European Journal of Soil Science

WILEY

European Journal of Soil Science

RESEARCH ARTICLE OPEN ACCESS

Improved Broad-Scale Modelling of Soil Organic Carbon Dynamics Following Land-Use Changes

Daria Seitz | Rene Dechow | David Emde | Florian Schneider | Axel Don

Thünen Institute of Climate-Smart Agriculture, Braunschweig, Germany

Correspondence: Daria Seitz (daria.seitz@thuenen.de)

Received: 30 April 2025 | **Revised:** 27 June 2025 | **Accepted:** 1 July 2025

Funding: This work was supported by the European Union's Horizon 2020 Research and Innovation Programme (862695) as part of the EJP Soil Programme.

Keywords: carbon sequestration | emissions reporting | machine learning | NIR | process-based modelling | RothC

ABSTRACT

Land-use changes (LUCs) strongly impact soil organic carbon (SOC) stocks over decades. However, there are too few long-term field experiments where these SOC dynamics have been observed long enough to validate process-based models for large-scale use. We have developed a new data-driven space-for-time approach for model validation using empirical data from over 3000 sites in the German Agricultural Soil Inventory, including 212 sites with LUC between cropland and grassland. Machine-learning models trained on sites under permanent land use were used to predict equilibrium SOC stocks for similar sites with changed land use. We used this derived data set to assess how well the process-based model RothC describes SOC dynamics following LUC. The default version of RothC struggled to capture the fast changes in SOC following LUC since it was mainly driven by differences in carbon input quantity and quality. Losses in SOC after converting grassland into cropland occurred faster than modelled, and SOC accrual after converting cropland to grassland was faster than simulated. This suggested an additional carbon stabilisation mechanism connected to grassland land use. We extended the RothC model with an additional carbon pool that builds up rapidly after grassland establishment, similar to aggregate-protected SOC. This improved the model efficiency from 0.49 to 0.80 for transitional croplands and from −3.39 to 0.90 after establishing grassland. This improved model version, RothC-LUC, is suitable for simulating SOC dynamics following LUC between cropland and grassland on a broad scale, such as in national inventory reports on greenhouse gas emissions.

1 | Introduction

Land-use changes (LUCs) have a strong impact on soil organic carbon (SOC) stocks and affect the climate (Guo and Gifford 2002; Li et al. 2024). Converting grasslands to cropland generally results in SOC losses and CO_2 emissions. Establishing grassland on former croplands generally increases SOC stocks and promotes carbon (C) sequestration in soils (Guo and Gifford 2002; Poeplau and Don 2013). The effect on SOC stocks lasts decades, with a new equilibrium reached after 80–140 years (Emde et al. 2024; Poeplau et al. 2011). This makes establishing grasslands on former croplands a long-term C sequestration measure (De et al. 2020) and converting grasslands to croplands a long-term source of CO_2 to the atmosphere (De Rosa et al. 2024; Guillaume et al. 2021; Sanderman et al. 2017).

The main drivers of C sequestration, stabilisation and mineralisation in soils strongly depend on land use (Dignac et al. 2017). In contrast to croplands, grasslands have a perennial vegetation cover and constantly provide C input via root turnover and rhizodeposition over the course of a year (Dignac et al. 2017; Poeplau et al. 2021). Root-derived C input has been shown to be more effective in forming and stabilising SOC than aboveground C input (Kätterer et al. 2011; Rasse et al. 2005; Sokol et al. 2019).

Summary

- We modelled soil organic carbon (SOC) dynamics following conversion between grassland and cropland.
- A novel data-driven space-for-time approach enabled model validation for broad-scale applications.
- Our newly developed process-based model, RothC-LUC, included SOC stabilisation in grasslands.
- In RothC-LUC, SOC stabilisation is a major reason for higher SOC stocks in grasslands.

Higher root-derived C inputs (Jacobs et al. 2020), increased aggregate stability in grasslands (Poeplau et al. 2024), reduced soil disturbance due to the lack of tillage (Haddaway et al. 2017), as well as the altered microbial community (Szoboszlay et al. 2017) and increased microbial carbon use efficiency (Liang et al. 2017; Schroeder et al. 2022) may contribute to stabilising SOC in grasslands.

Process-based SOC decomposition models, such as RothC (Coleman and Jenkinson 1996), are valuable for gaining a mechanistic understanding of SOC dynamics. They can be used to predict SOC changes and the corresponding climate effects of changes in arable management (Dechow et al. 2019; Launay et al. 2021; Riggers et al. 2019). RothC is a well-known SOC turnover model that can be used on a local to global scale (Peralta et al. 2022). Some of the land-use dependent drivers of SOC changes are implemented in RothC. This includes the varying decomposability of the C input depending on land use. Regular disturbance due to tillage is implicitly accounted for, as the model was calibrated on tilled croplands (Falloon 2001). By default, decomposition rates are modified based on a soil cover factor. Physical stabilisation is only implemented via the balance between stabilisation and mineralisation during decomposition depending on the clay content, representing adsorption. However, this does not change with LUC and the implementation of additional stabilisation mechanisms has been suggested (Dignac et al. 2017). There are contrasting findings as to whether the default version of RothC can sufficiently describe grassland SOC dynamics and LUC effects on SOC stocks (Heikkinen et al. 2014) or whether modifications are needed to account for land-use dependent SOC stabilisation (Gottschalk et al. 2010). Gottschalk et al. (2010) found that RothC reacted too slowly to accurately simulate SOC stock changes and added a pool to accelerate the SOC dynamics after deforestation to arable land. As SOC loss can also occur rapidly after LUC from grassland to cropland, it remains unclear whether the original version of RothC is suitable to accurately simulate the SOC effects of LUC between cropland and grassland.

The United Nations Framework Convention on Climate Change encourages using dynamic SOC models to report SOC changes in national greenhouse gas inventories, if the models are evaluated with observations (Paustian et al. 2006). For land-use categories without LUC (e.g., cropland remaining cropland), such process-based modelling approaches—also referred to as Tier-3—are already used to quantify the CO_2 emissions of several countries, for example, in Switzerland (Wüst-Galley et al. 2020), Finland (Statistics Finland 2024) and Sweden (Swedish Environmental Protection Agency 2024). However, LUC categories are generally only reported using simpler Tier-1 or Tier-2 approaches assuming a linear SOC change over, for example, 20 years as suggested in the IPCC Guidelines (Paustian et al. 2006). Using Tier-3 modelling approaches to quantify CO_2 emissions following LUC on a national scale would require model validation for a wide range of pedo-climatic conditions (Le Noë et al. 2023). But there are only a few long-term field experiments where SOC dynamics have been observed long enough to get close to equilibrium after LUC (Poeplau et al. 2011), with Rothamsted, UK being the only long-term field experiment for both LUC directions between cropland and grassland and with observations over 50 years (Johnston et al. 2009). This lack of data makes validating dynamic SOC models for large-scale use challenging.

Data-driven approaches offer novel opportunities to answer various research questions. They have also been discussed to complement process-based modelling approaches (Reichstein et al. 2019). The large amounts of data needed for data-driven approaches can be obtained from soil inventories. The German Agricultural Soil Inventory provides soil and management information (Jacobs et al. 2020; Poeplau et al. 2020) and information on the land-use history (Emde et al. 2024) for over 3000 sites distributed in an 8 × 8 km grid across Germany. This makes it a comprehensive data set to apply data-driven methods. It has successfully been used to train machine-learning models to predict SOC stocks (Sakhaee et al. 2022). Schneider et al. (2021) introduced data-driven reciprocal modelling whereby machine-learning models are trained on reference objects (e.g., grassland sites), then applied to similar objects (e.g., cropland sites) that have received treatment (e.g., arable use) and lie within the area of applicability, meaning they are similar enough to assess treatment effects. We hypothesised this approach could be used to generate an artificial data set enabling the model validation of process-based models for broad-scale use.

The overall goal of this study was to explore RothC's ability to simulate LUC effects on SOC stocks across various pedo-climatic conditions in the temperate region and to suggest model improvements if needed. In the first step, we used empirical data from the first German Agricultural Soil Inventory and data-driven reciprocal modelling to predict equilibrium SOC stocks on sites with historical LUC between cropland and grassland. In a second step, we used the derived data set to test the ability of three different versions of the RothC model to simulate SOC dynamics after LUC between cropland and grassland. We evaluated the accuracy of the models to fit observed SOC dynamics and the plausibility of modelled C input rates.

2 | Materials and Methods

2.1 | Data on Soils, Land-Use History and Climate

We used data from the first German Agricultural Soil Inventory containing over 3000 sites in an 8 × 8 km grid where soils were sampled to a depth of up to 100 cm (Poeplau et al. 2020). Carbon

inputs were estimated based on recent management data obtained via farmers' questionnaires (Jacobs et al. 2020). Land-use history dating back at least 62 years and up to 250 years was assessed by combining farmer questionnaires, historical maps and orthophotos, and soil surveyor observations at the time of sampling (Emde et al. 2024).

In this study, we only considered sites on mineral soils (< 8% SOC content in the topsoil, 0–30 cm) under the land uses of cropland ($n = 2093$) and grassland ($n = 678$), as these are the most common agricultural land uses in Germany. For the purpose of this study, we classified a site as grassland as soon as grasses or leys had been grown for 1 year. This was done in order to also cover the initial dynamics after LUC. Croplands with ley rotations, that is, with grasses or leys in the rotation several times, were not considered in this study and were removed from the data set. The total data set was reduced based on site properties due to the known limitations of the RothC model (cf. Supporting Information S1). Hydromorphic soils (soil types according to the German soil classification system KA-5 (Ad-Hoc AG Boden 2005): 'Marschen', 'Gleye', 'Erd- und Mulmmoore' and 'Naturnahe Moore') and soils with a groundwater level of < 80 cm below the surface were excluded because RothC is only parameterised for non-waterlogged soils.

For the purpose of this study, we assumed for *permanent land use* SOC stocks close to equilibrium at sites where there was no documented LUC. Following this assumption, we split the data set into *permanent* and *transitional* sites, where LUC has happened once within 223 years and 90 years for croplands and grasslands, respectively. The thresholds were based on the upper 95% confidence intervals (CIs) from Emde et al. (2024), who found German croplands reaching equilibrium SOC stocks after 180 years (95% CI: 151–223 years) and grasslands reaching equilibrium SOC stocks after 83 years (95% CI: 79–90 years). We excluded sites with any historic land use other than grassland or cropland and sites with more than one LUC. For the process-based modelling approach, we assumed a constant C input over time. Therefore, we excluded croplands older than 100 years from the transitional data set, since changes in their management due to plant breeding and fertilisation may have changed C inputs. Further sites were excluded during the modelling process based on the area of applicability assessment and plausibility checks (Section 2.2). The final data set used in this study included 49 transitional croplands, 160 transitional grasslands, 1373 permanent croplands and 137 permanent grasslands. The locations and pedo-climatic properties of these sites are shown in Figure 1 and Supporting Information S2.

