Supplemental information for: Soils' dirty little secret: Depth-based comparisons can be inadequate for quantifying changes in soil organic carbon and other mineral soil properties

Authors: Adam C. von Haden[1,2], Wendy H. Yang[1,2,3,4], Evan H. DeLucia[1,2,4]

Affiliations: [1]Center for Advanced Bioenergy and Bioproducts Innovation, [2]Institute for Sustainability, Energy, and Environment, [3]Department of Geology, [4]Department of Plant Biology

University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States

Corresponding Author:

Evan H. DeLucia
Center for Advanced Bioenergy and Bioproducts Innovation
University of Illinois at Urbana-Champaign
1206 W Gregory Drive
Urbana, Illinois 61801, United States
Phone: +1 (217) 333-6177
Email: delucia@illinois.edu

Contents:

Supplemental methods
Supplemental script
Supplemental references
Supplemental figures (4)

Supplemental methods

Site description and field sampling

The field site was located at the University of Illinois Energy Farm (40.064° N, 88.195° W) in Urbana, IL, USA. Site 30-year mean annual temperature and precipitation are 10.9 °C and 1051 mm, respectively (National Climatic Data Center, 2019). Soils at the site are silt loam and silty clay loam Mollisols including Dana silt loam (Fine-silty, mixed, superactive, mesic Oxyaquic Argiudolls), Flanagan silt loam (Fine, smectitic, mesic Aquic Argiudolls), and Drummer silty clay loam (Fine-silty, mixed, superactive, mesic Typic Endoaquolls) (Soil Survey Staff, 2019). In 2008, maize-soybean (*Zea mays* L.-*Glycine max* (L.) Merr.) rotation, switchgrass (*Panicum virgatum* L.), giant miscanthus (*Miscanthus* × *giganteus* J.M.Greef , Deuter ex Hodk., Renvoize), and native prairie treatments were initiated in five replicated blocks (Anderson-Teixeira et al., 2013). Four blocks were comprised of 0.7 ha plots, and the fifth block consisted of 3.8 ha plots.

Soil cores were collected using a Giddings hydraulic soil probe (Giddings Machine Company Inc., Windsor, CO, USA) in 2015. Ten cores were extracted from random locations in each of the 0.7 ha plots, and 40 cores were extracted from each of the 3.8 ha plots. The cores were taken to a nominal depth of 150 cm and were segmented into 0 to 10 cm, 10 to 30 cm, 30 to 50 cm, and 50 to 100 cm. After soils were air-dried, the coarse root fragments were removed, and soils were crushed using a Dynacrush DC-5 (Custom Laboratory Equipment Inc., Holden, MO USA). Soils were then sieved to 2 mm and oven-dried soil weight was calculated. When necessary, inorganic C was removed via acid fumigation. Soil organic carbon (SOC), total soil N, and $\delta^{13}C$ were determined using a Costech 4010 CHNSO Elemental Analyzer (Costech Analytical Technologies, Valencia, CA USA) in line with a Thermo Scientific DELTA V Advantage Isotope Ratio Mass Spectrometer (Thermo Fisher Scientific, Waltham, MA USA). $\delta^{13}C$ was reported with respect to VPDB. Fifty-four cores (17% of the total) that did not contain contiguous segments to 100 cm were removed from subsequent analysis.

Bulk density change simulations

To assess the potential effect of bulk density changes on fixed depth (FD)- versus equivalent soil mass (ESM)-based soil properties, we applied a simulation model to the soil samples collected from the University of Illinois Energy Farm. The model simulated compaction and expansion of soils where the changes were concentrated near the soil surface and diminished exponentially throughout the profile, as would be expected from surface disturbances such as tillage or heavy machinery traffic. This approach allows for bulk density to change while holding other factors constant. The following equation was applied to each contiguous soil core

$$D_{sim} = D_{actual} - D_{alter}[1 - \exp(-0.161\, D_{actual})] \quad \text{(S1)}$$

where D_{sim} is the simulated lower segment depth, D_{actual} is the lower depth of each segment within the soil core, and D_{alter} is the total depth of compaction or expansion within the core. D_{alter} ranged from -2.5 cm to 2.5 cm by 0.5 cm increments, which thus simulated a range of bulk density changes from 2.5 cm of expansion to 2.5 cm compaction and included zero bulk density change. For example, a D_{alter} value of 2.5 cm applied to a D_{actual} value of 10 cm results in D_{sim} of 8.0 cm, indicating that 2.0 cm (10 cm - 8.0 cm) of the 2.5 cm of total compaction occurred in the top 10 cm. Thus, the top 8 cm of the simulated compacted soil contains the equivalent cumulative soil masses and stocks of the top 10 cm of the uncompacted soil. Cubic spline functions were then used to fit the relationships between D_{sim} and cumulative soil masses and stocks within each core, and predictions from the spline were made at D_{actual}. The resulting soil bulk densities throughout the soil profile for 2.5 cm expansion, 2.5 cm compaction, and no bulk density change are shown in Fig. S2.

For elemental stocks, elemental mass percentages, and elemental ratios, which are reported on a ratio scale, absolute percentage error was calculated for each core segment under each scenario

$$absolute\ percentage\ error\ (\%) = \left|\frac{(V_{method} - V_{known})}{V_{known}}\right| 100 \qquad \text{(S2)}$$

where V_{method} is the value produced by the method (FD or ESM) under the simulated bulk density scenario, and V_{known} is the FD-based value with no change in bulk density. Absolute percentage error, rather than percentage error, was reported because percentage errors could be positive or negative, and thus the average of many errors may underestimate the true magnitude of error; calculating the absolute value of each error mitigates this issue.

For $\delta^{13}C$, which is reported on an interval scale, absolute error was calculated as

$$absolute\ error\ (‰) = |\delta_{method} - \delta_{known}| \qquad \text{(S3)}$$

where δ_{method} is the $\delta^{13}C$ resulting from the FD or ESM method in the simulation, and δ_{known} is the FD-based $\delta^{13}C$ value under no bulk density change.

Soil organic matter change simulations

We used two simulations to evaluate the consequences of using total soil mass (i.e., mineral soil plus SOM) rather than mineral soil mass for ESM-based calculations when changes in SOM occur. The baseline soil properties (i.e., bulk densities and SOM mass percentages) for both simulations were the same as “time 0” shown in main text Fig. 3, with SOM calculated as SOC/0.58 (Eq. A7).

