Posted on 18 Jan 2024 — The copyright holder is the author/funder. All rights reserved. No reuse without permission. — https://doi.org/10.22541/essoar.170560666.61809171/v1 — This is a preprint and has not been peer-reviewed. Data may be preliminary.

Projections of Soil Organic Carbon in China: The Role of Carbon Fluxes Revealed by Explainable Artificial Intelligence

Yongkun Zhang[1], Feini Huang[1], Xingjie Lu[1], Wei Shangguan[1], Gaosong Shi[2], Ye Zhang[1], Zhangcai Qin[1], Zhongwang Wei[1], Hua Yuan[1], and Yongjiu Dai[1]

[1]Sun Yat-sen University
[2]Sun Yat-sen Univerisity

January 18, 2024

Abstract

The impact of carbon fluxes on soil organic carbon (SOC) remains underexplored. We employed machine learning to model SOC dynamics. Our findings project an increase in China's SOC through to the year 2100 across various Shared Socioeconomic Pathways. Sensitivity analyses have identified carbon fluxes as the main drivers for this projected rise, followed by climate and land use. Further examination using an explainable artificial intelligence method, Shapley Additive Explanations, has uncovered both spatial and temporal variations in how gross primary production (GPP) influences SOC levels. Notably, GPP's contribution on SOC is initially negative at low levels, turning positive once a threshold of approximately 3 gC m-2d-1 is surpassed. Beyond a GPP of about 7 gC m-2d-1, its positive contribution to SOC plateaus. Critical zones for soil carbon sequestration are located around 400 mm annual precipitation line.

Hosted file

`983176_0_supp_11724512_s6zw6g.docx` available at https://authorea.com/users/715501/articles/700152-projections-of-soil-organic-carbon-in-china-the-role-of-carbon-fluxes-revealed-by-explainable-artificial-intelligence

Projections of Soil Organic Carbon in China: The Role of Carbon Fluxes Revealed by Explainable Artificial Intelligence

Yongkun Zhang[1], Feini Huang[1], Xingjie Lu[1], Wei Shangguan[1], Gaosong Shi[1], Ye Zhang[1], Zhangcai Qin[1], Zhongwang Wei[1], Huan Yuan[1], and Yongjiu Dai[1]

[1]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai, 519082, China.

Corresponding author: Wei Shangguan (shgwei@mail.sysu.edu.cn)

Key Points:

- The influence of carbon flux on SOC is more pronounced than that of climate change and land use change
- We identify two critical thresholds in the relationship between gross primary production and SOC
- Critical zones for soil carbon sequestration are located around 400 mm annual precipitation line

Abstract

The impact of carbon fluxes on soil organic carbon (SOC) remains underexplored. We employed machine learning to model SOC dynamics. Our findings project an increase in China's SOC through to the year 2100 across various Shared Socioeconomic Pathways. Sensitivity analyses have identified carbon fluxes as the main drivers for this projected rise, followed by climate and land use. Further examination using an explainable artificial intelligence method, Shapley Additive Explanations, has uncovered both spatial and temporal variations in how gross primary production (GPP) influences SOC levels. Notably, GPP's contribution on SOC is initially negative at low levels, turning positive once a threshold of approximately 3 gC $m^{-2}d^{-1}$ is surpassed. Beyond a GPP of about 7 gC $m^{-2}d^{-1}$, its positive contribution to SOC plateaus. Critical zones for soil carbon sequestration are located around 400 mm annual precipitation line.

Plain Language Summary

Soil's ability to absorb carbon is key to reducing atmospheric carbon dioxide, a major greenhouse gas. Yet, the influence of carbon fluxes—the exchange of carbon between the soil and the atmosphere—on soil carbon storage is not well understood. Our study utilized machine learning to estimate potential soil carbon storage in China by 2100, considering various global socioeconomic trajectories. We anticipate an uptick in soil carbon, largely due to carbon fluxes, with climate and land use changes also playing significant roles. Through explainable artificial intelligence, we've gained insights into how plant growth impacts soil carbon levels. We discovered that minimal plant growth correlates with lower soil carbon storage. As plants grow more, they enhance soil carbon storage until reaching a certain growth level, after which the effect plateaus. Zones critical for maximizing soil carbon storage correspond with areas receiving about 400 mm of rainfall annually. This understanding of plant growth's effect on soil carbon is invaluable for developing land management strategies aimed at maximizing carbon sequestration, thereby contributing to climate change mitigation efforts.

1 Introduction

Soil Organic Carbon (SOC) is a fundamental constituent of terrestrial ecosystems, performing an essential function in bolstering the resilience and productivity of ecosystems (Batjes, 2014; Lal, 2003; Minasny et al., 2017). SOC is not only crucial for providing nutrients that support plant growth and yield but also for retaining water and mitigating soil erosion (Trivedi et al., 2018). Even slight changes in the soil carbon pool can result in significant impacts on atmospheric carbon (Smith et al., 2008). Soil carbon sequestration, through its ability to capture and retain environmental carbon, acts as a powerful antidote against the intensification of the greenhouse effect (Lal et al., 2015). Therefore, predicting future SOC and identifying its key drivers are essential for understanding the evolving patterns of carbon stock distribution over time.

Methods for studying SOC are generally categorized into two types: process-based models and empirical models such as AI (Artificial Intelligence) methods. Process-based models simulate SOC dynamics based on detailed representations of internal biochemical and physical processes (Le Quéré et al., 2013). The Earth System Model (ESM) is an example of such a model, integrating carbon cycle processes with climate models (Intergovernmental Panel on Climate Change, 2023). These models are capable of projecting SOC distribution and temporal changes. However, due to the still uncertain physio-ecological mechanisms of SOC in terrestrial system, different ESMs have shown discrepancies in both historical and future SOC estimations (Ito et al., 2020).

Recently, AI methods have become powerful tools to for mapping and predicting SOC (McBratney et al., 2019). The SCORPAN framework (McBratney et al., 2003), introduced for Digital Soil Mapping (DSM), suggests that soil types or properties can be inferred from a combination of environmental factors (i.e., covariates). These include soil, climate, organisms, topography, parent material, age, spatial location, and other environmental variables (Chen et al., 2022; Lamichhane et al., 2019). The application of DSM technology to project future SOC changes relies on the space-for-time substitution concept (Pickett, 1989), which has been employed to anticipate SOC trends in regions such as Europe, China and Argentina (Heuvelink et al., 2021; Yigini & Panagos, 2016; Zhang et al., 2023). Among the various methods, Random

Forest (RF) has emerged as the most popular method for SOC mapping and prediction, demonstrating its effectiveness in this domain (Lamichhane et al., 2019; Padarian et al., 2020).

Significant research has been conducted on the anticipated changes in SOC, with climate change and land use change commonly recognized as the primary factors influencing future SOC variability (Davidson & Janssens, 2006). SOC are controlled by both carbon input and residence time (Luo et al, 2022). However, the role of carbon fluxes in shaping SOC dynamics has not been thoroughly investigated. The CO2 fertilization effect suggests that as atmospheric CO2 concentrations increase, carbon fluxes to ecosystems also rise (Baldocchi et al., 2001; Litton & Giardina, 2008). Yet, this additional carbon input may also enhance SOC decomposition, potentially leading to increased SOC loss (Crow et al., 2009; Kuzyakov, 2010; Sayer et al., 2011). Consequently, it remains uncertain whether such fertilization will result in soils becoming net carbon sources or sinks in the future (Field, 2001; Karnosky, 2003; Nowak et al., 2004; Liang et al., 2018).

Explainable Artificial Intelligence (XAI) has been successfully applied to attribute analysis in soil carbon studies (Luo et al., 2019; Patoine et al., 2022). To dissect the impact of various factors on SOC, with a focus on the influence of carbon flux, we integrated two XAI methods into our analysis: Random Forest Importance (RFI) and Shapley Additive Explanations (SHAP) (Huang et al., 2023). These methods will allow us to unravel the complex interactions between carbon flux and SOC, providing a clearer understanding of their relationship. This study seeks to elucidate three essential scientific questions: (1) What degree of variation in SOC levels can be expected in China from 2021 to 2100 under multiple Shared Socioeconomic Pathways (SSPs)? (2) What is the relative contribution of carbon fluxes to changes in SOC compared to the effect of climate change and land use change? (3) How will carbon fluxes shape the trajectory of SOC in the future?

2 Materials and Methods

2.1 Materials and Processing

In this study, we utilized a dataset comprising 8,979 soil profile records from the Second National Soil Survey of China, with most soil profile data being collected between 1979 and 1984 (Shangguan et al., 2013). SOC data from 2000 to 2014 were obtained from the carbon density dataset for China's terrestrial ecosystems (Xu et al., 2019). We focused our analysis on data from the 0-100 cm soil layer, converting the profile data to SOC density using a specified equation:

$$SOC = SOM \times 0.58 \times D \times BD \times (1 - G) \quad (1)$$

where SOM is soil organic matter, D denotes soil layer depth, BD is bulk density and G is gravel content (>2 mm). To standardize the depth of the data, we employed equal-area second-order spline interpolation (Odgers et al., 2012).

All covariates were resampled to a uniform resolution of 2.5 arc-minutes. The input covariates for our analysis were divided into static and dynamic categories, with a comprehensive list provided in Table S1. For static variables, such as those derived from the digital elevation model (DEM), we operated under the assumption that relief and parent material factors would remain constant over long-time scales, and that soil factors would undergo minimal changes over the span of hundreds of years (Grunwald, 2010).

The dynamic datasets were divided into historical and future periods for analysis. Historical data were synchronized with the timing of soil profile collection and assessed for each subsequent twenty-year interval. Covariate data from the periods 1980-1999 and 2000-2015 were aligned with the soil profile data. For projections into the future, we sourced data from four ESMs (ACCESS-ESM1-5 (Ziehn et al., 2020), EC-Earth3-Veg (Döscher et al., 2022), IPSL-CM6A-LR (Boucher et al., 2020), and MPI-ESM1-2-LR (Mauritsen et al., 2019)) under four Shared Socioeconomic Pathways (SSPs), for each twenty-year segment extending from 2021 to 2100, as only these ESMs provide outputs of carbon fluxes in the Coupled Model Intercomparison Project Phase 6 (CMIP6). The dynamic covariates encompassed climate variables, land use patterns, and carbon fluxes. Climate data, which included monthly maximum and minimum temperatures and precipitation, were sourced from the WorldClim2 database (Fick & Hijmans, 2017). Land use information was provided by the Land Use Harmonization project (LUH2) (Hurtt et al., 2020)， covering four SSP scenarios（SSP126, SSP248, SSP370, and SSP585) as well as historical periods. Carbon flux data were represented by two key variables: gross primary productivity (GPP) and net ecosystem productivity (NEP). Historical carbon flux data were acquired from the Global Carbon Fluxes dataset (GCFD, (Shangguan et al., 2023)), while future data were processed from the corresponding ESM data. Given the for analysis. We subsequently employed the following formula to calculate future carbon fluxes:

$$CF^{*}_{i,j} = CF_{history} + (CF_{i,j} - CF_{history,j}) \quad (2)$$

where CF denotes carbon flux covariates, i denotes each period in the future, and j denotes each ESM.

2.2 Model Building and Prediction

We developed RF models to predict SOC using data from two historical periods: 1980-1999 and 2000-2015. The regression matrix was constructed using soil profiles and covariates. After eliminating correlated variables, we selected the thirty most influential variables for inclusion in the model. The model's accuracy was evaluated using ten-fold cross-validation and three indicators: R^2, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Our experiments spanned six timeframes: the historical periods of 1980-1999 and 2000-2015, and the future period of 2021-2100, divided into four twenty-year intervals. Following the space-for-time substitution strategy (Blois et al., 2013; Liu et al., 2020), we treated the dynamic covariates from these six periods as sequential input data for the model. This approach allowed us to map historical SOC levels and project future SOC across different Earth System Models (ESMs) and Shared Socioeconomic Pathways (SSPs).

2.3 Attribution Analysis

We conducted a sensitivity analysis as follows. Initially, we held each of the three types of dynamic covariates—climate, land use, and carbon flux—at their historical values from the period 2000-2015 and then projected future SOC levels. By comparing these projections with the original predictions, we evaluated the influence of each covariate type on future SOC levels.

To delve deeper into the impact of specific variables on SOC, we employed XAI tools—RFI and SHAP—for detailed attribution analysis. RFI gauges the significance of variables in tree-based models by aggregating the decrease in entropy across all trees (Breiman, 2001). A variable that effectively partitions the data and substantially lowers entropy is deemed crucial for prediction. For robustness, importance values were averaged over 10 iterations. SHAP, conceptualized by Lundberg and Lee (Lundberg & Lee, 2017) based on game theory principles, computes the marginal contributions of each feature, treating feature values as players in a coalition. Given the current set of feature value, the estimated SHAP value is the contribution of a feature value to the difference between the actual prediction and the mean prediction. A low absolute value signifies that the impact of a feature on the deviation from the mean prediction is relatively minor. The sign (+/-) does not denote a positive or negative feedback mechanism; rather, it indicates whether the effect of the feature increases or decreases the deviation from the mean. For each input variable, we generated SHAP value maps analogous to the SOC predictions. Due to the intensive computational demands of SHAP, we consolidated the input covariates into 10x10 patches for analysis.

3 Results

3.1 Model Performance and Predictions

The performance metrics for our RF model indicated an R^2 of 0.41, an RMSE of 0.30 gC cm^{-2}, and a MAE of 0.22 gC cm^{-2}. By incorporating covariates from various time periods into the model, we were able to map SOC for historical periods and project SOC for future periods. During the two historical periods analyzed, regions with high SOC values were predominantly located in the mountainous area of Northeast China and the eastern part of the Qinghai-Tibet Plateau (Figure 1a and 1b).

We assessed the changes in SOC relative to the historical period (2000-2015) for each subsequent time interval (Figure 1c-1f). The spatial distribution patterns indicated a modest increase in SOC across most of the country in the future, although the regions experiencing significant increases were fewer under the lower carbon emission scenarios. Notably, SOC declines were primarily observed in the northeastern areas, while increases were concentrated in the eastern and southern regions of the Qinghai-Tibet Plateau. Under the SSP585 scenario (Figure 1f), the northeastern region experienced a more pronounced decrease in SOC compared to other SSPs, and the areas of increase were noticeably smaller than those under SSP370 (Figure 1e). This disparity may contribute to the overall lower SOC projections for SSP585 relative to SSP370.

