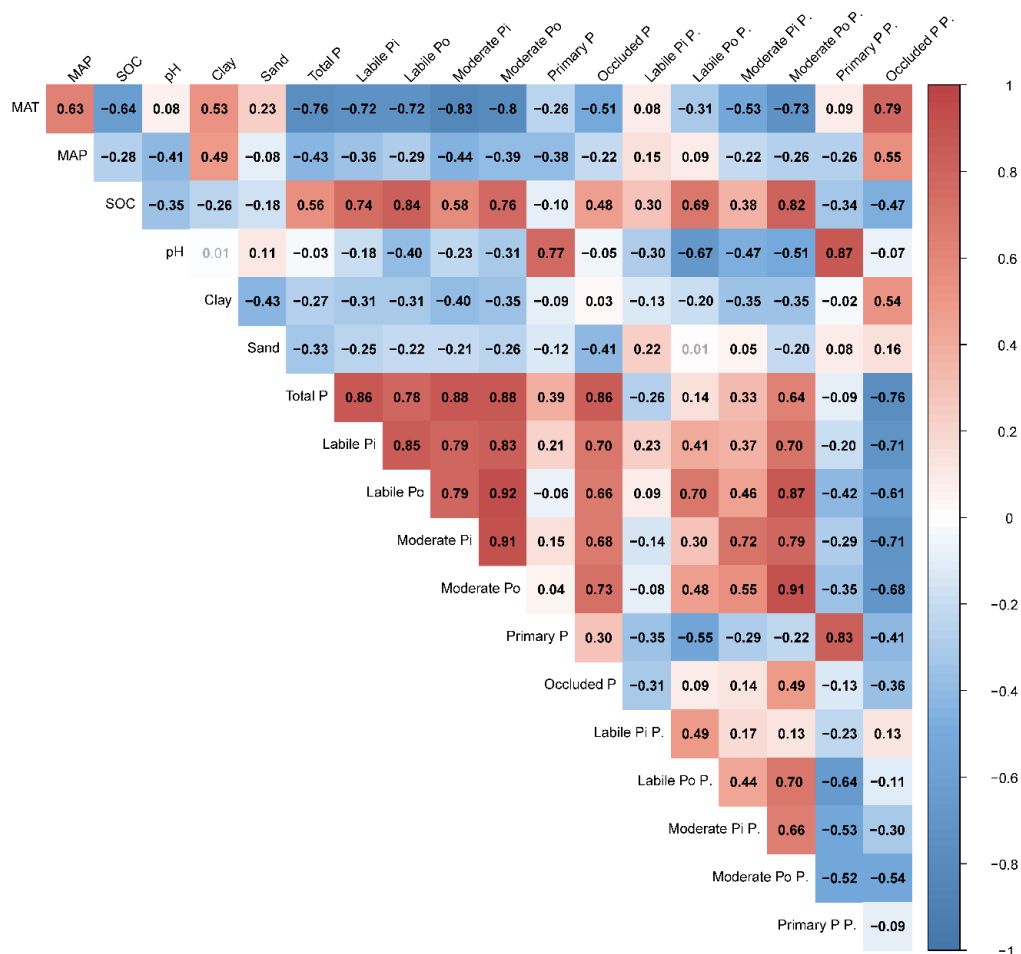
652

Biogeosciences
Discussions

653 **Table 2. Statistical summary of P pools in global (semi-)natural soils.** Results based on our collected sites database. P10,