The first simulation assessed the impact of varying quantities of SOM removal or addition on ESM-based calculations. The following equation was applied to the baseline soil properties

$$\Delta M_{SOM} = (SOM_{alter} - 1)\ M_{SOM,10} \qquad \text{(S4)}$$

where ΔM_{SOM} is the simulated change in the surface SOM stock (g SOM cm^{-2}), $M_{SOM,10}$ is the baseline SOM stock in the 0 to 10 cm depth interval (g SOM cm^{-2}), and SOM_{alter} is an SOM alteration factor that ranged from 0.5 to 1.5 by increments of 0.1. The resulting ΔM_{SOM} values thus spanned from a 50% loss to a 50% gain of the SOM stock that was originally contained within the 0 to 10 cm increment. The change in SOC was then applied to each depth interval

$$M_{SOM,sim} = M_{SOM} + \Delta M_{SOM} \quad \text{(S5)}$$

where $M_{SOM,sim}$ is the simulated cumulative SOM stock (g SOM cm^{-2}) following the SOM addition, and M_{SOM} is the baseline cumulative SOM stock (g SOM cm^{-2}). Thus, the simulated changes in SOM occurred only in the upper soil increment, but the changes are reflected in the cumulative SOM stocks throughout the soil profile. Soil mass in each interval was

$$M_{soil,sim} = M_{SOM,sim} + M_{mineral} \quad \text{(S6)}$$

where $M_{soil,sim}$ is the simulated cumulative soil mass (g soil cm^{-2}), and $M_{mineral}$ is the cumulative mineral soil mass (g mineral soil cm^{-2}). Note that while $M_{soil,sim}$ was affected by the SOM change (i.e., $M_{soil,sim}$ contains $M_{SOM,sim}$), $M_{mineral}$ was unaltered by the simulation, and thus mineral soil mass was conserved within the simulated soil profile. To simulate sampling to identical depths as the baseline scenario with no change in bulk density within each depth interval, a cubic spline was fit between $M_{SOM,sim}$ and $M_{soil,sim}$, and $M_{SOM,sim}$ values were predicted at the baseline M_{soil} values. $M_{mineral}$ was the difference between $M_{soil,sim}$ and $M_{SOM,sim}$. Next, a cubic spline model was fit

$$M_{SOM,sim} = S(M_{soil,sim}) \quad \text{(S7)}$$

The spline was then used to predict $M_{SOM,sim}$ at the baseline cumulative M_{soil} values, thus giving the total soil-based ESM predictions, $M_{SOM,sim,soil}$. A second spline model was fit

$$M_{SOM,sim} = S(M_{mineral}) \quad \text{(S8)}$$

The spline model was then used to predict $M_{SOM,sim}$ at the baseline $M_{mineral}$ values, thereby giving the mineral soil-based ESM predictions, $M_{SOM,sim,mineral}$. The errors for each ESM approach were then calculated as

$$error_{ESM,soil} = (M_{SOM,sim,soil} - M_{SOM}) - \Delta M_{SOM} \quad \text{(S9)}$$

$$error_{ESM,mineral} = (M_{SOM,sim,mineral} - M_{SOM}) - \Delta M_{SOM} \quad \text{(S10)}$$

The second simulation assessed how the initial SOM stock affected ESM-based calculations when additional SOM was added to the soil profile. The quantity of SOM added between timepoints was standardized among all scenarios

$$\Delta M_{SOM,added} = (0.25)\, M_{SOM,10} \quad \text{(S11)}$$

where $\Delta M_{SOM,added}$ is the quantity of SOM (g SOM cm^{-2}) added between sampling timepoints (initial and simulated), and $M_{SOM,10}$ is the baseline SOM (g SOM cm^{-2}). Thus, the amount of SOM added to the initial SOM stock is equivalent to 25% of the baseline SOM stock. The difference between baseline and initial SOM stocks were calculated as

$$\Delta M_{SOM,initial} = (SOM_{alter} - 1)\, M_{SOM,10} \quad \text{(S12)}$$

where $\Delta M_{SOM,initial}$ is the SOM difference (g SOM cm^{-2}), $M_{SOM,10}$ is the baseline SOM stock in the 0 to 10 cm depth interval (g SOM cm^{-2}), and SOM_{alter} is an SOM alteration factor that ranged from 0.5 to 1.5 by increments of 0.1. The resulting $\Delta M_{SOM,initial}$ values therefore spanned from a

50% decrease to a 50% increase of the SOM stock contained within the baseline 0 to 10 cm increment. Cumulative initial SOM stocks at each interval was then

$$M_{SOM,initial} = M_{SOM} + \Delta M_{SOM,initial} \tag{S13}$$

where $M_{SOM,initial}$ is the initial cumulative SOM stock (g SOM cm^{-2}) and M_{SOM} is the baseline cumulative SOM stock. $M_{soil,initial}$ was calculated as the sum of $M_{SOM,initial}$ and $M_{mineral}$, thus conserving the $M_{mineral}$ mass. The simulated addition of SOM was then

$$M_{SOM,sim} = M_{SOM,initial} + \Delta M_{SOM,added} \tag{S14}$$

where $M_{SOM,sim}$ is the simulated cumulative SOM stock following SOM addition. Cumulative soil mass, $M_{soil,sim}$, was calculated as the sum of $M_{SOM,sim}$ and $M_{mineral}$. To reproduce sampling to the same depths as the baseline scenario without bulk density changes, a cubic spline was fit between $M_{SOM,sim}$ and $M_{soil,sim}$, and new $M_{SOM,sim}$ values were predicted at the baseline M_{soil} values. $M_{mineral}$ was the difference between $M_{soil,sim}$ and $M_{SOM,sim}$. Cubic spline models to predict $M_{SOM,sim}$ were then fit using either $M_{soil,sim}$ (Eq. S7) or $M_{mineral}$ (Eq. S8), and predictions were made at $M_{soil,sim}$ ($M_{SOM,sim,soil}$) or $M_{mineral}$ ($M_{SOM,sim,mineral}$), respectively. Errors were calculated as

$$error_{ESM,soil} = \left(M_{SOM,sim,soil} - M_{SOM,initial}\right) - \Delta M_{SOM,added} \tag{S15}$$

$$error_{ESM,mineral} = \left(M_{SOM,sim,mineral} - M_{SOM,initial}\right) - \Delta M_{SOM,added} \tag{S16}$$

