Reviewer 3

Overall comment: Batool et al. have conducted a comprehensive analysis of the phosphorus budget in Europe, compiling a valuable dataset by integrating various assumptions and parameters from a range of publications. The inclusion of multiple sources of phosphorus inputs and outputs from pasture and non-agriculture provides a broader perspective on the phosphorus budget within terrestrial ecosystems. However, several key assumptions—particularly those related to time series reconstruction and spatial allocation—raise concerns, as they diminish the dataset's robustness and weaken the overall conclusions. Below are my specific comments for further improvement:

Reply: We thank the Reviewer for their useful comments to the manuscript. We reply below to the points raised by the Reviewer. We have carefully revised the manuscript, addressing the reviewers' concerns and suggestions, and we hope it now meets their expectations.

1.1 —

- Although the datasets have been allocated to gridded maps, they are primarily based on national-level aggregates. Caution should be exercised when claiming that this dataset offers high spatial resolution.

  Reply:

  Thank you for your insightful comment regarding the spatial resolution of our dataset. We acknowledge that while our dataset has been downscaled to a high-resolution grid, much of the underlying data, especially for inputs like fertilizer use and manure application, are indeed based on national-level aggregates. We would also like to emphasis here that we also used several other sources of spatially refined datas in our dowscaling methodology, such as different crop types, HYDE database on agriculture land development, the Global Land Cover database, among others. While we acknowledge the reviewer's point, we believe the reviewer would agree that many existing nutrient databases (Zhang et al. (2017); Lu and Tian (2017); Xu et al. (2019)) are based on national estimates in one way or another, given the long-term consistency and availability of data at this scale.

  In response to this point, we have clarified in the revised manuscript that although the dataset offers gridded outputs, users should exercise caution when interpreting these results as high-resolution spatially explicit data. The spatial patterns are influenced by the downscaling of national-level data, and local variations might not be fully captured. We have adjusted the language in the section 5 of the revised manuscript to reflect this more clearly, emphasizing that the dataset is suitable for basin-level, aggregated regional, national and continental-scale assessments but may not be as reliable for detailed local-scale analysis without further refinement of underlying data. Additionally, we have revised the discussion to address the limitations in the spatial allocation process explicitly, particularly with respect to the uncertainties that arise from using national aggregates for gridded outputs. The added text at line 861 under section 4 reads as follows:

  Second, while the dataset has been downscaled to a high-resolution grid, several underlying inputs, such as fertilizer use and manure application, originate from country-level aggregates. This approach limits the ability of gridded outputs to capture local variability in P surplus. Therefore, while the dataset is valuable at coarser basin, regional, national and continental scales, caution should be exercised for high-resolution applications. We recommend using the dataset at aggregate levels, such as countries, European socio-economic regions (e.g., NUTS levels), or river basin scales (see Section "Spatio-temporal variation in P surplus, P inputs and P outputs" and Figure 4) to support land and water management activities.

We would also like to emphasis here that the grid resolution of our dataset provides users with the flexibility to aggregate data at different spatial scales depending on their research needs. This allows the dataset to be used at different spatial scales, with the most robust applications being at a larger spatial scale. This point emphasizes the potential use of our dataset and has been added under section 4 in line 894 of the revised manuscript as follows:

Furthermore, our gridded dataset supports a greater degree of flexibility for aggregating data at various spatial scales, such as national, regional, and river basin levels, which is helpful in analyzing trans-boundary nutrient flows across Europe. This flexibility allows the dataset to address the needs of cross-regional and trans boundary applications, including in major European river basins like the Elbe, Danube and Rhine, thereby facilitating joint nutrient management in shared water bodies (Müller-Karulis et al., 2024). By identifying critical regions of high P surplus or its components (P fertilizer/manure), users can pinpoint locations where nutrient management improvements could have the greatest environmental and economic impact (Malagó and Bouraoui, 2021).

By revising the text in line with the comment of the reviewer, we hope that this will provide readers with a better understanding of the dataset's strengths and potential limitations for high-resolution applications and its potential use. We appreciate the feedback, as it will help to ensure that users of the dataset understand the appropriate context for its use.

- Pasture definition: Clarification is needed on whether "pasture" refers to both grazed and natural pasture, and what is meant by "semi-natural vegetation." The distinction between these categories could significantly affect the phosphorus budget.

Reply:

Thank you for your comment regarding the definitions of "pasture" and "semi-natural vegetation".For clarification, we have based these definitions on established sources, as described below:

- **Pasture Definition:** The definition of pasture in our study aligns with the FAO Land Use classification, which defines pasture area under permanent meadow and pasture as *"land used permanently (five years or more) to grow herbaceous forage crops, either cultivated or naturally occurring (e.g., wild prairie or grazing land)"* (FAOSTAT, 2021). This is the definition used in our dataset of pasture area, ensuring consistency with widely accepted land-use classifications and comprehensively accounting for land designated for long-term herbaceous forage, whether actively grazed or in a natural state.
- **Semi-Natural Vegetation Definition:** The term "semi-natural vegetation" is based on the definition from Land Cover Classification System (LCCS), which describes it as *"vegetation not planted by humans but influenced by human actions, such as grazing or selective logging, which alters the floristic composition"* (Di Gregorio, 2005). To spatially delineate semi-natural vegetation, we use the Global Land Cover (GLC) dataset (Bartholomé and Belward, 2005), which offers a high-resolution (300 m) classification of different land-cover classes. Following the GLC database, in our analysis, semi-natural vegetation includes:
  - Naturally occurring tree, shrub, and herbaceous cover
  - Shrubland (including both evergreen and deciduous types)
  - Lichens and mosses
  - Sparse vegetation (such as sparse tree, shrub, and herbaceous cover)
  - Vegetation in flooded areas (including shrub or herbaceous cover in freshwater, saline, or brackish water zones)

By following the FAO definition and utilizing the GLC dataset, we capture the range of vegetation types that fall under semi-natural land. To make this aspect clear, we have revised the definition of pasture in section 2.2.1 (line 156) which now reads as follows:

Pasture area is the land under permanent meadow and pasture and is defined as land used permanently (five years or more) to grow herbaceous forage crops, either cultivated or naturally occurring (e.g., wild prairie or grazing land) (FAOSTAT, 2021).

Similarly, we have added definition of semi-natural vegetation in section 2.2.2 (line 206) which now read as follows:

We used the classification of land cover categories from global land cover (GLC) (Bartholomé and Belward, 2005) that is available at a spatial resolution of 300 m. GLC includes 23 land cover classes that we grouped into 5 categories namely, cropland, semi-natural-vegetation (i.e. vegetation not planted by humans but influenced by human actions (Di Gregorio, 2005) including tree, shrub-land, herbaceous cover, Lichen and mosses), forest (broad-leaved, evergreen and deciduous forest), non-vegetation (bare areas, water bodies) and urban area. The proportions of these categories were then applied to the non-agricultural area to estimate their annual development from 1850 to 2019.

– Section 2.2.1: The authors mention using HYDE to derive temporal variation in cropland and pasture. It remains unclear whether this variation was applied on a grid-by-grid basis. If it was applied to grids, how did the authors "maintain the spatial distribution from Ramankutty et al. (2008) while accounting for annual temporal changes from HYDE"? Alternatively, if the country-level annual variation was used by aggregating grids within each country, it seems redundant, as FAOSTAT already rescaled the grids. Clarification is needed.

Reply:

Thank you for this insightful comment regarding the use of HYDE and Ramankutty et al. datasets for cropland and pasture reconstruction. For clarification, we first utilized the spatial distribution of cropland and pasture areas from Ramankutty et al. (2008), which are available at a spatial resolution of 5 arcmin for the year around 2000. This static spatial dataset serves as the base spatial distribution of agricultural areas. To derive the temporal variability in the gridded values of cropland and pasture area over the period 1850–2019, we employed the HYDE dataset, which provides decadal (and post-2000, annual) grid-level cropland and pasture areas. We generated the temporal variability in cropland and pasture with respect to the year 2000 area from the HYDE dataset for each year during 1850 – 2019 and each grid cell. We then applied these gridded normalised (temporal) values of cropland and pasture area to the respective land use area of Ramankutty et al. (2008) in 2000 at grid-level. Then, we harmonized the resulting grid-level cropland and pasture areas with FAOSTAT's country-level data for 1961–2019 to ensure consistency. This harmonization was achieved by calculating country-level adjustment ratios between FAOSTAT's reported areas and our reconstructed areas. Then, we applied these calculated ratio to our gridded estimates of cropland and pasture areas to ensure that the FAOSTAT country totals are maintained for the period 1961 – 2019.

We understand that this aspect may not have been clear in the initial version of the manuscript, as we referred to our previous study by Batool et al. (2022), where the land-cover-related reconstruction was presented in detail. To ensure the current paper stands alone in terms of necessary information, we have provided additional clarification in the section 2.2.1 of the revised manuscript and included equations illustrating this process. This will help readers understand how we integrated different databases to create the temporal and spatial dynamics of cropland and pasture areas. This revision now reads as follow:

Cropland is defined as land used for the cultivation of crops, including arable crops and land under permanent crops (Ramankutty et al., 2008; FAOSTAT, 2021). Pasture area is the land under permanent meadow and pasture and is defined as land used permanently (five years or more) to grow herbaceous forage crops, either cultivated or naturally occurring (e.g., wild prairie or grazing land) (FAOSTAT, 2021). To represent the spatial distribution of cropland and pasture areas, we utilized the dataset from Ramankutty et al. (2008), which provides gridded estimates at a 5-arcminute resolution for the year 2000. These gridded values serve as the baseline for cropland and pasture area in our analysis. To account for temporal changes in cropland and pasture areas, we used data from the History Database of the Global Environment (HYDE version 3.2) (Goldewijk et al., 2017). HYDE provides global decadal estimates of cropland and pasture areas from 1700 to 2000, as well as annual values from 2000 to 2017. We *generated annual time series of cropland and pasture areas for the period 1850–2019 using linear interpolation for the decadal estimates. For the years 2018 and 2019, we used the same values as 2017 due to a lack of available data.*

To combine the data from Ramankutty et al. (2008) and from HYDE, we first calculated temporal ratios for the HYDE data for each grid cell using the year 2000 as the reference year. These ratios represent the relative change in cropland $R_{\textit{HYDE-cr}}$ (-) and pasture area $R_{\textit{HYDE-past}}$ (-) over time, normalized to the year 2000:

$$R_{\text{HYDE-cr}}(i, y_{1850,-,2019}) = \frac{A_{\text{HYDE-cr}}(i, y_{1850,-,2019})}{A_{\text{HYDE-cr}}(i, y_{2000})} \tag{R1}$$

$$R_{\text{HYDE-past}}(i, y_{1850,-,2019}) = \frac{A_{\text{HYDE-past}}(i, y_{1850,-,2019})}{A_{\text{HYDE-past}}(i, y_{2000})} \tag{R2}$$

Where $A_{\textit{HYDE-cr}}(ha))$ and $A_{\textit{HYDE-past}}(ha))$ are the gridded cropland and pasture areas, respectively.

Next, we applied these normalized ratios to the baseline gridded values from Ramankutty et al. (2008) to derive annual cropland and pasture areas for each grid cell, as follows:

$$A_{\text{cr}}(i, y_{1850,-,2019}) = A_{\text{Ramankutty-cr}}(i, y_{2000}) \times R_{\text{HYDE-cr}}(i, y_{1850,-,2019}) \tag{R3}$$

$$A_{\text{past}}(i, y_{1850,-,2019}) = A_{\text{Ramankutty-past}}(i, y_{2000}) \times R_{\text{HYDE-past}}(i, y_{1850,-,2019}) \tag{R4}$$

Where $A_{\textit{Ramankutty-cr}}(ha)$ and $A_{\textit{Ramankutty-past}}(ha)$ are the gridded cropland and pasture areas from Ramankutty et al. (2008) for the year 2000, and $A_{\textit{cr}}(ha)$ and $A_{\textit{past}}(ha)$ are the estimated cropland and pasture areas.