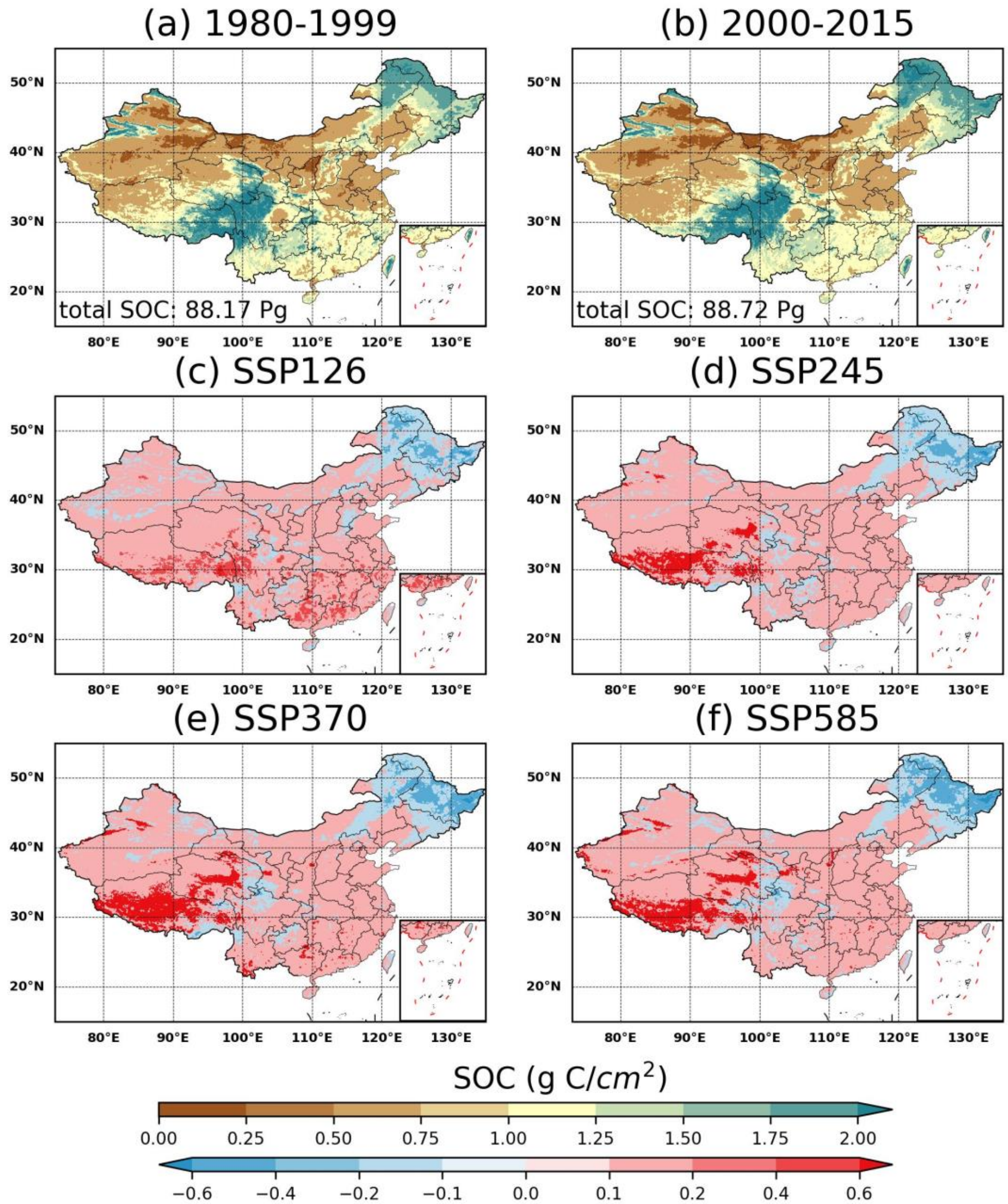


Figure 1. Maps depicting the distribution of soil organic carbon (SOC) density during the historical period and the projected changes in SOC density from 2000-2015 to 2081-2100 under various Shared Socioeconomic Pathways (SSPs). The data are averaged across multiple Earth System Models (ESMs).

3.2 Sensitivity Analysis

The aggregated results from four ESMs indicated an upward trend in total SOC stock across different SSPs (Figure 2a-2c), with the most substantial increase in SOC observed under SSP370 (Figure 2c). A notable variation was found in the SOC estimates produced by the different ESMs. Specifically, the SOC values from ACCESS-ESM1-5 and IPSL-CM6A-LR were markedly lower than those from the other two models, with SOC even showing a decline when compared to historical levels for ACCESS-ESM1-5.

The influences of climate, land use and carbon fluxes on future SOC were discerned by comparing the differences between the results obtained after holding these three types of variables constant and the original predictions (Figure 2e-2p). The discrepancy attributed to carbon fluxes was the greatest among the three sets of results and their positive effects on SOC increased as carbon emission increased. When carbon fluxes or land use were held constant, the projections were lower than the original predictions, whereas the fixed climate variables resulted in higher projections. This suggests that carbon flux and land use are likely to have a positive impact on future SOC, while climate variables may exert a negative influence due to faster soil decomposition. For climate variables, the difference under SSP126 initially increased and then decreased, implying that the adverse effects of climate on SOC first intensified and then diminished under the low carbon emission scenario. This phenomenon can be attributed to the complex interplay between warming effects: while it can lengthen the growing season and enhance productivity, it accelerates the decomposition rate of SOC. Since all ESMs utilized the same land use data, their results were consistent. However, as time advanced, the disparities among the different SSPs grew more pronounced. Notably, under SSP585, the variation due to land use was significantly less than under the other SSPs, suggesting that land use had a minimal impact on SOC changes under this pathway, with the other two variable types being more influential. Furthermore, the trend in SSP585 revealed that the positive contribution of land use initially rose and then fell over time.

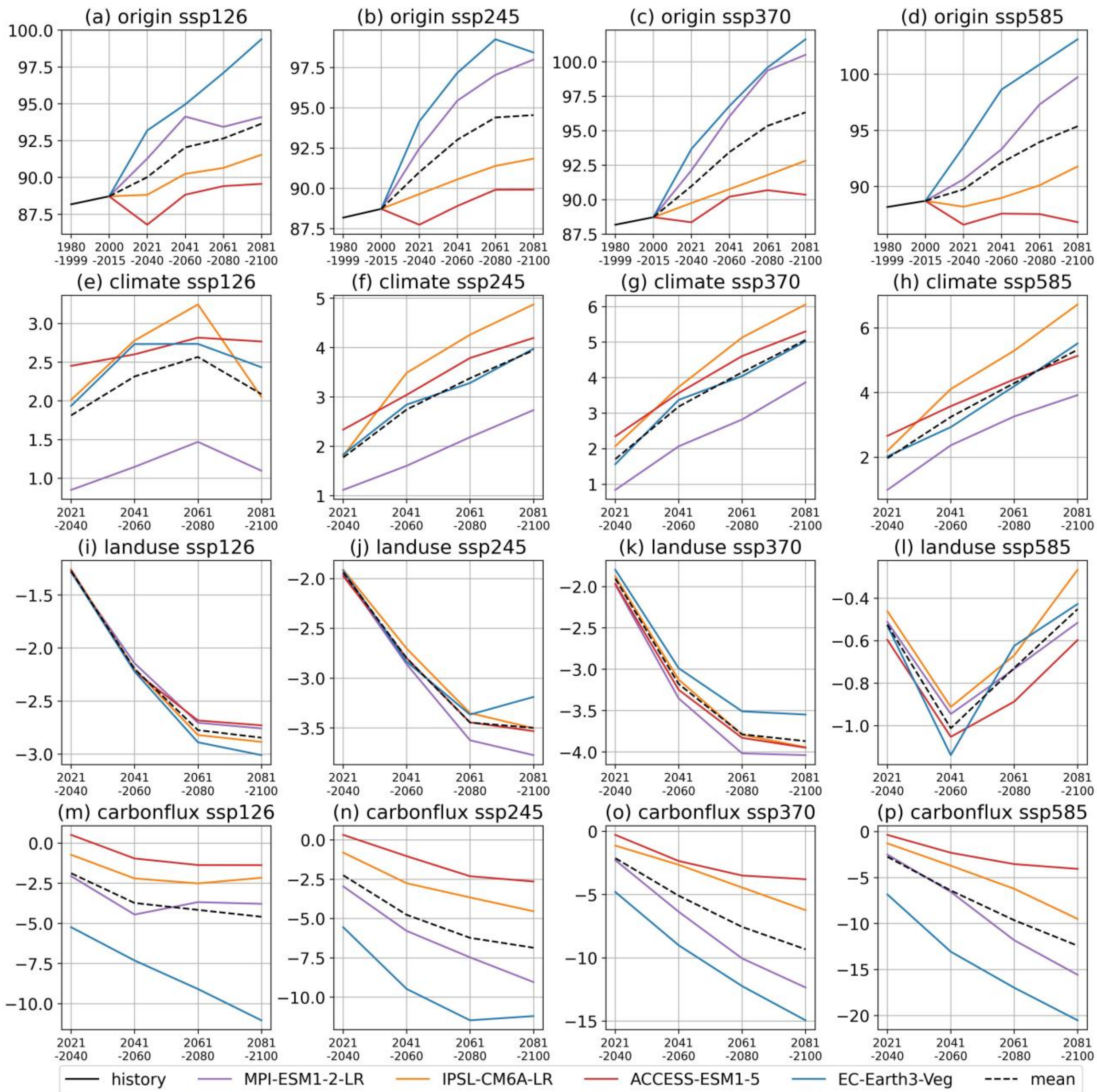


Figure 2. The temporal evolution of total SOC stock in peta-grams of carbon (Pg C) and discrepancies arising from holding specific variables constant. (a-d) the original predictions of SOC stock; variations resulting from fixing (e-h) climate, (i-l) land use and (m-p) carbon fluxes.

3.3 Attribution Analysis with XAI

Figure S1 displays RFI and SHAP values for various variables, featuring the top two variables in each category based on RFI for brevity. The results are based on outputs from EC-Earth3-Veg under SSP585, with similar observations across other ESMs and SSPs (not shown).

The April maximum temperature (tmax_04) emerged as the most important variable, followed by DEM and carbon flux variables. Notably, the SHAP values of certain variables exhibited temporal changes. Specifically, the SHAP values for tmax_04 and the carbon flux variables underwent inversion over time, indicating a reversal in their contributions to SOC. In the period 2081-2100, tmax_04 transitioned from positive contribution to negative contribution, while GPP of summer (GPP_S2) exhibited the opposite trend.

In Figure 3a-3d, the tmax_04 variable exhibited a gradual decrease in SHAP over time, while secmb (secondary mean biomass density), GPP_S2, and NEP_S2 showed a progressive increase. The SHAP of tmax_04 decreased more rapidly with higher carbon emissions. Interestingly, under the lowest carbon emission scenario SSP126, the SHAP value showed recovery in the last two periods. Notably, the SHAP value of secmb peaked under SSP245, indicating a non-linear relationship to carbon emissions. Initially, as carbon emissions rose, the SHAP values of GPP_S2 and NEP_S2 increased rapidly. However, in the more distant future, under some low carbon emission SSPs, their SHAP values decreased.

In Figure 3e-3h, the relationship between the SHAP values of GPP_S2, feature values, and SOC values is depicted. SHAP values tend to increase with feature values, plateauing thereafter with minimal change in contribution. Notably, around a feature value of approximately 3 gC $m^{-2}d^{-1}$, the SHAP values for different SSPs change sign from negative to positive. Beyond a GPP of about 7 gC $m^{-2}d^{-1}$, its positive contribution to SOC plateaus. Additionally, regions with a substantial negative contribution of GPP are primarily associated with data exhibiting low SOC values.

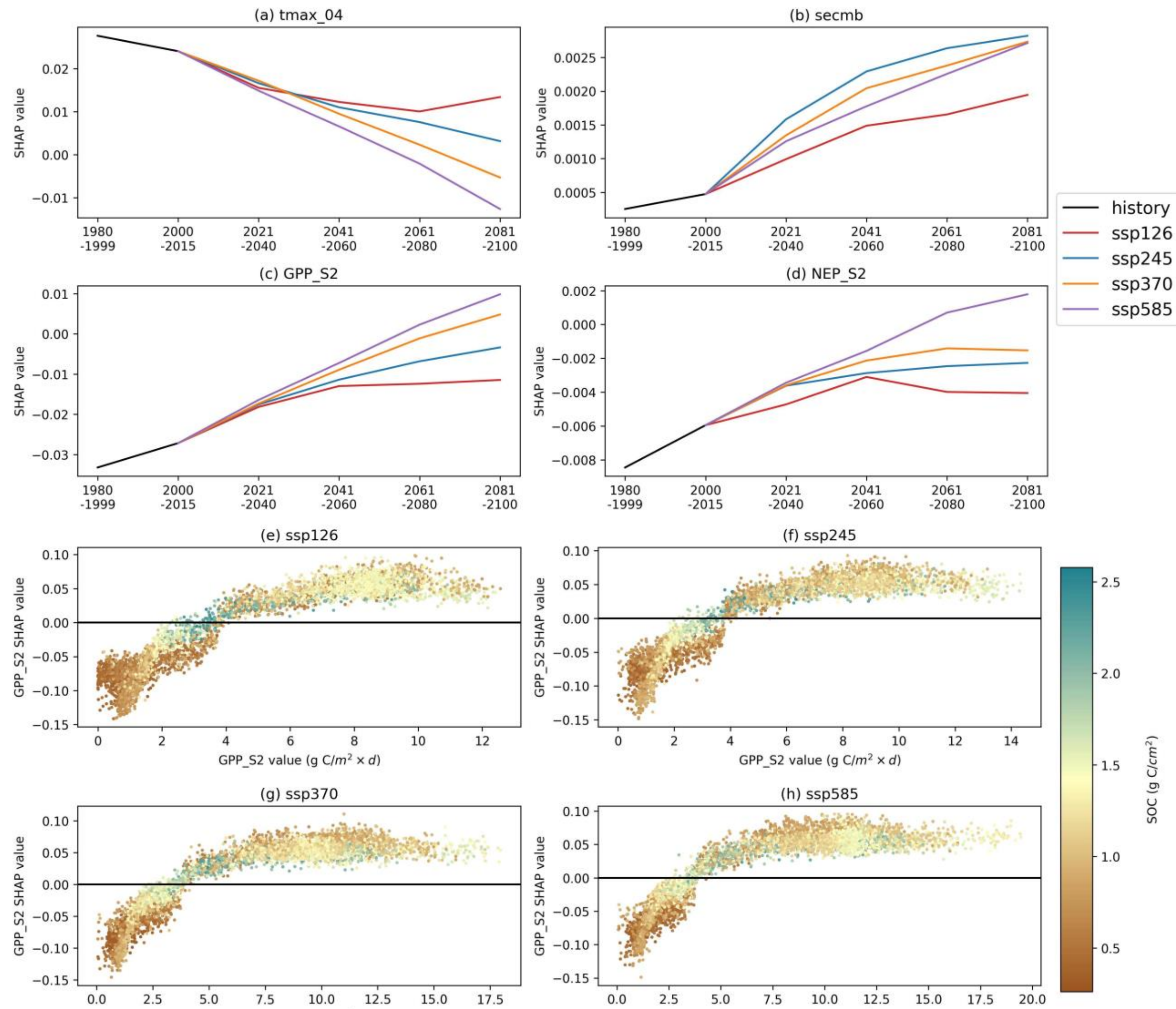


Figure 3. National average SHAP value changes for various variables (a-d) and SHAP values for four SSPs using EC-Earth3-Veg model outputs for 2081-2100 (e-h).

Figure 4 demonstrates that the contribution of GPP to SOC remained either negative or positive throughout the entire period in most regions of China. Interestingly, as carbon emissions increased, the areas undergoing a change from negative to positive contribution (depicted in red) expanded. Furthermore, the timing of this transition was progressively delayed from scenarios with lower to higher carbon emissions, indicating that in a future with higher carbon emissions, areas initially showing negative contributions may eventually shift to positive ones. Notably, the regions experiencing this sign change were mainly situated around the 400 mm annual precipitation line, emphasizing the importance of these zones as key areas for sequential soil carbon sequestration.

