Dartmouth College
Dartmouth Digital Commons

Dartmouth College Master's Theses | Theses and Dissertations

Spring 5-11-2023

Soil Respiration Measurements Reveal High Retention of Organic Carbon from Corn Residue Derived High-Lignin Fermentation Byproduct Enabling Sustainable Lignocellulosic Biofuel Production

Michelle Sun Wang
Dartmouth College, michelle.s.wang.th@dartmouth.edu

Follow this and additional works at: https://digitalcommons.dartmouth.edu/masters_theses

 Part of the Agricultural Science Commons, Agronomy and Crop Sciences Commons, Biology and Biomimetic Materials Commons, Biomaterials Commons, Bioresource and Agricultural Engineering Commons, Other Engineering Science and Materials Commons, and the Systems and Integrative Engineering Commons

Recommended Citation

Wang, Michelle Sun, "Soil Respiration Measurements Reveal High Retention of Organic Carbon from Corn Residue Derived High-Lignin Fermentation Byproduct Enabling Sustainable Lignocellulosic Biofuel Production" (2023). *Dartmouth College Master's Theses*. 87.
https://digitalcommons.dartmouth.edu/masters_theses/87

This Thesis (Master's) is brought to you for free and open access by the Theses and Dissertations at Dartmouth Digital Commons. It has been accepted for inclusion in Dartmouth College Master's Theses by an authorized administrator of Dartmouth Digital Commons. For more information, please contact dartmouthdigitalcommons@groups.dartmouth.edu.

Soil Respiration Measurements Reveal High Retention of Organic Carbon from Corn Residue Derived High-Lignin Fermentation Byproduct Enabling Sustainable Lignocellulosic Biofuel Production

A Thesis
Submitted to the Faculty
in partial fulfillment of the requirements for the
degree of

Master of Science

in

Engineering Sciences

by Michelle Sun Wang

Thayer School of Engineering
Guarini School of Graduate and Advanced Studies
Dartmouth College
Hanover, New Hampshire

May 2023

Examining Committee:

Chairman_____
Lee Lynd

Member_____
Caitlin Hicks Pries

Member_____
Mark Laser

Member_____
Armen Kemanian

F. Jon Kull, Ph.D.
Dean of the Guarini School of Graduate and Advanced Studies

Abstract

While 2G biofuel production can utilize non-edible, lignocellulosic feedstocks such as agricultural residues to produce liquid fuel, harvesting crop residues is unsustainable without careful management of the soil underneath. By harvesting a fraction of the crop residues left in the field after harvest, soil health can diminish and critically, the soil organic carbon (SOC) stored in agricultural fields can decrease. Currently, in the most popular 2G process models published, the issue of soil degradation remains unresolved with residue harvest strategies receiving considerable attention in the literature and other SOC management strategies receiving far less. Specifically, the strategy of returning the high lignin fermentation byproduct (HLFB) from ethanol production to soil has been sparsely modelled and only tested experimentally once. Our study endeavors to expand on this literature by evaluating the SOC storage potential of various HLFBs and anaerobic digestates and comparing them to their unprocessed corn stover feedstocks using soil incubation experiments, isotope analysis, and simple modelling techniques. For both a 267-day and a 135-day incubation experiment, we measured the amount of carbon lost through microbial respiration and the amount of carbon remaining at the end. We found that in all but one case, for the same initial amounts of substrate inputs, the incubated digestate and HLFBs respired away less carbon and persisted longer in the soil than the incubated corn stover. Then, by applying multi-pool exponential decay models to our data, we found that the incubated corn stover respired away to completion substantially quicker than the biologically processed materials in our projected timespan of 100 years. We then approximated the steady-state SOC levels for a scenario in which the same bioprocessed materials were annually re-added to an incubation with our

preliminary results indicating that the biologically processed materials formed .95-4.8 more SOC than their unprocessed counterparts. Emboldened by our experimental results and tenuously strengthened by our preliminary modelling results, we believe that our work supports the feasibility of returning HLFB to soil to restore SOC and opens the door to the increased circularity and viability of biofuels in a future low carbon economy.

Acknowledgements

This document represents the culmination of a three-year long project that has defined the latter half of my time at Dartmouth. This period has been one marked by an incredible amount of both personal and professional growth that I am so delighted to have experienced in the Upper Valley. I am grateful for all my supporters far and close (too many to be satisfyingly thanked here) who have encouraged me before, during, and after the writing of this thesis. Thanks to those who bought me cold brew coffee at 10 p.m., who sent encouraging messages from all over the globe, and who regularly pretended to listen and find the minutiae of my work interesting.

Thank you to my advisors, Lee Lynd and Caitlin Hicks Pries, who have showed me what it means to pursue excellence in academia while maintaining zest for life and work. Thank you to Lee for taking a chance on me when starting this research so many years ago despite my little experience and at times, little aptitude for the subject. Your passion for science and grand vision for this project specifically have always inspired me. Thank you to Caitlin for actively guiding me in the research process and for the tremendous amount of time you invested in my education. Your belief in me combined with your generosity made my success possible. I could not have done this without your guidance.

Thank you also to my collaborators at Penn State and at the National Renewable Energy Laboratory as well as my supporters at the Arthur L. Irving Institute for Energy and Society. Thank you specifically to Armen Kemanian, Tom Richard, and Brooke Goggins for being founding members in the development of this project and for being kind,

generous, supporters throughout this effort. Thank you to Xiaowen Chen, Yannick Bomble, Nancy Dowle, Matt Fowler, Neal Hennge, and others for hosting me at NREL in the summer of 2022 and producing some great HLFB and analyses. Thank you to Stephen Doig, April Salas, and others at the Irving Institute who saw the potential of this project and provided the funding and legitimacy to support my efforts. Finally, thank you especially to Audrey Adamchak from Dartmouth for joining the project and doing such great work on our second incubation's microbial biomass analyses this past year.

Thank you of course to my many friends at Dartmouth, specifically those at the Hicks Pries Lab, Sustainability Office, Outdoor Programs Office, Machine Shop, and friends in the Thayer School faculty. Thank you for your unwavering support of my ideas and my endeavors no matter how much longer they may have taken than promised. Thank you to everyone who fed me snacks and Tuesday lunch, who offered advice on bridges and life, and who reassured me in my journey through academia. I am grateful for all of you.

Thank you to my friends and family spread across the world but particularly those at 236 Heater Road who made the attainment of this degree possible. Without you all, I would not have felt as happy, healthy, and loved in the Upper Valley these past few years. I am so thankful for the community we brought together at our house, feeding each other, singing karaoke, debriefing on the couch, and so many more moments of friendship that I will cherish. Thank you, finally, to my parents for their unwavering support of me in all of my ambitions; I would not be the person I am without you both. Finally, thank you to anyone reading this far; I hope you find some value in this document.

Table of Contents

List of Tables

List of Figures

List of Acronyms

Soil Organic Carbon	SOC
Soil Organic Matter	SOM
High Lignin Fermentation Byproduct	HLFB
Dilute Acid Steam Explosion	DASE
Anaerobic Digestion	AD
Corn Stover	CS
National Renewable Energy Laboratory	NREL
Incubation	INC
Carbon	C
Carbon to Nitrogen Ratio	C:N

Introduction

This thesis endeavors to provide the beginnings of an answer to the question, "How might soil organic carbon be maintained while 2nd generation biofuels are produced?" While we do not claim to be able to answer this question definitively in this document, we hope to provide a preliminary investigation rooted in experimental laboratory data into this topic that at this point lacks considerable experimentation attention. Because this thesis was originally written to serve as the basis for a manuscript to be submitted to Nature Sustainability, the structure of this thesis still largely conforms to Nature Publishing Group guidelines and its readership is presumed to include a cross section of those familiar with biofuel production and/or the soil sciences with some overlap in between. The Background section is intended to inform this particular readership of basic concepts central to understanding the topic from each discipline's point of view i.e. the need for bioenergy, processes in biofuel production, mechanisms driving SOC formation, etc.

To begin this effort, we will define common terms and ideas to be referenced throughout this paper starting with soil organic matter (SOM) and soil organic carbon (SOC). SOM is defined as the fraction of soil that consists of plant or animal tissue in various stages of decomposition such as decomposing agricultural crop residues, while SOC refers to purely the amount of carbon stored within the SOM[1,2]. Globally, SOM contains more than three times the carbon stock as contained in either the atmosphere or all terrestrial vegetation[3]. Soil quality, however, is defined by the Soil Society of America as the "capacity of a specific kind of soil to function, within natural or managed ecosystem boundaries, to sustain plant and animal productivity, maintain or enhance water and air quality, and

support human health and habitation[4].” Soil quality metrics include nutrient content, stable soil structure (water stable aggregates, water infiltration, gas exchange), cation exchange capacity, and water availability, etc[5]. While soil quality is not explicitly related to a specific SOC threshold, higher levels of SOC are often associated with increased soil quality metrics[4,6,7].

Thus, this thesis is interested in increased SOC as a desirable metric by which to evaluate the success of various agricultural management strategies pertinent to biofuel production. We define these management strategies as follows. First, in Case 1 (No-Harvest), no material is harvested from a field, and 100% of the crop residue left after crop harvest is left on the field. Farmers can chop and till the crop residue into the soil in order to promote decomposition prior to the next season’s planting[8]. A review of farmer educational materials suggest that this scenario is similar to what many corn farmers in the US practice[9–11]. Because corn stover, the non-edible parts of corn, is not a dependably profitable commodity in the US, farmers prioritize adding soil organic matter, controlling erosion, building soil nutrients, and controlling soil temperature — functions of crop residue cover — over the minimal profits they could make selling corn stover[9]. In Case 2 (Harvest), 50% of crop residue left on a field is harvested for biofuel production with no biofuel byproduct (henceforth referred to as high lignin fermentation byproduct or HLFB) returned to the soil. In this case, HLFB is assumed to be either used as a coal substitute in an offsite power plant or burned onsite for process heat as described in popular 2G biofuel process models[12]. A 50% harvest rate is aligned with the median harvest percentage as suggested by a literature comparison detailed in Appendix A. In

Case 3 (Harvest with HLFB Return), 50% of crop residue left on a field is harvested and accompanied by the return of the HLFB to the field as a soil amendment. We anticipate that this case will lead to the second most if not most amount of SOC formed (relative to the other cases) as the HLFB is composed of only lignified material altered during fermentation and microbial necromass, and will decompose slower and possibly with higher retention of SOC as compared to their fresh counterparts. Case 3 is of particular interest to our study and will be evaluated in proxy form relative to a proxy of Case 1, the No-Harvest scenario in which SOC should be (as aligned with the conventional wisdom of increased SOM leading to increased SOC) at a maximum.