654 P25, P75, and P90 indicate percentile rank of 10%, 25%, 75%, and 90%, respectively.

	Count	Min	P10	P25	Median	Mean	P75	P90	Max
Concentration (mg kg^{-1})									
Labile Pi	1722	<0.1	2.2	6.2	14.3	37.1	34.3	78.6	961.5
Labile Po	1567	0.1	2.5	5.9	14.0	31.1	35.0	85.2	422.0
Moderately labile Pi	1742	<0.1	4.0	10.0	25.0	58.4	57.7	122.4	4520.9
Moderately labile Po	1588	0.2	8.3	22.1	60.8	120.3	155.1	333.4	1876.7
Primary P	1629	<0.1	1.2	4.7	38.9	106.8	145.0	328.3	1560.0
Occluded P	1453	0.8	34.5	86.2	178.0	260.5	309.6	532.9	2845.4
Proportion of total P (%)									
Labile Pi	1448	<0.1	0.6	1.7	4.0	5.9	7.7	13.6	54.5
Labile Po	1331	<0.1	0.8	1.7	4.1	5.9	7.8	13.1	63.4
Moderately labile Pi	1448	<0.1	0.9	3.0	7.5	9.3	12.9	20.2	55.6
Moderately labile Po	1384	0.1	3.1	8.0	18.0	19.5	27.1	38.5	74.1
Primary P	1448	<0.1	0.5	1.6	7.9	19.0	29.4	60.9	95.1
Occluded P	1448	0.8	15.4	26.8	42.4	41.9	56.4	67.9	92.3

655

Biogeosciences
Discussions
Open Access
EGU

Table 3. Coefficients of Pearson correlations among proportions and concentrations of soil P pools. Results based on the predicted maps for soils at depths of 0-30 cm. Coefficients with $P < 0.001$ are shown in black and bold. Labile Pi P. indicated the labile Pi proportion. The same meanings to the Labile Po P., Moderately labile Pi P., Moderately labile Po P., Primary P P., and Occluded P P..

|	MAP	SOC	pH	Clay	Sand	Total P	Labile Pi	Labile Po	Moderate Pi	Moderate Po	Primary P	Occluded P	Labile Pi P.	Labile Po P.	Moderate Pi P.	Moderate Po P.	Primary P P.	Occluded P P.
MAT	0.63	−0.64	0.08	0.53	0.23	−0.76	−0.72	−0.72	−0.83	−0.8	−0.26	−0.51	0.08	−0.31	−0.53	−0.73	0.09	0.79
MAP		−0.28	−0.41	0.49	−0.08	−0.43	−0.36	−0.29	−0.44	−0.39	−0.38	−0.22	0.15	0.09	−0.22	−0.26	−0.26	0.55
SOC			−0.35	−0.26	−0.18	0.56	0.74	0.84	0.58	0.76	−0.10	0.48	0.30	0.69	0.38	0.82	−0.34	−0.47
pH				0.01	0.11	−0.03	−0.18	−0.40	−0.23	−0.31	0.77	−0.05	−0.30	−0.67	−0.47	−0.51	0.87	−0.07
Clay					−0.43	−0.27	−0.31	−0.31	−0.40	−0.35	−0.09	0.03	−0.13	−0.20	−0.35	−0.35	−0.02	0.54
Sand						−0.33	−0.25	−0.22	−0.21	−0.26	−0.12	−0.41	0.22	0.01	0.05	−0.20	0.08	0.16
Total P							0.86	0.78	0.88	0.88	0.39	0.86	−0.26	0.14	0.33	0.64	−0.09	−0.76
Labile Pi								0.85	0.79	0.83	0.21	0.70	0.23	0.41	0.37	0.70	−0.20	−0.71
Labile Po									0.79	0.92	−0.06	0.66	0.09	0.70	0.46	0.87	−0.42	−0.61
Moderate Pi										0.91	0.15	0.68	−0.14	0.30	0.72	0.79	−0.29	−0.71
Moderate Po											0.04	0.73	−0.08	0.48	0.55	0.91	−0.35	−0.68
Primary P												0.30	−0.35	−0.55	−0.29	−0.22	0.83	−0.41
Occluded P													−0.31	0.09	0.14	0.49	−0.13	−0.36
Labile Pi P.														0.49	0.17	0.13	−0.23	0.13
Labile Po P.															0.44	0.70	−0.64	−0.11
Moderate Pi P.																0.66	−0.53	−0.30
Moderate Po P.																	−0.52	−0.54
Primary P P.																		−0.09

662   **Figure 1. Distribution of site-level training data.** The database contains 1838 observations covering 12 USDA soil orders