We harmonized our reconstructed cropland and pasture areas with FAOSTAT data available at country-level, which provides consistent information from 1961–2019. To do so, we calculated country-level ratios for cropland and pasture areas by comparing FAOSTAT data with the sum of our gridded estimates for each country. The ratios were calculated as follows:

$$R_{A_{\text{cr}}}(u, y_{1961,-,2019}) = \frac{A_{\text{FAO-cr}}(u, y_{1961,-,2019})}{\sum_{i=1}^{n_u} A_{\text{cr}}(i, y_{1961,-,2019})} \tag{R5}$$

$$R_{A_{\text{past}}}(u, y_{1961,-,2019}) = \frac{A_{\text{FAO-past}}(u, y_{1961,-,2019})}{\sum_{i=1}^{n_u} A_{\text{past}}(i, y_{1961,-,2019})} \tag{R6}$$

Whereas $R_{A_{cr}}(-)$ is the country-level ratio of cropland area, $A_{\textit{FAO-cr}}(ha)$ represents the country-level cropland area from FAOSTAT, n_u is the number of grid cells in country u, and $\sum_{i=1}^{n_u} A_{cr}(ha)$ is the sum of the gridded cropland areas in country u in year y. Similarly, $R_{A_{past}}(-)$ is the ratio of pasture area, $A_{\textit{FAO-past}}(ha)$ is the country-level pasture area from FAOSTAT, and $\sum_{i=1}^{n_u} A_{past}(ha)$ is the sum of the gridded pasture areas.

We applied these ratios to adjust our gridded estimates to match FAOSTAT's country-level data (all variables, except for ratios, are in ha):

$$A_{\text{cr}}^{\text{cor}}(i, y_{1961,-,2019}) = R_{A_{\text{cr}}}(u, y_{1961,-,2019}) \times A_{\text{cr}}(i, y_{1961,-,2019}) \tag{R7}$$

$$A_{\text{past}}^{\text{cor}}(i, y_{1961,-,2019}) = R_{A_{\text{past}}}(u, y_{1961,-,2019}) \times A_{\text{past}}(i, y_{1961,-,2019}) \tag{R8}$$

Whereas A_{cr}^{cor} represents the corrected gridded cropland, $R_{A_{cr}}$ is the country-level ratio of cropland area as given in equation R5, and A_{cr} is the original gridded cropland area as derived in equation R3. Similarly, A_{past}^{cor} represents the corrected gridded pasture area, $R_{A_{past}}$ is the country-level ratio of pasture area as shown in equation R6, and A_{past} is the original gridded pasture area as derived in equation R4.

A_{cr}^{cor} and A_{past}^{cor} represent the corrected gridded cropland and pasture areas, respectively. These corrections are based on country-level ratios—$R_{A_{cr}}$ for cropland and $R_{A_{past}}$ for pasture—calculated using equations R5 and R6. The original gridded areas, A_{cr} and A_{past}, are adjusted according to these ratios using equations R3 and R4.

For years prior to 1961, we used the same ratios as of 1961 to maintain consistency. In cases where FAOSTAT data were not available before 1992 (e.g., for Estonia, Croatia, Lithuania, Latvia, and Slovenia), we used the ratios from the year 1992 for the period 1850–1991. For countries like Luxembourg and Belgium, and Slovakia and Czech Republic, which were reported as single entities in historical records, we used combined ratios for the respective periods. Finally, the total agricultural area A_{agri}^{cor} (ha) for each grid cell was calculated by summing the corrected cropland A_{cr}^{cor} and pasture areas A_{past}^{cor} (all variables are in ha):

$$A_{\text{agri}}^{\text{cor}}(i, y_{1850,-,2019}) = A_{\text{cr}}^{\text{cor}}(i, y_{1850,-,2019}) + A_{\text{past}}^{\text{cor}}(i, y_{1850,-,2019}) \tag{R9}$$

We ensured physical consistency by checking that the agricultural area in each grid cell did not exceed the total physical area of the grid cell. In rare cases where this condition was violated due to inconsistencies in data sources (e.g., differences between FAOSTAT (FAOSTAT, 2021), HYDE (Goldewijk et al., 2017), and Ramankutty et al. (2008)), we redistributed the excess agricultural area to neighboring grid cells.

We hope that the revised manuscript presents the clarified our reconstruction of gridded cropland dynamics.

– Section 2.2.3: The authors refer to 17 non-fodder and 6 fodder crops. Do these cover all cropland? The area of these crops might be overestimated after harmonization with FAOSTAT if they do not cover all cropland. Also, how was the temporal dynamic of cropland area applied to Monfreda et al.—on a grid-by-grid or country level? Furthermore, how was the crop-specific area time series harmonized with FAOSTAT data on the map as the total cropland has been harmonized in section 2.21? Does each grid represent only one crop, or multiple crops?

Reply:

Thank you for your detailed comments. As we stated above, we understand that this aspect may not have been clear in the initial version of the manuscript, as we referred therein to our previous study by Batool et al. (2022), where the land-cover-related reconstruction was presented in detail. To ensure the current paper stands alone in terms of necessary information, we have made additional changes to the revised manuscript.

In the following, we have first clarified the points concerning the coverage of cropland by the 17 non-forage and 6 forage crops, the harmonization with FAOSTAT data and the handling of grid-specific cropland areas. We then detailed the changes made for the revised manuscript.

- **Coverage of all cropland by selected crops:** We obtained gridded crop specific harvested areas from Monfreda et al. (2008) which is available for 175 different crop types. Among these 175 crops, we selected 17 major crops for which fertilizer application rates are provided Heffer et al. (2017) and that have large production across Europe. These selected crops cover most of the cropland across Europe. However, minor crops not included within the 17 selected crops were excluded due to the lack of long-term, consistent data. This exclusion may result in some discrepancies in the overall coverage of certain crop types – a limitation we acknowledge in the manuscript.
- **Temporal dynamics applied at the grid level:** We used the cropland dataset from Ramankutty et al. (2008) available at a spatial resolution of 5 arcmin for the year around 2000 to get the spatial variability at gridded level. Then, we derived the temporal variability of cropland area at grid-level from the HYDE dataset for each year during 1850 – 2019, by referencing it to the year 2000, as in equation 14 of the revised manuscript. We then applied these normalized (temporal) gridded values of cropland area to the gridded cropland area of Ramankutty et al. (2008) in 2000, as in equation 16 of the revised manuscript. This approach ensured that each grid cell's cropland area of Ramankutty et al. (2008) are adjusted consistently over time with HYDE's cropland grid-level temporal dynamics.

– **Harmonization of crop-specific area time series with FAOSTAT:** The crop-specific harvested areas were harmonized with FAOSTAT country-level data using a ratio-based approach. The ratio was derived between FAOSTAT country-level data and the sum of our gridded estimates of crop-specific harvested area. This ratio was then applied to adjust the gridded estimates of crop-specific harvested areas for each grid cell, ensuring harmonization with FAOSTAT data. For years before 1961, we applied the 1961 ratios to maintain consistency in the estimates.

– **Crop representation in each grid:** Each grid cell represents multiple crops, not just a single crop. We used the crop distribution from Monfreda et al. (2008) to allocate the relative proportions of different crops within each grid cell. The proportional distribution of crops was adjusted over time using the temporal dynamics of cropland area, ensuring that the evolution of crop harvested areas in each grid cell is consistent with changes in total cropland.

To clarify these points, we have revised section 2.2.1 for the points related to cropland areas (the revision is mentioned in the reviewer's response above) and section 2.2.3 on crop-specific harvested area. The revised section 2.2.1 reads as follows:

*Reconstruction of crop-specific harvested area: We acquired gridded crop-specific harvested areas from Monfreda et al. (2008) for 175 different crops representing the year 2000. Among these, we selected 17 major non-fodder crops for which mineral fertilizer application rates are available (Heffer et al., 2017) and which are widely grown across Europe, as well as six fodder crop categories. Below we provide a more detailed overview on the selected crops (see also Table **??**). These selected crops cover most of the cropland across Europe. The harmonization process ensures that the total cropland area aligns with FAOSTAT estimates.*

To generate annual time series of crop-specific harvested areas, we applied the temporal dynamics of cropland areas, adjusting the spatial distribution of crops based on the Monfreda et al. (2008) dataset, while referencing FAOSTAT's country-level data to ensure consistency over time. The crop-specific harvested areas $A_{\textit{crops}}$ *(ha) were harmonized with FAOSTAT data* $A_{crops_{FAO}}$ *(ha) using a ratio-based approach. The ratio* R_A *(-) between FAOSTAT country-level data and the sum of gridded estimates was calculated as follows:*

$$R_A(u,c,y) = \frac{A_{crops_{FAO}}(u,c,y)}{\sum_{i=1}^{n_u} A_{\text{crops}}(i,c,y)} \tag{R10}$$

This ratio was then applied to adjust the gridded estimates of crop-specific harvested areas for each grid cell, ensuring harmonization with FAOSTAT data:

$$A^{\text{cor}}_{\text{crops}}(i,c,y) = A_{\text{crops}}(i,c,y) \times R_A(u,c,y) \tag{R11}$$

Where $A^{\textit{cor}}_{\textit{crops}}$ *is the corrected crop-specific harvested areas for grid cell* i*, crop* c*, and year* y*.*

For years prior to 1961, we applied the ratio from 1961 to maintain consistency across all years:

$$A^{\text{cor}}_{\text{crops}} = A_{\text{crops}}(i,c,y_{1850,-,1960}) \times R_A(u,c,y_{1961}) \tag{R12}$$

This method ensured that the crop-specific harvested areas were harmonized with FAOSTAT country-level data, with each grid representing multiple crops.

For fodder crops, we utilized country-level data from Einarsson et al. (2021), available from 1961 to 2019 for 26 European countries. This dataset includes six fodder crop categories, namely: temporary grassland, lucerne, other leguminous plants, green maize, root crops (forage beet, turnip, etc.), and other fodder plants harvested from cropland. For the period 1850–1960, we applied the temporal dynamics of reconstructed cropland areas to estimate fodder crop areas. For countries with missing data, we filled gaps by extrapolating ratios from neighboring countries with similar climatic and geographical conditions or using aggregated ratios from comparable regions.

- Section 2.3.3 and 2.3.8: The refinement of fertilizer application in the second approach requires further explanation. Did the authors multiply the rate by the percentage of treated area? Please clarify if cropland was only partially fertilized/manured, while pasture was 100% fertilized/manured in the second approach. I do not think these two fertilizer and manure datasets are two independent datasets. The one without considering the percentage of treated area is a biased estimate since it does not account this factor. Additionally, as the authors considered the percentage of treated area, the cropland (grid) that receives fertilizer would have greater surplus than the other cropland. Have the authors considered to allocate the fertilizer/manure only to those treated area?

Reply: Thank you for your detailed feedback on the allocation of fertilizer and manure in our study. We appreciate your observations and would like to clarify a few aspects regarding the distribution of these inputs to croplands and pastures.

