Supporting Information of

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

Bruno Ringeval*, Laurent Augusto, Hervé Monod, Dirk van Apeldoorn, Lex Bouwman, Xiaojuan Yang, David L. Achat, Louise P. Chini, Kristof Van Oost, Bertrand Guenet, Rong Wang, Bertrand Decharme, Thomas Nesme, Sylvain Pellerin

Global Change Biology

* Corresponding author: ***bruno.ringeval@inra.fr***

Supporting Text

Comparison between simulations and observations on sites

After calibration, we found that our model was roughly able to match the temporal evolution of total and different forms (e.g. References 1, 2, 3, 5, 7 in Table S3 and Figure S5). On some sites, the model had difficulty in capturing the change in either P_{TOT} (e.g. Reference 3) or the different forms contributing to P_{TOT} (e.g. Reference 6). The difficulty in capturing the observed temporal change in P_{TOT} with a soil budget computation is discussed in the reference paper of some sites (e.g. Reference 11) and could usually be related to a too thin observed soil horizon, that prevent to account for vertical migration or for the whole uptake. Specificities on some sites could also explain the mismatch (see last column of Table S3).

For few references (References 1, 9 and 10), the prescribed uptake was not consistent with the simulated size of P_{ILAB}: prescribing observed uptake to the model quickly led to the emptying of P_{ILAB} and the first pool contributing to its replenishment, i.e. P_{SEC}. In addition to the above mentioned issue about soil horizon thickness, the non-representation in the model of crop rotation and fallow periods (that could contribute to the regeneration of P_{ILAB} during a given year) or the addition of lime could contribute to the mismatch found. Also, we found that the mismatch between the observed uptake and simulated (P_{ILAB}+P_{SEC}) occurred in the context of poor P supply (e.g. unfertilized treatments in Reference 1), suggesting that some mechanisms used by plants to overcome limitation (e.g. biotic enhancement of weathering through root exudation) may be missing in our model. Nevertheless, this pattern was found on only small areas at the global scale in 2005, when the model was forced with BIOG and FARM from global datasets (Figure S8).

Summarising the evaluation, an observation vs. simulation comparison of the change in soil P between final and first years of observations was displayed on Figure 2 in the Main Text. We are aware that our final findings about driver contributions are dependent, not only on our soil P dynamics model, but also on the quality of the global datasets that we used. This is discussed in respective references (Table 1 of the Main Text).

Mathematical formalism of the contributions of the drivers

Here we provide a mathematical formalism to assist in the interpretation of the contribution of each driver (or of each interaction between drivers), computed thanks to equation (9) of the Main Text. Each driver encompasses several variables and drivers are potentially correlated spatially, which makes the method to estimate their contributions complicated. Also, methods used to estimate the contribution of *LUCC* is different from the ones applied to other drivers because of our inability to remove the spatial variability of the *LUCC* driver.

→ Notation

For purpose of simplicity, we considered only two drivers *A* and *B* in this short presentation. There are *N* grid cells and in each grid-cell *i*, a_i and b_i denote the values of the drivers *A* and *B*, respectively. $y_i=f(a_i,b_i)$ is the output of the soil P dynamics model (called *f* here) for the grid-cell *i*. In the Main Text, $f(a_i,b_i)$ corresponds to the soil P content. The variable of interest, noted *Y*, is the spatial variability of y_i among the grid-cells, measured, as in the Main Text, with the variance. Thus, $Y=Var(y_i)$.

→ Full factorial design

We performed a full factorial design in which each driver was either variable (in that case we note that this driver is equal to *+1*) or constant in space (in that case we note that this driver is equal to -1). In the case of a driver set to *-1*, all grid-cells have the same value of the driver (e.g. $\bar{a}$ in case of *A=-1*, where $\bar{a}$ is the median of *A* computed by using all grid cells). In the simplified case here, the full factorial design corresponds to the following:

A=-1 and *B=-1* (thus *A.B=+1*). For each grid-cell i, $y_i=f(\bar{a},\bar{b})$ and $Y_{-1,-1}=Var(y_i)=0$

A=-1 and *B=+1* (thus *A.B=-1*). For each grid-cell i, $y_i=f(\bar{a},b_i)$ and $Y_{-1,+1}=Var(y_i)$ is called VAR_B.

A=+1 and *B=-1* (thus *A.B=-1*). For each grid-cell i, $y_i=f(a_i,\bar{b})$ and $Y_{+1,-1}=Var(y_i)$ is called VAR_A.

A=+1 and *B=+1* (thus *A.B=+1*). For each grid-cell *i*, $y_i=f(a_i,b_i)$ and $Y_{+1,+1}=Var(y_i)$ is called VAR_{A*B}. Note that VAR_{A*B} correspond to the effect when *A* and *B* vary (and not to the interaction between *A* and *B* alone).

The contribution of each driver considered either alone or in interaction, and called *e* (for effect) in the following, is defined here thanks to the equation (9) of the Main Text:

$$e(F)=\frac{1}{n_{runs}/2}.\sum_{j=1}^{n_{runs}}[f_j.VAR_j]$$. In our case, this gives:

$$e(A)=\frac{1}{2}(-Y_{-1,-1}-Y_{-1,+1}+Y_{+1,-1}+Y_{+1,+1})$$

$$e(B)=\frac{1}{2}(-Y_{-1,-1}+Y_{-1,+1}-Y_{+1,-1}+Y_{+1,+1})$$

$$e(A.B)=\frac{1}{2}(+Y_{-1,-1}-Y_{-1,+1}-Y_{+1,-1}+Y_{+1,+1})$$

e(A) and *e(B)* are called the *first-order* contributions of drivers *A* and *B*, respectively while *e(A.B)* is called the interaction of *A* and *B*. *e(A)* and *e(B)* correspond to the left bar of each panel in Figure 4 of the Main Text while *e(A.B)* corresponds to the middle bar.

In our case,

$$e(A)=\frac{1}{2}(VAR_{A*B}+VAR_A-VAR_B)\quad,\quad e(B)=\frac{1}{2}(VAR_{A*B}+VAR_B-VAR_A)$$ and

$$e(A.B)=\frac{1}{2}(VAR_{A*B}-(VAR_A+VAR_B))$$.

We can see that: *e(A)* does not depend on VAR_A alone. The effect of the interaction could be seen as the difference between the effect of *A* and *B* when taken together and the sum of the variance

explained by each driver when considered alone. Such interaction effect could be positive or negative. A negative interaction could be interpreted as a compensation of two sources of variability; e.g. *A* tends to decrease the soil P where *B* tends to increase it, and *vice-versa.*
Two sources of interaction between *A* and *B* on *Y* could be mentioned:
- source 1: the equations in the model *f* introduce an interaction between *A* and *B*
- source 2: the drivers are not independent in space, i.e. a_i and b_i are correlated.

We can decompose *e(A)* as follows:

$$e(A)=\frac{1}{2}((Y_{+1,+1}-Y_{-1,+1})+(Y_{+1,-1}-Y_{-1,-1}))=\frac{1}{2}[e(A|(B=+1))+e(A|(B=-1))]$$

where *e(A | (B=+1))* is an estimate of the effect of *A* when *B* varies and *e(A | (B=-1))* is an estimate of the effect of *A* when *B* is fixed. Such effects could be called “elementary effects”. *e(A)* corresponds to the sum of the these elementary effects divided by $n_{runs}/2$. A first-order effect could be negative if, for instance, *e(A | (B=-1))* is small and *e(A | (B=+1))* is negative, meaning that *A* tends to reduce the variability when *B=+1*.
For information, by definition of the elementary effects, we found that:

$$e(A.B)=\frac{1}{2}[e(A|(B=+1))-e(A|(B=-1))]$$

→ <u>Indirect estimate of the contribution of *LUCC*</u>
We now note $Y_{+1,+1,+1,+1,+1,+1,+1}$ the variance of y_i in the case of the seven drivers considered in the Main Text, which are *BIOG*, *FARM*, *LOSS*, *CLIM*, *DEPO*, *BUFF* and *LUCC* respectively. We were not able to set *LUCC=-1* and thus, our full factorial design focused on the first six drivers. In all simulations performed in our full factorial design, *LUCC=+1* and one of the simulation of the full factorial design allowed us to compute $Y_{-1,-1,-1,-1,-1,-1,+1.}$ Because only drivers introduce spatial variability in simulated soil P in our approach, $Y_{-1,-1,-1,-1,-1,-1,-1}$ could be considered to be equal to 0. The estimates of $Y_{-1,-1,-1,-1,-1,-1,+1}$ and $Y_{-1,-1,-1,-1,-1,-1,-1}$ allow us to compute the effect of *LUCC* when all other drivers are constant spatially, i.e.:

$$e(LUCC|(BIOG=-1,FARM=-1,LOSS=-1,CLIM=-1,DEPO=-1,BUFF=-1))=(Y_{-1,-1,-1,-1,-1,-1,+1}-Y_{-1,-1,-1,-1,-1,-1,-1})$$

While it provides only a part of the first-order contribution of *LUCC*, this is the only effect that could be estimated in our approach. This elementary effect corresponds to the right bar of each panel in Figure 4 of the Main Text.

Temporal variation

In addition to drivers of the current spatial distribution in P_{TOT} and P_{ILAB} analysed in the Main Text, the drivers of the change in time of both variables during the 20th century were also assessed. For this, four simulations were carried out:

- in the "CTRL" run, all drivers for which information was available (Figure S1) vary over time. This is similar to simulations discussed in the Main Text.
- in the "LUCC=1900" run, no change in LUCC was taken into account from 1900 (i.e. for any y > 1900, $f^C(y)=f^C(1900); f^P(y)=f^P(1900); \Delta_{W1}^{W2}(y)=0$ in equation (1) of the Main Text)
- in the "LUCC=1900 & FARM=0" run, all FARM components (i.e. chemical fertilizer, manure, residue and uptake) are set to 0 from 1900 in addition to LUCC=1900

The change in soil P simulated for the 20th century in the 3 runs are plotted in the last two lines of Figure S10 and Figure S12 for cropland and pasture, respectively. In the last 2 runs, only grid-cells with a non null cropland (pasture) fraction in 1900 were considered. To allow a strict comparison between the 3 runs, a curve corresponding to the CTRL case but keeping only these grid-cells was added ("CTRL;1900 g-cells" in black dashed).

For cropland, simulated P_{TOT} increased during the 20th century at the global scale (+14%) and for all regions except Central and South America, which was characterized by a strong decrease during the first 30 years of the 20th century. The temporal evolution of P_{TOT} was driven by LUCC in the first half of the 20th, and then by FARM. The role of LUCC was found in all regions. In all regions except Africa, FARM played a role but with varying start during the 20th century as function of the regions. Prescribing LUCC=1900 tend to decrease P_{TOT} (as compared to CTRL) meaning that the areas converted into cropland have higher soil P than areas that were cropland previously. Note that in Oceania, the opposite occurs in comparison with other World regions and this is totally driven by a few grid-cells with a very positive FARM budget (not shown).

Simulated P_{ILAB} also increased during the 20th century at the global scale and this increase (+61%) was larger than that of P_{TOT}. FARM almost totally explained the temporal evolution of P_{ILAB}, in particular its large increase since 1970. This pattern is found throughout West Europe in particular.

Note that trends in P_{TOT} and P_{ILAB} remain after the suppression of the temporal variation of both LUCC and FARM ("LUCC=1900 & FARM=0" run). The negative trend found for P_{TOT} is totally explained by P losses (sensitivity run not shown), while the positive trend in P_{ILAB} is explained by the replenishment from some soil pools (apatite, organic or secondary). The evolution in time of deposition had no effect (not shown).

At the global scale, changes in P_{TOT} and P_{ILAB} in pasture (Figure S12) are negative (-4% and -2% respectively) and the absolute change is much lower than that of cropland. We found that LUCC totally drives the simulated evolution in P_{ILAB}.

Note about the representation of the uncertainty in the different drivers

As mentioned in the Main Text, the uncertainty in each driver was accounted for as follows: first, we estimated a range of uncertainty for each driver as defined by bottom and top boundaries (estimates 1 and 2 in Table S5), then the full factorial design 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 all values between estimates 1 and 2 were equally likely). This was done independently for each grid-cell.

Additional informations are provided below:

- while the range of uncertainty of some drivers could be approached by a constant value, the independence between grid-cells varied not only the average value of that driver but also its spatial distribution from one replication to the other.
- the independence between the grid-cells made VAR of both P_{TOT} and P_{ILAB} larger when the uncertainty in drivers was taken into account. However, this had no effect on the estimated contributions.
- arbitrary values were used to define bottom and top boundaries for BUFF and FARM because there was no available information. A relatively high uncertainty was considered (±30%). E.g. (Wang *et al.*, 2010) did not provide any information about the soil depth represented in their study, while (S_{max}, K_s) were expressed in gP/m^2.
- for a given grid-cell and a given driver, the value of uncertainty was considered as a constant in time. E.g. for a given grid-cell, if the randomly chosen value for the deposition in 1700 was:

DEPO(1700) = DEPO1(1700) + x*(DEPO2(1700) – DEPO1(1700)), where DEPO1 and DEPO2 are estimates 1 and 2 for the driver DEPO and x is the random value (between 0 and 1).

Then, for all years y, we prescribed:

DEPO(y)= DEPO1(y) + x*(DEPO2(y) – DEPO1(y))

- the uncertainty in FARM for the cropland and pasture fractions within the same grid-cell was considered independently.
- because of the independence between the uncertainty of two grid-cells, the value used to set a driver to -1 (defined as VAL in Figure S11) varied among the replications.
- finally, note that, because of the full factorial design chosen, we could not account for uncertainty in LUCC.

Global datasets

As mentioned in the Main Text, global datasets made it possible to represent the different drivers: the soil biogeochemical background corresponding to P inherited from natural soils at the conversion to agriculture (BIOG), 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); these are summarized in Table 1 of the Main Text. Each dataset has been published in a peer-review journal and readers should refer to the reference paper for more information about the datasets. In the following, we provide some brief general information on each dataset and focus on the treatments performed in our study to use it as input into our model.

Note that the spatial resolution of a few datasets (5 arcmin for erosion, 1° resolution for ISBA model output, 0.9375°x1.28570557° for deposition) was not consistent with our approach and thus, these datasets were re-gridded to half-degree resolution.

FARM

Soil input/output fluxes corresponding to farming practices (FARM; input: residues, chemical fertilizer, manure; output: uptake) are described in (Bouwman *et al.*, 2011). The dataset provides information about withdrawal, manure, chemical fertilizer with a cropland/pasture distinction for the following years: 1900, 1950, 1960, 1970, 1980, 1990, 1995 and 2000 as computed by IMAGE (Bouwman, 2006).

Chemical fertilizer application was used directly in our approach. Following various assumptions, we derived uptake and residue (the latter is defined here as the plant biomass that remains on/within the soil after harvesting and includes root biomass) from the withdrawal variable provided by (Bouwman *et al.*, 2011). This was required because uptake and residue concern different soil P pools (Figure 1 of the Main Text). To do this, we introduced ratios involving withdrawal, uptake and the different components of crop biomass provided in (Smil, 2000) (see the last two sections of the Supporting Information). Note also that a linear relationship was used to extrapolate the different variables between years for which datasets were available (Figure S1). The variation in time of uptake, residue, chemical fertilizer and manure was plotted for both cropland (Figure 6 of the Main Text) and pasture (Figure S13). For cropland, the global soil budget resulting from FARM was slightly negative and become positive from 1950. It reached a peak in 1980, with ~5kgp/ha/yr in 2000. For pasture, the soil budget resulting from FARM was still positive, related to the computation of withdrawal from pasture in (Bouwman *et al.*, 2011).

In addition, we estimated the distribution of residues and manure between P_{ILAB}, P_{OLAB} and P_{OSTA}. For simplicity, the parameters used were considered constant for all grid-cells. Following measurements reported from Hedley's fractionation method performed on manure (Table 1 of (Ylivainio & Turtola, 2013)), we considered that manure applied to the soil goes into P_{ILAB}, P_{OLAB}, and P_{OSTA} in the following proportions: 80, 10, and 10%, respectively. It has been shown that this proportion varies as a function of the livestock category (cows, pigs, etc.), manure treatment (compost, litter, etc.) and spatial variability in the proportion could not be accurately represented at this stage.

As opposed to manure, no direct Hedley measurements have been made on crop residues. Hedley measurements on soil, made immediately after the incorporation of crop residues, were also difficult to interpret because this incorporation promoted mobilization/immobilization of the P already present (Alamgir *et al.*, 2012). Study with RMN suggest that around 40 % of P in crop residues is orthophosphate, 40 % is labile organic (assuming labile is represented by acid nucleic and phytate) and 20 % is stable organic (Noack *et al.*, 2012). Such proportions were used in our

approach to represent the distribution of residues between P_{ILAB}, P_{OLAB} and P_{OSTA}. Note that the fraction of labile P in crop residue is much larger than that of labile N and this is related to the form of P reserves in the plant (Damon *et al.*, 2014).