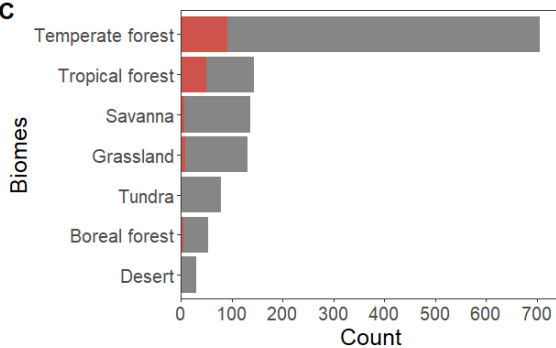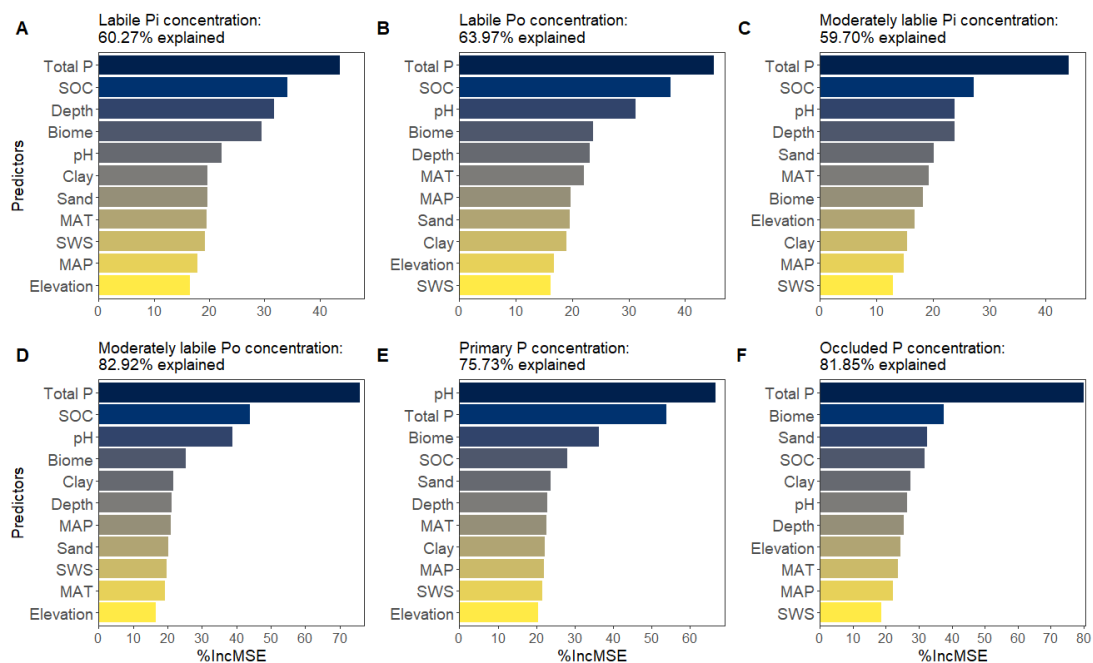663   (B) and all major terrestrial biomes (C).

664
665
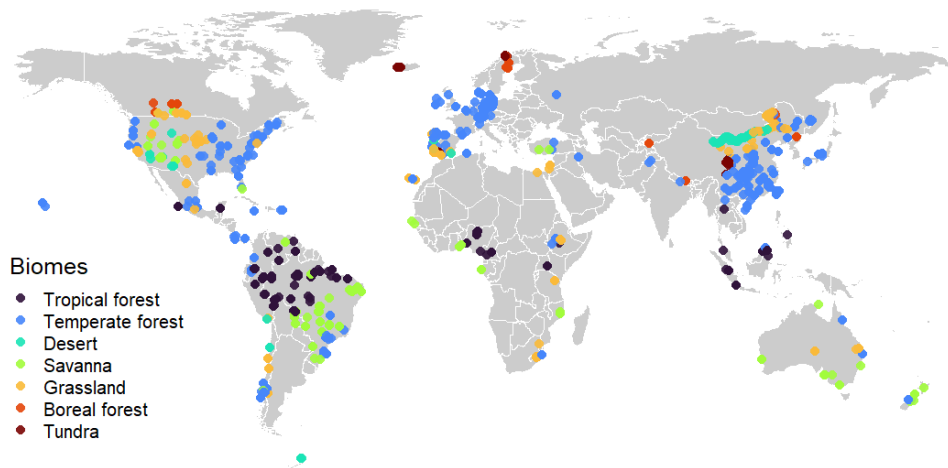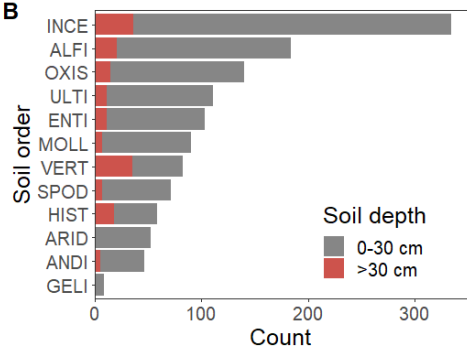
Figure 2. Relative importance of variables for predicting concentration of soil P pools quantified using random forest models. Mean decrease accuracy (%IncMSE) indicates the relative importance of each variable for predicting soil P pools. SWS: soil weathering stage.