Supplemental script

Background

This script is intended to provide a starting point for users to implement the equivalent soil mass (ESM) approach for calculating soil organic carbon (SOC) stocks and SOC mass percentages (commonly called “SOC concentration”). The user should read this guide and understand how the script works before using it. The script contains some basic error checking algorithms, but it cannot catch all data entry mistakes. It is possible to misuse the script and produce invalid results. The script was designed to work on individual soil cores that have been divided into three or more contiguous segments, as the cubic spline function contains four coefficients. The script has not been tested on soil cores containing fewer than three contiguous segments. If only one segment is present, interpolation will necessarily be linear rather than cubic, which is expected to result in higher predictive error (Wendt & Hauser, 2013). Considering the wide range of potential applications, the script should not be expected to function correctly in all situations. The user is encouraged to modify or rewrite the script to meet their specific needs.

Requirements

The R packages ‘openxlsx’ (Schauberger & Walker, 2019), ‘dplyr’ (Wickham et al., 2019), and ‘tidyr’ (Wickham & Henry, 2019) are required to be installed before running the script. The user is required to provide an Excel spreadsheet containing information about their samples. In the

“User input” section of the script, the “input_file_name” variable should be set to contain the XLSX spreadsheet path and filename, and the “input_file_sheet” variable should contain the sheet in the Excel spreadsheet in which the information is stored. The user also supplies the desired name of the output XLSX file as the “output_filename” variable. The input spreadsheet data table must contain the following columns:

ID: A name to identify a unique collection of samples. For example, this might include a combination of year, treatment, plot, and station. Additional examples are given in Fig. S4.

Rep: A name that can be used if multiple collections of samples (unique instances of ID + Rep) will be compared against a single Ref_ID. If this field is not needed, it can be set to a common value (e.g., 1). Examples are given in Fig. S4.

Ref_ID: A name to identify the sample or collection of samples from which the reference masses will be calculated. Each unique combination of “ID” and “Rep” should contain the same “Ref_ID.” Examples are given in Fig. S4.

Upper_cm: The upper depth (i.e., closest to the soil surface) for the sample (cm). This must be a number. Each collection of samples (i.e., ID + Rep) must contain a sample with 0 cm as the upper depth.

Lower_cm: The lower depth (i.e., furthest from the soil surface) for the sample. This must be a number. The lower depth must be equal to the upper depth of the adjacent sample within any collection of samples (i.e., ID + Rep). For example, if the lower depth is 10 cm for the first segment in a collection of samples, then the upper depth for the next segment must be 10 cm. Indeed, the primary purpose of the “Upper_cm” and “Lower_cm” is to provide a way to verify that collections of samples contain contiguous, adjacent depth increments, as cumulative values cannot be properly calculated from non-contiguous samples. The “Upper_cm” and “Lower_cm” values are also used to convert from apparent soil bulk density values to soil mass per unit area.

SOC_pct: The mass percent of soil organic carbon (SOC) in the soil sample (g SOC 100 g soil^{-1}). This is the format commonly given by elemental analyzers.

SOM_pct: The mass percent of soil organic matter (SOM) in the sample (g SOM 100 g soil^{-1}). If SOM was not measured (e.g., by loss on ignition), then the user may estimate SOM mass percent from SOC mass percent using a conversion factor (e.g., McBratney & Minasny, 2010). The SOM mass percent is needed to calculate mineral soil mass, and therefore a value is required to be supplied by the user.

BD_g_cm3: The apparent soil bulk density (g dry soil cm^{-3}). This is calculated based on the weight of dry fine earth (i.e., < 2 mm), which hence removes rocks and other coarse material from the mass. The apparent soil bulk density is calculated from the same sample in which SOC_pct and SOM_pct are measured.

In the “User input” section of the script, the user must also provide a value for “min_core_length_cm.” This determines the minimum length (cm) of a collection of samples (e.g., cores) that will be kept in the dataset. For example, if the target sampling depth was 100 cm, then the user may want to remove collections of samples that do not have a sample reaching 100 cm. This could be achieved by setting the “min_core_length_cm” to 100. If the user does not want to remove any cores, this value can be set to 0.

The user must also decide whether extrapolation is to occur outside of the measured range of soil mass values. For example, if the maximum reference mass is greater than the maximum cumulative mass of a particular collection of samples, then extrapolation must occur to calculate an ESM value at the maximum reference mass. If the reference mass is only slightly greater than the sample mass, then errors may be minimal. However, if the reference mass is much greater than the sample mass, then errors could be significant. To avoid the need for extrapolation, it is advisable to sample to one increment deeper than intended comparisons. For example, if comparisons are desired at a 100 cm reference soil mass, then taking an additional sample from 100 to 110 cm should give enough mass to provide another datapoint for interpolation rather than extrapolation.

Finally, the user must provide a vector for “ESM_depths_cm” that contains the depths (cm) in the reference samples at which the equivalent soil masses are calculated. All of the depths provided in the “ESM_depths_cm” must be present in the collection reference samples. The purpose of this input variable is to eliminate problems that can occur when collections of reference samples contain segments that were not cut to the intended length. For example, if the reference samples were intended to be cut at 10, 30, 50, and 100 cm, but an extra cut was accidentally made at 75 cm, the user probably does not want the cumulative mass of soil at 75 cm to act as one of the reference values. In this case, the user would input 10, 30, 50, and 100 into the “ESM_depths_cm” variable, and therefore the reference masses would only be calculated at 10, 30, 50, and 100 cm.

ESM procedure

After the spreadsheet is imported, some basic error checking is performed. Checks are made to ensure that all required columns are present and that each sample has proper identification and contains upper and lower depth information. The script also verifies that all values in the spreadsheet are numeric except for identification variables. If any of the checks fail, an error message is given, and the script stops. These checks will not catch all errors, so it is possible for errors and unforeseen issues to occur downstream.

Moving forward, it is critical to note that the cubic spline fitting procedure is performed on each group of samples that contain a unique “ID,” “Rep,” and “Ref_ID” (i.e., a group of samples). The code was designed for and tested with datasets where each instance of “ID,” “Rep,” and “Ref_ID” corresponded to one individual soil core that was extracted and segmented. With this approach, one cubic spline model is fit to each soil core separately, and thus a model will be fitted through each measured data point. While it may be possible to assign “ID,” “Rep,” and “Ref_ID” in a different way (e.g., aggregations or averages of cores), it is not recommended.

