

# Data from long-term experiments in temperate croplands to evaluate soil organic carbon models

Received: 12 September 2025

Accepted: 6 February 2026

Cite this article as: Fujisaki, K., Ferchaud, F., Clivot, H. *et al.* Data from long-term experiments in temperate croplands to evaluate soil organic carbon models. *Sci Data* (2026). <https://doi.org/10.1038/s41597-026-06863-7>

Kenji Fujisaki, Fabien Ferchaud, Hugues Clivot, Elisa Bruni, Bertrand Guenet, Christian Pichot, Antoine Versini, François Baudin, Antonio Bispo, Philippe Peylin, Manuel P. Martin, Johannes L. Jensen, Jørgen Eriksen, Claire Chenu, Andrew S. Gregory, Margaret J. Glendining, Ines Merbach, Nicolas Beaudoin, Bruno Mary, Alain Mollier, Gilles Tison, Christophe Montagnier, Abad Chabbi, Françoise Vertès, Alice Cadéro, Anne-Isabelle Graux, Sylvain Pellerin, Florent Levavasseur, Manon Gilles, Thierry Morvan, Camille Resseguier, Luis Milesi, Alicia Irizar, Adrià Andriulo, Marie-Noël Mistou, Arnaud Butier, Michel Bertrand, Bénédicte Autret, Marie-Hélène Jeuffroy, Gilles Grandeau, Thierry Doré, Vincent Cellier, Alain Berthier, Sébastien Darras, Guillaume Audebert, Ludovic Pasquier, Fabien Ecalte, Antoine Savoie, Marcus Schiedung, Christopher Poeplau, Nadia I. Maaroufi, Thomas Kätterer, Martin A. Bolinder, Jonathan Sanderman & Pierre Barré

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

## Title

Data from long-term experiments in temperate croplands to evaluate soil organic carbon models

## Abstract

Soil organic carbon (SOC) models need independent evaluation against field measurements, but those latter are rarely publicly available and harmonized. In this study, we collected and shared data from 167 agronomic treatments in 34 agronomic long-term experiments (LTEs) located in temperate croplands, allowing the evaluation of several soil organic C models such as RothC, Century, AMG, MIMICS, ICBM, Millennial, and CTOOL. The dataset includes climate data, soil properties, C inputs from crops (n=4588 records) and organic amendments, irrigation data, monthly soil cover, as well as SOC stock measurements in the topsoil layer (n=1328 records). Climate, soil moisture, and soil temperature data were extracted from daily climate databases. Carbon inputs from crops were calculated from observed yields and harvest index, with some harvest index values estimated, combined with crop allometric coefficients from the literature. Descriptions of LTE, agronomic treatments, methodological metadata, and a part of the code, accompanies the dataset. The dataset can be reused to evaluate single SOC models, or to evaluate an ensemble of models.

## Authors & Affiliations

Kenji Fujisaki <sup>1</sup>, Fabien Ferchaud <sup>2,13</sup>, Hugues Clivot <sup>3</sup>, Elisa Bruni <sup>4</sup>, Bertrand Guenet <sup>4</sup>, Christian Pichot <sup>5</sup>, Antoine Versini <sup>6</sup>, François Baudin <sup>7</sup>, Antonio Bispo <sup>1</sup>, Philippe Peylin <sup>8</sup>, Manuel P. Martin <sup>1</sup>, Johannes L. Jensen <sup>9</sup>, Jørgen Eriksen <sup>9</sup>, Claire Chenu <sup>10</sup>, Andrew S. Gregory <sup>11</sup>, Margaret J. Glendining <sup>11</sup>, Ines Merbach <sup>12</sup>, Nicolas Beaudoin <sup>13</sup>, Bruno Mary <sup>13</sup>, Alain Mollier <sup>14</sup>, Gilles Tison <sup>15</sup>, Christophe Montagnier <sup>10</sup>, Abad Chabbi <sup>10,16</sup>, Françoise Vertès <sup>17</sup>, Alice Cadéro <sup>18</sup>, Anne-Isabelle Graux <sup>18</sup>, Sylvain Pellerin <sup>14</sup>, Florent Levasseur <sup>10</sup>, Manon Gilles <sup>19</sup>, Thierry Morvan <sup>17</sup>, Camille Resseguier <sup>10</sup>, Luis Milesi <sup>20</sup>, Alicia Irizar <sup>20</sup>, Adriàn Andriulo <sup>20</sup>, Marie-Noël Mistou <sup>21</sup>, Arnaud Butier <sup>21</sup>, Michel Bertrand <sup>21</sup>, Bénédicte Autret <sup>22</sup>, Marie-Hélène Jeuffroy <sup>21</sup>, Gilles Grandeau <sup>21</sup>, Thierry Doré <sup>21</sup>, Vincent Cellier <sup>23</sup>, Alain Berthier <sup>23</sup>, Sébastien Darras <sup>24</sup>, Guillaume Audebert <sup>16</sup>, Ludovic Pasquier <sup>16</sup>, Fabien Ecalle <sup>16</sup>, Antoine Savoie <sup>25</sup>, Marcus Schiedung <sup>26</sup>, Christopher Poeplau <sup>26</sup>, Nadia I. Maaroufi <sup>27</sup>, Thomas Kätterer <sup>28</sup>, Martin A. Bolinder <sup>28</sup>, Jonathan Sanderman <sup>29</sup>, Pierre Barré <sup>4</sup>

<sup>1</sup> INRAE, Info&Sols, 45075, Orléans, France

<sup>2</sup> UMR Eco&Sols, Univ Montpellier, CIRAD, INRAE, IRD, Institut Agro, Montpellier, France

<sup>3</sup> Université de Reims Champagne-Ardenne, INRAE, FARE, Reims, France

- <sup>4</sup> Laboratoire de Géologie, École Normale Supérieure, CNRS, PSL Université, IPSL, Paris, France
- <sup>5</sup> INRAE, URFM, Avignon, 84000, France
- <sup>6</sup> CIRAD, UPR Recyclage et risque, 34000 Montpellier, France
- <sup>7</sup> Sorbonne Université, CNRS, IStEP 75005 Paris, France
- <sup>8</sup> Laboratoire des sciences du climat et de l'environnement (LSCE), IPSL, CEA/CNRS/UVSQ, Gif-sur-Yvette, France
- <sup>9</sup> Department of Agroecology, Aarhus University, Blichers Allé 20, Tjele 8830, Denmark
- <sup>10</sup> Université Paris-Saclay, INRAE, AgroParisTech, UMR ECOSYS, 91120 Palaiseau, France
- <sup>11</sup> Rothamsted Research, Harpenden, Hertfordshire, AL5 2JQ, United Kingdom
- <sup>12</sup> Department of Community Ecology, Helmholtz Centre for Environmental Research-UFZ, D-06120 Halle, Germany
- <sup>13</sup> BioEcoAgro Joint Research Unit, INRAE, Université de Liège, Université de Lille, Université de Picardie Jules Verne, 02000, Barenton-Bugny, France
- <sup>14</sup> INRAE, Bordeaux Sciences Agro, UMR ISPA, 33140, Villenave d'Ornon, France
- <sup>15</sup> INRAE, UE APC, Castanet-Tolosan, France
- <sup>16</sup> Unité de Recherche Pluridisciplinaire Prairies et Plantes Fourragères, INRAE - Centre de Recherche Nouvelle-Aquitaine-Poitiers, Lusignan, France
- <sup>17</sup> INRAE, Institut Agro, UMR SAS, 35000, Rennes, France
- <sup>18</sup> INRAE, Institut Agro, UMR PEGASE, 35590, Saint-Gilles, France
- <sup>19</sup> INRAE, UEAV, 68000, Colmar, France
- <sup>20</sup> INTA, Pergamino, Buenos Aires, Argentina
- <sup>21</sup> Université Paris-Saclay, INRAE, AgroParisTech, UMR Agronomie, 91120, Palaiseau, France
- <sup>22</sup> INRAE, Unité AgroSystèmes Territoires Ressources (ASTER), 88500, Mirecourt, France
- <sup>23</sup> INRAE, U2E, 21110, Bretenière, France
- <sup>24</sup> INRAE, GCIE, 80200, Estrées-Mons, France
- <sup>25</sup> INRAE, 11297 UE PAO, 37380, Nouzilly, France
- <sup>26</sup> Thünen Institute of Climate-Smart Agriculture, Braunschweig, Germany
- <sup>27</sup> Department of Soil and Environment, Swedish University of Agricultural Sciences, Uppsala, Sweden
- <sup>28</sup> Department of Ecology, Swedish University of Agricultural Sciences, Uppsala, Sweden
- <sup>29</sup> Woodwell Climate Research Center, Falmouth, Massachusetts, USA

Corresponding author: Kenji Fujisaki ([kenji.fujisaki@inrae.fr](mailto:kenji.fujisaki@inrae.fr))

## Background & Summary

Soil carbon (C) models are relevant tools to predict C dynamics in terrestrial ecosystems under future global changes. These models need nevertheless independent evaluation and validation using data from controlled long-term experiments (LTEs) with repeated measurements of soil organic C (SOC) stocks <sup>1</sup>. These data are labor expensive and time-consuming to acquire, as SOC changes slowly and soils are heterogeneous. Therefore, initiatives that disseminate and promote the use of data from LTEs for soil C modeling are crucial <sup>2</sup>. However, gathering and harmonizing input data for soil C modeling is a tedious task, and the input datasets supporting modeling research are rarely publicly available <sup>3</sup>. This data paper takes a step in this forward to overcome such constraints. We collected and estimated input data from 167 agronomic treatments in 34 LTEs, mainly located in France (n=23 LTEs), where annual crops and temporary grasslands were grown. The data matches required data needed to run and evaluate soil C models such as RothC <sup>4</sup>, Century <sup>5</sup>, AMG <sup>6</sup>, MIMICS <sup>7</sup>, ICBM <sup>8</sup>, Millennial <sup>9</sup>, and CTOOL <sup>10</sup>. The dataset includes climate data, soil properties, C inputs from crops and from organic amendments, irrigation, along with a description of the LTEs and the agronomic treatments. We also collected SOC stock measurements in the topsoil layer (ranging from 0-10 cm to 0-34 cm depth, depending on the site) at several sampling dates in these LTEs, allowing the validation of the SOC changes simulated by the models. The resulting dataset is freely available and reusable.

## Methods

The dataset was conceived according to the following steps. First, we gathered the input variables required to run the C models RothC, Century, AMG, MIMICS, ICBM, Millennial, and CTOOL, as well as the necessary time step associated with each variable. Then, we defined a common template to organize the datasets from the LTEs. We grouped the variables according to their category (climate, soil, C inputs, etc.) and to their required time step. Soil C models often need input variables that are not directly measured in the field and hence need to be estimated from intermediate variables or from independent models. This is for instance the case for the C inputs, which were calculated from crop yields and crop allometric coefficients. The template includes the variables used by the models as well as intermediate variables. The template is included in the dataset (soil\_carbon\_models\_template.xlsx file). The empty tables matching the template were then filled with observed data from the LTEs, depending on data availability. Some variables needed to be estimated, as explained further for each table. Some examples of data retrieving and estimation of variables are also included in the dataset. Part of the data gathered in the dataset were accessible and retrieved from several published datasets: this is the case for the LTEs of Kerbernez <sup>11</sup>, EFELE <sup>12</sup>, Colmar

PROspective<sup>13-16</sup>, SIC<sup>17</sup>, Rothamsted Highfield Bare Fallow<sup>18</sup>, Rothamsted Broadbalk<sup>19-23</sup>, Ultuna<sup>24</sup>, Lanna<sup>25</sup>, Lönnstorp<sup>25</sup>, NELITCSE<sup>26</sup>, Waite<sup>27</sup>, and KBS<sup>28,29</sup>.