672 **Figure 3. Relative importance of variables for predicting proportions of soil P pools quantified using random forest**

673 **models.** Mean decrease accuracy (%IncMSE) indicates the relative importance of each variable for predicting soil P pools.

674 SWS: soil weathering stage.
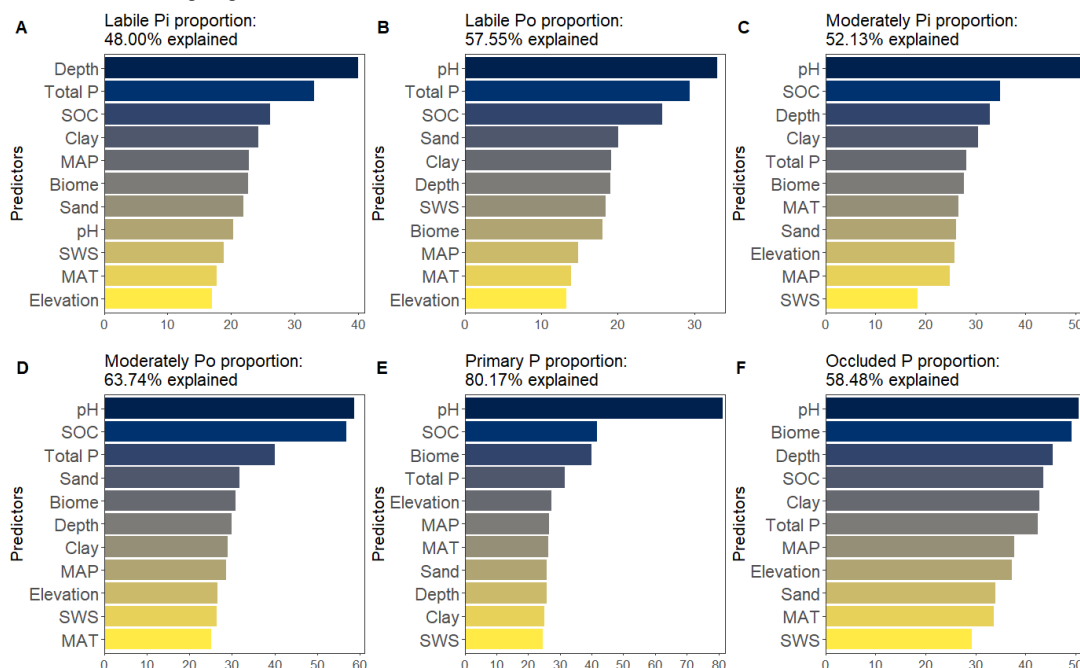
675

676

677 **Figure 4. Average concentrations of P pools and their proportions of total soil P concentration across soil weathering**

678 **stages and biomes. Labile and moderately labile Po form the organic pool.** Results based on global estimates for 0-30 cm