Our specific question answered more fully in this study is, "How much carbon is retained in the soil from the input of biologically processed materials like biofuel byproducts compared to unprocessed materials like biofuel feedstocks?" To answer this question, we conducted soil incubations to experimentally compare how these different residues decompose in soil. While this experiment does not resemble an actual field scenario per the inherent limitations of a bottle incubation, relative comparisons can be made amongst the residues. Furthermore, to attempt to analyze a closer to field scenario, we extrapolated our incubation results using simple models based off our incubation data. These extrapolations, both for a one-time input of material and an annual input scenario, allowed us to compare more generally how the accumulation and loss of SOC compares amongst residues in a more realistic space and timeframe.

Regarding its value to the intersection of biofuel production and soil sciences literature,

this thesis offers the following: the most comprehensive material characterization information to date of three experimental HLFBs, the second set of experimental data ever collected on the carbon retention properties of HLFBs, and the first set of carbon partitioning data on HLFBs decomposed in soil with delineation of the effect of soil priming. We intend for the data provided in this thesis to serve as the robust, experimental accompaniment to an impending literature review published on the topic of HLFB return and for the conclusions this thesis draws to be informative to the ever-evolving design of integrated biorefineries (Appendix O).

Background

In one of the four mitigation pathways the IPCC has identified to possibly keep global climate change to an eventual 1.5°C increase, the IPCC has defined a "Negative Emissions" scenario in which half of the future global energy supply consists of biomass derived energy[13]. The scenario is reliant on the mass deployment of traditional and new-age biomass derived energy generation sources accompanied by carbon dioxide removal technologies that result in net negative global GH emissions. While this scenario does not strictly consider liquid lignocellulosic biofuels, researchers across the fields of renewable energy generation have directed increasing attention on second generation (2G) biofuel technology as a potentially impactful negative emission technology[14–16]. Unlike first generation (1G) biofuels, which are derived from edible food crops, 2G biofuels are derived from non-food sources such as dedicated energy crops or agricultural crop residues, generally consisting of lignocellulosic material[17].

However, while there still is considerable interest by private and public organizations alike in developing 2G biofuels, no commercial 2G biofuel plants remain in operation in the US today due to a variety of factors including but not limited to unmet inflated expectations from venture capitalists and the U.S. Department of Energy alike in the early 2010s, nagging unsolved supply chain issues, and regulatory uncertainty such as with the instability of the Renewable Fuel Standard[18]. However, even with the gradual maturation of the technology since the closure of the first cellulosic-ethanol plants, critics of 2G biofuels have maintained that the supply chain could be inherently unsustainable as the removal of agricultural crop residues, one requirement of a possible supply chain (at least for ethanol derived from crop residues and not dedicated energy crops), can cause significant damage to soil quality and reduce the amount of carbon stored in the soil[19–21]. This argument is especially damaging to the proliferation of the crop residue derived form of 2G biofuel production as healthy soil has been well established as the foundation of human livelihood and the carbon stored inside the soil as the foundation of a habitable climate[22,23].

The soil organic carbon (SOC) contained in soil organic matter (SOM) is the largest terrestrial pool of carbon on Earth, storing three times as much carbon as the atmosphere directly helping to regulate the world's climate[3]. The mechanisms driving SOM persistence and SOC sequestration by extension are still not fully understood; however, it is now generally accepted that molecular structure alone does not control SOM stability, but also a variety of biogeochemical factors and environmental conditions including, but not limited to, climate, moisture, depth, rhizosphere inputs, and microbial

communities[3,24]. Due to humankind's historical agricultural land use, soils have lost a cumulative ~133 Pg carbon on par with estimates of carbon lost from deforestation, equivalent to ~17% of what our atmosphere currently holds[25]. Agricultural management practices that keep SOC underground are thus increasingly important and sought after in an era of globally declining soil quality and SOC stocks[26]. The United Nations' "4 per 1000" Initiative set an international goal to grow SOC stocks 0.4% annually primarily in highly managed agricultural soils[27]. If this goal is fulfilled, agricultural soils could store 2-3 Gt carbon annually offsetting 20-35% of global anthropogenic carbon dioxide emissions[28]. Practices that retain and increase SOC stocks have been long debated in the agriculture literature including rotation of annual crops with perennials, increasing carbon input through the addition of organic matter, and no-till farming (in certain soil types and climactic conditions)[29–32].

The practice of adding inputs of organic matter to a field is particularly relevant to the specific form of 2G biofuel production that necessitates crop residue harvest. The same requirement does not necessarily apply to 2G biofuel production from dedicated energy crops. Since the turn of the century, there have been many studies published in the soil sciences literature to support the reduced harvest of crop residues on soil as a method to maintain soil fertility and SOC stocks in agricultural fields[33–36]. While the exact amount of crop residue left on agricultural fields to maintain healthy soil and stable SOC stock is still a matter of debate, generally the literature has settled on a suggested crop residue harvest rate range of approximately 50% to minimize soil erosion[19,37–41]. However, in regards to SOC, it has been suggested that an even higher percentage of residue needs to

remain on the field to offset SOC losses[42]. Generally, harvesting crop residues decreases SOC over time and requires the addition of externally produced fertilizers to maintain soil quality[12,43,44]. However, there is a convenient opportunity for circularity of SOC stock with the practice of returning the byproducts of 2G biofuel production to the soil[20,43–45].

This byproduct, termed high lignin fermented byproduct (HLFB), is the end result of 2G bioprocessing and is often assumed in process models to be either combusted to offset energy demands of production or converted into a high-value product through currently immature technologies [12,46,47]. However, modelling work completed in 2015 showed that by amending soil with HLFB, SOC is not only greater relative to a harvest only (no HLFB return) scenario, but also greater than a non-harvest scenario[43]. In other terms, when strictly considering crop residue carbon flows, returning HLFB to the field can result in net positive carbon storage whereas simply harvesting crop residues may result in carbon losses. Additionally, accompanying modelling work completed in 2013 shows that on a life cycle basis, returning HLFB to the soil results in greater avoided GHGs than if no residue were removed due to the emissions displaced by replacing fossil fuels and the soil carbon stored through the application of HLFB[44]. While these modelling results show promise for HLFB transforming into increased SOC, there is a dearth of experimental work actually testing this hypothesis.

The transformation of HLFB into persistent SOC is one of two potential fates for HLFB-derived carbon in soils. The first fate of HLFB-derived carbon is not in the soil itself, but in the atmosphere in the form of respired carbon dioxide during microbial decomposition.

Additionally, as soil microbes consume the accessible carbon containing sugars in HLFB for energy, their activity can lead to the phenomenon known as soil priming whereupon inputs of new carbon stimulates the decomposition of old soil carbon[48–50]. The priming effect can be either positive or negative i.e. cause an increase or decrease in old SOC respired, respectively, with no conclusive general mechanisms attributed to its cause[48,51]. In a variety of ecosystems, the priming effect has been observed to be a relatively short-term phenomenon that is controlled by several factors including but not limited to microbial community composition, SOM chemical structure, and nutrient availability[51]. The second fate of HLFB-derived carbon is to remain in the soil in the form of SOC or SOM. This remaining carbon is what the previously referenced models consider as stored SOC. This remaining carbon can become adsorbed to soil minerals where it is relatively protected from further decomposition through abiotic leaching or the death of microbial biomass[52,53]. Or, the HLFB-derived carbon can remain relatively untransformed, perhaps due to its molecular structure. Lignin, an organic polymer found in plant tissue that is in high concentrations in HLFB, has been historically correlated with higher amounts of SOC sequestration in a variety of field and laboratory experiments[50,54–57]. While there has certainly been skepticism of the role of lignin in leading or lagging SOC sequestration, recent studies suggest that despite the complexity of lignin fates in soil, increased amounts of lignin in soil can assist in SOC accumulation[54,58,59].

Considering the two fates of HLFB in soil, experimental work on both the laboratory and field scale is needed. To our knowledge, there has only been a single published study where an HLFB was actually added to soils and its impacts on soil carbon storage and

soil health were studied[60]. In the experiment, HLFB was applied at the same rate to soil as its unprocessed feedstock, corn stover. Over 112 days, the HLFB released half of the carbon the corn stover released and showed small but statistically significant positive effects on soil quality such as decreased bulk density, increased water retention, and a greater percentage of water-stable aggregates. Additionally, the study included a fertility experiment in which crops were grown in HLFB amended and non-amended soils. Plant growth heights were found to be similar between the two soils suggesting that the HLFB amendments did not negatively harm soil fertility. While this is the only study to experimentally test HLFB as a soil amendment, there is an abundance of literature testing anaerobic digestates as soil amendments, an adjacent bioprocessed material also coproduced from a bioenergy production process. When equal masses of digested and undigested agricultural residues were added to soils, several laboratory soil incubation studies observed that digested residues released carbon more slowly than undigested residues, and that added digestate led to the formation of stored SOC[61–64]. Exemplified by the results of a 2021 study analyzing the efficacy of biosolid byproducts from anaerobic digestion as a soil amendment, soil quality improves and carbon sequestration increases when bioprocessed material is added to soil relative to unprocessed residue[65].