We sampled gridded meteorological maps providing the mean temperature (DWD Climate Data Center 2023b), total precipitation (DWD Climate Data Center 2023c) and radiation (DWD Climate Data Center 2023a) used to calculate the evapotranspiration after Turc (1961) for each site in monthly intervals. The time series of these meteorological data covered the years 1951–2022. We filled missing values with the mean from the same months of the other years (2 months at two sites in total). To account for recent increases in mean annual temperature due to climate change (cf. Supporting Information S3), we split the data set at the year of the LUC into one set 'prior to LUC' and one set 'after LUC'.

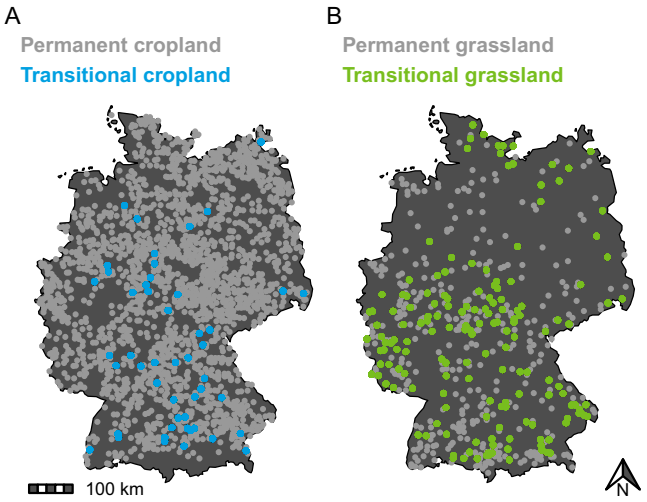

FIGURE 1 | Location of selected cropland (A) and grassland (B) sites. Grey dots represent permanent land use and coloured dots represent transitional sites (blue: Croplands, green: Grasslands).

2.2 | Data-Driven Reciprocal Modelling

We predicted SOC stocks in the topsoil (0–30 cm) at cropland and grassland equilibrium with regard to LUC (Figure 2) using a stacked ensemble model based on the R package *mlr3* (Lang et al. 2019) with random forest (ranger; Wright and Ziegler 2017), boosted regression trees (*gbm*; Ridgeway and GBM Developers 2024) and support vector regression (*svm* function of the *e1071* package; D. Meyer et al. 2023) base learners. The ensemble model is outlined in full in Sakhaee et al. (2024). In brief, a nested 5-fold spatial cross-validation approach was carried out 15 times drawing random samples from 50 strata across Germany (100 × 100 km^2 INSPIRE grid). Model hyperparameters were tuned across all base learners simultaneously using a random search algorithm. All 30 cross-validated SOC stock estimations (15 for cropland equilibrium, 15 for grassland equilibrium) were used as independent site estimates in the RothC modelling and are referred to as *predicted SOC stocks* throughout this manuscript.

We excluded transitional sites outside the area of applicability (*aoa* function from *CAST* package; H. Meyer et al. 2024; H. Meyer and Pebesma 2021) with regard to the following soil properties: clay content, sand content, organic carbon content or organic carbon stock, inorganic carbon, C:N ratio, WRB soil group, German soil classification, elevation, slope, mean annual temperature and mean annual precipitation. If the predicted cropland SOC stock was higher than the predicted grassland SOC stock, they were removed from the data set (two croplands, three grasslands).

We aggregated the data set into five duration intervals for each LUC direction. The five intervals were defined using the Fisher-Jenks algorithm (Fisher 1958; Jenks 1977). This clusters one-dimensional data into a given number of groups by minimising the within-group distance and maximising the between-group distance.

In order to compare the observed SOC stocks (SOC_{obs}) with the site-specific predictions of the cropland equilibrium SOC

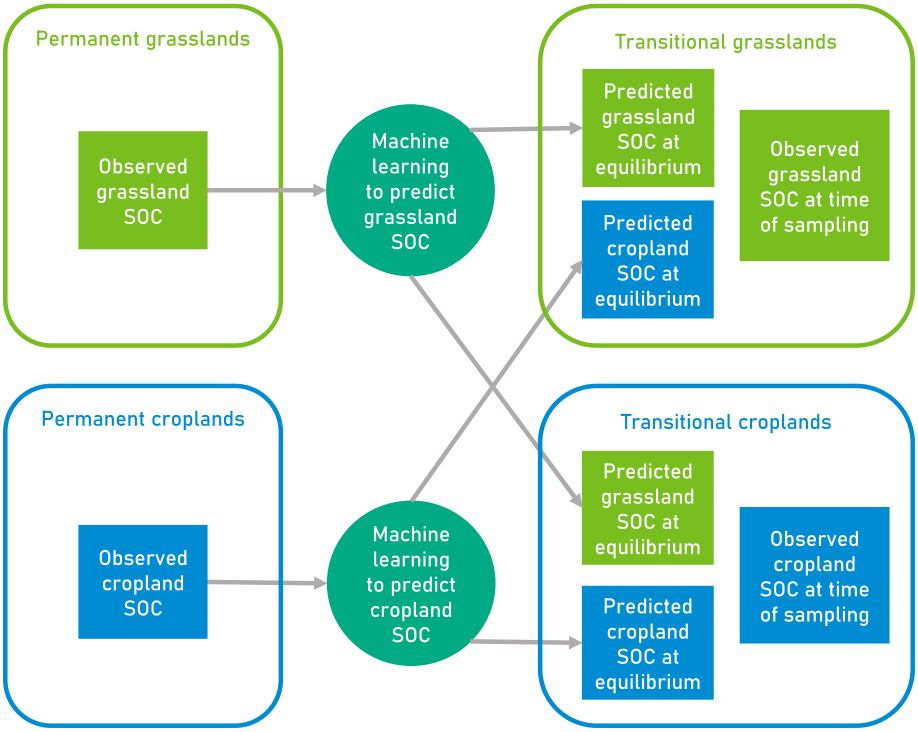

FIGURE 2 | Estimating land-use specific equilibrium SOC stocks of transitional sites by training machine-learning models to predict SOC stocks of sites with permanent land use.

stock ($SOC_{crop-pred}$) and grassland equilibrium SOC stock ($SOC_{grass-pred}$), we normalised the observations linearly using the following equation:

$$SOC_{norm} = \frac{SOC_{obs} - SOC_{crop-pred}}{SOC_{grass-pred} - SOC_{crop-pred}}$$

2.3 | Testing the RothC Model

2.3.1 | Model Description

We tested three different versions of RothC: *RothC-default* (Coleman and Jenkinson 1996), *RothC-silt* (Gottschalk et al. 2010) and *RothC-LUC* (new development in this study).

RothC (Coleman and Jenkinson 1996) is a dynamic, process-based SOC decomposition model. It has five pools: DPM (decomposable plant material), RPM (resistant plant material), HUM (humified organic matter), BIO (microbial biomass) and IOM (inert organic matter), each with varying decomposition rates (cf. Table 1). Carbon input from plants or organic fertiliser enters the system via the DPM and RPM pools with a specific DPM/RPM ratio. Carbon in the DPM or RPM pools is then decomposed and either (a) transferred to the HUM (54%) and BIO (46%) pools or (b) mineralised, leaving the system as CO_2 (Figure 3A). The ratio between the pool transfer and mineralisation is determined by the clay content of the soil, with less mineralisation in clayey soils. The decomposition rates are modified by the monthly temperature, precipitation and evapotranspiration, as well as by a soil cover factor that slows decomposition if growing plants are present. In the default model, grassland use is only accounted for by changing the amount and quality (DPM/RPM ratio) of the C input. We refer to this original version of RothC as *RothC-default*.

A modified version of RothC was introduced by Gottschalk et al. (2010) to improve modelled SOC dynamics after deforestation, here referred to as *RothC-silt*. RothC-silt has one additional pool, HUM-SILT, that gets filled with a fixed share of 0.39 of the total pool transfer from the other pools and whose decomposition rate changes according to land use, with an increased decomposition rate under cropland (Table 1, Figure 3B). The model was calibrated for LUC between forest, a carbon-rich system, and cropland, a carbon-poor system, and it is the only RothC version attempting to model SOC dynamics after LUC. In our tests focusing on agricultural LUCs between grassland and cropland, we applied the lower decomposition rate of forest in the case of grassland land use.

In this study, we propose a new version of RothC dedicated to modelling the SOC effects of LUC, RothC-LUC, in which we add a pool, AGG. This new pool represents a land-use dependent stabilisation process in the form of aggregates built up under grassland land use. AGG is a pool that slowly decomposes at a fixed rate (0.05 a^{-1}) and functions similarly to the HUM pool, but only receives input under grassland, not under cropland (Table 1, Figure 3C). If the site is used as grassland, a share of 0.30 moves from other pools to the AGG pool and is stabilised there. If it is used as cropland, a share of 0 moves to AGG. Thus, it performs exactly like *RothC-default* under long-term cropland, an advantage compared to *RothC-silt*. During the transfer to AGG under grassland, which represents the

TABLE 1 | Pools and their decomposition rates of RothC-default, RothC-silt and a new proposed version RothC-LUC.

Pool	Decomposition rate (a^{-1})	RothC-default	RothC-silt	RothC-LUC	Comment
DPM	10	X	X	X	
RPM	0.3	X	X	X	
BIO	0.66	X	X	X	Filled with 46% share from total pool transfer in RothC-default, less in other versions
HUM	0.02	X	X	X	Filled with 54% share from total pool transfer in RothC-default, less in other versions
IOM	0.0	X	X	X	
HUM-Silt	0.3 (cropland); 0.02 (grassland)		X		Filled with 39% share from total pool transfer; decomposition rate changes with land use
AGG	0.05			X	Only filled under grassland with 30% share from total pool transfer; has a maximum capacity

formation of aggregates, we assume that no SOC is lost as CO_2. The remaining 70% of decomposition products are transferred to HUM and BIO or are mineralised to CO_2 as in the default version. The AGG pool is defined to represent the main difference between the equilibrium SOC stocks of grassland and cropland. Its maximum capacity AGG_{cap} is quantified via the equation:

$$AGG_{cap} = \left(SOC_{grass-pred} - SOC_{crop-pred}\right) \times 0.8$$

with $SOC_{grass-pred}$ and $SOC_{crop-pred}$ representing the SOC stock at equilibrium under grassland and cropland use, respectively. If the maximum capacity of the AGG pool is reached, the C that is not needed to maintain the size of the AGG pool is (a) transferred to HUM and BIO or (b) mineralised to CO_2 as in the default version.

