



**HAL**  
open science

## Phosphorus in agricultural soils: drivers of its distribution at the global scale

Bruno Ringeval, Laurent Augusto, Herve Monod, Dirk van Apeldoorn, Lex Bouwman, Xiaojuan Yang, David Achat, Louise Chini, Kristof van Oost, Bertrand Guenet, et al.

### ► To cite this version:

Bruno Ringeval, Laurent Augusto, Herve Monod, Dirk van Apeldoorn, Lex Bouwman, et al.. Phosphorus in agricultural soils: drivers of its distribution at the global scale. *Global Change Biology*, 2017, 23 (8), pp.3418 - 3432. 10.1111/gcb.13618 . hal-01708836

**HAL Id: hal-01708836**

**<https://hal.science/hal-01708836>**

Submitted on 4 Jun 2018

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - ShareAlike 4.0 International License

DR. BRUNO RINGEVAL (Orcid ID : 0000-0001-8405-1304)

Received Date : 11-Aug-2016

Accepted Date : 27-Nov-2016

Article type : Primary Research Articles

## Phosphorus in agricultural soils: drivers of its distribution at the global scale

### Running head: Global distribution of agricultural soil P

Bruno Ringeval<sup>1\*</sup>, Laurent Augusto<sup>1</sup>, Hervé Monod<sup>2</sup>, Dirk van Apeldoorn<sup>3</sup>, Lex Bouwman<sup>3</sup>, Xiaojuan Yang<sup>4</sup>, David L. Achat<sup>1</sup>, Louise P. Chini<sup>5</sup>, Kristof Van Oost<sup>6</sup>, Bertrand Guenet<sup>7</sup>, Rong Wang<sup>7,8,9</sup>, Bertrand Decharme<sup>10</sup>, Thomas Nesme<sup>11</sup>, Sylvain Pellerin<sup>1</sup>

<sup>1</sup> INRA, UMR 1391 ISPA, F-33140 Villenave d'Ornon, France

<sup>2</sup> INRA, UMR MIAj, Domaine de Vilvert, F78352 Jouy-en-Josas Cedex, France

<sup>3</sup> Department of Earth Sciences–Geochemistry, Faculty of Geosciences, Utrecht University, 3508 TA, Utrecht, The Netherlands

<sup>4</sup> Environmental Science Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831-6335, USA

<sup>5</sup> Department of Geography, University of Maryland, College Park, MD 20742, USA

<sup>6</sup> George Lemaître Centre for Earth and Climate Research, Earth & Life Institute, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium

<sup>7</sup> Laboratoire des Sciences du Climat et de l'Environnement, LSCE/IPSL, CEA-CNRS-UVSQ, Université Paris-Saclay, F-91191 Gif-sur-Yvette, France

<sup>8</sup> Laboratory for Earth Surface Processes, College of Urban and Environmental Sciences, Peking University, 100871 Beijing, China

<sup>9</sup> Sino-French Institute for Earth System Science, College of Urban and Environmental Sciences, Peking University, 100871 Beijing, China

<sup>10</sup> Centre National de Recherches Meteorologiques, 42 Av. Gaspard Coriolis, 31057 Toulouse Cedex, France

<sup>11</sup> Univ. Bordeaux, Bordeaux Sciences Agro, UMR 1391 ISPA, F-33170 Gradignan, France

\* Corresponding author: [bruno.ringeval@inra.fr](mailto:bruno.ringeval@inra.fr), +33 5 57 12 25 05

Keywords: phosphorus, agricultural soils, modelling, global scale, biogeochemical cycles

Type of paper: Primary Research Articles

### Short Summary

Numerous processes drive the global spatial distribution of phosphorus (P) in agricultural soils, but their relative roles remain unclear. Thanks to a modelling approach, we found that

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/gcb.13618

This article is protected by copyright. All rights reserved.

almost all of the variability of the cropland soil total P could be explained by the distribution of the soil biogeochemical background (BIOG), while both BIOG and farming practices explained the spatial variability of the fraction of total P involved in plant nutrition. Our work showed how the different drivers were combined to explain the global distribution of agricultural soil P, which is useful for further investigations on the limiting effect of P on global crop production.

### **Abstract Figure**

Phosphorus in agricultural soils: drivers of its distribution at the global scale; B. Ringeval\*, L. Augusto, H. Monod, D. van Apeldoorn, L. Bouwman, X. Yang, D.L. Achat, L.P. Chini, K. Van Oost, B. Guenet, R. Wang, B. Decharme, T. Nesme, S. Pellerin.

Numerous processes drive the global spatial distribution of phosphorus (P) in agricultural soils, but their relative roles remain unclear. Thanks to a modelling approach, we found that almost all of the global spatial variability of total soil P in cropland soils ( $P_{TOT}$ ) could be explained by the distribution of the soil biogeochemical background (that determines the P content of soils at the conversion to agriculture, BIOG), while both BIOG and farming practices (FARM) explained the spatial variability of inorganic labile P ( $P_{ILAB}$ ) (~40% and ~60%, respectively).

### **Abstract**

Phosphorus (P) availability in soils limits crop yields in many regions of the world, while excess of soil P triggers aquatic eutrophication in other regions. Numerous processes drive the global spatial distribution of P in agricultural soils, but their relative roles remain unclear. Here, we combined several global datasets describing these drivers with a soil P dynamics model to simulate the distribution of P in agricultural soils and to assess the contributions of

the different drivers at the global scale. We analyzed both the labile inorganic P ( $P_{\text{ILAB}}$ ), a proxy of the pool involved in plant nutrition and the total soil P ( $P_{\text{TOT}}$ ). We found that the soil biogeochemical background corresponding to P inherited from natural soils at the conversion to agriculture (BIOG) and farming practices (FARM) were the main drivers of the spatial variability in cropland soil P content but that their contribution varied between  $P_{\text{TOT}}$  vs  $P_{\text{ILAB}}$ . When the spatial variability was computed between grid-cells at half degree resolution, we found that almost all of the  $P_{\text{TOT}}$  spatial variability could be explained by BIOG, while BIOG and FARM explained 38% and 63% of  $P_{\text{ILAB}}$  spatial variability, respectively. Our work also showed that the driver contribution was sensitive to the spatial scale characterizing the variability (grid-cell vs. continent) and to the region of interest (global vs. tropics for instance). In particular, the heterogeneity of farming practices between continents was large enough to make FARM contribute to the variability of  $P_{\text{TOT}}$  at that scale. We thus demonstrated how the different drivers were combined to explain the global distribution of agricultural soil P. Our study is also a promising approach to investigate the potential effect of P as a limiting factor for agroecosystems at the global scale.

## Introduction

Global nutrient management is a key question for agronomists working on global food security (Mueller *et al.*, 2012; Makowski *et al.*, 2013). This is crucial for phosphorus (P), which limits net primary production, and for which access to its sources is becoming increasingly uncertain due to limited rock phosphate supplies and potential geopolitical issues (Peñuelas *et al.*, 2013). In undisturbed terrestrial ecosystems, soil P is often limiting (Elser *et al.*, 2007), either because the amount of total P is low or because P is in forms that are not available for plants (Walker & Syers, 1976). In agroecosystems, the export of P due to harvesting tends to increase the soil P limitation and to maintain high crop yields, P is often supplied in the form

of chemical fertilizers or animal manure. In the 2000s, chemical P fertilizer application averaged 10kgP/ha/yr at the global scale, but showed a large heterogeneity, with a supply of 25 kgP/ha/yr in Europe vs 3 kgP/ha/yr in Africa (Liu *et al.*, 2008). A deficit in available soil P leads to a loss of attainable yield (Mueller *et al.*, 2012) while an excess of total P in soils could trigger aquatic eutrophication in surrounding water bodies through runoff and water erosion (Schindler *et al.*, 2008). The P content of agricultural soils is thus an index for agricultural soil fertility, as well as an important variable affecting water quality. High-resolution assessment of the current distribution of P in agricultural soils at the global scale is critical for the identification of areas where P is in a state of imbalance and mitigation is needed. However, such an assessment has not been possible up to now due to limited data availability and poor understanding of soil P dynamics at large scales.

The P-content of soils in currently unmanaged ecosystems can be considered to be the result of transformations that occur on a geological time-scale. While P in unmanaged soils can be predicted from local soil properties (Yang *et al.*, 2013), this approach cannot be used for agricultural systems because humans have altered the P cycle in these ecosystems (Elser & Bennett, 2011) and changed the spatial distribution of P in soils. In fact, numerous long-term, interacting and highly spatially variable processes drive the P content in agricultural soils across the world. For instance, farming practices (e.g. the amount of fertilizer applied) have the potential to re-shape both total and available soil P (Sattari *et al.*, 2012). Agricultural soils are subjected to erosion rates, which are typically one order of magnitude higher than those occurring in unmanaged soils, potentially leading to high P losses (Quinton *et al.*, 2010). Atmospheric P deposition has increased with anthropogenic emissions and is likely to play an increasing role in soil P budgets (Wang *et al.*, 2014). Internal soil dynamics processes, which are known to have small effects on total soil P, have the potential to modify the redistribution

of P within the different soil pools and thus, to modulate plant available P. Such soil dynamics depend on the soil buffering capacity (Vitousek *et al.*, 2010) and soil meteorological conditions (Buendía *et al.*, 2010), both varying spatially. Finally, land use and land cover change tend to modify the P content of agroecosystems due to the conversion of natural lands with their original soil P contents, into cultivated areas (Mackenzie *et al.*, 2002; MacDonald *et al.*, 2012). While in a given context some drivers could *a priori* be neglected as compared to others (e.g. local ratio of deposition to chemical fertilizer equal to 0.03 in countries with intensive agriculture), this could vary from one location to the other. In addition, many of these drivers have varied over time (Ringeval *et al.*, 2014), thereby making their relative contribution to the *current* global distribution of P in agricultural soils extremely difficult to decipher. Several studies have attempted to model part of the P cycle at the global scale. Some of them focused on a given driver, taken separately (e.g. (MacDonald *et al.*, 2011) focused on farming practices) and showed that the different drivers are characterized by a large spatial heterogeneity across the world. However, no study has provided an integrated assessment of the combination of these drivers. Here, we built a simple, process-based soil P dynamics model that enabled the combination of various recent spatially explicit global datasets to reconstruct the temporal evolution of P in agricultural soils during the 20<sup>th</sup> century. These global datasets enabled us to represent the drivers that are known to play a role in the current global distribution of P in agricultural soils. Our analyses focused on the distribution of total soil P and more labile P involved in plant nutrition. We first calibrated the soil P dynamics model against observations on sites. Simulated global maps of agricultural soil P were then analysed, particular attention being paid to uncertainty in the simulated soil P resulting from uncertainty in the drivers. Finally, as the core of our study, a sensitivity analysis based on a full factorial design where each driver was either variable or constant in space, was performed to estimate the contribution of each driver to the current distribution of agricultural soil P.