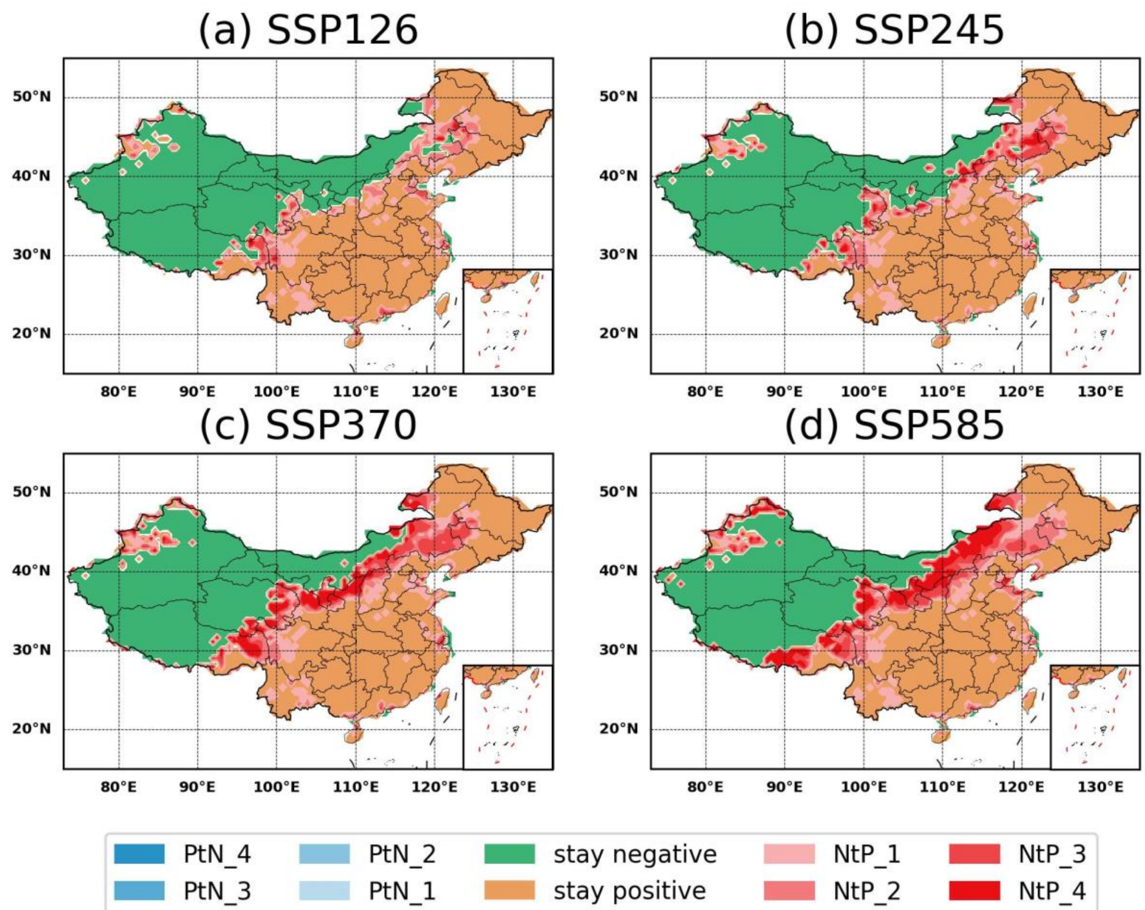


Figure 4. Sign shifts in GPP's contribution to SOC. Green signifies a consistent negative contribution, and orange indicates a consistent positive. Blue (PtN) denotes a shift from positive to negative, while red (NtP) indicates a shift from negative to positive. The numbers 1-4 represent the periods when the shift occurred: 1 for 2021-2040, 2 for 2041-2060, 3 for 2061-2080, and 4 for 2081-2100.

4 Discussions

Building upon the insights provided in our study, it is important to contextualize the projected increases in soil organic carbon (SOC) within the broader framework of global carbon cycling and climate change mitigation strategies. Our findings suggest that the carbon flux plays a pivotal role in determining SOC levels in China, with its impact varying across different SSPs.

Compared to other studies, the total soil organic carbon (SOC) stocks for historical periods reported in our study (Figure 1 and 2) are consistent with expected ranges (Liang et al., 2019; Li et al., 2022; Liu et al., 2022; Song et al., 2020; Yang et al., 2023; Zhang et al., 2023).

The increase in SOC under all SSPs, contrary to some studies predicting declines (Zhang et al., 2023), underscores the complexity of SOC dynamics and the need for comprehensive models that incorporate a wide array of variables. The carbon flux variables, which include both GPP and NEP, emerged as key drivers of SOC changes, potentially offsetting the negative effects of increased soil respiration due to rising temperatures.

Sensitivity analyses reveal that carbon flux is the variable with the most substantial impact on future SOC changes (Figure 2), with its promoting effect intensifying under SSPs with higher carbon emissions. This suggests that the CO_2 fertilization effect may continue to enhance SOC in environments with elevated CO_2 levels. This effect is particularly relevant in the context of global efforts to increase carbon sinks as a means to combat climate change. The diminishing negative impact of climate variables on SOC under SSP126 indicates that such effects may decrease only under scenarios with the lowest carbon emissions. Conversely, the positive influence of land use on SOC may weaken under SSP585, the scenario with the highest carbon emissions. Interestingly, the two ESMs with the largest total SOC stocks also exhibit the greatest reductions when carbon flux is held constant, suggesting that differences in SOC between ESMs may be partly due to the carbon flux simulation. It is generally accepted that warmer temperatures associated with climate change will increase soil respiration and reduce SOC, but some studies suggest that SOC may increase under conditions of higher atmospheric carbon dioxide due to increased carbon sequestration by vegetation (Terrer et al., 2021). Terrestrial carbon fluxes serve as a robust indicator of vegetation carbon sequestration. Figure S1 reveals that forthcoming enhancements in SOC predominantly stem from GPP and NEP. Notably, temperature emerges as the foremost adverse factor, exerting a significant negative influence by fostering the decomposition of soil carbon. The spatial analysis of GPP (Figure 4) reveals that areas with sign changes in its contribution to SOC are near the 400 mm precipitation line, indicates that precipitation patterns play a significant role in SOC sequestration. This finding has implications for land management practices, suggesting that regions with intermediate precipitation levels may be key targets for interventions aimed at increasing SOC stocks.

Furthermore, our study posits that GPP's contribution to SOC is negative at low values but becomes positive above a certain threshold. After reaching a peak, the positive contribution of GPP stabilizes and does not further increase. This finding implies that regions at the intersection of positive and negative contributions could enhance SOC accumulation through targeted interventions. Specifically, near the 400 mm precipitation line, vegetation restoration efforts could elevate GPP beyond the threshold, shifting its contribution from negative to positive. However, in areas already characterized by high GPP, additional planting may not yield further increases in SOC accumulation.

In our study, we opted to use carbon fluxes to represent the CO_2 fertilization effect rather than directly employing CO_2 concentration. Although we tested incorporating CO_2 concentration as a covariate in our machine learning model, it failed to accurately capture the fertilization effect. This discrepancy stems from the relatively minor spatial and seasonal changes in CO_2 concentration compared to the anticipated future increases, rendering the space-for-time substitution approach ineffective as it involves extrapolation. Consequently, our proposed method of utilizing carbon fluxes as covariates proves to be a valuable approach for addressing the CO_2 fertilization effect in machine learning models of SOC.

While our study provides valuable insights into the potential for SOC sequestration in China, it also highlights the inherent uncertainties in modeling the complex earth systems. The discrepancies between ESMs underscore the need for continued refinement of these tools and for the integration of diverse data sources to improve predictive accuracy. Additionally, the machine learning model trained on historical data face constraints in extrapolating future conditions due to potential alterations in the relationship between SOC and its influencing factors under climate change (Pickett, 1989). This constraint arises because changes in climate can introduce novel dynamics that may not be fully represented or captured by historical records, thereby impacting the predictive power of models for SOC behavior in a changing environment. Moreover, interpretive methods themselves may introduce additional uncertainty (Huang et al., 2023).

5 Conclusions

Our comprehensive study has provided valuable insights into the dynamics of SOC in China, projecting an overall increase in SOC stocks across various SSPs until the year 2100. This positive trend contrasts with some existing literature that anticipates declines in SOC under certain scenarios, highlighting the critical role of carbon flux, particularly GPP, in influencing SOC outcomes. Our findings underscore the significance of carbon flux as the most influential variable affecting future SOC changes, with its impact being more pronounced under higher carbon emission scenarios. This suggests that the CO2 fertilization effect may continue to play a vital role in enhancing SOC, even in high CO_2 environments. The spatial analysis within our study has revealed that areas near the 400 mm precipitation line are critical zones for SOC sequestration, indicating that precipitation patterns are key determinants in the carbon cycle. Our analyses identify thresholds in the GPP-SOC relationship, with GPP's contribution to SOC transitioning from negative to positive beyond a certain level. However, this positive contribution does not increase indefinitely, indicating a plateau effect that has important implications for land management and carbon sequestration strategies.

Despite the promising projections, our study acknowledges the inherent uncertainties associated with ESMs and the interpretation of complex environmental data. The variability between ESMs highlights the need for ongoing research and model refinement to enhance the accuracy of SOC predictions.

As the global community continues to seek solutions for climate change mitigation, understanding the factors that influence SOC is crucial for developing effective carbon management strategies. Our research contributes to this understanding by providing a nuanced view of the interactions between carbon fluxes, climate variables, land use, and SOC. It is our hope that these insights will inform future land management practices and policies aimed at maximizing the potential of soils as carbon sinks, thereby supporting global efforts to combat climate change and promote sustainable development.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Soil profile data from the Second National Soil Survey of China were derived from the China data set of soil properties for land surface modeling (Shangguan et al., 2013). China's terrestrial ecosystems data were derived from Xu et al. (2019). Relief data were calculated based on DEM derived from Multi-Error-Removed Improved Terrain DEM (Yamazaki et al., 2017). Landform data were from the European Soil Data Center (Iwahashi and Pike, 2007). Climate data were downloaded from the WorldClim version 2.1 (Fick & Hijmans, 2017). Landuse data were downloaded from the Land Use Harmonization project (Hurtt et al., 2020). Historical carbon flux data were derived from the Global Carbon Fluxes dataset (Shangguan et al., 2023), and future carbon flux data were derived from CMIP6 available in Earth System Grid Federation (ESGF, Lawrence Livermore National Laboratory, 2023). Soil maps were downloads from Shangguan et al. (2013) and the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012). The average soil and sedimentary deposit thickness map were download from (Pelletier, 2016).Rock type data was derived from USGS Geosciences and Environmental Change Science Center (GECSC) based on the Global Lithological Map database v1.1 (GLiM, Hartmann & Moosdorf, 2012). Bedrock depth data was downloaded from Yan et al. (2020).

Acknowledgments

The study was partially supported by the National Natural Science Foundation of China (42375144, 42088101 and U1811464), Guangdong Basic and Applied Basic Research Foundation (2021B0301030007) and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311020008).

References

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., … Wofsy, S. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Batjes, N. H. (2014). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, *65*(1). https://doi.org/10.1111/ejss.12114_2

Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T., & Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(23). https://doi.org/10.1073/pnas.1220228110

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., … Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, *12*(7). https://doi.org/10.1029/2019MS002010

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., & Walter, C. (2022). Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma*, *409*, 115567. https://doi.org/10.1016/j.geoderma.2021.115567

Crow, S. E., Lajtha, K., Filley, T. R., Swanston, C. W., Bowden, R. D., & Caldwell, B. A. (2009). Sources of plant-derived carbon and stability of organic matter in soil: Implications for global change. *Global Change Biology*, *15*(8). https://doi.org/10.1111/j.1365-2486.2009.01850.x

Davidson, E. A., & Janssens, I. A. (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, *440*(7081), 165–173. https://doi.org/10.1038/nature04514

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L. P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., … Zhang, Q. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development*, *15*(7). https://doi.org/10.5194/gmd-15-2973-2022

FAO/IIASA/ISRIC/ISS-CAS/JRC (2012), Harmonized World Soil Database (version1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria. http://dx.doi.org/10.3334/ORNLDAAC/1247

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12). https://doi.org/10.1002/joc.5086

Field, C. B. (2001). Plant Physiology of the “Missing” Carbon Sink. *Plant Physiology*, *125*(1), 25–28. https://doi.org/10.1104/pp.125.1.25

Grunwald, S. (2010). Current State of Digital Soil Mapping and What Is Next. In *Digital Soil Mapping* (pp. 3–12). Springer Netherlands. https://doi.org/10.1007/978-90-481-8863-5_1

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12). https://doi.org/10.1029/2012GC004370

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science*, *72*(4), 1607–1623. https://doi.org/10.1111/ejss.12998

Huang, F., Shangguan, W., Li, Q., Li, L., & Zhang, Y. (2023). Beyond prediction: An integrated post-hoc approach to interpret complex model in hydrometeorology. *Environmental Modelling and Software*, *167*. https://doi.org/10.1016/j.envsoft.2023.105762

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., … Zhang, X. (2020). Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6. *Geoscientific Model Development*, *13*(11). https://doi.org/10.5194/gmd-13-5425-2020

Intergovernmental Panel on Climate Change. (2023). Climate Change 2021 – The Physical Science Basis. In *Climate Change 2021 – The Physical Science Basis*. https://doi.org/10.1017/9781009157896

Ito, A., Hajima, T., Lawrence, D. M., Brovkin, V., Delire, C., Guenet, B., Jones, C. D., Malyshev, S., Materia, S., McDermid, S. P., Peano, D., Pongratz, J., Robertson, E., Shevliakova, E., Vuichard, N., Wårlind, D., Wiltshire, A., & Ziehn, T. (2020). Soil carbon sequestration simulated in CMIP6-LUMIP models: implications for climatic mitigation. *Environmental Research Letters*, *15*(12), 124061. https://doi.org/10.1088/1748-9326/abc912

Iwahashi, J., and R. J. Pike (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature, *Geomorphology*, 86(3), 409-440. https://doi.org/10.1016/j.geomorph.2006.09.012.

Karnosky, D. F. (2003). Impacts of elevated atmospheric CO2 on forest trees and forest ecosystems: Knowledge gaps. In *Environment International* (Vol. 29, Issues 2–3). https://doi.org/10.1016/S0160-4120(02)00159-9

Kuzyakov, Y. (2010). Priming effects: Interactions between living and dead organic matter. *Soil Biology and Biochemistry*, *42*(9). https://doi.org/10.1016/j.soilbio.2010.04.003

Lal, R. (2003). Soil erosion and the global carbon budget. In *Environment International* (Vol. 29, Issue 4). https://doi.org/10.1016/S0160-4120(02)00192-7

Lal, R., Negassa, W., & Lorenz, K. (2015). Carbon sequestration in soil. *Current Opinion in Environmental Sustainability*, *15*, 79–86. https://doi.org/10.1016/j.cosust.2015.09.002

Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, *352*, 395–413. https://doi.org/10.1016/j.geoderma.2019.05.031

Lawrence Livermore National Laboratory. (2023). Coupled Model Intercomparison Project 6 Dataset *[Dataset]*. Earth System Grid Federation (ESFG). Retrieved from https://aims2.llnl.gov/search

Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., Van Der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., … Zeng, N. (2013). The global carbon budget 1959-2011. *Earth System Science Data*, *5*(1). https://doi.org/10.5194/essd-5-165-2013

Liang, J., Zhou, Z., Huo, C., Shi, Z., Cole, J. R., Huang, L., Konstantinidis, K. T., Li, X., Liu, B., Luo, Z., Penton, C. R., Schuur, E. A. G., Tiedje, J. M., Wang, Y. P., Wu, L., Xia, J., Zhou, J., & Luo, Y. (2018). More replenishment than priming loss of soil organic carbon with additional carbon input. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05667-7

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., et al. (2022). Matrix approach to land carbon cycle modeling. Journal of Advances in Modeling Earth Systems, 14, e2022MS003008. https://doi.org/10.1029/2022MS003008

Liang, Z., Chen, S., Yang, Y., Zhao, R., Shi, Z., & Viscarra Rossel, R. A. (2019). National digital soil map of organic matter in topsoil and its associated uncertainty in 1980's China. *Geoderma*, *335*, 47–56. https://doi.org/10.1016/J.GEODERMA.2018.08.011

Li, H., Wu, Y., Liu, S., Xiao, J., Zhao, W., Chen, J., Alexandrov, G., & Cao, Y. (2022). Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Global Change Biology*, *28*(10). https://doi.org/10.1111/gcb.16154

Litton, C. M., & Giardina, C. P. (2008). Below-ground carbon flux and partitioning: global patterns and response to temperature. *Functional Ecology*, *22*(6), 941–954. https://doi.org/10.1111/j.1365-2435.2008.01479.x

Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J. L., Song, X., Shi, Z., Zhu, A. X., & Zhang, G. L. (2022). Mapping high resolution National Soil Information Grids of China. *Science Bulletin*, *67*(3). https://doi.org/10.1016/j.scib.2021.10.013

Liu, M., Han, G., & Zhang, Q. (2020). Effects of agricultural abandonment on soil aggregation, soil organic carbon storage and stabilization: Results from observation in a small karst catchment, Southwest China. *Agriculture, Ecosystems and Environment*, *288*. https://doi.org/10.1016/j.agee.2019.106719

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*.