Our study assesses the impact of organic matter addition on SOC with and without 2G biofuel specific biological processing. To expand on the sparse experimental data of HLFB return on soil, we incubated three more HLFBs with more documentation than currently presently in the literature. In total, we performed two soil incubations with the three HLFBs from different sources, two anaerobically digested residues, and two

unprocessed corn stover samples. We determined their decomposition rates, effect on native SOC, and potential for long term carbon sequestration. Additionally, we also tested the effect of soil type and substrate dosage to validate our experimental assumptions. Finally, we performed isotope analysis to partition between soil-derived SOC and residue-derived SOC which allowed us to approximate the amount of soil priming caused by HLFB addition. We expected that the HLFBs would release less carbon than their unprocessed counterparts, which would lead to more carbon retained even with the mass lost from bioconversion accounted for. Providing experimental data in support of this hypothesis will be a valuable addition to the biofuel literature and inform a larger conclusion in support of the agricultural management strategy of returning HLFB to soil as a means to restore SOC lost during biofuel production.

Methods

Material Characterization

Pre-Incubation

We analyzed the materials used in our incubation for percent carbon, percent nitrogen, 13C, and 15N using an Infrared Mass Spectrometer with an elemental analyzer attached (EA Isolink™ CNSOH IRMS System). Pre analysis, all material was dried and ground to fine powder on a rolling table or ball mill (Spex SamplePrep 8000M-115 Mill). For percent lignin and structural sugars, we sent our residue samples to the National Renewable Laboratory analytical team where they performed their Laboratory Analytical Procedure for the Determination of Structural Carbohydrates and Lignin in Biomass[66].

Post-Incubation

After the incubations concluded, we dried and ground a portion of our treatments for analysis of percent carbon, percent nitrogen, and ^{13}C following the same pre-incubation protocol. Additionally, we performed chloroform fumigations and potassium extractions on the non-dried portions of treatments to quantify microbial biomass carbon and nitrogen using an organic carbon analyzer (GE Sievers 900 Series Laboratory TOC Analyzer) and Lachat auto-analyzer. Results from these tests are not discussed further in this thesis but are included in Appendix B.

Soil Incubations

In our study, we tested the effect of two distinct soil types with contrasting amounts of organic matter. For both incubations, we added substrates to Palouse soil, a fine-silty, mixed, superactive, mesic Pachic Ultic Haploxerolls that had previously grown wheat in Pullman, Washington (USDA-ARS Palouse Conservation Farm). This soil had received inputs exclusively from C3 plant material and had not received to our knowledge any input of animal manure. For Incubation one only, we also added substrates to Vershire soil, a coarse-loamy, mixed, active, frigid Humic Dystrudepts that had previously been used for grazing in Vershire, Vermont. To our best knowledge, the soil had been multi-use and certainly received inputs of animal manure. More information on the soils used in this study can be found in Appendix M.

Two soil incubations experiments were conducted that spanned 267 and 135 days (any ranges given throughout the Methods section reflect slight differences in the two

incubations' experimental conditions). Incubation experiments entail the careful maintenance of standardized conditions and monitoring of a variety of experimental units and controls throughout the study's span. Each experimental unit, referred to as "treatment" going forward, consisted of a portion of wet soil and a portion of substrate. The treatments were kept in plastic sample cups and placed inside pint sized mason jars sealed with an airtight lid and stopcock valve. Gas measurements for determining CO_2 concentrations were taken using 60 mL syringes that drew from the stopcock valves. The experimental conditions across both incubations varied slightly as seen in Table 1. Per a preliminary incubation conducted in preparation of this study, we found particle size (within a .50-8.5 mm range) to have a non-significant effect on decomposition rates (Appendix C) [67].

Condition	Incubation 1	Incubation 2
Temperature	22 °C	25 °C
Light	Off	Off
Moisture	90% FC (Palouse: FC = 23.9%; Vershire: 35.5%)	90% FC (Palouse: FC = 23.9%)
Dryness of Amendments	Dried in oven	Dried in oven
Water Refill	Refilled water weekly	Refilled water once a week (last measurement of the week), eventually once every two weeks
Amendment to Soil Ratio	1.5g residue to 37.5g soil	1.5g residue to 37.5g soil and 0.75g residue to 37.5 g soil
Replicate Number	4 replicates per treatment except for AD1 treatments with 3 replicates	3 replicates per treatment
Timeframe	135 days	267 days
Stopcock Valves	Closed in between measurements	Open in between measurements

Table 1: Incubation conditions and differences. FC stands for field capacity, the amount of water a soil can hold without draining.

Substrates were dried, milled, and incubated with soil in individual jars in a mass ratio of 2:50, substrate to dry soil. By varying experimental parameters, we tested the effect of treatment, soil type, and substrate dosage on carbon respired by the treatment, fraction of carbon retained by the substrate, and fraction of carbon retained by the soil within the span of the incubation.

Preincubation

To build up a steady microbial population after rewetting dried soils (Palouse) or adjusting the water content of field moist soils (Vershire), we pre-incubated our incubation jars for 10 days for the Palouse soil and 3 days for the Vershire soil before adding the substrate of interest. The soil was moistened to the appropriate water content

before preincubation and water was added to account for loss after preincubation. Afterwards, each treatment's residue portion was mixed into the individual soil sample thoroughly by hand with a scoopula until the treatment appeared homogenously mixed.

Incubation Conditions

Incubations were conducted in a Thermo Scientific™ Precision™ Low Temperature BOD Refrigerated Incubator maintained at 22 and 25 °C for Incubations one and two, respectively. Soil moisture was maintained at 20% gravitational water content for Palouse soil samples and 32% for Vershire. This water content represents 90% of each soil's field capacity. We added deionized water to each jar weekly by weight to stay within 10% of the original water content.

Measurements and Calculations

We measured CO_2 concentrations in the jar headspace regularly, transitioning from a twice-a-day to a weekly to a biweekly schedule as the incubations progressed and microbial activity slowed. We sampled 30 mL of air from the headspace of each jar and ran that air through an infrared gas analyzer (IRGA; PP Systems EGM-5, Amesbury, Massachusetts). From our concentration measurements of time 0 and time 1 and using the ideal gas law to convert ppm CO_2 to grams C, we calculated the flux of carbon respired (mg/day) for each sampling interval. Then, we performed a trapezoidal Riemann sum integration to estimate the cumulative amount of carbon respired by day (mg) in between sampling times. Graphs showing both carbon flux and cumulative carbon respired for each incubation can be found in Appendix D. The data for making these graphs can be

found in Appendix Q. Additionally, the raw data consisting of IRGA measurements can also be found in the Hicks Pries Lab Github.

Additionally, we conducted isotopic analyses on our treatments pre and post incubation using an Infrared Mass Spectrometer with an elemental analyzer attached (EA Isolink™ CNSOH IRMS System). We analyzed the materials for percent carbon, percent nitrogen, 13C, and 15N. This allowed us to partition between residue-derived carbon and soil-derived carbon losses in our residue containing treatments. Using our 13C measurements and the following equations, we can find, $frac_{residue}$, or the fraction of total carbon derived from the residue. Correspondingly, $frac_{soil}$ can be solved for which represents the fraction of total carbon derived from the soil. The data and calculations for this analysis can be found in the Hicks Pries Lab Github and in Appendix P.

$$frac_{residue} = \frac{13C_{mix,post-inc} - 13C_{soil,pre-inc}}{13C_{residue,pre-inc} - 13C_{soil,pre-inc}}$$

$$1 = frac_{residue} + frac_{soil}$$

From this partitioning work, priming can be calculated. Soil priming is defined in this study as the difference between soil-derived carbon loss from a substrate containing treatment and soil-derived carbon loss from a comparable soil control pre and post incubation. Notably, there can be both positive and negative soil priming with positive priming meaning carbon loss of soil is stimulated by the addition of organic matter input while negative priming meaning that the addition of residue reduced the decomposition of soil carbon.

Statistics

We performed various statistical analyses on our study data including 2-way and 1-way ANOVAs, Tukey Honest Significant Differences (HSD) tests, and paired T-tests. To investigate significant differences amongst treatments in various metrics including carbon respired and partitioned carbon losses, we used the baseR functions *aov*, *Anova (Type III)*, and *TukeyHSD*[68]. To investigate differences amongst our incubation conditions including open and closed valves, soil controls, and a shared treatment between Incubation one and two, we used the baseR function *t.test*. We looked at the outputted p values to identify significant differences with $p < .05$ considered significant and ran a Tukey HSD posthoc test to determine significant differences between pairs of treatments. Appendix L and the Hicks Pries Lab Github contains the code we used for our statistical analyses.

Modelling

One-Time Input Modelling

We fit a variety of multi-pool models to our experimentally derived cumulative carbon respired data for our various treatments. The one-time input we reference here refers to the initial amount of carbon in our incubations consisting of soil and a portion of substrate. We utilized the SoilR package which contained built in functions and models to both fit and extrapolate our data[69]. We tested models with two-pool structures with the pools representing conceptual fractions of carbon that decompose at distinct rates (fast and slow). Two-pool models included both a series structure where a portion of the fast pool is transferred to the slow pool and a parallel structure where the pools decomposed

independently. The models partition carbon into the pools, agnostic to the ratio of substrate to soil carbon in our incubations. Based on the R^2 values (how well the model fit our data) for the various models, we chose to apply the two-pool parallel model to the soil controls and the two-pool series model to the substrate-containing treatments. The estimated best-fit parameters and R^2 values can be found in Appendix E. We used those parameters to project the amount of residue carbon remaining in the soil up to 100-year timespans.

Generally, in the SoilR package, two-pool models take the form:

$$\frac{dC}{dt} = I(t) + A * C(t) = A * C(t)$$

where C(t) is a 2x1 vector of carbon stores in two pools at a given time t and I(t) is a time-dependent column vector describing the amount of input to each pool. For our analysis, we assumed I(t) = 0 as there were no extra inputs beyond the initial input of soil and substrate carbon present in our incubations. This simplifies the solution to the previous equation to:

$$C(t) = C_0 * e^{A(t-t_0)}$$

Gamma, γ, which is included in our parameter tables in Appendix E represents the partitioning of C_0 into two pools with C_{total} being experimentally determined by our IRMS analysis detailed in Appendix P.

$$C_0 = C_{total} * \begin{bmatrix} \gamma \\ 1-\gamma \end{bmatrix}$$

A is a 2x2 square matrix containing decomposition rates for each pool and transfer coefficients between each pool. For the two-pool parallel model, A has the form:

$$A = \begin{bmatrix} -k_1 & 0 \\ 0 & -k_2 \end{bmatrix}$$

For the two-pool series model, A has the form:

$$A = \begin{bmatrix} -k_1 & 0 \\ \alpha_{2,1} & -k_2 \end{bmatrix}$$

We fit the models to our data using the *modFit* function in SoilR which performs a Nelder-Mead optimization to find best-fit parameters. Then, we reran the models to extrapolate our data into longer time series using the *getAccumulatedRelease* function on the *TwopParallelModel* and *TwopSeriesModel* statements we wrote. To reiterate, we chose to apply the two-pool parallel model to the soil controls and the two-pool series model to the substrate-containing treatments. The code used to accomplish the modelling described here can be found in Appendix F and in the Soil Incubation repository in the Hicks-Pries Lab GitHub.