## Materials and Methods

### *Modelling framework*

The system that was modelled comprised P in the top 0-0.3m horizon of agricultural soils. In agricultural soils, P availability and root uptake decline substantially below the plough layer (Lynch & Brown, 2001) and long term fertilization trials show that contrasted fertilization regimes mainly affect the P availability in the plough layer with very little effect in deeper layers (Messiga *et al.*, 2012; Li *et al.*, 2017). Thus, the system modelled should ideally be the plough layer. However, the thickness of the plough layer shows strong variations in time and space as a function of the farming practices (Brandsæter *et al.*, 2011; Gronle *et al.*, 2015; Alcántara *et al.*, 2016) and no spatially explicit dataset providing plough depth was available at the global scale. We thus chose to represent the soil horizon corresponding to a fixed thickness of 0.3m which, at the global scale, is considered as encompassing a major proportion of crop roots (Jackson *et al.*, 1996). The sensitivity of our results to the thickness of the soil horizon modelled is assessed and discussed in the following sections. No vertical discretisation was considered. Our modelling framework worked at half-degree fractional resolution (0.5° latitude x 0.5° longitude). Cropland and pasture soils were distinguished in each grid-cell. Soil P pools were expressed in kgP/ha while fluxes were given in kgP/ha/yr. The modelling framework, resulting from the combination of a soil P dynamics model and global datasets, was run at 1-year time intervals.

The spatially explicit global datasets used in this study enabled us to represent the drivers that are known to play a role in shaping the P in agricultural soils over the last centuries and explain the current global distribution: the soil biogeochemical background corresponding to P inherited from natural soils at the conversion to agriculture (BIOG), soil P input/output corresponding to farming practices (FARM), land use and land cover change (LUCC), soil water

content and temperature affecting weathering and mineralization of organic matter (CLIM), losses of P through soil erosion (LOSS), atmospheric P deposition (DEPO) and soil buffering capacity (BUFF). The different drivers and their representations are described in Table 1. The datasets resulted from the combination of different measurements (satellites, on site measurements, etc.), and/or model simulations. All datasets were re-gridded to a half-degree resolution. More details about each dataset are provided in the Supporting Information.

While we focused on soil P of cropland and pasture, four land fractions were considered in each grid-cell to represent the land use and land cover change. These land fractions were: cropland (C), pasture (P), non-agricultural vegetated areas (NA) and urban areas (U). NA actually included both undisturbed and secondary vegetation re-growing in its natural state. For a given grid-cell, at the beginning of each time-step  $y$  ( $y$  represents a specific year in the model), the effect of land use and land cover change on the evolution of each soil P pool was represented by equation (1):

$$P^C(y) = [P^C(y-1) \cdot (f^C(y-1) - \sum_{i \in P, NA, U} \Delta_C^i) + \sum_{i \in P, NA, U} P^i(y-1) \cdot \Delta_i^C] / f^C(y) \quad (1a)$$

$$P^P(y) = [P^P(y-1) \cdot (f^P(y-1) - \sum_{i \in C, NA, U} \Delta_P^i) + \sum_{i \in C, NA, U} P^i(y-1) \cdot \Delta_i^P] / f^P(y) \quad (1b)$$

where  $P^x$  represented a given soil P pool (apatite, occluded, etc.; see equation (3)) of the land fraction  $x$  (in kgP/ha);  $f^x$  was the grid-cell fraction covered by  $x$  (no unit) and  $\Delta_{x1}^{x2}$  was the land fraction covered by  $x1$  converted into  $x2$  from  $y-1$  to  $y$ .  $f^C$ ,  $f^P$  and the different land-conversions ( $\Delta$ ) were provided by the Land Use Harmonization dataset (LUHa.v1) (Hurtt *et al.*, 2011) and correspond to the LUCC driver. Note that transition from urban areas to cropland/pasture ( $\Delta_U^C$  and  $\Delta_U^P$ ) were considered as null.  $P^{NA}$  is the P inherited from natural soils at the conversion to agriculture and correspond to the driver BIOG.  $P^{NA}$  was also used as



initial conditions for agricultural soils (i.e. in 1700, see below). For each grid-cell and at any moment,  $P^{NA}$  was set to the

current P in unmanaged soil estimated by (Yang *et al.*, 2013), i.e.:

for any y,  $P^{NA}(y) = P^{Yang}$

(2)

Through this equation, we assumed that the P inherited from natural soils at the conversion to agriculture could be represented by prescribing the current P in unmanaged soils to all soils converted to agricultural soils over the last 300 years. This was a key-assumption of our approach. This could be limiting, in particular in regions where shifting cultivation occurs, leading to modify the P content of soils covered by natural vegetation. We also neglected soil P input corresponding to forest biomass left on soil at the time of conversion (Beck & Sanchez, 1996).

In each time-step, once equation (1) had been applied,  $P^C$  and  $P^G$  evolved for all soil pools according to other following drivers: soil input/output resulting from farming practices (FARM), P losses due to soil erosion (LOSS), and atmospheric P deposition (DEPO). In addition, the soil P dynamics model allowed us to redistribute P within the different soil P pools according to soil buffering capacity (BUFF) and soil conditions such as soil temperature and humidity (CLIM) (Figure 1).

#### *Soil P dynamics model*

The design of the soil P dynamics model (i.e. which pools are represented, Figure 1) was chosen to be consistent with the Hedley fractionation method (Hedley & Stewart, 1982; Tiessen *et al.*, 1984). The Hedley method fractionates soil P into different forms of P that are removed sequentially with successively stronger extraction agents. While relating the fractions re-

Accepted Article  
moved by chemical products to soil mechanisms is prone to difficulties (Tiessen *et al.*, 1984), these fractions are usually merged into various inorganic and organic pools and classified according to their plant availability, which gives a comprehensive picture of the different forms of P in soils (Cross & Schlesinger, 1995) (Table S1). In addition, to our knowledge, Yang *et al.* (2013) is the only published database providing information about the global distribution of P in unmanaged soils and it was based on the concepts developed in the Hedley method.

Total P ( $P_{TOT}$ ) was defined as the sum of all represented soil P pools:

$$P_{TOT} = P_{OCC} + P_{APA} + P_{SEC} + P_{ILAB} + P_{OSTA} + P_{OLAB} \quad (3)$$

where  $P_{OCC}$ ,  $P_{APA}$ ,  $P_{SEC}$ ,  $P_{ILAB}$ ,  $P_{OSTA}$ ,  $P_{OLAB}$  represented occluded P, P in apatite, P bound on secondary minerals, labile inorganic P, stable organic P and labile organic P, respectively. Our analysis focused on two variables: total soil P ( $P_{TOT}$ ) and inorganic labile P ( $P_{ILAB}$ ). We chose  $P_{TOT}$  because it represents the total amount of P and is therefore an integrative variable for several processes, whereas  $P_{ILAB}$  is used as a proxy for soil P available to plants on a short time-scale (Cross & Schlesinger, 1995).

As a first approach, we chose to represent as few soil P dynamics processes as possible. Processes like root-induced processes to overcome P limitation or P immobilization/release by microbial biomass were not taken into account in our model in particular because their representation at the global scale remains uncertain (see Discussion). The flux parameterizations used in our study were based on studies performed with Dynamic Global Vegetation Models (DGVMs) (Wang *et al.*, 2010; Goll *et al.*, 2012) because DGVMs have been used at the global scale and are process-based, which were two aims of our approach. Parameterizations are described in details below where  $fP_X^Y$  defines the P flux from pool X to pool Y.

Weathering, occlusion and mineralization are described thanks to a first order kinetic. Apatite weathering is represented by:

$$fP_{APA}^{SOLU} = k_w \cdot g_1(T) \cdot g_2(W) \cdot P_{APA} \quad (4)$$

where  $g_1$  and  $g_2$  (unitless) described the sensitivity to soil temperature (T, in °C) and relative soil water content (W, unitless), respectively. Before calibration,  $k_w = 0.0001 \text{ yr}^{-1}$  following Buendía *et al.* (2010).  $k_w$  is the base rate at a temperature of  $\sim 15^\circ\text{C}$  (Table 3 of Buendía *et al.* (2010)) thus this value is used as reference temperature in  $g_1(T)$ .  $g_1(T)$  could be approached by an Arrhenius equation following Hartmann *et al.* (2014) (activation energy = 61 kJ/mol; averaged for granitic rocks (Hartmann *et al.*, 2014)). We used a  $Q_{10}$  formulation to be consistent with equation (6). A  $Q_{10}$  of 2.4 gave the same analytical result as the Arrhenius equation described in Hartmann *et al.* (2014) for common temperature ranges and was thus chosen. Following Buendía *et al.* (2010),  $g_2(W) = W$ .