Luo, Z., Wang, G., & Wang, E. (2019). Global subsoil organic carbon turnover times dominantly controlled by soil properties rather than climate. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11597-9

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., … Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4). https://doi.org/10.1029/2018MS001400

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McBratney, A., de Gruijter, J., & Bryce, A. (2019). Pedometrics timeline. *Geoderma*, *338*. https://doi.org/10.1016/j.geoderma.2018.11.048

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z. S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., … Winowiecki, L. (2017). Soil carbon 4 per mille. In *Geoderma* (Vol. 292). https://doi.org/10.1016/j.geoderma.2017.01.002

Nowak, R. S., Ellsworth, D. S., & Smith, S. D. (2004). Functional responses of plants to elevated atmospheric CO2 - Do photosynthetic and productivity data from FACE experiments support early predictions? In *New Phytologist* (Vol. 162, Issue 2). https://doi.org/10.1111/j.1469-8137.2004.01033.x

Odgers, N. P., Libohova, Z., & Thompson, J. A. (2012). Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma*, *189–190*, 153–163. https://doi.org/10.1016/j.geoderma.2012.05.026

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. In *SOIL* (Vol. 6, Issue 1). https://doi.org/10.5194/soil-6-35-2020

Patoine, G., Eisenhauer, N., Cesarz, S., Phillips, H. R. P., Xu, X., Zhang, L., & Guerra, C. A. (2022). Drivers and trends of global soil microbial carbon over two decades. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-31833-z

Pelletier, J.D., Broxton, P.D., Hazenberg, P., Zeng, X., Troch, P.A., Niu, G., Williams, Z.C., Brunke, M.A. & Gochis, D. 2016. Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1304

Pickett, S. T. A. (1989). Space-for-Time Substitution as an Alternative to Long-Term Studies. In *Long-Term Studies in Ecology*. https://doi.org/10.1007/978-1-4615-7358-6_5

Sayer, E. J., Heard, M. S., Grant, H. K., Marthews, T. R., & Tanner, E. V. J. (2011). Soil carbon release enhanced by increased tropical forest litterfall. *Nature Climate Change*, *1*(6). https://doi.org/10.1038/nclimate1190

Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., … Zhang, Y. (2013). A China data set of soil properties for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *5*(2), 212–224. https://doi.org/10.1002/jame.20026

Shangguan, W., Xiong, Z., Nourani, V., Li, Q., Lu, X., Li, L., Huang, F., Zhang, Y., Sun, W., & Dai, Y. (2023). A 1 km Global Carbon Flux Dataset Using In Situ Measurements and Deep Learning. *Forests*, *14*(5), 913. https://doi.org/10.3390/f14050913

Smith, P., Fang, C., Dawson, J. J. C., & Moncrieff, J. B. (2008). Impact of Global Warming on Soil Organic Carbon. In *Advances in Agronomy* (Vol. 97, pp. 1–43). https://doi.org/10.1016/S0065-2113(07)00001-6

Song, X. D., Wu, H. Y., Ju, B., Liu, F., Yang, F., Li, D. C., Zhao, Y. G., Yang, J. L., & Zhang, G. L. (2020). Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma*, *363*. https://doi.org/10.1016/j.geoderma.2019.114145

Terrer, C., Phillips, R. P., Hungate, B. A., Rosende, J., Pett-Ridge, J., Craig, M. E., van Groenigen, K. J., Keenan, T. F., Sulman, B. N., Stocker, B. D., Reich, P. B., Pellegrini, A. F. A., Pendall, E., Zhang, H., Evans, R. D., Carrillo, Y., Fisher, J. B., Van Sundert, K., Vicca, S., & Jackson, R. B. (2021). A trade-off between plant and soil carbon storage under elevated CO2. *Nature*, *591*(7851). https://doi.org/10.1038/s41586-021-03306-8

Trivedi, P., Singh, B. P., & Singh, B. K. (2018). Soil carbon: Introduction, importance, status, threat, and mitigation. In *Soil Carbon Storage: Modulators, Mechanisms and Modeling*. https://doi.org/10.1016/B978-0-12-812766-7.00001-9

Xu, L., He, N., & Yu, G. (2019). A dataset of carbon density in Chinese terrestrial ecosystems (2010s). *China Scientific Data*, *1*, 4. https://doi.org/10.11922/csdata.2018.0026.zh

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., & Bates, P. D. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874

Yan, F., Shangguan, W., Zhang, J., & Hu, B. (2020). Depth-to-bedrock map of China at a spatial resolution of 100 meters. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-019-0345-6

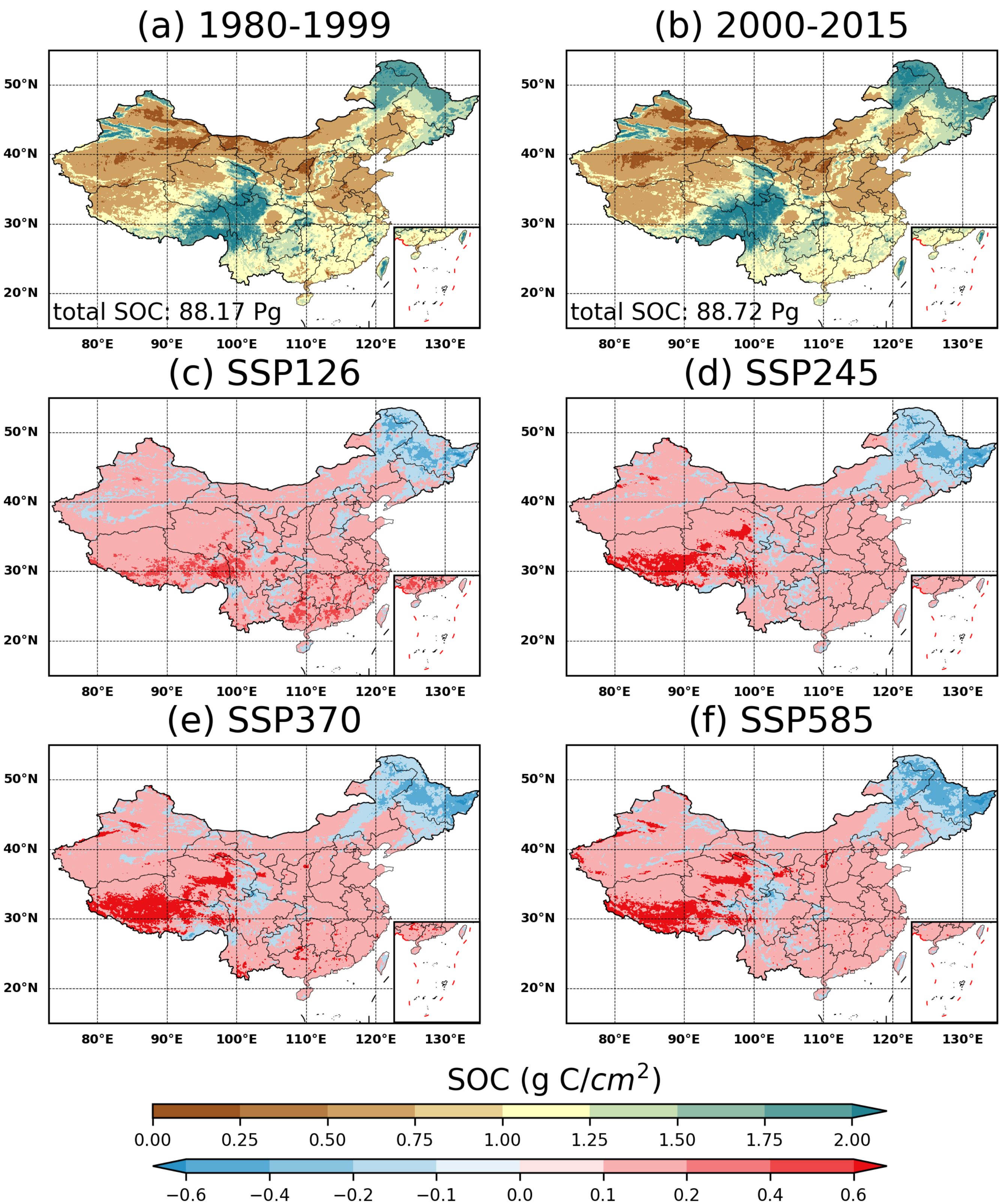
Yang, R. M., Huang, L. M., Zhang, X., Zhu, C. M., & Xu, L. (2023). Mapping the distribution, trends, and drivers of soil organic carbon in China from 1982 to 2019. *Geoderma*, *429*. https://doi.org/10.1016/j.geoderma.2022.116232

Yigini, Y., & Panagos, P. (2016). Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Science of The Total Environment*, *557–558*, 838–850. https://doi.org/10.1016/j.scitotenv.2016.03.085

Zhang, Z., Ding, J., Zhu, C., Wang, J., Ge, X., Li, X., Han, L., Chen, X., & Wang, J. (2023). Historical and future variation of soil organic carbon in China. *Geoderma*, *436*, 116557. https://doi.org/10.1016/j.geoderma.2023.116557

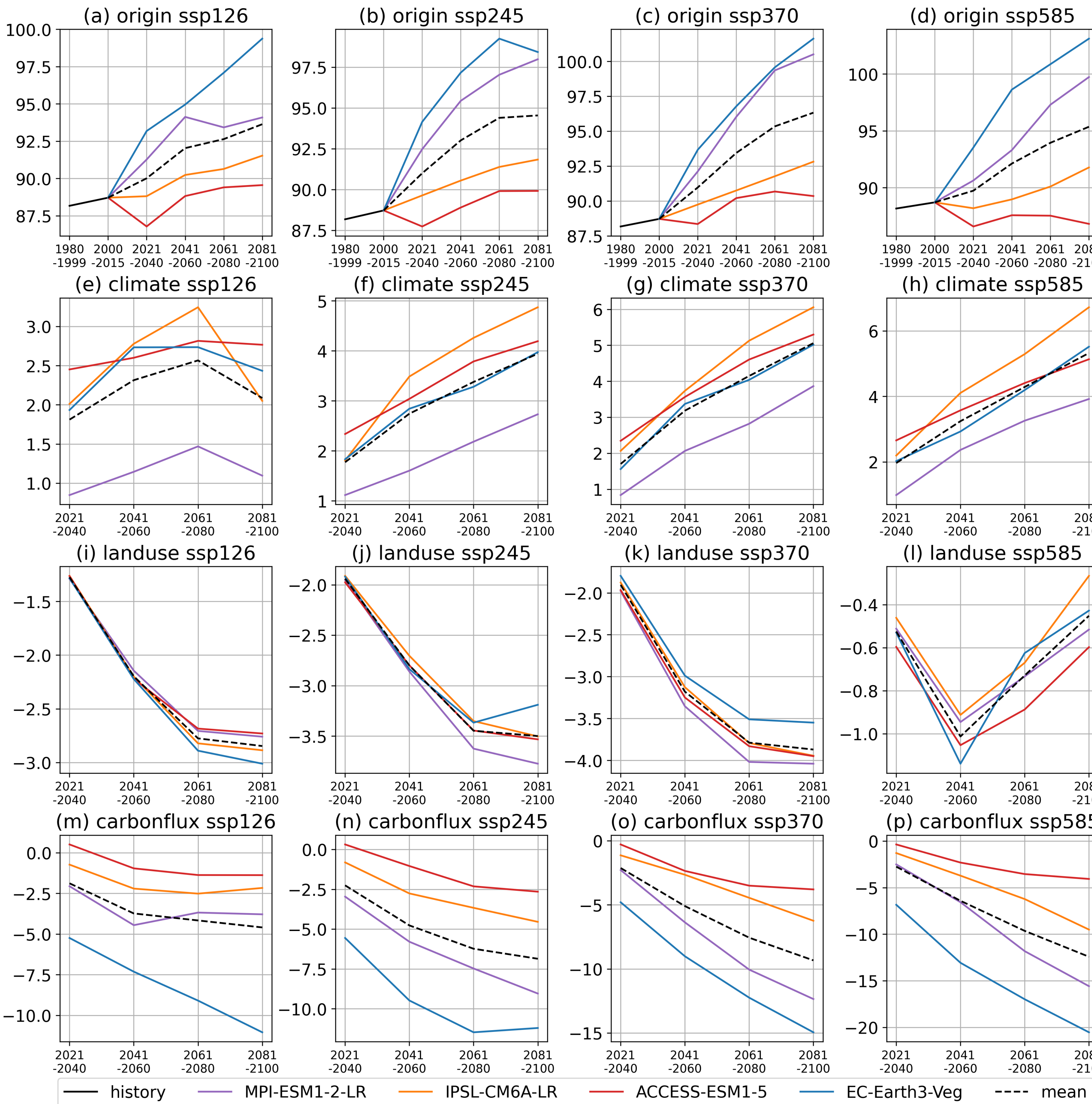
Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y. P., & Srbinovsky, J. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1). https://doi.org/10.1071/ES19035

Figure 1.



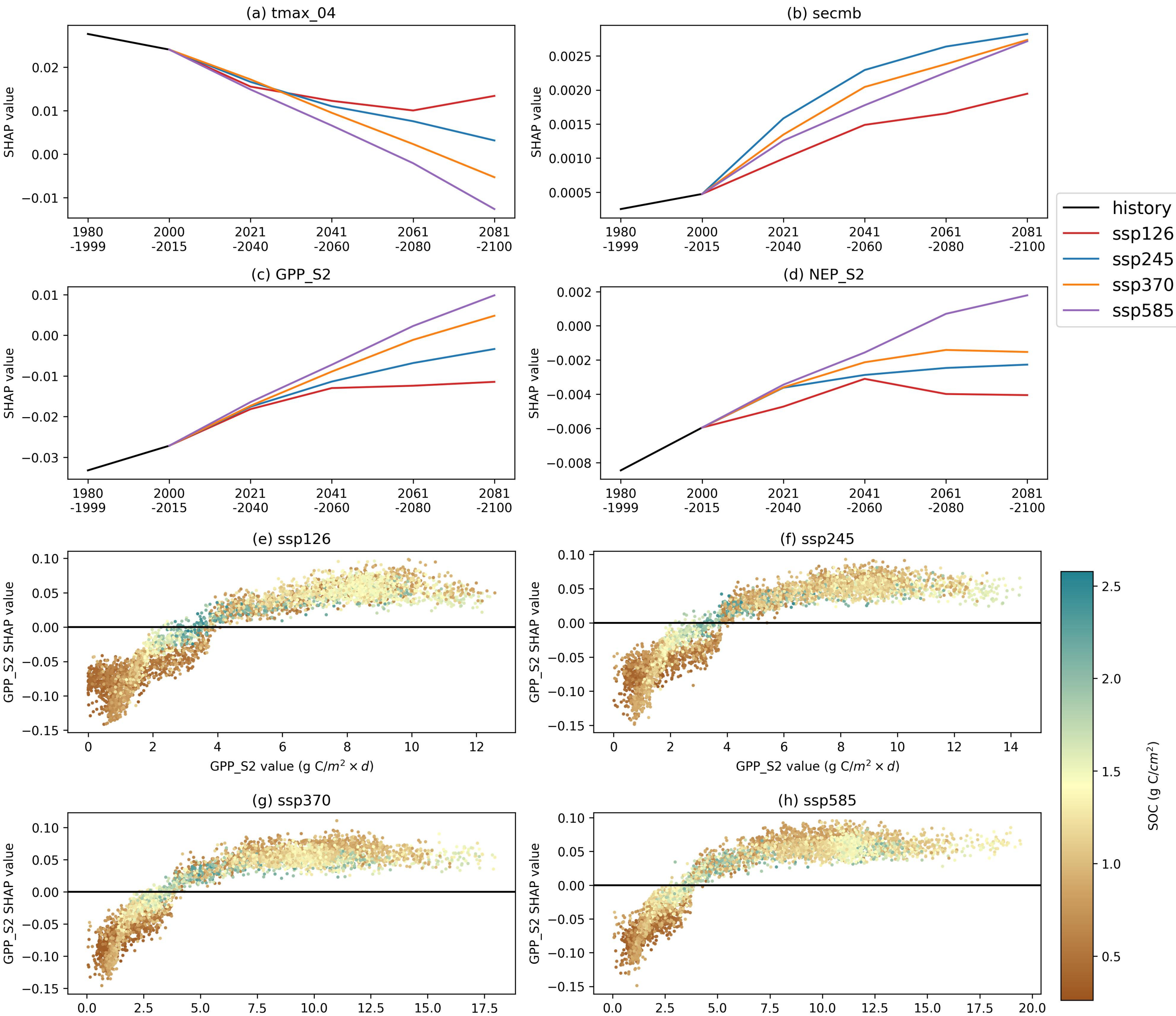
(a) 1980-1999
(b) 2000-2015
total SOC: 88.17 Pg
total SOC: 88.72 Pg
(c) SSP126
(d) SSP245
(e) SSP370
(f) SSP585
50°N
40°N
30°N
20°N
80°E
90°E
100°E
110°E
120°E
130°E
SOC (g C/cm^2)
0.00
0.25
0.50
0.75
1.00
1.25
1.50
1.75
2.00
−0.6
−0.4
−0.2
−0.1
0.0
0.1
0.2
0.4
0.6

Figure 2.