- **Fertilizer Application (Section 2.3.3) :** First of all, we would like to clarify that we considered that fertilizer is applied to 100% of the cropland and pasture, since we did not have more detailed data to determine the spatial variability of fertilizer application rates within a given country. We have clarified this in the revised manuscript in line 271 in section 2.3.3. In addition, in our study, the term **"treated"** is only applied to manure inputs and refers to the share of manure processed through specific management systems (e.g., lagoons, slurry, solid storage) as per FAOSTAT (FAOSTAT, 2023) and the Intergovernmental Panel on Climate Change (IPCC) guidelines (Dong et al., 2020). This concept of treated vs. non-treated does not apply to fertilizers, which are distributed to cropland and pasture areas based on two approaches to account fo underlying uncertainty in application rates. These approaches are elaborated below:
  - **First approach:** In this method, we determined country-specific fertilizer application rates for various crops and grassland using data from the International Fertilizer Industry Association (IFA; https://www.ifastat.org), combined with FAOSTAT's cropland and grassland area statistics. To capture spatial variations, we applied these rates for different crops and pasture to gridded respective areas over the period from 1850 to 2019. This approach provided annual fertilizer application amounts for each crop type (non-fodder and fodder), pastures, and the overall total for each grid cell. Next, the fertilizer application totals were adjusted to ensure consistency with the FAO country-level fertilizer amounts applied to soil during 1850−2019.
  - **Second approach:** To refine the distribution of fertilizer, we considered country-specific data that provides proportion of fertilizer applied to cropland and pasture areas, as reported by Ludemann et al. (2023). This approach follows national-level statistics, which indicate that in some countries, not all fertilizer is applied to croplands and pastures. For instance, a majority of the European countries apply 100% of their fertilizer to croplands, while the proportion differs for a few European countries (e.g., 90% for Austria, Finland, France, Germany, the Netherlands, and Poland)
- **Manure Application (Section 2.3.8):** The term **"treated manure"** refers to the manure processed through different manure management systems (MMS), such as lagoon or slurry storage, based on FAOSTAT and IPCC guidelines. It represents manure that is applied to soils after treatment, excluding manure left on pastures or used as fuel. We would like to clarify that we assume that all of the manure treated within a given grid cell is applied to soils within that grid cell. Therefore, all cropland and pasture areas located in grid cells where the manure treated is not equal to zero receive manure. We used three methodologies for manure distribution between cropland and pasture within each grid cell:
  - **Equal distribution:** Manure is distributed evenly between cropland and pasture within each grid cell, following Xu et al. (2019).
  - **Country-specific proportions:** Using country-level data from Ludemann et al. (2023), we applied national ratios to adjust manure distribution. Some countries apply nearly all manure to croplands (e.g., 100% for several European countries), while others allocate a portion to pastures.

- **Time-varying data:** We also used dynamic country-specific data from Einarsson et al. (2021) to assign manure based on actual practices in each country over time. This method captures the evolving practices of manure application across croplands and pastures, accounting for historical changes.

By employing two distinct approaches for fertilizer and three for manure allocation, we provide a comprehensive representation of uncertainties, allowing users to compare, for example, the simplistic uniform distribution to a more refined national data scenarios. We have also revised the manuscript in line 427 of Section 2.3.7 to clarify that the term "treated" refers specifically to manure-related datasets as per FAOSTAT (FAOSTAT, 2023) and IPCC guidelines (Dong et al., 2020), and not to other aspects (e.g., fertilizers, croplands, or pasture areas), in order to avoid potential confusion, which reads as follows:

Specifically, from FAOSTAT, we used the 'Treated manure N' estimates, which represent the quantity of manure processed through specific manure management systems (e.g., lagoons, slurry, solid storage) prior to N loss in these systems (FAOSTAT, 2023). Since P losses in these systems are minimal (FAOSTAT, 2023), we considered that the entire amount of treated P manure is applied to soil. It is important to clarify that in this context, the term 'Treated' refers exclusively to manure management and does not extend to fertilizers, which are directly distributed to cropland and pasture areas without similar classification.

1.2 — Concerns regarding time-series reconstruction and spatial allocation:

The reconstruction of time-series spatial maps presents several issues, particularly when relying on a single reference year for spatial distribution. This approach is problematic for periods before 1961 due to a lack of country-level control data. I recommend trimming the study period to 1961–2019, as the 1850–1960 period is based on unsupported assumptions. The extrapolation lacks the necessary historical data and should be omitted unless stronger justifications can be provided.

Reply:

We appreciate the reviewer's concern regarding the reliability of the time-series reconstruction for the period 1850–1960. While we acknowledge the limitations of available data for this early period, we believe that retaining the full dataset from 1850 to 2019 is helpful for several reasons:

- **Use of Historical and Proxy Data**: Although pre-1961 data are limited, we have employed reliable sources (e.g., Cordell et al. (2009), Bayliss-Smith and Wanmali (1984)) to infer reasonable temporal dynamics. Specifically, we used global phosphate rock production data (Cordell et al., 2009) to estimate P fertilizer inputs and historical wheat yield trends (Bayliss-Smith and Wanmali, 1984) as a proxy for other crops. While detailed, spatially explicit data for this period are unavailable, these proxies provide a robust basis for capturing temporal trends and aligning with established historical patterns. We acknowledge the uncertainties inherent in using proxy data but emphasize that these methods offer the most reliable approach for reconstructing historical patterns given the limitations of existing dataset.
- **Historical Context and Long-term assessment**: Including the period from 1850 allows us to capture pivotal shifts in agricultural practices, land use, and industrial development that directly influenced phosphorus (P) dynamics prior to the Green Revolution (Guejjoud et al., 2023). These early changes laid the groundwork for modern nutrient management (Pratt and El Hanandeh, 2023; Ringeval et al., 2024; Sharpley et al., 2013). If we were to omit the 1850–1960 period, we would lose critical insights into the pre-industrial and early industrial phases of agricultural intensification, which are essential for understanding how historical P inputs have shaped current nutrient surpluses – albeit at a crude level, it is still useful for analyzing and understanding long-term developments at regional scales.

- **Phosphorus Legacy and Policy Implications**: P applied during historical periods continues to influence present-day P dynamics due to legacy effects (McCrackin et al., 2018; Guejjoud et al., 2023; Sharpley et al., 2013). By excluding historical database, we would risk overlooking the long-term environmental consequences of historical P accumulation leading to eutrophication and nutrient runoff, which remain critical issues today. Understanding the legacy of early P inputs is essential for designing policies aimed at mitigating both historical and current P surpluses (McCrackin et al., 2018; Sharpley et al., 2013).

- **Cross-Disciplinary Relevance**: Retaining the period from 1850–1960 enhances the dataset's utility for cross-disciplinary studies, including historical agriculture, environmental change, and biogeochemical cycling. Research into the historical impacts of land use, industrialization, and nutrient legacies often requires long-term datasets. As stated above, while the dataset prior to 1961 represents crude estimates, we believe it will still useful for analyzing and understanding long-term developments at regional scales.

While we recognize the uncertainty associated with the early period, we emphasize that the historical trends from 1850 onward provide valuable context for understanding long-term P dynamics. Nevertheless, we understand the reviewer's concern and have issued a cautionary note accordingly. In section 4 of the revised manuscript, we have made it clear that the 1850–1960 period is based on a combination of proxy data and extrapolations, and we recommend that these early estimates be interpreted with caution. We have balanced the text by detailing cautionary notes (at line 849) and the utility on the use our datasets (at line 910) as follows:

First, the reconstruction of P surplus before 1961 is constrained by limited historical data. For the period 1850–1960, we relied on proxy information and extrapolations based on data from 1961 onward, which inherently introduces higher uncertainty. For instance, national wheat production trends were used to estimate other crop productions, but this method may not fully capture the variations of each specific crop types. We assume that the relative values of the fertilizer application rates taken from the International Fertilizer Association (IFA) (Heffer et al., 2017) for 2014–2015 remained constant for the period 1850–2019, and pre-1961 temporal variations were inferred from global phosphate rock production trends. Livestock distributions were based on GLW3 (Gilbert et al., 2018) data circa 2010 to estimate manure production for the entire 1850–2019 period, and simplifying assumptions were made before 1961, since no country-level manure data were available. Such simplifications may not accurately reflect historical livestock numbers or distribution patterns, influencing P surplus estimates. Furthermore, spatial datasets, especially for land use and crop production, are more detailed and reliable from the mid-1990s onward, making the P surplus estimates more robust for recent decades. Thus, while historical estimates provide general trend insights, recent data (from the mid-1990s) offer greater reliability.

Although pre-1961 data are limited, we have employed reliable sources (e.g., Cordell et al. (2009), Bayliss-Smith and Wanmali (1984)) to infer reasonable temporal dynamics. By covering the period from 1850, our dataset captures pivotal shifts in agricultural practices, land use, and industrial development that directly influenced phosphorus (P) dynamics prior to the Green Revolution (Guejjoud et al., 2023). These early changes laid the groundwork for modern nutrient management (Pratt and El Hanandeh, 2023; Ringeval et al., 2024; Sharpley et al., 2013), making the dataset useful not only for current policy analysis but also as a historical baseline for exploring how shifts in climate and agricultural practices affect nutrient cycles over time. Coupled with our nitrogen (N) surplus dataset (Batool et al., 2022), this dataset enables integrated nutrient studies, facilitating the development of comprehensive management strategies that support both P and N sustainability goals. Additionally, the dataset's detailed historical record could support climate adaptation studies, enabling stakeholders to examine how nutrient budgets respond to evolving climate conditions and assess the long-term sustainability of various agricultural practices under changing environmental conditions.

1.3 — Additionally, a limitation section addressing these issues is highly recommended. Cropland and pasture: The authors used cropland and pasture distributions circa 2000 from Ramankutty et al. for 1850-2019. There is no country-level data control before 1960.Non-agriculture: The ratios of these non-agricultural area in each grid cell from GCL circa 2000 were used for 1850-2019, again with no supporting data before 1960. Crop-specific harvest area: the distribution of specific crops from Monfreda et al circa 2000 was used for 1850-2019. There is a lack of country-level data control before 1960.Fertilizer: Crop-specific fertilizer use was derived from IFA circa 2014-2015 and was rescaled throughout 1961-2019. The temporal trend before 1961 was based on global fertilizer production. There is a no country-level data control before 1960. Manure: The animal distribution was based on GLW3 circa 2010 for 1850-2019. Crop production: The annual trend of production for all other crops was based on wheat production, which is not reasonable. P removal by pasture: The P removal was calculated by multiplying 0.6 (or even using NUE) with P input. The removal of P is more likely influenced by herd size and grazing frequency rather than the P input.

Reply:

Thank you for the valuable feedback on the potential and limitations of the dataset. We have carefully considered each of the points raised by the reviewer and addressed them in the revised manuscript. Additionally, we have incorporated a new section that highlights both the limitations and future avenues for improving the dataset. Below, we provide responses to the specific points raised:

- **Cropland and Pasture (1850-2019)**: We acknowledge the limitation in using (Ramankutty et al., 2008)'s cropland and pasture distribution circa 2000 for the entire period from 1850 to 2019, especially in the absence of pre-1960 country-level data. This assumption introduces uncertainty in the spatial distribution of agricultural land before 1960. In the limitation section of the manuscript, we now emphasize that the pre-1960 estimates should be interpreted with caution due to the lack of direct historical land-use data. We also suggest future efforts focus on integrating more granular historical data sources to refine the estimates for earlier periods.
- **Non-Agricultural Areas (1850-2019)**: Similar to the cropland and pasture data, the use of GLC (Global Land Cover) ratios circa 2000 for non-agricultural areas introduces limitations, especially before 1960. We recognize that non-agricultural areas, such as forests and urban zones, likely underwent significant changes over the historical period that are not fully captured in our dataset. This limitation has been highlighted in the manuscript, and we propose that future work consider improved historical land-use reconstructions.
- **Crop-Specific Harvest Areas (1850-2019)**: The use of Monfreda et al.'s crop-specific harvest areas circa 2000 for the entire period similarly introduces uncertainty before 1960. We acknowledge that this method does not account for shifts in crop distributions and varieties over time, which could affect the accuracy of P surplus estimates. This limitation has been explicitly mentioned in the revised manuscript.
- **Fertilizer Use (1850-2019)**: The derivation of crop-specific fertilizer use from IFA data for 2014-2015, rescaled to cover 1850-1960, introduces uncertainty for the period prior to 1961 when global dynamics of phosphate rock production were used. While we agree that there is no country-level data control before 1960, we believe that using global phosphate rock production trends is a reasonable approximation given the lack of alternative data. Nevertheless, we have added this as a limitation and encourage future work to focus on refining early-period fertilizer estimates with more historical data.
- **Manure Production (1850-2019)**: The use of livestock distribution from GLW3 circa 2010 to estimate manure production for the entire 1850-2019 period poses challenges, particularly before 1960 when detailed livestock data are sparse. This simplification may not accurately reflect historical herd sizes or distribution patterns, which could influence P surplus estimates. We have explicitly noted this in the limitations section and recommend future studies incorporate more detailed historical livestock data where available.
- **Crop Production (1850-1960)**: We acknowledge that using wheat production as a proxy for other crops may not fully capture the nuances of different crop production systems. However, in the absence of detailed crop-specific

production data before 1961, this approach provides a useful, albeit simplified, estimation. To assess the applicability of this method, we analyzed the temporal alignment between wheat production and other crop categories for the period 1961-2019 using scatter plots and correlation coefficients for the EU28 region. These analyses, presented in Figure R1 (Supplementary Figure S2), showed reasonable correlations for most crops, supporting the use of wheat production dynamics as a proxy during the reconstruction period. Nonetheless, variations in correlation strength among crops highlight the potential for refinement by incorporating additional crop-specific data where available. We have highlighted this limitation in the manuscript and suggest that future work could improve upon this method by incorporating additional crop data where possible.