The occlusion is parameterized as follows:

$$fP_{SEC}^{OCC} = k_{occ} \cdot P_{SEC} \quad (5)$$

The pre-calibrated  $k_{occ}$  is  $1.2e^{-5} \text{ yr}^{-1}$  (Yang *et al.*, 2014). In this first attempt, we did not consider de-occlusion processes. Thus, once a molecule of P belongs to the  $P_{OCC}$  pool, it cannot be involved in the plant nutrition any more.

The mineralization of organic matter is described thanks to:

$$fP_{OSTA}^{ILAB} = k_{m1} \cdot h_1(T) \cdot h_2(W) \cdot P_{OSTA} \quad (6a)$$

$$fP_{OLAB}^{ILAB} = k_{m2} \cdot h_1(T) \cdot h_2(W) \cdot P_{OLAB} \quad (6b)$$

where  $h_1$  and  $h_2$  represents the sensitivity to soil temperature and soil water content and  $k_{m1}$ ,  $k_{m2}$  are turnover rates for stable and labile pools (in  $\text{yr}^{-1}$ ). These pools are considered to be biologically available over different time scales (“short”, “intermediate”) (Cross & Schlesinger, 1995) but the literature about the Hedley method did not provide any accurate estimates of these time scales. Models of soil carbon usually consider that only few percent of

the recalcitrant pool could be mineralized each year. On the other hand, the comparison between labile organic P measured using a Hedley fractionation method (fraction  $\text{NaHCO}_3$  in (Oberson *et al.*, 1993)) and a kinetic of mineralization (Oehl *et al.*, 2004) on the same site suggests that the labile organic pool is mineralized very quickly. Following this information, we prescribed  $k_{m1}=0.02\text{yr}^{-1}$  and  $k_{m2}=0.60\text{yr}^{-1}$  as default values before the calibration step. Following our strategy, the same functions as used in the ORCHIDEE (Krinner *et al.*, 2005) DGVM to characterize the carbon mineralization were prescribed to  $h_1(T)$  and  $h_2(W)$  ( $Q_{10}=2$  and reference temperature =  $30^\circ\text{C}$  for  $h_1$ ;  $h_2(W)=-1.1*W^2+2.4*W-0.29$ ).

Following (Wang *et al.*, 2010; Goll *et al.*, 2012), we used a Langmuir equation to describe the equilibrium between labile inorganic P and P bound on secondary minerals equilibrium:

$$P_{SEC} = S_{max} \cdot \left[ \frac{K_s \cdot P_{ILAB}}{1 + K_s \cdot P_{ILAB}} \right] \quad (7)$$

where  $S_{max}$  is the maximum capacity of secondary minerals to bind P (in kgP/ha) and  $K_s$  is a coefficient (in ha/kgP). The suitability of this equation to describe the equilibrium between  $P_{ILAB}$  and  $P_{SEC}$  is discussed in the last section. ( $K_s, S_{max}$ ) vary as function of soil order following (Wang *et al.*, 2010) (Table 1 and Table S2).

#### *Temporal variations of the drivers and initial conditions*

Our analyses mainly focused on the soil P simulated for the year of 2005. Because the current soil P is a result of the changes in the different drivers during the past, we performed simulations for the 1700-2005 period. Variations in time of the different drivers are described in Figure S1. Overall, only the temporal variations of FARM, LUCC, DEPO and in a lesser extent LOSS were considered. The choice of 1700 as the starting year corresponds to a compromise between time-periods for which information about landuse (Hurtt *et al.*, 2011) and farming practices (Bouwman *et al.*, 2011) are available (from 1500 and 1900 respectively). We prescribed the values provided by (Yang *et al.*, 2013) for unmanaged soils ( $P^{Yang}$ ) to agri-

cultural soil P pools in 1700 (hereafter referred to as the initial conditions), i.e.:

$$P^C(1700) = P^P(1700) = P^{NA}(1700) = P^{Yang}$$

(8)

where P represented a given soil P pool (apatite, occluded, etc.; see equation (3)). We assumed that farming practices before 1700 led to negligible modifications of the soil P pools of agricultural soil in 2005 since most P was supposed to be recycled locally; the effect of this assumption on the results was tested by varying the starting year (Supporting Information).

#### *Calibration of the soil P dynamics model*

The calibration focused on base rates used in equations 4-6:  $k_w$ ,  $k_{occ}$ ,  $k_{m1}$ ,  $k_{m2}$ . The parameters used in equation 7 correspond to the driver BUFF and are considered in next sections. To calibrate our soil P dynamics model, we compiled Hedley measurements recorded in long-term experiments on cropland/pasture sites from the literature. We found 11 references providing local and observed variables to force our model instead of the ones provided by global datasets. This allowed us to calibrate the intrinsic abilities of our model to reproduce the observed temporal soil P dynamics. Simulations performed in that case are called “simulations on sites” hereafter. The 11 references found correspond to 12 locations and 49 “treatments” including variation in fertilizers applied (frequency, amount and form), crop rotation and soil type (Table S3). The 12 locations encompass a relative diversity in latitude (Figure S2). The different treatments in each location were considered independently. Few treatments on the same site (e.g. different crop rotations in Ref 9 in Table S3 while crop rotation is not represented in the model) could be considered as pseudo-replicates, but removing them did not change the calibration or final match between observations and the model (not shown).

Simulations were performed on the period of available measurements. Measurements of ei-

ther soil P in uncultivated sites at the end of the experiments or soil P under fertilization treatments at the beginning of the experiments have been used as initial conditions (BIOG).

Observed farming practices (fertilizer rate application, etc.) were used as input into our model to prescribe FARM. In simulations on sites, soil erosion, CLIM, DEPO were provided by global datasets while no change in land use and land cover change (LUCC) was prescribed.

The observed USDA soil order was used to define the parameters in the Langmuir equation (BUFF). For simulations on sites, the model was run on the soil horizon characterized by a bottom depth closest to -0.3m (i.e. 0.3m below the soil surface) for which measurements were available and which encompasses the plough layer. For a few sites, we have no information about the plough depth and measurements were available only for a very thin soil layer (e.g. 0-0.1m; Table S3). The bulk density required to translate observations from mgP/kg to kgP/ha and to compute losses were derived from observations, either directly or indirectly (i.e. as a function of the available information, by using carbon content (Augusto *et al.*, 2010), soil organic matter or the bulk density measured at different depths or in different treatments for the same location).

For each treatment, we computed the Root Mean Square Error (RMSE) between observed and modelled soil P pools for the last year of available observations. Then, we summed the RMSE of all treatments, expressed it in percent of averaged observations and used this value as criteria for calibration. The calibration was made in two stages: first, we focused on each parameter independently, taken alone and searched for a value that minimized the criteria for one given soil P pool. For each parameter, the pool chosen is the pool whose size is sensitive (almost) only to that parameter:  $P_{APA}$  for  $k_w$ ,  $P_{OCC}$  for  $k_{occ}$ ,  $P_{OSTA}$  for  $k_{m1}$ ,  $P_{OLAB}$  for  $k_{m2}$ . In a second stage, we made all parameters vary together and searched for parameter values minimizing the criteria for  $P_{ILAB}$ . If necessary, a compromise between the calibrated values of these two stages was made.

### *Estimation of the contribution of each driver*

The current spatial variability of both  $P_{TOT}$  and  $P_{ILAB}$  was quantified thanks to the variance (named VAR hereafter) of both variables simulated for 2005. The variance was computed at two spatial scales: either at the “grid-cell scale” (i.e. using all grid-cells at half degree resolution with a non-zero cropland (or pasture) fraction in our simulation for 2005) or at “continental scale” (i.e. using 7 large World regions, namely: North America, Central and South America, Africa, Oceania, Western Europe, Asia, Russia; Figure S2). Note that VAR was weighted by the cropland/pasture area of each grid-cell in the computation at grid-cell scale. Cropland/pasture area of each grid-cell was also used to compute averaged  $P_{TOT}$  and  $P_{ILAB}$  for each World region.

Each driver encompassed several variables and drivers were potentially correlated spatially, which made the method of estimating their contributions complicated. To measure the contribution of each driver to the current spatial variability of  $P_{TOT}$  and  $P_{ILAB}$ , we assessed the effect of removing the spatial variability of a given driver on the simulated VAR of  $P_{TOT}$  and  $P_{ILAB}$ . A full factorial design was generated with six factors at two levels, resulting in  $n_{runs}=2^6$  simulation runs. The six factors (called F hereafter) corresponded to all drivers except land use and land cover change (i.e. BIOG, LOSS, BUFF, CLIM, FARM, DEPO). For each factor F, level (F=+1) meant that the spatial variability of F was taken into account in the simulation run, while level (F=-1) meant that its spatial variability was suppressed by prescribing a constant value to the cropland (or pasture) fractions of all grid-cells. As it was not possible to remove the spatial variability of the land use and land cover change artificially (i.e. LUCC=-1), we restricted the factors of the full factorial design to all drivers except LUCC. Indeed, using a constant value for the cropland/pasture fractions of all grid-cells at a given time-step did not allow us to maintain a temporal consistency between fractions and fraction conversions.