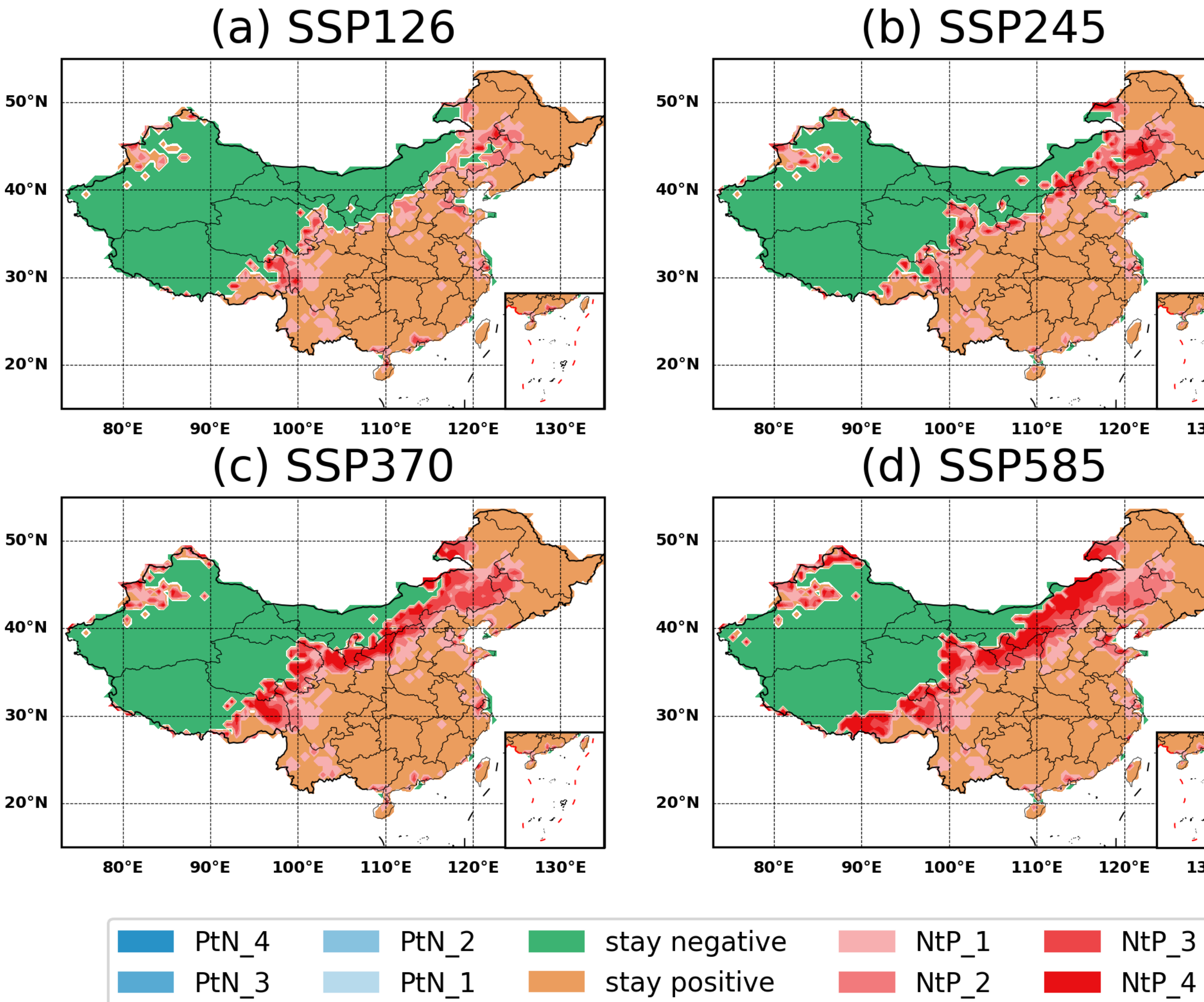
(a) origin ssp126
(b) origin ssp245
(c) origin ssp370
(d) origin ssp585
(e) climate ssp126
(f) climate ssp245
(g) climate ssp370
(h) climate ssp585
(i) landuse ssp126
(j) landuse ssp245
(k) landuse ssp370
(l) landuse ssp585
(m) carbonflux ssp126
(n) carbonflux ssp245
(o) carbonflux ssp370
(p) carbonflux ssp585
1980 -1999
2000 -2015
2021 -2040
2041 -2060
2061 -2080
2081 -2100
history
MPI-ESM1-2-LR
IPSL-CM6A-LR
ACCESS-ESM1-5
EC-Earth3-Veg
mean

Figure 3.



(a) tmax_04
(b) secmb
(c) GPP_S2
(d) NEP_S2
(e) ssp126
(f) ssp245
(g) ssp370
(h) ssp585
SHAP value
GPP_S2 SHAP value
GPP_S2 value (g C/$m^2 \times d$)
SOC (g C/cm^2)
1980 -1999
2000 -2015
2021 -2040
2041 -2060
2061 -2080
2081 -2100
history
ssp126
ssp245
ssp370
ssp585

Figure 4.


(a) SSP126
(b) SSP245
(c) SSP370
(d) SSP585
50°N
40°N
30°N
20°N
80°E
90°E
100°E
110°E
120°E
130°E
PtN_4
PtN_3
PtN_2
PtN_1
stay negative
stay positive
NtP_1
NtP_2
NtP_3
NtP_4

Projections of Soil Organic Carbon in China: The Role of Carbon Fluxes Revealed by Explainable Artificial Intelligence

Yongkun Zhang[1], Feini Huang[1], Xingjie Lu[1], Wei Shangguan[1], Gaosong Shi[1], Ye Zhang[1], Zhangcai Qin[1], Zhongwang Wei[1], Huan Yuan[1], and Yongjiu Dai[1]

[1]Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai), Guangdong Province Key Laboratory for Climate Change and Natural Disaster Studies, School of Atmospheric Sciences, Sun Yat-Sen University, Zhuhai, 519082, China.

Corresponding author: Wei Shangguan (shgwei@mail.sysu.edu.cn)

Key Points:

- The influence of carbon flux on SOC is more pronounced than that of climate change and land use change
- We identify two critical thresholds in the relationship between gross primary production and SOC
- Critical zones for soil carbon sequestration are located around 400 mm annual precipitation line

Abstract

The impact of carbon fluxes on soil organic carbon (SOC) remains underexplored. We employed machine learning to model SOC dynamics. Our findings project an increase in China's SOC through to the year 2100 across various Shared Socioeconomic Pathways. Sensitivity analyses have identified carbon fluxes as the main drivers for this projected rise, followed by climate and land use. Further examination using an explainable artificial intelligence method, Shapley Additive Explanations, has uncovered both spatial and temporal variations in how gross primary production (GPP) influences SOC levels. Notably, GPP's contribution on SOC is initially negative at low levels, turning positive once a threshold of approximately 3 gC $m^{-2}d^{-1}$ is surpassed. Beyond a GPP of about 7 gC $m^{-2}d^{-1}$, its positive contribution to SOC plateaus. Critical zones for soil carbon sequestration are located around 400 mm annual precipitation line.

Plain Language Summary

Soil's ability to absorb carbon is key to reducing atmospheric carbon dioxide, a major greenhouse gas. Yet, the influence of carbon fluxes—the exchange of carbon between the soil and the atmosphere—on soil carbon storage is not well understood. Our study utilized machine learning to estimate potential soil carbon storage in China by 2100, considering various global socioeconomic trajectories. We anticipate an uptick in soil carbon, largely due to carbon fluxes, with climate and land use changes also playing significant roles. Through explainable artificial intelligence, we've gained insights into how plant growth impacts soil carbon levels. We discovered that minimal plant growth correlates with lower soil carbon storage. As plants grow more, they enhance soil carbon storage until reaching a certain growth level, after which the effect plateaus. Zones critical for maximizing soil carbon storage correspond with areas receiving about 400 mm of rainfall annually. This understanding of plant growth's effect on soil carbon is invaluable for developing land management strategies aimed at maximizing carbon sequestration, thereby contributing to climate change mitigation efforts.

1 Introduction

Soil Organic Carbon (SOC) is a fundamental constituent of terrestrial ecosystems, performing an essential function in bolstering the resilience and productivity of ecosystems (Batjes, 2014; Lal, 2003; Minasny et al., 2017). SOC is not only crucial for providing nutrients that support plant growth and yield but also for retaining water and mitigating soil erosion (Trivedi et al., 2018). Even slight changes in the soil carbon pool can result in significant impacts on atmospheric carbon (Smith et al., 2008). Soil carbon sequestration, through its ability to capture and retain environmental carbon, acts as a powerful antidote against the intensification of the greenhouse effect (Lal et al., 2015). Therefore, predicting future SOC and identifying its key drivers are essential for understanding the evolving patterns of carbon stock distribution over time.

Methods for studying SOC are generally categorized into two types: process-based models and empirical models such as AI (Artificial Intelligence) methods. Process-based models simulate SOC dynamics based on detailed representations of internal biochemical and physical processes (Le Quéré et al., 2013). The Earth System Model (ESM) is an example of such a model, integrating carbon cycle processes with climate models (Intergovernmental Panel on Climate Change, 2023). These models are capable of projecting SOC distribution and temporal changes. However, due to the still uncertain physio-ecological mechanisms of SOC in terrestrial system, different ESMs have shown discrepancies in both historical and future SOC estimations (Ito et al., 2020).

Recently, AI methods have become powerful tools to for mapping and predicting SOC (McBratney et al., 2019). The SCORPAN framework (McBratney et al., 2003), introduced for Digital Soil Mapping (DSM), suggests that soil types or properties can be inferred from a combination of environmental factors (i.e., covariates). These include soil, climate, organisms, topography, parent material, age, spatial location, and other environmental variables (Chen et al., 2022; Lamichhane et al., 2019). The application of DSM technology to project future SOC changes relies on the space-for-time substitution concept (Pickett, 1989), which has been employed to anticipate SOC trends in regions such as Europe, China and Argentina (Heuvelink et al., 2021; Yigini & Panagos, 2016; Zhang et al., 2023). Among the various methods, Random

Forest (RF) has emerged as the most popular method for SOC mapping and prediction, demonstrating its effectiveness in this domain (Lamichhane et al., 2019; Padarian et al., 2020).

Significant research has been conducted on the anticipated changes in SOC, with climate change and land use change commonly recognized as the primary factors influencing future SOC variability (Davidson & Janssens, 2006). SOC are controlled by both carbon input and residence time (Luo et al, 2022). However, the role of carbon fluxes in shaping SOC dynamics has not been thoroughly investigated. The CO2 fertilization effect suggests that as atmospheric CO2 concentrations increase, carbon fluxes to ecosystems also rise (Baldocchi et al., 2001; Litton & Giardina, 2008). Yet, this additional carbon input may also enhance SOC decomposition, potentially leading to increased SOC loss (Crow et al., 2009; Kuzyakov, 2010; Sayer et al., 2011). Consequently, it remains uncertain whether such fertilization will result in soils becoming net carbon sources or sinks in the future (Field, 2001; Karnosky, 2003; Nowak et al., 2004; Liang et al., 2018).

Explainable Artificial Intelligence (XAI) has been successfully applied to attribute analysis in soil carbon studies (Luo et al., 2019; Patoine et al., 2022). To dissect the impact of various factors on SOC, with a focus on the influence of carbon flux, we integrated two XAI methods into our analysis: Random Forest Importance (RFI) and Shapley Additive Explanations (SHAP) (Huang et al., 2023). These methods will allow us to unravel the complex interactions between carbon flux and SOC, providing a clearer understanding of their relationship. This study seeks to elucidate three essential scientific questions: (1) What degree of variation in SOC levels can be expected in China from 2021 to 2100 under multiple Shared Socioeconomic Pathways (SSPs)? (2) What is the relative contribution of carbon fluxes to changes in SOC compared to the effect of climate change and land use change? (3) How will carbon fluxes shape the trajectory of SOC in the future?

2 Materials and Methods

2.1 Materials and Processing

In this study, we utilized a dataset comprising 8,979 soil profile records from the Second National Soil Survey of China, with most soil profile data being collected between 1979 and 1984 (Shangguan et al., 2013). SOC data from 2000 to 2014 were obtained from the carbon density dataset for China's terrestrial ecosystems (Xu et al., 2019). We focused our analysis on data from the 0-100 cm soil layer, converting the profile data to SOC density using a specified equation:

$$SOC = SOM \times 0.58 \times D \times BD \times (1 - G) \quad (1)$$

where SOM is soil organic matter, D denotes soil layer depth, BD is bulk density and G is gravel content (>2 mm). To standardize the depth of the data, we employed equal-area second-order spline interpolation (Odgers et al., 2012).

All covariates were resampled to a uniform resolution of 2.5 arc-minutes. The input covariates for our analysis were divided into static and dynamic categories, with a comprehensive list provided in Table S1. For static variables, such as those derived from the digital elevation model (DEM), we operated under the assumption that relief and parent material factors would remain constant over long-time scales, and that soil factors would undergo minimal changes over the span of hundreds of years (Grunwald, 2010).

The dynamic datasets were divided into historical and future periods for analysis. Historical data were synchronized with the timing of soil profile collection and assessed for each subsequent twenty-year interval. Covariate data from the periods 1980-1999 and 2000-2015 were aligned with the soil profile data. For projections into the future, we sourced data from four ESMs (ACCESS-ESM1-5 (Ziehn et al., 2020), EC-Earth3-Veg (Döscher et al., 2022), IPSL-CM6A-LR (Boucher et al., 2020), and MPI-ESM1-2-LR (Mauritsen et al., 2019)) under four Shared Socioeconomic Pathways (SSPs), for each twenty-year segment extending from 2021 to 2100, as only these ESMs provide outputs of carbon fluxes in the Coupled Model Intercomparison Project Phase 6 (CMIP6). The dynamic covariates encompassed climate variables, land use patterns, and carbon fluxes. Climate data, which included monthly maximum and minimum temperatures and precipitation, were sourced from the WorldClim2 database (Fick & Hijmans, 2017). Land use information was provided by the Land Use Harmonization project (LUH2) (Hurtt et al., 2020), covering four SSP scenarios（SSP126, SSP248, SSP370, and SSP585) as well as historical periods. Carbon flux data were represented by two key variables: gross primary productivity (GPP) and net ecosystem productivity (NEP). Historical carbon flux data were acquired from the Global Carbon Fluxes dataset (GCFD, (Shangguan et al., 2023)), while future data were processed from the corresponding ESM data. Given the for analysis. We subsequently employed the following formula to calculate future carbon fluxes:

$$CF_{i,j}^{*} = CF_{history} + (CF_{i,j} - CF_{history,j}) \tag{2}$$

where CF denotes carbon flux covariates, i denotes each period in the future, and j denotes each ESM.

2.2 Model Building and Prediction

We developed RF models to predict SOC using data from two historical periods: 1980-1999 and 2000-2015. The regression matrix was constructed using soil profiles and covariates. After eliminating correlated variables, we selected the thirty most influential variables for inclusion in the model. The model's accuracy was evaluated using ten-fold cross-validation and three indicators: R^2, Root Mean Square Error (RMSE), and Mean Absolute Error (MAE). Our experiments spanned six timeframes: the historical periods of 1980-1999 and 2000-2015, and the future period of 2021-2100, divided into four twenty-year intervals. Following the space-for-time substitution strategy (Blois et al., 2013; Liu et al., 2020), we treated the dynamic covariates from these six periods as sequential input data for the model. This approach allowed us to map historical SOC levels and project future SOC across different Earth System Models (ESMs) and Shared Socioeconomic Pathways (SSPs).

2.3 Attribution Analysis

We conducted a sensitivity analysis as follows. Initially, we held each of the three types of dynamic covariates—climate, land use, and carbon flux—at their historical values from the period 2000-2015 and then projected future SOC levels. By comparing these projections with the original predictions, we evaluated the influence of each covariate type on future SOC levels.