– **P Removal by Pasture**: We agree with the reviewer that P removal from pasture is influenced by factors beyond P input, such as herd size and grazing frequency. While we based our estimates on phosphorus use efficiency (PUE), we recognize that PUE is an approximation that may not fully capture these complexities. In response, we have revised the manuscript to discuss this limitation and have included it in the expanded limitations section (refers to section 4 of the revised manuscript). Future work should aim to incorporate more dynamic models that account for herd size, grazing intensity, and climatic factors to improve P removal estimates.

In response to the reviewer's suggestion, we have added a dedicated section —*Potential use and limitations of the dataset* in the revised manuscript. This section thoroughly explains these issues by clearly outlining the assumptions and uncertainties related to land use, crop production, fertilizer, and manure data. As stated above, we believe that retaining the historical period (1850-1960) provides valuable insights into long-term P surplus trends, but we have issued a cautionary note on the use of early estimates.

To address the concern regarding the use of wheat production as a proxy, Figure R1 (Supplementary Figure S2) has been added to show the relationship for the temporal alignment between wheat production and other crop categories for the period 1961-2019. The moderate to high correlation coefficient ranging from 0.3 to more than 0.9 for different crops, supports the assumption of our methodology for using the wheat production dynamic as a proxy for other crops. Nevertheless, we also acknowledge the fact that further refinements can be done by incorporating additional crop-specific data where available. To ensure clarity, we have also included the following text in Section 2.4.2 at line 569:

The temporal alignment between wheat production and other crop categories was assessed for the EU28 region (Supplementary Figure S2) and resulting correlation coefficients were estimated. Most crops showed a reasonable positive correlation with wheat production, ranging from 0.3 to more than 0.9, indicating consistent temporal dynamics across different crop types. These results supports the use of wheat production dynamics as a proxy for other crops during the reconstruction period (1850–1960). However, variations in correlation strength among crops suggest that future refinements could benefit from incorporating additional crop-specific data where available.

1.4 — Other sources

– While the authors aimed to provide a comprehensive phosphorus budget, additional sources of phosphorus emissions, such as those from burning (both agricultural and non-agricultural) and urban phosphorus use (e.g., gardens, golf courses), should be considered. Has phosphorus from fertilizer and manure applied to urban areas or human waste been accounted for? These could be significant sources, and their omission weakens the comprehensiveness of the dataset relative to other inputs like deposition and weathering.

Reply: Thank you for your insightful feedback regarding additional sources to account within the phosphorus (P) budget. Regarding the P emissions from burning, we would like to emphasis that they are accounted for in our analysis within the atmospheric deposition component of our dataset. The underlying data explicitly include atmospheric P deposition from various sources such as mineral dust, primary biogenic aerosol particles, sea salt, natural combustion and anthropogenic combustion (e.g. agricultural residue burning, forest fires, logging fires and fossil fuel burning) (Ringeval et al., 2024). To clarify the inclusion of burning-related emissions, we have revised Section 2.3.9 (line 495) as follows:

In our study, we assessed P inputs from atmospheric deposition for different land types, including agricultural land (cropland and pastures) and non-agricultural land. To estimate P deposition for agricultural soils, we used the dataset provided by Ringeval et al. (2024) which represents global atmospheric deposition rates of P to cropland and pasture from 1900 to 2018 at a spatial resolution of 0.5 degrees. This dataset accounts for various sources, including mineral dust, primary biogenic aerosol particles, sea salt, natural combustion, and anthropogenic combustion (e.g., agricultural residue burning, forest fires, logging fires, and fossil fuel burning) (Ringeval et al., 2024).

We agree that urban P inputs, such as those from gardens, golf courses, urban fertilizer use, could be more nuanced and should be considered in a comprehensive P budget. In the current version of our dataset, we have primarily focused on P inputs from agricultural sources (e.g., fertilizer and manure) and natural processes (e.g., atmospheric deposition and chemical weathering) that are generally well-documented at large spatial scales (Panagos et al., 2022; Ringeval et al., 2024).

P inputs from urban areas especially those from human waste (i.e., sewage and wastewater) are not accounted for in our analysis, as these often represent point source inputs. Much of this waste often ends up as direct discharges rather than diffuse sources in soil. Our study focuses on characterizing major diffuse sources in the P surplus budget. In parallel, we have also developed a long-term database on nutrient inputs from point sources (urban areas), detailed in a separate study by Sarrazin et al. (2024). By focusing on diffuse P sources, our dataset complements existing datasets that address point source nutrient contributions, such as the European Pollutant Release and Transfer Register (E-PRTR) (Roberts, 2009) and the nutrient load database by Vigiak et al. (2020). Together, these datasets contribute to ongoing efforts to comprehensively understand P dynamics across terrestrial ecosystems, spanning both diffuse and point sources. Moving forward, we plan to explore the integration of different sources in future effort to characterize total P inputs to terrestrial system.

To clarify these distinctions, we have revised relevant sections of the manuscript to highlights the focus on diffuse sources as follows.

In Abstract as: *This study reconstructs and analyzes the annual long-term P surplus for both agricultural and non-agricultural soils from diffuse sources across Europe at a 5 arcmin ($\approx$ 10 km at the equator) spatial resolution from 1850 to 2019.*

In the Introduction (line 64), we have emphasized the scope of the study: *To address these limitations, we present here a database of yearly long-term P budgets, termed "P surplus" - defined as the difference between P inputs (mineral fertilizer, animal manure, atmospheric deposition and chemical weathering) and P removals (crop and pasture removals), covering both agricultural (cropland and pastures) and non-agricultural soils at a 5 arcmin (1/12°; approximately 10 km at the equator) spatial resolution from 1850 to 2019 across Europe, focusing only on diffuse sources.*

Finally, we have added the following paragraph in Section 5 (line 947) to discuss on point sources in more detail:

P inputs from urban areas especially those from human waste (i.e., sewage and wastewater) are not accounted for in our analysis. These inputs are typically classified as point sources, with much of the waste directly discharged rather than contributing to diffuse soil inputs. Our study focuses on characterizing major diffuse sources in the P surplus budget. In parallel, we have also developed a long-term database on nutrient inputs from point sources (urban areas), detailed in a separate study by Sarrazin et al. (2024). Additional datasets, such as the European Pollutant Release and Transfer Register (E-PRTR) (Roberts, 2009) and the nutrient load database by Vigiak et al. (2020), provide valuable information on nutrient contributions from urban and industrial sources across Europe. Our current dataset on diffuse sources of P complements existing datasets on point sources and contributes to ongoing efforts to comprehensively understand P dynamics across terrestrial ecosystems, spanning both diffuse *and point sources. Integrating these point sources with our existing dataset can provide a more comprehensive characterization of P inputs to receiving water bodies (e.g., rivers, lakes, wetlands, etc.) from different sources.*

1.5 — Show and publish inputs and outputs:

– The phosphorus surplus represents the balance between inputs and outputs. I recommend including the temporal and spatial changes of individual input and output categories alongside the surplus to help readers understand the drivers of these trends. Additionally, publishing the input and output datasets would be valuable for broader research use.

Reply: Thank you for the valuable suggestion. We agree that providing both temporal and spatial changes for P inputs and outputs alongside P surplus greatly enhances the interpretation of the data.

To address this comment, we have expanded the dataset to include not only P surplus but also the underlying P inputs and outputs for the entire study period. This allows for a further exploration of how specific input and output categories drive P surplus trends. These datasets are made publicly available to support further research and detailed analysis of changing P dynamics across Europe.

In the revised manuscript, we have added new visualizations to better represent these dynamics. Figure R2 (which corresponds to Figure 2 in the revised manuscript) now includes gridded maps for P inputs and outputs alongside P surplus, providing a comprehensive spatial overview. The snapshots of variations in mineral fertilizer and animal manure have been provided in Figure R4 (which corresponds to Supplementary Figure S3). Additionally, Figure R3 (which corresponds to Figure 4 in the revised manuscript) shows the decadal trajectories of agricultural and total P inputs (orange) and outputs (blue), with the P surplus represented by a red line for each decade. Figure R5 and R6 (which corresponds to Supplementary Figures S8 and S9) depict decadal P input and output trajectories for individual countries, illustrating how these factors evolve over time and contribute to P surplus trends.

In this regard, we have revised additional text in the revised manuscript at lines 597-650 in section 3.1.

The spatio-temporal variations in our P surplus, inputs, outputs at the gridded level is illustrated in Fig R2 for the selected years: 1900, 1930, 1960, 1990 and 2015 (See Supplementary Figure R4 for the corresponding variations in mineral fertilizer and animal manure). These plots show that, while Northern Europe consistently exhibits a positive P surplus with relatively stable P inputs and outputs, most of Central and Western Europe experiences variable P fluxes dynamics over time. For example, in 1900 and 1930, there are notable areas in Central and Western Europe with negative P surplus (P deficit), where P outputs exceeds P inputs, particularly in agricultural regions. As time progresses, the pattern shifts. By 1960 and 1990, the P surplus becomes more positive across these regions. During this time periods, Northern Europe continues to show a positive P surplus, with values ranging from approximately 0 to 4 $kg\ ha^{-1}$ of physical area yr^{-1}, and with a balanced P inputs and outputs between 2–4 $kg\ ha^{-1}$ of physical area yr^{-1}. Conversely, the mid-latitude areas, particularly in Central and Western Europe, exhibit higher P surplus and inputs, from 10 to over 18 $kg\ ha^{-1}$ of physical area yr^{-1}, with moderate outputs (4–14 $kg\ ha^{-1}$ of physical area yr^{-1}) in most of the grids, whereas Southern Europe presents moderate P surplus and outputs, between 4 and 8 $kg\ ha^{-1}$ of physical area yr^{-1}, with higher P inputs (10–16 $kg\ ha^{-1}$ of physical area yr^{-1}). Notably, industrialized countries like Germany, France, and the Netherlands experienced a peak in P surplus and inputs around 1990, followed by a decline except in the Netherlands, where P surplus exceeded 20 $kg\ ha^{-1}$ of physical area yr^{-1}. P outputs in some regions also continued to rise. By 2015, an increase in grid cells with negative P surplus (P deficit) was observed, particularly in areas like central France and Germany, reflecting a situation where P outputs exceeds P inputs, similar to a century ago, as can be seen in central France and Germany. Central European countries mainly rely on mineral fertilizers, except regions like the Netherlands, Belgium, and Denmark, where animal manure dominates due to high livestock densities (See Supplementary Figure R4). Overall, over the period from 1850 to 2019, our analysis identifies large temporal fluctuations in P fluxes across most European *regions, except for the north, where P fluxes levels have remained stable at a low level. This underscores the importance of long-term datasets in capturing such variations.*