### Climate data

Climate data were obtained for each LTE using gridded datasets. Average daily and monthly air temperature, daily and monthly precipitations, and monthly reference evapotranspiration (calculated with the Penman-Monteith method<sup>30</sup> which requires net radiation, air temperature, wind speed, vapor pressure deficit, and atmospheric pressure) were extracted for the periods between the first and last SOC stock measurement. For sites located in mainland France, weather data were extracted from the SAFRAN reanalysis<sup>31</sup> (source: Météo-France), which delivers gridded daily outputs on a 8×8 km grid, from 1958 to the present (<https://www.data.gouv.fr/fr/datasets/donnees-changement-climatique-sim-quotidienne>, last accessed 11<sup>th</sup> August 2025). SAFRAN data were downloaded and processed using an Application Programming Interface (API) developed by GéoSAS ([https://geoslas.fr/web/?page\\_id=6345](https://geoslas.fr/web/?page_id=6345), last accessed 11<sup>th</sup> August 2025). For sites located in other countries, weather data were extracted and processed from the ERA5<sup>32</sup> and ERA5-Land<sup>33</sup> reanalyses. These reanalysis products were extracted using the Open-Meteo API<sup>34</sup> with the openmeteo package in R<sup>35</sup>. Average daily air temperature was extracted from the ERA5-Land reanalysis, delivered on a 11×11 km grid, from 1950 to the present<sup>33</sup>. Daily precipitation was extracted from the ERA5 daily reanalysis, delivered on a 25×25 km grid, from 1940 to the present<sup>32</sup>. Reference evapotranspiration available through the Open-Meteo API is calculated with both ERA5 and ERA5-Land data, since some ERA-Land variables are purely interpolated from ERA5 and not a result of the ERA5-Land replay. This is the case for precipitation, short wave radiation, and wind speed, which were internally retrieved from ERA5 by the Open-Meteo API and combined with air temperature retrieved from ERA5-Land to calculate reference evapotranspiration.

Daily soil temperature and soil moisture data were extracted from the ERA5-Land through the Open-Meteo API, with the same method detailed above for climate data. Daily soil temperature and moisture are available for the 0-7 cm and 7-28 cm soil layers; they were weight-averaged to generate data for the 0-28 cm layer.

Climate, soil moisture, and soil temperature data were also extracted before the first measured SOC stock data as it is needed to initialize SOC pool sizes in some models. SAFRAN, ERA5, and ERA5-Land data were extracted with the methodology described above but then averaged over several decades. Mean daily air temperature, daily precipitation, monthly reference evapotranspiration, soil moisture, and soil temperature were calculated for each day and month of the year by averaging all available daily values using data extracted 30 years before the first SOC stock measurement. For some sites, data extraction 30 years before the beginning of the experiment was not possible since SAFRAN and ERA5 data were not

available before 1958 and 1940, respectively. For some sites, the period of data extraction was therefore reduced to 16 years. For the oldest LTEs (Folleville, Issoudun, Grignon LTBF, Versailles LTBF), average weather data were calculated using data extracted from SAFRAN from 1959 to 1989. Details of the period considered for the calculation of normal climate data before the beginning of the experiments are provided in the `climate_site_metadata` table.

For the Rothamsted LTEs site, where the first SOC stock was estimated for 1843, ERA5-Land data are not available in the first decades of the experiment. Therefore, we used the climate data collected at the Rothamsted station (<https://www.era.rothamsted.ac.uk/station/rms#datasets>), available since the beginning of the experiment. However, only rainfall data were collected at the beginning of the experiment. We thus used the climate data from 1915, where at least wind speed, air temperature, and rainfall were measured. Reference evapotranspiration (Penman-Monteith or FAO-56 method) was calculated using an established method<sup>36</sup>. To match the timeline of SOC stocks and weather data in Rothamsted Broadbalk LTE, SOC stocks data were taken from 1914 to 2010. The climate data for the year 1914, initially missing, was estimated by averaging daily climate data between 1915 and 1945.

### Soil organic carbon stocks

Soil organic carbon (SOC) stocks were computed for each LTE, at each sampling date and for each agronomic treatment. SOC stocks were expressed in  $\text{Mg ha}^{-1}$  and summed for the topsoil layer, which ranged from 0-10 cm to 0-34 cm depth, depending on the tillage depth. SOC stocks were calculated according to equivalent soil mass principles where possible, in order to take into account possible variation of bulk density over time<sup>37</sup>. The equivalent soil mass correction was mainly done with the SimpleESM R script<sup>38</sup>. For some LTEs, bulk density measurements were not available and hence were estimated using pedo-transfer functions. The methodology used to calculate SOC stocks (bulk density assessment and SOC stocks calculation methods) is provided in the `soc` table along with SOC stocks values.

### *Thermal fractionation*

For some LTEs, centennially stable and active fractions of SOC measured by Rock-Eval® thermal analysis<sup>39</sup> were available (e.g. ref<sup>40</sup>) and were therefore gathered in the dataset, as these data can be used to evaluate or initialize the soil C models. The partition between centennially stable and active fractions of SOC was achieved with the PARTY<sub>SOC</sub>V2.0<sub>EU</sub> model<sup>41</sup>. In the dataset, we reported the stock of the active and centennially stable SOC fraction.

### Physical fractionation

For Lanna, Lönnstorp, Kerbernez, and Versailles LTBF LTEs, physical SOC fractions data were available. We reported in the dataset the relative proportion of organic C (OC) associated with fine fraction ( $<50 \mu\text{m}$ ;  $oc\_fine\_fraction$  in  $\text{g g}^{-1}$ ), the stock of OC ( $oc\_fine\_stock$  in  $\text{Mg ha}^{-1}$ ), and the stock of OC associated with the coarse fraction ( $> 50\mu\text{m}$ ,  $oc\_coarse\_stock$  in  $\text{Mg ha}^{-1}$ ). To derive the two fractions for Lanna, Lönnstorp and Kerbernez LTEs, suspended soils were treated with  $100 \text{ J ml}^{-1}$  of ultrasonication followed by wet sieving to  $50 \mu\text{m}$  to obtain the coarse and fine fraction, as described by Begill et al. <sup>42</sup> and Just et al. <sup>43</sup>. For Versailles LTBF LTE, physical SOC fractions data come from Lutfalla et al. <sup>44</sup>. Soil samples were fractionated into sand, silt, and clay fractions. The coarse fraction  $> 50 \mu\text{m}$  was separated by sieving at  $50 \mu\text{m}$  after dispersion of the sand-sized aggregates (overnight shaking with 20 glass beads in 180 mL of deionized water). The silt and clay fractions were then separated by centrifugation after ultrasonic dispersion. The clay and silt fractions were combined in the present dataset in the OC fine fraction.

### Soil properties

Soil properties (particle size distribution, pH,  $\text{CaCO}_3$  content, C:N ratio) were collected for each site in the topsoil layer. Clay content is expressed after sample decarbonation, as required by the AMG model. pH was measured in water. pH measurement method was not harmonized in the dataset, with soil to water ratio varying from 1:1 to 1:5; this ratio was included in the `site_metadata` table. C:N ratio was calculated from contents assessed by dry combustion methods. Field capacity and wilting point were only available for a few LTEs (Fagnières LTBF, Bad Lauchstädt LTBF, Ultuna, and NELITCSE). If not available, these variables were estimated with pedotransfer functions. We used Eq. 1 and Eq. 2 which were calibrated and validated in French croplands <sup>45</sup>:

$$fc = 0.278 + 0.00245 \times clay\_decarb - 0.00315 \times sand \quad (1)$$

$$wp = 0.08 + 0.00401 \times clay\_decarb - 0.000293 \times sand \quad (2)$$

where  $fc$  and  $wp$  are the volumetric soil moisture ( $\text{cm}^3 \text{ cm}^{-3}$ ) at field capacity ( $\text{pF}=2.0$ ) and wilting point ( $\text{pF}=4.2$ ), respectively, with  $\text{pF}$  denoting the logarithm (base 10) of the soil water tension in cm of water, and  $clay\_decarb$  and  $sand$  are the clay and sand contents ( $\text{g } 100 \text{ g}^{-1}$ ), respectively. For the PROspective LTE where  $\text{CaCO}_3$  content in soil is about  $130 \text{ g kg}^{-1}$ ,  $fc$  and  $wp$  were calculated with particle-size fractions measured on samples without decarbonation <sup>46</sup>.

## Carbon inputs

### *Carbon inputs from crops*

The C inputs from crops were calculated using the approach of Clivot et al. <sup>47</sup>, which adapted the Bolinder et al. <sup>48</sup> to the French cropping context.

Yields of harvested products from each main crop, cover crop, harvested cover crop, or weeds (if measured), were expressed as Mg dry matter (DM) ha<sup>-1</sup>. Some yield data were occasionally missing and were estimated using averages for the crop in the corresponding treatment. These estimations were sometimes needed to match the timeline of SOC stocks measurements.

Harvest index (HI) of each crop, defined as the fraction of the harvested yield product in the total aboveground biomass was derived from measurements if available (for annual grain crops, measurements are done by cutting the aboveground biomass at ground level at the time of the harvest), or from default values for French crops <sup>47,49</sup>.

Aboveground C inputs (*abg\_c\_input*) for main crops were calculated from yield and HI, and expressed in Mg C ha<sup>-1</sup> (Eq. 3):

$$abg\_c\_input = yield \times \frac{(1 - HI)}{HI} \times residue\_fraction \times residue\_c\_cont \times 0.001 \quad (3)$$

where *yield* is the yield of the agricultural product for main crops and harvested cover crops (Mg DM ha<sup>-1</sup>), *residue\_fraction* is the fraction of non-harvested biomass returned to soil (unitless; between 0 and 1) either measured or based on fixed values per crop, when part of the residues are exported from the field, and *residue\_c\_cont* is the C content of the non-harvested biomass, typically crop residues, either measured or from a default value for French crops (440 g C kg<sup>-1</sup>).

As there is no agricultural product for cover crops and weeds, HI was set to 0 for these crops and therefore not used in the calculation of *abg\_c\_input* for these crops, and *residue\_fraction* was set to 1 (Eq. 4):

$$abg\_c\_input = yield \times residue\_fraction \times residue\_c\_cont \times 0.001 \quad (4)$$

Belowground C inputs by roots (*root\_c\_input*) were calculated from crop specific shoot:root ratios, as no measurements of root biomass or rhizodeposition were available. First, the C inputs from root biomass at harvest were calculated as follows and expressed in Mg C ha<sup>-1</sup> (Eq. 5):

$$root\_c\_input = \frac{yield}{shoot\_root \times HI} \times root\_c\_cont \times 0.001 \quad (5)$$

where *yield* is the yield of the agricultural product for main crops and harvested cover crops, or the aboveground biomass produced by the cover crops and weeds (Mg DM ha<sup>-1</sup>), *shoot\_root* is the shoot:root ratio of the crop (unitless), and *root\_c\_cont* is the C content of the roots (g C kg<sup>-1</sup>). The shoot:root values were taken from the literature <sup>47,49</sup>. As *root\_c\_cont* was not available, we used a default value of 400 g C

kg<sup>-1</sup>. As there is no agricultural product for cover crops and weeds, HI was set to 0 for these crops and therefore not used in the calculation of *root\_c\_input* for these crops (Eq. 6):

$$root\_c\_input = \frac{yield}{shoot\_root} \times root\_c\_cont \times 0.001 \quad (6)$$

The extra-root C, produced by root turnover and rhizodeposition (*rhizodep\_c\_input*), was calculated as follow, according to Bolinder et al.<sup>48</sup> (Eq. 7):

$$rhizodep\_c\_input = root\_c\_input \times 0.65 \quad (7)$$

The sum of *root\_c\_input* and *rhizodep\_c\_input* represents the total belowground C input in the whole soil profile, but it should be scaled to match the sampling depth of SOC stock measurements. This adjustment was done with an asymptotic equation representing the root distribution profile of each crop<sup>50,51</sup>. The belowground C inputs *bg\_c\_input* was therefore adjusted for sampling depth and calculated as follows and expressed in Mg C ha<sup>-1</sup> (Eq. 8):

$$bg\_c\_input = (root\_c\_input + rhizodep\_c\_input) \times (1 - \beta^{sampling\ depth}) \quad (8)$$

where  $\beta$  is the crop coefficient for root distribution along the whole soil profile, ranging from 0 to 1, and *sampling depth* is the sampling depth used for SOC stock calculation (cm).

Lignin:N ratio of the crop parts (aboveground and root parts), used in the Century model, were taken or estimated from a dataset of chemical quality of crop residues<sup>52</sup>. We associated for each crop a humification coefficient for aboveground carbon inputs that can be used in the the AMG model (table *h\_coef\_crop.csv* table), taken from a published AMG parameters dataset<sup>49</sup>.