Annual Input Steady State Modelling

We additively combined our one-time input modelling results to project long-term steady state values of carbon retained in our treatments in which the initial amount of substrate carbon present in the treatment is added to the soil every year for 100 years. This effort imitates a simple bioenergy cropping scenario in which organic matter is added to soil every year after crop residue harvest. We estimated the average steady state amount of residue-derived carbon present and quantitatively compared these results amongst our treatments. Our primary assumptions are that soil priming has little to no effect on 100-year scale projections, that the original soil-derived carbon will degrade to a point approaching zero despite new additions of organic material, and that the residue dosage

does not affect the values of the modeled parameters. For computing efficiency, we modelled the treatments using a timespan of 100 years with one fifth year time step. Since the model fitted parameters reflect experimental scenarios and are unable to differentiate between soil- or substrate-derived carbon, we isolated the substrate by subtracting the modeled soil control projections from the modeled substrate-containing treatment projections. From that result, we calculated the amount of carbon retained at any time step by subtracting the cumulative carbon respired from the cumulative carbon added. The calculations used to produce the work described here is shown in Appendix G and can be found in the Incubation Modelling repository in the Hicks Pries Lab Github.

Results

Preparation of High Lignin Fermentation Byproduct

We sourced five biologically processed residues prepared at lab, pilot, and industrial scale operations for our soil incubation experiments. We used corn stover derived materials because corn stover is considered the most abundant crop residue for 2G biofuel production in the US[70]. Corn stover from the leading U.S. biofuel producing company, POET, was subject to either anaerobic digestion or a dilute-acid steam explosion (DASE) pretreatment, saccharification, and fermentation aligned with lignocellulosic liquid biofuel production protocols unique to each preparer. Hereafter, we refer to residues prepared via anaerobic digestion as digestate, and residues prepared via DASE as HLFB. Details on the preparation of each residue can be found in Table 2.

Substrate	Preparer	Feedstock	Pretreatment	Hydrolysis	Fermentation	Mass Yield
AD1	Lynd Lab, Dartmouth College	Prepared with CS1 milled to a cut-off size of 0.5 mm	N/A		- Residence time of 480 hours - Final carbohydrate content of 0.23 g carbohydrate/g dry solid	-
AD2	Lynd Lab, Dartmouth College	Prepared with CS2 milled to a cut-off size of 0.5 mm	N/A		- Residence time of 240 hours with once-a-day renewals of 10% media - Carbohydrate solubilization of 82 ± 5.5% - Solids mass lass of 61 ± 6.0% - Final carbohydrate content of 0.36 g carbohydrate/g dry solid	-
HLFB1	Wyman Lab, UC Riverside	Prepared with CS1 milled to a cut-off size of 1/8"	- Impregnated with 0.5% sulfuric acid overnight - Steam exploded for 20 minutes at 160 deg C	- Simultaneous saccharification carried out over 10 days - Inoculated with Cellic® CTec2 at 15 mg enzyme/g glucan - Fermented with *Saccharomyces cerevisiae* D5A at 37 deg C		26%
HLFB2	Integrated Biorefinery Research Facility, NREL	Prepared with CS2 milled to a cut-off size of 1/8"	- Impregnated with 1% sulfuric acid overnight - Steam exploded for 10 minutes at 170 deg C	Inoculated with Cellic® CTec3 at 119.6 g/L of glucose; 1.4 g/L of xylose	Fermented with *Saccharomyces cerevisiae* D5A at room temperature for 25.5 hours	15%
HLFB3	Project Liberty, POET	Prepared with Project Liberty Corn Stover; particle size ranging from 2.75" to 6"	Details to preparation as described in Project Liberty: Launch of an Integrated Bio-Refinery with Eco-Sustainable and Renewable Technologies. Conversion of Corn Stover Biomass to Bio-Ethanol, Final Report. *			47%

*Project Liberty preparation details are documented in Martin et al. 2021.[71]

Table 2. Biologically processed residues in this study with relevant preparation details. Mass yield refers to the ratio of dried HLFB produced from dried feedstock. Any information not included in the table is due to a lack of documentation. CS1 and CS2 refer to the two different corn stover feedstocks used to produce HLFB1 and HLFB2, which were incubated alongside the HLFB treatments.

The three HLFBs (HLFB1, HLFB2, HLFB3) were intentionally sourced from a variety of operational scales to reflect how diverse residues result from similar DASE protocols in the evolving lignocellulosic biofuel industry. We were interested specifically in residues produced via DASE pretreatment as NREL has consistently included dilute acid pretreatment in its reports on the state-of-the-art process design and economics of integrated biorefinery pathways[12,72,73]. We included anaerobic digestates in our study as there is sufficient literature supporting anaerobic digestate's promise as a carbon storing

soil amendment[5163,65,74,75]. To understand the material characteristics of the biofuel byproducts and to address the 2G biofuel space's dearth of information on HLFBs, we determined the composition of these residues following the methods described in Methods (Table 3). Despite the HLFBs undergoing technically similar bioconversion processes, the resulting materials varied considerably in our metrics of interest: C:N, lignin:N and average total sugar concentrations.

All of the bioprocessed materials save for HLB3 were produced from dried corn stover (CS1, CS2) that we incubated in conjunction with the biologically processed residues. Corn stover includes all the non-edible aerial parts of the maize plant including cobs, husks, leaves, and stalks left after crop harvesting. Thus, there can be considerable variability in feedstock carbohydrate levels, which affect the maximum theoretical biofuels yield, optimum pretreatment, saccharification conditions, and ultimately, the composition of the residues[76]. Importantly, the majority of the fermentable sugars in corn stover are shielded by enzyme resistant carbohydrate-lignin linkages, which are targeted for breakdown by various modern pretreatment processes[77]. Amongst our samples, corn stover contained two to three times more structural sugars and two to three times less lignin by mass fraction compared to its processed counterparts. Thus, lignin is left primarily inert through biochemical conversion as referenced by the term, HLFB.

				Carbon and Nitrogen[1]			Lignin[2]		Structural Sugars[2]					Solubilization[3]
Incubation	Substrate	Substrate	Feedstock	%C	%N	C:N	%Lignin	Lignin: N	%Glucan	%Xylan	%Galactan	%Arabinan	Average Total % Sugar	General Solubilization
1	CS1	Corn Stover	CS1	44 (.37)	0.49 (.03)	89	18 (.05)	37	39 (.02)	26 (.41)	1.5 (.24)	3.5 (.10)	71 (.57)	0%
	AD1	Anaerobic Digestate	CS1	36 (.35)	1.6 (.02)	22	40 (.25)	24	13 (.21)	8 (.08)	1.3 (.01)	2.2 (.03)	24 (.33)	-
	HLFB1	HLFB	CS1	44 (1.1)	3 (03)	15	56 (.33)	19	4.8 (.03)	0.5 (.03)	0 (.00)	1.8 (.04)	7.1 (.01)	74%
2	CS2	Corn Stover	CS2	46 (.14)	0.77 (.04)	60	16 (.08)	21	39 (.16)	28 (.24)	1.8 (.29)	3.9 (.01)	73 (.39)	0%
	AD2	Anaerobic Digestate	CS2	44 (.43)	3.4 (.03)	13	32 (.04)	9.6	19 (.08)	12 (.14)	1.4 (.09)	2.1 (.00)	35 (.13)	65%
	HLFB2*	HLFB	CS2	50 (.04)	1.3 (.02)	39	56	43	26	3.1	0.00	0.00	29	84%
	HLFB3	HLFB	CS**	45 (.03)	2 (.01)	22	45 (.48)	22	23 (.02)	3.9 (.01)	0 (.00)	0.54 (.00)	27 (.03)	52%

*Analysis of Lignin and Structural Sugars for HLFB2 only contained 1 replicate and therefore standard error for those tests is not included for HLFB2.
**Corn stover used to produce HLFB3 was not incubated in this study as the HLFB was leftover from Project Liberty and the original feedstock was unavailable.
[1]Carbon and nitrogen content of the samples were measured at Dartmouth College in the Hicks Pries Lab using an EA Isolink™ CNSOH IRMS System.
[2]Lignin content and structural sugar content were analyzed at NREL following their Laboratory Analytical Procedure for the Determination of Structural Carbohydrates and Lignin in Biomass[66].
[3]General solubilization was calculated from the respective mass yields and %C information presented. General solubilization is defined here as $\%Solubilization = 1 - mass\ yield * \frac{\%C\ substrate}{\%C\ feedstock}$

Table 3: Material characteristics of the various residues. General solubilization is an analog for biological degradation with greater solubilization equaling greater amount of biological degradation. Standard error is represented in parentheticals next to mean values with three to four replicates analyzed for Carbon and Nitrogen tests and two replicates for Lignin and Structural Sugars analysis.

Despite being prepared in the same bioreactors, AD1 and AD2 differed significantly in their material characteristics due to AD1's double amount of residence time and corresponding biological degradation. Despite this difference, the amount of lignin in the material is comparable. Relatedly, dilute acid steam explosion is sometimes used a pretreatment for anaerobic digestion as the physical attack on plant cell walls makes the fermentable sugars bound in lignin more easily accessible[78].

Comparison of Carbon Retention of HLFBs Using Soil Incubation Data and Isotope Analysis

The HLFBs, digestate, and corresponding corn stover feedstock were incubated with soil as part of either a 267-day (Incubation one) or 135-day (Incubation two) experiment. The incubation results showed the expected exponential decay of C production over time of the treatments (Appendix D).