In the next steps, the script removes any unnecessary columns and filters out any groups of samples (i.e., “ID” + “Rep”) that do not have an “Upper_cm” value equal to 0. Then, any samples with missing values (NAs) are removed. Any segments within a group of samples that are taken from below a contiguity break are removed. For example, if a core contains 0 to 10 cm, 10 to 30 cm, 35 to 50 cm, and 50 to 100 cm segments, the 35 to 50 cm and 50 to 100 cm segments would be removed because of the continuity break between 30 and 35 cm. Next, any cores that do not meet the “min_core_length_cm” requirement are removed. Finally, a check is performed to determine whether any observations are left in the data set; if no observations remain, an error is thrown.

In the following step, the script inserts “zero” sample for each group of samples (i.e. each instance of “ID” and “Rep”). That is, a sample with Upper_cm, Lower_cm, SOC_pct, SOM_pct, and BD_g_cm3 all equal to zero is placed into each group of samples. The purpose of this is to improve the cubic spline fitting procedure by providing an extra data point to represent the soil surface (i.e., 0 cm) where there is zero cumulative soil mass, zero cumulative SOC, and zero cumulative SOM. Ideally, the first segment of each group of samples would be short (e.g., 0 to 5 cm) to better constrain the model in the area close to the soil surface.

Next, calculations of areal soil mass, areal SOC mass, areal SOM mass, and areal mineral soil mass are performed for each soil sample. These values are in typical fixed depth (FD) form, and they are saved for future reference. The values are then converted to cumulative form for each group of samples (i.e., each instance of “ID” and “Rep”).

A loop cycles through each unique instance of “ID,” “Rep,” and “Ref_ID.” A group of reference samples is subset based on the “Ref_ID.” In case the group of reference samples contains more than one unique contiguous sample (i.e., more than one core), the cumulative mineral soil weights from the group of reference samples is averaged by sampling depth intervals. The sampling depth intervals must be contained in the previously input ESM_depths_cm variable.

If extrapolation is set to FALSE, then any samples with cumulative mass greater than the maximum cumulative mass in the reference group of samples are removed prior to spline model fitting. If extrapolation is set to TRUE, then only samples that contain a deeper increment than the reference sample set are removed.

A cubic spline model is fit between cumulative mineral soil mass and cumulative SOC mass for each group of samples. A cubic spline is also fit between cumulative mineral soil mass and cumulative SOM mass. Predictions for cumulative SOC mass and cumulative SOM mass are made at the mineral soil masses provided in the reference group of samples. The cumulative mineral soil mass, cumulative SOC mass, and cumulative SOM mass are converted to non-cumulative form. Total soil mass, soil bulk density, SOC mass percent, and SOM mass percent are then calculated based on the mineral soil mass, SOC mass, and SOM mass within each segment. These ESM data are saved, and the loop continues to the next group of samples.

After the ESM calculation has been performed for each group of samples, the “zero” samples are removed. Columns are re-ordered, and the FD and ESM datasets are output to a spreadsheet.

FD and ESM output spreadsheets

The output spreadsheet contains two sheets: one for FD and one for ESM. Both sheets contain the same column headers. The first 8 columns contain the same headers as the input spreadsheet: ID, Rep, Ref_ID, Upper_cm, Lower_cm, SOC_pct, SOM_pct, and BD_g_cm3. The Ref_ID contains blank values in the FD sheet because reference samples are not used for FD calculations. All soil properties values given in the ESM sheet are ESM-based.

The “Type” column indicates whether the observation is for FD or ESM. The remaining columns are as follows:

Soil_g_cm2: Total sample soil mass (g soil cm^{-2}).

SOC_g_cm2: SOC mass within the sample (g SOC cm^{-2}).

SOM_g_cm2: SOM mass within the sample (g SOM cm^{-2}).

Min_Soil_g_cm2: Mineral soil mass within the sample (g mineral soil cm^{-2}).

Cum_Soil_g_cm2: The cumulative soil within each group of samples (g soil cm^{-2}).

Cum_SOC_g_cm2: The cumulative SOC mass within each group of samples (g SOC cm^{-2}).

Cum_SOM_g_cm2: The cumulative SOM mass within each group of samples (g SOM cm^{-2}).

Cum_Min_Soil_g_cm2: The cumulative mineral soil mass within each group of samples (g mineral soil cm^{-2}).

It is important to note that even though the ESM sheet contain depth interval information (i.e., Upper_cm and Lower_cm), the ESM soil properties given in the sheet are based on “ESM depths” rather than fixed depths. That is, the values reported in the ESM sheet do not necessarily pertain to those that were measured within the sampled depth interval, but instead they are the estimated values that would have been obtained if the soil samples were taken to a common mineral soil mass (i.e., the reference mass) rather than a common depth. This is a very important distinction to make when interpreting ESM-based values.

Example datasets

Example datasets are provided within a supplemental XLSX file. The file contains four spreadsheets that correspond to the four examples shown in Fig. S4. Each spreadsheet contains six hypothetical soil cores with four depth increments each. The first three soil cores in each sheet correspond to Y0 (Fig. S4a, b, and d) or T1 (Fig. S4c). The last three soil cores in each sheet correspond to Y5 (Fig. S4a, b, and d) or T2 (Fig. S4c). The cores from Y0 or T1 are congruent to the cores from Y5 or T2, respectively, except that increased bulk density was simulated in the later cores by applying 1.5 cm of soil surface compaction with Eq. S1. Thus, when using the FD approach, the SOC stocks and SOC mass percentages appear different between Y0 and Y5 or T1 and T2. However, when using ESM, the SOC stocks and SOC mass percentages are nearly equivalent between Y0 and Y5 or T1 and T2. These datasets can be used to test the script and may provide a template for user data.