679 depth. Dry vegetation combines grassland and savanna biomes to simplify the figure.
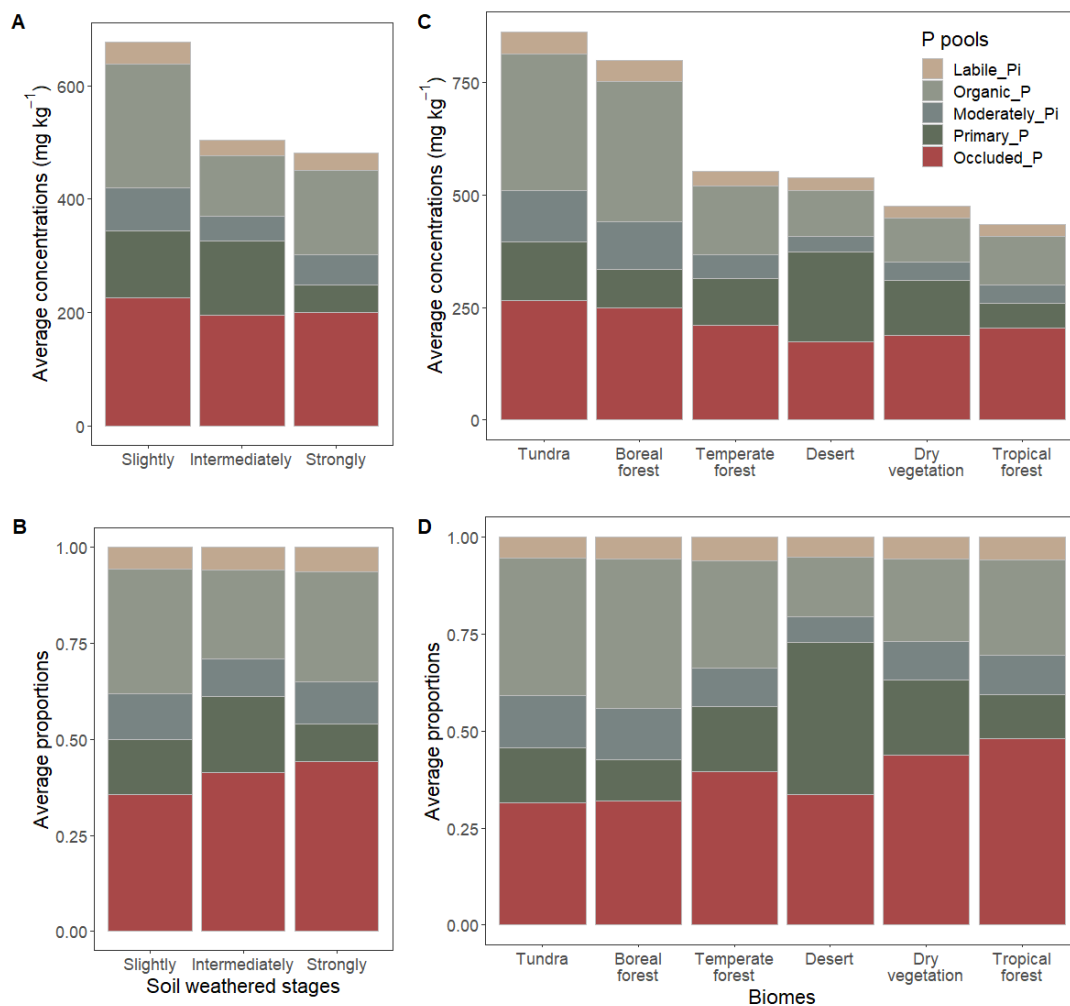

680

681

682

683 **Figure 5. Partial dependence plots showing dependence of centralized soil P pool concentrations on predictors.** To