		Initial Amounts of C		135 days				267 days			
		Soil	Residue	Treatment (Soil + Residue)			Residue	Treatment (Soil + Residue)			Residue
Incubation	Substrate	mg C	mg C	mg C Respired	% Incubation Control	% C Respired Out of Initial	% C Respired Out of Initial	mg C Respired	% Incubation Control	% C Respired Out of Initial	% C Respired Out of Initial
1 (Palouse Soil, Standard Dosage)	Soil Control	514 ± 1.7	-	42 ± .43	100%	8%	N/A	66 ± 1.2	100%	13%	-
	CS1	513 ± 2.1	633 ± 7.2	320 ± 19	774%	28%		458 ± 14	691%	40%	59%
	AD1	515 ± 2.3	544 ± 14	100 ± .33	240%	9%		144 ± 1.9	218%	14%	14%
	HLFB1	515 ± .58	673 ± 9.8	106 ± .45	256%	9%		139 ± 1.2	210%	12%	18%
1 (Vershire Soil, Standard Dosage)	Soil Control	1113 ± 6.5	-	42 ± .53	100%	4%		68 ± .13	100%	6%	-
	CS1	1117 ± 4.8	635 ± 7.6	267 ± 5.1	638%	15%		395 ± 11	577%	23%	44%
	AD1	1112 ± 7.8	539 ± 3.6	143 ± 2.2	341%	9%		199 ± 1.7	292%	12%	11%
	HLFB1	1110 ± 5.3	673 ± 9.1	104 ± .44	250%	6%		136 ± 1.4	199%	8%	14%
2 (Palouse Soil, Standard Dosage)	Soil Control	541 ± .73	-	38 ± 1.5	100%	7%	-	N/A			
	AD2	535 ± .64	668 ± .65	195 ± 5.0	508%	16%	45%				
	HLFB1	535 ± .67	664 ± 1.8	142 ± 2.9	382%	12%	25%				
	HLFB2	534 ± .76	746 ± .74	209 ± .87	542%	16%	38%				
	HLFB3	534 ± .80	675 ± .36	94 ± 1.2	244%	8%	33%				
2 (Palouse Soil, Reduced Dosage)	CS2	535 ± .43	346 ± .38	201 ± 8.4	523%	23%	60%				
	AD2	536 ± .41	334 ± .04	117 ± 2.5	305%	14%	45%				
	HLFB2	534 ± 1.0	373 ± .20	137 ± 2.6	355%	15%	33%				
	HLFB3	535 ± .39	337 ± .39	62 ± 2.2	161%	7%	15%				

Table 4: Carbon respired across various treatments and incubation-condition groups. Isotope analysis was performed at the end of each incubation to differentiate between losses in soil derived carbon and residue derived carbon as shown in the Residue, % C Respired Out of Initial columns. Since Incubation one lasted for a total of 267 days, carbon partitioning data was unavailable for the 135-day case though amount of cumulative carbon until this point was calculatable and displayed. Results are to be compared with respect to the other treatments within their own incubation-condition group and not across different incubation-condition groups. Reduced dosage treatments are treatments incubated with half the residue dosage of other (i.e., in a ratio of .75 g residue to 37.5 g soil versus the standard dosage of 1.5 g residue to 37.5 g soil).

Across the four incubation-condition groups (i.e., a group of treatments defined by the same soil type, incubation, and dosage), corn stover treatments consistently released the most carbon (5-7 times more than the soil control) relative to all other treatments within the group (Table 4). This was true regardless of differences in timeframe, dosage, or soil

type across the four groups. From graphs of the carbon respired by treatments over time, the corn stover treatments had the highest initial slope and most delayed approach to an asymptote if an asymptote was approached at all (Appendix D). Our separate two-way ANOVAs of Incubation one and Incubation two showed that different substrates had a significant effect on the amount of carbon respired (two-way ANOVA, Incubation one substrate effect, $df = 4$, $p << 0.05$; Incubation two substrate effect, $df = 2$, $p << 0.05$). Soil type did not have a significant effect in Incubation one on the amount of carbon respired (two-way ANOVA, soil effect, $df = 1$, $p = 0.78$). Dosage, on the other hand, had a significant effect on amount of carbon respired in Incubation two (two-way ANOVA, dosage effect, $df = 1$, $p << 0.05$), but the effect was not directly proportional. On average, a two-fold increase in dosage translated to a $160 \pm 8\%$ increase in amount of carbon respired across AD2, HLFB2, and HLFB3 (Appendix H).

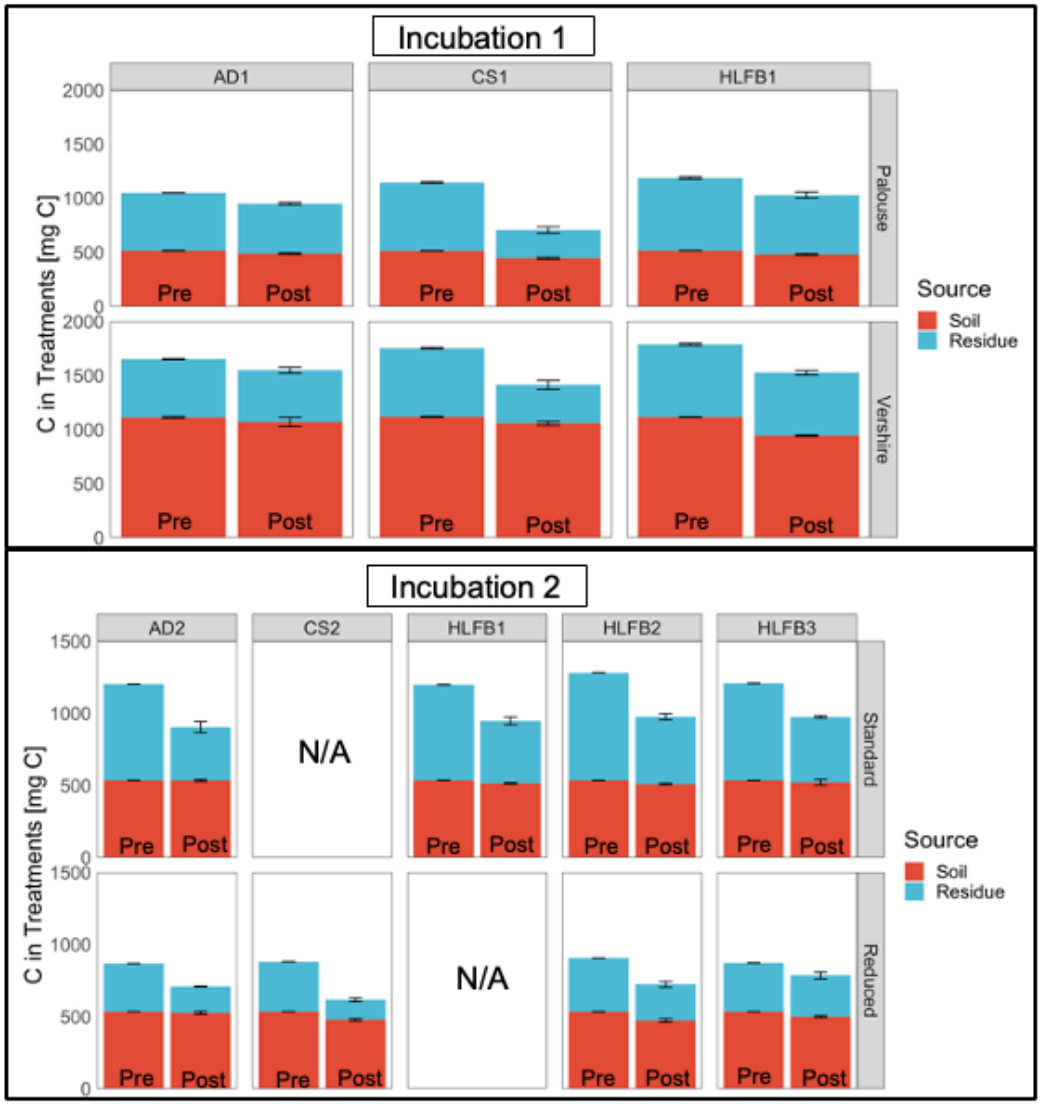


Figure 1: The amount of soil and residue derived carbon contained in treatments pre and post incubation. We used isotope analysis to partition the amounts of carbon between residue and soil pre and post incubation for the various treatments. The stacked bars represent the amount of carbon in treatments partitioned by source of carbon (i.e., soil derived carbon versus residue derived carbon). In each box, the left bar represents the total carbon in treatment pre incubation, and the right bar represents the total carbon in treatment post incubation. The top boxed figure shows Incubation one data, while the bottom boxed figure shows Incubation two data.

Isotope analysis of the treatments pre and post incubations revealed how soil-derived and residue-derived carbon transformed throughout the incubations (Figure 1). Generally, losses in total treatment carbon from our incubation treatments were primarily losses of residue-derived carbon with the amount of residue carbon lost being sometimes as great as 291 times larger than soil-derived carbon losses for a given treatment (Appendix I and J). For every comparable incubation-condition group, corn stover lost the most residue-derived carbon. Our statistical analyses of the carbon remaining showed that the effect of substrate type on differences in remaining residue-derived carbon were always significant (two-way ANOVA, Incubation one substrate effect, $df = 2$, $p << 0.05$; Incubation two substrate effect, $df = 3$, $p << 0.05$) but not always for differences in soil-derived carbon (Appendix K and L).

For Palouse soils in Incubation one, substrate effect was significant on soil-derived carbon losses between CS1 and both AD1 and HLFB1 respectively (two-way ANOVA, treatment effect, $df = 2$, $p << .05$ with CS1 = 67 mg C, AD1 = 25 mg C, and HLFB1 = 34 mg C). For Vershire soils in Incubation one, substrate effect was not significant save for the comparison between CS1 and AD1 (two-way ANOVA, $df = 2$, $p = .66$ with CS1 = 60 mg C, AD1 = 42 mg C, and HLFB1 = 169 mg C). For Incubation two standard dosage treatments, substrate effect was not significant on the group's soil-derived carbon losses (two-way ANOVA, treatment effect, $df = 3$, $p = 0.06$). For Incubation two reduced dosage treatments, substrate effect on soil-derived losses were significant (two-way

ANOVA, treatment effect, df = 3, p << 0.05 with CS2 = mg C, AD2 = mg C, HLFB2 = mg C, HLFB3 = mg C).