ESM Script

```r
# ----
# Supplemental R script for:
# Soils' dirty little secret: Depth-based comparisons can be inadequate
# for quantifying changes in soil organic carbon
# and other mineral soil properties.
# By:
# Adam C. von Haden, Wendy H. Yang, Evan H. DeLucia

# Purpose: To calculate SOC stocks and SOC mass percentages
# using the equivalent soil mass approach.

# Author: Adam C. von Haden

# Date: 2020-03-23

# Notes:
# Example_datasets.xlsx is available as a supplemental file
# See Supplemental Information for full documentation

# Load required libraries ----
# These libraries may first need to be installed by the user
library(openxlsx)
library(dplyr)
library(tidyr)

# Get input from user ----
# Filename of the input XLSX spreadsheet
input_file_name <- "Example_datasets.xlsx"

# Name of the sheet on the spreadsheet that contains the data
input_file_sheet <- "a_temporal_paired"

# The minimum core length acceptable to use for ESM
min_core_length_cm <- 0

# Determines if extrapolation will be allowed outside of sample mass
# Options are TRUE and FALSE
extrapolation <- TRUE

# Sets the (lower) depths at which reference masses are calculated
# Depths must be present in references
ESM_depths_cm <- c(10, 30, 50, 100)

# Desired filename of the ouput XLSX spreadsheet
output_filename <- "FD_ESM_output.xlsx"

# Import XLSX file ----
raw_FD <- read.xlsx(input_file_name, sheet=input_file_sheet)
```

```
# Check input for basic errors ----
is_error <- FALSE
required_colnames <- c("ID", "Ref_ID", "Rep", "Upper_cm", "Lower_cm",
                       "SOC_pct", "SOM_pct", "BD_g_cm3")

if (!all(required_colnames %in% colnames(raw_FD))){
  is_error <- TRUE
  stop(gettextf("Missing or misspelled column name(s)"))
}

if (any(is.na(raw_FD$ID) | is.na(raw_FD$Ref_ID) | is.na(raw_FD$Rep))){
  is_error <- TRUE
  stop(gettextf("Missing ID, Rep, or Ref_ID values"))
}

if (any(is.na(raw_FD$Upper_cm) | is.na(raw_FD$Lower_cm))  {
  is_error <- TRUE
  stop(gettextf("Missing Upper_cm or Lower_cm values"))
}

if (any(!is.numeric(raw_FD$Upper_cm) | !is.numeric(raw_FD$Lower_cm))  {
  is_error <- TRUE
  stop(gettextf("Non-numeric Upper_cm or Lower_cm values"))
}

if (any(!is.numeric(raw_FD$SOC_pct) | !is.numeric(raw_FD$BD_g_cm3) |
        !is.numeric(raw_FD$SOM_pct))){
  is_error <- TRUE
  stop(gettextf("Non-numeric SOC_pct, BD_g_cm3, or SOM_pct values"))
}


# Stop script if there is an error with the input file ----
if (is_error){
  warning(gettextf("ESM script has failed due to error(s) listed above"))
} else {

  # Begin processing input file ----

  # Remove extra columns and add FD type
  reduced_FD <- subset(raw_FD, select=c(ID, Rep, Ref_ID, Upper_cm, Lower_cm,
                                        SOC_pct, SOM_pct, BD_g_cm3))
  reduced_FD$Type <- "FD"

  # Order columns
  reduced_FD <- reduced_FD[, c("Type", "ID", "Rep", "Ref_ID", "Upper_cm",
                               "Lower_cm", "SOC_pct", "SOM_pct", "BD_g_cm3")]

  # Remove cores that do not have a surface (0 cm) sample
  filtered_FD <- reduced_FD %>% group_by(ID, Rep) %>%
    filter(min(Upper_cm) == 0)

  # Remove samples that have NAs
  filtered_FD <- filtered_FD  %>% drop_na

  # Remove any samples that are below a zone of non-contiguity
```

```r
filtered_FD <- filtered_FD %>% arrange(ID, Rep, Upper_cm, Lower_cm)

all_contiguous_FD <- filtered_FD %>%
  group_by(ID, Rep) %>%
  filter(all(Upper_cm==dplyr::lag(Lower_cm, default=FALSE)))

noncontiguous_FD <- filtered_FD %>%
  group_by(ID, Rep) %>%
  filter(any(Upper_cm!=dplyr::lag(Lower_cm, default=FALSE)))

removed_noncontiguous_FD <- noncontiguous_FD %>%
  group_by(ID, Rep) %>%
  filter(Upper_cm < Upper_cm[which(Upper_cm!=dplyr::lag(Lower_cm,
                                                        default=FALSE))])

filtered_FD <- rbind(all_contiguous_FD, removed_noncontiguous_FD)

# Remove cores that do not have a sample deeper than min_core_length_cm
filtered_FD <- filtered_FD %>%
  group_by(ID, Rep) %>%
  filter(!max(Lower_cm) < min_core_length_cm)

# Throw an error if there are no observations left in dataset
if (nrow(filtered_FD)==0){
  stop(gettextf("No observations remaining in dataset"))
}

# Add in zero masses at zero cm (for interpolation of first interval)
modified_FD <- filtered_FD %>%
  group_by(ID, Ref_ID, Rep) %>%
  summarise() %>%
  mutate(Type = "FD", Upper_cm=0, Lower_cm=0,
         SOC_pct=0, SOM_pct=0, BD_g_cm3=0) %>%
  bind_rows(filtered_FD, .) %>%
  arrange(ID, Ref_ID, Rep, Upper_cm, Lower_cm)

# Begin FD-based calculations ----

# Calculate soil mass in each interval
modified_FD$Soil_g_cm2 <-
  (modified_FD$Lower_cm-modified_FD$Upper_cm)*modified_FD$BD_g_cm3

# Calculate SOC mass in each interval
modified_FD$SOC_g_cm2 <- (modified_FD$SOC_pct/100)*modified_FD$Soil_g_cm2

# Calculate SOM mass in each interval
modified_FD$SOM_g_cm2 <- (modified_FD$SOM_pct/100)*modified_FD$Soil_g_cm2

# Calculate mineral soil mass in each interval
modified_FD$Min_Soil_g_cm2 <- modified_FD$Soil_g_cm2-modified_FD$SOM_g_cm2

# Calculate cumulative masses
cumulative_FD <- modified_FD %>%
  group_by(ID, Rep) %>%
  mutate(Cum_Soil_g_cm2 = cumsum(Soil_g_cm2),
         Cum_SOC_g_cm2 = cumsum(SOC_g_cm2),
         Cum_SOM_g_cm2 = cumsum(SOM_g_cm2),
```

```
        Cum_Min_Soil_g_cm2 = cumsum(Min_Soil_g_cm2))

# Begin ESM-based calculations
cumulative_ESM <- data.frame()

for (i in 1:nrow(distinct(cumulative_FD, ID, Ref_ID, Rep))){
  current_vals <- distinct(cumulative_FD, ID, Ref_ID, Rep)[i,]
  current_Rep <- subset(cumulative_FD, ID==current_vals$ID &
                            Ref_ID==current_vals$Ref_ID &
                            Rep==current_vals$Rep)

  # Subset the reference set of values
  current_refs <- subset(cumulative_FD, ID==current_vals$Ref_ID)

  # Average the reference values (in case of multiple values per depth)
  current_refs_mean <- current_refs %>% group_by(Upper_cm, Lower_cm) %>%
    filter(Lower_cm %in% ESM_depths_cm) %>%
    mutate_at(vars(-Upper_cm, -Lower_cm, -Cum_Min_Soil_g_cm2),
              function(x) x = NA) %>%
    mutate(Cum_Min_Soil_g_cm2 = mean(Cum_Min_Soil_g_cm2, na.rm=TRUE)) %>%
    summarise_all(mean) %>%
    mutate(ID=current_vals$ID, Ref_ID=current_vals$Ref_ID,
           Rep=current_vals$Rep, Type="ESM")

  #Determine whether extrapolation outside of maximum mass occurs
  if (extrapolation == FALSE) {
    # Remove references where mineral mass is greater than the sample max
    # Completely avoids extrapolation outside of spline model
    current_refs_filtered <-
      current_refs_mean[which(current_refs_mean$Cum_Min_Soil_g_cm2
                              <= max(current_Rep$Cum_Min_Soil_g_cm2)),]
  } else {
    # Remove references that have a depth greater than sample max
    # Extraoplates only to the maximum depth of the samples
    current_refs_filtered <-
      current_refs_mean[which(current_refs_mean$Lower_cm <=
                                max(current_Rep$Lower_cm)),]
  }

  # Interpolate SOC and SOM using cubic spline models
  current_refs_filtered$Cum_SOC_g_cm2 <-
    spline(x=current_Rep$Cum_Min_Soil_g_cm2,
           y=current_Rep$Cum_SOC_g_cm2,
           xout=current_refs_filtered$Cum_Min_Soil_g_cm2,
           method="hyman")$y

  current_refs_filtered$Cum_SOM_g_cm2 <-
    spline(x=current_Rep$Cum_Min_Soil_g_cm2,
           y=current_Rep$Cum_SOM_g_cm2,
           xout=current_refs_filtered$Cum_Min_Soil_g_cm2,
           method="hyman")$y

  # Calculate non-cumulative masses
  current_refs_final <- current_refs_filtered %>%
    group_by(ID, Ref_ID, Rep) %>%
    mutate(Min_Soil_g_cm2 =
             Cum_Min_Soil_g_cm2-dplyr::lag(Cum_Min_Soil_g_cm2, default=0),
```

```r
            SOC_g_cm2 = Cum_SOC_g_cm2-dplyr::lag(Cum_SOC_g_cm2, default=0),
            SOM_g_cm2 = Cum_SOM_g_cm2-dplyr::lag(Cum_SOM_g_cm2, default=0))

    current_refs_final$Soil_g_cm2 <-
      current_refs_final$Min_Soil_g_cm2 + current_refs_final$SOM_g_cm2

    current_refs_final$BD_g_cm3 <-
      current_refs_final$Soil_g_cm2/
      (current_refs_final$Lower_cm-current_refs_final$Upper_cm)

    current_refs_final$SOC_pct <-
      current_refs_final$SOC_g_cm2/current_refs_final$Soil_g_cm2*100

    current_refs_final$SOM_pct <-
      current_refs_final$SOM_g_cm2/current_refs_final$Soil_g_cm2*100

    current_refs_final$Cum_Soil_g_cm2 <-
      cumsum(current_refs_final$Soil_g_cm2)

    current_ESM <- data.frame(current_refs_final)
    cumulative_ESM <- rbind(cumulative_ESM, current_ESM)
  }

  # Post-processing cleanup and output ----
  # Remove zero masses at zero depth ESM and FD
  cumulative_FD <- subset(cumulative_FD, !(Upper_cm == 0 & Lower_cm == 0))
  cumulative_ESM <- subset(cumulative_ESM, !(Upper_cm == 0 & Lower_cm == 0))

  # Add NA for FD reference ID
  cumulative_FD$Ref_ID <- NA

  # Re-order column names in ESM dataset
  cumulative_ESM <- cumulative_ESM[colnames(cumulative_FD)]

  # Output datasets to XLSX file
  output_wb <- createWorkbook()
  addWorksheet(output_wb, "FD")
  addWorksheet(output_wb, "ESM")
  writeData(output_wb, "FD", cumulative_FD)
  writeData(output_wb, "ESM", cumulative_ESM)
  saveWorkbook(output_wb, output_filename)
  print(paste("Results have been saved to", output_filename, sep=" "))

  # Clear objects from R environment
  rm(list = ls())
}
```