The name of crops in the dataset were harmonized using a naming convention with only subscript letters. When applicable, crops were distinguished between spring-sown and winter-sown crops. When possible, we mapped crop names in the *crop\_allom.csv* table to AGROVOC thesaurus<sup>53</sup> URIs.

#### *Carbon inputs from exogenous organic matter applications*

Several LTEs in the dataset included applications of exogenous organic matter (EOM) such as manures, slurries, sludges, composts, digestates, etc. C inputs from EOM (*c\_eom\_input*) were calculated as (Eq. 9):

$$c\_eom\_input = eom\_fm \times eom\_dm\_cont \times eom\_c\_cont \quad (9)$$

where *eom\_fm* is the amount of EOM applied on the field (Mg fresh matter ha<sup>-1</sup>), *eom\_dm\_cont* is the dry matter content of the fresh EOM product (g kg<sup>-1</sup>), and *eom\_c\_cont* is the C content of the dried EOM product (g kg<sup>-1</sup>).

Each EOM input was associated with an application date, either recorded (e.g. from published datasets<sup>12,23</sup>) or estimated, using the LTE descriptions found in the literature. Lignin:N ratio of EOM were taken from

measured values if available, or estimated from a dataset of C & N mineralization data of several EOMs<sup>54</sup>. Humification coefficients of EOMs for AMG model were taken from measurements if available, or published values<sup>55</sup>. The EOM type names were harmonized by distinguishing the type of product and their origin, such as *bovine manure*, *pig slurry*, *sewage sludge*, etc.

### Monthly soil cover

Monthly soil cover is a parameter used by the RothC model<sup>4</sup>. This parameter was computed for each site and treatment based on the sowing and harvest dates of each crop. For most sites, sowing and harvest dates were available. For some sites, sowing and harvest dates were estimated from usual regional practices. When no crop was growing on the field, the *soil\_cover* value was set to 0, whereas when a crop was growing on the field, *soil\_cover* value was set to 1.

### Irrigation

Irrigation data were reported at daily time step, as irrigation affects the water balance in the field which has an impact on modelled soil C dynamics.

### Metadata collection

#### *Site metadata*

Each site of the LTEs was characterized in the table *site\_metadata.csv*, using the template developed by Bonares and EJP soil project<sup>56</sup>. The variables in this table include location of the experiment, type of experiment, research themes, institution responsible of the experiment, soil type, land use type, experimental design, and dates of beginning and ending. Some fields in the table were not completed as they were not relevant for soil C modeling, however we added to this table several fields relevant for soil C models: the dominant plant functional type of the experiment; the land use before the beginning of the experiment, and if relevant, the date of land use change. This information is useful to initialize soil C models.

#### *Treatments metadata*

We characterized each agronomic treatment in the dataset, with basic information regarding soil management practices: mineral nitrogen fertilization (yes/no), organic fertilization (yes/no), irrigation (yes/no), presence of cover crop in the crop rotation (yes/no) and identification of bare fallow treatments. We also characterized the tillage intensity of each agronomic treatment with three classes<sup>57,58</sup>: high-intensity tillage (presence of inversion tillage like ploughing or non-inversion tillage practices at 40 cm

depth or below ), intermediate intensity tillage (tillage events without inversion and above 40 cm depth), and no-tillage.

## Data Record

The dataset is accessible from a ZIP archive deposited on Recherche Data Gouv<sup>59</sup>. The tables of the dataset are stored in three folders: data, parameters, and metadata. The dataset contains 18 csv tables listed in Table 1. The description of variables in the dataset is provided in the *variables\_metadata.csv* table in the metadata folder. A code folder provides R scripts allowing the reproduction of some tables, as detailed in Methods and Code availability sections.

Table 1. Description of tables in the dataset

Table name	Table description	Parameters
climate_data_daily	Daily climate parameters for each site during the experiment	Temperature (°C), precipitation (mm)
climate_data_monthly	Monthly climate parameters for each site during the experiment	Temperature (°C), precipitation (mm), Reference evapotranspiration (mm)
climate_data_daily_spinup	Daily normal temperature and precipitation 16 to 30 years before the beginning of each experiment	Temperature (°C), precipitation (mm)
climate_data_monthly_spinup	Monthly normal temperature and precipitation 16 to 30 years before the beginning of each experiment	Reference evapotranspiration (mm)
soil_moist_temp	Daily soil temperature and moisture in topsoil	Soil temperature (°C), soil moisture (cm <sup>3</sup> cm <sup>-3</sup> )
soil_moist_temp_spinup	Daily normal soil temperature and moisture in topsoil 16 to 30 years before the beginning of each experiment	Soil temperature (°C), soil moisture (cm <sup>3</sup> cm <sup>-3</sup> )
soil_properties	Soil properties for each LTE	Particle size distribution, soil pH, CaCO <sub>3</sub> content (g kg <sup>-1</sup> ), C:N ratio, wilting point (cm <sup>3</sup> cm <sup>-3</sup> ), field capacity (cm <sup>3</sup> cm <sup>-3</sup> )
soc	Measurements of SOC stocks	Sampling date, upper and lower sampling depth (cm), SOC stocks (Mg ha <sup>-1</sup> ), thermal fractions, physical fractions
c_inputs_crop	C inputs from crops for each crop cycle in the dataset	Crop name, sowing and harvest dates, yield (Mg ha <sup>-1</sup> ), harvest index, C content of crop residues (g kg <sup>-1</sup> ), aboveground C inputs (Mg ha <sup>-1</sup> ), belowground C inputs (Mg ha <sup>-1</sup> )

crop_allom	Parameters for each crop, used to estimate C inputs or needed by some C models	Crop type, crop name, crop allometry, default harvest index, shoot:root ratio, Beta parameter
h_coef_crop	Humification coefficient for crops	Crop name, AMG humification coefficient
c_inputs_eom	C inputs from each application of exogenous organic matter (EOM)	EOM type, application date, C input, lignin:N ratio, AMG humification coefficient
irrigation	Daily irrigation data for each irrigation event	Date, irrigation amount (mm),
soil_cover	Monthly soil cover during each experiment	Year, month, soil cover (0 or 1)
site_metadata	Metadata of each site	Location, periods of experiment, institution, goal of the LTE, experimental design, type of vegetation, soil type, previous land use
treatments_info	Characteristics of each agronomic treatment	Tillage intensity, bare fallow, organic amendments, mineral nitrogen fertilization, cover crop
climate_site_metadata	Climate data source and method to derive climate data before the beginning of experiment for each site	Site, climate data sources
variables_metadata	Description of dataset variables	Variables names, description, time step, unit

The dataset gathers data from 34 LTEs, including 167 agronomic treatments. The LTE are described in Table 2 and locations are shown in Figure 1. Most of the LTE are located in France (n=23), other LTE come from United Kingdom (n=2), United States (n=2), Sweden (n=3), Denmark (n=1), Germany (n=1), Australia (n=1), and Argentina (n=1). The duration of data on these LTEs varies from 7 to 96 years. The main research themes of the LTEs are cropping systems experiments (n=7), bare fallow experiments (n=7), and mineral fertilization experiments (n=5). In the 167 agronomic treatments, 37 received EOM amendments, 14 treatments were irrigated, 11 treatments were bare fallows, and 44 treatments included cover crops in the crop rotations. The dataset contains 1328 SOC stock records, and 4588 crop cycle records.

*Table 2 Long term experiments (LTEs) in the dataset. Lon. and Lat. Columns refer to coordinates of the LTEs; d is the duration in years, between the first and last SOC stock measurement in the dataset – some LTEs may have a larger duration than the duration in the dataset*

LTE	Country	Lon.	Lat.	d	Agronomic treatments	Ref.
auzeville	France	1.51	43.53	35	P fertilization	60
doazit	France	-0.63	43.70	12	K fertilization	61
folleville	France	1.94	48.84	19	P&K fertilization	62
kerbernez	France	-4.13	47.95	27	temporary grassland rotations	63

LTE	Country	Lon.	Lat.	d	Agronomic treatments	Ref.
mant	France	-0.50	43.59	17	P fertilization	64
tartas	France	-0.73	43.87	21	K fertilization	64
colmar_prospective	France	7.33	48.06	18	EOM application and N fert.	46
feucherolles_qualiagro	France	1.97	48.90	23	EOM application and N fert.	65
le_rheu_efele_pros	France	-1.79	48.11	9	EOM application and N fert.	66
le_rheu_efele_tsmo	France	-1.79	48.11	9	EOM application and soil tillage	66
lusignan_acbb	France	0.12	46.42	15	duration and fertilization of temporary grasslands	66
estrees_mons_sdc	France	3.03	49.87	7	crop rotation and soil tillage	67
versailles_la_cage	France	2.13	48.80	22	cropping system	68
grignon_sic	France	1.95	48.84	6	cropping system	69
issoudun	France	1.99	46.95	21	crop residue management	70
estrees_mons_be	France	3.01	49.87	13	bioenergy crops	71
pergamino	Argentina	-60.67	-33.85	39	soybean monoculture	6
breteniere_res0pest	France	5.10	47.24	10	cropping system	72
estrees_mons_res0pest	France	3.03	49.87	10	cropping system	72
lusignan_res0pest	France	0.12	46.42	10	cropping system	72
nouzilly_res0pest	France	0.79	47.54	10	cropping system	72
askov_ltbf	Denmark	9.12	55.47	29	long term bare fallow	73
bad_lauchstadt_ltbf	Germany	11.88	51.40	32	long term bare fallow	74
grignon_ltbf	France	1.92	48.85	48	long term bare fallow	73
rothamsted_ltbf	UK	-0.36	51.80	64	long term bare fallow	73
versailles_ltbf	France	2.13	48.80	80	long term bare fallow	73
fagnieres_ltbf	France	4.31	48.94	45	long term bare fallow	75
ultuna	Sweden	17.65	59.82	63	EOM application and N fert.	76
lanna	Sweden	13.13	58.35	28	crop residue management and N fert.	77
lonnstorp	Sweden	13.08	55.67	24	crop residue management and N fert.	77
broadbalk	UK	-0.37	51.81	95	fertilization and EOM application	78
nelitcse	USA	-96.47	40.85	22	tillage and cropping system	79

LTE	Country	Lon.	Lat.	d	Agronomic treatments	Ref.
waite	Australia	138.63	-34.97	30	crop rotation	80
kbs	USA	-84.40	42.40	10	cropping system	81

## Technical Validation

The dataset was checked by importing all tables in an R environment and running several verification tests. The *check\_dataset.R* script reproduces this first technical validation of the dataset. The absence of missing data was checked, as well as the consistency between LTE names and treatments across all tables. The timelines consistency between climate tables, soc table, and C input tables was also confirmed. However, for one LTE (“Cropping systems and soil structure” experiment), the timeline of SOC stocks measurements differed across treatments (n=28), causing minor mismatch between the different table timelines. Modelers using the dataset should therefore sometimes filter the climate tables to match exactly the SOC stock timeline.

The dataset was also validated by checking data distribution, outliers, and observed trends in SOC stocks and C inputs. The *dataset\_desc.RMD* file provided in the code folder of the dataset describes and checks the dataset with a few verification plots from the dataset. The location of each LTE was checked by plotting them on a map (Figure 1). The physical consistencies of climate forcings from processed climate data were also checked by plotting mean average air temperature and mean average precipitation (Figure 2). The *check\_dataset.R* file also provides code generating climate diagrams at monthly scale for all LTEs.