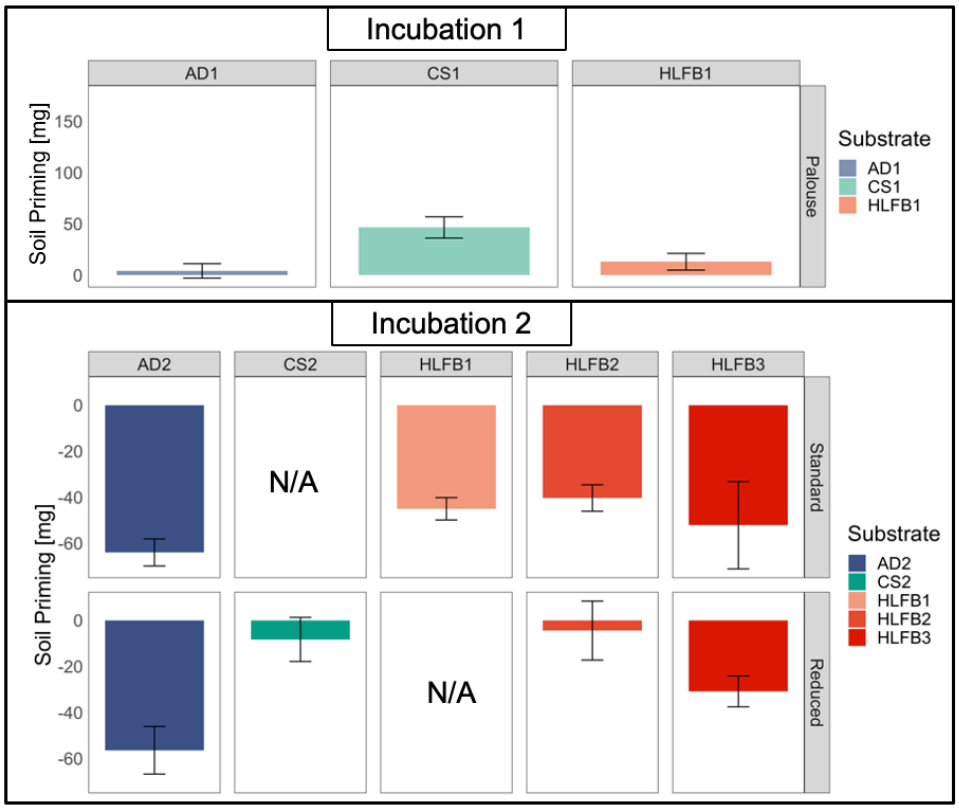


Figure 2: Soil priming across the incubations. Soil priming is the difference between the loss of soil-derived carbon from a substrate containing treatment and the loss of soil-derived carbon from a comparable soil control. Both positive and negative soil priming can occur. Boxed figures show results for different Incubations and conditions.

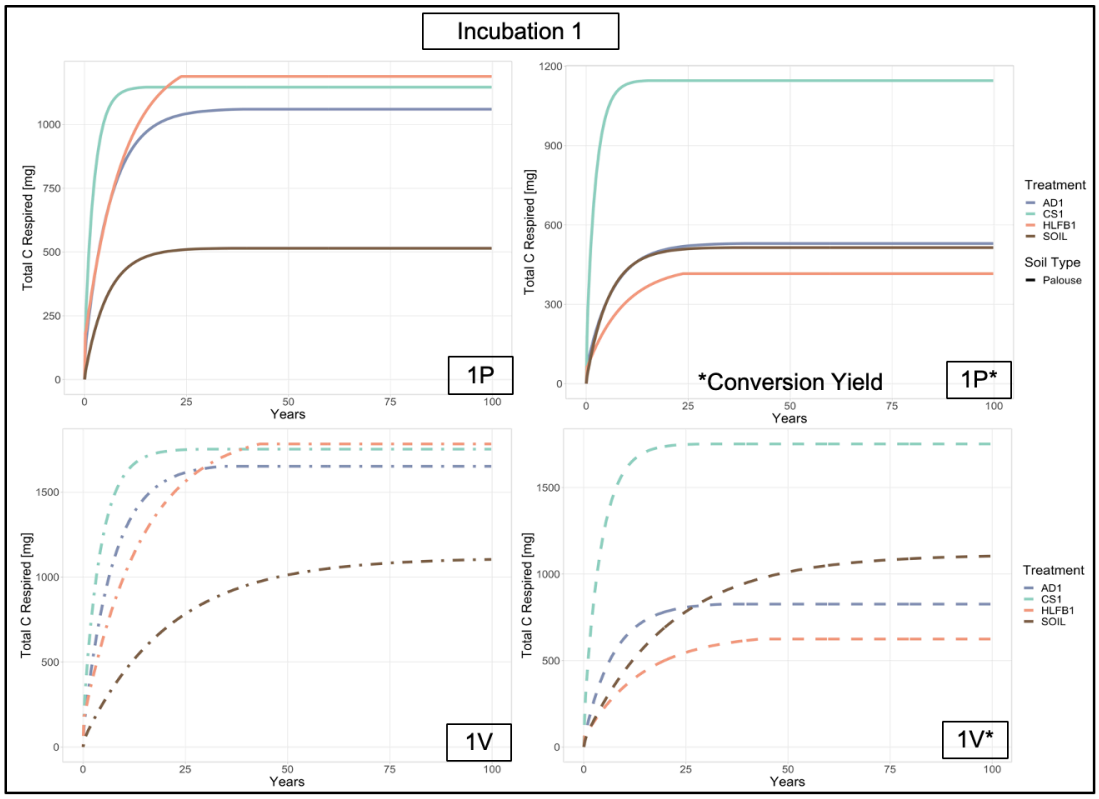
Drawing from the soil-derived carbon loss data, we found the effect of soil priming for our treatments to be much smaller than total soil-derived carbon losses relative to residue-derived carbon losses (Figure 2). Soil priming is defined in this study as the difference between soil-derived carbon loss from a substrate containing treatment and

soil-derived carbon loss from a comparable soil control pre and post incubation. Notably, there can be both positive and negative soil priming with positive priming meaning carbon loss of soil is stimulated by the addition of organic matter input while negative priming meaning that the addition of residue reduced the decomposition of soil carbon. Incubation one treatments only exhibited positive priming while Incubation two treatments only exhibited negative priming perhaps due to the very different timespans the incubations occurred in. Due to both measurement sensitivity errors using the IRMS when measuring the Vershire soil controls and the inherent error introduced by using the Vershire soil, a soil that was not rigorously controlled for foreign 13C sources, priming for the Vershire treatments could not be quantified. Incubation two treatments which only contained Palouse soils varied in magnitudes of priming across the dosages except for AD2 treatments, which caused similar amounts of priming (-56 mg C for Reduced and -64 mg C for Standard). Because of this study's definition of priming as the difference between the soil-derived carbon loss of a residue containing treatment and a soil control, the statistical differences within incubation-condition groups i.e. Incubation one Palouse treatments or Incubation two standard dosage treatments, are the same as compared to the statistical differences of the other soil-derived carbon losses (Appendix L).

Modelling One-Time Inputs and Steady State SOC Scenarios of Byproduct Return

To extrapolate beyond the timescales of our incubation, we fit multi-pool models to the carbon respired data from our incubations. Essentially, we extended the lengths of our incubations indefinitely and compared both the time elapsed for the treatments to respire fully away and the magnitude of their respective carbon loss. Because we applied the

two-pool series structure to each residue containing treatment, we could compare carbon fates by treatment within incubation-condition groups. Additionally, we compared carbon fates by treatment after accounting for bioconversion losses (i.e., carbon fates of reduced mass inputs of anaerobic digestate and HLFBs as compared to corn stover). We found that in every incubation-condition group containing corn stover, the corn stover treatment respired away the quickest (Figure 3).