Calibrating the scaling factor 0.8, pool-transfer share (0.30) and decomposition rate (0.05 a^{-1}) was carried out using data from the Rothamsted long-term field experiments Highfield and Fosters (Johnston et al. 2009) and equations describing the SOC dynamics following LUC over time derived from the first German Agricultural Soil Inventory (Emde et al. 2024) (cf. Supporting Information S4).

2.3.2 | The Workflow

The general workflow for transitional croplands is illustrated in Figure 4. For transitional grasslands, land uses are reversed. Validation data for each transitional site consisted of 31 SOC stock values for the topsoil (0–30 cm): one observed SOC stock value with known years since LUC at the time of sampling, 15 estimates of SOC stock at equilibrium under grassland land use, and 15 estimates of SOC stock at equilibrium under cropland land use.

In the first step (Section 2.3.3), we ran RothC in inverse mode to equilibrium to calibrate the averaged C input needed to reach the predicted equilibrium SOC stocks for both cropland and grassland. This step also gave the pool distributions in equilibrium. In the next step (Section 2.3.4), we used (1) the predicted SOC stock before LUC, (2) the pool distribution before LUC and (3) the C input after LUC. We ran RothC in forward mode for the known number of years since LUC. To evaluate the model (Section 2.4), we compared the modelled SOC stocks with the observed SOC stocks at the time of sampling.

2.3.3 | C Input Estimation

For each transitional site, we calibrated the averaged C inputs needed to reach and maintain the predicted SOC stocks in equilibrium (400 years) for both cropland and grassland use. For RothC default, we used an analytical solution (Dechow et al. 2019). For RothC-silt (Gottschalk et al. 2010) and RothC-LUC, we used the Brent optimisation algorithm (with a lower limit of 0 and an upper limit of 30 t C ha^{-1} a^{-1}). We calibrated the C input for both the 15 predicted grassland SOC stocks and the 15 predicted cropland SOC stocks using the meteorological data prior to LUC (see Section 2.1) for the land use prior to LUC and vice versa. For the cropland runs, we assumed bare soil in 3 out of 12 months (January, August and September) based on management data from the German Agricultural Soil Inventory (Jacobs et al. 2020). For grassland runs, we assumed soil cover throughout the year. The assumed C input quality (DPM/RPM ratio) was derived by weighting C input components reported

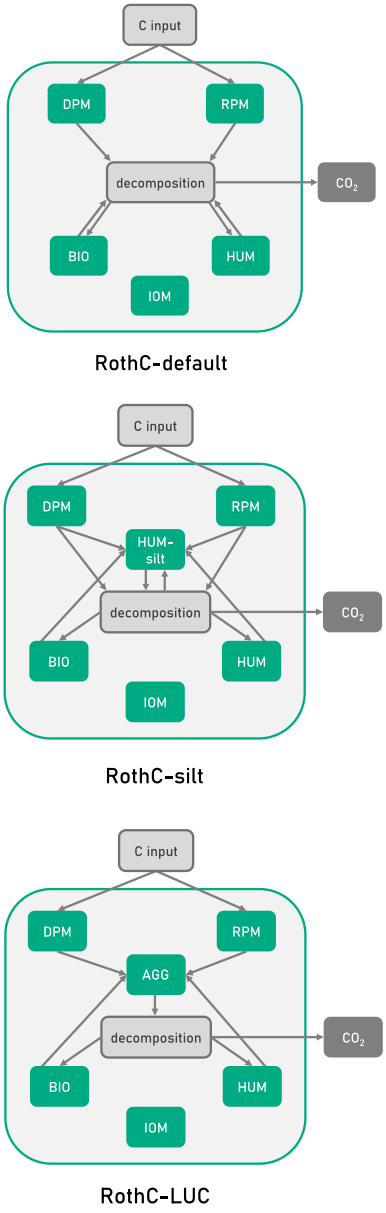

FIGURE 3 | Schematic overview of default RothC (A), RothC-silt as proposed by Gottschalk et al. (2010) (B) and our proposed model RothC-LUC (C). The carbon (C) input enters the systems via the DPM (decomposable plant material) and RPM (resistant plant material) and gets translocated to the BIO (microbial biomass) and HUM (humified organic matter) pools or mineralised to CO_2 during decomposition. The IOM (inert organic matter) pool is neither decomposed nor filled. The proposed AGG pool (C) representing aggregates is filled only under grassland with a share of 0.30 from other pools and slowly decomposes at a rate similar to HUM. The decomposition rate of the HUM-silt pool in RothC-silt changes depending on land use and is higher under cropland.

for the German Agricultural Soil Inventory (Jacobs et al. 2020; Riggers et al. 2019). Table 2 shows the partitioning of the input in the DPM, RPM and HUM pools based on parameterisations from Dechow et al. (2019). The size of the IOM pool was estimated based on Falloon et al. (1998).

As a result, we obtained the averaged C input rates for both land uses, that is, a grassland C input for each grassland prediction and a cropland C input for each cropland prediction. Additionally, we obtained the pool distribution among the SOC pools for each predicted SOC stock.

2.3.4 | Forward Runs

To simulate the SOC dynamics following LUC, each version of RothC was run in forward mode starting from the predicted SOC stock and its pool distribution under the land use *before* LUC (e.g., grassland in transitional croplands). The forward runs were performed with the C input that had been calibrated for the corresponding predicted SOC stock of the land use *after* LUC (e.g., cropland in transitional croplands). A total of 225 repetitions were simulated for each site (15 cropland predictions × 15 grassland predictions).

We used the same model parameters for the forward runs as for the equilibrium runs: bare soil assumptions according to land use, IOM estimation according to Falloon et al. (1998), averaged DPM/RPM based on reported management and meteorological data for the period after LUC. We ran the model for 400 years, where we assumed that equilibrium was reached. Then, we sampled the modelled SOC stock at the known time after LUC to compare it with the observed SOC stock.

2.4 | Model Evaluation and Statistics

We compared our RothC-modelled SOC stocks after time t_i with observed SOC stocks after time t_i, where t_i is the time since LUC in years. We used the mean bias error (MBE), mean absolute percentage error (MAPE), root mean squared error (RMSE), Lin's concordance correlation coefficient (CCC) (Liao and Lewis 2000; Lin 1989) and the Nash–Sutcliffe model efficiency (NSE) (Nash and Sutcliffe 1970) to evaluate the model performance. Uncertainties are given as standard deviation if not stated otherwise.

2.5 | Software

We ran all simulations, analyses and visualisations in R version 4.2.2 (R Core Team 2022), where we used the following packages:

- SoilR (Sierra et al. 2012) to run RothC and RothC-silt, and rcpp (Eddelbuettel et al. 2025) to run RothC-LUC, which was implemented in C++.
- mlr3 (Lang et al. 2019) ensemble modelling framework to predict equilibrium SOC stocks with ranger (Wright and Ziegler 2017), gbm (Ridgeway and GBM Developers 2024) and svm (D. Meyer et al. 2023) base learners
- plyr, dplyr, tidyr, forcats and purrr from the tidyverse (Wickham et al. 2019) for data processing
- classInt (Bivand 2023) to create the intervals
- ggplot2 (Wickham et al. 2019), giscoR (Hernangómez 2025), patchwork (Pedersen 2024) and sf (Pebesma 2018) for visualisation and
- ie2misc (Embry et al. 2023), DescTools (Signorell 2024) and Hmisc (Harrell Jr. 2023) for model evaluation.

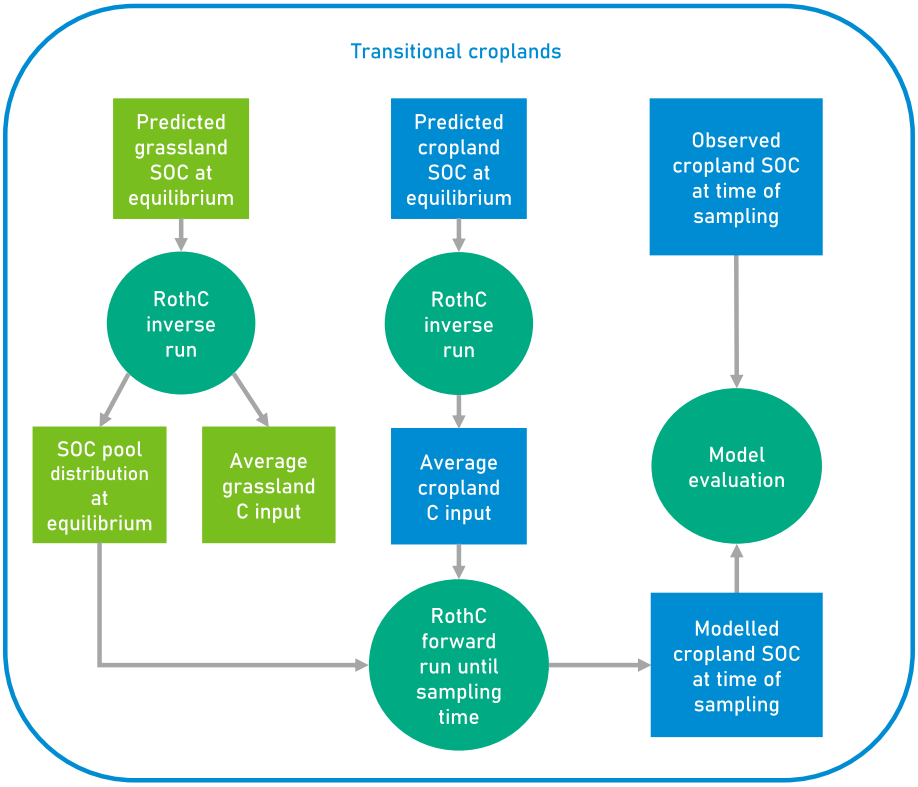

FIGURE 4 | Schematic overview of the RothC-modelling workflow for transitional croplands.

TABLE 2 | Land-use specific average DPM/RPM ratio of C input used in the C input calibration.