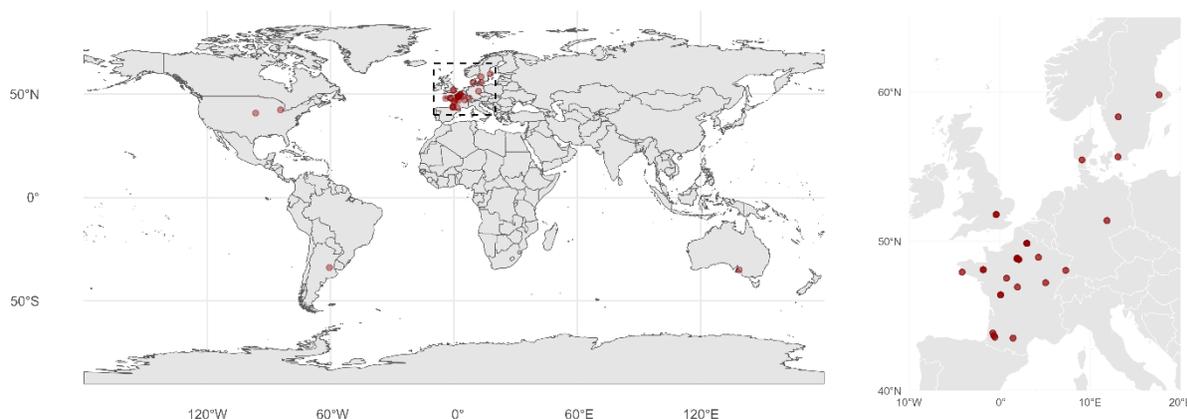


Figure 1. Location of the 34 LTEs. The right panel shows a detailed view of the European sites.

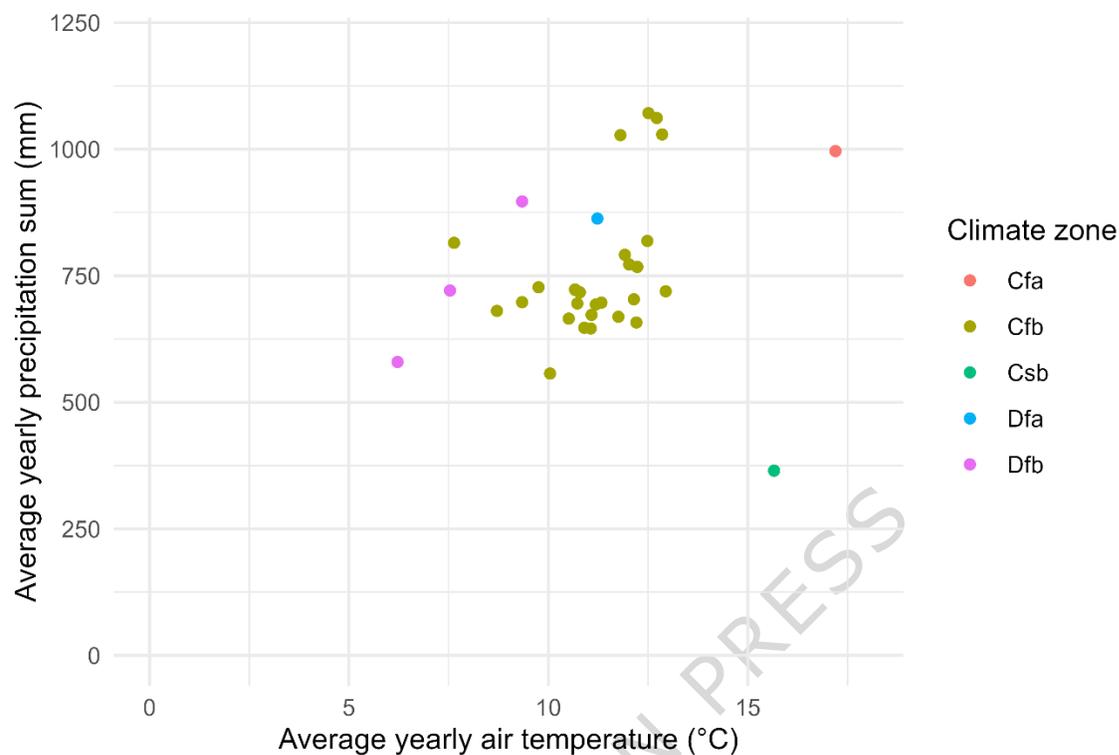


Figure 2. Average yearly climate data and Köppen-Geiger climate zones of the LTEs. Climate zone codes refer to Humid subtropical (Cfa), Temperate oceanic (Cfb), Warm-summer Mediterranean (Csb), Hot-summer humid continental (Dfa), and Warm-summer humid continental (Dfb).

The particle size distribution in the *soil\_properties.csv* table was checked and is represented in Figure 3. Most of the soils have large silt content, with only 6 LTEs having a silt content below 30 %.

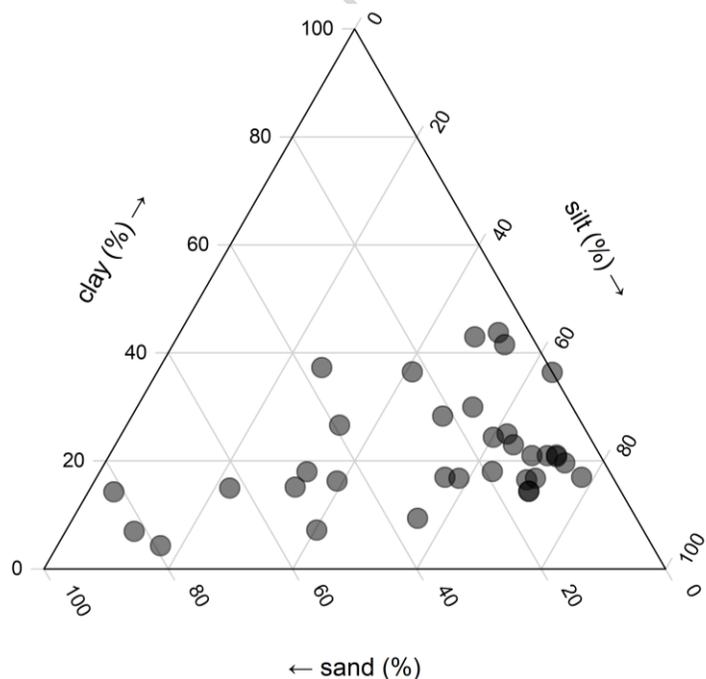


Figure 3. Particle size distribution in the LTEs.

The SOC stocks distribution and dynamics were checked with several plots. The initial SOC stocks distribution according to soil layer thickness was consistent, with smaller SOC stocks values for the shallowest soil layer thickness (Figure 4). The SOC stock dynamics over time, shown in Figure 5A, shows a wide range of variation, from sharp SOC decreases in bare fallow treatments, to large SOC stock increases observed in LTEs where large amounts of EOM were applied. On average, the SOC stock changes between the last and first measurement were slightly negative (Figure 5B).

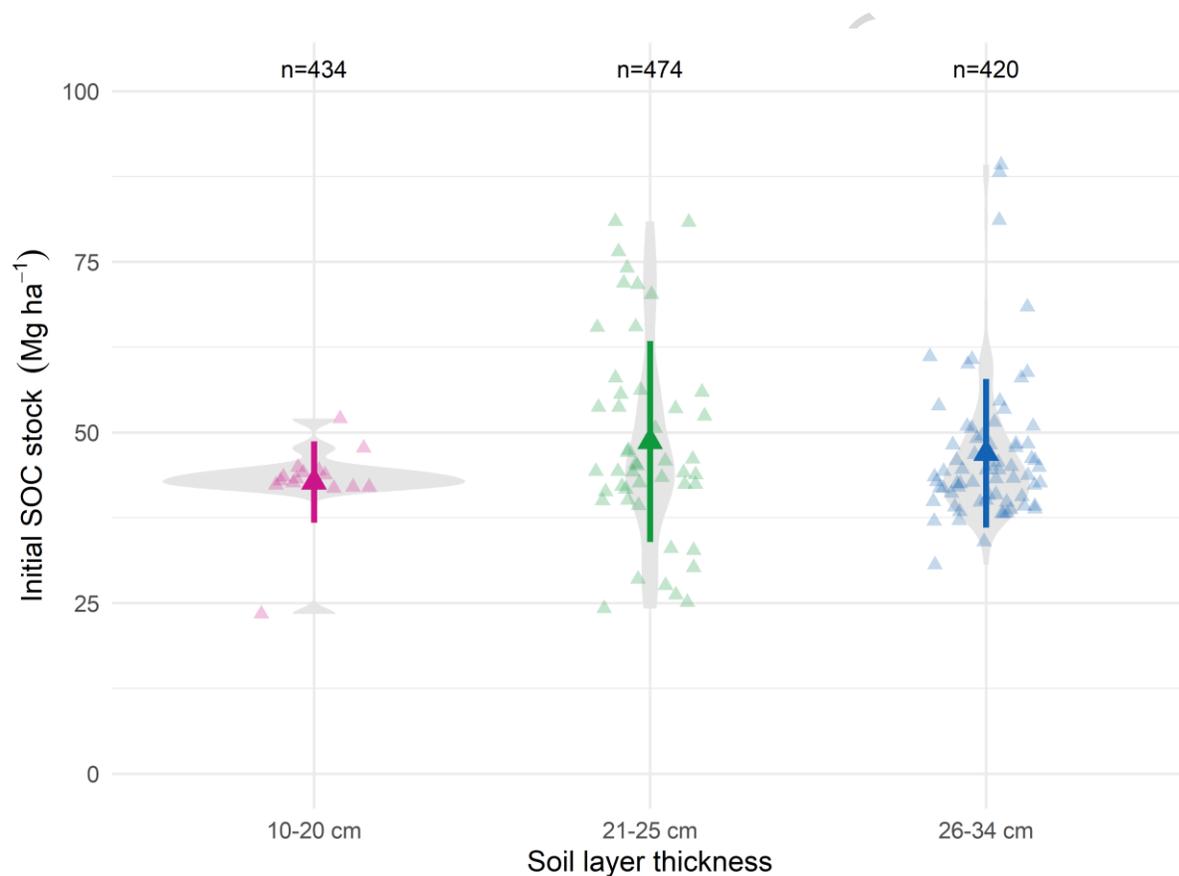


Figure 4. Initial soil organic carbon (SOC) stocks in the LTEs. Each triangle represents an observation for each agronomic treatment in the dataset. Grey area is a smoothed violin plot of SOC stock distribution, including the mean and standard deviation; the width of the violin plot reflects the density of observations.

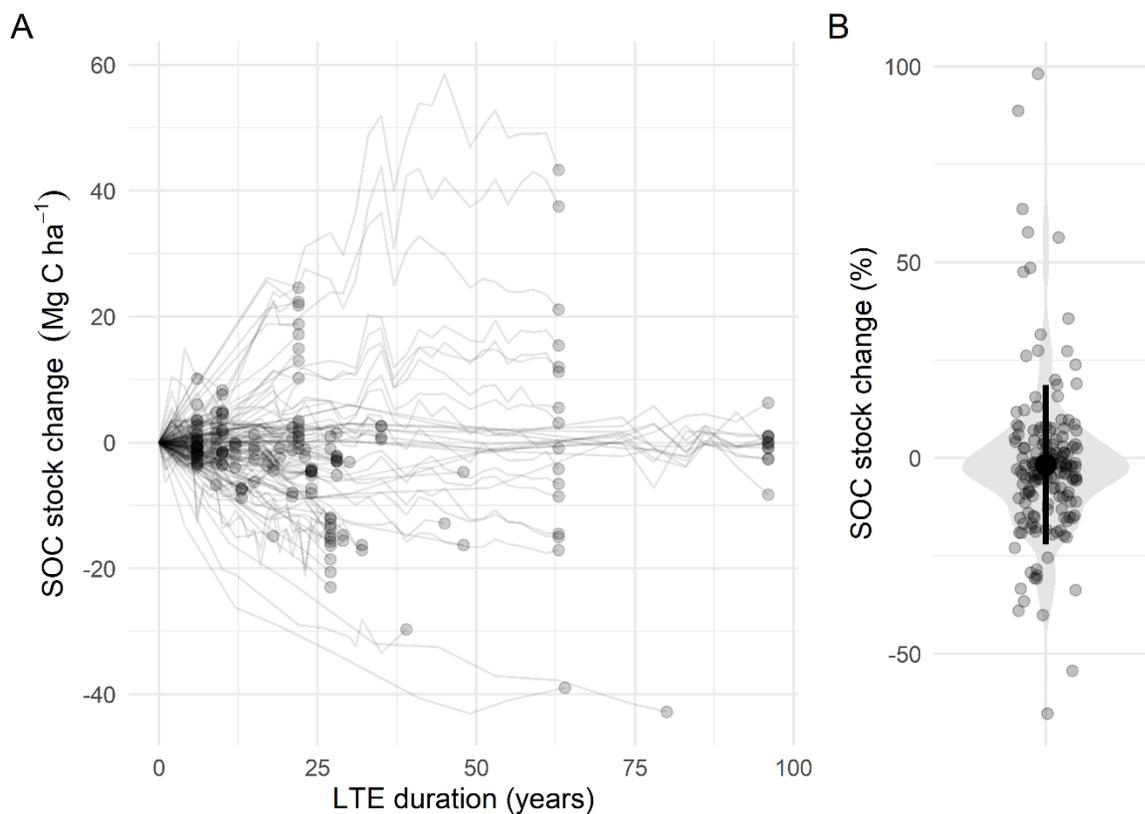


Figure 5. SOC stock changes over time in the LTEs relative to the initial stock (A), and distribution of SOC stock changes between last and first measurements (B). In panel A, intermediate observations between the first and last SOC stock measurements are masked; lines represent the pure interpolation between observations. In panel B, grey area is a smoothed violin plot of SOC stock changes, including the mean and standard deviation; the width of the violin plot reflects the density of observations.