To delve deeper into the impact of specific variables on SOC, we employed XAI tools—RFI and SHAP—for detailed attribution analysis. RFI gauges the significance of variables in tree-based models by aggregating the decrease in entropy across all trees (Breiman, 2001). A variable that effectively partitions the data and substantially lowers entropy is deemed crucial for prediction. For robustness, importance values were averaged over 10 iterations. SHAP, conceptualized by Lundberg and Lee (Lundberg & Lee, 2017) based on game theory principles, computes the marginal contributions of each feature, treating feature values as players in a coalition. Given the current set of feature value, the estimated SHAP value is the contribution of a feature value to the difference between the actual prediction and the mean prediction. A low absolute value signifies that the impact of a feature on the deviation from the mean prediction is relatively minor. The sign (+/-) does not denote a positive or negative feedback mechanism; rather, it indicates whether the effect of the feature increases or decreases the deviation from the mean. For each input variable, we generated SHAP value maps analogous to the SOC predictions. Due to the intensive computational demands of SHAP, we consolidated the input covariates into 10x10 patches for analysis.

3 Results

3.1 Model Performance and Predictions

The performance metrics for our RF model indicated an R^2 of 0.41, an RMSE of 0.30 gC cm^{-2}, and a MAE of 0.22 gC cm^{-2}. By incorporating covariates from various time periods into the model, we were able to map SOC for historical periods and project SOC for future periods. During the two historical periods analyzed, regions with high SOC values were predominantly located in the mountainous area of Northeast China and the eastern part of the Qinghai-Tibet Plateau (Figure 1a and 1b).

We assessed the changes in SOC relative to the historical period (2000-2015) for each subsequent time interval (Figure 1c-1f). The spatial distribution patterns indicated a modest increase in SOC across most of the country in the future, although the regions experiencing significant increases were fewer under the lower carbon emission scenarios. Notably, SOC declines were primarily observed in the northeastern areas, while increases were concentrated in the eastern and southern regions of the Qinghai-Tibet Plateau. Under the SSP585 scenario (Figure 1f), the northeastern region experienced a more pronounced decrease in SOC compared to other SSPs, and the areas of increase were noticeably smaller than those under SSP370 (Figure 1e). This disparity may contribute to the overall lower SOC projections for SSP585 relative to SSP370.

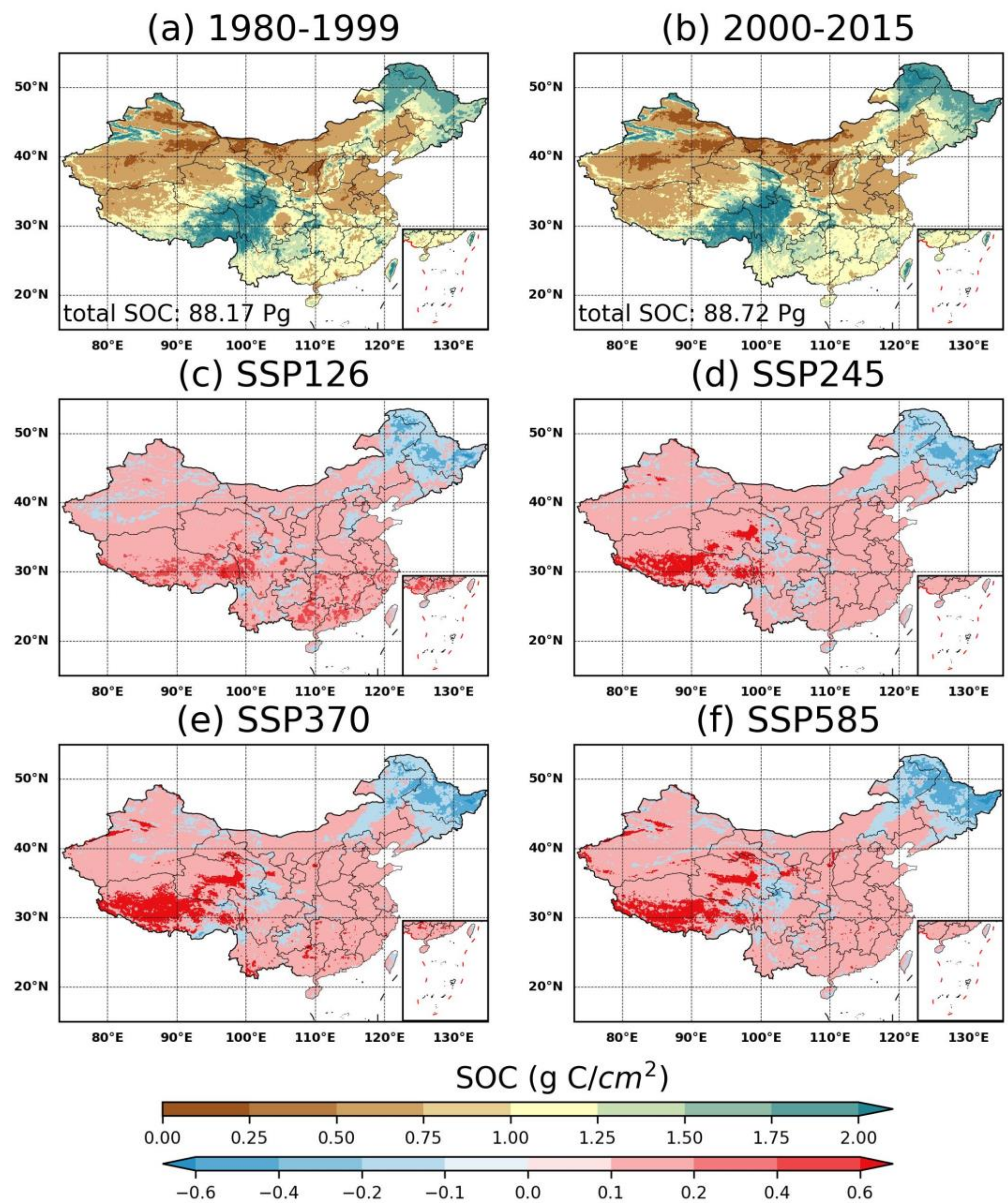


Figure 1. Maps depicting the distribution of soil organic carbon (SOC) density during the historical period and the projected changes in SOC density from 2000-2015 to 2081-2100 under various Shared Socioeconomic Pathways (SSPs). The data are averaged across multiple Earth System Models (ESMs).

3.2 Sensitivity Analysis

The aggregated results from four ESMs indicated an upward trend in total SOC stock across different SSPs (Figure 2a-2c), with the most substantial increase in SOC observed under SSP370 (Figure 2c). A notable variation was found in the SOC estimates produced by the different ESMs. Specifically, the SOC values from ACCESS-ESM1-5 and IPSL-CM6A-LR were markedly lower than those from the other two models, with SOC even showing a decline when compared to historical levels for ACCESS-ESM1-5.

The influences of climate, land use and carbon fluxes on future SOC were discerned by comparing the differences between the results obtained after holding these three types of variables constant and the original predictions (Figure 2e-2p). The discrepancy attributed to carbon fluxes was the greatest among the three sets of results and their positive effects on SOC increased as carbon emission increased. When carbon fluxes or land use were held constant, the projections were lower than the original predictions, whereas the fixed climate variables resulted in higher projections. This suggests that carbon flux and land use are likely to have a positive impact on future SOC, while climate variables may exert a negative influence due to faster soil decomposition. For climate variables, the difference under SSP126 initially increased and then decreased, implying that the adverse effects of climate on SOC first intensified and then diminished under the low carbon emission scenario. This phenomenon can be attributed to the complex interplay between warming effects: while it can lengthen the growing season and enhance productivity, it accelerates the decomposition rate of SOC. Since all ESMs utilized the same land use data, their results were consistent. However, as time advanced, the disparities among the different SSPs grew more pronounced. Notably, under SSP585, the variation due to land use was significantly less than under the other SSPs, suggesting that land use had a minimal impact on SOC changes under this pathway, with the other two variable types being more influential. Furthermore, the trend in SSP585 revealed that the positive contribution of land use initially rose and then fell over time.

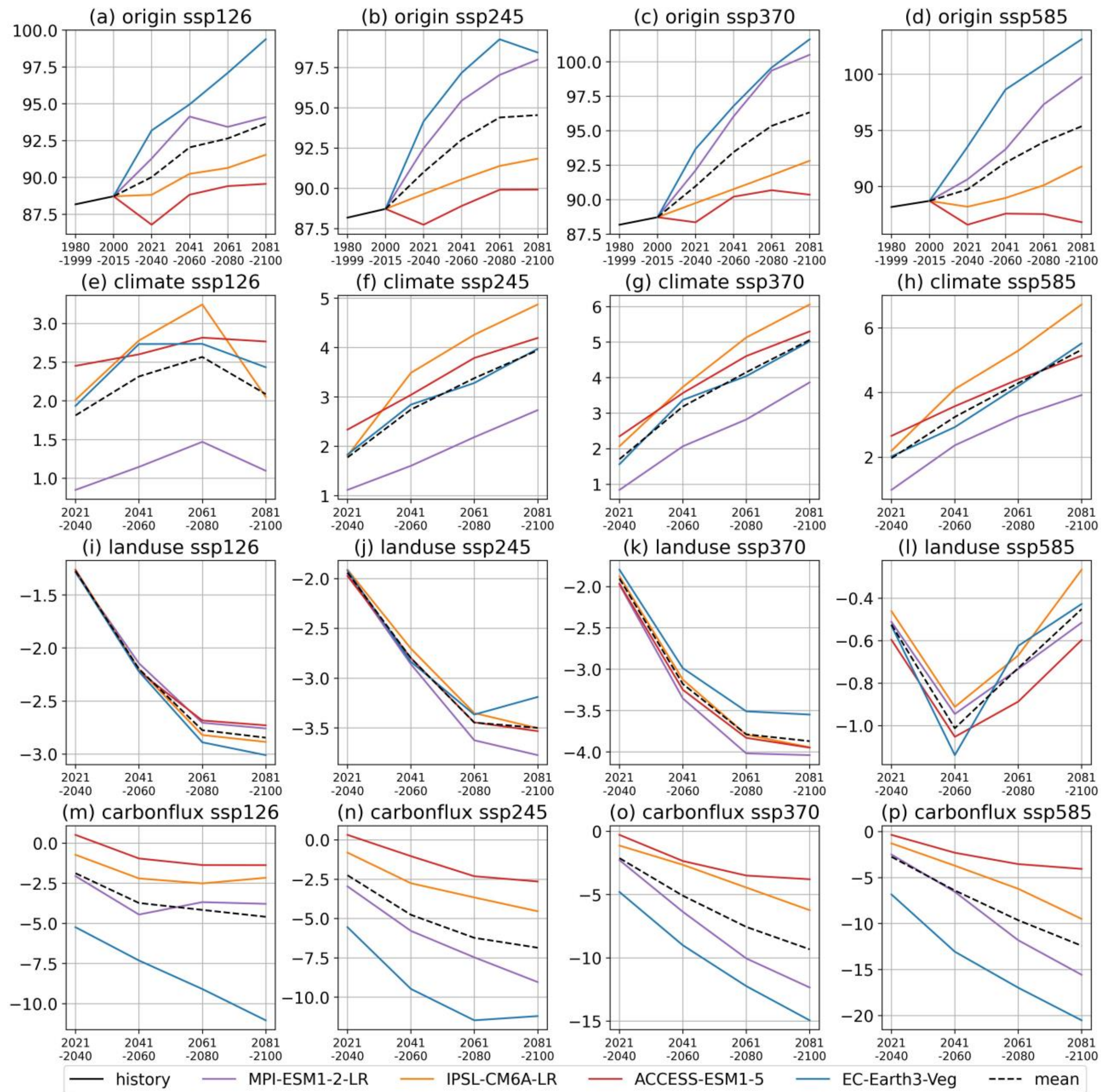


Figure 2. The temporal evolution of total SOC stock in peta-grams of carbon (Pg C) and discrepancies arising from holding specific variables constant. (a-d) the original predictions of SOC stock; variations resulting from fixing (e-h) climate, (i-l) land use and (m-p) carbon fluxes.

3.3 Attribution Analysis with XAI

Figure S1 displays RFI and SHAP values for various variables, featuring the top two variables in each category based on RFI for brevity. The results are based on outputs from EC-Earth3-Veg under SSP585, with similar observations across other ESMs and SSPs (not shown).

The April maximum temperature (tmax_04) emerged as the most important variable, followed by DEM and carbon flux variables. Notably, the SHAP values of certain variables exhibited temporal changes. Specifically, the SHAP values for tmax_04 and the carbon flux variables underwent inversion over time, indicating a reversal in their contributions to SOC. In the period 2081-2100, tmax_04 transitioned from positive contribution to negative contribution, while GPP of summer (GPP_S2) exhibited the opposite trend.

In Figure 3a-3d, the tmax_04 variable exhibited a gradual decrease in SHAP over time, while secmb (secondary mean biomass density), GPP_S2, and NEP_S2 showed a progressive increase. The SHAP of tmax_04 decreased more rapidly with higher carbon emissions. Interestingly, under the lowest carbon emission scenario SSP126, the SHAP value showed recovery in the last two periods. Notably, the SHAP value of secmb peaked under SSP245, indicating a non-linear relationship to carbon emissions. Initially, as carbon emissions rose, the SHAP values of GPP_S2 and NEP_S2 increased rapidly. However, in the more distant future, under some low carbon emission SSPs, their SHAP values decreased.

In Figure 3e-3h, the relationship between the SHAP values of GPP_S2, feature values, and SOC values is depicted. SHAP values tend to increase with feature values, plateauing thereafter with minimal change in contribution. Notably, around a feature value of approximately 3 gC $m^{-2}d^{-1}$, the SHAP values for different SSPs change sign from negative to positive. Beyond a GPP of about 7 gC $m^{-2}d^{-1}$, its positive contribution to SOC plateaus. Additionally, regions with a substantial negative contribution of GPP are primarily associated with data exhibiting low SOC values.

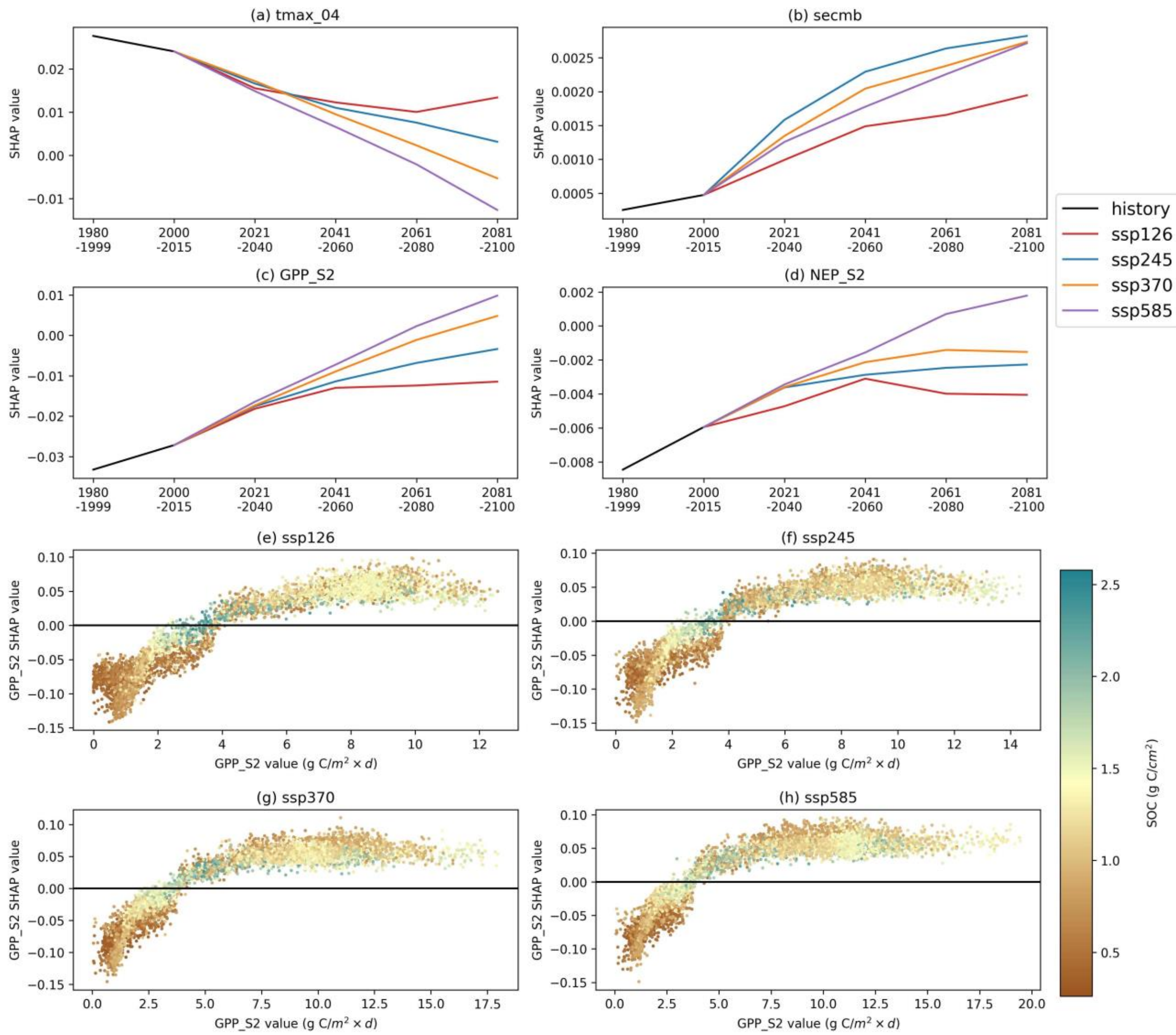


Figure 3. National average SHAP value changes for various variables (a-d) and SHAP values for four SSPs using EC-Earth3-Veg model outputs for 2081-2100 (e-h).