Land use	Pool	Partitioning fraction
Cropland	DPM	0.76
	RPM	0.24
	HUM	0.00
Grassland	DPM	0.685
	RPM	0.315
	HUM	0.00

3 | Results

3.1 | SOC Stocks at Equilibrium

The predicted cropland SOC stock assumed at equilibrium was 58.2 ± 9.2 Mg C ha^{-1} (on transitional grasslands) and 58.6 ± 9.0 Mg C ha^{-1} (on transitional croplands), on average (Figure 5). The predicted grassland SOC stock assumed at equilibrium was higher at 78.9 ± 13.3 Mg C ha^{-1} on transitional grasslands and 87.2 ± 12.8 Mg C ha^{-1} on transitional croplands. These values seemed reasonable compared to the measured SOC stocks of permanent grasslands (84.6 ± 23.9 Mg C ha^{-1}) and permanent croplands (56.2 ± 18.6 Mg C ha^{-1}).

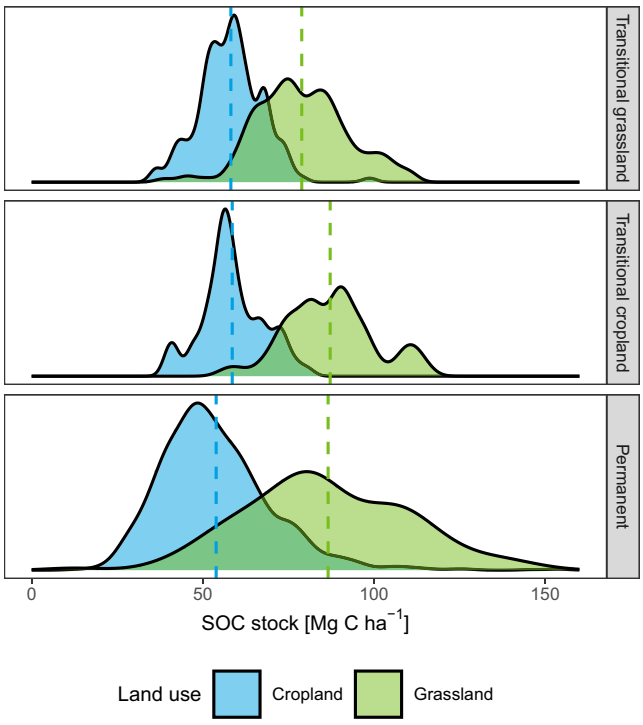

FIGURE 5 | Predicted SOC stocks for assumed-permanent land use on transitional grasslands (n = 2400) and transitional croplands (n = 870) in upper facets and observed SOC stocks on permanent croplands (blue, n = 1373) and grasslands (green, n = 137) in the lower facet. Dashed coloured lines represent the mean.

Since, for some transitional sites, the observed SOC stock was lower than the cropland prediction or higher than the grassland prediction (16 out of 49 croplands, 83 out of 160 grasslands, shown in Supporting Information S5), we aggregated the data set in five duration intervals for each LUC direction. Figure 6 shows how the aggregated, measured SOC stocks approach the new equilibrium SOC stocks over time. In transitional croplands, the measured SOC stocks decrease over time, while they increase in transitional grasslands.

3.2 | RothC Model Comparison

3.2.1 | Calibrated Carbon Inputs

The C input calibrated with RothC default was on average 3.22 ± 0.64 Mg C ha^{-1} a^{-1} for croplands and 3.76 ± 0.64 Mg C ha^{-1} a^{-1} for grasslands (Table 3). Using RothC-silt doubled the calibrated C input for croplands to 7.88 ± 1.60 Mg C ha^{-1} a^{-1} due to the increased decomposition rate under cropland, while grassland C inputs remained in a similar range as RothC default.

With the proposed model version RothC-LUC, the cropland C inputs remained the same as the default version at 3.22 ± 0.64 Mg C ha^{-1} a^{-1}. In contrast, the calibrated C input under grassland decreased to 2.81 ± 0.52 Mg C ha^{-1} a^{-1} due to the new stabilising AGG pool.

3.2.2 | Model Performance

RothC default systematically underestimated SOC stocks on transitional grasslands and overestimated them on transitional croplands (MBE of −8.24 and +3.36 Mg C ha^{-1}, respectively, Table 4). The modelled dynamics were too slow compared to the measurements (Figure 7). The new model version RothC-LUC accelerated the dynamics of SOC stock changes for both directions of LUC. This decreased the MBE to −0.24 Mg C ha^{-1}

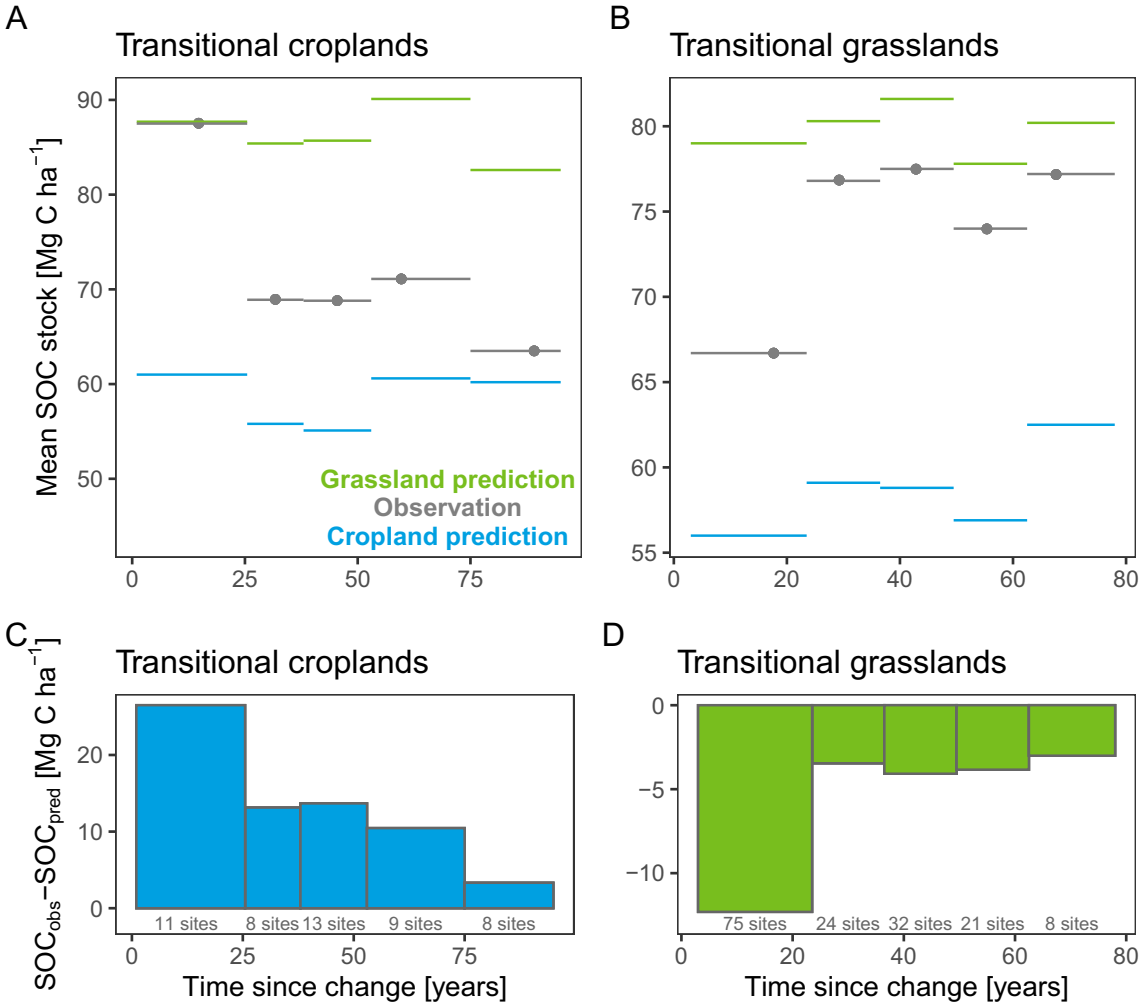

FIGURE 6 | Mean observed SOC stocks of transitional croplands (A) and transitional grasslands (B) over time (grey lines and dots), aggregated for five duration intervals, where dots represent the mean duration and grey lines the time span of the interval. Coloured lines show the mean predicted SOC stocks (aggregated in five duration intervals) representing the equilibrium conditions of each land use (green: Grassland and blue: Cropland). Bar plots show the difference between observed SOC stock (SOC_{obs}) and predicted SOC stock of the same land use (SOC_{pred}) (C: Transitional croplands, D: Transitional grasslands).

TABLE 3 | Mean carbon inputs for each land use estimated using different versions of RothC: RothC-default, RothC-silt and RothC-LUC. Uncertainty is given as the standard deviation representing site variability.

Assumed land use	Model	Mean C input (Mg C ha^{-1} a^{-1})	Standard deviation (Mg C ha^{-1} a^{-1})
Cropland	RothC default	3.22	0.64
Cropland	RothC-silt	7.88	1.60
Cropland	RothC-LUC	3.22	0.64
Grassland	RothC default	3.76	0.64
Grassland	RothC-silt	3.71	0.63
Grassland	RothC-LUC	2.81	0.52

TABLE 4 | Model performance indicators for intervals, comparing RothC default, RothC-silt and RothC-LUC and both directions of land-use change: The mean bias error (MBE), mean absolute percentage error (MAPE), root mean squared error (RMSE), concordance correlation coefficient (CCC) and Nash–Sutcliffe model efficiency (NSE).