Data relative to crops and C inputs from crops were also checked with basic tests (absence of NAs, date consistency, absence of duplication). The number of observations per crop and per crop type, shown in Figure 6, highlights that the most frequently occurring crops in the dataset are winter wheat, spring barley, silage maize, and grain maize. Total C inputs from crops (comprising aboveground and belowground parts) show a wide distribution, with Q1, median, and Q3 values of 0.5, 1.16, 2.45, and a mean of 1.74 Mg C ha<sup>-1</sup> yr<sup>-1</sup> (Figure 7). These C inputs are a slightly lower than current estimates derived from country-scale SOC modeling in croplands in Western Europe, such as 2.14 Mg C ha<sup>-1</sup> yr<sup>-1</sup> reported for France<sup>82</sup> and 2.7 Mg C ha<sup>-1</sup> yr<sup>-1</sup> for Germany<sup>83</sup>. This discrepancy is largely attributable to the significant weight of the Broadbalk (UK) and Ultuna (Sweden) LTEs in the dataset, which are characterized by low carbon inputs from crops due to limited yields and/or intensive crop residue removal. The distribution of C inputs from aboveground or belowground parts is shown in Figure 8. The consistency of calculated aboveground C inputs from crops was checked by comparing the aboveground C inputs with crop yields and fraction of crop residues returned

to the field (Figure 9). Overall, the aboveground C inputs are, as expected, smaller when the crop residues are exported from the field, and for crops with a large HI (e.g. silage maize). There are however cases where aboveground C inputs are small even with large crop residue retention (e.g., spring barley, spring oat, and winter rapeseed); these data points derive from the Ultuna LTE, where all crop residues are exported from the field<sup>84</sup>. For this LTE, the yield was considered to be the total aboveground biomass, associated with an estimated HI of 0.95. This effectively results in smaller aboveground C inputs.

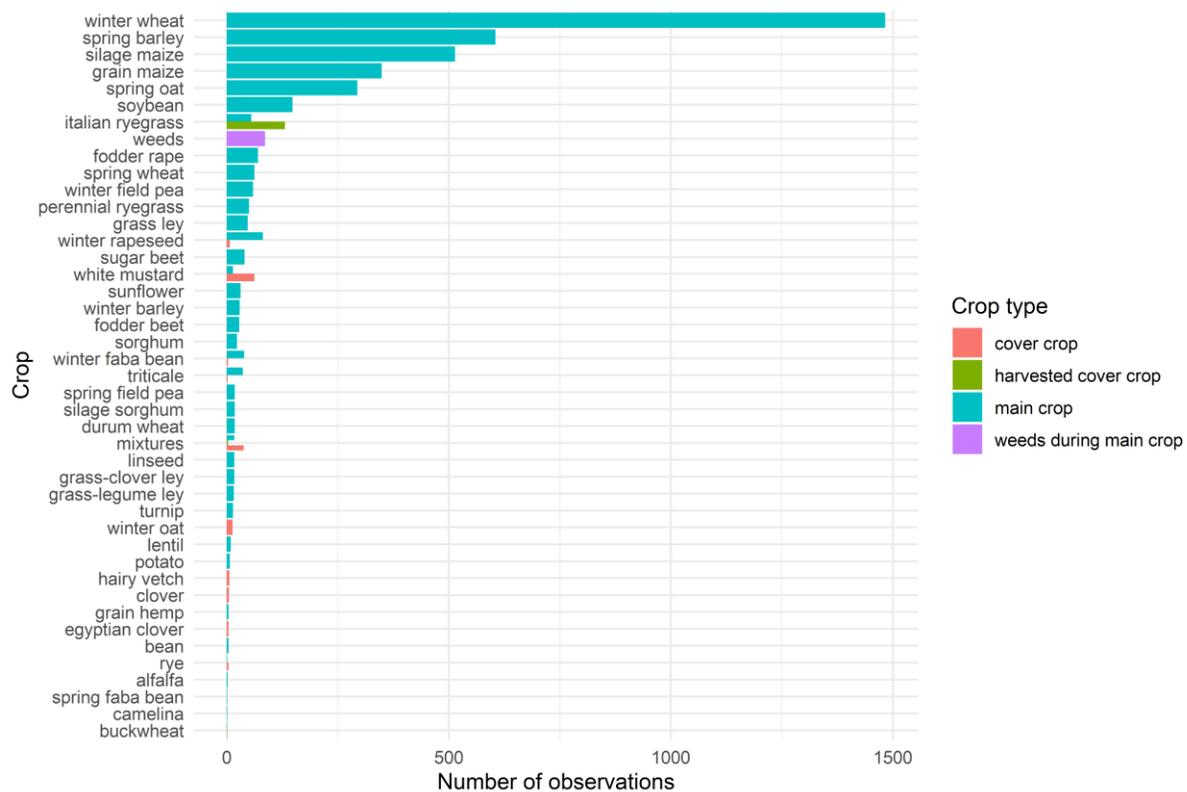


Figure 6. Number of observations per crop and crop type.

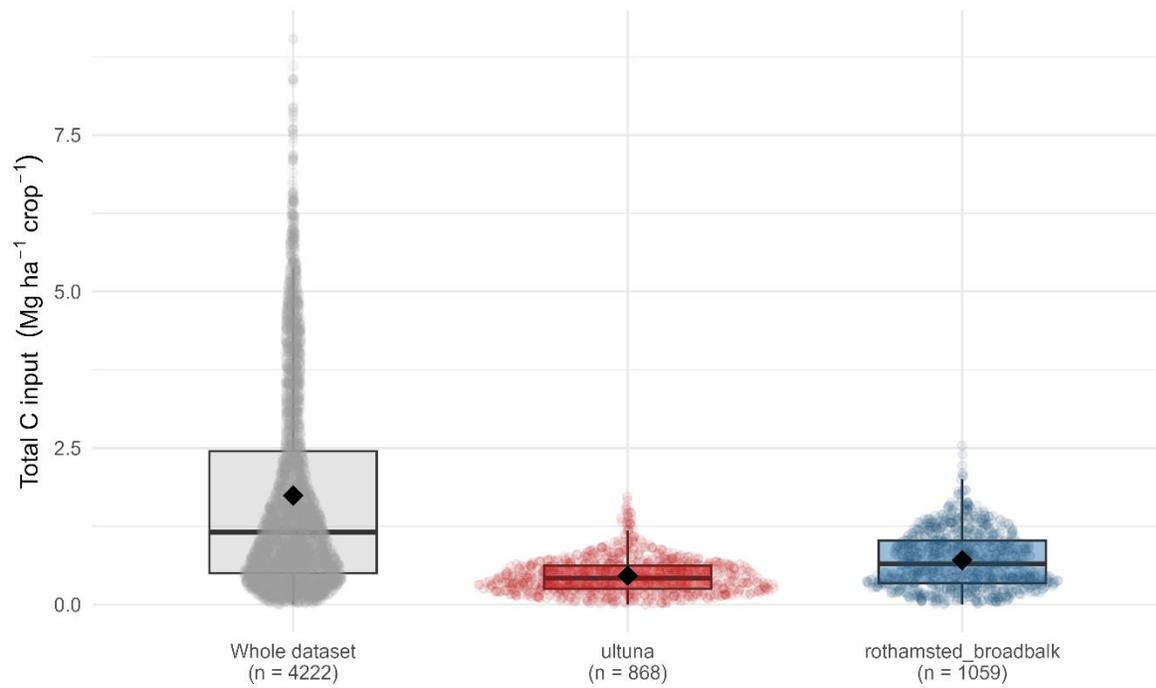


Figure 7. Distribution of total carbon inputs from crops in the whole dataset and in Ultuna and Broadbalk LTEs. Data points are shown as a sina plot; boxplots show medians and interquartile ranges, while diamonds indicate means.

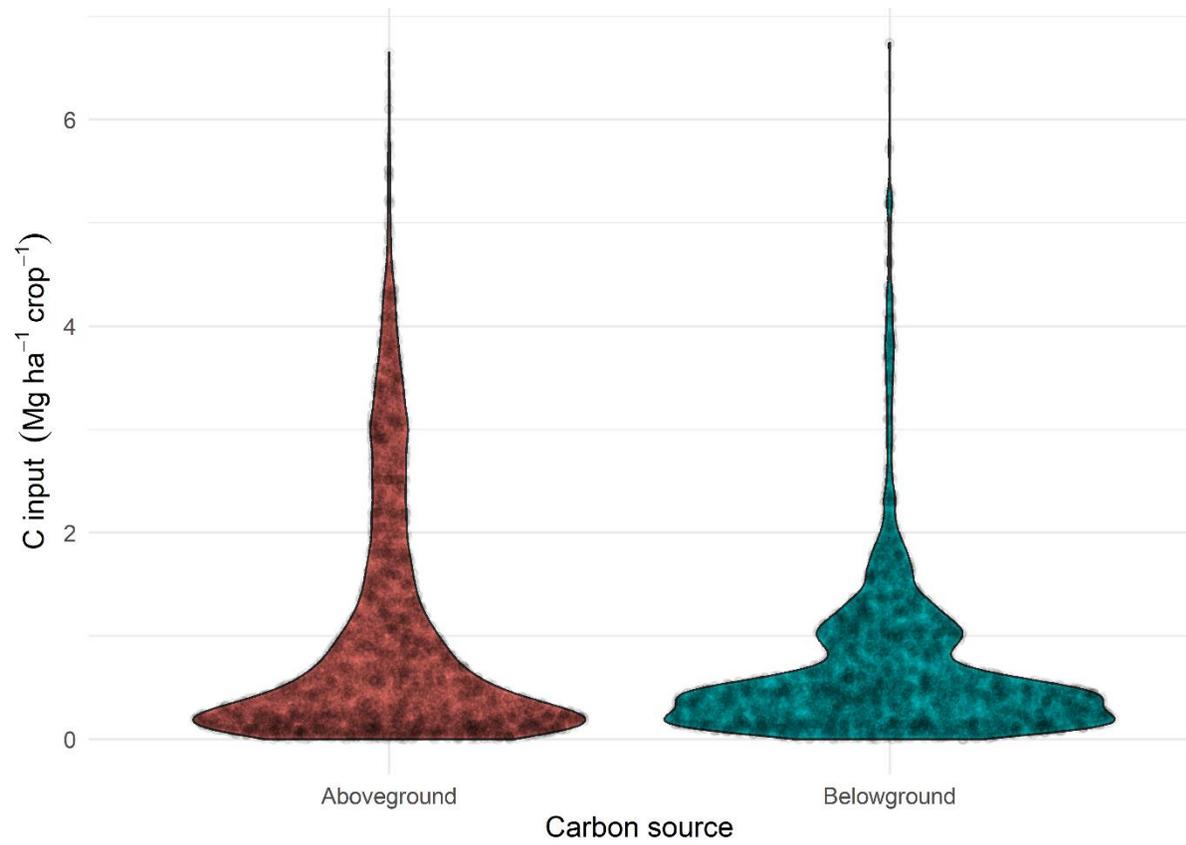


Figure 8. Distribution of aboveground and belowground C inputs from crops.