A table of the full factorial design is provided in Table S4. The first-order effect of factor F on

VAR was estimated by (Kobilinsky & Monod, 1991):

$$e(F) = \frac{1}{n_{runs}/2} \cdot \sum_{j=1}^{n_{runs}} [f_j \cdot VAR_j] \quad (9)$$

where  $f_j$  denotes the value +1/-1 of F at run  $j$  and  $VAR_j$  denotes the VAR at run  $j$ . To estimate interactions, the same equation was used, with  $f_j$  now the product of the +1/-1 values of the factors involved in the interaction. Note that the contribution was then expressed in percent of the VAR computed when all factors were set to +1. Only a fraction of the contribution of LUCC could be estimated thanks to the run where the 6 factors (BIOG, LOSS, BUFF, CLIM, FARM, DEPO) were set to -1. More details on the mathematical formalism are provided in the Supporting Information. The sensitivity of our results to some model assumptions (e.g. the thickness of the soil horizon modelled) or analysis setting (e.g. the value prescribed to all grid-cells to set a factor to -1) was evaluated (Supporting Information).

#### *Uncertainty in the representation of the drivers*

At the global scale, each driver is prone to uncertainty and we assessed how this uncertainty propagated to the simulation of  $P_{ILAB}$  and  $P_{TOT}$  and to the contribution of the drivers to their spatial variability. First, we estimated a range of uncertainty for each driver defined by bottom and top boundaries (estimates 1 and 2 in Table S5) following available information. No information about the uncertainty was available for FARM and BUFF, and we arbitrarily assumed that this uncertainty reached  $\pm 30\%$ . Then, the full factorial design described in the previous section was replicated 30 times. For a given replication and a given run within a replication, the value of each driver was chosen randomly within the range between the two estimates by assuming a uniform distribution (i.e. assuming that all values between estimates 1 and 2 were equally likely). This was done independently for each grid-cell.

Each replication of the full factorial design was used to compute the contribution of the different drivers to the spatial variability in  $P_{TOT}$  and  $P_{ILAB}$  (see equation 9). The 30 replications



of the full factorial design were used to compute an average and standard deviation of the contributions. This standard deviation was used as an estimate of the uncertainty in the contribution.

How the uncertainty in the representation of the drivers affected our confidence in the simulated  $P_{TOT}$  and  $P_{ILAB}$  was assessed as follows. First, from the full factorial design we selected the 30 simulations where all drivers=+1. Then, we computed for each grid-cell, a coefficient of variation of these 30 simulations. Finally, a global average of these coefficients of variation (the average was weighted by the cropland/pasture area of each grid-cell) was computed and used as a global indicator of uncertainty in  $P_{TOT}$  and  $P_{ILAB}$ . The contribution of the uncertainty of a given driver (e.g. BIOG) on the uncertainty in simulated  $P_{TOT}$  and  $P_{ILAB}$  was estimated by using the same estimate for that driver in the 30 simulations and by measuring the resulting decrease in the global indicator of uncertainty.

## Results

### *Calibration of the soil P dynamics model*

Varying  $k_w$  and  $k_{occ}$  did not improve the agreement between model and observations for any soil P pools ( $P_{APA}$ ,  $P_{OCC}$  or  $P_{ILAB}$ ) (Figures S3 and S4). This could be explained by the fact that observations made on relatively few years lead to low constraints on pools characterized by long time-scales. Pre-optimized values were kept for these parameters ( $k_w = 1.e^{-4} yr^{-1}$  ;  $k_{occ} = 1.2e^{-5} yr^{-1}$ ). Organic pools and  $P_{ILAB}$  were more sensitive to  $k_{m1}$  and to a larger extent to  $k_{m2}$ . Calibrated  $k_{m1}$  and  $k_{m2}$  were  $1e^{-2} yr^{-1}$  and  $1e^{-1} yr^{-1}$ , respectively. These values correspond to turnover times of 100 and 10 yr for  $P_{OSTA}$  and  $P_{OLAB}$ , respectively. An observation vs. simulation comparison of the change in  $P_{TOT}$  and  $P_{ILAB}$  between final and first years of observations for all treatments (n=49) after calibration is displayed on Figure 2. A comparison detailed for each soil P pool and each treatment is plotted in Figure S5. We found a fairly good

agreement between the simulated and observed temporal evolution of  $P_{TOT}$  and  $P_{ILAB}$ , despite some discrepancies at a few sites. Different reasons could be put forward to explain such mismatches, such as: an inconsistency between observed soil horizon thickness and the processes represented in our model, or missing mechanisms in our model (such as fallow periods or plant processes to overcome limitations) (Table S3). However, we considered the agreement to be reasonable enough to use the model as a tool to assess the drivers of the current global P distribution.

#### *Distribution of $P_{TOT}$ and $P_{ILAB}$ at the global scale and associated uncertainty*

In this section, we focused primarily on cropland (see Figure S7 for pasture). Global averages of  $P_{TOT}$  and  $P_{ILAB}$  for cropland amounted to  $2211 \pm 13$  and  $246 \pm 1$  kgP/ha for the top 0-0.3m soil, respectively (average  $\pm$  std over the simulations performed for the uncertainty analysis, values weighted by the cropland area). Using an averaged global bulk density of  $1.3 \text{g.cm}^{-3}$ , these numbers correspond to concentration of 567 and 63 mgP per kg of soil for  $P_{TOT}$  and  $P_{ILAB}$ , respectively. Highest  $P_{TOT}$  was found in northern America and western India, while lowest  $P_{TOT}$  was found in Argentina (Figure 3a), where simulated soil P content cannot sustain the prescribed uptake (Figure S8). The distribution of  $P_{TOT}$  was heterogeneous over the surface of the world with a close proximity of areas with the highest and lowest  $P_{TOT}$ s. Overall, the areas of highest (or lowest)  $P_{TOT}$  corresponded to areas of highest (or lowest)  $P_{ILAB}$  (Figure 3b), but some discrepancies existed between the distribution of the two variables. This was visible using the percentile distribution of both variables (Figure S9), and the  $P_{ILAB}/P_{TOT}$  ratio (Figure 3c). The global average of the  $P_{ILAB}/P_{TOT}$  ratio amounted to  $9 \pm 5\%$  but reached values close to 0 in northern America and 30% in north India. The coefficient of correlation between both variables computed using all grid-cells was 0.73 for cropland, which confirmed that  $P_{ILAB}$  could not simply be approached by a constant fraction of  $P_{TOT}$ .

As each driver was prone to strong uncertainty, we investigated how this uncertainty propagated to simulated  $P_{TOT}$  and  $P_{ILAB}$ . We found a 47 and 61% global uncertainty for cropland  $P_{TOT}$  and  $P_{ILAB}$ , respectively (Figure 3; right column). The uncertainty in the simulation of  $P_{TOT}$  was high in North America and western Russia (coefficient of variation of the simulations performed with drivers varying within their range of uncertainty ~60%) and low in China and Europe (~20%) (Figure 3d). A similar spatial pattern was found for  $P_{ILAB}$  despite higher uncertainty in North America, Africa and Argentina (Figure 3e). We found that the uncertainty in BIOG played a key role in the uncertainty in simulated  $P_{TOT}$  and  $P_{ILAB}$ , explaining more than half of the uncertainty in  $P_{ILAB}$  and almost all the uncertainty in  $P_{TOT}$  (Table S6). We found that the uncertainty in FARM had a small effect on the uncertainty in  $P_{TOT}$  and  $P_{ILAB}$  but this could be related to a potential underestimation of the uncertainty of FARM. Nevertheless, at the current stage, our results stress the need for more accurate estimates of P in natural soils at the global scale. The increasing resolution of soil datasets (e.g. GlobalSoilMap (Mulder *et al.*, 2016), SoilGrid1km (Hengl *et al.*, 2014)) could help to increase our confidence in simulated soil P maps. Despite the large uncertainty in the distribution of  $P_{TOT}$  and  $P_{ILAB}$ , we were able to robustly decipher the relative contribution of the drivers of global soil P distribution (see following and small error-bars in Figure 4).

#### *Contributions of 'natural' and 'anthropogenic' drivers*

In this section, we focused on cropland and only the last paragraph is related to pasture. We found that the simulated spatial variability of  $P_{ILAB}$  was larger than that of  $P_{TOT}$ : for instance, the coefficient of variation of all cropland grid-cells computed when all drivers varied spatially was 1.67 for  $P_{ILAB}$  vs 0.81 for  $P_{TOT}$ . Thanks to the full factorial design performed, we were able to estimate the relative contributions of the different drivers to the spatial variability computed at either grid-cell or continental scale. Positive contributions were interpreted as

leading to an increase in the spatial variability of soil P (Supporting Information). Each driver made a first-order contribution to the spatial variability of soil P (left bar of each panel in Figure 4). At the grid-cell scale (first line of Figure 4), we found that BIOG alone explained almost all of the variability of  $P_{TOT}$  (with a first order contribution of 103%) while FARM made a preponderant contribution as compared to BIOG for  $P_{ILAB}$  (first-order contribution of 63 and 38% for FARM and BIOG, respectively). Thus, BIOG explained the differences between  $P_{ILAB}$  and  $P_{TOT}$  maps (Figure 3). Other drivers made only minor first-order contributions: LOSS for  $P_{TOT}$  (-8%) and LOSS (-7%), BUFF (4%) and CLIM (3%) for  $P_{ILAB}$ . All other drivers (including LUCC) made no significant first order contribution at the grid-cell scale. The behaviour of  $P_{TOT}$  was explained by  $P_{APA}$ ,  $P_{OCC}$  and  $P_{OSTA}$  : the sum of these pools corresponded to 74% of  $P_{TOT}$  and their spatial variability was totally driven by BIOG (not shown).