LUCC

The Land-Use Harmonization data (Hurtt *et al.*, 2011) (called LUHa hereafter) provides global, gridded, fractional (at half-degree spatial resolution) land-use states and land-use transitions annually for the years 1500-2100 (*http://luh.umd.edu*). The model used in (Hurtt *et al.*, 2011) is constrained with data inputs including the HYDE historical cropland, pasture and urban data (Goldewijk, 2001), historical national wood harvest reconstructions, potential biomass and recovery rates, and future projections of land-use from Integrated Assessment Models. Because these inputs do not fully constrain the problem, additional assumptions were made, including the priority of primary or secondary land for wood harvesting and agricultural conversion, the inclusiveness in wood harvest statistics of wood-cut in the conversion of forest to agricultural land use, the spatial pattern of wood harvest, and the residence time of land in agricultural use.

While LUHa is based upon the HYDE historical land-use dataset, which is used itself within IMAGE, some differences could be observed in the agricultural grid-cell fractions between LUHa and (Bouwman *et al.*, 2011). This could be attributed to differences in the HYDE version used and the original resolution. In particular, many grid-cells have a very low cropland fraction in LUHa and no cropland fraction at all in IMAGE. In our study, these discrepancies could lead to a false estimate of the contribution of FARM and LUCC to the spatial distribution of soil P. We solved this issue by making LUHa consistent to IMAGE. To do this, for the grid-cells mentioned above, we: i) set cropland and pasture fractions to 0 for years of mismatch and ii) prescribed transition with natural vegetation in the first/last year of mismatch in order to keep transition variables consistent.

BUFF

The spatial variability of (S_{max}, K_s) corresponds to the soil buffering capacity driver. Following (Wang *et al.*, 2010), (S_{max}, K_s) varies with the soil order (Table S2). (Wang *et al.*, 2010) calibrated (S_{max}, K_s), so that the resulting P distribution in unmanaged soil matched the dataset described in (Cross & Schlesinger, 1995). The global distribution of soil orders is obtained from the United States Department of Agriculture (USDA) website (http://soils.usda.gov/use/worldsoils/mapindex/order.html). As in (Yang *et al.*, 2013), the USDA map for Latin America has been replaced by the soil order map based on the Soil and Terrain database for Latin America and the Caribbean (SOTERLAC, http://www.isric.org/).

DEPO

(Wang *et al.*, 2014) provides atmospheric P deposition resulting from mineral dust, primary biogenic aerosol particles, seasalt and combustion. Atmospheric P deposition from combustion is representative of the 1960-2007 period while other variables are representative of the 2000s. Combustion includes both natural fires and anthropogenic combustion.

For the purpose of our study, we separated atmospheric P deposition from combustion, into: i) deposition from anthropogenic combustion and ii) deposition from natural fires, as follows:

$$\overline{D_{comb}} = \overline{D_{comb}^{anth}} + \overline{D_{comb}^{nat}} \quad \text{and,} \quad \overline{D_{comb}^{anth}} = \overline{D_{comb}} \cdot \frac{\overline{E_{comb}^{anth}}}{\overline{E_{comb}^{anth}} + \overline{E_{comb}^{nat}}}$$

where D and E correspond to P deposition and P emission, respectively; both averaged over the 1960-2007 period. Emissions are provided by (Wang *et al.*, 2014) and the $\frac{\overline{E_{comb}^{anth}}}{\overline{E_{comb}^{anth}} + \overline{E_{comb}^{nat}}}$ ratio is computed for 12 large regions (~33% for Total Eastern and Southern Africa, 15% for Total Northern

Africa, 24% for Total Western and Central Africa, 92% for Total East Asia, 97% for Total South and South-east Asia, 69% for Total Western and Central Asia, 94% for Total Europe, 100% for Total Caribbean, 82% for Total Central America; 91% for Total North America, 9% for Total Oceania and 54% for Total South America).

We assumed that deposition from mineral dust, primary biogenic aerosol particles, seasalt and natural fires are constant in time and thus, we used the temporal average for each year of our simulations (Figure S1).

We added year-to-year variability to P deposition from anthropogenic combustion by using year-to-year variability in emissions provided by (Wang *et al.*, 2014):

$D_{comb}^{anth}(y)=\frac{E_{comb}^{anth}(y)}{\overline{E_{comb}^{anth}}}.\overline{D_{comb}^{anth}}$ where y is within 1960-2007. The $\frac{E_{comb}^{anth}(y)}{\overline{E_{comb}^{anth}}}$ ratio is computed for each World regions mentioned above. For years before 1960, we assumed that:
$D_{comb}^{anth}(y)=D_{comb}^{anth}(1960)$ (Figure S1).

10% of P deposition from dust and 50% of deposition from other sources falls within P_{ILAB} while rest falls within P_{APA} (Mahowald *et al.*, 2008).

BIOG

The P in unmanaged soils provided by (Yang *et al.*, 2013) was used in our study. This dataset provides values of P for the top 0.5m. We used the ratio of bulk density provided by Soilgrids50km (ISRIC – World Soil Information, 2016) for top 0.3m and top 0.5m to derive P in unmanaged soils for top 0.3m. Note that no information about soil bulk density (but about parent material bulk density) was used in (Yang *et al.*, 2013) to generate the data. The grid-cells whose unmanaged soil P pools values are not provided in (Yang *et al.*, 2013) (e.g. grid-cells characterized by the gelisol soil order) have been excluded from our computation.

The dataset about P in unmanaged soils was built using three main steps as described in detail in (Yang *et al.*, 2013) :
- a parent material map was combined with rock P concentration to generate a map of parent material concentration
- a map of total P content in the top 0-0.5m was then derived by combining the map of parent material concentration generated in the previous step with an index that quantitatively describes the cumulative total P loss during soil development. The map of that index (called PPDI for P pedogenic depletion index) was built by using a soil order map (used to classify soils at different weathering stages) and by prescribing a PPDI to each soil order based on 8 chronosequences.
- lastly, (Yang *et al.*, 2013) derived maps of the different forms of P in soils by applying the relationship between soil order and the fractions of total P held in different P forms based on a literature review of Hedley's measurements (Yang & Post, 2011).

LOSS

Two main processes lead to losses of P from agricultural soils to the surrounding environment and water bodies: i) erosion+runoff and ii) leaching. We neglected leaching and gave priority to erosion+runoff processes (Senthilkumar *et al.*, 2012). Thus, only erosion+runoff was represented in the model. Losses through erosion+runoff from pool X were defined as:

$$fP_X^{out} = \frac{P_X}{P_{TOT}} \cdot \frac{P_{TOT}}{Soil} \cdot f_{sediment}$$ (Supp. equation 1)

where *X* belongs to all soil P forms (i.e. *X* in {APA, SEC, OCC, ILAB, OSTA, OLAB}), $f_{sediment}$ is the flux of eroded sediment (in kg of soil.ha^{-1}.yr^{-1}) and *Soil* is the weight of top 0-0.3m soil (in kg of soil.ha^{-1}) derived from the bulk density provided by Soilgrids50km for the same soil horizon (ISRIC – World Soil Information, 2016). Given the unavailability of datasets focusing on agricultural soils, we assumed that that bulk density could be also applied to cropland and pasture soils, even though it is known that soil treatment has an effect on soil physical properties (Bronick & Lal, 2005).

$f_{sediment}$ was provided by (Van Oost *et al.*, 2007). (Van Oost *et al.*, 2007) computed global estimates of sediment mobilized by water erosion, with a cropland/pasture distinction as a function of slope, climate and soil erodibility. This flux corresponds to a gross erosion rate and not to the net flux, which results from both mobilization and processes, such as deposition, storage and burial. However, we considered that these later processes were not relevant to the system represented here, either because they happened in non-agricultural fractions of the grid-cell (e.g. deposition on river banks) or because they concerned soil horizons below 0.3 m (burial).
In addition, we assumed that the estimates from (Van Oost *et al.*, 2007) (in kg of soil.ha^{-1}.yr^{-1}), corresponding to years ~2000, are representative of erosion fluxes for the whole century. This did not allow us to represent change in erosion due to change in farming practices. However, given that (i) land use is the main driver for global agricultural erosion (Van Oost *et al.*, 2007) and (ii) that our approach accounts for historical changes in agricultural land area (Figure S1), we argue that we have captured the essence of this process.

CLIM
The mean annual relative liquid soil water content and soil temperature, both averaged for the top 0.3m were prescribed from simulations performed with two Dynamic Global Vegetation Models (ISBA and ORCHIDEE). The mean annual values used were kept maintained constant in time over the whole period in our approach (Figure S1) and were representative of the 1979-2010 period during which the ISBA and ORCHIDEE simulations were performed. The relative soil water content was computed with values from 0 (wilting point) to 1 (saturation). A brief description of each model and the name of climate datasets used to force them are given below.
The ISBA model uses a multi-layer approach for the snowpack and the soil energy and mass budgets, as well as a comprehensive set of sub-grid parameterizations for hydrology (Decharme *et al.*, 2013, 2015). It explicitly solves soil freezing/melting at each soil node of the soil grid. Here, the model was forced by the 3-hourly global meteorological forcing from Princeton University (http://hydrology.princeton.edu) at a 1 degree resolution (Sheffield *et al.*, 2006).
The ORCHIDEE model is a spatially explicit process-based model calculating the fluxes of CO_2, H_2O, and heat exchanged between the land surface and the atmosphere on a half-hourly basis, and the variations of water and carbon pools on a daily basis (Krinner *et al.*, 2005). Soil hydrology is computed following a physical description of water diffusion and retention in unsaturated soils, stemming from the Richards equation (de Rosnay, 2002). Freeze processes are not represented in the ORCHIDEE version used. A global simulation was performed at the resolution of 0.5x0.5° with transient land use (Hurtt *et al.*, 2006) using the climate data from the CRU-NCEP (N. Viovy et al., personal communication, 2009, ftp://nacp.ornl.gov/synthesis/2009/frescati/temp/land_use_change/original/readme.htm).

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

$U = H + R_{tot}$ (1)

$R_{tot} = R_{not-rm} + R_{rm}$ (2)

$W = H + R_{rm}$ (3)

where U: uptake; H: harvest; R_{tot}: total residue; R_{rm}: residue removed from field; R_{not-rm}: biomass remaining on/within the soil after harvesting; W: withdrawal. All variables are P fluxes. U and R_{not-rm} are the variables required as input to our model. Note that R_{not-rm} is called 'residue' in the 'Global datasets' section (and includes root biomass if the harvest/withdrawal is aboveground).

We defined the following ratios:

$r_{U/W} = U/W$

$r_{U/H} = U/H$

$r_R = R_{not-rm}/R_{tot}$

The values of such parameters, considered as constant in space, were computed thanks to (Smil, 2000) (see next section).

In the case of knowledge about W:

By definition of $r_{U/W}$, $U = r_{U/W}.W$.

By subtracting (3) to (1) and by definition of R_{not-rm}, we get $R_{not-rm} = U - W$ then

$R_{not-rm} = (r_{U/W} - 1).W$

(Bouwman *et al.*, 2011) provides withdrawal for pasture and the above equations are thus appropriate to compute U and R_{not-rm} for pasture. In that case, $r_{U/W}$ is required. Note that, in (Bouwman *et al.*, 2011), W for pasture has been estimated as 87.5 % of total applied fertilizer.

In the case of knowledge about H:

By definition of $r_{U/H}$, $U = r_{U/H}.H$.

By injecting the definition of r_R in (1) and then rearranging, we get: $R_{not-rm} = (U - H).r_R$. By injecting the definition of $r_{U/H}$, we obtained: $R_{not-rm} = (r_{U/H} - 1).r_R.H$

(Bouwman *et al.*, 2011) provide harvest for cropland and the above equations are thus appropriate to compute U and R_{not-rm} for cropland. Both $r_{U/H}$ and r_R are required.

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

These 3 parameters do not vary spatially. The computation of $r_{u/H}$ and r_R for cropland requires global cropland estimates of U, H, R_{not-rm} and R_{tot} (here, '*tot*' means the sum of residue removed and residue remaining on the field). These estimates were based on values found in (Smil, 2000). For some crop categories, (Smil, 2000) provided the following variables at the global scale : shoot P uptake (U^{shoot}), harvest (H), total shoot residue (R_{tot}^{shoot}) (Table 5 of Ref (Smil, 2000)). Values of U^{shoot}, H and R_{tot}^{shoot} corresponding to global cropland are estimated by computing a sum of the different crop categories.

Global cropland U is then computed by using U^{shoot} and a ratio of P content in roots to P content in shoots ($r_{Root/Shoot}$): $U = U^{shoot}.(1 + r_{Root/Shoot})$. We used a $r_{Root/Shoot}$ equal to 0.25. That value is sensitive to crop species (Manlay *et al.*, 2002) and the level of P in soil, but for simplicity, we chose a value for all plants which was close to that provided in (Manlay *et al.*, 2002) for rice. Finally, we

found a global cropland $r_{U/H}$ of 2.

By definition, $r_R = R_{not-rm} / R_{tot}$.
By decomposing this into roots and shoots and by assuming that all root biomass remains within the soil as residue (i.e. $R_{not-rm}^{root} = R_{tot}^{root}$ and $R_{tot}^{root} = U^{root}$), we get:
$r_R = \frac{R_{not-rm}^{shoot} + U^{root}}{R_{tot}^{shoot} + U^{root}}$. (Smil, 2000) approached aboveground crop residue removed from the field by half of the total aboveground residue, thus:
$r_R = \frac{0.5 * R_{tot}^{shoot} + U^{root}}{R_{tot}^{shoot} + U^{root}}$. Finally, by using $r_{Root/Shoot}$ as described above, we find:
$r_R = \frac{0.5 * R_{tot}^{shoot} + r_{Root/Shoot} . U^{shoot}}{R_{tot}^{shoot} + r_{Root/Shoot} . U^{shoot}}$ and a global cropland r_R of 0.7.

For pasture, we assumed that withdrawal corresponds to the whole shoot biomass, thus $r_{U/W} = U / U^{shoot}$ and we get $r_{U/W} = (1 + r_{Root/Shoot})$. Thus, we found a global grassland $r_{u/w}$ of 1.25.

References

Alamgir M, McNeill A, Tang C, Marschner P (2012) Changes in soil P pools during legume residue decomposition. *Soil Biology and Biochemistry*, **49**, 70–77.

Bouwman A., Kram (2006) *Integrated modelling of global environmental change: an overview of IMAGE 2.4*. Netherlands Environmental Assessment Agency, Bilthoven.

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*.

Bronick CJ, Lal R (2005) Soil structure and management: a review. *Geoderma*, **124**, 3–22.

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.

Damon PM, Bowden B, Rose T, Rengel Z (2014) Crop residue contributions to phosphorus pools in agricultural soils: A review. *Soil Biology and Biochemistry*, **74**, 127–137.

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.

Decharme B, Brun E, Boone A, Delire C, Le Moigne P, Morin S (2015) Impacts of snow and organic soils parameterization on North-Eurasian soil temperature profiles simulated by the ISBA land surface model. *The Cryosphere Discussions*, **9**, 6733–6790.

Goldewijk KK (2001) Estimating global land use change over the past 300 years: the HYDE database. *Global Biogeochemical Cycles*, **15**, 417–433.

Hurtt GC, Frolking S, Fearon MG *et al.* (2006) The underpinnings of land-use history: three centuries of global gridded land-use transitions, wood-harvest activity, and resulting secondary lands. *Global Change Biology*, **12**, 1208–1229.

Hurtt GC, Chini LP, Frolking 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.

ISRIC – World Soil Information (2016) SoilGrids: an automated system for global soil mapping. Available for download at https://soilgrids.org; Aggregated data (~50km) was downloaded from ftp://ftp.soilgrids.org/data/aggregated/.

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.

Mahowald N, Jickells TD, Baker AR *et al.* (2008) Global distribution of atmospheric phosphorus sources, concentrations and deposition rates, and anthropogenic impacts. *Global Biogeochemical Cycles*, **22**, n/a–n/a.

Manlay RJ, Chotte J-L, Masse D, Laurent J-Y, Feller C (2002) Carbon, nitrogen and phosphorus allocation in agro-ecosystems of a West African savanna: III. Plant and soil components under continuous cultivation. *Agriculture, ecosystems & environment*, **88**, 249–269.

Noack SR, McLaughlin MJ, Smernik RJ, McBeath TM, Armstrong RD (2012) Crop residue phosphorus: speciation and potential bio-availability. *Plant and Soil*, **359**, 375–385.

Van Oost K, Quine TA, Govers G *et al.* (2007) The Impact of Agricultural Soil Erosion on the Global Carbon Cycle. *Science*, **318**, 626–629.

De Rosnay P (2002) Impact of a physically based soil water flow and soil-plant interaction representation for modeling large-scale land surface processes. *Journal of Geophysical Research*, **107**.