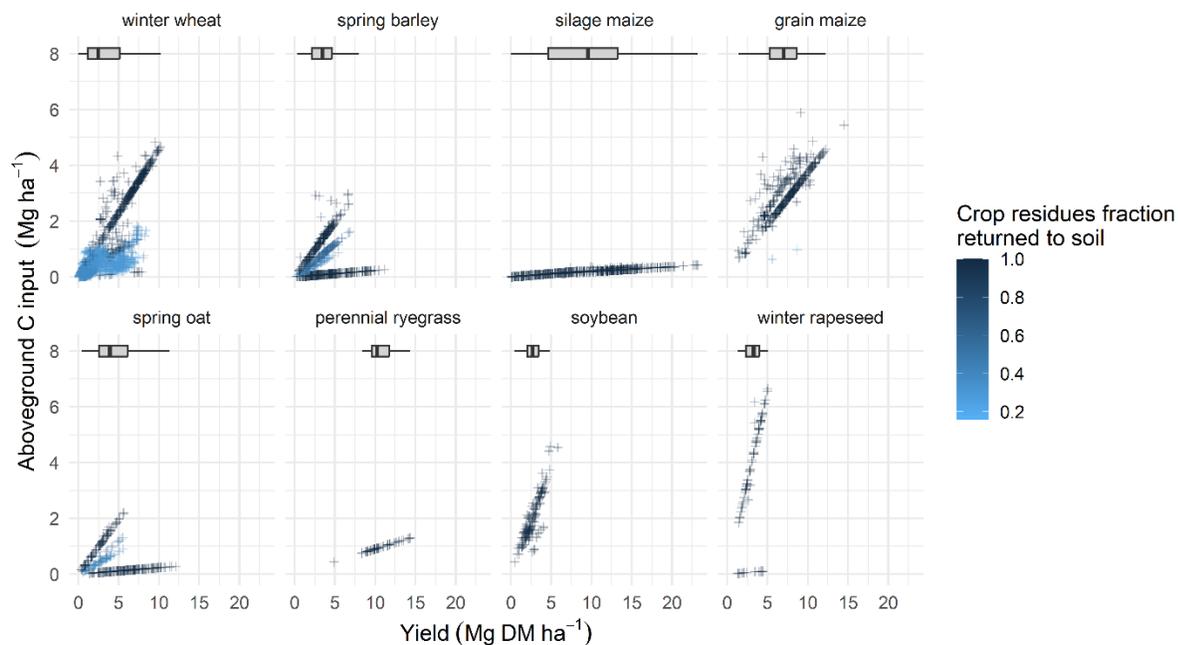


Figure 9. Distribution of yields of the major crops in the dataset, and relationship between aboveground C inputs, yields, and crop residues fraction returned to soil. Each subplot includes a simplified boxplot without outliers of yield distribution.

A limitation of the dataset is the need to estimate some required variables for SOC models, especially for C inputs. Belowground C inputs are a major source of C flux into the soil, yet they are very rarely measured in LTEs. We relied on estimates of belowground C produced by the crops scaled with the aboveground biomass (shoot:root ratio), i.e., belowground carbon inputs vary with yield. Other studies used fixed estimations of belowground C inputs, yield-independent and not scaled with aboveground biomass variations<sup>85,86</sup>. However, future users may use their own assumptions and methods to estimate C inputs based on measured yields and the HI provided in the dataset. Another limitation of the dataset is that most climatic variables were not directly measured, but came from reanalysis products, which could lead to uncertainties in soil C modeling exercises. Among the models that we identified to define the necessary input variables, only AMG incorporates soil pH as an input variable to calculate the mineralization rate of the active carbon pool<sup>47</sup>. In the present dataset, pH values were not harmonized across different soil-to-water measurement ratios. Previous findings in Polish soils indicate that such differences in ratio resulted in a difference of 0.1 pH units, which was not statistically significant<sup>87</sup>.

The present dataset does not encompass all temperate cropland systems and is heavily biased toward the Western European context. Evaluation of soil carbon models in other temperate regions such as the Americas, Asia, or Eastern Europe using this dataset should therefore be approached with caution. The inclusion of LTE data from these regions would enhance the dataset's representativeness and reduce geographical bias.

## Usage notes

The tables of the dataset can be joined by the *site\_name* variable, present in all data tables. The link between parameters tables and data tables can be done with the *crop\_name* and *crop\_type* variables for *crop\_allom.csv* and *c\_inputs\_crop.csv* tables, and with *eom\_type* for *h\_coef\_eom.csv* and *c\_inputs\_eom.csv* tables. The dataset is provided as a common dataset to run with several soil C models. We did not provide individual input files specifically formatted for each model, as several tools allow to run several models simultaneously<sup>88,89</sup>. Therefore, users who wish to reuse the dataset to evaluate one given model would need to rename and reformat some data to match the input data requirements of the model. Soil layer thickness used in SOC stock calculation varies between sites, from 0-10 cm to 0-34 cm. SOC data were not normalized for a specific soil thickness, as soil thickness normalization may not be required in multi-model simulations in the topsoil layer if model parameters are recalibrated<sup>88</sup>. Users of the dataset dealing with models that require a specific soil thickness (e.g., Century was originally parameterized for 0-20 cm topsoil layer) may need to normalize the SOC stocks and C inputs data.

Users wanting to add new datasets from other temperate LTEs can contact the corresponding author that may decide to update the data paper. In order to facilitate the reuse of these data by placing them in a broader context of interoperability with other data from environmental observations or experiments, their semantic modeling is currently underway. Using methodologies and tools developed by AnaEE France<sup>90</sup>, the variables and their acquisition contexts are modelled based on the generic OBOE ontology<sup>91</sup> extended to cover the soil carbon domain. This ongoing work will be described in a separate, forthcoming methodology paper, presenting the methodology and semantic resources used and delivering the semantized versions of the data.

## Data availability

The dataset is accessible from a ZIP archive deposited on Recherche Data Gouv<sup>59</sup>: <https://doi.org/10.57745/WKQHW2>.

## Code availability

Several R scripts are provided in the code folder of the dataset. Scripts for climate data extraction, aggregation and formatting from SAFRAN (*get\_safran\_data\_geosas\_api.R*), ERA5 & ERA5-Land (*get\_ERA5\_data\_openmeteo.R*), Rothamsted dataset (*climate\_data\_broadbalk.R*) are provided and can be

used to reproduce the climate tables, and soil moisture and temperature tables. The final tables are generated with *merge\_climate\_data.R*.

The calculation of C inputs from observed data (yields and HI) can be found in the *c\_input\_calculation.R* file. The *dataset\_check.R* script reproduces the checking of the dataset. Description of the dataset including code for the figures used in this paper is found in the *dataset\_desc.Rmd* file. The R version used was 4.5.1. The R environment used in this work was encapsulated with the *renv* package, ensuring code reproducibility within a Rstudio project. This R environment can be restored by running the *setup.R* script.

## Acknowledgements, Author Contributions & Competing Interests

### Acknowledgements

This work is part of project ALAMOD of the exploratory research program FairCarboN and received government funding managed by the Agence Nationale de la Recherche under the France 2030 program, reference ANR-22-PEXF-0002.

We thank the IMMORTAL project, funded by a joint CNRS/INRAE call for proposals « Cycle du carbone dans les écosystèmes terrestres ».

We thank Rothamsted Research for information and data from the e-RA database. The Rothamsted Long-Term Experiments - National Bioscience Research Infrastructure (RLTE-NBRI) is funded by the UK Research and Innovation - Biotechnology and Biological Sciences Research Council (UKRI-BBSRC) under award BBS/E/RH/23NB0007 (2023-2028). The RLTE-NBRI is also supported by the Lawes Agricultural Trust.

Support for this research was also provided by the USDA Long-Term Agroecosystem Research (LTAR) Program and the NSF Long-Term Ecological Research Program (DEB 2224712) at the Kellogg Biological Station, and by Michigan State University AgBioResearch.

We also thank the Swedish University of Agricultural Sciences for supporting the maintenance of the Swedish LTEs including the associated database and soil sample archive.

USDA-ARS Material Transfer Agreement #18655. USDA is an equal opportunity provider and employer. Mention of trade names or commercial products in this publication does not imply recommendation or endorsement by the U.S. Department of Agriculture.

The QualiAgro, PROspective and EFELE experiments are part of the SOERE-PRO (network of long-term experiments dedicated to the study of the impact of the recycling of organic waste products) and are integrated as a service of the “Investment in the Future” infrastructure AnaEE-France, overseen by the

French National Research Agency (ANR-11-INBS-0001). The QualiAgro experiment was founded and is still supported by INRAE and Veolia. The PROspective experiment was founded and is still supported by INRAE and SMRA68. The La Cage and the 42 plots experiments at Versailles were created and are supported by INRAE. The Lusignan experiment is part of the SOERE-ACBB (French national observatory on Agroecosystems, Biogeochemical Cycles and Biodiversity) and is integrated as a service of the “Investment in the Future” infrastructure AnaEE-France, overseen by the French National Research Agency (ANR-11-INBS-0001). It was founded and is still supported by INRAE. Rés0Pest is an experimental network of zero-pesticide farming systems for field crops and mixed farming, comprising nine sites and which aims to generate knowledge that can be used to design innovative, pesticide-efficient farming systems.

We acknowledge Hubert Boizard, Pascal Cocandea, Caroline Colnenne-David, Marie-Laure Decau, Pascal Deneroy, Christophe Desvignes, Rosemonde Devaux, Jérôme Guérif, Claire Jouany, Dominique Le Floch, Emilie Mignot, Christian Morel, Brice Mosa, Bernard Nicolardot, Bernard Nicoullaud, Daniel Plenet, Guy Richard, Véronique Tanneau, and Eric Venet, for their contribution to site management and/or data collection in the LTE of the dataset. We thank Andy Geschwendtas and Mingming Zong for support of the SOC fractionation.

#### Author contributions

Kenji Fujisaki – project conception, data processing, data validation, data analysis, original draft writing

Fabien Ferchaud – project conception, data processing, data validation, data acquisition

Hugues Clivot – project conception, data processing, data validation

Elisa Bruni – project conception, data validation

Bertrand Guenet - project conception, data validation

Christian Pichot – project conception, data validation

Antoine Versini – project conception

François Baudin – data collection, data processing

Antonio Bispo – project conception, data validation

Manuel P. Martin – project conception, data validation

Johannes L. Jensen – site management, data acquisition

Jørgen Eriksen – site management, data acquisition

Claire Chenu – site management, data acquisition

Andrew S. Gregory – site management, data acquisition

Margaret J. Glendining – data acquisition, data processing

Ines Merbach – site management, data acquisition  
Nicolas Beaudoin – site management, data acquisition  
Bruno Mary – site management, data acquisition, data processing  
Alain Mollier – site management, data acquisition  
Gilles Tison – site management, data acquisition  
Christophe Montagnier – site management, data acquisition  
Abad Chabbi – site management, data acquisition  
Françoise Vertes – site management, data acquisition  
Alice Cadéro – data processing  
Anne-Isabelle Graux – data processing  
Sylvain Pellerin – site management, data acquisition  
Florent Levavasseur – site management, data acquisition, data processing  
Manon Gilles – site management, data acquisition  
Thierry Morvan – site management, data acquisition  
Camille Resseguier – site management, data acquisition  
Luis Milesi – site management, data acquisition  
Alicia Irizar – site management, data acquisition  
Adriàn Andriulo – site management, data acquisition  
Marie-Noël Mistou – site management, data acquisition  
Arnaud Butier – site management, data acquisition  
Michel Bertrand – site management, data acquisition  
Bénédicte Autret – data acquisition, data processing  
Marie-Hélène Jeuffroy – site management, data acquisition  
Gilles Grandeau – site management, data acquisition  
Thierry Doré – site management, data acquisition  
Vincent Cellier – site management, data acquisition, data processing  
Alain Berthier – site management, data acquisition  
Sébastien Darras – site management, data acquisition  
Guillaume Audebert – site management, data acquisition  
Ludovic Pasquier – site management, data acquisition  
Fabien Ecalte – site management, data acquisition  
Antoine Savoie – site management, data acquisition  
Marcus Schiedung – data acquisition, data validation  
Christopher Poeplau – data acquisition