Model	Direction of land-use change	MBE (Mg C ha^{-1})	MAPE (%)	RMSE (Mg C ha^{-1})	CCC (—)	NSE (—)
RothC default	Transitional cropland	3.36	8.2	5.8	0.63	0.49
RothC-silt	Transitional cropland	−16.6	22.4	18.0	0.08	−3.84
RothC-LUC	Transitional cropland	−0.39	3.4	3.7	0.85	0.80
RothC default	Transitional grassland	−8.24	11.0	8.5	0.23	−3.39
RothC-silt	Transitional grassland	−14.0	18.7	14.2	0.10	−11.2
RothC-LUC	Transitional grassland	−0.24	1.37	1.3	0.94	0.90

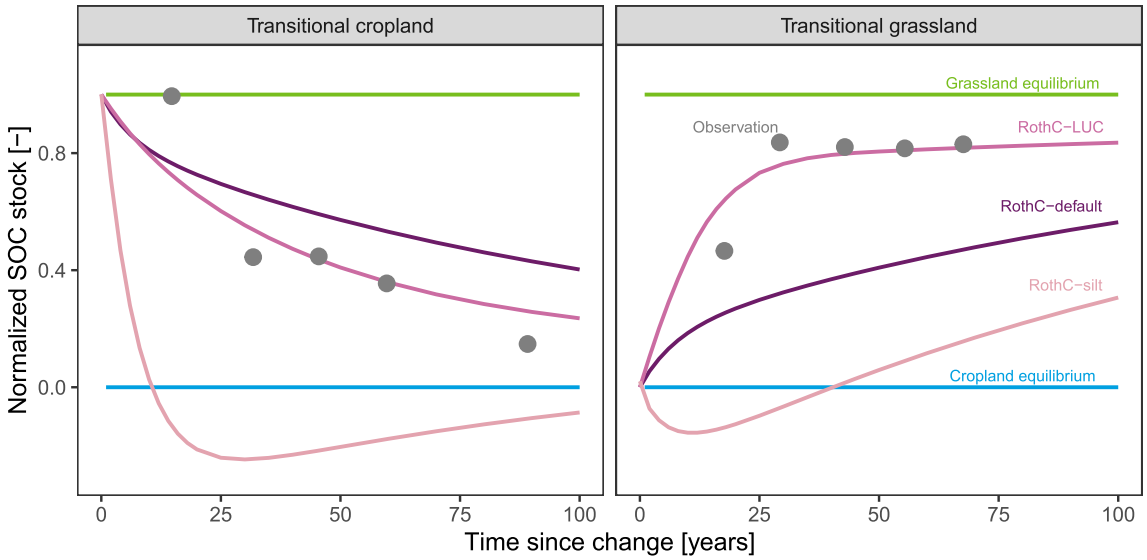

FIGURE 7 | RothC-LUC (pink curve) accelerated the modelled soil organic carbon stock dynamics after LUC compared with RothC-default (violet curve). Model accuracy thus improved compared to the observations (grey dots). RothC-silt (light pink curve) did not improve the model accuracy compared to the other versions and observations. Modelled and measured SOC stocks were normalised between the predicted cropland SOC stock (blue line) and the predicted grassland SOC stock (green line).

(transitional grasslands) and −0.39 Mg C ha^{-1} (transitional croplands). In contrast, RothC-silt overestimated SOC losses for transitional croplands (Figure 7) and did not improve model accuracy (Table 4), which could be attributed to its calibration for a different case of LUC (forest-cropland).

The proposed model version RothC-LUC improved the modelling accuracy compared to RothC-default (Figure 8, cf. Supporting Information S6). For transitional grasslands, the NSE (ranging between $-\infty$ and 1, with 1 representing a perfect fit) improved from −3.39 to 0.90 (Table 4, cf. Supporting

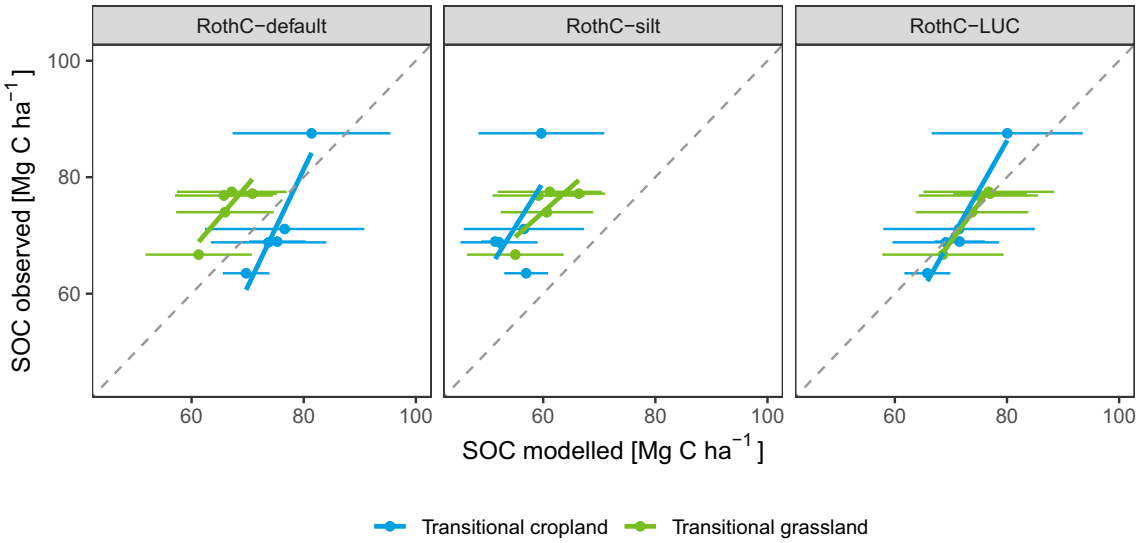

FIGURE 8 | The new proposed model version RothC-LUC (on the right) improved the model fit compared to RothC-default (on the left) and RothC-silt (in the centre) for both transitional grasslands (green) and transitional croplands (blue). Points show the mean of the duration intervals, error bars represent the intervals' standard deviation, and coloured lines show linear regressions of the intervals' mean. The dashed grey line marks the 1:1 line.

Information S7). For transitional croplands, the NSE improved from 0.49 to 0.80 (Table 4).

4 | Discussion

4.1 | Predicted SOC Stocks at Equilibrium

The data-driven reciprocal modelling approach produced plausible results on average: Measured SOC stocks showed increasing SOC stocks in transitional grasslands and decreasing SOC stocks in transitional croplands that approached predicted equilibrium SOC stocks over time. This general pattern and temporal timescale are in line with the expected SOC dynamics and findings in other studies (Emde et al. 2024; Johnston et al. 2009; Poeplau et al. 2011). This allowed the use of this approach to validate models on a broad range of pedo-climatic conditions, confirming their applicability on a national scale, an advantage over using individual long-term field experiments for model validation.

Extrema with unexpected high or low SOC stocks were not accurately predicted due to the central tendency bias of machine-learning models. The use of ensemble modelling (Sakhaee et al. 2024) aimed to reduce this bias. To avoid the remaining bias from single extrema in the model evaluation, we aggregated the data into five duration intervals. Consequently, our approach cannot validate models for individual sites, but only for multiple sites on a larger scale, where over- and underestimations compensate each other. At the same time, this approach reduces the risk of overfitting to individual sites and is at the preferred scale of model validation for large-scale model applications (Le Noë et al. 2023).

For any future application of this approach to creating artificial data sets for model validation, the data on which the models are trained must be comparable to the data to which the models are applied (Schneider et al. 2021). In this study, we ensured this by (1) applying all site-selection criteria to the total data set, including those criteria only necessary for applying RothC (e.g., no water-logged soils), and (2) removing all transitional sites outside of the area of applicability (H. Meyer and Pebesma 2021) which were too different from the training data set with respect to the most important predictor variables.

4.2 | SOC Dynamics Modelled Using RothC-Default Were Too Slow

The default version of the RothC model was not suitable to accurately simulate SOC dynamics following LUC between cropland and grassland (Table 4, Figure 7). It reacted too slowly and systematically underestimated transitional grassland SOC stocks (i.e., modelled SOC stocks did not increase fast enough) and overestimated transitional cropland SOC stocks (i.e., modelled SOC stocks decreased too slowly).

This comparatively slow dynamics of RothC was also addressed by Gottschalk et al. (2010), who found systematic biases after LUC between forest and cropland. To accelerate the modelled SOC dynamics, they increased the decomposition rate under cropland use. We tested this version (RothC-silt) with our data set, but it did not improve the model accuracy compared to RothC-default (Table 4). In order to maintain the same SOC stocks despite more decomposition, higher C input rates were needed. Consequently, cropland C inputs calibrated using RothC-silt were almost twice as high (7.9 ± 1.6 Mg C ha^{-1} a^{-1}) as other RothC versions and data from the literature, which seemed implausible. For example, Jacobs et al. (2020) found a mean C input of 3.7 Mg C ha^{-1} a^{-1} for German cropland sites based on allometric functions and management data. Schulze

et al. (2009) calculated a mean C input of 3.2 Mg C ha^{-1} a^{-1} on European croplands. With a mean C input of 3.2 ± 0.6 Mg C ha^{-1} a^{-1} (Table 3), RothC-default and RothC-LUC both gave plausible C input estimates for croplands compared to the literature. Our results suggest that RothC-silt is poorly suited to simulating SOC change in response to LUC between cropland and grassland.

In contrast, Heikkinen et al. (2014) found an acceptable match of the RothC modelled SOC increase after LUC from cropland to grassland. However, they used the actual evapotranspiration instead of the potential evaporation suggested in the RothC manual (Coleman and Jenkinson 1996) in order to decrease initially simulated SOC stocks, which they considered unrealistically high (Heikkinen et al. 2014). The site was located in the Moscow region of Russia, where the climate was continental, the cropland had been degraded, and the C inputs were estimated to be low. The typical management and pedo-climatic conditions were not easily comparable to those of this study. This may have resulted in different SOC dynamics.

4.3 | RothC-LUC Improved Modelled SOC Stock Dynamics After LUC

The newly proposed RothC-LUC considerably improved the fit of the SOC dynamics following LUC (Figure 8). The introduction of a stabilising AGG pool, constantly decomposing at a low decomposition rate but only filled under grassland and not under cropland, led to accelerated SOC dynamics overall after LUC. This improved all model evaluation indicators (Table 4). Our results suggest that the default version of RothC was indeed lacking a land-use dependent stabilisation mechanism, as also suggested by others (Dignac et al. 2017; Gottschalk et al. 2010). This would explain the temporal dynamics of SOC changes after LUC.

This improvement leads to faster increases of modelled SOC stocks when croplands are changed to grasslands. After 20 years, 70% of equilibrium SOC stocks are reached using RothC-LUC compared to the default RothC of only 25%, an underestimation of roughly 10 Mg C ha^{-1}. The Tier-1 approach, currently mostly used in greenhouse gas emissions reporting, significantly underestimates the time needed to reach the new equilibrium in a temperate climate. After 20 years, only 70% of the equilibrium SOC stocks are reached using RothC-LUC compared to 100% assumed in the Tier-1 approach. For the LUC from grassland to cropland, this corresponds to an overestimation of the SOC effect of around 20 Mg C ha^{-1} within 20 years and, therefore, an overestimation of greenhouse gas emissions if the Tier-1 approach is used.