Senthilkumar K, Nesme T, Mollier A, Pellerin S (2012) Conceptual design and quantification of

phosphorus flows and balances at the country scale: The case of France. *Global Biogeochemical Cycles*, **26**, GB2008.
Sheffield J, Goteti G, Wood EF (2006) Development of a 50-year high-resolution global dataset of meteorological forcings for land surface modeling. *Journal of Climate*, **19**, 3088–3111.
Smil V (2000) Phosphorus in the environment: natural flows and human interferences. *Annual review of energy and the environment*, **25**, 53–88.
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.
Yang X, Post WM (2011) Phosphorus transformations as a function of pedogenesis: A synthesis of soil phosphorus data using Hedley fractionation method. *Biogeosciences*, **8**, 2907–2916.
Yang X, Post WM, Thornton PE, Jain A (2013) The distribution of soil phosphorus for global biogeochemical modeling. *Biogeosciences*, **10**, 2525–2537.
Ylivainio K, Turtola E (2013) Solubility and plant-availability of P in manure. *Baltic Manure WP4 Standardisation of Manure Types with Focus on Phosphorus*.

Caption of Supporting Tables

Table S1: Table of correspondence between pools represented in our model (Figure 1 in the Main Text) and Hedley fractions measured on sites used for model calibration. Fractions within brackets are considered only on sites where information about their content is provided. Bicarbonate = $NaHCO_3$; Hydroxide = NaOH.

Table S2: Value of ($1/K_s$) and S_{max} per soil order class (taken from Table 2 of Wang et al. 2010). Values provided by Wang et al. (2010) (in gP/m2) were converted into kgP/ha by assuming that they are representative of the soil horizon (0-0.3m) studied here.

Table S3: "Simulations on sites": description of observations on sites (references, latitude/longitude, soil horizon, soil order, treatments) and information used to prescribe BIOG and FARM in the model. FARM is defined thanks to following variables: uptake, residues, chemical fertilizer and manure applied on field. In the table, we focused in particular on the computations of uptake and residues from available observations (usually withdrawal or harvest). Relationships between uptake (U), residues (R_{not-rm}), withdrawal or harvest ($WorH$) were written in a general form for both cropland and pasture as:

$$U = r_{U/WorH} . WorH \qquad R_{not-rm} = (r_{U/WorH} - 1) . r_R . WorH$$

where $r_{U/WorH}$ correspond either to $r_{U/H}$ or $r_{U/W}$. ($r_{U/H}$, $r_{U/W}$ and r_R) were parameters defined in the last two sections of the Supporting Information. While spatially constant values for $r_{U/H}$, $r_{U/W}$ and r_R were used in global simulations, we made these values vary among sites in the "simulations on sites" by using information from the corresponding reference. If no information is available for a given site, global parameterizations were nevertheless used. References used in the "simulations on sites" are listed below.

Beck, M. A., and P. A. Sanchez (1996), Soil phosphorus movement and budget after 13 years of fertilized cultivation in the Amazon basin, *Plant Soil*, *184*(1), 23–31.

Bailey, L. D., E. D. Spratt, D. W. L. Read, F. G. Warder, and W. S. Ferguson (1977), Residual effects of phosphorus fertilizer. II. For wheat and flax grown on chernozemic soils in Manitoba, *Can. J. Soil Sci.*, *57*(3), 263–270.

Crews, T. E., and P. C. Brookes (2014), Changes in soil phosphorus forms through time in perennial versus annual agroecosystems, *Agric. Ecosyst. Environ.*, *184*, 168–181, doi:10.1016/j.agee.2013.11.022.

Lan, Z. M., X. J. Lin, F. Wang, H. Zhang, and C. R. Chen (2012), Phosphorus availability and rice grain yield in a paddy soil in response to long-term fertilization, *Biol. Fertil. Soils*, *48*(5), 579–588, doi:10.1007/s00374-011-0650-5.

McKenzie, R. H., J. W. B. Stewart, J. F. Dormaar, and G. B. Schaalje (1992a), Long-term crop rotation and fertilizer effects on phosphorus transformations: II. In a Luvisolic soil, *Can. J. Soil Sci.*, *72*(4), 581–589.

McKenzie, R. H., J. W. B. Stewart, J. F. Dormaar, and G. B. Schaalje (1992b), Long-term crop rotation and fertilizer effects on phosphorus transformations: I. In a Chernozemic soil, *Can. J. Soil Sci.*, *72*(4), 569–579.

Oberson, A., D. K. Friesen, I. M. Rao, S. Bühler, and E. Frossard (2001), Phosphorus transformations in an oxisol under contrasting land-use systems: the role of the soil microbial biomass, *Plant Soil*, *237*(2), 197–210.

Otabbong, E., J. Persson, O. Iakimenko, and L. Sadovnikova (1997), The Ultuna long-term soil organic matter experiment, *Plant Soil*, *195*(1), 17–23.

Vu, D. T., R. D. Armstrong, P. J. Newton, and C. Tang (2011), Long-term changes in phosphorus fractions in growers' paddocks in the northern Victorian grain belt, *Nutr. Cycl. Agroecosystems*, *89*(3), 351–362, doi:10.1007/s10705-010-9400-6.

Wagar, B. I., J. W. B. Stewart, and J. O. Moir (1986), Changes with time in the form and availability of residual fertilizer phosphorus on Chernozemic soils, *Can. J. Soil Sci.*, *66*(1), 105–119.

Wang, X., D. W. Lester, C. N. Guppy, P. V. Lockwood, and C. Tang (2007), Changes in phosphorus fractions at various soil depths following long-term P fertiliser application on a Black Vertosol from south-eastern Queensland, *Aust. J. Soil Res.*, *45*(7), 524, doi:10.1071/SR07069.

Zhang, T. Q., A. F. MacKenzie, B. C. Liang, and C. F. Drury (2004), Soil test phosphorus and phosphorus fractions with long-term phosphorus addition and depletion, *Soil Sci. Soc. Am. J.*, *68*(2), 519–528.

Table S4: Table of the full factorial design. As an example, the mean and variance of cropland grid-cells for P_{ILAB} are given for each run. The variance was used to estimate the contribution of each driver to the spatial variability of P_{ILAB} at grid-cell scale thanks to equation (9) of the Main Text. In the study, 30 simulations were performed for each line in order to represent the uncertainty in the driver estimate and the value provided here (mean and variance) were averaged over these simulations.

Table S5: Description of the two estimates used to assess the uncertainty associated with each driver.

Table S6: Change in the uncertainty in simulated P_{TOT} and P_{ILAB} after removing the uncertainty in each driver. The global indicator of the uncertainty in P_{TOT} and P_{ILAB} was computed as follows (Main Text): "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 coefficient of variations (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}."

Table S7: Variability of proxies of BIOG and FARM at different spatial scales. The variability was computed by using the variance (VAR) between grid-cells, countries, World regions or continents. Only the two extremes spatial scales (grid-cell and continents) were discussed in the Main Text. Because of the difference in the kind of variables used as proxies (fluxes or pools), only the ratio $VAR_{Continent}/VAR_{Grid\text{-}cells}$ (last column) could be compared between the two lines.

Caption of Supporting Figures

Figure S1: Representation of the temporal variation of the different drivers in this study. The hatched bar indicates time periods for which temporal variation was available from global datasets. Blue bars indicate periods for which a given driver was considered as constant. In the latter case, the text written on the blue bars gives the values used during the simulation. Only DEPO corresponding to atmospheric deposition resulting from anthropogenic combustion was indicated on the figure. Other depositions are assumed constant during the whole time-period. Soil input/output resulting from farming practices (FARM) were available for years 1900, 1950, 1960, 1970, 1980, 1990, 1995, and 2000. A linear interpolation is used for each grid-cell over 1900-1950, 1950-1960, 1960-1970, 1970-1980, 1980-1990, 1980-1990 and 1990-1995. FARM equal to 1900 (respectively 2000) was used for the period from the simulation starting year to 1900 (resp. over 2000-2005 period). Note that despite a time constant erosion flux ($f_{sediment}$ in kg of soil/ha), the losses (in kgP/ha) for each grid-cell evolved in time as a function of the land use and land cover change and simulated change in soil P content.

Figure S2: Locations of sites used in the calibration of the soil P dynamics model. The color-pallet corresponds to the number of treatments for each location. Background color shows the 7 large World regions used in the computation of variability (North America, Central and South America, Africa, Oceania, Western Europe, Asia, Russia).

Figure S3: Stage 1 of the calibration. The RMSE between model and observation over 49 treatments (expressed in percent of the averaged observations) for four soil pools was plotted as function the parameter evolved in the main/unique influx (or outflux) of that pool.

Figure S4: Stage 2 of the calibration. The RMSE between model and observation over 49 treatments (expressed in percent of the averaged observations) for P_{ILAB} was plotted as function of change in k_w (line), k_{occ} (column), k_{m1} (y-axis of each panel) and k_{m2} (x-axis of each panel). Masked values correspond to simulations where $k_{m2}<k_{m1}$.

Figure S5: Temporal variation of the different components of P_{TOT} on sites: comparison between observations and simulations (indicated by black arrows) on sites. Each line corresponds to one reference and the different panels of a given line correspond to the different treatments (refer to Table S3 for a full description). For each panel, the 1st bar corresponds to the oldest observation and this observation was used as initial conditions in the model (BIOG). Simulated soil P corresponds to soil P given by the model at the end of the year indicated on the x-axis. The difference between observed uptake and simulated (P_{ILAB}+P_{SEC}) (difference cumulated over the period of simulation) is given in hatched bar on the right of each panel (see the caption of Figure S8 for more information).

Figure S6: Influence plot of the linear regression for P_{ILAB} given in the Figure 2 of the main text. Each symbol correspond to one treatment used in the linear regression. Influence of a given treatment (size of the symbol) is a function of the leverage (x-axis) and residuals (y-axis). The numbers displayed for a treatment with large influence is the number of the reference that reports this treatment (Table S3).

Figure S7: Simulated P_{TOT} (a-d), P_{ILAB} (b-e) and P_{ILAB}/P_{TOT} (c-f) for pasture: 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. Irregular colour pallet corresponding to the 0,20,40,60,80,99th percentiles was chosen for panels a and b. This Figure is similar to the Figure 3 of the Main Text but applied on pasture.

Figure S8: Differences computed for 2005 between the uptake estimated from database (U_{data}) and the simulated ($P_{ILAB}+P_{SEC}$). Only grid-cells with positive ($U_{data}-(P_{ILAB}+P_{SEC})$) are displayed.

Our model was designed such that, within a given time interval, the soil pools were successively affected by: the addition of fertilizers (both chemical and organic), atmospheric deposition, weathering, uptake, residues, occlusion, mineralization, equilibrium between P_{ILAB} and P_{SEC}, and losses. Overall, we prevented the net P output of a given pool from being larger than the pool size, i.e. a soil P pool cannot be negative. An exception to this rule concerns P_{ILAB}, which could be negative *within* a given time step. In particular, the plant uptake (considered as output of P_{ILAB}) is prescribed by the global dataset and could be larger than P_{ILAB}. In that case, P_{ILAB} acts as a sink for P_{SEC}, which shifts the $P_{SEC} \leftrightarrow P_{ILAB}$ equilibrium towards P_{ILAB} before the end of the considered time-step. The variable plotted here is the difference between the uptake estimated from database (U_{data}) and the sum of P_{ILAB} and P_{SEC} before the computation of the $P_{SEC} \leftrightarrow P_{ILAB}$ equilibrium. This difference is plotted for 2005 as example. In the case of $U_{data}>(P_{SEC}+P_{ILAB})$, P_{SEC} and P_{ILAB} are set to 0 in the model.

Figure S9: Percentiles distribution of grid-cells with a non-null cropland fraction in 2005 for P_{TOT} (panel a) and P_{ILAB} (panel b). For both variables, ~13300 grid-cells are concerned. Median and mean of cropland grid-cells were plotted in blue and red, respectively. Error-bars represent 1 std of the 30 simulations used to take the uncertainty in the representation of the drivers into account. Note that the plotted median and mean were not weighted by the cropland area of each grid-cell.

Figure S10: Temporal variation (1900-2005) of some variables involved in the driver representation (lines 1 and 2) and simulated P_{TOT} (line 3) and P_{ILAB} (line 4) for 7 large World regions and at the global scale.

While constant for each grid-cell (Figure S1), CLIM and BIOG, averaged for each World region, varied in time because of the change in the cropland (or pasture) area. P losses (LOSS) vary in time because i) the fraction of soil varied according to LUCC and ii) the soil P content varied according to our simulation. Cropland area shows some discontinuities, due to filters from IMAGE applied on LUHA to make FARM and LUCC consistent (cf. Section “Global datasets” in Supporting Information).

Different sensitivity tests are plotted for P_{TOT} and P_{ILAB}: the “CTRL”, “LUCC=1900” and “LUCC =1900 & FARM=0”. These runs are defined in the “Temporal variation” section of the Supporting Information.

Figure S11: Sensitivity of the results to some model assumptions and analysis settings. The figure displays the change in the driver contributions to the global spatial variability of cropland P_{TOT} (left panels) and P_{ILAB} (right panels) computed at grid-cell scale following a change in:

- the value (VAL) prescribed to all grid-cells to suppress the driver variability, i.e. to set a driver F equal to -1 in the full factorial design (line 1: VAL=median of all cropland grid-cells as in the Main Text vs. line 2: VAL=mean of all cropland grid-cells). Note the following remarks: two different values of VAL were used for cropland and pasture; VAL varied in time during the simulation as function of the cropland/pasture distribution at the global scale; in both lines (1^{st} and 2^{nd}), losses are set to 0 when LOSS=-1.

- the starting year (FIRSTYEAR) of the simulation (line 1: FIRSTYEAR=1700 as in the Main Text vs. line 3: FIRSTYEAR=1500).

- the thickness of the soil horizon that was modelled (line 1: SOILLAYER=0.3m as in the Main Text vs. line 4: SOILLAYER=0.2m).

Each line differs from the reference (first line) due to a change in only one model setting. No uncertainty in the drivers was taken into account here, which explains the lack of error-bars. The different statistical variables (VAL, variance to compute the spatial variability of P_{TOT} and P_{ILAB}) were computed by considering all grid-cells with a non-null cropland fraction and by using the cropland area of each grid-cell as a weight.

Figure S12: The same as Figure S10 but for pasture.

Figure S13: Variation in time of pasture soil P budget and budget components resulting from farming practices (FARM) for 7 large World regions and at the global scale. This Figure is similar to the Figure 6 of the Main Text but applied on pasture.