684 simplify comparison, partial dependent analysis results are centralized in this plot.
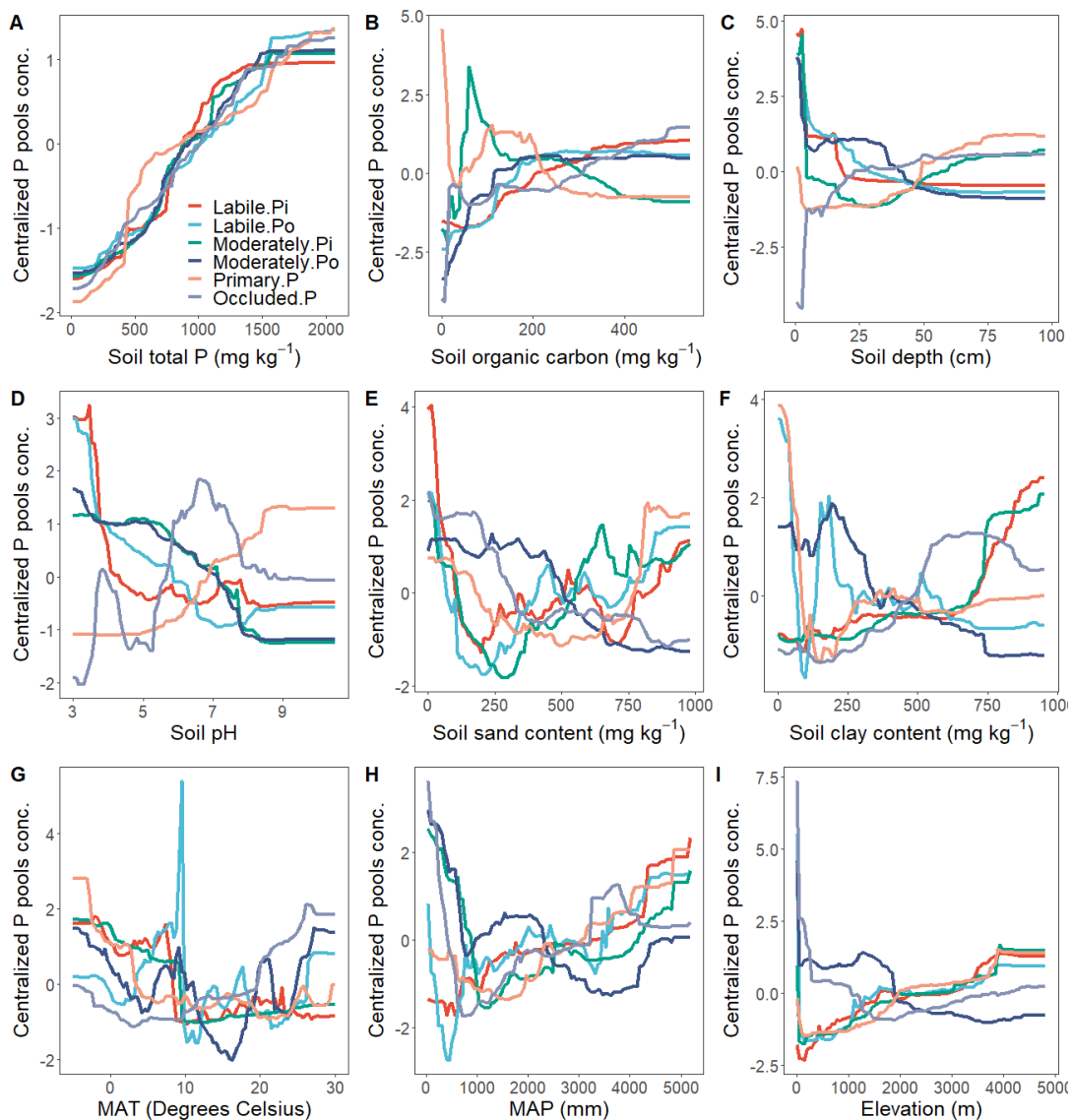

685

686

687

688 **Figure 6. Partial dependence plots showing the dependence of soil P pool proportions on predictors.** The sum of all six