Figure 4 demonstrates that the contribution of GPP to SOC remained either negative or positive throughout the entire period in most regions of China. Interestingly, as carbon emissions increased, the areas undergoing a change from negative to positive contribution (depicted in red) expanded. Furthermore, the timing of this transition was progressively delayed from scenarios with lower to higher carbon emissions, indicating that in a future with higher carbon emissions, areas initially showing negative contributions may eventually shift to positive ones. Notably, the regions experiencing this sign change were mainly situated around the 400 mm annual precipitation line, emphasizing the importance of these zones as key areas for sequential soil carbon sequestration.

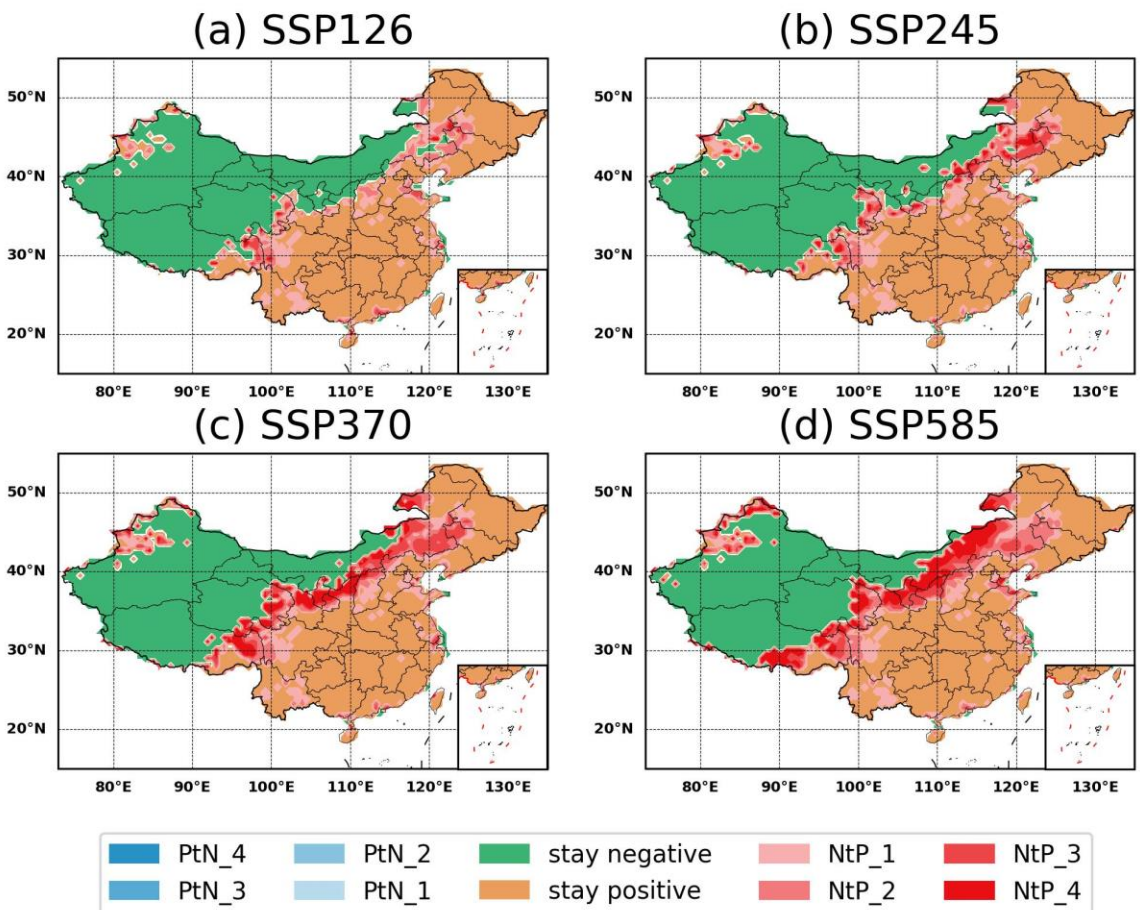


Figure 4. Sign shifts in GPP's contribution to SOC. Green signifies a consistent negative contribution, and orange indicates a consistent positive. Blue (PtN) denotes a shift from positive to negative, while red (NtP) indicates a shift from negative to positive. The numbers 1-4 represent the periods when the shift occurred: 1 for 2021-2040, 2 for 2041-2060, 3 for 2061-2080, and 4 for 2081-2100.

4 Discussions

Building upon the insights provided in our study, it is important to contextualize the projected increases in soil organic carbon (SOC) within the broader framework of global carbon cycling and climate change mitigation strategies. Our findings suggest that the carbon flux plays a pivotal role in determining SOC levels in China, with its impact varying across different SSPs.

Compared to other studies, the total soil organic carbon (SOC) stocks for historical periods reported in our study (Figure 1 and 2) are consistent with expected ranges (Liang et al., 2019; Li et al., 2022; Liu et al., 2022; Song et al., 2020; Yang et al., 2023; Zhang et al., 2023).

The increase in SOC under all SSPs, contrary to some studies predicting declines (Zhang et al., 2023), underscores the complexity of SOC dynamics and the need for comprehensive models that incorporate a wide array of variables. The carbon flux variables, which include both GPP and NEP, emerged as key drivers of SOC changes, potentially offsetting the negative effects of increased soil respiration due to rising temperatures.

Sensitivity analyses reveal that carbon flux is the variable with the most substantial impact on future SOC changes (Figure 2), with its promoting effect intensifying under SSPs with higher carbon emissions. This suggests that the CO_2 fertilization effect may continue to enhance SOC in environments with elevated CO_2 levels. This effect is particularly relevant in the context of global efforts to increase carbon sinks as a means to combat climate change. The diminishing negative impact of climate variables on SOC under SSP126 indicates that such effects may decrease only under scenarios with the lowest carbon emissions. Conversely, the positive influence of land use on SOC may weaken under SSP585, the scenario with the highest carbon emissions. Interestingly, the two ESMs with the largest total SOC stocks also exhibit the greatest reductions when carbon flux is held constant, suggesting that differences in SOC between ESMs may be partly due to the carbon flux simulation. It is generally accepted that warmer temperatures associated with climate change will increase soil respiration and reduce SOC, but some studies suggest that SOC may increase under conditions of higher atmospheric carbon dioxide due to increased carbon sequestration by vegetation (Terrer et al., 2021). Terrestrial carbon fluxes serve as a robust indicator of vegetation carbon sequestration. Figure S1 reveals that forthcoming enhancements in SOC predominantly stem from GPP and NEP. Notably, temperature emerges as the foremost adverse factor, exerting a significant negative influence by fostering the decomposition of soil carbon. The spatial analysis of GPP (Figure 4) reveals that areas with sign changes in its contribution to SOC are near the 400 mm precipitation line, indicates that precipitation patterns play a significant role in SOC sequestration. This finding has implications for land management practices, suggesting that regions with intermediate precipitation levels may be key targets for interventions aimed at increasing SOC stocks.

Furthermore, our study posits that GPP's contribution to SOC is negative at low values but becomes positive above a certain threshold. After reaching a peak, the positive contribution of GPP stabilizes and does not further increase. This finding implies that regions at the intersection of positive and negative contributions could enhance SOC accumulation through targeted interventions. Specifically, near the 400 mm precipitation line, vegetation restoration efforts could elevate GPP beyond the threshold, shifting its contribution from negative to positive. However, in areas already characterized by high GPP, additional planting may not yield further increases in SOC accumulation.

In our study, we opted to use carbon fluxes to represent the CO_2 fertilization effect rather than directly employing CO_2 concentration. Although we tested incorporating CO_2 concentration as a covariate in our machine learning model, it failed to accurately capture the fertilization effect. This discrepancy stems from the relatively minor spatial and seasonal changes in CO_2 concentration compared to the anticipated future increases, rendering the space-for-time substitution approach ineffective as it involves extrapolation. Consequently, our proposed method of utilizing carbon fluxes as covariates proves to be a valuable approach for addressing the CO_2 fertilization effect in machine learning models of SOC.

While our study provides valuable insights into the potential for SOC sequestration in China, it also highlights the inherent uncertainties in modeling the complex earth systems. The discrepancies between ESMs underscore the need for continued refinement of these tools and for the integration of diverse data sources to improve predictive accuracy. Additionally, the machine learning model trained on historical data face constraints in extrapolating future conditions due to potential alterations in the relationship between SOC and its influencing factors under climate change (Pickett, 1989). This constraint arises because changes in climate can introduce novel dynamics that may not be fully represented or captured by historical records, thereby impacting the predictive power of models for SOC behavior in a changing environment. Moreover, interpretive methods themselves may introduce additional uncertainty (Huang et al., 2023).

5 Conclusions

Our comprehensive study has provided valuable insights into the dynamics of SOC in China, projecting an overall increase in SOC stocks across various SSPs until the year 2100. This positive trend contrasts with some existing literature that anticipates declines in SOC under certain scenarios, highlighting the critical role of carbon flux, particularly GPP, in influencing SOC outcomes. Our findings underscore the significance of carbon flux as the most influential variable affecting future SOC changes, with its impact being more pronounced under higher carbon emission scenarios. This suggests that the CO2 fertilization effect may continue to play a vital role in enhancing SOC, even in high CO2 environments. The spatial analysis within our study has revealed that areas near the 400 mm precipitation line are critical zones for SOC sequestration, indicating that precipitation patterns are key determinants in the carbon cycle. Our analyses identify thresholds in the GPP-SOC relationship, with GPP's contribution to SOC transitioning from negative to positive beyond a certain level. However, this positive contribution does not increase indefinitely, indicating a plateau effect that has important implications for land management and carbon sequestration strategies.

Despite the promising projections, our study acknowledges the inherent uncertainties associated with ESMs and the interpretation of complex environmental data. The variability between ESMs highlights the need for ongoing research and model refinement to enhance the accuracy of SOC predictions.

As the global community continues to seek solutions for climate change mitigation, understanding the factors that influence SOC is crucial for developing effective carbon management strategies. Our research contributes to this understanding by providing a nuanced view of the interactions between carbon fluxes, climate variables, land use, and SOC. It is our hope that these insights will inform future land management practices and policies aimed at maximizing the potential of soils as carbon sinks, thereby supporting global efforts to combat climate change and promote sustainable development.

Conflict of Interest

The authors declare no conflicts of interest relevant to this study.

Data Availability Statement

Soil profile data from the Second National Soil Survey of China were derived from the China data set of soil properties for land surface modeling (Shangguan et al., 2013). China's terrestrial ecosystems data were derived from Xu et al. (2019). Relief data were calculated based on DEM derived from Multi-Error-Removed Improved Terrain DEM (Yamazaki et al., 2017). Landform data were from the European Soil Data Center (Iwahashi and Pike, 2007). Climate data were downloaded from the WorldClim version 2.1 (Fick & Hijmans, 2017). Landuse data were downloaded from the Land Use Harmonization project (Hurtt et al., 2020). Historical carbon flux data were derived from the Global Carbon Fluxes dataset (Shangguan et al., 2023), and future carbon flux data were derived from CMIP6 available in Earth System Grid Federation (ESGF, Lawrence Livermore National Laboratory, 2023). Soil maps were downloads from Shangguan et al. (2013) and the Harmonized World Soil Database (FAO/IIASA/ISRIC/ISS-CAS/JRC, 2012). The average soil and sedimentary deposit thickness map were download from (Pelletier, 2016).Rock type data was derived from USGS Geosciences and Environmental Change Science Center (GECSC) based on the Global Lithological Map database v1.1 (GLiM, Hartmann & Moosdorf, 2012). Bedrock depth data was downloaded from Yan et al. (2020).

Acknowledgments

The study was partially supported by the National Natural Science Foundation of China (42375144, 42088101 and U1811464), Guangdong Basic and Applied Basic Research Foundation (2021B0301030007) and the Innovation Group Project of Southern Marine Science and Engineering Guangdong Laboratory (Zhuhai) (311020008).

References

Baldocchi, D., Falge, E., Gu, L., Olson, R., Hollinger, D., Running, S., Anthoni, P., Bernhofer, C., Davis, K., Evans, R., Fuentes, J., Goldstein, A., Katul, G., Law, B., Lee, X., Malhi, Y., Meyers, T., Munger, W., Oechel, W., … Wofsy, S. (2001). FLUXNET: A New Tool to Study the Temporal and Spatial Variability of Ecosystem–Scale Carbon Dioxide, Water Vapor, and Energy Flux Densities. *Bulletin of the American Meteorological Society*, *82*(11), 2415–2434. https://doi.org/10.1175/1520-0477(2001)082<2415:FANTTS>2.3.CO;2

Batjes, N. H. (2014). Total carbon and nitrogen in the soils of the world. *European Journal of Soil Science*, *65*(1). https://doi.org/10.1111/ejss.12114_2

Blois, J. L., Williams, J. W., Fitzpatrick, M. C., Jackson, S. T., & Ferrier, S. (2013). Space can substitute for time in predicting climate-change effects on biodiversity. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(23). https://doi.org/10.1073/pnas.1220228110

Boucher, O., Servonnat, J., Albright, A. L., Aumont, O., Balkanski, Y., Bastrikov, V., Bekki, S., Bonnet, R., Bony, S., Bopp, L., Braconnot, P., Brockmann, P., Cadule, P., Caubel, A., Cheruy, F., Codron, F., Cozic, A., Cugnet, D., D'Andrea, F., … Vuichard, N. (2020). Presentation and Evaluation of the IPSL-CM6A-LR Climate Model. *Journal of Advances in Modeling Earth Systems*, *12*(7). https://doi.org/10.1029/2019MS002010

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Chen, S., Arrouays, D., Leatitia Mulder, V., Poggio, L., Minasny, B., Roudier, P., Libohova, Z., Lagacherie, P., Shi, Z., Hannam, J., Meersmans, J., Richer-de-Forges, A. C., & Walter, C. (2022). Digital mapping of GlobalSoilMap soil properties at a broad scale: A review. *Geoderma*, *409*, 115567. https://doi.org/10.1016/j.geoderma.2021.115567

Crow, S. E., Lajtha, K., Filley, T. R., Swanston, C. W., Bowden, R. D., & Caldwell, B. A. (2009). Sources of plant-derived carbon and stability of organic matter in soil: Implications for global change. *Global Change Biology*, *15*(8). https://doi.org/10.1111/j.1365-2486.2009.01850.x

Davidson, E. A., & Janssens, I. A. (2006). Temperature sensitivity of soil carbon decomposition and feedbacks to climate change. *Nature*, *440*(7081), 165–173. https://doi.org/10.1038/nature04514