Table S1

Pool name	Model pool description	P fractions measured on sites thanks to the Hedley method and its derivatives
P_{OLAB}	Labile organic P	[H_2O Po + Resin Po +] Bicarbonate Po
P_{OSTA}	Stable organic P	Hydroxide Po [+ Sonic Po + HCl Po]
P_{ILAB}	Labile inorganic P	[H_2O Pi + Resin Pi +] Bicarbonate Pi
P_{SEC}	Inorganic P bound on secondary minerals	Hydroxide Pi [+ Sonic Pi]
P_{APA}	Apatite	HCl Pi (HCl diluated or not)
P_{OCC}	Occluded inorganic P	Residual P [+ Hot HCl]

Table S2

Soil order	$1/K_S$ (in kgP/ha)	S_{max} (in kgP/ha)
Alfisol/Spodosol	750	1340
Andisol/Aridisol	780	800
Entisol	640	500
Gelisol/Histosol/Inceptisol	650	770
Mollisol	540	740
Oxisol	100	1450
Ultisol	640	1330
Vertisol	320	320

Table S3

Site nb	Reference	Latitude, longitude	Modeled soil horizon for that site (depth of plough layer is indicated in brackets)	Years of soil sampling	Cropland or pasture	Treatments	Observed soil order	Note about initial conditions regarding soil P used for simulations (BIOG)	Way to prescribe soil P budget terms corresponding to FARM (chemical fertilizer, manure, uptake, residues)				Mismatch between simulations and observations? If yes, potential reason
									Note about some P budget terms	Computation of uptake (U) and residues (R_{not-rm}) from available observations (usually withdrawal or harvest, WorH)			
										WorH	$r_{U/WorH}$	r_R	
1	(Crews & Brookes, 2014)	(+51.81; -0.356)	0-23cm (plough layer = 0-23cm)	1893, 2009	pasture and cropland	Unfertilized vs fertilized (35 kgP/ha/yr) regime for both pasture and cropland sites from 1844 (cropland) or 1856 (pasture)	Alfisol	- The P Hedley residual was not provided in Crews and Brook [2014]: we approached it as the total P minus the sum of all other fractions		We assumed that 'estimated P removal in harvest' corresponded to P in grain yield for cropland and P in exported biomass for pasture	From global parameterization (=2 for cropland, 1.25 for pasture)	From global parameterization (=0.7 for cropland; 1 for pasture)	Difficulty in sustaining the observed harvest in unfertilized treatments. Contrary to the observations, we simulated an increase in P_{OLAB} and P_{OSTA}. This could be done to the detriment of P_{ILAB} and could be explained by a not appropriate way to derive P budget terms of FARM.
2	(Zhang *et al.*, 2004)	(+45.42; -73.93)	0-20cm (plough layer = 0-10cm)	1988, 1993, 1997	cropland	Combination of: - Different fertilizing rates (0, 44, 132 kgP/ha/yr), and - Different application periods (continuous fertilization (C) vs only during the first period of the trial (D))	Inceptisol			P in grain yield provided	From global parameterization (=2)	1 (all plant residue returned to the soil as indicated in the reference)	

3	(Wang *et al.*, 2007a)	(-27.43; +150.45)	0-30cm (no information aboutt the depth of the plough layer)	1994 2003	cropland	Different fertilizer application rates (0 and 20kgP/ha/yr)	Vertisol	- The current P in a site uncropped from 1985 (called 'reference' site in *Wang et al.* [2007]) has been used as initial conditions for 1985	- We did not account for the fact that there was no harvest in 1995 and 2000 (we used the averaged yield over 1994-2003 for each year of that period)	P in grain yield provided	From global parameterization (=2)	1 (all plant residue returned to the soil as indicated in the reference)	Difficulty in representing the observed change in P_{TOT}: as suggested in *Wang et al.* [2007], the exchange of P with soil layer below 0.3m could occur on the 'reference' site because of grass growing. That could falsify the use of the 'reference' site in 2003 as the initial conditions. Difficulty in sustaining the observed harvest: role played by acid and residual P pools: lower role of P_{APA} and P_{OCC} in the model than in the observations
4	(Wagar *et al.*, 1986)	Site 1 : (+49.1; -100.6); Site 2 : (+52.146 ; -106.65)	0-15cm (plough layer = 0-15cm for site 1, no information for site 2)	Site 1: 1966, 1970, 1971, 1972, 1973 ; Site 2: 1980, 1981, 1982, 1983, 1984	croppland	Evolution in different sites following single large broadcast P treatment in 1965 for site 1 (0, 200 and 400 kgP/ha) and 1979 for site 2 (0 and 160 kgP/ha)	Mollisol			Site 1: P in grain yield provided in *Bailey et al.* [1977] Site 2: we used the grain yield provided in *Wagar et al.* [1986] and we assumed a P content of wheat grain of 0.284%	From global parameterization (=2)	From global parameterization (=0.7)	Large single P application (1965 or 1979) could lead to large losses (e.g. in one site, 30% of P applied moved below the top 15cm of soil over 5 yr of cropping [*Wagar et al.*, 1986]). This is linked to the soil layer studied, which is very shallow (0-15cm). These losses were not accounted for in our model, which has been calibrated for 0-0.3m.

5	(Vu *et al.*, 2011)	(-36; +142.9)	0-10cm (no information aboutt the depth of the plough layer)	1996, 2005	cropland	Different soil orders	- The soil group was provided in Australian classification and not in USDA classification. Vertosol, Calcarosol, Dermosol, Sodosol were (more or less arbitrarly) appraoched in the model by Vertisol, Mollisol, Inceptisol, Alfisol, respectively. Chromosol was excluded.		- We assumed that the provided P soil inputs were applied in the form of chemical fertilizer	P in grain yield provided (averaged over the different sites within the same soil group)	From global parameterization (=2)	From global parameterization (=0.7)	
6	(Otabbong *et al.*, 1997)	(+59.8; +17.6)	0-20cm (plough layer = 0-20cm)	1956, 1991	Cropland	Different treatments: inorganic P addition only (PK) or in combination with straw (S), green manure (GM), manure (FYM) or sewage sludge (SS). Site with PK treatment and continuous fallow was not considered because no information on the production biomass	Inceptisol		- The rate of chemical P applied was not clearly stated for treatments GM, FYM, SS: we assumed the same rate as in the PK treatment. - We assimilated sewage sludge as manure - We did not consider year-to-year change in parametrization for residues and uptake while it was noted that the rotation included non-leguminous fodder crops - We used the averaged yield over the period redistributed for each year while strong year-to-year variability was noted in the reference paper (e.g. no harvest in 1964 and 1976 due to damage by birds or drought)	We assumed that 'P removed in crops' corresponded to P in grain yield	From global parameterization (=2 for cropland, 1.25 for pasture)	From global parameterization (=0.7 for cropland; 1 for pasture)	Larger increase of $P_{IL,AB}$ in simulations than in observations : - No representation of the rotation and its potential effect on residues, uptake, etc. - Potential wrong characterization of sewage sludge.

7	(Oberson *et al.*, 2001)	(+4.5; -71.317)	0-10cm (plough layer = 0-15cm)	1993, 1997	Cropland and pasture	Soils with contrasting land-use systems (pasture vs cropland) after burning of savanna in 1993	Oxisol	- We used the savannah plot as initial conditions (1993)		We assumed that 'P exports' corresponded to P in grain yield for cropland and P in exported biomass for pasture	From global parameterization (=2 for cropland, 1.25 for pasture)	We assumed that all plant residue returned to the soil for cropland (=1)). From global parameterization for pasture (=1)	Measurements concerned only the 0-0.1m horizon layer. Migration of P below 0.1m (suggested in *Oberson et al.* [2001]) could be high and was not represented in our model, which has been calibrated for 0-0.3m.
8	(McKenzie *et al.*, 1992a)	(+53.1; -114.5)	0-10cm (plough layer = between 0-14 and 0-24cm)	1930, 1986	cropland	Combinations of: - Different rotations (continuously cropped grain-forage rotation (CC) vs. wheat-fallow rotation (WF)) and, - Fertilizer treatments (unfertilized vs. fertilized [6 and 22 kgP/ha/yr for respectively 1930-1979 and 1980-1985])	Alfisol	- We used the uncultivated plot in 1986 as initial conditions (1930)	- We distributed the different P budget terms equally over all years while fallow was not fertilized with P during the wheat-fallow rotation. - We accounted for the change in P fertilizing rate in 1980	P removed in aboveground biomass	Adapted from global parametrization (=1.25) according to the nature of WH (= aboveground biomass and not grain)	1 to estimate R_{not_rm} as P in root biomass	Measurements concerned only the 0-0.1m horizon layer. Migration of P below 0.1m (as losses) or contribution of deeper soil P to the plant uptake were not represented in our model, which has been calibrated for 0-0.3m.
9	(McKenzie *et al.*, 1992b)	(+49.7; -112.8)	0-10cm (no information about the depth of the plough layer)	1912, 1985	cropland	Combinations of: - Different rotations (continuous wheat (W), wheat-fallow (WF), wheat-wheat-fallow (WWF) rotations) - Fertilizer treatments (two rates of N (0 and 45kgN/ha/yr) and two rates of P (0 and 20kgP/ha/yr))	Mollisol	- We used the uncultivated plot in 1985 as initial conditions (1912)	- We distributed the different P budget terms equally over all years while fallow was not fertilized with P during the wheat-fallow rotation. - We accounted for the change in the management of P residue in 1941 and in P fertilizing rate in 1972	P in grain yield provided	From global parameterization (=2)	0.4 during 1912-1941 (to represent the export of straw as mentioned in *McKenzie et al.* [1992]) then = 1	Difficulty in sustaining the observed harvest: - The WH prescribed is provided by *McKenzie et al.* [1992]. It has been computed with very high P content in grain (0.4%) that contributed to empty P_{ILAB}. - Potential temporal trend in the yield was not taken into account while such variability could have contributed to P_{ILAB} regeneration. - Fallow and its potential effect on the regeneration of P_{ILAB} was not represented in the model.

10	(Lan *et al.*, 2012)	(+26.22; +119.07)	0-20cm (plough layer = 0-20cm)	1983, 1987, 1991, 1996, 2000, 2004, 2007, 2009	cropland	Control without fertilization vs. NPK vs NPK+cattle manure vs NPK+rice straw. NPK: chemical fertilizer with level of 12kgP/ha/crop. Shift from double crops harvest to single crop harvest in 2005	Ultisol		- Only available P content of straw and manure was provided: we derived total P in straw and manure from parametrizations used in the model - NaOH II included within sonic	Grain yield was provided and we assumed a P content of rice grain of 0.262%	From global parameterization (=2)	From global parameterization (=0.7)	Model not able to capture the change in P_{TOT}: - Problem with numbers provided in *Lan et al.* [2012] (confusion between numbers provided per crop or per year) ? - Measurements concerned the 0-0.2m horizon layer alone. The contribution of deeper soil P to the plant uptake was not represented in our model, which was calibrated for 0-0.3m. Difficulty in sustaining the observed harvest: - explained by the non-representation of double crops harvests over the 1983-2005 period ?
11	(Beck & Sanchez, 1996)	(-5.75; -76.1)	0-40cm (plough layer = 0-15cm)	1973, 1985	cropland	Unfertilized and fertilized cropland after slash and burning of a secondary forest in 1972	Ultisol	- Use of soil P content of forest as initial conditions (1973)	- NaOH II included within sonic	- P uptake and P in residue provided directly (Table 6 of *Beck and Sanchez* [1996]) following assumptions about P in root biomass - P in root biomass computed thanks to Tables 2 (grain/total biomass and P content) and 6 (P removed in grain and in residue) of *Beck and Sanchez* [1996](~0.23 and 5.4kgP/ha/yr for non-fertilized and fertilized trials respectively).			Difficulty in capturing the change in P_{TOT}: already mentioned in *Beck and Sanchez* [1996]. Higher labile fraction in the model than in observations: - Inappropriate labile/stable fraction of soil inputs (in particular, *Beck and Sanchez* [1996] did not provide a distinction between P applied in the form of biomass or in the form of ash)

Table S4

	Unique factors						Interactions between two factors															Interactions between three factors																P_{ILAB} : mean of cropland grid-cells	P_{ILAB} : variance of cropland grid-cells
N° run	BIOG	LOSS	BUFF	CLIM	FARM	DEPO	BIOG, LOSS	BIOG, BUFF	BIOG, CLIM	BIOG, FARM	BIOG, DEPO	LOSS, BUFF	LOSS, CLIM	LOSS, FARM	LOSS, DEPO	BUFF, CLIM	BUFF, FARM	BUFF, DEPO	CLIM, FARM	CLIM, DEPO	FARM, DEPO	BIOG, LOSS, BUFF	BIOG, LOSS, CLIM	BIOG, LOSS, FARM	BIOG, LOSS, DEPO	BIOG, BUFF, CLIM	BIOG, BUFF, FARM	BIOG, BUFF, DEPO	BIOG, CLIM, FARM	BIOG, CLIM, DEPO	BIOG, FARM, DEPO	LOSS, BUFF, CLIM	LOSS, BUFF, FARM	LOSS, BUFF, DEPO	LOSS, CLIM, FARM	LOSS, CLIM, DEPO	LOSS, FARM, DEPO	BUFF, CLIM, FARM	BUFF, CLIM, DEPO	BUFF, FARM, DEPO	CLIM, FARM, DEPO		
1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	100.1	1307.9
2	-1	-1	-1	-1	-1	1	1	1	1	1	-1	1	1	1	-1	1	1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	1	1	-1	-1	1	-1	1	1	1	102.1	1159.4
3	-1	-1	-1	-1	1	-1	1	1	1	-1	1	1	1	-1	1	1	-1	1	-1	1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	179.9	113414.1
4	-1	-1	-1	-1	1	1	1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	-1	1	1	1	1	-1	-1	1	1	1	1	-1	-1	183.9	104943.3
5	-1	-1	-1	1	-1	-1	1	1	-1	1	1	1	-1	1	1	-1	1	1	-1	-1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	1	-1	-1	1	1	-1	1	107.1	1452.5
6	-1	-1	-1	1	-1	1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	1	1	-1	1	1	-1	1	-1	110.5	1521.4
7	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	191.5	111607.2
8	-1	-1	-1	1	1	1	1	1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	-1	1	1	1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1	195.2	110197.4
9	-1	-1	1	-1	-1	-1	1	-1	1	1	1	-1	1	1	1	-1	-1	-1	1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1	97.6	2145.7
10	-1	-1	1	-1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	-1	-1	1	1	1	1	-1	-1	1	1	-1	99.9	2018.2
11	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	179.9	117503.4
12	-1	-1	1	-1	1	1	1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	1	1	1	-1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	183.9	121175.3
13	-1	-1	1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	1	-1	-1	1	1	1	1	-1	-1	105.4	2412.2
14	-1	-1	1	1	-1	1	1	-1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	108.3	2577.3
15	-1	-1	1	1	1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	1	-1	-1	-1	1	-1	1	1	-1	-1	1	-1	1	1	1	191.4	120811.5
16	-1	-1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	195.0	111549.5
17	-1	1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	92.4	1393.9
18	-1	1	-1	-1	-1	1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	1	-1	-1	-1	1	-1	1	1	1	1	-1	1	-1	-1	-1	94.6	1271.4
19	-1	1	-1	-1	1	-1	-1	1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	166.2	107372.7
20	-1	1	-1	-1	1	1	-1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	1	1	-1	169.5	100764.1

21	-1	1	-1	1	-1	-1	-1	1	-1	1	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	1	-1	1	98.0	1478.7
22	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	101.0	1500.8
23	-1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	174.8	99019.3
24	-1	1	-1	1	1	1	-1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	178.6	106039.9
25	-1	1	1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	-1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	91.1	2192.5
26	-1	1	1	-1	-1	1	-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	93.5	2104.8
27	-1	1	1	-1	1	-1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	166.4	108042.2
28	-1	1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	-1	169.7	104324.3
29	-1	1	1	1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	-1	1	1	1	1	-1	1	-1	-1	-1	-1	1	-1	-1	1	1	96.8	2357.1
30	-1	1	1	1	-1	1	-1	-1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	-1	99.6	2443.8
31	-1	1	1	1	1	-1	-1	-1	-1	-1	1	1	1	1	-1	1	1	-1	1	-1	-1	-1	-1	-1	1	-1	-1	1	-1	1	1	1	1	-1	1	-1	-1	1	-1	-1	-1	175.7	107454.0
32	-1	1	1	1	1	1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	179.1	108334.9
33	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	-1	-1	-1	-1	178.8	64772.7
34	1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	1	1	1	-1	1	1	-1	1	-1	-1	1	1	1	-1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	1	1	1	181.5	64142.3
35	1	-1	-1	-1	1	-1	-1	-1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	1	256.6	172211.3
36	1	-1	-1	-1	1	1	-1	-1	-1	1	1	1	1	-1	-1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	-1	-1	-1	-1	1	-1	1	1	1	1	-1	1	1	-1	-1	259.9	179635.4
37	1	-1	-1	1	-1	-1	-1	-1	1	-1	-1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1	-1	1	193.9	76716.8
38	1	-1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	1	-1	1	-1	196.9	76627.1
39	1	-1	-1	1	1	-1	-1	-1	1	1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	1	1	-1	275.0	185003.6
40	1	-1	-1	1	1	1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	279.7	198783.8
41	1	-1	1	-1	-1	-1	-1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	1	1	1	-1	179.6	70602.8
42	1	-1	1	-1	-1	1	-1	1	-1	-1	1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	1	182.6	70927.8
43	1	-1	1	-1	1	-1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	258.7	190144.8
44	1	-1	1	-1	1	1	-1	1	-1	1	1	-1	1	-1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	1	1	-1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	-1	262.8	195518.8
45	1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	-1	1	1	1	1	-1	-1	-1	1	1	195.0	83772.9
46	1	-1	1	1	-1	1	-1	1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	1	-1	1	-1	1	-1	-1	197.8	83909.1
47	1	-1	1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	1	-1	1	1	1	-1	-1	-1	277.5	214022.7
48	1	-1	1	1	1	1	-1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	280.8	205477.1