Supplemental references

Anderson-Teixeira, K. J., Masters, M. D., Black, C. K., Zeri, M., Hussain, M., Bernacchi, C. J., DeLucia, E. H. (2013). Altered belowground carbon cycling following land-use change to perennial bioenergy crops. *Ecosystems*, 16, 508–520. doi: 10.1007/s10021-012-9628-x

McBratney, A. B., Minasny, B. (2010). Comment on "Determining soil carbon stock changes: Simple bulk density corrections fail". *Agriculture, Ecosystems & Environment*, 136, 185–186. doi: 10.1016/j.agee.2009.12.010

National Climatic Data Center (2019). National Centers for Environmental Information, National Oceanic and Atmospheric Administration. https://www.ncdc.noaa.gov/

Soil Survey Staff (2019). Web Soil Survey. Natural Resources Conservation Service, United States Department of Agriculture. https://websoilsurvey.sc.egov.usda.gov/

Schauberger, S., Walker, A. (2019). openxlsx: Read, write and edit XLSX files. R package version 4.1.4.

Wendt, J. W., Hauser, S. (2013). An equivalent soil mass procedure for monitoring soil organic carbon in multiple soil layers. *European Journal of Soil Science*, 64, 58–65. doi: 10.2136/sssaj2008.0063

Wickham, H., François, R., Henry, L., Müller, K. (2019). dplyr: A grammar of data manipulation. R package version 0.8.3.