At continental scales (second line of Figure 4), we found that BIOG and FARM were the main drivers for both  $P_{TOT}$  and  $P_{ILAB}$ . We found that FARM made a larger contribution to the spatial variability of  $P_{TOT}$  than at the grid-cell scale (first-order contribution of 33%). Many drivers were involved at the continental scale for  $P_{ILAB}$ . In particular, BUFF, CLIM and LUCC had a significant role. When shifting from grid-cells to continents, the larger contribution of FARM could be explained by a lower decrease in the variability of FARM than that of BIOG (Table S7). We also found that the larger contribution of FARM to  $P_{TOT}$  variability when shift to continental scale was mainly explained by soil organic P pools ( $P_{OLAB}+P_{OSTA}$ ) (not shown), underlying the role of farming practices affecting these pools (application of manure, residues).

Drivers could contribute to the spatial variability of soil P, not only through first-order effects but, also as a result of interaction effects (middle bar of each panel in Figure 4). Overall, the interactions between two drivers made only small contributions to the spatial variability of soil P at grid-cell scale. The largest interaction was found between BIOG and LOSS, contrib-

uting for -9% and -5% to the variability of  $P_{TOT}$  and  $P_{ILAB}$ , respectively. Such negative contributions could be interpreted as LOSS tending to decrease the soil P where BIOG tends to increase it, and *vice-versa*. At the continental scale, a positive BIOGxFARM interaction contributing for 11% and 6% to the variability of  $P_{TOT}$  and  $P_{ILAB}$ , respectively, was found.

In addition to the variability computed by using all cropland grid-cells around the World, we computed the driver contributions (at the grid-cell scale) in different latitude bands (Figure 5). We found that the spatial variability of  $P_{TOT}$  did not change as function of the latitude band contrary to  $P_{ILAB}$  whose the variability within the tropical band was lower than the one within the  $>30^{\circ}N$  band (not shown). We also found higher CLIM, BUFF and DEPO contributions and a lower FARM contribution in the tropics than in  $>30^{\circ}N$  or at the global scale. Later access to chemical fertilizer in South America than in Europe or North America (Figure 6) and specificities of tropical soils (e.g. richness in iron oxides) could explain the lower FARM and larger BUFF contributions found, respectively.

Finally, we found that  $P_{TOT}$  and  $P_{ILAB}$  had increased in cropland during the 20<sup>th</sup> century (+14 and +61 %, respectively) due to past variations of some drivers and to soil P dynamics. The increase in  $P_{TOT}$  was mainly explained by LUCC in 1900-1950, but interestingly, was then driven by FARM, while FARM explained the increase in  $P_{ILAB}$  almost totally, confirming the key role of fertilization practices in driving soil P availability at the global scale (Supporting Information).

For pasture (Figure 4e-h), we found that BIOG was the primary driver of soil P content. At the grid-cell scale, the first-order contribution of BIOG reached 104% for  $P_{TOT}$  and 101% for  $P_{ILAB}$ ; the latter variable being also moderately driven by LOSS and BUFF. Interestingly, we found that the contribution of the FARM driver to the spatial variability of soil P was much lower in pasture than in cropland which suggests a much lower anthropogenic perturbation of

these ecosystems from a biogeochemical point of view. However, FARM made a large contribution to  $P_{TOT}$  if focused on the continental scale.

Our results showed little sensitivity to the model assumptions and the analysis settings that were tested, except for the thickness of the soil horizon considered (Figure S11). Reducing the thickness of the soil horizon that was modelled (from 0-0.3m to 0-0.2m) led to significantly increases in the contribution of FARM (e.g. from ~70 to 90% for the contribution to the variability of  $P_{ILAB}$  at the grid-cell scale) to the detriment of BIOG.

## **Discussion**

### *Main drivers and interactions*

Previous studies showed that the possible different drivers of soil P content have strong spatial heterogeneity at the global scale. However, all these studies focused only on a single driver, thus providing a partial picture of the soil P content and preventing any analysis of the relative contribution of the different potential drivers. Our study is the first to show how the different drivers contribute interactively to the soil P content at the global scale. The fact that BIOG and FARM appeared as the main drivers of soil P content in croplands is a critical result for further modelling efforts, and increases our understanding of the distribution of P in agricultural soils at the global scale. Even if the importance of BIOG and FARM was intuitive in a given local context, our study has demonstrated that this also applies at the global scale.

BIOG plays a key role for both  $P_{TOT}$  and  $P_{ILAB}$  whatever the agroecosystems (cropland or pasture) and the spatial scale of the variability (grid-cell or continent) considered. The role of FARM is different, making a major contribution to the variability of  $P_{ILAB}$  but having a minor role in the variability of  $P_{TOT}$ . The contribution of FARM also varies between cropland vs

pasture, as a function of the latitude and as a function of the spatial scale of the variability.

Total soil P is partly driven by pools with long residence times (stable organic, apatite and occluded) which show small sensitivity to human perturbation. This explains the smaller sensitivity of  $P_{TOT}$  to FARM when considering all grid-cells. However, the heterogeneity between farming practices affecting soil organic pools (i.e. manure application, residues) found between continents is large enough to make FARM a major driver of the  $P_{TOT}$  variability at this scale. The greater contribution of FARM to labile P in croplands than in pastures confirmed that differences in fertilization strategies on croplands vs. pastures had strong consequences on soil P spatial distribution (Sattari *et al.*, 2016).

Overall, our results underline the complex spatial patterns of the drivers and ultimately of the soil P. The lack of negative interaction between BIOG and FARM suggested that fertilization practices did not tend to reduce the spatial variability in soil P content due to the biogeochemical background. At the continental scale, we even found a positive interaction. The easier access to the P resources in rich countries, independently from the natural soil P content, could be invoked to explain this decoupling between BIOG and FARM at the global scale.

Finally, the fact that FARM and LUCC were identified as key drivers of the temporal evolution of P in croplands demonstrated the long-term legacy effect of human activities on the current soil P, as previously found at the country scale (Ringeval *et al.*, 2014).

#### *Limitations of the model framework*

Our modelling framework has some limitations, which have the potential to modify our results slightly. The main limitations concerns are i) the lack of representation of some soil P pools or dynamics processes, ii) a relatively arbitrary definition of the drivers represented in our study and iii) the sensitivity of our results to the thickness of the modelled soil horizon.

Accepted Article

First, some pools are not explicitly represented (the pool of orthophosphate ions) or not represented at all (the P in microbial biomass) in our soil P dynamic model. Following the approach commonly used at the global scale (Wang *et al.*, 2010; Goll *et al.*, 2012), we did not represent explicitly a pool of orthophosphate ions ( $\text{H}_2\text{PO}_4^-$  and  $\text{HPO}_4^{2-}$ ) in the soil solution, while it is the main form of P taken up by the plant roots. Instead, our  $\text{P}_{\text{ILAB}}$  pool included both orthophosphate ions and more labile P (Table S1). Parameters involved in equations describing P in soil solution should be optimized against measurements complementary to those made by the Hedley method (e.g.  $^{32}\text{P}$  labelling and isotopic dilution approaches; e.g. (Morel *et al.*, 2014)) because soluble P is not commonly measured by the Hedley method. While this is not straightforward, we think that such representation would be more process-based. It could lead to a higher contribution of BUFF to the global distribution of available P than the one found here. The P in microbial biomass that may function as a buffer by immobilizing orthophosphate ions from soil solution and potentially preventing them from bonding to oxides of iron, aluminum and manganese (Achat *et al.*, 2010) is missing in our approach. Representing microbes in global models started only recently and many uncertainties remain (Wieder *et al.*, 2015). Processes used by plants to overcome limitation (such as root exudates and their potential effects on mycorrhizae, de-occlusion or weathering; (Buendía *et al.*, 2014)) were not represented in our model. A representation of root exudates would require a coupling with the carbon cycle (Buendía *et al.*, 2014) which is challenging (Reed *et al.*, 2015) and beyond the scope of our paper. In our approach, grid-cells where our simulations predicted an important P limitation (i.e. where simulated  $\text{P}_{\text{ILAB}} + \text{P}_{\text{SEC}}$  was smaller than the plant uptake prescribed by FARM dataset) were few (Figure S8) and thus, implementing processes to overcome limitation would probably not lead to a serious modification of the contribution of drivers found in our study.



Second, how the drivers were distinguished in our study is somewhat arbitrary and was constrained by the availability of global datasets. The availability and quality of databases are often limiting for studies performed at the global scale (Makowski *et al.*, 2013). Despite this constraint, we considered that the drivers represented enabled us to take into account all processes known to play a role in shaping the P in agricultural soils over the last centuries and explaining the current global distribution. For instance, we recognize that the soil biogeochemical background corresponding to P inherited from natural soils at the conversion to agriculture (BIOG) results from a huge diversity of processes that are considered in other drivers, such as P losses (LOSS), soil buffering capacity (BUFF) or sensitivity of some soil processes to soil water and temperature (CLIM). However, because temporal scales characterizing BIOG and other drivers are different, it makes sense to distinguish them.

Finally, the contributions of drivers found in our study were sensitive to the thickness of the soil horizon that was modelled. The 0-0.3m soil horizon was chosen because it encompasses a major proportion of roots. A better representation would consist of representing the plough layer in a spatially explicit way. However, no information was currently available at the global scale and our work highlighted the need for such a dataset in the modelling of nutrients in agricultural soils.