49	1	1	-1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	-1	-1	1	1	1	1	1	1	-1	-1	-1	-1	1	1	1	1	1	1	1	1	1	1	1	1	-1	-1	-1	-1	159.2	49104.4
50	1	1	-1	-1	-1	1	1	-1	-1	-1	1	-1	-1	-1	1	1	1	-1	1	-1	-1	-1	-1	-1	1	1	1	-1	1	-1	-1	1	1	-1	1	-1	-1	-1	1	1	1	162.2	49001.7
51	1	1	-1	-1	1	-1	1	-1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	-1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	1	1	231.1	149388.5
52	1	1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	-1	-1	1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	1	-1	-1	235.3	160713.4
53	1	1	-1	1	-1	-1	1	-1	1	-1	-1	-1	1	-1	-1	-1	1	1	-1	-1	1	-1	1	-1	-1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	1	1	-1	1	171.1	56552.1
54	1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	174.0	57111.7
55	1	1	-1	1	1	-1	1	-1	1	1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	-1	1	1	-1	246.3	164267.8
56	1	1	-1	1	1	1	1	-1	1	1	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	-1	-1	-1	1	249.7	172533.9
57	1	1	1	-1	-1	-1	1	1	-1	-1	-1	1	-1	-1	-1	-1	-1	-1	1	1	1	1	-1	-1	-1	-1	-1	-1	1	1	1	-1	-1	-1	1	1	1	1	1	1	-1	161.4	55602.3
58	1	1	1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	1	-1	-1	1	-1	-1	1	-1	-1	1	1	-1	-1	-1	-1	1	1	-1	-1	1	-1	-1	1	164.0	55307.7
59	1	1	1	-1	1	-1	1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	-1	1	-1	1	234.2	173025.4
60	1	1	1	-1	1	1	1	1	-1	1	1	1	-1	1	1	-1	1	1	-1	-1	1	1	-1	1	1	-1	1	1	-1	-1	1	-1	1	1	-1	-1	1	-1	-1	1	-1	237.6	166749.8
61	1	1	1	1	-1	-1	1	1	1	-1	-1	1	1	-1	-1	1	-1	-1	-1	-1	1	1	1	-1	-1	1	-1	-1	-1	-1	1	1	-1	-1	-1	-1	1	-1	-1	1	1	172.6	62269.6
62	1	1	1	1	-1	1	1	1	1	-1	1	1	1	-1	1	1	-1	1	-1	1	-1	1	1	-1	1	1	-1	1	-1	1	-1	1	-1	1	-1	1	-1	-1	1	-1	-1	175.2	63380.9
63	1	1	1	1	1	-1	1	1	1	1	-1	1	1	1	-1	1	1	-1	1	-1	-1	1	1	1	-1	1	1	-1	1	-1	-1	1	1	-1	1	-1	-1	1	-1	-1	-1	249.0	184322.0
64	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	252.2	179245.6

Table S5

Driver name	Description of the two estimates used to assess the uncertainty associated with the driver
BIOG (natural soil biogeochemical background)	Use of the mean and standard deviation associated to P_{TOT} provided by *Yang et al.* [2013]: BIOG1 = $\overline{P_{TOT}} - \sqrt{3}.\ std_{PTOT}$ BIOG2 = $\overline{P_{TOT}} + \sqrt{3}.\ std_{PTOT}$ Only the uncertainty associated with P_{TOT} was accounted for (i.e. the same contribution of the different forms to P_{TOT} was used in BIOG1 and BIOG2).
LUCC (land use and land cover change)	Uncertainty not taken into account.
FARM (farming practices)	We made the soil budget (FARM input – output) vary around the value provided by *Bouwman et al.* [2011]. FARM1 = 70% of the soil budget computed by *Bouwman et al.* [2011] FARM2 = 130% of the soil budget computed by *Bouwman et al.* [2011]. The same contribution of the different fluxes (chemical fertilizer, manure, residue, uptake) to the soil budget was used in FARM1 and FARM2.
CLIM (soil temperature and soil water content)	Use of the simulations provided by 2 different DGVMs CLIM1 = soil water content and temperature simulated by ISBA [*Decharme et al.*, 2013] CLIM2 = soil water content and temperature simulated by ORCHIDEE [*Krinner et al.*, 2005]
LOSS (P losses through soil erosion)	Use of two 'scenarios' about flux of eroded sediment ($f_{sediment}$) provided by *Van Oost et al.* [2007] LOSS1 = computation of losses using the lower 'scenario' provided by *Van Oost et al.* [2007] (called “normal” in the Supporting Information of *Van Oost et al.* [2007]) LOSS2 = computation of losses using the higher scenario provided by *Van Oost et al.* [2007] ($f_{sediment}$ in LOSS2 = $f_{sediment}$ in LOSS1*1.26)
DEPO (atmospheric deposition)	Use of the mean and upper/lower boundaries provided in *Wang et al.* [2014] DEPO1=25% of mean deposition provided by *Wang et al.* [2014] DEPO2=175 % of mean deposition provided by *Wang et al.* [2014] The uncertainty concerns the total deposition (and not the proportions of the total deposition going into P_{ILAB} and P_{APA}).
BUFF (soil buffering capacity)	BUFF1 = 70% of S_{max} provided by *Wang et al.* [2007] BUFF2 = 130**%** of S_{max} provided by *Wang et al.* [2007] We considered K_s and S_{max} as co-varying and described both BUFF1 and BUFF2 by the same K_S/S_{max} ratio as the one provided by *Wang et al.* [2007].

Table S6

Drivers whose the uncertainty was removed	Global indicator of uncertainty in P_{TOT}	Global indicator of uncertainty in P_{ILAB}
None	46.7	60.8
BIOG	2.3	23.5
FARM	46.4	55.7
LOSS	46.5	60.9
BUFF	46.5	56.6
CLIM	46.5	60.8
DEPO	46.6	60.8

Table S7

Driver	Proxy of driver	Variability (VAR) computed between:				$VAR_{Continents}/VAR_{Grid\text{-}cells}$
		Grid-cells (n=13,352)	Countries (n=226)	Regions (n=25)	Continents (n=7)	
FARM	Sum of annual soil P budget resulting from farming practices over 1956-2005	163,704	2,438,563	90,098	12,763	0.08
BIOG	P_{TOT} in unmanaged soils	1,955,292	926,769	270,156	31,625	0.02

Supporting Figures

Figure S1

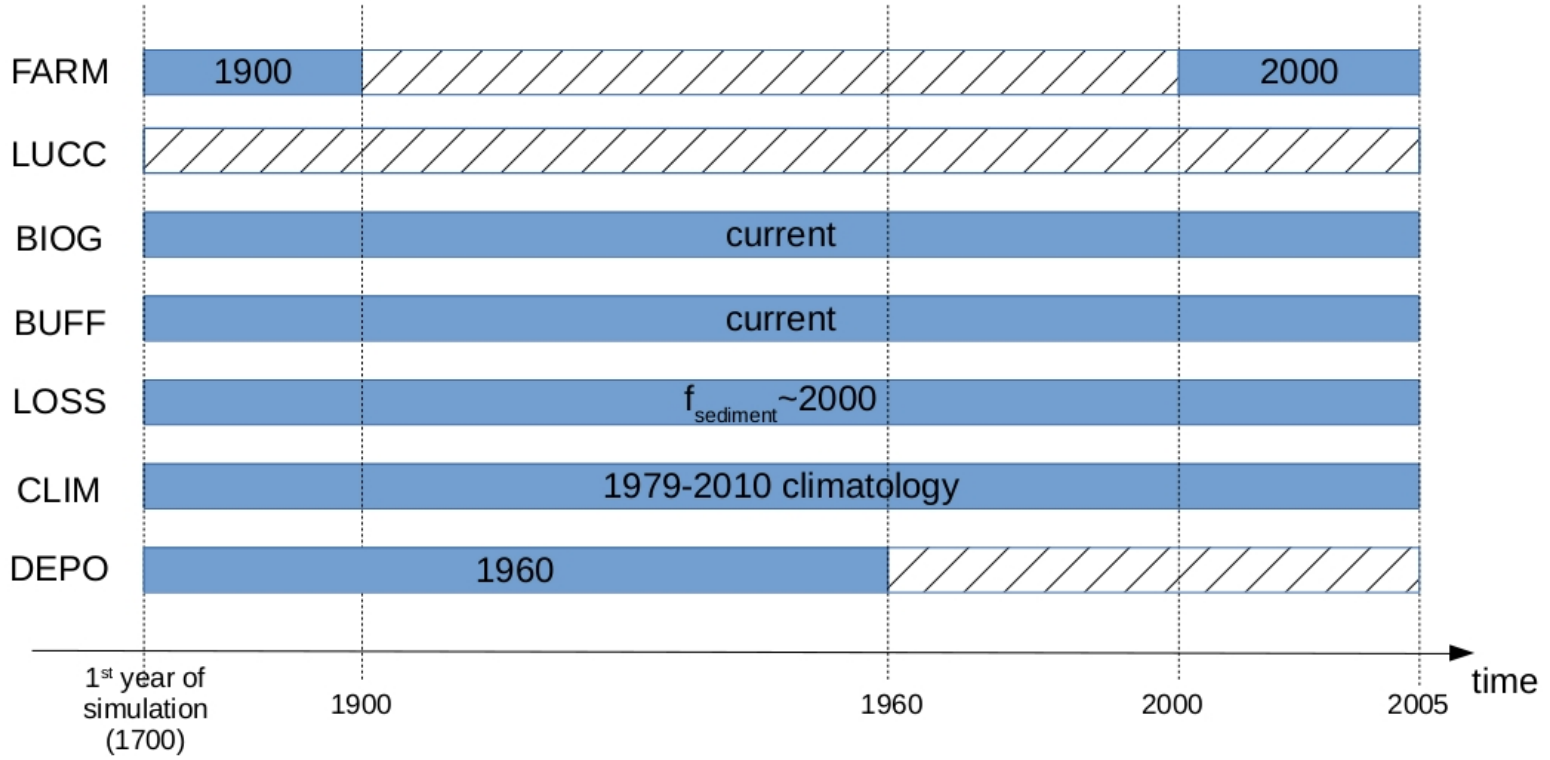


Figure S2

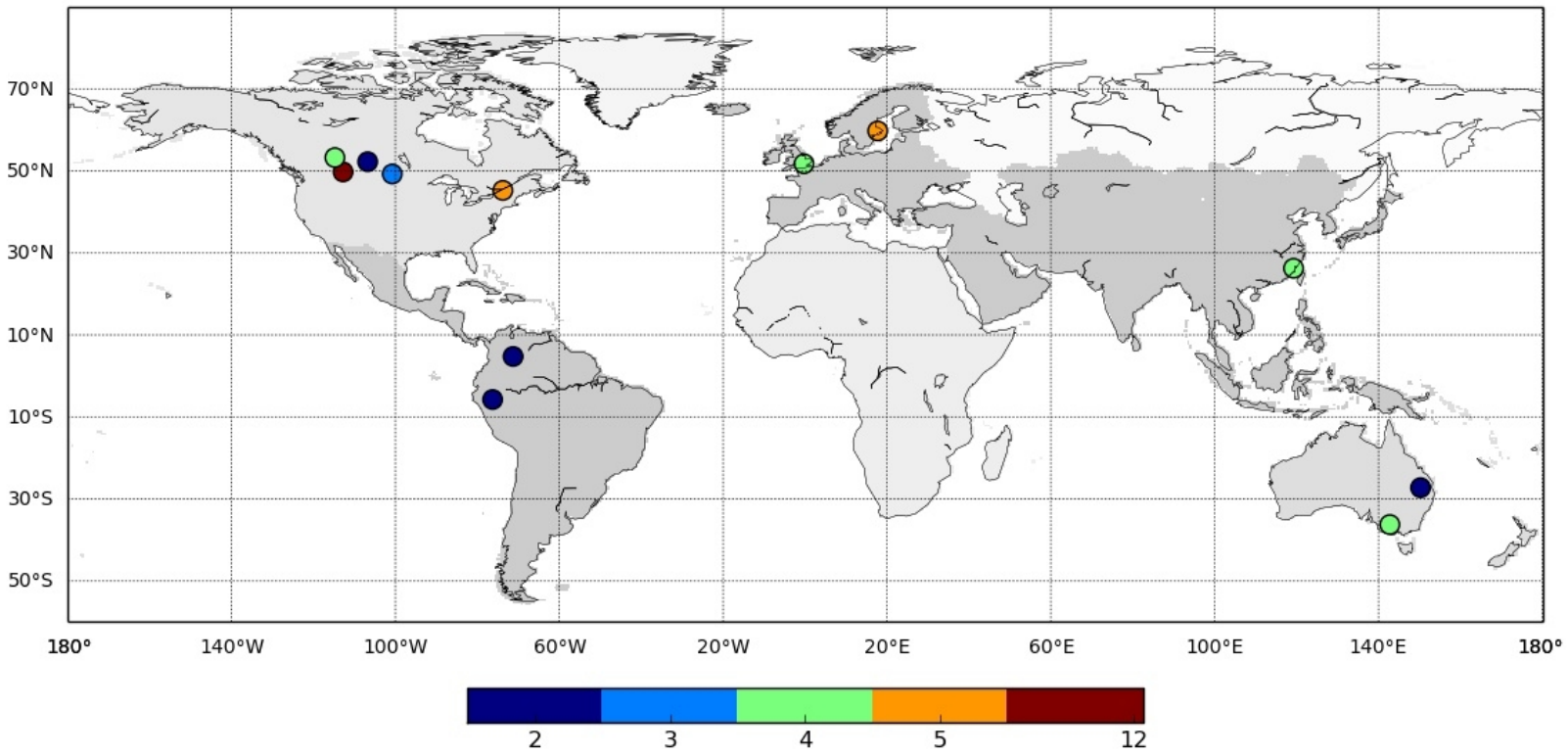

70°N
50°N
30°N
10°N
10°S
30°S
50°S
180°
140°W
100°W
60°W
20°W
20°E
60°E
100°E
140°E
180°
2
3
4
5
12

Figure S3

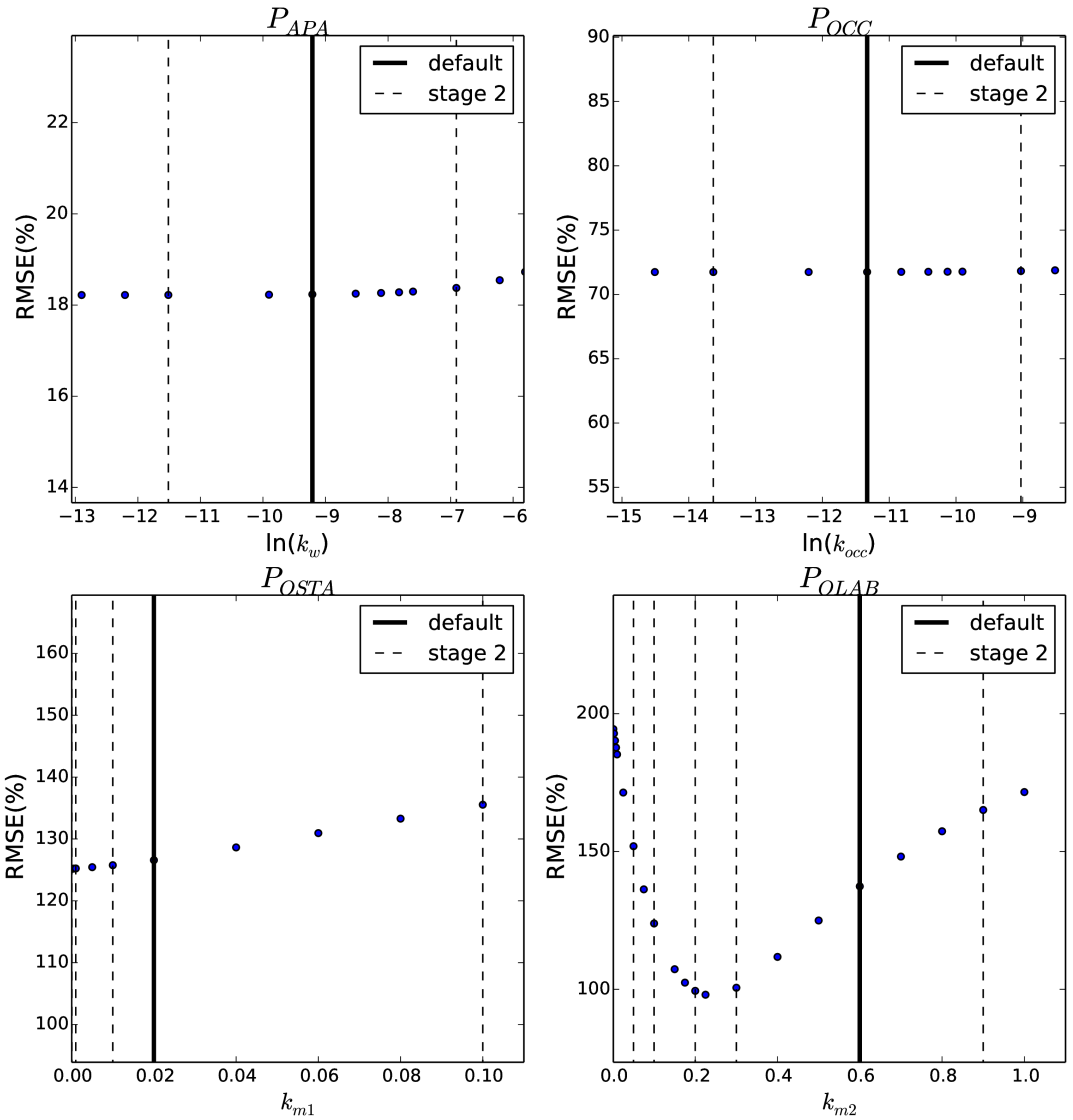

P_{APA}
default
stage 2
RMSE(%)
$\ln(k_w)$
P_{OCC}
default
stage 2
RMSE(%)
$\ln(k_{occ})$
P_{OSTA}
default
stage 2
RMSE(%)
k_{m1}
P_{OLAB}
default
stage 2
RMSE(%)
k_{m2}