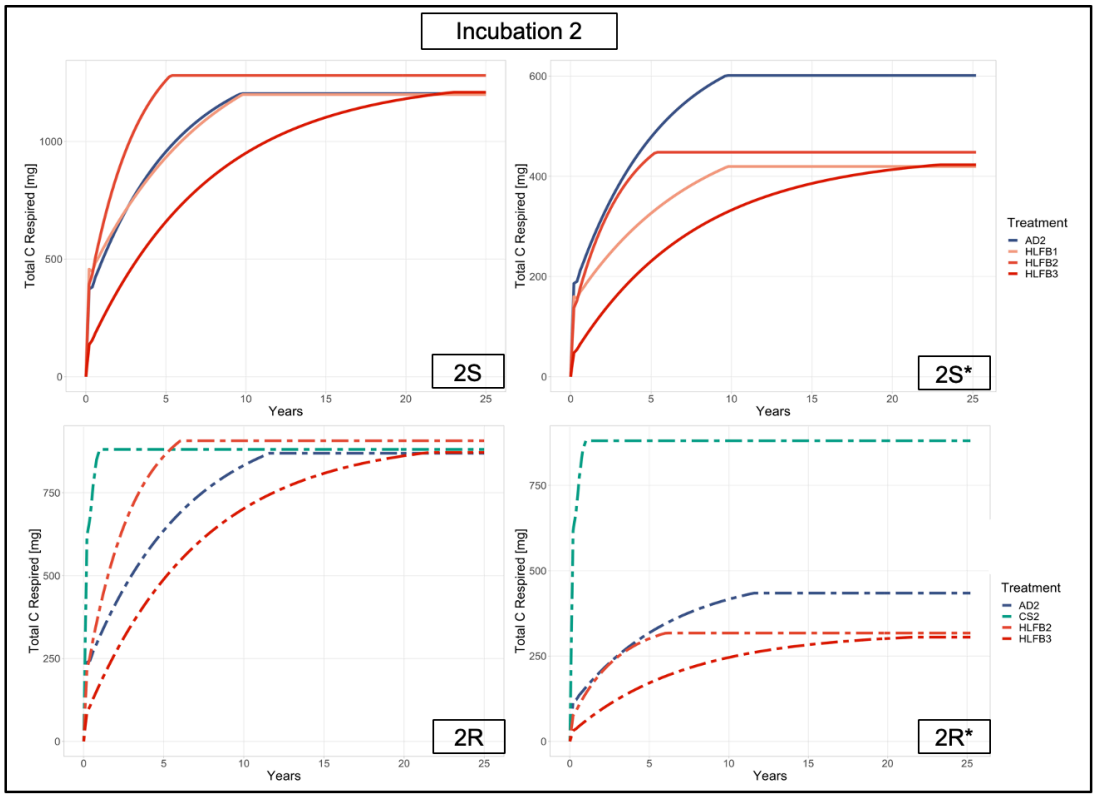


Figure 3: One-time input modelling of various treatments in different incubation-condition groups. The graphs represent projections of our incubation data, essentially showing how carbon is respired indefinitely. P and V refer to the differing soil types of Palouse and Vershire respectively. S and R refer to the differing dosages of residue to soil, termed standard and reduced respectively. Asterisk labelled graphs represent results from adjusted mass inputs reflecting bioconversion mass yields of 50% for anaerobic digestion and 35% for HLFBs. The y-axis represents the amount of carbon respired as a function of time. The same model structure was applied to each treatment thus allowing for careful comparison within incubation-condition groups. The graphs reflect projections with timesteps of .20 years and manual corrections for any model overshooting as shown by the sharp transition to plateaus among the Incubation two graphs.

For both Incubations one and two, corn stover treatments showed an immediate release in carbon much faster than the comparable bioprocessed residue containing treatments. For Palouse soils in Incubation one with and without bioconversion accounted for, the corn stover treatment released 90% of its total carbon by year 5 as opposed to AD1, which reached this state at 14.2 years and HLFB1 at 16 years (Figure 3, 1P and 1P*). For Vershire soils in Incubation one with and without bioconversion, the trends were similar with CS1 reaching 90% release by 9.6 years, AD1 by 16 years, and HLFB1 by 27.2 years (Figure 3, 1V and 1V*. For Incubation 2 with reduced dosages, CS2 was the quickest to achieve 90% release by far at .80 years, HLFB2 at 4.4 years, AD2 at 8.4 years, and HLFB3 at 13.6 years (Figure 3, 2R and 2R*). Finally, for Incubation two with standard dosages where there was no CS treatment to compare to, comparisons amongst the three HLFBs showed HLFB2 achieving 90% release by year 4, AD2 by year 7, HLFB1 by year 7.4, and HLFB3 by year 14.4 (Figure 3, 2S and 2S*). Once conversion yields (100% for corn stover, 50% for anaerobic digestate, and 35% for HLFB) were considered in our projections, we found that the magnitude of carbon respired by the corn stover treatments relative to the other treatments was especially pronounced (see Figure 3 plots 1P*, 1V*, and 2R*). Since the two-pool series model predicts that the total initial carbon is eventually completely respired, the greater magnitude corn stover carbon loss is not unexpected. However, due to its exceptionally high rates of carbon release, this modeling shows that corn stover does not retain carbon in the same way as its biologically processed counterparts.

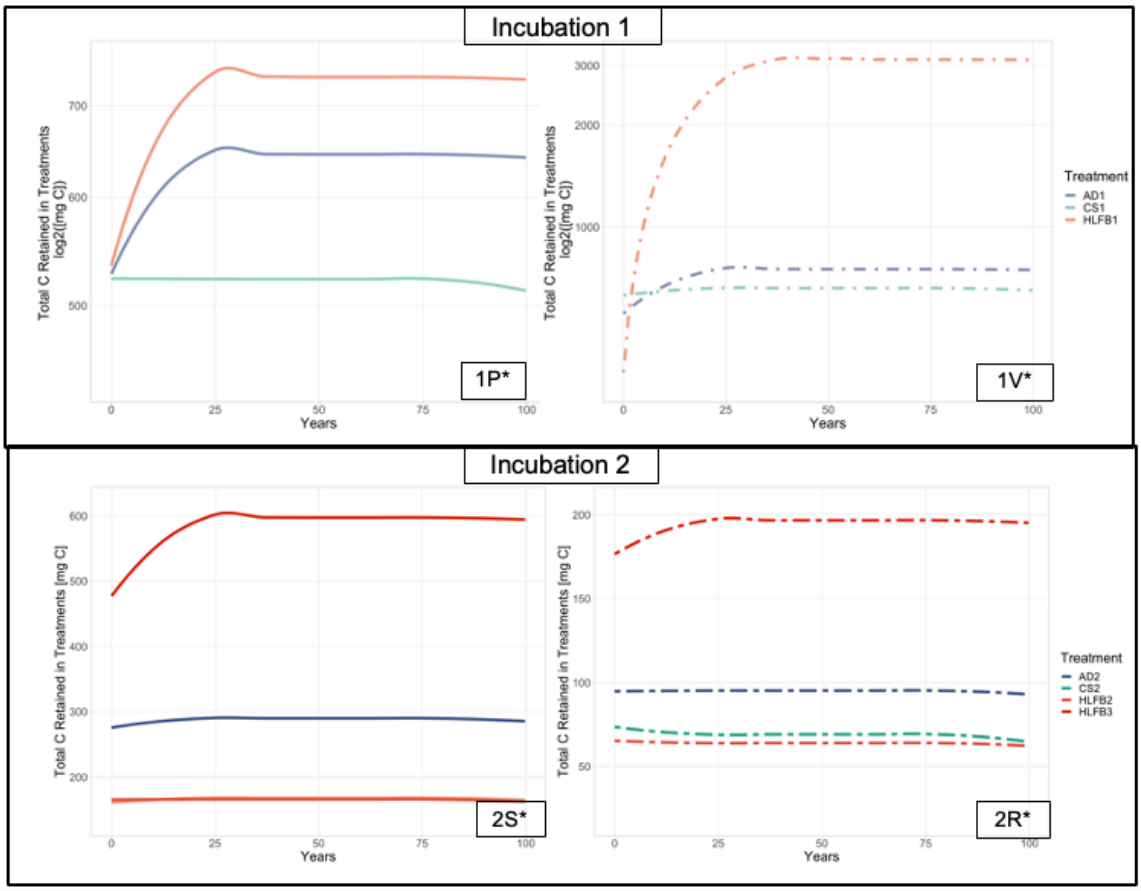


*Figure 4: Steady state levels of carbon in treatment from annual inputs. Organic inputs, equal to the Figure 3 * plots (i.e., bioconversion yield adjusted mass inputs) were modelled to be annually readded to the treatments. These annual input graphs represent a simple, bioenergy cropping scenario in which continuous carbon accumulation and respiration reach differing steady state levels that can be compared amongst substrates. Additionally, instead of carbon released, the y-axis shows the amount of carbon retained by the treatments with higher curves indicating more carbon retained within the treatment. To reduce the noisiness of the graph, geom_smooth from the ggplot2 package was used to plot these values[79]. Like the Figure 3 naming scheme, Incubation one graphs differ by soil type with 1P* = Palouse and 1V* = Vershire soil types respectively.*

Incubation two graphs differ by dosage with 2S = Standard dosage while 2R* = Reduced dosage.*

Using our one-time input modelling data, we calculated the steady state carbon levels in our treatments in 100-year time scales imitating a bioenergy cropping scenario in which organic matter is added to the soil on a yearly basis. From our steady state modelling results, we found that all the biologically processed residues tested except HLFB2 formed more steady state carbon than corn stover in every comparable incubation-condition group. Because the one-time input models of corn stover projected especially fast releases of carbon, despite higher amounts of initial carbon, at steady state, corn stover did not accumulate as much carbon as other materials. While bioprocessed materials added less initial carbon to the soil relative to corn stover, when conversion yields were accounted for, their slower decay allowed for higher soil carbon accumulation.

Group	Treatment	Steady State Carbon (mg)	SS_{res} / SS_{CS}
INC1, Palouse, Standard	CS1	513	1.0
	AD1	642	1.3
	HLFB1	732	1.4
INC1, Vershire, Standard	CS1	651	1.0
	AD1	747	1.1
	HLFB1	3151	4.8
INC2, Palouse, Reduced	CS2	65	1.0
	AD2	93	1.4
	HLFB2	62	.95
	HLFB3	195	3.0

Table 5: Ratios of steady state carbon formation of biologically processed residues to corn stover while considering reduced mass inputs of residues. Steady state carbon values were extracted from Figure 2 at time approaching 100 years.

By comparing the ratios of steady state carbon formed from processed residue input to unprocessed input material, we found that the processed residues formed more steady state carbon than corn stover in every comparison except HLFB2 (Table 5). For the other materials, we found that this ratio ranged from 1.1 to 1.3 for AD1, 1.4 to 4.8 for HLFB1, and were 1.4 for AD2 and 3.0 for HLFB3 respectively (Table 5). HLFB2 was estimated to form approximately 95% of the steady state carbon formed by corn stover due to both its rate of decomposition being the closest to corn stover (4.4 years for HFLB2 vs 0.80 years for CS2, Figure 4) combined with its relatively smaller carbon input after bioconversion. Pre-bioconversion adjustment, HLFB2 contained slightly more carbon than corn stover, (373 mg carbon for HLFB2 vs. 343 mg carbon for CS2, Table 2). However, by adjusting for the requisite 35% conversion yield, this effect was negated.