#### *A perspective for assessing P limitation on global yields*

Whereas previous studies (Elser & Bennett, 2011) have only described major changes in global biogeochemical cycles arising from agricultural practices, we are able to provide actual estimates of these perturbations by quantifying and mapping soil nutrient status at the global scale. Our modelling framework is a critical step towards the assessment of the P limitation in agricultural ecosystems and for global food production. A combination of our global estimates of soil P status with global crop models (Elliott *et al.*, 2015; Reed *et al.*, 2015)

would permit a more accurate exploration of the consequences of farming practices on the carbon cycle in agroecosystems. Such combinations would require different modelling developments. First, the P uptake by plants should be represented in a mechanistic way (while prescribed from global dataset in our current study). Then, the effects of P limitation on crop development and yield have to be incorporated. Finally, the confidence in simulated maps of  $P_{ILAB}$  has to be increased by reducing the uncertainty of some drivers (in particular in BIOG). While such uncertainty has no effect on the contributions found in our study, this limits the future use of simulated maps of  $P_{TOT}$  and  $P_{ILAB}$ . Once these developments achieved, our approach could be combined with global crop models. Such coupled model could be used under different farming scenarios to investigate, besides consideration of local nuances of farm operations (Sharpley *et al.*, 2016), how policies could be oriented towards more sustainable P resource management and yield gap closure.

### **Acknowledgements**

This research was supported by the “Institut National de la Recherche Agronomique” (INRA) and the Environnement et Agronomie (EA) division. We are grateful to Mark Irvine for help with the computing aspects. We thank Christophe Nguyen, Christian Morel, André Schneider, Pascal Denoroy, Alain Mollier, Benjamin Nowak, Corina Buendia, Axel Kleidon, Adrien Rusch and Noémie Schaller for helpful discussions about the phosphorus cycle, farming practices or statistics. Finally, we would like to thank Aldyth Nys for her assistance with the linguistic aspects of this paper. Modelling of the soil P dynamics model and analysis were performed in using Python (Python Software Foundation. Python Language Reference, version 2.7. Available at <http://www.python.org>). The authors declare no conflict of interest.

## References

- Achat DL, Morel C, Bakker MR, Augusto L, Pellerin S, Gallet-Budynek A, Gonzalez M (2010) Assessing turnover of microbial biomass phosphorus: Combination of an isotopic dilution method with a mass balance model. *Soil Biology and Biochemistry*, **42**, 2231–2240.
- Alcántara V, Don A, Well R, Nieder R (2016) Deep ploughing increases agricultural soil organic matter stocks. *Global Change Biology*, **22**, 2939–2956.
- Augusto L, Bakker MR, Morel C *et al.* (2010) Is “grey literature” a reliable source of data to characterize soils at the scale of a region? A case study in a maritime pine forest in southwestern France. *European Journal of Soil Science*, **61**, 807–822.
- Beck MA, Sanchez PA (1996) Soil phosphorus movement and budget after 13 years of fertilized cultivation in the Amazon basin. *Plant and Soil*, **184**, 23–31.
- Bouwman L, Goldewijk KK, Van Der Hoek KW *et al.* (2011) Exploring global changes in nitrogen and phosphorus cycles in agriculture induced by livestock production over the 1900–2050 period. *Proceedings of the National Academy of Sciences*.
- Brandsæter LO, Bakken AK, Mangerud K, Riley H, Eltun R, Fykse H (2011) Effects of tractor weight, wheel placement and depth of ploughing on the infestation of perennial weeds in organically farmed cereals. *European Journal of Agronomy*, **34**, 239–246.
- Buendía C, Kleidon A, Porporato A (2010) The role of tectonic uplift, climate, and vegetation in the long-term terrestrial phosphorous cycle. *Biogeosciences*, **7**, 2025–2038.
- Buendía C, Arens S, Hickler T, Higgins SI, Porada P, Kleidon A (2014) On the potential vegetation feedbacks that enhance phosphorus availability &ndash; insights from a process-based model linking geological and ecological timescales. *Biogeosciences*, **11**, 3661–3683.
- Cross AF, Schlesinger WH (1995) A literature review and evaluation of the Hedley fractionation: Applications to the biogeochemical cycle of soil phosphorus in natural ecosystems. *Geoderma*, **64**, 197–214.
- Decharme B, Martin E, Faroux S (2013) Reconciling soil thermal and hydrological lower boundary conditions in land surface models. *Journal of Geophysical Research: Atmospheres*, **118**, 7819–7834.
- Elliott J, Müller C, Deryng D *et al.* (2015) The Global Gridded Crop Model Intercomparison: data and modeling protocols for Phase 1 (v1.0). *Geoscientific Model Development*, **8**, 261–277.
- Elser J, Bennett E (2011) A broken biogeochemical cycle. *Nature*, **478**, 29–31.
- Elser JJ, Bracken MES, Cleland EE *et al.* (2007) Global analysis of nitrogen and phosphorus limitation of primary producers in freshwater, marine and terrestrial ecosystems. *Ecology Letters*, **10**, 1135–1142.
- Goll DS, Brovkin V, Parida BR *et al.* (2012) Nutrient limitation reduces land carbon uptake in simulations with a model of combined carbon, nitrogen and phosphorus cycling. *Biogeosciences*, **9**, 3547–3569.
- Gronle A, Lux G, Böhm H *et al.* (2015) Effect of ploughing depth and mechanical soil loading on soil physical properties, weed infestation, yield performance and grain quality in sole and intercrops of pea and oat in organic farming. *Soil and Tillage Research*, **148**, 59–73.
- Hartmann J, Moosdorf N, Lauerwald R, Hinderer M, West AJ (2014) Global chemical weathering and associated P-release — The role of lithology, temperature and soil properties. *Chemical Geology*, **363**, 145–163.
- Hedley MJ, Stewart JWB (1982) Method to measure microbial phosphate in soils. *Soil Biology*

*gy and Biochemistry*, **14**, 377–385.

- Hengl T, de Jesus JM, MacMillan RA *et al.* (2014) SoilGrids1km — Global Soil Information Based on Automated Mapping (ed Bond-Lamberty B). *PLoS ONE*, **9**, e105992.
- Hurtt GC, Chini LP, Frohling S *et al.* (2011) Harmonization of land-use scenarios for the period 1500–2100: 600 years of global gridded annual land-use transitions, wood harvest, and resulting secondary lands. *Climatic Change*, **109**, 117–161.
- Jackson RB, Canadell J, Ehleringer JR, Mooney HA, Sala OE, Schulze ED (1996) A global analysis of root distributions for terrestrial biomes. *Oecologia*, **108**, 389–411.
- Kobilinsky, Monod (1991) Experimental Design Generated by Group Morphisms: An Introduction. *Scandinavian Journal of Statistics*, **18**, 119–134.
- Krinner G, Viovy N, de Noblet-Ducoudré N *et al.* (2005) A dynamic global vegetation model for studies of the coupled atmosphere-biosphere system. *Global Biogeochemical Cycles*, **19**, n/a–n/a.
- Li H, Mollier A, Ziadi N, Shi Y, Parent L-É, Morel C (2017) Soybean root traits after 24 years of different soil tillage and mineral phosphorus fertilization management. *Soil and Tillage Research*, **165**, 258–267.
- Liu Y, Villalba G, Ayres RU, Schroder H (2008) Global Phosphorus Flows and Environmental Impacts from a Consumption Perspective. *Journal of Industrial Ecology*, **12**, 229–247.
- Lynch JP, Brown KM (2001) Topsoil foraging—an architectural adaptation of plants to low phosphorus availability. *Plant and Soil*, **237**, 225–237.
- MacDonald GK, Bennett EM, Potter PA, Ramankutty N (2011) Agronomic phosphorus imbalances across the world’s croplands. *Proceedings of the National Academy of Sciences USA*, **108**, 3086–3091.
- MacDonald GK, Bennett EM, Taranu ZE (2012) The influence of time, soil characteristics, and land-use history on soil phosphorus legacies: a global meta-analysis. *Global Change Biology*, **18**, 1904–1917.
- Mackenzie FT, Ver LM, Lerman A (2002) Century-scale nitrogen and phosphorus controls of the carbon cycle. *Chemical Geology*, **190**, 13–32.
- Makowski D, Nesme T, Papy F, Doré T (2013) Global agronomy, a new field of research. A review. *Agronomy for Sustainable Development*.
- Messiga AJ, Ziadi N, Morel C, Grant C, Tremblay G, Lamarre G, Parent L-E (2012) Long term impact of tillage practices and biennial P and N fertilization on maize and soybean yields and soil P status. *Field Crops Research*, **133**, 10–22.
- Morel C, Ziadi N, Messiga A *et al.* (2014) Modeling of phosphorus dynamics in contrasting agroecosystems using long-term field experiments. *Canadian Journal of Soil Science*, **94**, 377–387.
- Mueller ND, Gerber JS, Johnston M, Ray DK, Ramankutty N, Foley JA (2012) Closing yield gaps through nutrient and water management. *Nature*, **490**, 254–257.
- Mulder VL, Lacoste M, Richer-de-Forges AC, Martin MP, Arrouays D (2016) National versus global modelling the 3D distribution of soil organic carbon in mainland France. *Geoderma*, **263**, 16–34.
- Oberson A, Fardeau JC, Besson J-M, Sticher H (1993) Soil phosphorus dynamics in cropping systems managed according to conventional and biological agricultural methods. *Biology and Fertility of Soils*, **16**, 111–117.
- Oehl F, Frossard E, Fliessbach A, Dubois D, Oberson A (2004) Basal organic phosphorus mineralization in soils under different farming systems. *Soil Biology and Biochemistry*, **36**, 667–675.
- Peñuelas J, Poulter B, Sardans J *et al.* (2013) Human-induced nitrogen–phosphorus imbalances alter natural and managed ecosystems across the globe. *Nature Communica-*

tions, **4**.