We found that the predicted equilibrium SOC stock was never reached over our simulation time of 400 years, that is, modelled SOC stocks in equilibrium were too low in transitional grasslands and too high in transitional croplands (cf. Supporting Information S1). There are two reasons for this pattern, which are discussed here for transitional grasslands. First, when the capacity of the fast-filling AGG pool is reached, any further increase is only controlled by the slower increase of the HUM pool. Second, the C input we used in the forward runs was calibrated for the higher grassland SOC stock. However, the forward runs are started from the SOC stock of cropland at equilibrium, which had a smaller IOM pool than the grassland equilibrium SOC stock. As the IOM is inert by definition and it neither receives C input nor decomposes, it cannot increase and SOC stocks starting from cropland can never reach the predicted grassland equilibrium. Alternatively, the IOM pool could be estimated based on the predicted cropland SOC stock for both land uses. This would result in modelled SOC stocks reaching the predicted equilibrium SOC stocks. However, we kept the IOM quantification by Falloon et al. (1998) in our approach because we interpret the IOM as a very slow pool whose changes do not have large effects in the time periods we usually model (10–100 years). Longer simulation periods would involve large uncertainties due to the increasingly uncertain development of, for instance, land management, crop breeding and climate (Bruni et al. 2021; Riggers et al. 2021). These uncertainties would exceed the uncertainty of the IOM estimation. An alternative estimation of the IOM pool would be required based purely on static soil parameters, such as the texture.

4.4 | Less Carbon Input Needed if Stabilisation Is Increased

Grassland C inputs calibrated using RothC-LUC were 25% lower (2.8 Mg C ha^{-1} a^{-1}) compared to RothC-default (3.8 Mg C ha^{-1} a^{-1}) and RothC-silt (3.7 Mg C ha^{-1} a^{-1}). This can be explained by the newly introduced AGG pool, which stabilised incoming carbon in the system by slowing down the total mineralisation in the system under grassland use. Therefore, less C input was needed to maintain the same SOC stock. The AGG pool added a new land-use-specific stabilisation process. This could represent aggregation (Poeplau et al. 2024; Spohn and Giani 2011), mycorrhizal fungi symbiosis (Zhang et al. 2025) or a shift in the microbial turnover towards a more stable anabolism (Liang et al. 2017; Liang and Zhu 2021) and an increase in microbial carbon use efficiency (Schroeder et al. 2022). Such a microbial stabilisation process under grassland was also found by Poeplau et al. (2019), who showed that SOC increases in grasslands could not be explained by C input increases alone, but were also attributed to stabilisation via microbial anabolism. Also, Hu and Chabbi (2022) found large SOC increases in French grasslands compared to cropland, which could not be explained by the amount or quality of C input estimated to be the same for both land uses. Instead, they attributed the SOC increases to a lower decomposition of SOC in permanent grasslands.

Separating the contribution of stabilisation mechanisms and C input on the overall SOC increase in (transitional) grasslands is challenging. Compared to croplands, estimating the C input in perennial grasslands is uncertain due to the uncertain input parameters. These include the root biomass, root turnover and rhizodeposition under a perennial vegetation cover. Consequently, estimates in the literature show a larger variability. Schulze et al. (2009) estimated a mean C input of 5.7 ± 1.7 Mg C $^{-1}$ ha^{-1} on European grasslands based on some eddy flux measurements. This was almost twice as high as their estimates for croplands (3.2 Mg C ha^{-1} a^{-1}). In contrast, Jacobs et al. (2020) estimated a mean C input of 3.7 ± 1.3 Mg C $^{-1}$ ha^{-1} for grassland sites from the German Agricultural Soil Inventory, which did not differ from

the average cropland C input (3.7 Mg C ha^{-1} a^{-1}). They assumed a fixed belowground C input of 2.2 Mg C $^{-1}ha^{-1}$ based on Poeplau et al. (2018), who sampled roots from seven long-term grassland experiments in Germany. A 50% lower belowground C input ranging between 1.0 and 1.9 Mg C $^{-1}ha^{-1}$ was found in French long-term grassland experiments (Hu and Chabbi 2022) based on root sampling. Compared to the highly variable estimates from the literature, our mean grassland C input calibrated using RothC-LUC seems low (2.8 ± 0.5 Mg C ha^{-1} a^{-1}), but not implausible.

4.5 | Quantifying the Aggregate Protected Carbon Pool and Limitations of RothC-LUC

When applying the RothC-LUC model, the capacity of the AGG pool needs to be estimated. In this study, it was calculated via the difference between equilibrium SOC stocks. However, in follow-up applications, where land-use specific equilibrium SOC stocks are unknown, it is necessary to estimate the capacity of the AGG pool independently from equilibrium SOC stocks. Based on our data set, we propose approximating the pool capacity via the pH measured in H_2O, which explained 24% of the variation, or via the sand content (%), explaining 16% of the variation (Figure 9). If no additional parameters are known, it is also possible to approximate the AGG capacity via the clay content. This explained 14% of the variation and is already needed to run RothC-default. These correlations stem from the machine-learning ensemble used to create the data set and are not empirically measured. However, other studies have found that the pH and clay content were among the main factors determining aggregate stability (Boix-Fayos et al. 2001; Poeplau et al. 2024; Rivera and Bonilla 2020) and increasing the pH via liming led to higher C loadings of small aggregate fractions in grasslands (Egan et al. 2018). Lime promotes aggregates and the stabilisation of SOC via Ca^{2+} bridging (Rowley et al. 2017). This could explain the higher capacity to stabilise carbon in soils with higher pH.

The pool structure of RothC-LUC may lead to unexpected patterns if equilibrium SOC stocks are too close together (< 5 Mg C ha^{-1}, cf. Supporting Information S8 and S9). However, applying this approach in the real world is not prone to this potential issue, since measured grassland and cropland SOC changes tend to be far in excess of this problematic threshold at > 10 t C ha^{-1} (Guo and Gifford 2002; Poeplau and Don 2013). The RothC-LUC model was calibrated and tested for Central European pedo-climatic conditions and intensive agricultural management. It could be necessary to modify it before using it in a larger range of pedo-climatic conditions and agricultural practices. Calibration and validation were done for topsoils (0–30 cm) where microbial decomposition, stabilisation, as well as C input by plants are main processes affecting SOC stocks. In the subsoil, other processes like advective or diffusive transport and reduced microbial activity add to the overall fate of subsoil C stocks (Sierra et al. 2024). These processes are not implemented in the models we tested, restricting the applicability of the RothC-LUC model to the topsoil.

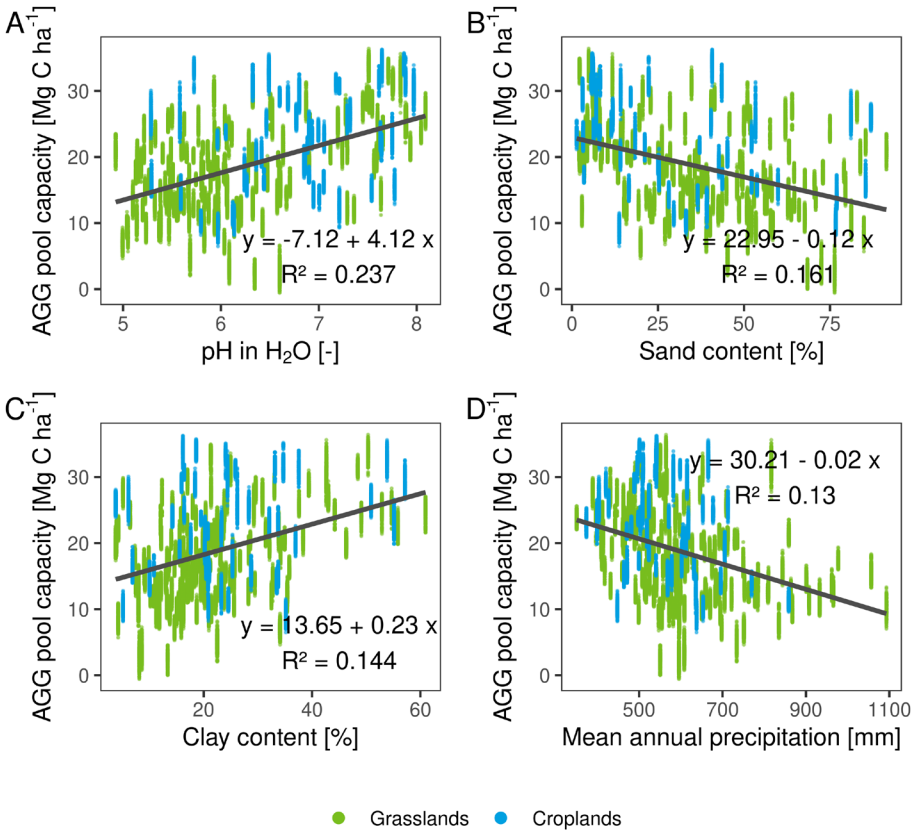

FIGURE 9 | The AGG pool capacity (Mg C ha^{-1}) over pH in H_2O (A), sand content (%, B), clay content (%, C) and mean annual precipitation (mm, D). Green and blue dots represent estimates for transitional grasslands and croplands, respectively. Grey lines represent linear regression models with the equation and R^2 given in each plot.

5 | Conclusions

National inventory reports under the United Nations Framework Convention on Climate Change are developing towards Tier-3 approaches, where greenhouse gas emissions are estimated using process-based models. RothC is one of these models and a popular choice to report fluxes of CO_2 between the soil-atmosphere interface. With the default model version, efforts to sequester C in soils will be underestimated, since C accrual under grassland is faster than described by the model. However, C loss will also be faster if C-rich ecosystems, such as grassland, are converted into C-depleted croplands. Such systematic bias in reporting greenhouse gas emissions can be avoided with our proposed RothC modification, RothC-LUC. Our study suggests that SOC dynamics following LUC are driven by stabilisation mechanisms that are not captured in the default model. This challenges the current perception that higher C inputs are the main reason for higher SOC stocks in grasslands.

Modelling offers the chance to explore complex systems for which direct measurements are not available, such as for soil carbon dynamics with timescales far beyond the project period and also beyond the working lifetime of a scientist. The temporal dimension of many decades to more than a century during which soils are affected by LUC was tackled in our study by combining empirical data on the status of SOC with land-use history timelines. Our combination of a novel data-driven space-for-time approach with process-based models will allow the prediction of unprecedented soil carbon futures after LUC in an era of climate change.

Author Contributions

Daria Seitz: investigation, writing – original draft, formal analysis, software, visualization, writing – review and editing. **Rene Dechow:** conceptualization, methodology, investigation, software, validation, writing – review and editing. **David Emde:** data curation, software, writing – review and editing. **Florian Schneider:** methodology, data curation, writing – review and editing. **Axel Don:** conceptualization, supervision, validation, funding acquisition, writing – review and editing.

Acknowledgements

This study received funding from the European Union's Horizon 2020 Research and Innovation Programme (grant agreement no. 862695) as part of the EJP Soil Programme. Open Access funding enabled and organized by Projekt DEAL.

Conflicts of Interest

The authors declare no conflicts of interest.