Furthermore, cumulative P fluxes, including P surplus, inputs, and outputs, are presented for four distinct time periods, which we term as following: (i) 1850–1920 (Pre-modern agriculture), (ii) 1921–1960 (Industrialization before the Green Revolution), (iii) 1961–1990 (Green Revolution and synthetic fertilizer expansion), and (iv) 1991–2019 (Environmental awareness and policy intervention phase) (Supplementary Figure S4). These plots revealed marked shift in P dynamics across Europe over time. During 1850-1920, P surplus was relatively low, averaging 8-10 $t\ yr^{-1}$ in much of the Central and Eastern Europe, with some Western Europe regions like France, the Netherlands, and Denmark exceeding 16 $t\ yr^{-1}$. Northern Europe typically showed much lower values of 2-4 $t\ yr^{-1}$. In the subsequent period (1921–1960), P inputs began to rise modestly, averaging 50-70 $t\ yr^{-1}$, driven by early industrialization and chemical fertilizer use, though P surplus remained moderate due to relatively high P outputs. The Green Revolution period (1961–1990) saw a sharp increase in P inputs, exceeding 80 $t\ yr^{-1}$ in many regions due to agricultural intensification, resulting in substantial P surplus, with most areas surpassing 18 $t\ yr^{-1}$. In the most recent phase (1991–2019), P inputs declined steadily due to improved agricultural practices and environmental policies like the EU Nitrates Directive, while P outputs increased, narrowing P surplus. In some Western and Eastern Europe, P surplus even turned negative, reflecting P mining. These temporal and spatial trends highlight the importance of sustainable nutrient management practices and policies in reducing P surplus over time. Moving forward, strategies like reallocating nutrients inputs based on regional needs, improving the integration of crop and livestock systems could help to further optimize nutrient use efficiency. Such measures, coupled with continued monitoring of P indicators-P surplus and PUE- are essential to address P-related environmental challenges and promote sustainable agricultural practices Zou2022.

The peak in P surplus observed around 1980 likely aligns with the intensified fertilizer use of the Green Revolution (Supplementary Figure S5). The subsequent decline in P surplus after 1990 reflects multiple factors, including policy shifts in Western Europe (e.g., Nitrate Directive (Directive 91/676/EEC) (European Commission, 2000b) and Water Framework Directive (Directive 2000/60/EC) (European Commission, 2000a)), regional legislations that restricted P fertilization (Amery and Schoumans, 2014)), economic adjustments, and increased awareness of sustainable nutrient management (Ludemann et al., 2023; Senthilkumar et al., 2012; Cassou, 2018). Country-specific legislation also played a role, since a few European countries, including the Netherlands, Ireland, Norway, and Sweden, have specific legislation limiting P applications (Bouraoui et al., 2011). In some cases, the decrease in P surplus began even earlier, as in Denmark and the UK, where P was not a major limiting factor for crop yield since soil P levels had likely reached sufficient levels for crop production without additional inputs (Bouraoui et al., 2011). On the other side, in Central and Eastern European regions, the collapse of the Soviet Union and subsequent (agro-)economic restructuring may have led to reduced P inputs, as indicated by a sharp drop in fertilizer use (Csathó et al., 2007; Ludemann et al., 2023) (Supplementary Figure S5) and subsequently reflected in corresponding P surplus budgets. Such distinct P surplus patterns observed across Europe appear to have been shaped by these combined influences, and disentangling the different factors will require careful consideration in future studies. On a global scale, Zou et al. (2022) discussed the distinct roles of socioeconomic and environmental factors governing the dynamics of long-term P surplus evolution across different countries.

These enhancements help clarify the major drivers of P surplus and provide better insights into the evolving P dynamics in Europe. By making both the datasets and visualizations publicly accessible, we hope to offer the research community a valuable resource for understanding the interplay between P inputs and outputs across time and regions.

1.6 — Technique corrections:

- Line 49: “difference” should be “different”.

Reply: Thank you for pointing this out. We have corrected the typo in Line 49 by replacing “difference” with “different” as suggested.

- Equation 12 and 13: Wrong equations.

Reply: We have reviewed and corrected Equations 12 and 13 in the manuscript to ensure they accurately reflect the intended calculations. The errors in these equations have been rectified, and the revised versions are now correctly presented.

References

Amery, F. and Schoumans, O.: Agricultural phosphorus legislation in Europe, Instituut voor Landbouw-, Visserij-en Voedingsonderzoek, 2014.

Bartholomé, E. and Belward, A. S.: GLC2000: A new approach to global land cover mapping from earth observation data, Int. J. Remote Sens., 26, 1959–1977, https://doi.org/10.1080/01431160412331291297, 2005.

Batool, M., Sarrazin, F. J., Attinger, S., Basu, N. B., Van Meter, K., and Kumar, R.: Long-term annual soil nitrogen surplus across Europe (1850–2019), Scientific Data, 9, 1–22, https://doi.org/10.1038/s41597-022-01693-9, 2022.

Bayliss-Smith and Wanmali: Understanding Green Revolutions: Agrarian Change and Development Planning in South Asia, Cambridge University Press, 1984.

Bouraoui, F., Grizzetti, B., Aloe, A., et al.: Long term nutrient loads entering European seas, European Commission, Luxembourg, 72, 2011.

Cassou, E.: The Greening of Farm Support Programs: International Experience s with Agricultural Subsidy Reform, World Bank, Washington, DC, 2018.

Cordell, D., Drangert, J.-O., and White, S.: The story of phosphorus: global food security and food for thought, Global environmental change, 19, 292–305, 2009.

Csathó, P., Sisák, I., Radimszky, L., Lushaj, S., Spiegel, H., Nikolova, M., Nikolov, N., Čermák, P., Klir, J., Astover, A., et al.: Agriculture as a source of phosphorus causing eutrophication in Central and Eastern Europe, Soil Use and Management, 23, 36–56, 2007.

Di Gregorio, A.: Land cover classification system: classification concepts and user manual: LCCS, vol. 2, Food & Agriculture Org., 2005.

Dong, H., Mangino, J., McAllister, T. A., Hatfield, J. L., Johnson, D. E., Bartram, D., Gibb, D., and Martin, J. H.: 2006 IPCC Guidelines for National Greenhouse Gas inventories - Chapter 10: Emissions from livestock and manure management, Report, 2020.

Einarsson, R., Sanz-co, A., Aguilera, E., Billen, G., Josette, G., G, H. J. M. V., and Lassaletta, L.: Crop production and nitrogen use in European cropland and grassland 1961 – 2019, Scientific Data, pp. 1–29, https://doi.org/10.1038/s41597-021-01061-z, 2021.

European Commission: Water Framework Directive 2000/60/EC, Official Journal of the European Communities, L 269, 1–15, 2000a.

European Commission: Protection of water against pollution caused by nitrates from agricultural sources, Official Journal of the European Communities, L 269, 1–15, 2000b.

FAOSTAT: FAOSTAT (Food and Agriculture Organization Corporate Statistical Database): Land Use domain, available at:, https://www.fao.org/faostat/en/#data/RL/, [Accessed: 10-August-2021], 2021.

FAOSTAT: FAOSTAT (Food and Agriculture Organization Corporate Statistical Database): FAOSTAT Domain Soil Nutrient Budget Metadata, release November 2022, available at:, https://www.fao.org/faostat/en/#data/ESB, [Accessed: 01-May-2023], 2023.

Gilbert, M., Nicolas, G., Cinardi, G., Van Boeckel, T. P., Vanwambeke, S. O., Wint, G. R., and Robinson, T. P.: Global distribution data for cattle, buffaloes, horses, sheep, goats, pigs, chickens and ducks in 2010, Scientific Data, 5, https://doi.org/10.1038/sdata.2018.227, 2018.

Goldewijk, K. K., Beusen, A., Doelman, J., and Stehfest, E.: Anthropogenic land use estimates for the Holocene - HYDE 3.2, Earth Syst. Sci. Data, 9, 927–953, https://doi.org/10.5194/essd-9-927-2017, 2017.

Guejjoud, H., Curie, F., and Grosbois, C.: Analyzing a century of agricultural phosphorus surplus and its long-term key drivers in France, Nutrient Cycling in Agroecosystems, https://doi.org/10.1007/s10705-023-10300-1, 2023.

Heffer, P., Gruère, A., and Roberts, T.: Assessment of Fertilizer Use by Crop at the Global Level, Tech. rep., International Fertilizer Association (IFA) and International Plant Nutrition Institute (IPNI), https://www.fertilizer.org/resource/assessment-of-fertilizer-use-by-crop-at-the-global-level-2014-2014-15/, 2017.

Lu, C. and Tian, H.: Global nitrogen and phosphorus fertilizer use for agriculture production in the past half century: shifted hot spots and nutrient imbalance, Earth System Science Data, 9, 181–192, 2017.

Ludemann, C. I., Wanner, N., Chivenge, P., and ...: A global reference database in FAOSTAT of cropland nutrient budgets and nutrient use efficiency: nitrogen, phosphorus and potassium, 1961–2020, Earth System Science Data, pp. 1–24, https://essd.copernicus.org/preprints/essd-2023-206/%0Ahttps://essd.copernicus.org/preprints/essd-2023-206/essd-2023-206.pdf, 2023.

Malagó, A. and Bouraoui, F.: Global anthropogenic and natural nutrient fluxes: from local to planetary assessments, Environmental Research Letters, 16, 054 074, 2021.

McCrackin, M. L., Muller-Karulis, B., Gustafsson, B. G., Howarth, R. W., Humborg, C., Svanbäck, A., and Swaney, D. P.: A century of legacy phosphorus dynamics in a large drainage basin, Global Biogeochemical Cycles, 32, 1107–1122, 2018.

Monfreda, C., Ramankutty, N., and Foley, J. A.: Farming the planet: 2. Geographic distribution of crop areas, yields, physiological types, and net primary production in the year 2000, Global Biogeochem. Cycles, 22, 1–19, https://doi.org/10.1029/2007GB002947, 2008.

Müller-Karulis, B., McCrackin, M. L., Dessirier, B., Gustafsson, B. G., and Humborg, C.: Legacy nutrients in the Baltic Sea drainage basin: How past practices affect eutrophication management, Journal of Environmental Management, 370, 122 478, 2024.

Panagos, P., Muntwyler, A., Liakos, L., Borrelli, P., Biavetti, I., Bogonos, M., and Lugato, E.: Phosphorus plant removal from European agricultural land, Journal fur Verbraucherschutz und Lebensmittelsicherheit, 17, 5–20, https://doi.org/10.1007/s00003-022-01363-3, 2022.

Pratt, C. and El Hanandeh, A.: The untapped potential of legacy soil phosphorus, Nature Food, 4, 1024–1026, https://doi.org/10.1038/s43016-023-00890-y, 2023.

Ramankutty, N., Evan, A. T., Monfreda, C., and Foley, J. A.: Farming the planet: 1. Geographic distribution of global agricultural lands in the year 2000, Global Biogeochem. Cycles, 22, 1–19, https://doi.org/10.1029/2007GB002952, 2008.

Ringeval, B., Demay, J., Goll, D. S., He, X., Wang, Y. P., Hou, E., Matej, S., Erb, K. H., Wang, R., Augusto, L., Lun, F., Nesme, T., Borrelli, P., Helfenstein, J., McDowell, R. W., Pletnyakov, P., and Pellerin, S.: A global dataset on phosphorus in agricultural soils, Scientific Data, 11, 1–34, https://doi.org/10.1038/s41597-023-02751-6, 2024.