- Quinton JN, Govers G, Van Oost K, Bardgett RD (2010) The impact of agricultural soil erosion on biogeochemical cycling. *Nature Geoscience*, **3**, 311–314.
- Reed SC, Yang X, Thornton PE (2015) Incorporating phosphorus cycling into global modeling efforts: a worthwhile, tractable endeavor. *New Phytologist*, n/a–n/a.
- Ringeval B, Nowak B, Nesme T, Delmas M, Pellerin S (2014) Contribution of anthropogenic phosphorus to agricultural soil fertility and food production. *Global Biogeochemical Cycles*, **28**, 743–756.
- Sattari SZ, Bouwman AF, Giller KE, van Ittersum MK (2012) Residual soil phosphorus as the missing piece in the global phosphorus crisis puzzle. *Proceedings of the National Academy of Sciences USA*, **109**, 6348–6353.
- Sattari SZ, Bouwman AF, Martínez Rodríguez R, Beusen AHW, van Ittersum MK (2016) Negative global phosphorus budgets challenge sustainable intensification of grasslands. *Nature Communications*, **7**, 10696.
- Schindler DW, Hecky RE, Findlay DL *et al.* (2008) Eutrophication of lakes cannot be controlled by reducing nitrogen input: results of a 37-year whole-ecosystem experiment. *Proceedings of the National Academy of Sciences*, **105**, 11254–11258.
- Sharpley A, Kleinman P, Jarvie H, Flaten D (2016) Distant Views and Local Realities: The Limits of Global Assessments to Restore the Fragmented Phosphorus Cycle. *Agricultural & Environmental Letters*, **1**.
- Tiessen H, Stewart JWB, Cole CV (1984) Pathways of Phosphorus Transformations in Soils of Differing Pedogenesis. *Soil Science Society of America Journal*, **48**, 853–858.
- Vitousek PM, Porder S, Houlton BZ, Chadwick OA (2010) Terrestrial phosphorus limitation: mechanisms, implications, and nitrogen-phosphorus interactions. *Ecological Applications*, **20**, 5–15.
- Walker TW, Syers JK (1976) The fate of phosphorus during pedogenesis. *Geoderma*, **15**, 1–19.
- Wang YP, Law RM, Pak B (2010) A global model of carbon, nitrogen and phosphorus cycles for the terrestrial biosphere. *Biogeosciences*, **7**, 2261–2282.
- Wang R, Balkanski Y, Boucher O, Ciais P, Peñuelas J, Tao S (2014) Significant contribution of combustion-related emissions to the atmospheric phosphorus budget. *Nature Geoscience*, **8**, 48–54.
- Wieder WR, Allison SD, Davidson EA *et al.* (2015) Explicitly representing soil microbial processes in Earth system models: Soil microbes in earth system models. *Global Biogeochemical Cycles*, **29**, 1782–1800.
- Yang X, Post WM, Thornton PE, Jain A (2013) The distribution of soil phosphorus for global biogeochemical modeling. *Biogeosciences*, **10**, 2525–2537.
- Yang X, Thornton PE, Ricciuto DM, Post WM (2014) The role of phosphorus dynamics in tropical forests – a modeling study using CLM-CNP. *Biogeosciences*, **11**, 1667–1681.

### Supporting Information

The Supporting information file contains 13 additional figures, 7 additional tables and following text sections: “Comparison between simulations and observations on sites”, “Mathematical formalism of the contributions of the drivers”, “Temporal variation”, “Note about the representation of the uncertainty in the different drivers”, “Global datasets”, “Computation of

uptake and residue for each grid-cell based on variables provided in Bouwman *et al.*, 2011”,

“Computation of some parameters involved in the computation of uptake and residue:  $r_{U/H}$

and  $r_R$  for cropland,  $r_{U/W}$  for pasture”.

### Table and Figures captions

Table 1: Drivers: name, description of its effect on soil P and representation in our approach. Variables given in the 3<sup>rd</sup> column correspond to variables whose spatial variability was removed in the full factorial design to compute the driver contribution to the spatial variability of  $P_{TOT}$  and  $P_{ILAB}$ . More information about the different datasets (last column) is provided in Supporting Information.

Figure 1: Modelling framework: the soil P dynamics model and drivers of the spatial distribution of P in agricultural soils. The model allowed us to represent the evolution in time of different soil P pools (inorganic – blue – and organic – orange – pools) as function of simulated soil P dynamics fluxes (green arrows) and soil P input/output (black arrows). The name of the drivers and place where they are involved in the modelling framework are given in magenta (boxes and arrows, respectively). The effects of BIOG and LUCC do not appear on the figure and are described thanks to equations 1, 2 and 8.

Figure 2: Change in soil P between the first and the last dates of observations on sites: comparison between observations (y-axis) and simulations (x-axis). Both changes in  $P_{TOT}$  (triangle) and  $P_{ILAB}$  (circles) are given and each symbol corresponds to one treatment. The right panel displays a zoom of the area of the left panel defined with [-600,2000] boundaries. The colour pallet corresponds to the number of the reference reporting the observations (Table S3). Linear regressions provided in the right panel were computed excluding one treatment of Reference 9. Influence plot of the regression for  $P_{ILAB}$  is given in Figure S6. Simulations were performed with calibrated  $k_{m1}$  and  $k_{m2}$ .

Figure 3: Simulated agricultural soil P for cropland. Simulated  $P_{TOT}$  (a-d),  $P_{ILAB}$  (b-e) and  $P_{ILAB}/P_{TOT}$  (c-f) for cropland: mean (left panels) and coefficient of variations (CV, right panels) computed using the 30 simulations performed to take into account the uncertainty in the global datasets used. An irregular colour pallet corresponding to 0 and the 20,40,60,80,99<sup>th</sup> percentiles was chosen for panels a and b.

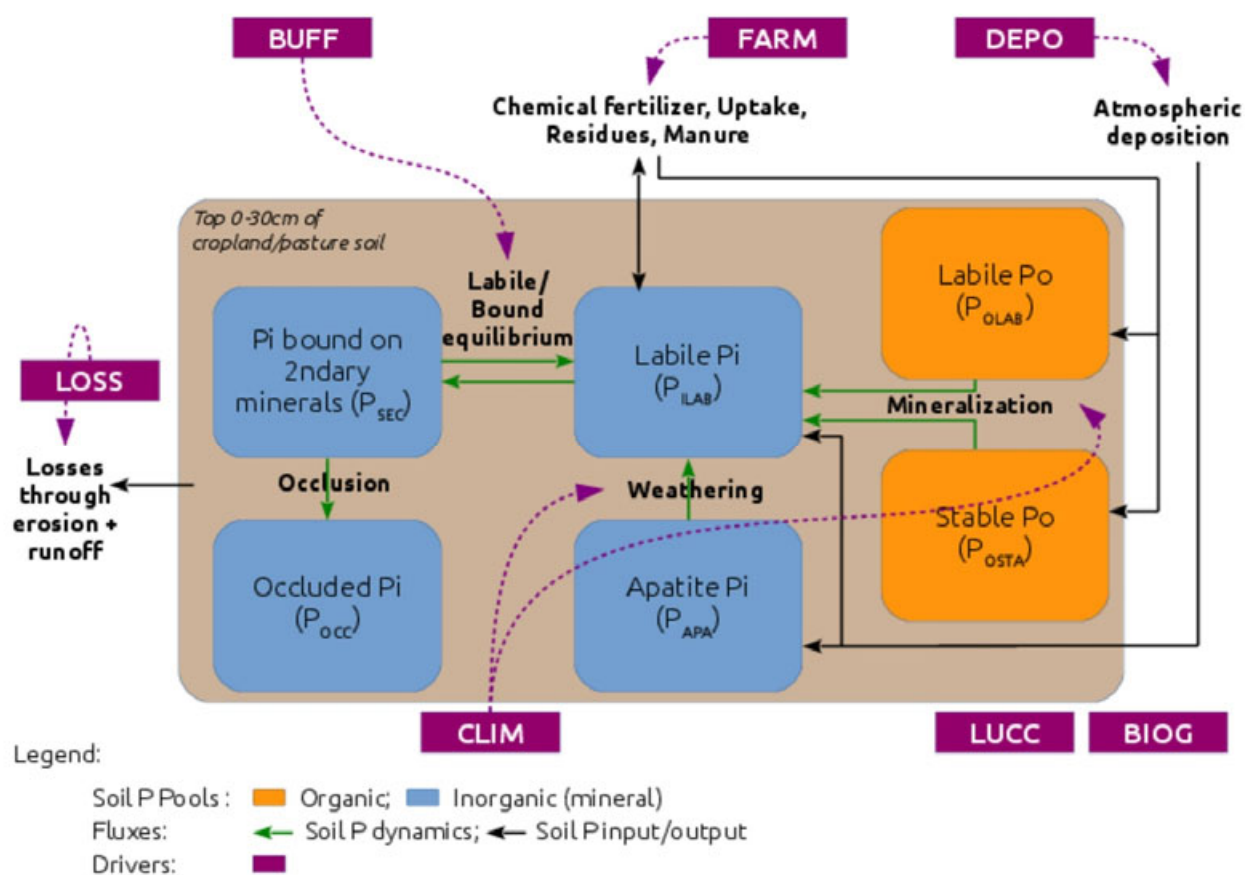
Figure 4: Contributions of the different drivers to the current spatial variability of agricultural soil P. The contributions are given for both  $P_{TOT}$  (left panels) and  $P_{ILAB}$  (right panels) for cropland (top) and pasture (bottom). Variability was computed either at grid-cell (1<sup>st</sup> and 3<sup>rd</sup> lines) or continental scale (2<sup>nd</sup> and 4<sup>th</sup> lines). For each panel, the first-order effect (left bar) and the effect of interaction between two factors (middle bar) were computed thanks to equation (4) for BIOG, LOSS, BUFF, CLIM, FARM and DEPO. The contribution of LUCC (right bar) was estimated from the remaining spatial variability of  $P_{TOT}$  and  $P_{ILAB}$  when the 6 drivers above were set to -1 (Supporting Information). Error-bars represent 1 std of the 30 repetitions of the full factorial design used to take the uncertainty in the representation of the drivers into account.