689 P pool proportions was standardized to 1, though the sum from partial dependent analysis is usually not equal to 1 due to
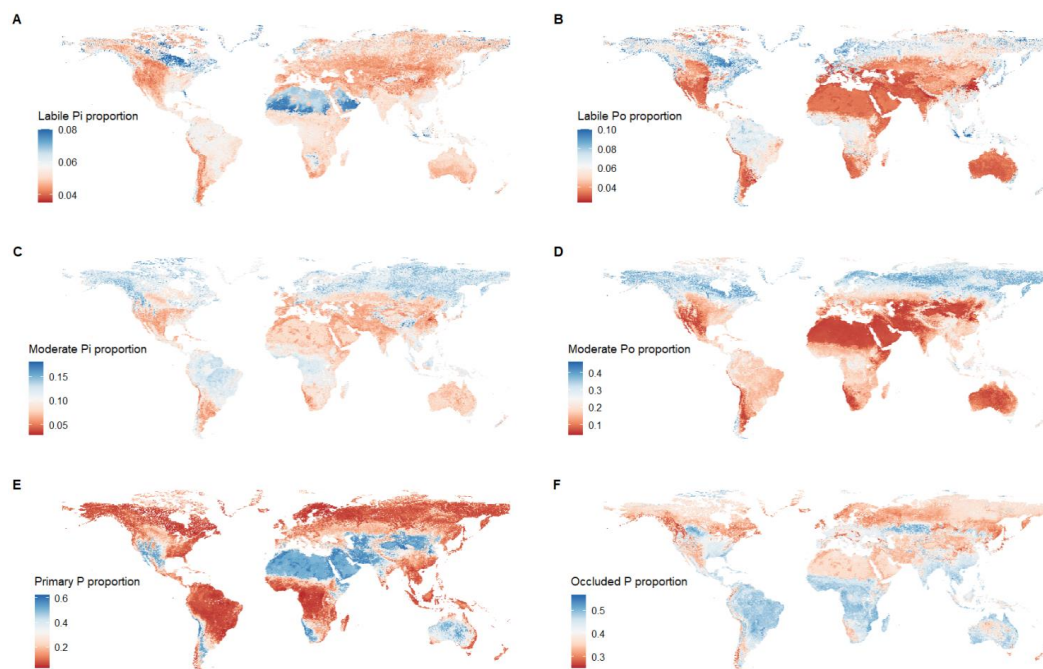
690 uncertainty.
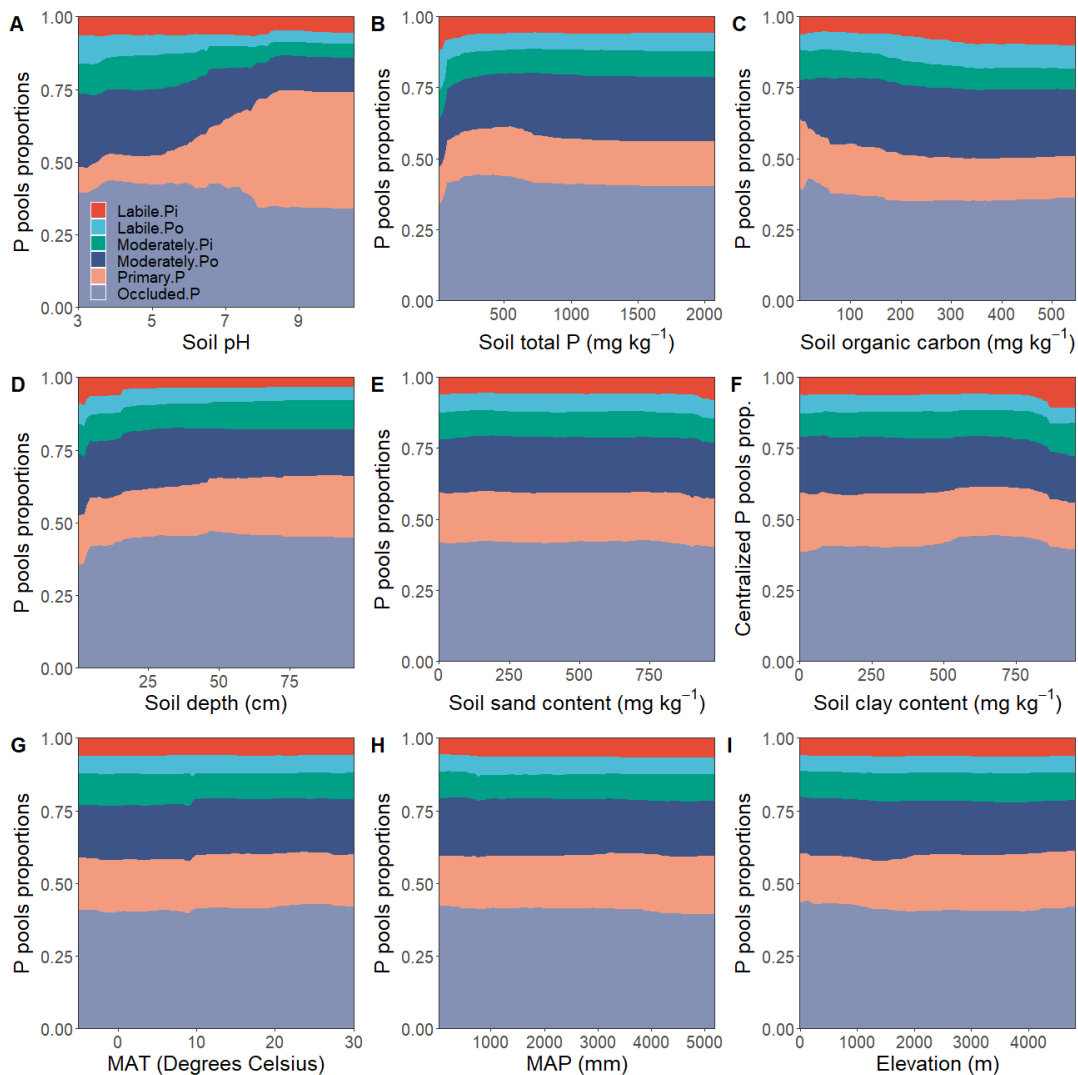
691
692