Roberts, P.: European Pollutant Release and Transfer Register (E-PRTR). Capturing and reporting industrial pollutant emissions data, CONCAWE Review, 19, 2009.

Sarrazin, F. J., Attinger, S., and Kumar, R.: Gridded dataset of nitrogen and phosphorus point sources from wastewater in Germany (1950–2019), Earth System Science Data, 16, 4673–4708, 2024.

Senthilkumar, K., Nesme, T., Mollier, A., and Pellerin, S.: Regional-scale phosphorus flows and budgets within France: The importance of agricultural production systems, Nutrient Cycling in Agroecosystems, 92, 145–159, https://doi.org/10.1007/s10705-011-9478-5, 2012.

Sharpley, A., Jarvie, H. P., Buda, A., May, L., Spears, B., and Kleinman, P.: Phosphorus legacy: overcoming the effects of past management practices to mitigate future water quality impairment, Journal of environmental quality, 42, 1308–1326, 2013.

Vigiak, O., Grizzetti, B., Zanni, M., Aloe, A., Dorati, C., Bouraoui, F., and Pistocchi, A.: Domestic waste emissions to European waters in the 2010s, Scientific Data, 7, 33, 2020.

Xu, R., Tian, H., Pan, S., Dangal, S. R., Chen, J., Chang, J., Lu, Y., Skiba, U. M., Tubiello, F. N., and Zhang, B.: Increased nitrogen enrichment and shifted patterns in the world's grassland: 1860–2016, Earth System Science Data, 11, 175–187, 2019.

Zhang, B., Tian, H., Lu, C., Dangal, S. R., Yang, J., and Pan, S.: Global manure nitrogen production and application in cropland during 1860–2014: a 5 arcmin gridded global dataset for Earth system modeling, Earth System Science Data, 9, 667–678, 2017.

Zou, T., Zhang, X., and Davidson, E.: Global trends of cropland phosphorus use and sustainability challenges, Nature, 611, 81–87, 2022.

List of Figures

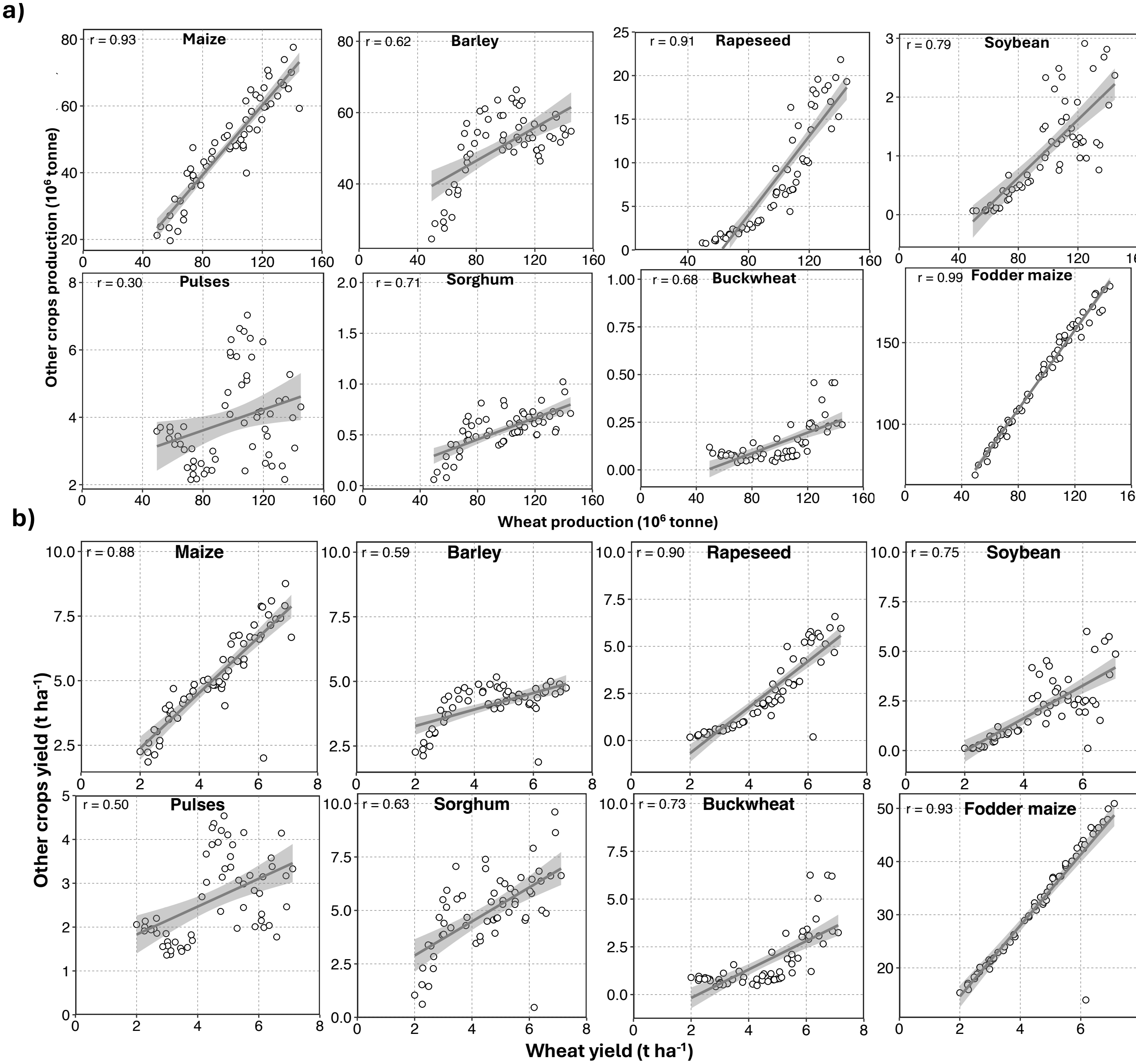

Figure R1. Scatter plots illustrating the relationship between wheat production and the production of various crops (panel a, in 10^6 t) and between wheat yield and the yield of various crops (panel b, in $t\ ha^{-1}$) across the European Union (EU28) from 1961 to 2019. Each circle represents a single year, illustrating the temporal relationship between wheat and other crops. The correlation coefficient (r) is annotated in each panel, highlighting the degree of alignment in temporal dynamics. The eight crops shown—maize, barley, rapeseed, soybean, pulses, sorghum, buckwheat, and fodder maize—were selected based on their wide cultivation across the EU28 and/or their relatively high P content.

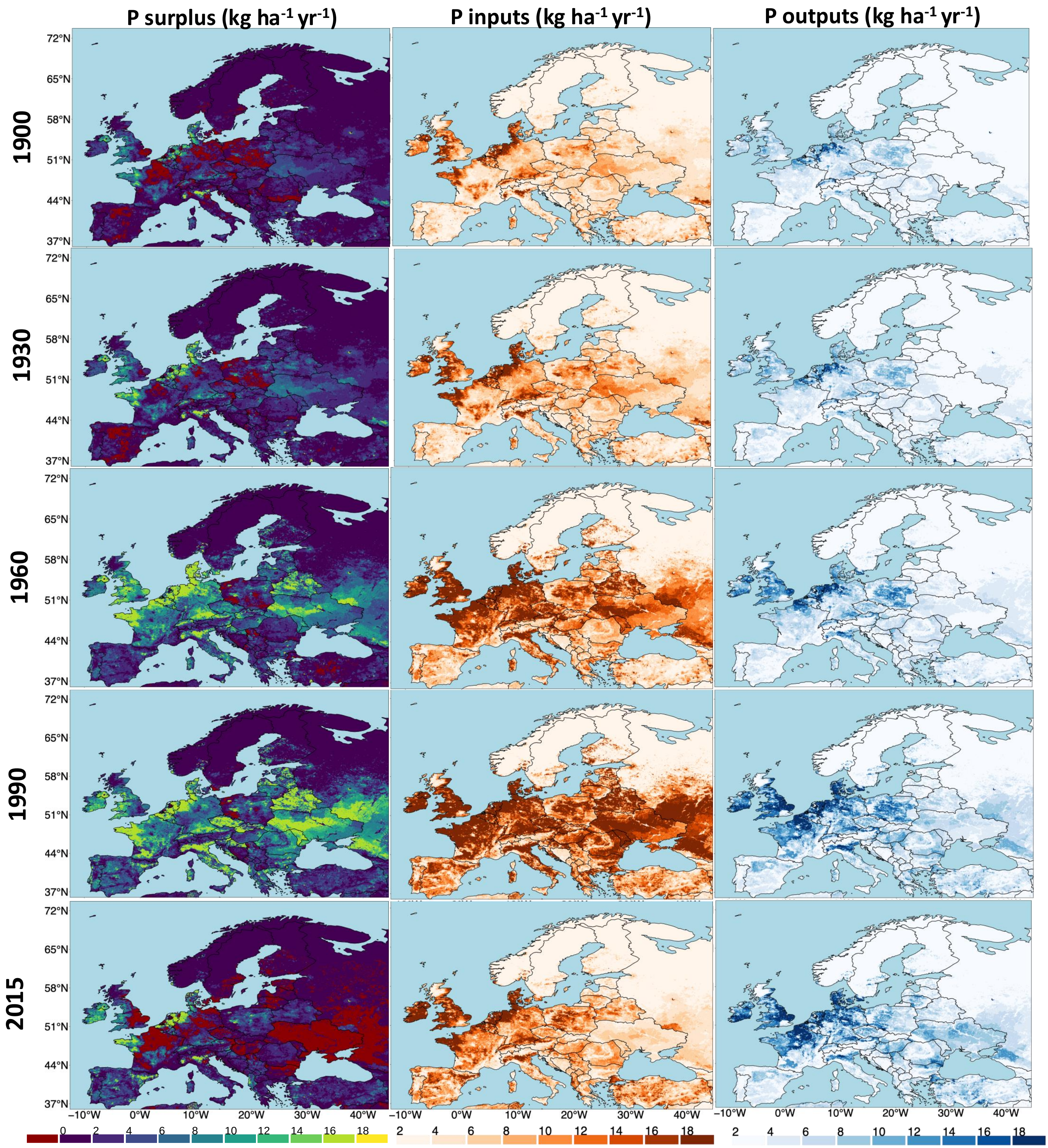

Figure R2. Snapshots of P surplus, P inputs and P outputs ($kg\ ha^{-1}$ of grid physical area yr^{-1}) across Europe. The figure shows the annual spatial variation in P surplus, P inputs and P outputs given as the mean of our 48 P surplus, P inputs and P outputs estimates for the selected years.