Figure 5: Driver contributions to the current spatial variability of soil P at the grid-cell scale for different latitude bands. Only first-order contributions are displayed and they were made for cropland  $P_{TOT}$  (panel a) and  $P_{ILAB}$  (panel b). In each panel, the bars correspond (from left to right) to the contributions found for the global scale, the tropical band (30°S-30°N) and the northern band (>30°N), respectively. Note that no uncertainty in the drivers was taken into account here.

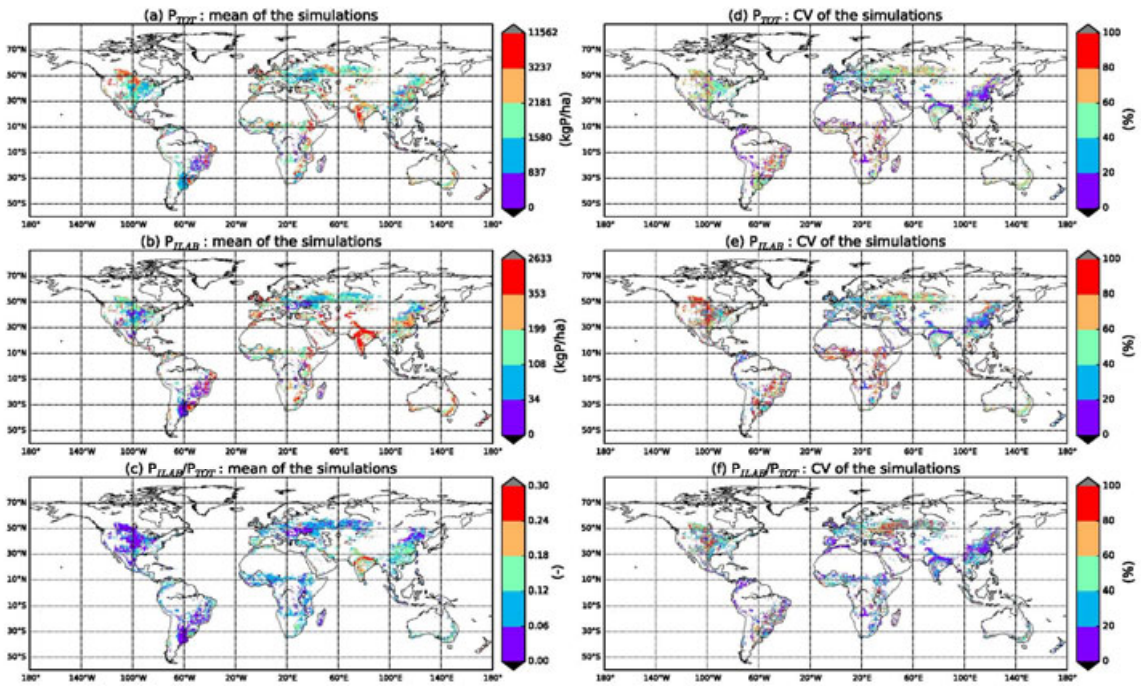
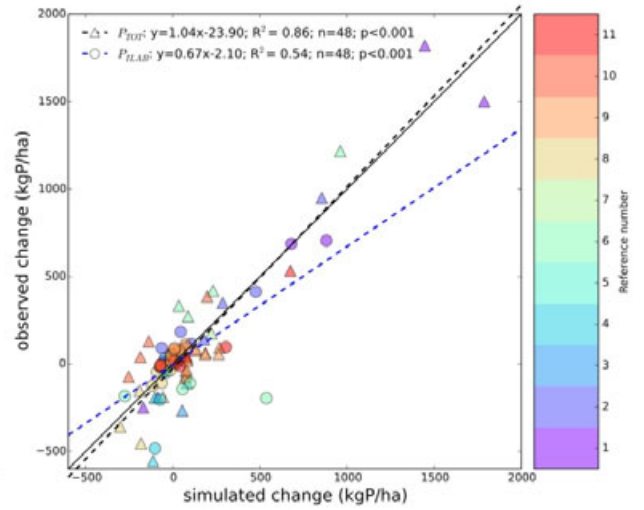
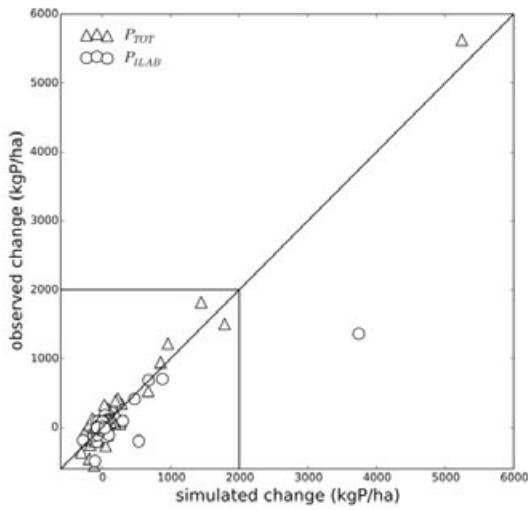
Figure 6: Variation in time of cropland soil P budget and budget components resulting from farming practices (FARM) for 7 large World regions and at the global scale.

Driver name	Description of the effect on soil P	Corresponding variables	Datasets used to estimate the corresponding variables and their variation in space and time
BIOG (natural soil biogeochemical background)	P inherited from natural soils at the time of conversion to agriculture. P in natural soils was also used to approach P in agricultural soils at the beginning of the simulation (in 1700, initial conditions).	$P_{OCC}^{NA}$ , $P_{SEC}^{NA}$ , $P_{OSTA}^{NA}$ , $P_{APA}^{NA}$ , $P_{ILAB}^{NA}$ , $P_{OLAB}^{NA}$ (equations 1 and 8)	(Yang <i>et al.</i> , 2013) estimate of current P in unmanaged soils
FARM (farming practices)	Soil P input/output corresponding to farming practices	P in: chemical fertilizer, manure applied on agricultural soils, residues, and plant uptake. Here, 'residues' refer to plant biomass that remains on/within the soil after the harvest and include root biomass.	Derived from P withdrawal and P fertilizer estimated in (Bouwman <i>et al.</i> , 2011). Basic assumptions were made to compute residues and uptake from withdrawal.
LUCC (land use and land cover change)	Decrease/increase of agricultural soil P resulting from land conversion within the same grid-cell	Cropland and pasture fractions ( $f^C$ , $f^P$ ) and the different land-conversions ( $\Delta$ ) (equation 1)	Land Use Harmonization dataset (Hurtt <i>et al.</i> , 2011)
CLIM (soil temperature and soil water content)	Effect of soil temperature and soil water content on P weathering and P mineralization	Relative soil water content and soil temperature of top 0-0.3m soil horizon (equations 4 and 6)	Simulated by two Dynamic Global Vegetation Models (ISBA (Decharme <i>et al.</i> , 2013) and ORCHIDEE (Krinner <i>et al.</i> , 2005)). The annual average of the climatology computed for the 1979-2010 period (i.e. no year-to-year variability) was used.
LOSS (P losses through soil erosion)	Losses of P due to water erosion and runoff processes	Losses of P for each soil P pool ( $fP_X^{out}$ in Supp. equation 1).	Losses were computed using the soil P content simulated in our approach, the sediment fluxes resulting from erosion provided by (Van Oost <i>et al.</i> , 2007) and the weight of top 0-0.3m soil computed from bulk density of the same horizon provided by Soilgrids (ISRIC – World Soil Information, 2016).

DEPO (atmospheric deposition)	Soil P input resulting from deposition of atmospheric P	P deposition falling within labile P ( $P_{ILAB}$ ) and apatite ( $P_{APA}$ )	Derived from the (Wang <i>et al.</i> , 2014) dataset that provides atmospheric P deposition resulting from mineral dust, primary biogenic aerosol particles, sea salt and combustion, averaged over different time-periods.
BUFF (soil buffering capacity)	Properties of soil describing its ability to replenish the $P_{ILAB}$ pool and to adsorb labile P on its particles ( $P_{SEC}$ )	Parameters involved in the Langmuir equation used to describe the equilibrium between $P_{ILAB}$ and $P_{SEC}$ ( $K_S$ and $S_{max}$ ) (equation 7)	( $K_S, S_{max}$ ) estimated for each soil order by (Wang <i>et al.</i> , 2010). The global distribution of soil orders is similar to the one used in (Yang <i>et al.</i> , 2013).







contribution (%) of variability at the scale of