Döscher, R., Acosta, M., Alessandri, A., Anthoni, P., Arsouze, T., Bergman, T., Bernardello, R., Boussetta, S., Caron, L. P., Carver, G., Castrillo, M., Catalano, F., Cvijanovic, I., Davini, P., Dekker, E., Doblas-Reyes, F. J., Docquier, D., Echevarria, P., Fladrich, U., … Zhang, Q. (2022). The EC-Earth3 Earth system model for the Coupled Model Intercomparison Project 6. *Geoscientific Model Development*, *15*(7). https://doi.org/10.5194/gmd-15-2973-2022

FAO/IIASA/ISRIC/ISS-CAS/JRC (2012), Harmonized World Soil Database (version1.2). FAO, Rome, Italy and IIASA, Laxenburg, Austria. http://dx.doi.org/10.3334/ORNLDAAC/1247

Fick, S. E., & Hijmans, R. J. (2017). WorldClim 2: new 1-km spatial resolution climate surfaces for global land areas. *International Journal of Climatology*, *37*(12). https://doi.org/10.1002/joc.5086

Field, C. B. (2001). Plant Physiology of the "Missing" Carbon Sink. *Plant Physiology*, *125*(1), 25–28. https://doi.org/10.1104/pp.125.1.25

Grunwald, S. (2010). Current State of Digital Soil Mapping and What Is Next. In *Digital Soil Mapping* (pp. 3–12). Springer Netherlands. https://doi.org/10.1007/978-90-481-8863-5_1

Hartmann, J., & Moosdorf, N. (2012). The new global lithological map database GLiM: A representation of rock properties at the Earth surface. *Geochemistry, Geophysics, Geosystems*, *13*(12). https://doi.org/10.1029/2012GC004370

Heuvelink, G. B. M., Angelini, M. E., Poggio, L., Bai, Z., Batjes, N. H., Bosch, R., Bossio, D., Estella, S., Lehmann, J., Olmedo, G. F., & Sanderman, J. (2021). Machine learning in space and time for modelling soil organic carbon change. *European Journal of Soil Science*, *72*(4), 1607–1623. https://doi.org/10.1111/ejss.12998

Huang, F., Shangguan, W., Li, Q., Li, L., & Zhang, Y. (2023). Beyond prediction: An integrated post-hoc approach to interpret complex model in hydrometeorology. *Environmental Modelling and Software*, *167*. https://doi.org/10.1016/j.envsoft.2023.105762

Hurtt, G. C., Chini, L., Sahajpal, R., Frolking, S., Bodirsky, B. L., Calvin, K., Doelman, J. C., Fisk, J., Fujimori, S., Goldewijk, K. K., Hasegawa, T., Havlik, P., Heinimann, A., Humpenöder, F., Jungclaus, J., Kaplan, J. O., Kennedy, J., Krisztin, T., Lawrence, D., … Zhang, X. (2020). Harmonization of global land use change and management for the period 850-2100 (LUH2) for CMIP6. *Geoscientific Model Development*, *13*(11). https://doi.org/10.5194/gmd-13-5425-2020

Intergovernmental Panel on Climate Change. (2023). Climate Change 2021 – The Physical Science Basis. In *Climate Change 2021 – The Physical Science Basis*. https://doi.org/10.1017/9781009157896

Ito, A., Hajima, T., Lawrence, D. M., Brovkin, V., Delire, C., Guenet, B., Jones, C. D., Malyshev, S., Materia, S., McDermid, S. P., Peano, D., Pongratz, J., Robertson, E., Shevliakova, E., Vuichard, N., Wårlind, D., Wiltshire, A., & Ziehn, T. (2020). Soil carbon sequestration simulated in CMIP6-LUMIP models: implications for climatic mitigation. *Environmental Research Letters*, *15*(12), 124061. https://doi.org/10.1088/1748-9326/abc912

Iwahashi, J., and R. J. Pike (2007). Automated classifications of topography from DEMs by an unsupervised nested-means algorithm and a three-part geometric signature, *Geomorphology*, 86(3), 409-440. https://doi.org/10.1016/j.geomorph.2006.09.012.

Karnosky, D. F. (2003). Impacts of elevated atmospheric CO2 on forest trees and forest ecosystems: Knowledge gaps. In *Environment International* (Vol. 29, Issues 2–3). https://doi.org/10.1016/S0160-4120(02)00159-9

Kuzyakov, Y. (2010). Priming effects: Interactions between living and dead organic matter. *Soil Biology and Biochemistry*, *42*(9). https://doi.org/10.1016/j.soilbio.2010.04.003

Lal, R. (2003). Soil erosion and the global carbon budget. In *Environment International* (Vol. 29, Issue 4). https://doi.org/10.1016/S0160-4120(02)00192-7

Lal, R., Negassa, W., & Lorenz, K. (2015). Carbon sequestration in soil. *Current Opinion in Environmental Sustainability*, *15*, 79–86. https://doi.org/10.1016/j.cosust.2015.09.002

Lamichhane, S., Kumar, L., & Wilson, B. (2019). Digital soil mapping algorithms and covariates for soil organic carbon mapping and their implications: A review. *Geoderma*, *352*, 395–413. https://doi.org/10.1016/j.geoderma.2019.05.031

Lawrence Livermore National Laboratory. (2023). Coupled Model Intercomparison Project 6 Dataset *[Dataset]*. Earth System Grid Federation (ESFG). Retrieved from https://aims2.llnl.gov/search

Le Quéré, C., Andres, R. J., Boden, T., Conway, T., Houghton, R. A., House, J. I., Marland, G., Peters, G. P., Van Der Werf, G. R., Ahlström, A., Andrew, R. M., Bopp, L., Canadell, J. G., Ciais, P., Doney, S. C., Enright, C., Friedlingstein, P., Huntingford, C., Jain, A. K., … Zeng, N. (2013). The global carbon budget 1959-2011. *Earth System Science Data*, *5*(1). https://doi.org/10.5194/essd-5-165-2013

Liang, J., Zhou, Z., Huo, C., Shi, Z., Cole, J. R., Huang, L., Konstantinidis, K. T., Li, X., Liu, B., Luo, Z., Penton, C. R., Schuur, E. A. G., Tiedje, J. M., Wang, Y. P., Wu, L., Xia, J., Zhou, J., & Luo, Y. (2018). More replenishment than priming loss of soil organic carbon with additional carbon input. *Nature Communications*, *9*(1). https://doi.org/10.1038/s41467-018-05667-7

Luo, Y., Huang, Y., Sierra, C. A., Xia, J., Ahlström, A., Chen, Y., et al. (2022). Matrix approach to land carbon cycle modeling. Journal of Advances in Modeling Earth Systems, 14, e2022MS003008. https://doi.org/10.1029/2022MS003008

Liang, Z., Chen, S., Yang, Y., Zhao, R., Shi, Z., & Viscarra Rossel, R. A. (2019). National digital soil map of organic matter in topsoil and its associated uncertainty in 1980's China. *Geoderma*, *335*, 47–56. https://doi.org/10.1016/J.GEODERMA.2018.08.011

Li, H., Wu, Y., Liu, S., Xiao, J., Zhao, W., Chen, J., Alexandrov, G., & Cao, Y. (2022). Decipher soil organic carbon dynamics and driving forces across China using machine learning. *Global Change Biology*, *28*(10). https://doi.org/10.1111/gcb.16154

Litton, C. M., & Giardina, C. P. (2008). Below-ground carbon flux and partitioning: global patterns and response to temperature. *Functional Ecology*, *22*(6), 941–954. https://doi.org/10.1111/j.1365-2435.2008.01479.x

Liu, F., Wu, H., Zhao, Y., Li, D., Yang, J. L., Song, X., Shi, Z., Zhu, A. X., & Zhang, G. L. (2022). Mapping high resolution National Soil Information Grids of China. *Science Bulletin*, *67*(3). https://doi.org/10.1016/j.scib.2021.10.013

Liu, M., Han, G., & Zhang, Q. (2020). Effects of agricultural abandonment on soil aggregation, soil organic carbon storage and stabilization: Results from observation in a small karst catchment, Southwest China. *Agriculture, Ecosystems and Environment*, *288*. https://doi.org/10.1016/j.agee.2019.106719

Lundberg, S., & Lee, S.-I. (2017). *A Unified Approach to Interpreting Model Predictions*.

Luo, Z., Wang, G., & Wang, E. (2019). Global subsoil organic carbon turnover times dominantly controlled by soil properties rather than climate. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-11597-9

Mauritsen, T., Bader, J., Becker, T., Behrens, J., Bittner, M., Brokopf, R., Brovkin, V., Claussen, M., Crueger, T., Esch, M., Fast, I., Fiedler, S., Fläschner, D., Gayler, V., Giorgetta, M., Goll, D. S., Haak, H., Hagemann, S., Hedemann, C., … Roeckner, E. (2019). Developments in the MPI-M Earth System Model version 1.2 (MPI-ESM1.2) and Its Response to Increasing CO2. *Journal of Advances in Modeling Earth Systems*, *11*(4). https://doi.org/10.1029/2018MS001400

McBratney, A. B., Mendonça Santos, M. L., & Minasny, B. (2003). On digital soil mapping. *Geoderma*, *117*(1–2), 3–52. https://doi.org/10.1016/S0016-7061(03)00223-4

McBratney, A., de Gruijter, J., & Bryce, A. (2019). Pedometrics timeline. *Geoderma*, *338*. https://doi.org/10.1016/j.geoderma.2018.11.048

Minasny, B., Malone, B. P., McBratney, A. B., Angers, D. A., Arrouays, D., Chambers, A., Chaplot, V., Chen, Z. S., Cheng, K., Das, B. S., Field, D. J., Gimona, A., Hedley, C. B., Hong, S. Y., Mandal, B., Marchant, B. P., Martin, M., McConkey, B. G., Mulder, V. L., … Winowiecki, L. (2017). Soil carbon 4 per mille. In *Geoderma* (Vol. 292). https://doi.org/10.1016/j.geoderma.2017.01.002

Nowak, R. S., Ellsworth, D. S., & Smith, S. D. (2004). Functional responses of plants to elevated atmospheric CO2 - Do photosynthetic and productivity data from FACE experiments support early predictions? In *New Phytologist* (Vol. 162, Issue 2). https://doi.org/10.1111/j.1469-8137.2004.01033.x

Odgers, N. P., Libohova, Z., & Thompson, J. A. (2012). Equal-area spline functions applied to a legacy soil database to create weighted-means maps of soil organic carbon at a continental scale. *Geoderma*, *189–190*, 153–163. https://doi.org/10.1016/j.geoderma.2012.05.026

Padarian, J., Minasny, B., & McBratney, A. B. (2020). Machine learning and soil sciences: A review aided by machine learning tools. In *SOIL* (Vol. 6, Issue 1). https://doi.org/10.5194/soil-6-35-2020

Patoine, G., Eisenhauer, N., Cesarz, S., Phillips, H. R. P., Xu, X., Zhang, L., & Guerra, C. A. (2022). Drivers and trends of global soil microbial carbon over two decades. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-31833-z

Pelletier, J.D., Broxton, P.D., Hazenberg, P., Zeng, X., Troch, P.A., Niu, G., Williams, Z.C., Brunke, M.A. & Gochis, D. 2016. Global 1-km Gridded Thickness of Soil, Regolith, and Sedimentary Deposit Layers. ORNL DAAC, Oak Ridge, Tennessee, USA. http://dx.doi.org/10.3334/ORNLDAAC/1304

Pickett, S. T. A. (1989). Space-for-Time Substitution as an Alternative to Long-Term Studies. In *Long-Term Studies in Ecology*. https://doi.org/10.1007/978-1-4615-7358-6_5

Sayer, E. J., Heard, M. S., Grant, H. K., Marthews, T. R., & Tanner, E. V. J. (2011). Soil carbon release enhanced by increased tropical forest litterfall. *Nature Climate Change*, *1*(6). https://doi.org/10.1038/nclimate1190

Shangguan, W., Dai, Y., Liu, B., Zhu, A., Duan, Q., Wu, L., Ji, D., Ye, A., Yuan, H., Zhang, Q., Chen, D., Chen, M., Chu, J., Dou, Y., Guo, J., Li, H., Li, J., Liang, L., Liang, X., … Zhang, Y. (2013). A China data set of soil properties for land surface modeling. *Journal of Advances in Modeling Earth Systems*, *5*(2), 212–224. https://doi.org/10.1002/jame.20026

Shangguan, W., Xiong, Z., Nourani, V., Li, Q., Lu, X., Li, L., Huang, F., Zhang, Y., Sun, W., & Dai, Y. (2023). A 1 km Global Carbon Flux Dataset Using In Situ Measurements and Deep Learning. *Forests*, *14*(5), 913. https://doi.org/10.3390/f14050913

Smith, P., Fang, C., Dawson, J. J. C., & Moncrieff, J. B. (2008). Impact of Global Warming on Soil Organic Carbon. In *Advances in Agronomy* (Vol. 97, pp. 1–43). https://doi.org/10.1016/S0065-2113(07)00001-6

Song, X. D., Wu, H. Y., Ju, B., Liu, F., Yang, F., Li, D. C., Zhao, Y. G., Yang, J. L., & Zhang, G. L. (2020). Pedoclimatic zone-based three-dimensional soil organic carbon mapping in China. *Geoderma*, *363*. https://doi.org/10.1016/j.geoderma.2019.114145

Terrer, C., Phillips, R. P., Hungate, B. A., Rosende, J., Pett-Ridge, J., Craig, M. E., van Groenigen, K. J., Keenan, T. F., Sulman, B. N., Stocker, B. D., Reich, P. B., Pellegrini, A. F. A., Pendall, E., Zhang, H., Evans, R. D., Carrillo, Y., Fisher, J. B., Van Sundert, K., Vicca, S., & Jackson, R. B. (2021). A trade-off between plant and soil carbon storage under elevated CO2. *Nature*, *591*(7851). https://doi.org/10.1038/s41586-021-03306-8

Trivedi, P., Singh, B. P., & Singh, B. K. (2018). Soil carbon: Introduction, importance, status, threat, and mitigation. In *Soil Carbon Storage: Modulators, Mechanisms and Modeling*. https://doi.org/10.1016/B978-0-12-812766-7.00001-9

Xu, L., He, N., & Yu, G. (2019). A dataset of carbon density in Chinese terrestrial ecosystems (2010s). *China Scientific Data*, *1*, 4. https://doi.org/10.11922/csdata.2018.0026.zh

Yamazaki, D., Ikeshima, D., Tawatari, R., Yamaguchi, T., O'Loughlin, F., Neal, J. C., Sampson, C. C., Kanae, S., & Bates, P. D. (2017). A high-accuracy map of global terrain elevations. *Geophysical Research Letters*, *44*(11), 5844–5853. https://doi.org/10.1002/2017GL072874

Yan, F., Shangguan, W., Zhang, J., & Hu, B. (2020). Depth-to-bedrock map of China at a spatial resolution of 100 meters. *Scientific Data*, *7*(1). https://doi.org/10.1038/s41597-019-0345-6

Yang, R. M., Huang, L. M., Zhang, X., Zhu, C. M., & Xu, L. (2023). Mapping the distribution, trends, and drivers of soil organic carbon in China from 1982 to 2019. *Geoderma*, *429*. https://doi.org/10.1016/j.geoderma.2022.116232

Yigini, Y., & Panagos, P. (2016). Assessment of soil organic carbon stocks under future climate and land cover changes in Europe. *Science of The Total Environment*, *557–558*, 838–850. https://doi.org/10.1016/j.scitotenv.2016.03.085

Zhang, Z., Ding, J., Zhu, C., Wang, J., Ge, X., Li, X., Han, L., Chen, X., & Wang, J. (2023). Historical and future variation of soil organic carbon in China. *Geoderma*, *436*, 116557. https://doi.org/10.1016/j.geoderma.2023.116557

Ziehn, T., Chamberlain, M. A., Law, R. M., Lenton, A., Bodman, R. W., Dix, M., Stevens, L., Wang, Y. P., & Srbinovsky, J. (2020). The Australian Earth System Model: ACCESS-ESM1.5. *Journal of Southern Hemisphere Earth Systems Science*, *70*(1). https://doi.org/10.1071/ES19035