Data Availability Statement

The data that support the findings of this study and code for the RothC-LUC model are openly available on Github (https://github.com/DSeitz/RothC-LUC) and Zenodo (https://doi.org/10.5281/zenodo.15853391). The soil data set from the German Agricultural Soil Inventory used in this study is available in the OpenAgrar repository (https://doi.org/10.3220/DATA20200203151139).

References

Ad-Hoc AG Boden. 2005. *Bodenkundliche Kartieranleitung*. Schweizerbart'sche Verlagsbuchhandlung.

Bivand, R. 2023. "classInt: Choose Univariate Class Intervals." R Package Version 0.4-10. https://CRAN.R-project.org/package=classInt.

Boix-Fayos, C., A. Calvo-Cases, A. Imeson, and M. Soriano-Soto. 2001. "Influence of Soil Properties on the Aggregation of Some Mediterranean Soils and the Use of Aggregate Size and Stability as Land Degradation Indicators." *Catena* 44, no. 1: 47–67. https://doi.org/10.1016/S0341-8162(00)00176-4.

Bruni, E., B. Guenet, Y. Huang, et al. 2021. "Additional Carbon Inputs to Reach a 4 Per 1000 Objective in Europe: Feasibility and Projected Impacts of Climate Change Based on Century Simulations of Long-Term Arable Experiments." *Biogeosciences* 18, no. 13: 3981–4004. https://doi.org/10.5194/bg-18-3981-2021.

Coleman, K., and D. Jenkinson. 1996. "RothC-26.3—A Model for the Turnover of Carbon in Soil." In *Evaluation of Soil Organic Matter Models*, 237–246. Springer-Verlag.

De Rosa, D., C. Ballabio, E. Lugato, M. Fasiolo, A. Jones, and P. Panagos. 2024. "Soil Organic Carbon Stocks in European Croplands and Grasslands: How Much Have We Lost in the Past Decade?" *Global Change Biology* 30, no. 1: e16992. https://doi.org/10.1111/gcb.16992.

De, M., J. A. Riopel, L. J. Cihacek, et al. 2020. "Soil Health Recovery After Grassland Reestablishment on Cropland: The Effects of Time and Topographic Position." *Soil Science Society of America Journal* 84, no. 2: 568–586. https://doi.org/10.1002/saj2.20007.

Dechow, R., U. Franko, T. Kätterer, and H. Kolbe. 2019. "Evaluation of the RothC Model as a Prognostic Tool for the Prediction of SOC Trends in Response to Management Practices on Arable Land." *Geoderma* 337: 463–478. https://doi.org/10.1016/j.geoderma.2018.10.001.

Dignac, M.-F., D. Derrien, P. Barré, et al. 2017. "Increasing Soil Carbon Storage: Mechanisms, Effects of Agricultural Practices and Proxies. A Review." *Agronomy for Sustainable Development* 37, no. 2: 14. https://doi.org/10.1007/s13593-017-0421-2.

DWD Climate Data Center. 2023a. "Gridded Monthly Sums of Incoming Shortwave Radiation (Global Radiation) on the Horizontal Plain for Germany Based on Ground and Satellite Measurements, Version V003."

DWD Climate Data Center. 2023b. "Raster der Monatsmittel der Lufttemperatur (2m) für Deutschland, Version v1.0."

DWD Climate Data Center. 2023c. "Raster der Monatssumme der Niederschlagshöhe für Deutschland, Version v1.0."

Eddelbuettel, D., R. Francois, J. Allaire, et al. 2025. "Rcpp: Seamless R and C++ Integration." R Package Version 1.0.14.

Egan, G., M. J. Crawley, and D. A. Fornara. 2018. "Effects of Long-Term Grassland Management on the Carbon and Nitrogen Pools of Different Soil Aggregate Fractions." *Science of the Total Environment* 613: 810–819. https://doi.org/10.1016/j.scitotenv.2017.09.165.

Embry, I., A. Hoos, and T. Diehl. 2023. "ie2misc: Irucka Embry's Miscellaneous USGS Functions." R Package Version 0.9.1. https://CRAN.R-project.org/package=ie2misc.

Emde, D., C. Poeplau, A. Don, S. Heilek, and F. Schneider. 2024. "The Centennial Legacy of Land-Use Change on Organic Carbon Stocks of German Agricultural Soils." *Global Change Biology* 30, no. 8: e17444. https://doi.org/10.1111/gcb.17444.

Falloon, P. 2001. "Large Scale Spatial Modelling of Soil Organic Matter Dynamics." Accessed April 22, 2025. https://eprints.nottingham.ac.uk/12338/1/364398.pdf.

Falloon, P., P. Smith, K. Coleman, and S. Marshall. 1998. "Estimating the Size of the Inert Organic Matter Pool From Total Soil Organic

Carbon Content for Use in the Rothamsted Carbon Model." *Soil Biology and Biochemistry* 30, no. 8–9: 1207–1211. https://doi.org/10.1016/S0038-0717(97)00256-3.

Fisher, W. D. 1958. "On Grouping for Maximum Homogeneity." *Journal of the American Statistical Association* 53, no. 284: 789–798.

Gottschalk, P., J. Bellarby, C. Chenu, et al. 2010. "Simulation of Soil Organic Carbon Response at Forest Cultivation Sequences Using 13C Measurements." *Organic Geochemistry* 41, no. 1: 41–54. https://doi.org/10.1016/j.orggeochem.2009.04.017.

Guillaume, T., L. Bragazza, C. Levasseur, Z. Libohova, and S. Sinaj. 2021. "Long-Term Soil Organic Carbon Dynamics in Temperate Cropland-Grassland Systems." *Agriculture, Ecosystems & Environment* 305: 107184. https://doi.org/10.1016/j.agee.2020.107184.

Guo, L. B., and R. M. Gifford. 2002. "Soil Carbon Stocks and Land Use Change: A Meta Analysis." *Global Change Biology* 8, no. 4: 345–360. https://doi.org/10.1046/j.1354-1013.2002.00486.x.

Haddaway, N. R., K. Hedlund, L. E. Jackson, et al. 2017. "How Does Tillage Intensity Affect Soil Organic Carbon? A Systematic Review." *Environmental Evidence* 6, no. 1: 30. https://doi.org/10.1186/s13750-017-0108-9.

Harrell, F. E., Jr. 2023. "Hmisc: Harrell Miscellaneous." R Package Version 5.1-1. https://CRAN.R-project.org/package=Hmisc.

Heikkinen, J., I. Kurganova, V. Lopes de Gerenyu, T. Palosuo, and K. Regina. 2014. "Changes in Soil Carbon Stock After Cropland Conversion to Grassland in Russian Temperate Zone: Measurements Versus Model Simulation." *Nutrient Cycling in Agroecosystems* 98, no. 1: 97–106. https://doi.org/10.1007/s10705-014-9599-8.

Hernangómez, D. 2025. "{giscoR}: Download Map Data From GISCO API—Eurostat (Version 0.4.0)."

Hu, T., and A. Chabbi. 2022. "Grassland Management and Integration During Crop Rotation Impact Soil Carbon Changes and Grass-Crop Production." *Agriculture, Ecosystems & Environment* 324: 107703. https://doi.org/10.1016/j.agee.2021.107703.

Jacobs, A., C. Poeplau, C. Weiser, A. Fahrion-Nitschke, and A. Don. 2020. "Exports and Inputs of Organic Carbon on Agricultural Soils in Germany." *Nutrient Cycling in Agroecosystems* 118, no. 3: 249–271. https://doi.org/10.1007/s10705-020-10087-5.

Jenks, G. F. 1977. *Optimal Data Classification for Choropleth Maps*. Department of Geographiy, University of Kansas Occasional Paper.

Johnston, A. E., P. R. Poulton, and K. Coleman. 2009. "Soil Organic Matter: Its Importance in Sustainable Agriculture and Carbon Dioxide Fluxes." *Advances in Agronomy* 101: 1–57. https://doi.org/10.1016/S0065-2113(08)00801-8.

Kätterer, T., M. A. Bolinder, O. Andrén, H. Kirchmann, and L. Menichetti. 2011. "Roots Contribute More to Refractory Soil Organic Matter Than Above-Ground Crop Residues, as Revealed by a Long-Term Field Experiment." *Agriculture, Ecosystems & Environment* 141, no. 1–2: 184–192. https://doi.org/10.1016/j.agee.2011.02.029.

Lang, M., M. Binder, J. Richter, et al. 2019. "mlr3: A Modern Object-Oriented Machine Learning Framework in R." *Journal of Open Source Software* 4, no. 44: 1903. https://doi.org/10.21105/joss.01903.

Launay, C., J. Constantin, F. Chlebowski, et al. 2021. "Estimating the Carbon Storage Potential and Greenhouse Gas Emissions of French Arable Cropland Using High-Resolution Modeling." *Global Change Biology* 27, no. 8: 1645–1661. https://doi.org/10.1111/gcb.15512.

Le Noë, J., S. Manzoni, R. Abramoff, et al. 2023. "Soil Organic Carbon Models Need Independent Time-Series Validation for Reliable Prediction." *Communications Earth & Environment* 4, no. 1: 158. https://doi.org/10.1038/s43247-023-00830-5.

Li, L., T. Awada, Y. Zhang, and K. Paustian. 2024. "Global Land Use Change and Its Impact on Greenhouse Gas Emissions." *Global Change Biology* 30, no. 12: e17604. https://doi.org/10.1111/gcb.17604.

Liang, C., and X. Zhu. 2021. "The Soil Microbial Carbon Pump as a New Concept for Terrestrial Carbon Sequestration." *Science China Earth Sciences* 64, no. 4: 545–558. https://doi.org/10.1007/s11430-020-9705-9.

Liang, C., J. P. Schimel, and J. D. Jastrow. 2017. "The Importance of Anabolism in Microbial Control Over Soil Carbon Storage." *Nature Microbiology* 2: 17105. https://doi.org/10.1038/nmicrobiol.2017.105.

Liao, J. J., and J. W. Lewis. 2000. "A Note on Concordance Correlation Coefficient." *PDA Journal of Pharmaceutical Science and Technology* 54, no. 1: 23–26.

Lin, L. I.-K. 1989. "A Concordance Correlation Coefficient to Evaluate Reproducibility." *Biometrics* 45: 255–268.

Meyer, D., E. Dimitriadou, K. Hornik, A. Weingessel, and F. Leisch. 2023. "e1071: Misc Functions of the Department of Statistics, Probability Theory Group (Formerly: E1071), TU Wien." https://doi.org/10.32614/CRAN.package.e1071.