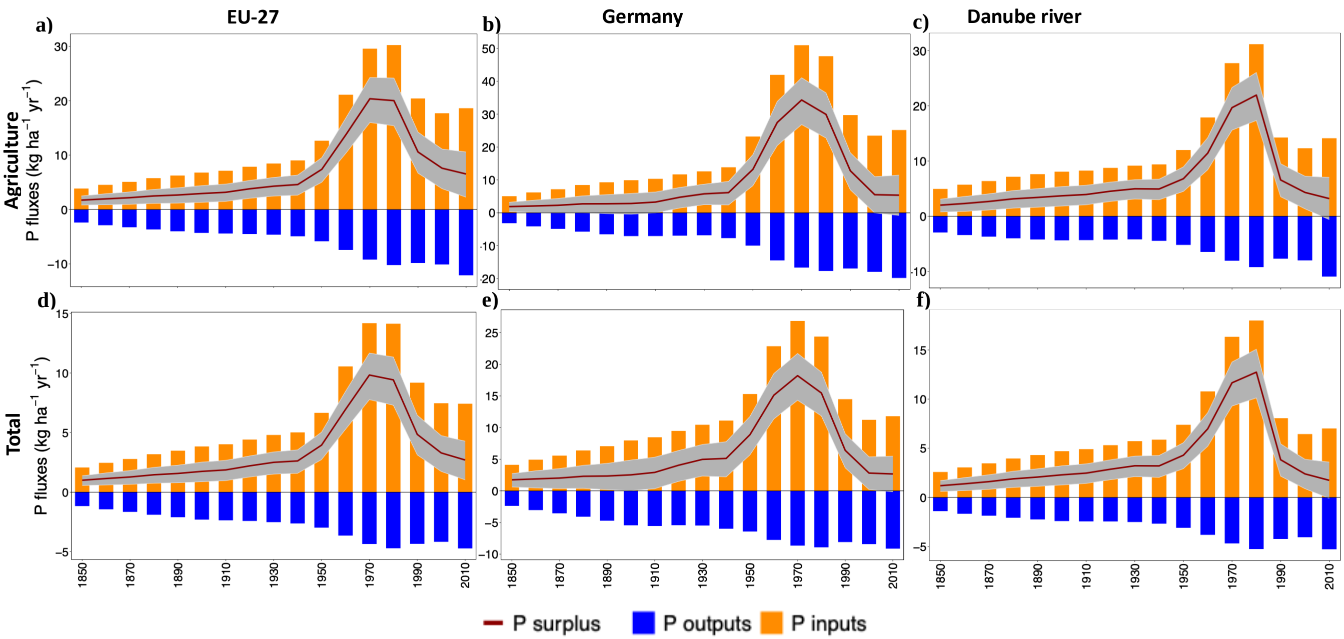

Figure R3. Decadal trajectory of agricultural P surplus ($kg\ ha^{-1}$ of agricultural area yr^{-1}) and total P surplus ($kg\ ha^{-1}$ of physical area yr^{-1}) and its contributing components for the EU-27, Germany, and the Danube river basin from 1850 to 2019. Upward orange bars represent the average of 48 P inputs, while downward blue bars indicate the average of 48 P outputs, showing decadal means. The grey ribbon shows the range (min and max) of the 48 P surplus estimates, with the red line representing the average value for each decade. **(a-c)** Agricultural P surplus for EU-27, Germany and Danube river, **(d-f)** Total P surplus for EU-27, Germany and Danube river

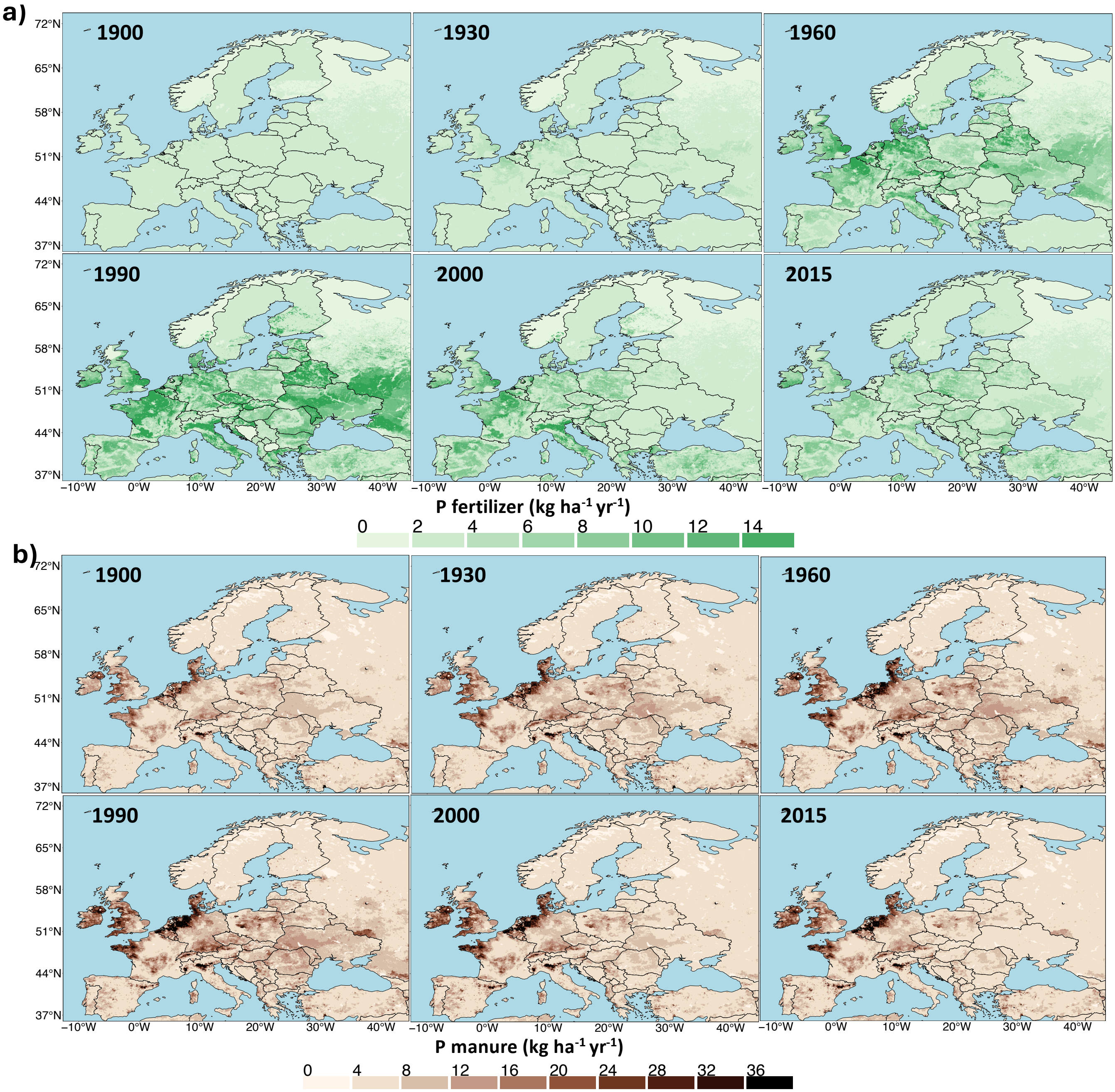

Figure R4. Snapshots of P inputs from mineral fertilizer (a) and animal manure (b) ($kg\ ha^{-1}$ of physical area yr^{-1}) across Europe. The figure shows the annual spatial variation in P inputs from mineral fertilizer and animal manure given estimates for the selected years.

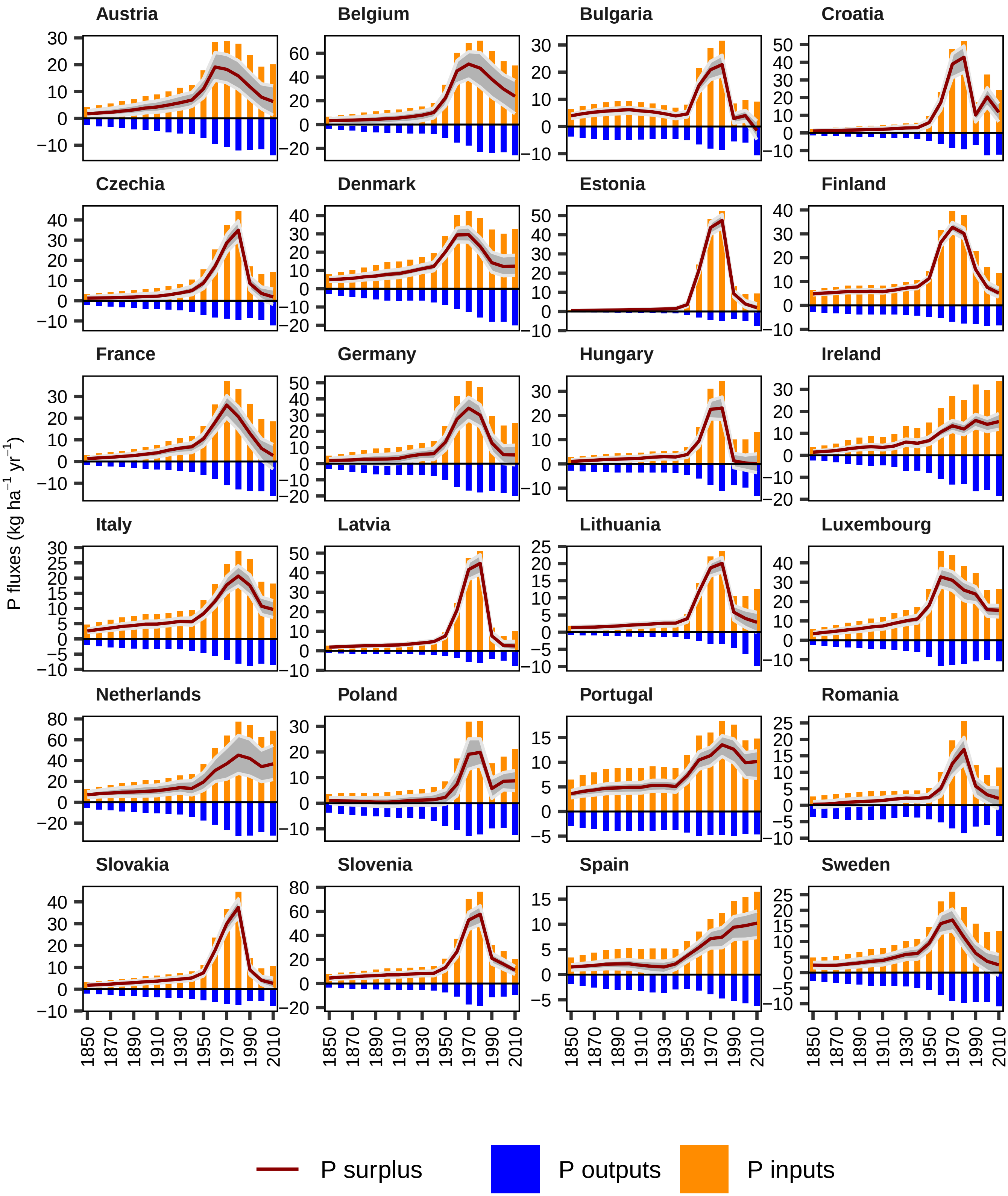

Figure R5. Decadal trajectory of agricultural P surplus ($kg\ ha^{-1}$ of agricultural area yr^{-1}) and its contributing components for different European countries from 1850 to 2019. Upward orange bars represent the average of 48 P inputs, while downward blue bars indicate the average of 48 P outputs, showing decadal means. The grey ribbon shows the range (min and max) of the 48 P surplus estimates, with the red line representing the average value for each decade.