Wickham, H., Henry, L. (2019). tidyr: Tidy Messy Data. R package version 1.0.0.

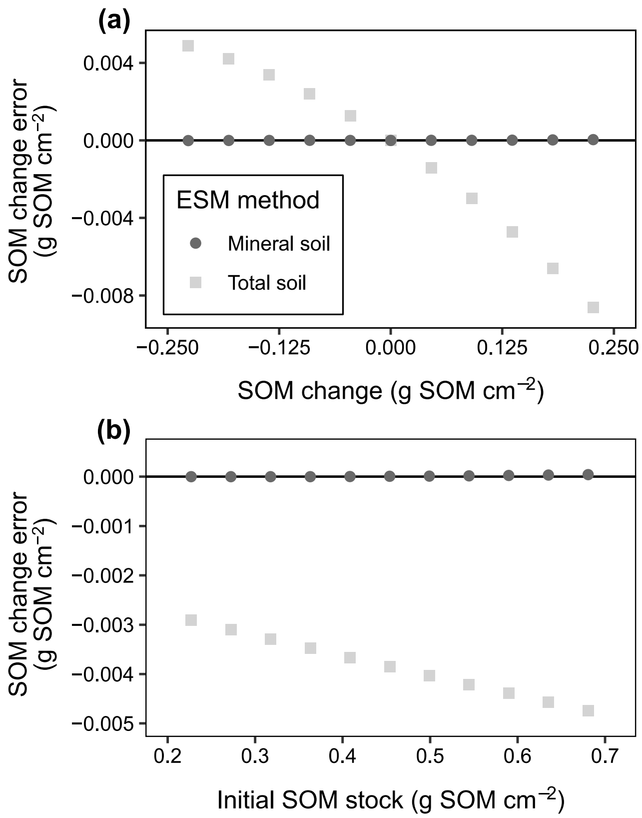

Figure S1 – Errors in the 0 to 10 cm SOM stock calculation resulting from using cumulative mineral soil versus cumulative total soil as the ESM reference mass under different simulated scenarios of SOM change (a) and initial SOM stocks (b). For panel a, the initial soil properties are the same as those given for "time 0" in Fig. 3 (main text), with SOM stocks calculated as SOC/0.58. The SOM stock changes range from a 50% decrease to a 50% increase (Supplemental methods). For panel b, the initial SOM stocks ranged from -50% to +50% of the SOM stocks contained within the "time 0" depth increments shown in Fig. 3 (main text), with SOM calculated from SOC stocks. The amount of SOM added in all scenarios in panel b was equivalent to 25% of the SOM stock in the "time 0" depth increments shown in Fig. 3 (main text).

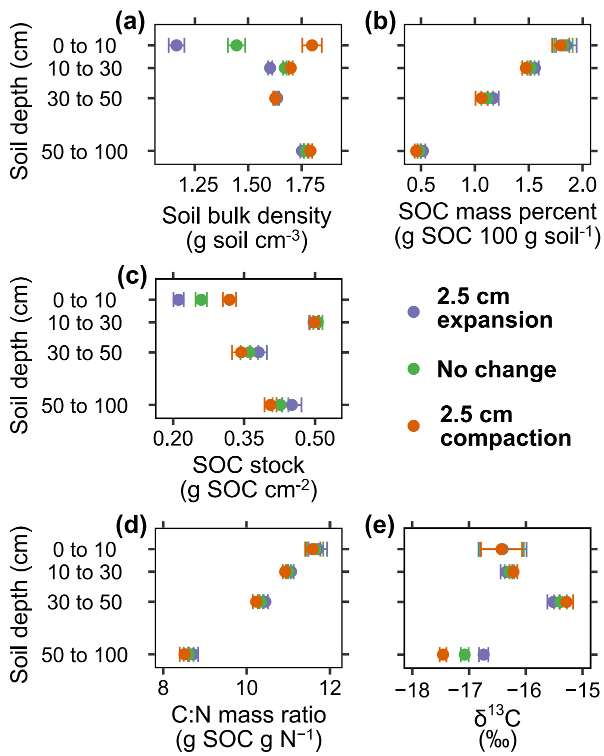

Figure S2 – Fixed depth-based soil properties for three scenarios including 2.5 cm of soil surface expansion, 2.5 cm of soil surface compaction, and no change. The no change scenario shows baseline values measured within four bioenergy cropping systems to which the expansion and compaction simulations were applied. Error bars show the standard error among the cropping systems.