Nadia I. Maaroufi – site management, data acquisition

Thomas Kätterer – data acquisition, data processing

Martin A. Bolinder – data acquisition, data processing

Jonathan Sanderman – site management, data collection, data processing

Pierre Barré – project conception, data processing, data validation

### Competing Interests

None

## References

1. Le Noë, J. *et al.* Soil organic carbon models need independent time-series validation for reliable prediction. *Commun Earth Environ* **4**, 1–8 (2023) <https://doi.org/10.1038/s43247-023-00830-5>.
2. Blanchy, G. *et al.* An open-source metadataset of running European mid- and long-term agricultural field experiments. *Soil Use and Management* **40**, e12978 (2024) <https://doi.org/10.1111/sum.12978>.
3. Jandl, R. *et al.* Current status, uncertainty and future needs in soil organic carbon monitoring. *Science of the Total Environment* **468**, 376–383 (2014) <http://www.sciencedirect.com/science/article/pii/S0048969713009406>.
4. Coleman, K. & Jenkinson, D. S. RothC-26.3 - A Model for the turnover of carbon in soil. in *Evaluation of Soil Organic Matter Models* (eds Powelson, D. S., Smith, P. & Smith, J. U.) 237–246 (Springer Berlin Heidelberg, 1996). [http://link.springer.com/chapter/10.1007/978-3-642-61094-3\\_17](http://link.springer.com/chapter/10.1007/978-3-642-61094-3_17).
5. Parton, W. J., Stewart, J. W. B. & Cole, C. V. Dynamics of C, N, P and S in grassland soils: a model. *Biogeochemistry* **5**, 109–131 (1988) <https://doi.org/10.1007/BF02180320>.
6. Andriulo, A., Mary B. & Guerif, J. Modelling soil carbon dynamics with various cropping. *Agronomie* **19**, 365–377 (1999).
7. Wieder, W. R., Grandy, A. S., Kallenbach, C. M., Taylor, P. G. & Bonan, G. B. Representing life in the Earth system with soil microbial functional traits in the MIMICS model. *Geoscientific Model Development* **8**, 1789–1808 (2015) <https://doi.org/10.5194/gmd-8-1789-2015>.

8. Andr n, O. & K tterer, T. ICBM: The Introductory Carbon Balance Model for Exploration of Soil Carbon Balances. *Ecological Applications* **7**, 1226–1236 (1997) <https://doi.org/10.2307/2641210>.
9. Abramoff, R. *et al.* The Millennial model: in search of measurable pools and transformations for modeling soil carbon in the new century. *Biogeochemistry* **137**, 51–71 (2018) <https://doi.org/10.1007/s10533-017-0409-7>.
10. Giannini-Kurina, F. *et al.* Modelling and validating soil carbon dynamics at the long-term plot scale using the rCTOOL R package. *Environmental Modelling & Software* **183**, 106229 (2025) <https://doi.org/10.1016/j.envsoft.2024.106229>.
11. Graux, A.-I., Cadero, A., Ferchaud, F. & Vert s, F. Kerbernez: a long-term experiment to study the effect of different forage cropping systems on crop yields and soil organic matter in a temperate oceanic climate. Recherche Data Gouv <https://doi.org/10.57745/P8NNZK> (2024).
12. Morvan, T., Le Roy, P. & Gaillard, F. A dataset of crop yields and exports of C, N, P, K, and Mg over 9 years at the EFELE long-term field-experiment site. Recherche Data Gouv <https://doi.org/10.57745/MZ79D8> (2025).
13. Michaud, A., Montenach, D., Levvasseur, F. & Houot, S. PROspective - agronomic management dataset. Recherche Data Gouv <https://doi.org/10.57745/77AFBH> (2023).
14. Michaud, A., Montenach, D., Levvasseur, F. & Houot, S. PROspective - organic waste products physicochemical dataset. Recherche Data Gouv <https://doi.org/10.57745/6GCPX6> (2023).
15. Michaud, A., Montenach, D., Levvasseur, F. & Houot, S. PROspective - plant physicochemical dataset. Recherche Data Gouv <https://doi.org/10.57745/BPB34G> (2023).
16. Montenach, D., Michaud, A., Levvasseur, F. & Houot, S. PROspective - soil physicochemical dataset. Recherche Data Gouv <https://doi.org/10.57745/IOSBW6> (2023).
17. Colnenne-David, C. Innovative cropping systems designed to reach both environmental and production targets: data set of biotic and abiotic variables from a twelve-year French field trial. Recherche Data Gouv <https://doi.org/10.57745/5TJJZA> (2024).

18. Poulton, P., Gregory, A., Glendining, M. & Ostler, R. Highfield Bare Fallow soil chemical properties, 1959-2023. Electronic Rothamsted Archive, Rothamsted Research <https://doi.org/10.23637/rrs1-SOILCN-2> (2025).
19. Rothamsted Research. Broadbalk Wheat annual grain and straw yields 1852-1925. Electronic Rothamsted Archive, Rothamsted Research <https://doi.org/10.23637/rbk1-1796346264-1> (2021).
20. Glendining, M. & Poulton, P. Broadbalk Wheat annual grain and straw yields 1926-1967. Electronic Rothamsted Archive, Rothamsted Research <https://doi.org/10.23637/rbk1-yld2667-01> (2023).
21. Glendining, M. & Poulton, P. Broadbalk Wheat annual grain and straw yields 1968-2022. Electronic Rothamsted Archive, Rothamsted Research <https://doi.org/10.23637/rbk1-yld6822-01> (2023).
22. Rothamsted Research. Broadbalk soil organic carbon content 1843-2010. Electronic Rothamsted Archive, Rothamsted Research <https://doi.org/10.23637/KeyRefOABKsoc> (2014).
23. Glendining, M., Gregory, A., Poulton, P. & Wilmer, W. Broadbalk Wheat Experiment organic manure chemical composition. Electronic Rothamsted Archive, Rothamsted Research <https://doi.org/10.23637/rbk1-FYM-01> (2024).
24. Swedish University of Agricultural Sciences. SLU long-term field experiments: The frame trial (R3-RAM56), crop and soil data from 1956 and onwards (Version 1) [Data set]. Swedish University of Agricultural Sciences <https://doi.org/10.5878/7f5n-vp94> (2025).
25. Swedish University of Agricultural Sciences. SLU long-term field experiments: Soil organic matter in a cereal-only cropping system (R3-0020), crop and soil data from 1971 and onwards (Version 1) [Data set]. Swedish University of Agricultural Sciences <https://doi.org/10.5878/09pa-x212> (2025).
26. Jin, V. Long-term tillage and cropping system experiment for Greenhouse gas Reduction through Agricultural Carbon Enhancement network and Nutrient Use and Outcome Network in Lincoln, Nebraska. U.S. Department of Agriculture - Agricultural Research Service <https://doi.org/10.15482/USDA.ADC/1503991> (2020).
27. Sanderman, J., David, R., Moore, A., Keith, H. & Farquharson, R. Waite Permanent Rotation Trial. CSIRO <https://doi.org/10.4225/08/55E5165EC0D29> (2015).

28. Robertson, G., Haddad, N. & Snapp, S. Agronomic Yields in Row Crop Agriculture at the Kellogg Biological Station, Hickory Corners, MI (1989 to 2021). Environmental Data Initiative <https://doi.org/10.6073/PASTA/88B4C974BDC9183F784F3A91B01086C6> (2022).
29. Robertson, G. & Simmons, J. Main Cropping System Experiment Field Logs and treatment descriptions at the Kellogg Biological Station, Hickory Corners, MI (1988 to 2020). Environmental Data Initiative <https://doi.org/10.6073/PASTA/E5642426F9200404B6BA3AF45AD9C423> (2020).
30. Allen, R. G., Pereira, L. S., Raes, D., Smith, M., & others. Crop evapotranspiration-Guidelines for computing crop water requirements-FAO Irrigation and drainage paper 56. *Fao, Rome* **300**, D05109 (1998).
31. Vidal, J.-P., Martin, E., Franchistéguy, L., Baillon, M. & Soubeyroux, J.-M. A 50-year high-resolution atmospheric reanalysis over France with the Safran system. *International Journal of Climatology* **30**, 1627–1644 (2010) <https://doi.org/10.1002/joc.2003>.
32. Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* **146**, 1999–2049 (2020) <https://doi.org/10.1002/qj.3803>.
33. Muñoz-Sabater, J. *et al.* ERA5-Land: a state-of-the-art global reanalysis dataset for land applications. *Earth System Science Data* **13**, 4349–4383 (2021) <https://doi.org/10.5194/essd-13-4349-2021>.
34. Zippenfenig, P. Open-Meteo.com Weather API. Zenodo <https://doi.org/10.5281/zenodo.7973572> (2024).
35. Pisel, T. openmeteo: Retrieve Weather Data from the Open-Meteo API. (2023).
36. Zotarelli, L., Dukes, M., Romero, C., Migliaccio, K. & Morgan, K. Step by Step Calculation of the Penman-Monteith Evapotranspiration (FAO-56 Method). *EDIS* **AE459**, (2010) <https://doi.org/10.32473/edis-ae459-2010>.
37. Ellert, B. H. & Bettany, J. R. Calculation of organic matter and nutrients stored in soils under contrasting management regimes. *Canadian Journal of Soil Science* **75**, 529–538 (1995) <http://pubs.aic.ca/doi/abs/10.4141/cjss95-075>.

38. Ferchaud, F., Chlebowski, F. & Mary, B. *SimpleESM: R Script to Calculate Soil Organic Carbon and Nitrogen Stocks at Equivalent Soil Mass*. <https://hal.science/hal-04013158> (2023) <https://hal.science/hal-04013158>.
39. Disnar, J. R., Guillet, B., Keravis, D., Di-Giovanni, C. & Sebag, D. Soil organic matter (SOM) characterization by Rock-Eval pyrolysis: scope and limitations. *Organic Geochemistry* **34**, 327–343 (2003) [https://doi.org/10.1016/S0146-6380\(02\)00239-5](https://doi.org/10.1016/S0146-6380(02)00239-5).
40. Kpemoua, T. P. I. *et al.* What is the stability of additional organic carbon stored thanks to alternative cropping systems and organic waste product application? A multi-method evaluation. *SOIL* **10**, 533–549 (2024) <https://doi.org/10.5194/soil-10-533-2024>.
41. Cécillon, L. *et al.* Partitioning soil organic carbon into its centennially stable and active fractions with machine-learning models based on Rock-Eval® thermal analysis (PARTYSOCv2.0 and PARTYSOCv2.0EU). *Geoscientific Model Development* **14**, 3879–3898 (2021) <https://doi.org/10.5194/gmd-14-3879-2021>.
42. Begill, N., Don, A. & Poeplau, C. No detectable upper limit of mineral-associated organic carbon in temperate agricultural soils. *Global Change Biology* **29**, 4662–4669 (2023) <https://doi.org/10.1111/gcb.16804>.
43. Just, C. *et al.* A Simple Approach to Isolate Slow and Fast Cycling Organic Carbon Fractions in Central European Soils—Importance of Dispersion Method. *Front. Soil Sci.* **1**, (2021) <https://doi.org/10.3389/fsoil.2021.692583>.
44. Lutfalla, S. *et al.* Multidecadal persistence of organic matter in soils: multiscale investigations down to the submicron scale. *Biogeosciences* **16**, 1401–1410 (2019) <https://doi.org/10.5194/bg-16-1401-2019>.
45. Román Dobarco, M., Cousin, I., Le Bas, C. & Martin, M. P. Pedotransfer functions for predicting available water capacity in French soils, their applicability domain and associated uncertainty. *Geoderma* **336**, 81–95 (2019) <https://doi.org/10.1016/j.geoderma.2018.08.022>.