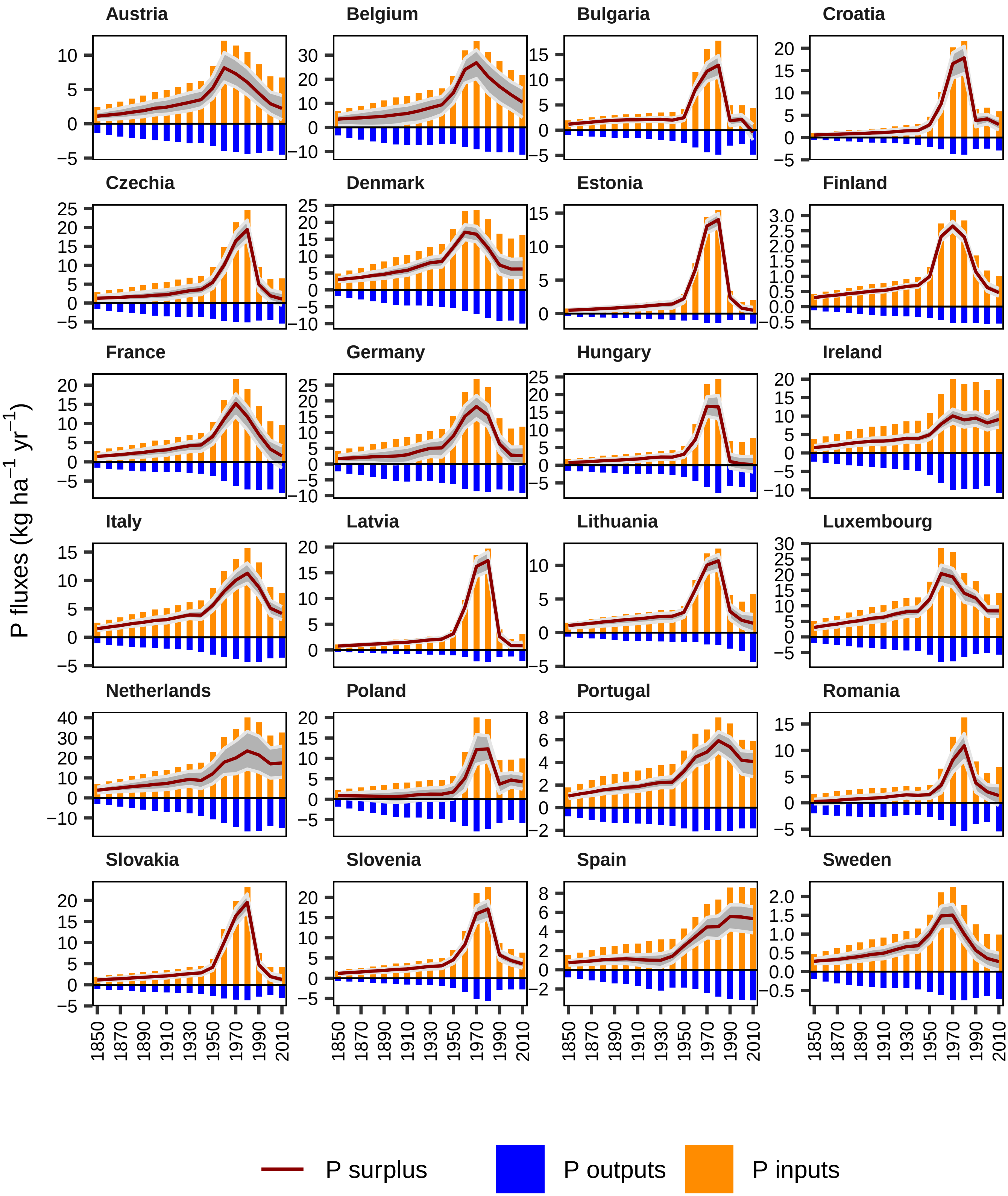

Figure R6. Decadal trajectory of total P surplus ($kg\ ha^{-1}$ of physical area yr^{-1}) and its contributing components for different European countries from 1850 to 2019. Upward orange bars represent the average of 48 P inputs, while downward blue bars indicate the average of 48 P outputs, showing decadal means. The grey ribbon shows the range (min and max) of the 48 P surplus estimates, with the red line representing the average value for each decade.

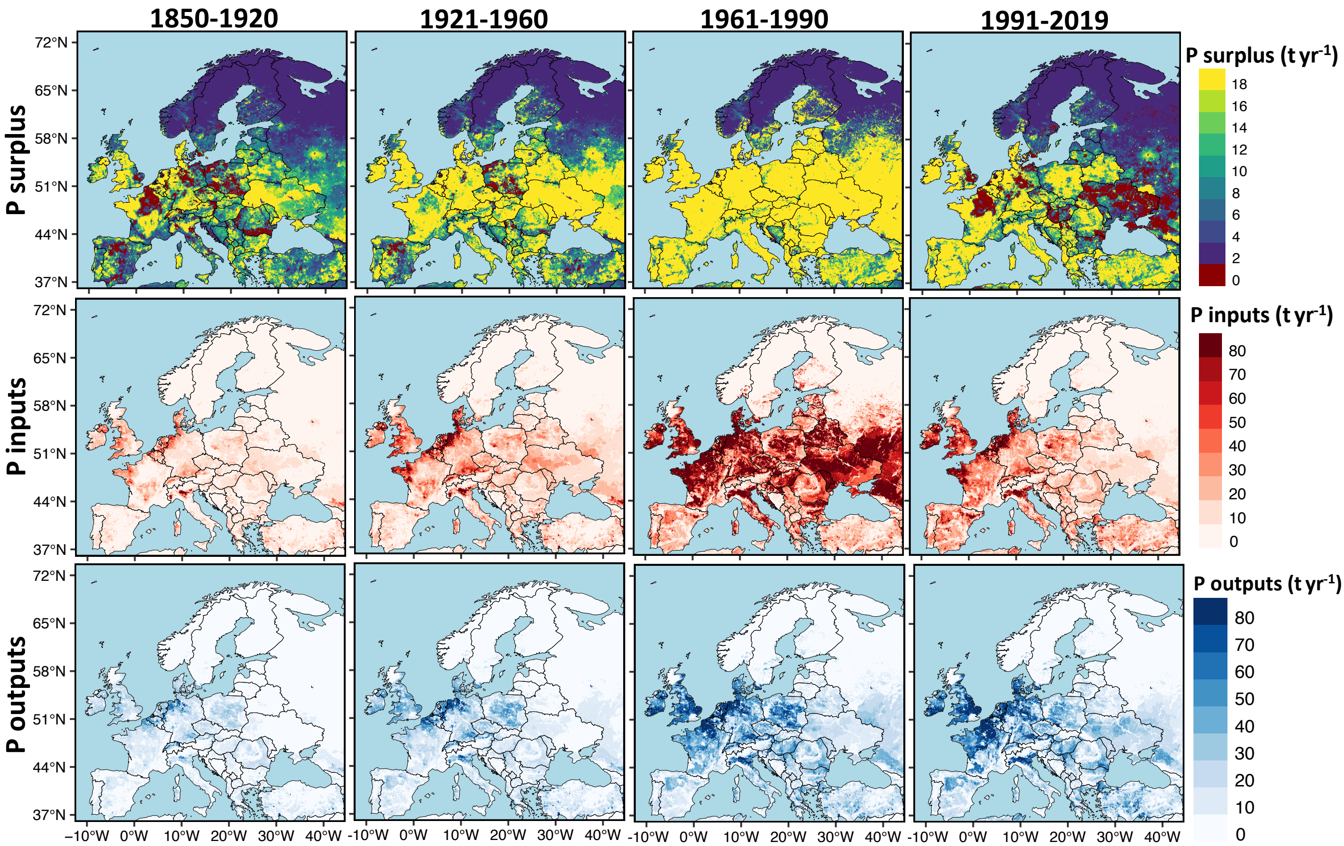

Figure R7. Cumulative total P surplus, P inputs, and P outputs over four historical periods across Europe. The first row shows the accumulated phosphorus (P) surplus, the second row displays P inputs, and the third row illustrates P outputs across Europe for four distinct periods, which we term as following: (i) 1850–1920 (Pre-modern agriculture), (ii) 1921–1960 (Industrialization before the Green Revolution), (iii) 1961–1990 (Green Revolution and expansion of synthetic fertilizers), and (iv) 1991–2019 (Environmental awareness and policy intervention phase). All values are normalized per year within each time period, with units in tonnes per year.

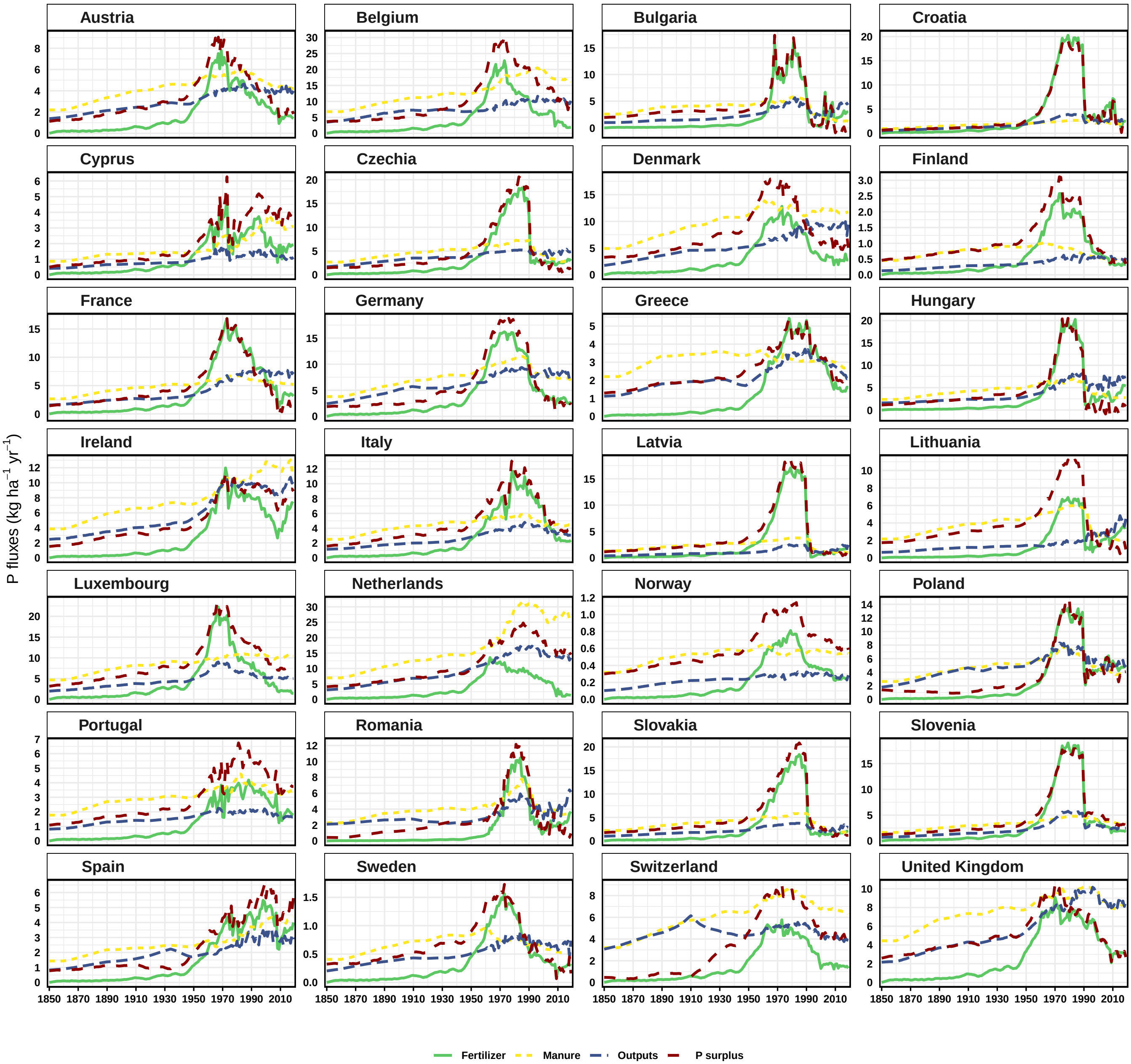

Figure R8. Time series of phosphorus (P) inputs from fertilizer and manure, P outputs, and P surplus ($kg\ ha^{-1}$ of physical area yr^{-1}) across various European countries from 1850 to 2019. This figure highlights changes in P fluxes over time, showing a peak in P surplus around 1980 followed by a decline after 1990 for most countries. These patterns illustrate the influence of agricultural intensification during the Green Revolution, as well as subsequent policy, economic, and environmental shifts in both Western and Eastern Europe. The red line represents the mean of 48 P surplus estimates, while green, yellow, and blue lines depict temporal dynamics for P inputs from fertilizer, manure, and P outputs, respectively. Together, these pattern provide insight into how various factors may have influenced P surplus dynamics over time.