693 **Figure 7. Global maps of P pool concentrations at depths of 0-30 cm.** Note that croplands and other heavily influenced
694 areas were not masked from the maps, so soils in these areas can be used to represent soils without extensive anthropogenic
695 activity.
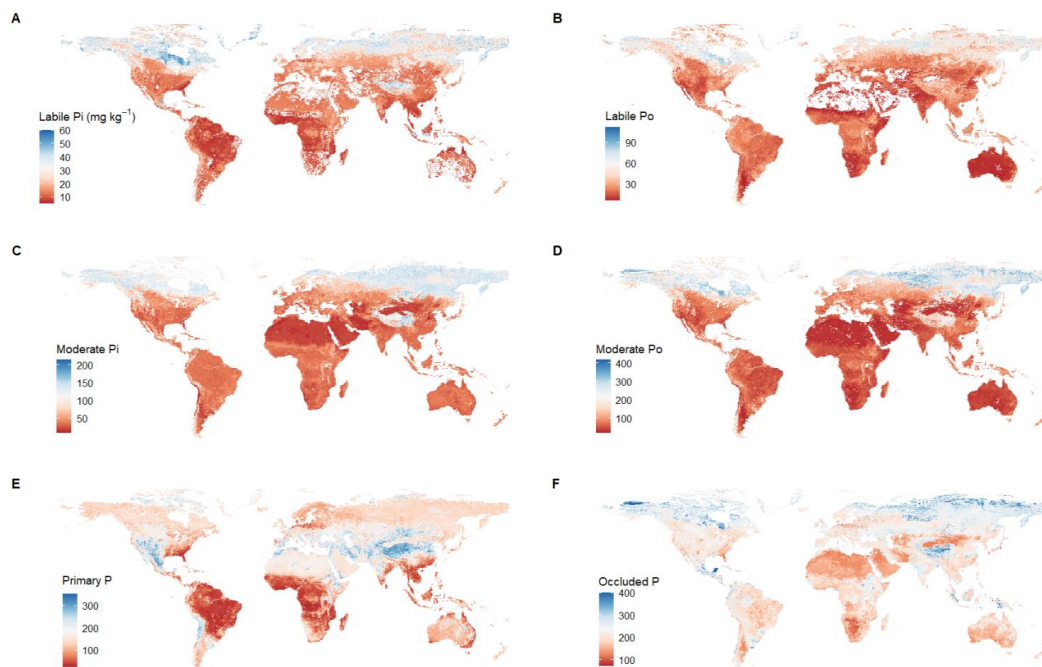
696
697
698

699 **Figure 8. Global maps of P pool proportions at depths of 0-30 cm.** Note that croplands and other heavily influenced areas

700 were not masked from the maps.

701

702