Figure S4

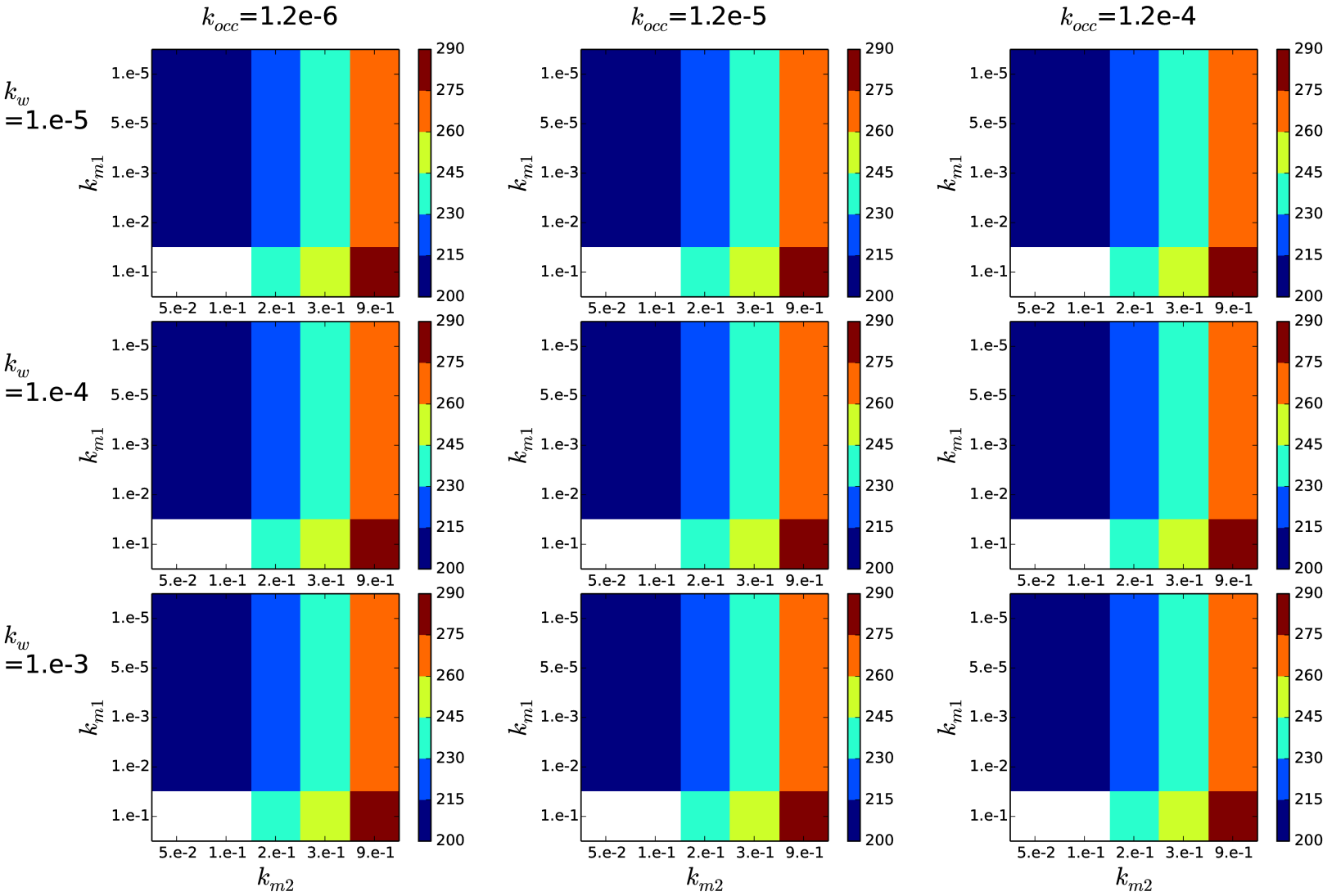

k_{occ}=1.2e-6
k_{occ}=1.2e-5
k_{occ}=1.2e-4
k_w =1.e-5
k_w =1.e-4
k_w =1.e-3
k_{m1}
k_{m2}
1.e-5
5.e-5
1.e-3
1.e-2
1.e-1
5.e-2 1.e-1 2.e-1 3.e-1 9.e-1
290
275
260
245
230
215
200

Figure S5 (1/3)

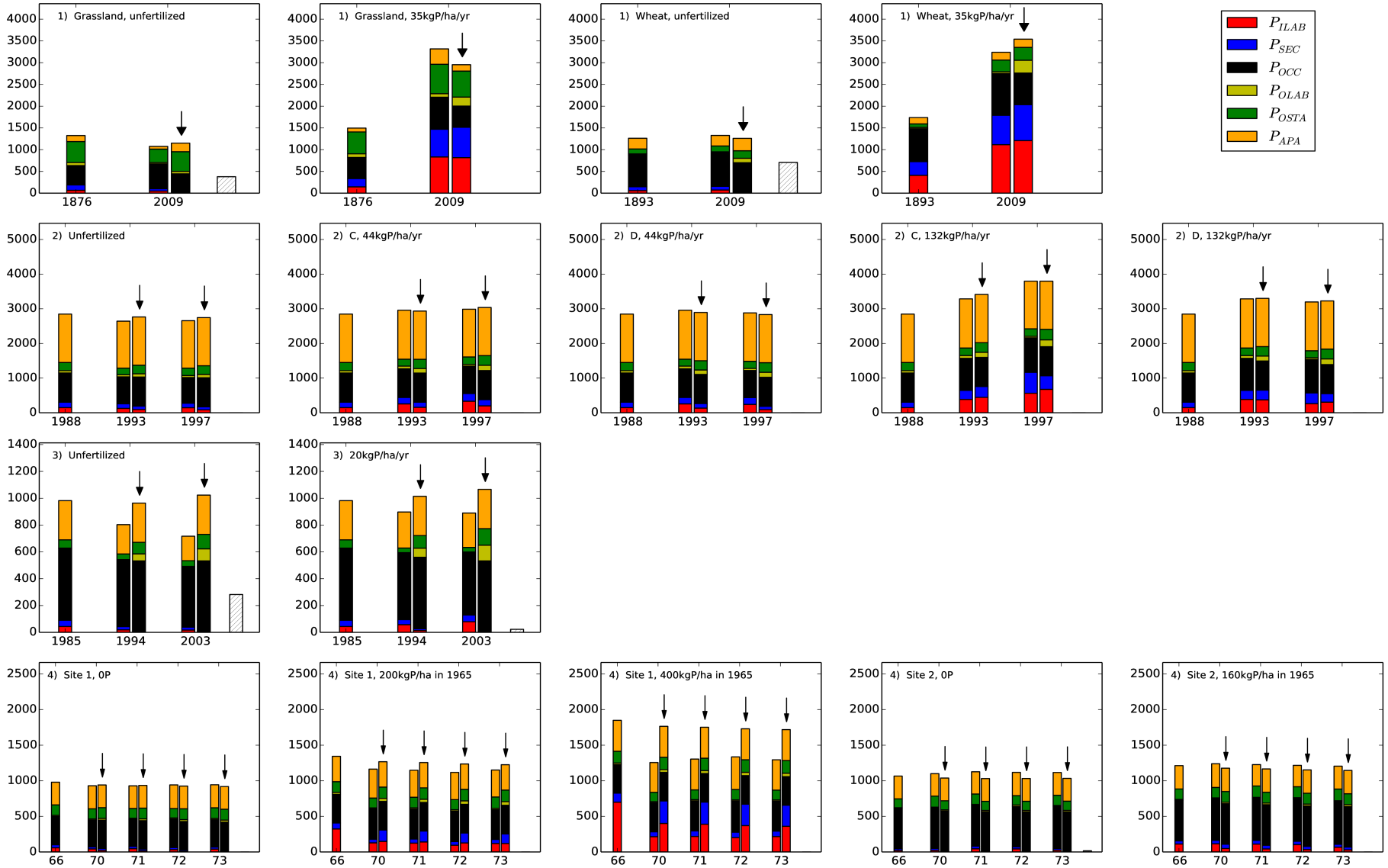

1) Grassland, unfertilized
1) Grassland, 35kgP/ha/yr
1) Wheat, unfertilized
1) Wheat, 35kgP/ha/yr
P_{ILAB}
P_{SEC}
P_{OCC}
P_{OLAB}
P_{OSTA}
P_{APA}
2) Unfertilized
2) C, 44kgP/ha/yr
2) D, 44kgP/ha/yr
2) C, 132kgP/ha/yr
2) D, 132kgP/ha/yr
3) Unfertilized
3) 20kgP/ha/yr
4) Site 1, 0P
4) Site 1, 200kgP/ha in 1965
4) Site 1, 400kgP/ha in 1965
4) Site 2, 0P
4) Site 2, 160kgP/ha in 1965

Figure S5 (2/3)

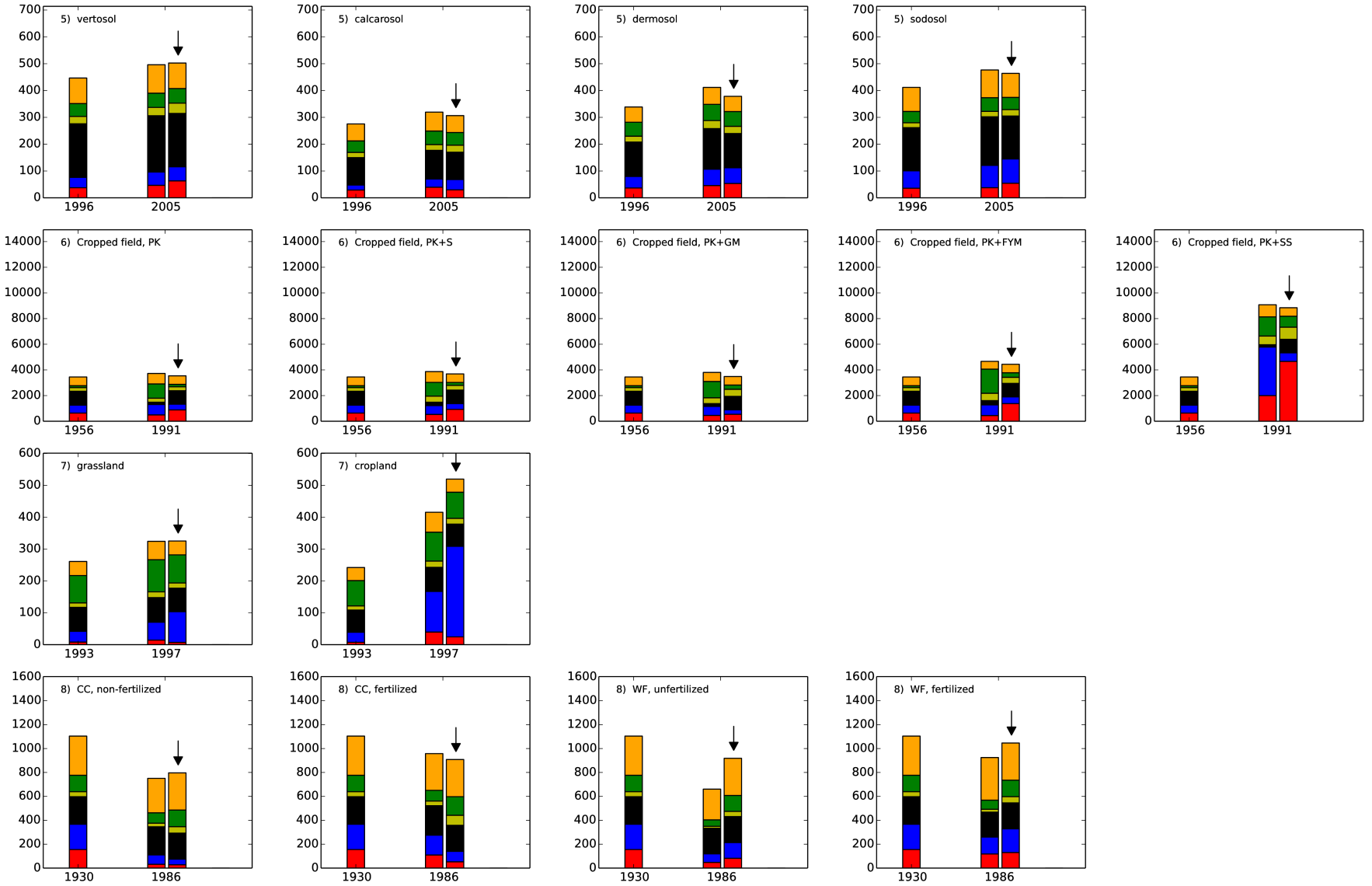

5) vertosol
5) calcarosol
5) dermosol
5) sodosol
6) Cropped field, PK
6) Cropped field, PK+S
6) Cropped field, PK+GM
6) Cropped field, PK+FYM
6) Cropped field, PK+SS
7) grassland
7) cropland
8) CC, non-fertilized
8) CC, fertilized
8) WF, unfertilized
8) WF, fertilized
1996
2005
1956
1991
1993
1997
1930
1986

Figure S5 (3/3)

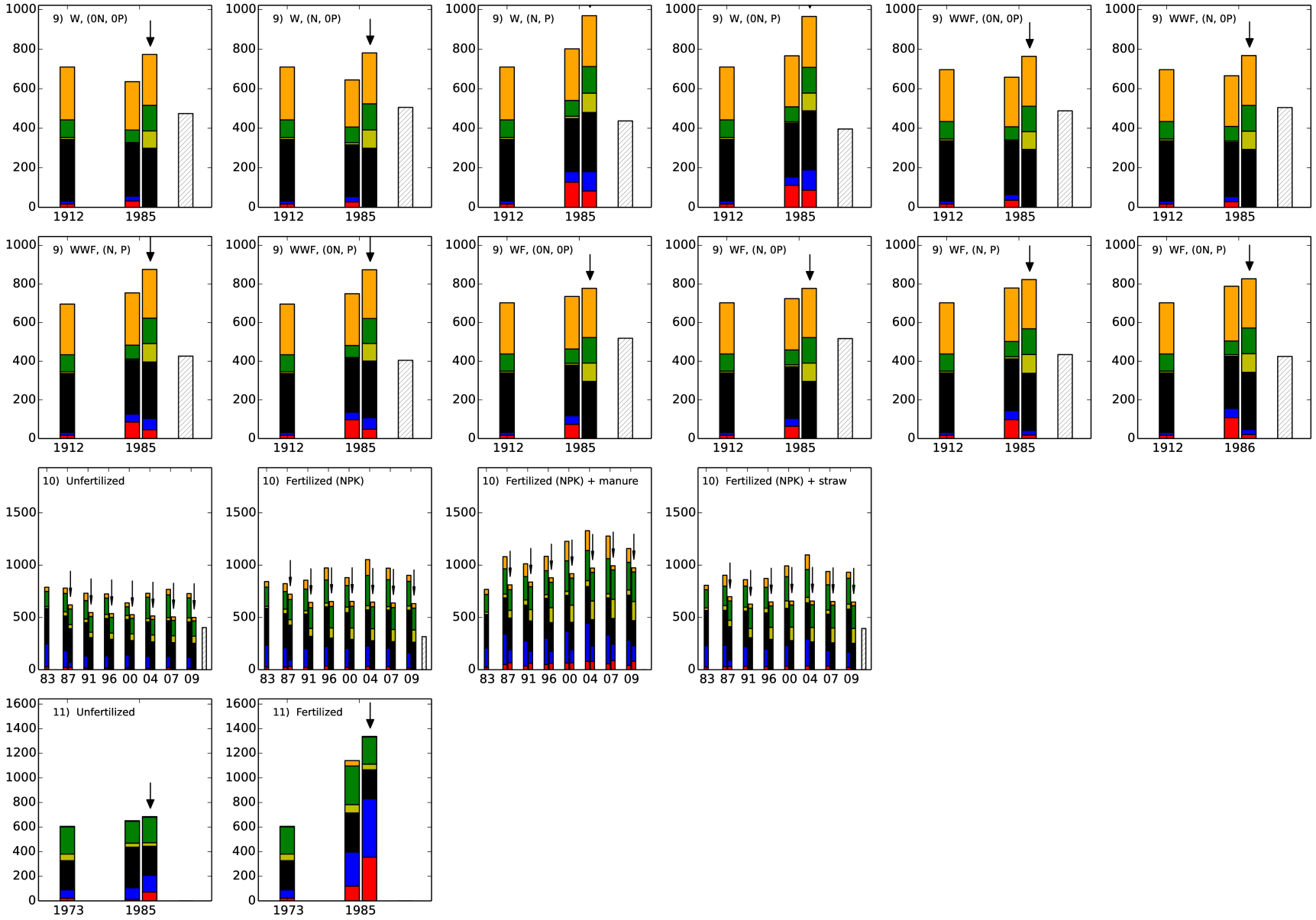

9) W, (0N, 0P)
9) W, (N, 0P)
9) W, (N, P)
9) W, (0N, P)
9) WWF, (0N, 0P)
9) WWF, (N, 0P)
9) WWF, (N, P)
9) WWF, (0N, P)
9) WF, (0N, 0P)
9) WF, (N, 0P)
9) WF, (N, P)
9) WF, (0N, P)
10) Unfertilized
10) Fertilized (NPK)
10) Fertilized (NPK) + manure
10) Fertilized (NPK) + straw
11) Unfertilized
11) Fertilized
1912
1985
1986
1973
83 87 91 96 00 04 07 09