46. Chen, H. *et al.* An 18-year field experiment to assess how various types of organic waste used at European regulatory rates sustain crop yields and C, N, P, and K dynamics in a French calcareous soil. *Soil and Tillage Research* **221**, 105415 (2022) <https://doi.org/10.1016/j.still.2022.105415>.
47. Clivot, H. *et al.* Modeling soil organic carbon evolution in long-term arable experiments with AMG model. *Environmental Modelling & Software* **118**, 99–113 (2019) <https://doi.org/10.1016/j.envsoft.2019.04.004>.
48. Bolinder, M. A., Janzen, H. H., Gregorich, E. G., Angers, D. A. & VandenBygaart, A. J. An approach for estimating net primary productivity and annual carbon inputs to soil for common agricultural crops in Canada. *Agriculture, Ecosystems & Environment* **118**, 29–42 (2007) <https://doi.org/10.1016/j.agee.2006.05.013>.
49. Clivot, H. *et al.* AMGv2 parameters. Recherche Data Gouv <https://doi.org/10.57745/MEQQIX> (2023).
50. Fan, J., McConkey, B., Wang, H. & Janzen, H. Root distribution by depth for temperate agricultural crops. *Field Crops Research* **189**, 68–74 (2016) <https://doi.org/10.1016/j.fcr.2016.02.013>.
51. Gale, M. R. & Grigal, D. F. Vertical root distributions of northern tree species in relation to successional status. *Can. J. For. Res.* **17**, 829–834 (1987) <https://doi.org/10.1139/x87-131>.
52. Thiébeau, P., Jensen, L. S., Ferchaud, F. & Recous, S. Dataset of biomass and chemical quality of crop residues from European areas. *Data in Brief* **37**, 107227 (2021) <https://doi.org/10.1016/j.dib.2021.107227>.
53. Caracciolo, C. *et al.* The AGROVOC Linked Dataset. *Semantic Web* **4**, 341–348 (2013) <https://doi.org/10.3233/SW-130106>.
54. Levavasseur, F. *et al.* Carbon and nitrogen mineralization data of various exogenous organic matter (EOM) under controlled conditions. Recherche Data Gouv <https://doi.org/10.57745/JGHVYV> (2025).
55. Levavasseur, F. & Houot, S. Predicting the short- and long-term effects of recycling organic wastes in cropping systems with the PROLEG tool. *Soil Use and Management* **39**, 535–556 (2023) <https://doi.org/10.1111/sum.12856>.

56. Donmez, C. *et al.* Provision of metadata of European agricultural long-term experiments through BonaRes and EJP SOIL collaboration. *Data in Brief* **42**, 108226 (2022) <https://doi.org/10.1016/j.dib.2022.108226>.
57. Haddaway, N. R. *et al.* How does tillage intensity affect soil organic carbon? A systematic review. *Environmental Evidence* **6**, (2017) <https://doi.org/10.1186/s13750-017-0108-9>.
58. Fujisaki, K. *et al.* Semantics about soil organic carbon storage: DATA4C+, a comprehensive thesaurus and classification of management practices in agriculture and forestry. *SOIL* **9**, 89–100 (2023) <https://doi.org/10.5194/soil-9-89-2023>.
59. Fujisaki, K. *et al.* Dataset from long-term experiments in temperate croplands to evaluate soil organic carbon models. Recherche Data Gouv <https://doi.org/10.57745/WKQHW2> (2026).
60. Colomb, B., Debaeke, P., Jouany, C. & Nolot, J. M. Phosphorus management in low input stockless cropping systems: Crop and soil responses to contrasting P regimes in a 36-year experiment in southern France. *European Journal of Agronomy* **26**, 154–165 (2007) <https://doi.org/10.1016/j.eja.2006.09.004>.
61. Soyer, J. P., Lubet, E., Menet, M. & Chignon, R. Entretien de la fertilité phosphopotassique du sol en culture céréalière à rotation simple (monoculture maïs ou rotation blé/maïs). *C.R. Acad. Agric. France* 843–857 (1976).
62. Morel, C., Plénet, D. & Mollier, A. Calibration of maize phosphorus status by plant-available soil P assessed by common and process-based approaches. Is it soil-specific or not? *European Journal of Agronomy* **122**, 126174 (2021) <https://doi.org/10.1016/j.eja.2020.126174>.
63. Graux, A.-I., Cadéro, A., Ferchaud, F. & Vertès, F. The Kerbernez long-term experiment: A dataset on crop yield and soil organic matter evolution in forage crop rotations and permanent grasslands in a temperate oceanic climate. *Data in Brief* **58**, 111259 (2025) <https://doi.org/10.1016/j.dib.2024.111259>.
64. Morel, C., Jouany, C., Denoroy, P. & Montagnier, C. Functional and mechanistic assessment of plant-available soil phosphorus greatly improved the multisite diagnosis of maize yield response. *Field Crops Research* **317**, 109539 (2024) <https://doi.org/10.1016/j.fcr.2024.109539>.

65. Levavasseur, F. *et al.* The simple AMG model accurately simulates organic carbon storage in soils after repeated application of exogenous organic matter. *Nutr Cycl Agroecosyst* **117**, 215–229 (2020) <https://doi.org/10.1007/s10705-020-10065-x>.
66. Morvan, T. *et al.* A comprehensive dataset on nitrate, Nitrite and dissolved organic carbon leaching losses from a 4-year Lysimeter study. *Data in Brief* **32**, 106029 (2020) <https://doi.org/10.1016/j.dib.2020.106029>.
67. Lamichhane, J. R., Boizard, H., Dürr, C., Richard, G. & Boiffin, J. Effect of cropping systems and climate on soil physical characteristics, field crop emergence and yield: A dataset from a 19-year field experiment. *Data in Brief* **39**, 107581 (2021) <https://doi.org/10.1016/j.dib.2021.107581>.
68. Autret, B. *et al.* Alternative arable cropping systems: A key to increase soil organic carbon storage? Results from a 16 year field experiment. *Agriculture, Ecosystems & Environment* **232**, 150–164 (2016) <https://doi.org/10.1016/j.agee.2016.07.008>.
69. Colnenne-David, C., Jeuffroy, M.-H., Grandeau, G., Ferchaud, F. & Doré, T. Innovative cropping systems designed to reach both environmental and production targets: Data set of biotic and abiotic variables from a twelve-year French field trial. *Data in Brief* **54**, 110398 (2024) <https://doi.org/10.1016/j.dib.2024.110398>.
70. Saffih-Hdadi, K. & Mary, B. Modeling consequences of straw residues export on soil organic carbon. *Soil Biology and Biochemistry* **40**, 594–607 (2008) <https://doi.org/10.1016/j.soilbio.2007.08.022>.
71. Ferchaud, F., Vitte, G. & Mary, B. Changes in soil carbon stocks under perennial and annual bioenergy crops. *GCB Bioenergy* **8**, 290–306 (2016) <https://doi.org/10.1111/gcbb.12249>.
72. Cellier, V. *et al.* Analysis of the socio-economic and environmental sustainability of a network of zero-pesticide cropping systems (RésoPest) after 10 years of experimentation. *Innovations Agronomiques* **98**, 142 (2025) <https://doi.org/10.17180/ciag-2025-Vol98-art11-GB>.
73. Barré, P. *et al.* Quantifying and isolating stable soil organic carbon using long-term bare fallow experiments. *Biogeosciences* **7**, 3839–3850 (2010) <https://doi.org/10.5194/bg-7-3839-2010>.

74. Schlüter, S. *et al.* Divergence in physical, chemical, and biological soil properties caused by different long-term bare fallow management and natural succession. *Geoderma* **459**, 117361 (2025) <https://doi.org/10.1016/j.geoderma.2025.117361>.
75. Yin, X. *et al.* Long-term modelling of soil N mineralization and N fate using STICS in a 34-year crop rotation experiment. *Geoderma* **357**, 113956 (2020) <https://doi.org/10.1016/j.geoderma.2019.113956>.
76. Pold, G., MacDonald, E., Braun, S. & Herrmann, A. M. Soil and vegetation property data from the Ultuna R3-RAM56 long-term soil amendment experiment, 1956–2023. *Data in Brief* **59**, 111350 (2025) <https://doi.org/10.1016/j.dib.2025.111350>.
77. Börjesson, G., Bolinder, M. A., Kirchmann, H. & Kätterer, T. Organic carbon stocks in topsoil and subsoil in long-term ley and cereal monoculture rotations. *Biol Fertil Soils* **54**, 549–558 (2018) <https://doi.org/10.1007/s00374-018-1281-x>.
78. Poulton, P. R. *et al.* The Broadbalk Wheat Experiment, Rothamsted, UK: Crop yields and soil changes during the last 50 years. *Advances in Agronomy* 173–298 (2024) <https://doi.org/10.1016/bs.agron.2023.11.003>.
79. Varvel, G. E. & Wilhelm, W. W. No-tillage increases soil profile carbon and nitrogen under long-term rainfed cropping systems. *Soil and Tillage Research* **114**, 28–36 (2011) <https://doi.org/10.1016/j.still.2011.03.005>.
80. Grace, P., Oades, J., Keith, H. & Hancock, T. Trends in wheat yields and soil organic carbon in the Permanent Rotation Trial at the Waite Agricultural Research Institute, South Australia. *Aust. J. Exp. Agric.* **35**, 857 (1995) <https://doi.org/10.1071/EA9950857>.
81. Syswerda, S. P., Corbin, A. T., Mokma, D. L., Kravchenko, A. N. & Robertson, G. P. Agricultural Management and Soil Carbon Storage in Surface vs. Deep Layers. *Soil Science Society of America Journal* **75**, 92 (2011) <https://doi.org/10.2136/sssaj2009.0414>.
82. Martin, M. P. *et al.* Feasibility of the 4 per 1000 aspirational target for soil carbon: A case study for France. *Global Change Biology* **27**, 2458–2477 (2021) <https://doi.org/10.1111/gcb.15547>.

83. Riggers, C., Poeplau, C., Don, A., Frühauf, C. & Dechow, R. How much carbon input is required to preserve or increase projected soil organic carbon stocks in German croplands under climate change? *Plant Soil* **460**, 417–433 (2021) <https://doi.org/10.1007/s11104-020-04806-8>.
84. Kätterer, T., Bolinder, M. A., Andrén, O., Kirchmann, H. & Menichetti, L. 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**, 184–192 (2011) <https://doi.org/10.1016/j.agee.2011.02.029>.
85. Taghizadeh-Toosi, A. *et al.* Visiting dark sides of model simulation of carbon stocks in European temperate agricultural soils: allometric function and model initialization. *Plant Soil* **450**, 255–272 (2020) <https://doi.org/10.1007/s11104-020-04500-9>.
86. Hirte, J., Leifeld, J., Abiven, S., Oberholzer, H.-R. & Mayer, J. Below ground carbon inputs to soil via root biomass and rhizodeposition of field-grown maize and wheat at harvest are independent of net primary productivity. *Agriculture, Ecosystems & Environment* **265**, 556–566 (2018) <https://doi.org/10.1016/j.agee.2018.07.010>.
87. Kabała, C., Muszyfaga, E., Gałka, B., Łabuńska, D. & Mańczyńska, P. Conversion of Soil pH 1:2.5 KCl and 1:2.5 H<sub>2</sub>O to 1:5 H<sub>2</sub>O: Conclusions for Soil Management, Environmental Monitoring, and International Soil Databases. *Pol. J. Environ. Stud.* **25**, 647–653 (2016) <https://doi.org/10.15244/pjoes/61549>.
88. Bruni, E. *et al.* Multi-modelling predictions show high uncertainty of required carbon input changes to reach a 4‰ target. *European Journal of Soil Science* **73**, e13330 (2022) <https://doi.org/10.1111/ejss.13330>.
89. Sierra, C. A., Müller, M. & Trumbore, S. E. Models of soil organic matter decomposition: the SoilR package, version 1.0. *Geosci. Model Dev.* **5**, 1045–1060 (2012) <https://doi.org/10.5194/gmd-5-1045-2012>.

90. Pichot, C. *et al.* Semantic Management of Data from Biodiversity and Ecosystem Studies: Toward an Integrated Workflow from Collection to Publication. Application to Plankton Data from Lake Geneva. in *JOWO* (2021). <https://ceur-ws.org/Vol-2969/paper11-s4biodiv.pdf>.
91. Madin, J. *et al.* An ontology for describing and synthesizing ecological observation data. *Ecological Informatics* **2**, 279–296 (2007) <https://doi.org/10.1016/j.ecoinf.2007.05.004>.

ARTICLE IN PRESS