Meyer, H., and E. Pebesma. 2021. "Predicting Into Unknown Space? Estimating the Area of Applicability of Spatial Prediction Models." *Methods in Ecology and Evolution* 12, no. 9: 1620–1633. https://doi.org/10.1111/2041-210x.13650.

Meyer, H., C. Milà, M. Ludwig, J. Linnenbrink, and F. Schumacher. 2024. "CAST: 'caret' Applications for Spatial-Temporal Models." https://doi.org/10.32614/CRAN.package.CAST.

Nash, J. E., and J. V. Sutcliffe. 1970. "River Flow Forecasting Through Conceptual Models Part I—A Discussion of Principles." *Journal of Hydrology* 10, no. 3: 282–290. https://doi.org/10.1016/0022-1694(70)90255-6.

Paustian, K., N. H. Ravindranath, and A. van Amstel. 2006. "2006 IPCC Guidelines for National Greenhouse Gas Inventories." Volume 4: Agriculture, Forestry and Other Land Use. https://www.ipcc-nggip.iges.or.jp/public/2006gl/index.html.

Pebesma, E. 2018. "Simple Features for R: Standardized Support for Spatial Vector Data." *R Journal* 10, no. 1: 439–446. https://doi.org/10.32614/RJ-2018-009.

Pedersen, T. L. 2024. "patchwork: The Composer of Plots." R Package Version 1.2.0.

Peralta, G., L. Di Paolo, I. Luotto, et al. 2022. *Global Soil Organic Carbon Sequestration Potential Map (GSOCseq v1. 1)—Technical Manual*. Food and Agriculture Organization.

Poeplau, C., A. Don, and F. Schneider. 2021. "Roots Are Key to Increasing the Mean Residence Time of Organic Carbon Entering Temperate Agricultural Soils." *Global Change Biology* 27, no. 19: 4921–4934. https://doi.org/10.1111/gcb.15787.

Poeplau, C., A. Don, L. Vesterdal, et al. 2011. "Temporal Dynamics of Soil Organic Carbon After Land-Use Change in the Temperate Zone—Carbon Response Functions as a Model Approach." *Global Change Biology* 17, no. 7: 2415–2427. https://doi.org/10.1111/j.1365-2486.2011.02408.x.

Poeplau, C., A. Jacobs, A. Don, et al. 2020. "Stocks of Organic Carbon in German Agricultural Soils—Key Results of the First Comprehensive Inventory." *Journal of Plant Nutrition and Soil Science* 183, no. 6: 665–681. https://doi.org/10.1002/jpln.202000113.

Poeplau, C., and A. Don. 2013. "Sensitivity of Soil Organic Carbon Stocks and Fractions to Different Land-Use Changes Across Europe." *Geoderma* 192: 189–201. https://doi.org/10.1016/j.geoderma.2012.08.003.

Poeplau, C., D. Zopf, B. Greiner, et al. 2018. "Why Does Mineral Fertilization Increase Soil Carbon Stocks in Temperate Grasslands?"

Agriculture, Ecosystems & Environment 265: 144–155. https://doi.org/10.1016/j.agee.2018.06.003.

Poeplau, C., M. Helfrich, R. Dechow, et al. 2019. "Increased Microbial Anabolism Contributes to Soil Carbon Sequestration by Mineral Fertilization in Temperate Grasslands." *Soil Biology and Biochemistry* 130: 167–176. https://doi.org/10.1016/j.soilbio.2018.12.019.

Poeplau, C., T. Riefling, M. Schiedung, and R. Anlauf. 2024. "Land Use and Soil Property Effects on Aggregate Stability Assessed by Three Different Slaking Methods." *European Journal of Soil Science* 75, no. 4: e13549. https://doi.org/10.1111/ejss.13549.

R Core Team. 2022. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing.

Rasse, D. P., C. Rumpel, and M.-F. Dignac. 2005. "Is Soil Carbon Mostly Root Carbon? Mechanisms for a Specific Stabilisation." *Plant and Soil* 269, no. 1: 341–356. https://doi.org/10.1007/s11104-004-0907-y.

Reichstein, M., G. Camps-Valls, B. Stevens, et al. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566, no. 7743: 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Ridgeway, G., and GBM Developers. 2024. "gbm: Generalized Boosted Regression Models." https://doi.org/10.32614/CRAN.package.gbm.

Riggers, C., C. Poeplau, A. Don, C. Bamminger, H. Höper, and R. Dechow. 2019. "Multi-Model Ensemble Improved the Prediction of Trends in Soil Organic Carbon Stocks in German Croplands." *Geoderma* 345: 17–30. https://doi.org/10.1016/j.geoderma.2019.03.014.

Riggers, C., C. Poeplau, A. Don, C. Frühauf, and R. Dechow. 2021. "How Much Carbon Input Is Required to Preserve or Increase Projected Soil Organic Carbon Stocks in German Croplands Under Climate Change?" *Plant and Soil* 460: 417–433. https://doi.org/10.1007/s11104-020-04806-8.

Rivera, J. I., and C. A. Bonilla. 2020. "Predicting Soil Aggregate Stability Using Readily Available Soil Properties and Machine Learning Techniques." *Catena* 187: 104408. https://doi.org/10.1016/j.catena.2019.104408.

Rowley, M. C., S. Grand, and É. P. Verrecchia. 2017. "Calcium-Mediated Stabilisation of Soil Organic Carbon." *Biogeochemistry* 137, no. 1–2: 27–49. https://doi.org/10.1007/s10533-017-0410-1.

Sakhaee, A., A. Gebauer, M. Ließ, and A. Don. 2022. "Spatial Prediction of Organic Carbon in German Agricultural Topsoil Using Machine Learning Algorithms." *Soil* 8, no. 2: 587–604. https://doi.org/10.5194/soil-8-587-2022.

Sakhaee, A., T. Scholten, R. Taghizadeh-Mehrjardi, M. Ließ, and A. Don. 2024. "Spatial Prediction of Organic Matter Quality in German Agricultural Topsoils." *Agriculture* 14, no. 8: 1298. https://doi.org/10.3390/agriculture14081298.

Sanderman, J., T. Hengl, and G. J. Fiske. 2017. "Soil Carbon Debt of 12,000 Years of Human Land Use." *Proceedings of the National Academy of Sciences of the United States of America* 114, no. 36: 9575–9580. https://doi.org/10.1073/pnas.1706103114.

Schneider, F., C. Poeplau, and A. Don. 2021. "Predicting Ecosystem Responses by Data-Driven Reciprocal Modelling." *Global Change Biology* 27, no. 21: 5670–5679. https://doi.org/10.1111/gcb.15817.

Schroeder, J., T. Peplau, F. Pennekamp, E. Gregorich, C. C. Tebbe, and C. Poeplau. 2022. "Deforestation for Agriculture Increases Microbial Carbon Use Efficiency in Subarctic Soils." *Biology and Fertility of Soils* 60, no. 1: 17–34. https://doi.org/10.1007/s00374-022-01669-2.

Schulze, E. D., S. Luyssaert, P. Ciais, et al. 2009. "Importance of Methane and Nitrous Oxide for Europe's Terrestrial Greenhouse-Gas Balance." *Nature Geoscience* 2, no. 12: 842–850. https://doi.org/10.1038/ngeo686.

Sierra, C. A., B. Ahrens, M. A. Bolinder, et al. 2024. "Carbon Sequestration in the Subsoil and the Time Required to Stabilize Carbon for Climate Change Mitigation." *Global Change Biology* 30, no. 1: e17153. https://doi.org/10.1111/gcb.17153.

Sierra, C. A., M. Müller, and S. E. Trumbore. 2012. "Models of Soil Organic Matter Decomposition: The SoilR Package, Version 1.0." *Geoscientific Model Development* 5, no. 4: 1045–1060. https://doi.org/10.5194/gmd-5-1045-2012.

Signorell, A. 2024. "DescTools: Tools for Descriptive Statistics." R Package Version 0.99.54. https://CRAN.R-project.org/package=DescTools.

Sokol, N. W., S. E. Kuebbing, E. Karlsen-Ayala, and M. A. Bradford. 2019. "Evidence for the Primacy of Living Root Inputs, Not Root or Shoot Litter, in Forming Soil Organic Carbon." *New Phytologist* 221, no. 1: 233–246. https://doi.org/10.1111/nph.15361.

Spohn, M., and L. Giani. 2011. "Impacts of Land Use Change on Soil Aggregation and Aggregate Stabilizing Compounds as Dependent on Time." *Soil Biology and Biochemistry* 43, no. 5: 1081–1088. https://doi.org/10.1016/j.soilbio.2011.01.029.

Statistics Finland. 2024. "Greenhouse Gas Emissions in Finland 1990 to 2022." Accessed April 22, 2025. https://unfccc.int/sites/default/files/resource/FI_NID_UN_2022_2024-12-16.pdf.

Swedish Environmental Protection Agency. 2024. "National Inventory Report Sweden 2024." Accessed April 22, 2025. https://unfccc.int/sites/default/files/resource/NIR%20SE%20submission%202024.pdf.

Szoboszlay, M., A. B. Dohrmann, C. Poeplau, A. Don, and C. C. Tebbe. 2017. "Impact of Land-Use Change and Soil Organic Carbon Quality on Microbial Diversity in Soils Across Europe." *FEMS Microbiology Ecology* 93, no. 12: fix146. https://doi.org/10.1093/femsec/fix146.

Turc, L. 1961. "Estimation of Irrigation Water Requirements, Potential Evapotranspiration: A Simple Climatic Formula Evolved Up to Date." *Annales Agronomiques* 12, no. 1: 13–49.

Wickham, H., M. Averick, J. Bryan, et al. 2019. "Welcome to the Tidyverse." *Journal of Open Source Software* 4, no. 43: 1686. https://doi.org/10.21105/joss.01686.

Wright, M. N., and A. Ziegler. 2017. "Ranger: A Fast Implementation of Random Forests for High Dimensional Data in C++ and R." *Journal of Statistical Software* 77, no. 1: 1–17. https://doi.org/10.18637/jss.v077.i01.

Wüst-Galley, C., S. G. Keel, and J. Leifeld. 2020. "A Model-Based Carbon Inventory for Switzerland's Mineral Agricultural Soils Using RothC." Accessed April 22, 2025. https://ira.agroscope.ch/de-CH/publication/45789.

Zhang, E., Y. Wang, T. W. Crowther, et al. 2025. "Mycorrhiza Increases Plant Diversity and Soil Carbon Storage in Grasslands." *Proceedings of the National Academy of Sciences of the United States of America* 122, no. 7: e2412556122. https://doi.org/10.1073/pnas.2412556122.

Supporting Information

Additional supporting information can be found online in the Supporting Information section.