Figure S6

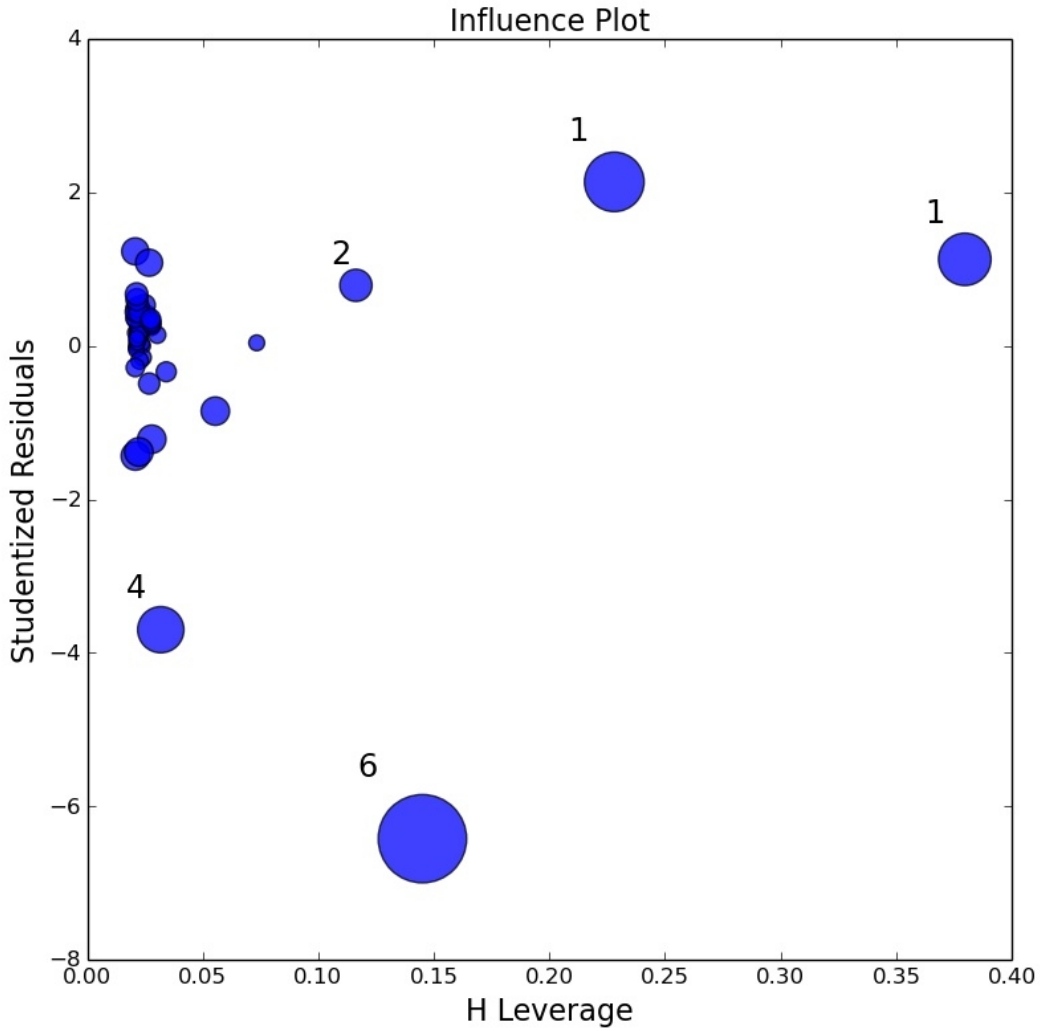

Influence Plot
Studentized Residuals
H Leverage
1
1
2
4
6
4
2
0
−2
−4
−6
−8
0.00
0.05
0.10
0.15
0.20
0.25
0.30
0.35
0.40

Figure S7

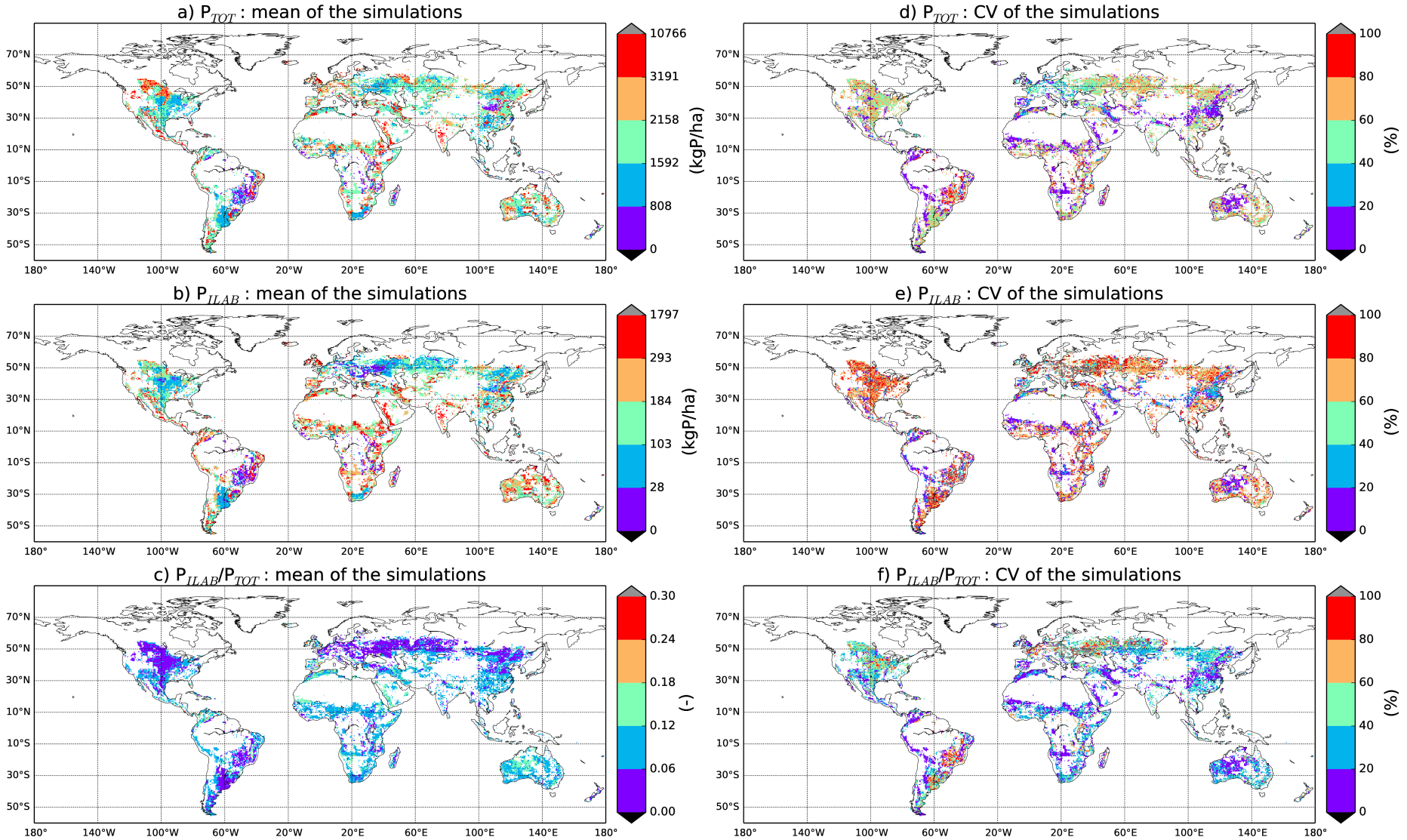

a) P_{TOT} : mean of the simulations
10766
3191
2158
1592
808
0
(kgP/ha)
d) P_{TOT} : CV of the simulations
100
80
60
40
20
0
(%)
b) P_{ILAB} : mean of the simulations
1797
293
184
103
28
0
(kgP/ha)
e) P_{ILAB} : CV of the simulations
100
80
60
40
20
0
(%)
c) P_{ILAB}/P_{TOT} : mean of the simulations
0.30
0.24
0.18
0.12
0.06
0.00
(-)
f) P_{ILAB}/P_{TOT} : CV of the simulations
100
80
60
40
20
0
(%)
70°N
50°N
30°N
10°N
10°S
30°S
50°S
180°
140°W
100°W
60°W
20°W
20°E
60°E
100°E
140°E
180°

Figure S8

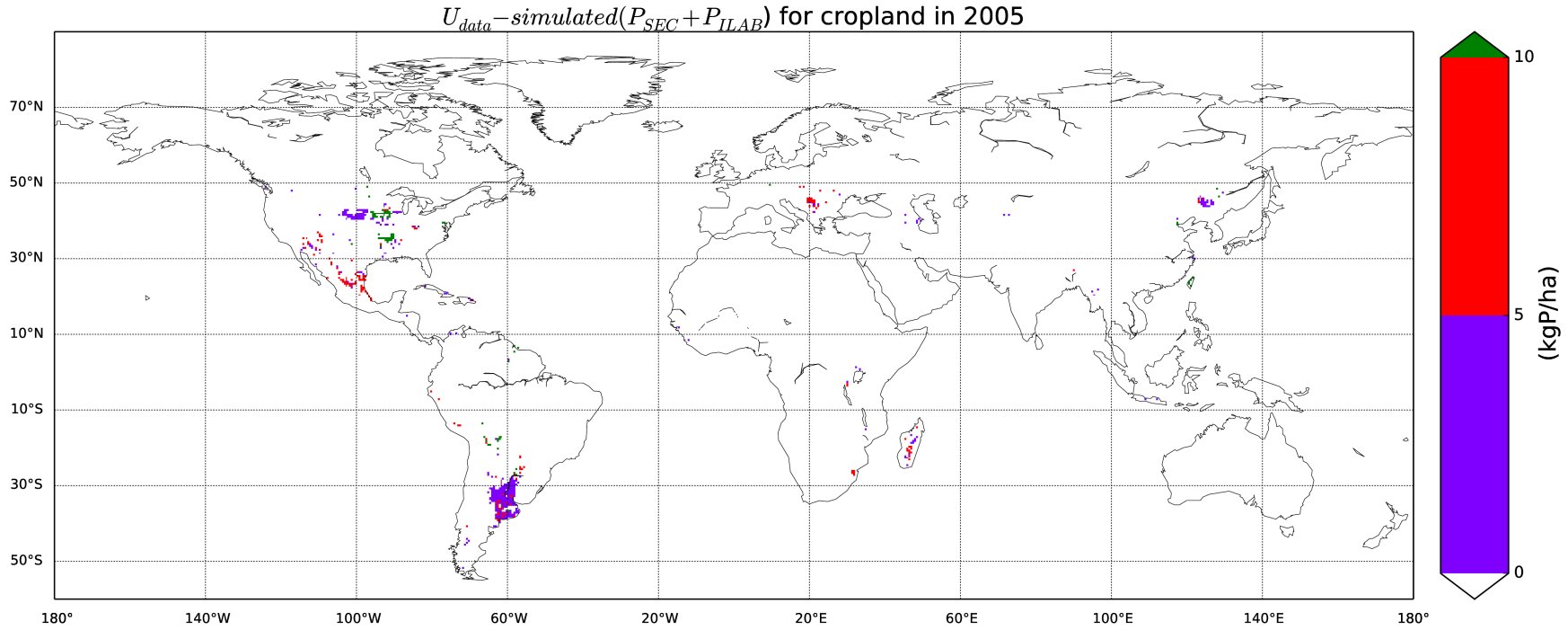

$U_{data} - simulated(P_{SEC} + P_{ILAB})$ for cropland in 2005
70°N
50°N
30°N
10°N
10°S
30°S
50°S
180°
140°W
100°W
60°W
20°W
20°E
60°E
100°E
140°E
180°
10
5
0
(kgP/ha)

Figure S9

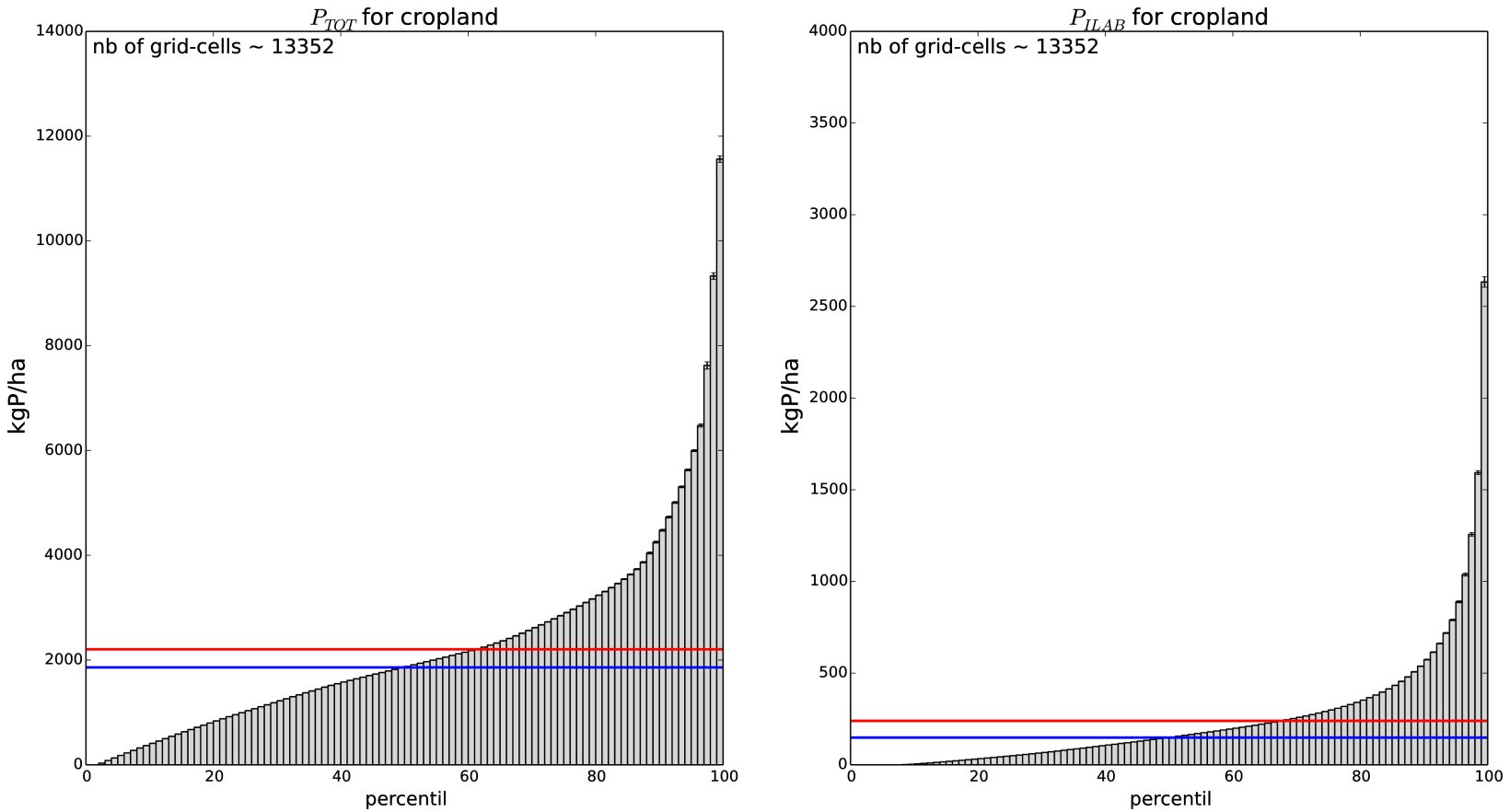

P_{TOT} for cropland
nb of grid-cells ~ 13352
kgP/ha
percentil
P_{ILAB} for cropland
nb of grid-cells ~ 13352
kgP/ha
percentil