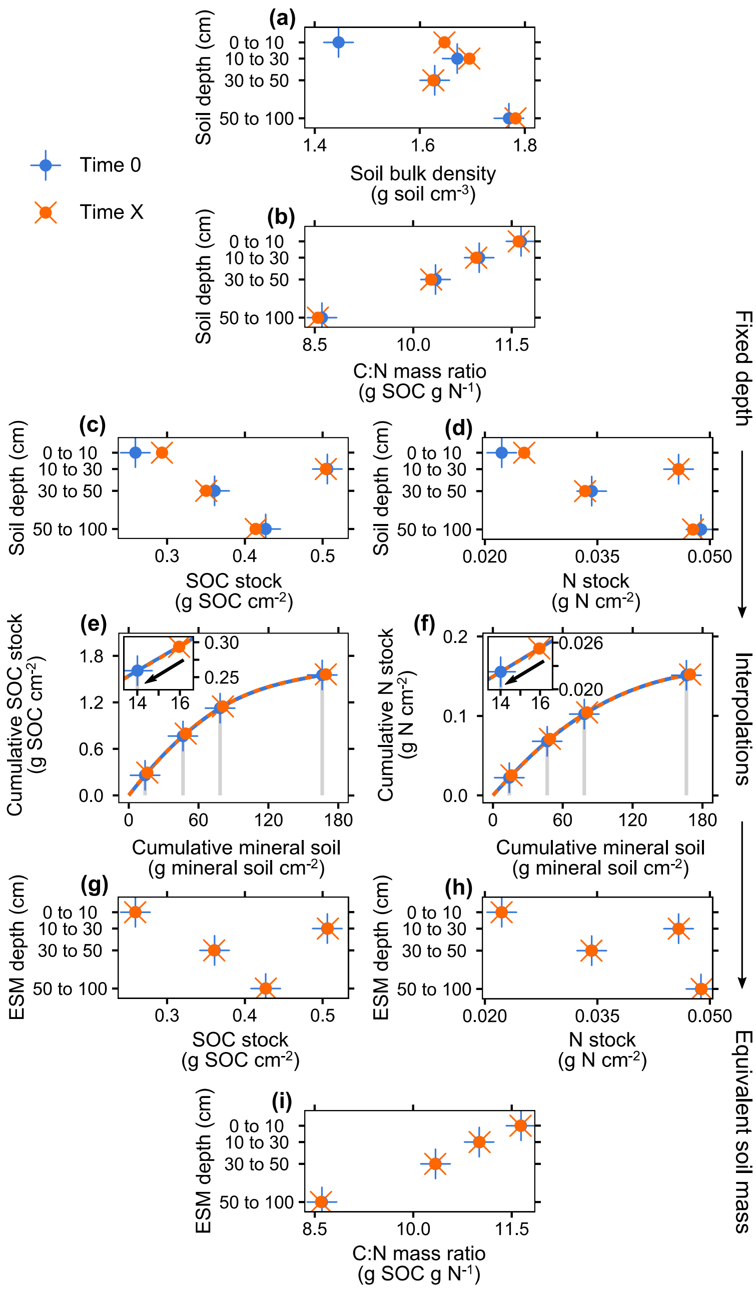

Figure S3 – Soil C:N mass ratio in fixed-depth (FD) and equivalent soil mass (ESM) form in a scenario where 1.5 cm of surface compaction occurs between time 0 and time X (a). Time 0 values were calculated from the average of four bioenergy cropping systems in Illinois, USA. Time X values were calculated based on simulated compaction of the soil surface (Supplemental methods). FD-based C:N values (b) are converted to ESM-based C:N values (i) by first calculating FD-based SOC and FD-based N stocks (c, d). The stocks are then converted to cumulative form and plotted against cumulative mineral soil (e, f). Time X values are interpolated at the time 0 reference mineral soil masses shown as the vertical gray lines (e, f). Non-cumulative ESM SOC and N stocks are calculated (g, h), and ESM C:N mass ratios are calculated within each depth increment (i).

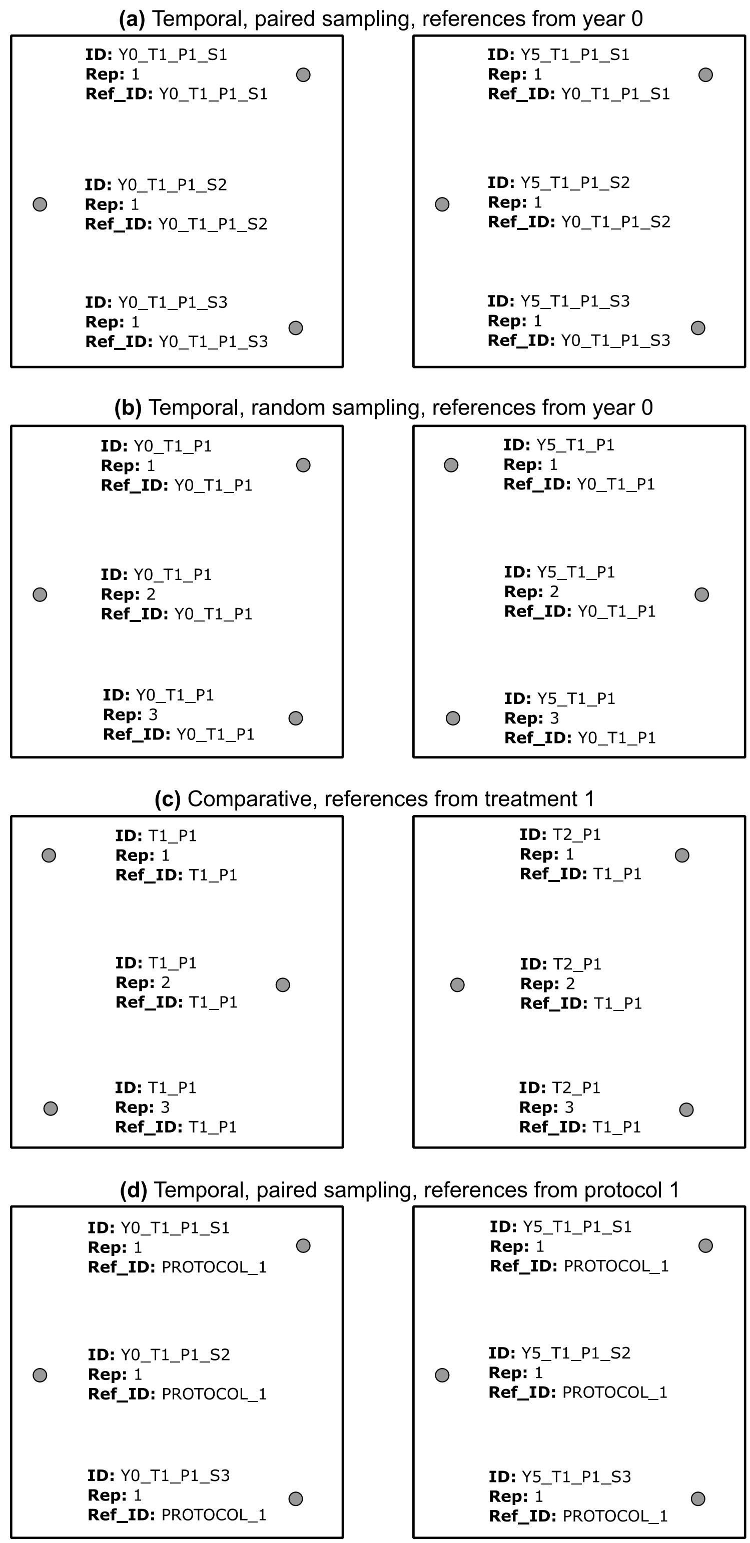

Figure S4 – Example sampling and labeling schemes (ID, Rep, and Ref_ID) for various several experimental designs. Large squares represent field plots, and circles show the locations where soil cores are taken in each plot. Y = year, T = treatment, P = plot, S = station. The experimental designs shown are intended for illustrative purposes only, as details will vary by experiment.