Figure S10

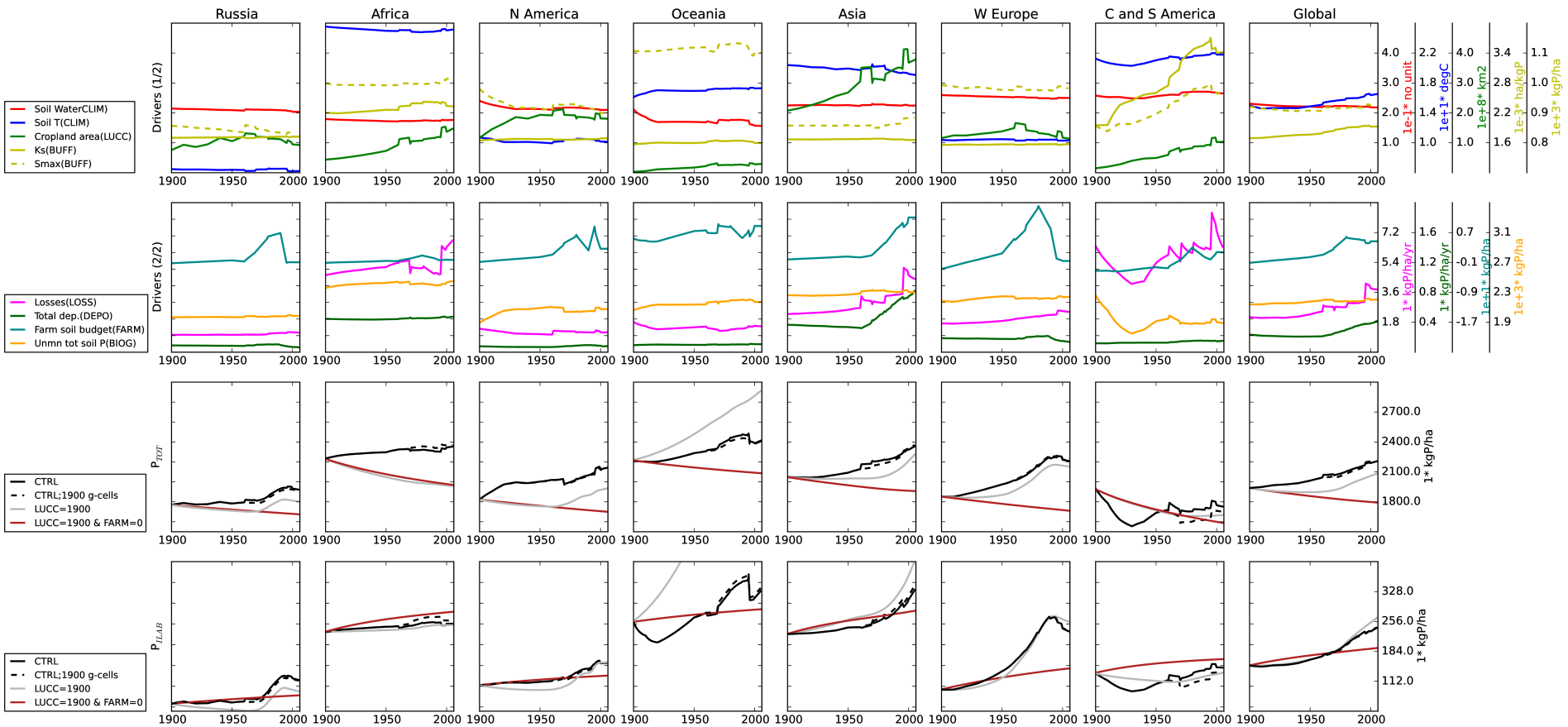

Russia
Africa
N America
Oceania
Asia
W Europe
C and S America
Global
Drivers (1/2)
Soil WaterCLIM)
Soil T(CLIM)
Cropland area(LUCC)
Ks(BUFF)
Smax(BUFF)
1e-1* no unit
1e+1* degC
1e+8* km2
1e-3* ha/kgP
1e+3* kgP/ha
Drivers (2/2)
Losses(LOSS)
Total dep.(DEPO)
Farm soil budget(FARM)
Unmn tot soil P(BIOG)
1* kgP/ha/yr
1* kgP/ha/yr
1e+1* kgP/ha
1e+3* kgP/ha
P_TOT
CTRL
CTRL;1900 g-cells
LUCC=1900
LUCC=1900 & FARM=0
1* kgP/ha
P_LAB
CTRL
CTRL;1900 g-cells
LUCC=1900
LUCC=1900 & FARM=0
1* kgP/ha
1900
1950
2000

Figure S11

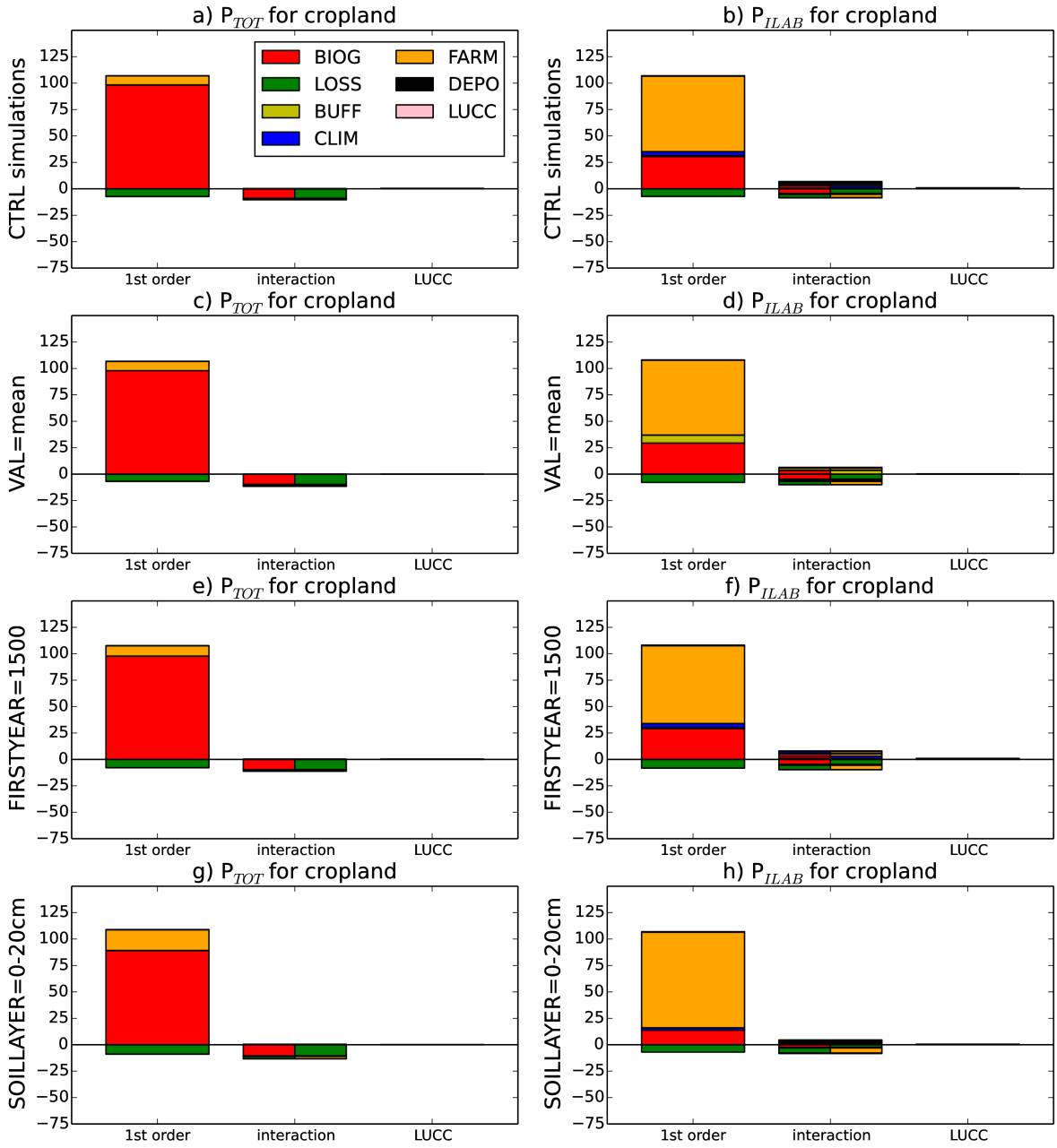

a) P$_{TOT}$ for cropland
b) P$_{ILAB}$ for cropland
c) P$_{TOT}$ for cropland
d) P$_{ILAB}$ for cropland
e) P$_{TOT}$ for cropland
f) P$_{ILAB}$ for cropland
g) P$_{TOT}$ for cropland
h) P$_{ILAB}$ for cropland
CTRL simulations
VAL=mean
FIRSTYEAR=1500
SOILLAYER=0-20cm
BIOG
LOSS
BUFF
CLIM
FARM
DEPO
LUCC
1st order
interaction
LUCC
125
100
75
50
25
0
−25
−50
−75

Figure S12

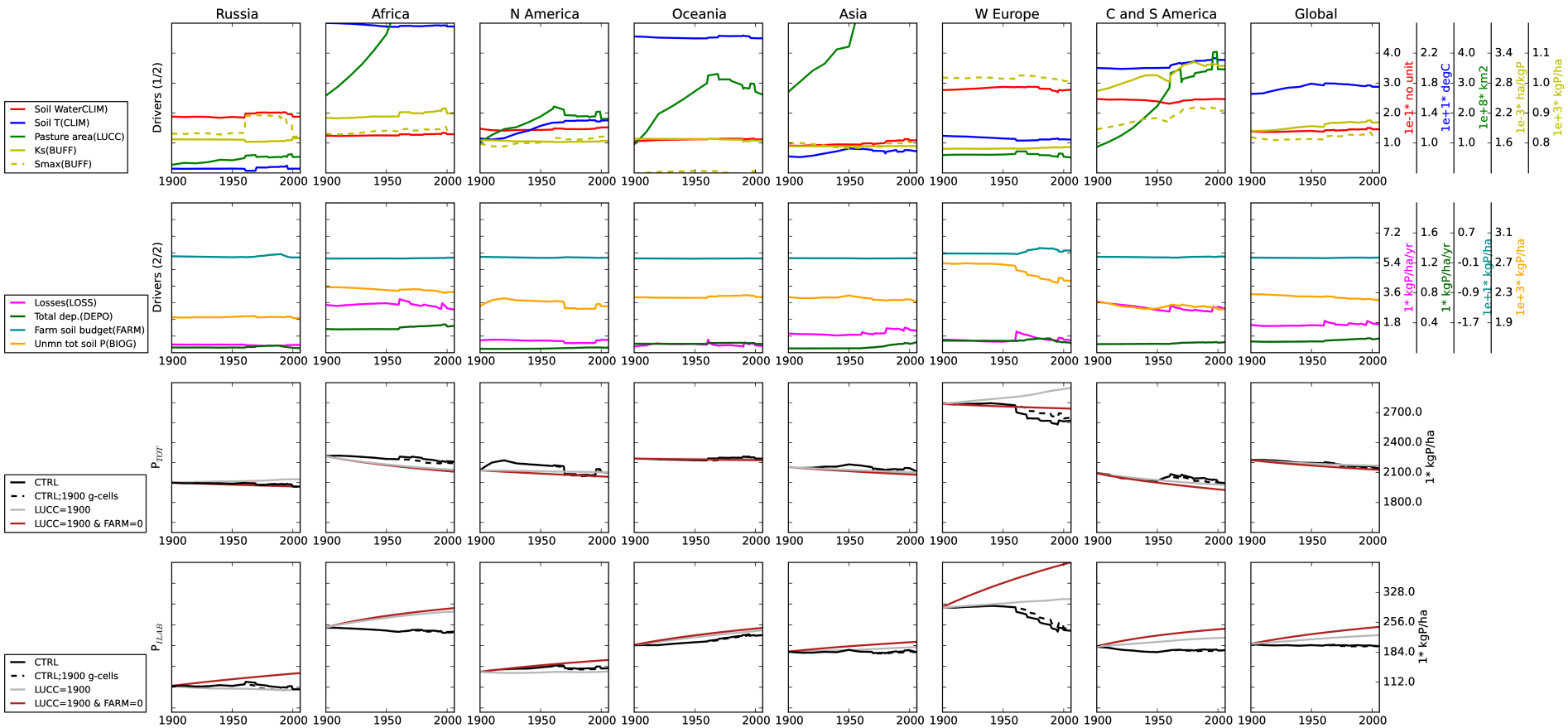
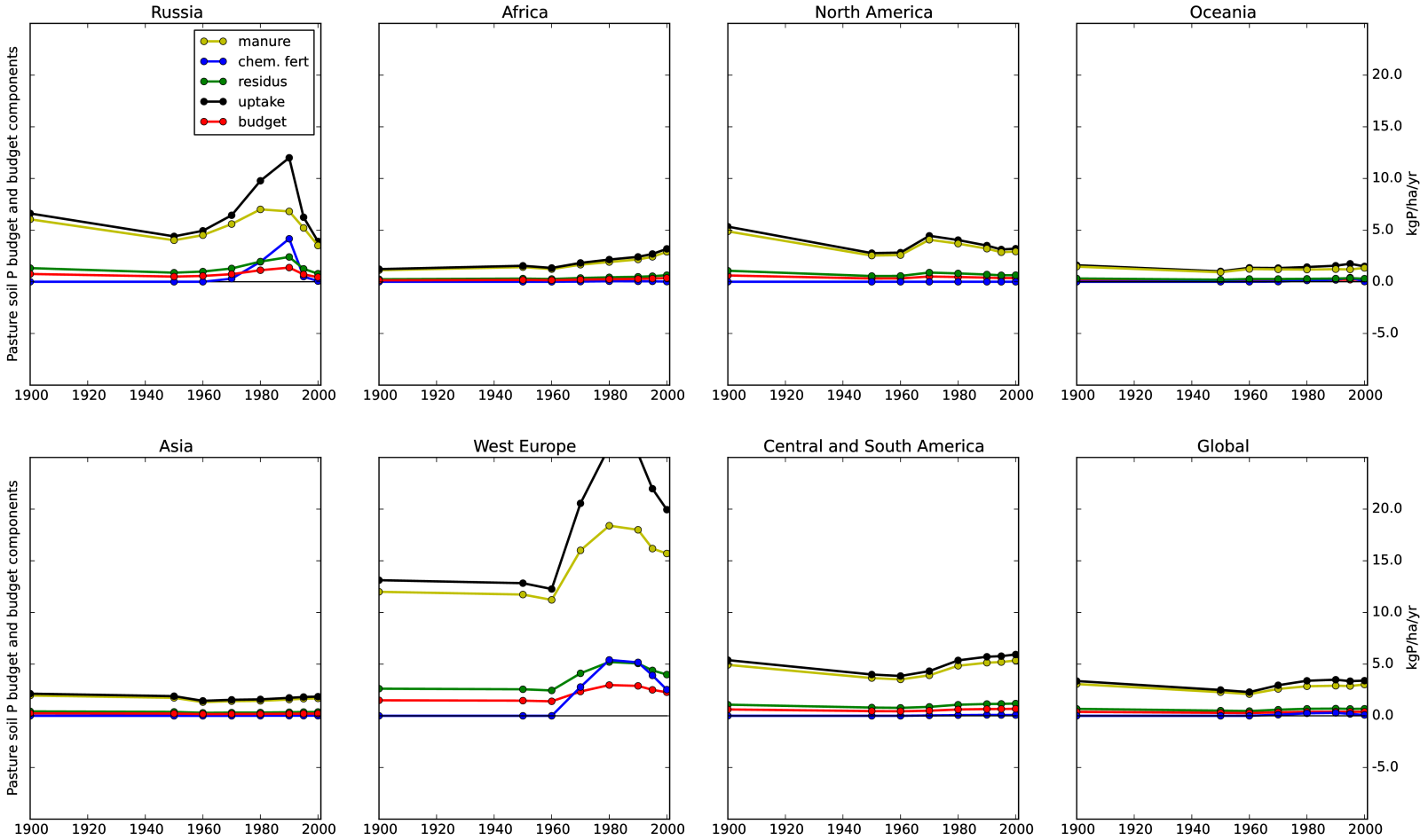

Russia
Africa
N America
Oceania
Asia
W Europe
C and S America
Global
Drivers (1/2)
Soil WaterCLIM)
Soil T(CLIM)
Pasture area(LUCC)
Ks(BUFF)
Smax(BUFF)
1e-1* no unit
1e+1* degC
1e+8* km2
1e-3* ha/kgP
1e+3* kgP/ha
Drivers (2/2)
Losses(LOSS)
Total dep.(DEPO)
Farm soil budget(FARM)
Unmn tot soil P(BIOG)
1* kgP/ha/yr
1* kgP/ha/yr
1e+1* kgP/ha
1e+3* kgP/ha
P_{TOT}
CTRL
CTRL;1900 g-cells
LUCC=1900
LUCC=1900 & FARM=0
1* kgP/ha
P_{LAB}
CTRL
CTRL;1900 g-cells
LUCC=1900
LUCC=1900 & FARM=0
1* kgP/ha
1900
1950
2000

Figure S13


Russia
Africa
North America
Oceania
Asia
West Europe
Central and South America
Global
manure
chem. fert
residus
uptake
budget
Pasture soil P budget and budget components
kgP/ha/yr
1900
1920
1940
1960
1980
2000
20.0
15.0
10.0
5.0
0.0
-5.0