Discussion

Our incubation results join Johnson et al. (2007) as the second empirical soil experiments ever conducted with HLFBs[60]. While we both performed controlled, laboratory-scale, multi-day soil incubation experiments with HLFB, the scope and focus of our study differed. Conducting our study over 15 years after Johnson et al. allowed us access to higher quality HLFBs (i.e., HLFBs that had more rigorous documentation on their preparation and more reflective of the current state of the art of lignocellulosic biofuel production). Consequently, we tested three HLFBs as opposed to one enabling us to derive some generalized relationships between material characteristics of HLFB and carbon retention metrics (Figure 5). Since our focus was to specifically compare SOC

formation potential between HLFB and unprocessed feedstock, our efforts went towards quantifying and partitioning the carbon retained in our treatments whereas Johnson et al. focused on quantifying and qualifying the various soil quality metrics that changed in their treatments. Though we did not explicitly measure their metrics of interests, which included changes in bulk density, water retention characteristics, humic acid concentration, and water-stable aggregates percentage; we offer that the higher SOC formation potential we identified (relative to corn stover) for the HLFB's is strongly and positively associated with these soil quality metrics[4,6,80]. While we cannot make a direct comparison to Johnson et al.'s carbon released results considering we tested different incubation conditions and application rates, as a rudimentary comparison, we can look at the ratio of carbon respired by the corn stover to carbon respired by the HLFB for both their data and ours. For Johnson et al.'s 112 day incubation where they incubated both HLFB and corn stover at the same application rate of 1.0 kg material per m^2 soil, the ratio of carbon respired by corn stover to HLFB was 1.43-1.54. For the most apt comparison, we can look to the most similar HLFB we tested, HLFB2 (C:N = 39 and lignin:N = 43), which was most similar to the Johnson et al. HLFB (C:N = 30 and lignin:N = 30) and was also produced at NREL. Taking our HLFB2 incubation results at day 112 resulted in a ratio of carbon respired by corn stover to HLFB equaled 1.46, within the range of Johnson et al. (Appendix N). Overall, our results strengthen Johnson et al.'s conclusions. Applying HLFB as a land amendment to soil may increase SOC and enhance positive soil qualities relative to a base case of leaving corn stover on the field.

This potential is explored in depth in a soon to be published manuscript by authors

affiliated with this study[81]. This manuscript, to be published in 2023 as a literature review about HLFB return, defines a new and specific metric for steady state SOC formation potential (Appendix O). Specifically of interest is the quantity $Y^c\varepsilon$ where Y^c represents the conversion yield of a bioprocessed residue (i.e., 0.35 for HLFB) and ε represents the relative efficiency of steady-state SOC formation per standardized unit of input from soil-applied organic matter for the No-Harvest and Harvest with HLFB Return cases. Applying the results of this thesis to this framework, we find that our SS_{res} / SS_{CS} parameter (Table 5) effectively equals $Y^c\varepsilon$. Since we calculated SS_{res} / SS_{CS} from annual-input modelling scenarios that factor in bioconversion yields, the case of No-Harvest is represented by SS_{CS} and Harvest with HLFB Return is represented by SS_{res}. Taken together, Table 5 offers multiple values for $Y^c\varepsilon$ given varying soil types and conditions. We hope that further research can be conducted in this intersection of new theory and experimental data.

Priming, while observed in our experiments, is not satisfyingly and conclusively quantified in this study. Between Incubation one and two, priming in our incubations does not appear to follow any consistent trends beyond being positive for Incubation one treatments and negative for Incubation two treatments. However, even this difference is not conclusive as AD1, CS2, and HLFB2 are within one standard deviation of being either positive or negative. Additionally, a key difference between the two treatments is the length of the incubations with Incubation two being almost twice as long as Incubation one (267 days versus 135 days). This temporal difference could explain the apparent difference in priming effect between the incubations[49]. We offer both a process-

based and mechanistic explanation. Process wise, since we calculated priming by subtracting out the carbon respired from the soil control at the end of each incubation, Incubation one may have exhibited positive priming only because of the averaging effect inherent in a longer incubation. The longer the incubation, the less weight the initial short-term acceleration or retardation in microbial activity caused by priming has on the overall amount of carbon respired as the residue containing treatments released much larger magnitudes of carbon in comparison to the soil controls. The opposite effect could then be applied to Incubation two. Since Incubation two was half as short as Incubation one, the effect of soil priming could have been more pronounced and apparent especially from our simple calculation. Mechanistically, Incubation's two negative priming can also be explained by microbes experiencing preferential substrate utilization in which microbes switch from consuming poorly degradable SOM to more easily decomposable organic matter input, slowing their decomposition of the original SOM[51]. Over time, as the new input is consumed, the larger population of microbes may switch back to the original SOM or start mining the original organic matter for nutrients. Overall, while we have approximated the priming effect for our treatments, we are unable to conclusively comment on how exactly priming may work with other HLFBs.

Relative to low lignin organic matter inputs, high lignin inputs have been shown to be associated with less soil-derived carbon loss and more residue-derived loss[50]. This is supported by our results in both Incubations one and two where positive priming is greatest and negative priming is smallest for the low lignin corn stover, respectively. In both cases, corn stover exhibits more soil-derived carbon loss than the comparable

bioprocessed materials. Higher amounts of positive priming / lower amounts of lower negative priming for corn stover support the hypothesis that bioprocessed materials retain more carbon than their unprocessed counterparts overall. In future research, we would hope to more certainly ascertain the effect of priming with more rigorous isotopic analysis throughout the incubation and not just at the beginning and end.

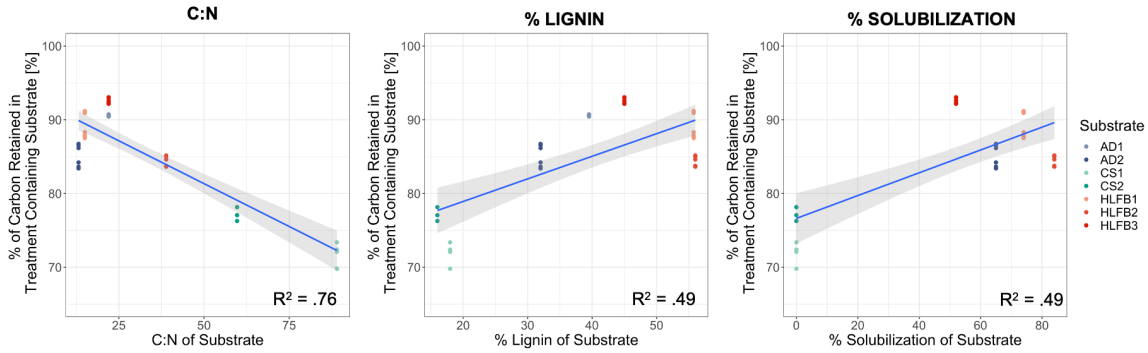


Figure 5: Correlations of residue C:N, percentage lignin, and percentage solubilization versus percentage of total carbon respired in each treatment from our incubations. Decreased C:N, increased lignin content, and increased solubilization are analogous to increased amounts of bioprocessing.

As an outcome of the biofuel production process, bioprocessed materials are essentially pre-decomposed relative to their input in the soil. Thus, upon addition to soil, bioprocessed materials tend to decompose at a slower rate than fresh residues or unprocessed material and may possibly retain more carbon by mass percentage. This is true especially in the beginning stages of decomposition as shown by our one-time input projections whereupon corn stover released its carbon the quickest in every comparable incubation-condition group. Decreased C:N, increased lignin percentage, and increased solubilization rates are analogous to increased amounts of bioprocessing or effectively,

pre-decomposition. Plotting these metrics relative to percent carbon retained, we found that our hypothesis that increased amounts of pre-decomposition may correlate with increased amounts of carbon retained (Figure 5). However, we acknowledge that our correlations, especially for the solubilization graph, may be dominated by the cluster of corn stover data and thus, is not a definitive correlative relationship. We include these preliminary correlations not for their conclusive, inherent value, but because we believe this data may inform future research diretions on SOC formation of HLFBs as future HLFBs will vary in these plotted metrics and may be contextualized by the generalized relationships presented here.

Conclusions and Future Work

Supported by multiple lines of evidence – incubation data, short-and-long term modelling, and partitioning data – at a minimum, our results indicate that returning the same amount by mass of biologically processed material to soil leads to increased amounts of carbon retained in soil as compared to unprocessed material. Taken further, we estimate that returning even a reduced amount of mass reflective of bioconversion yields would still lead at least similar amounts of carbon retained in soil as compared to full amounts of unprocessed material. We find that in all but one incubation-condition group, the digestate and HLFBs respire less carbon, store more carbon, and persist longer in the soil.

		Soil Incubation	Isotope Analysis	One-Time Input Modelling		Steady State Modelling
Incubation Condition	Substrate	%C Retained (end of incubation)	%C Retained of Residue C (end of incubation)	Slowest to Respire Away Completely	Slowest to Respire Away Completely (with conversion)	Highest Steady State SOC
1 (Palouse Soil, Standard Dosage)	CS1	**3**	**3**	**3**	**3**	**3**
	AD1	2	1	2	1	2
	HLFB1	1	2	1	2	1
1 (Vershire Soil, Standard Dosage)	CS1	**3**	**3**	**3**	**3**	**3**
	AD1	2	1	2	2	2
	HLFB1	1	2	1	1	1
2 (Palouse Soil, Reduced Dosage)	CS2	**4**	**4**	**4**	**4**	3
	AD2	2	3	2	2	2
	HLFB2	3	2	3	3	**4**
	HLFB3	1	1	1	1	1

Table 6: Summary table of the carbon retention properties found from our study's tests of various substrate containing treatments relative to incubation-condition groups. 1 indicates that the treatment retained the most carbon relevant to the corresponding test while 4 indicates the least carbon relative. The least retentive material is bolded in each comparison.

From our modelling efforts, we see that the rankings of the unprocessed material relative to the processed material remain relatively unchanging despite any magnitude differences. From our partitioning work, we conclude that residue-derived carbon is both more easily accessible to soil microbes in the form of fermentable sugars and in relative abundance compared to soil derived carbon. We estimate that priming as an effect on general SOC levels is relatively minimal. Overall, we attribute these differences in SOC

levels both experimentally determined and modelled be due to the unprocessed nature of corn stover versus the biologically processed residues and not as an outcome influenced by our various incubation conditions. Thus, through our experimental and theoretical study of HLFB's decomposition in soil, we find that returning HLFB to soil increases SOC relative to a non-harvest case as represented by our various projections of unprocessed material decomposition. Placed in the broader context of 2G biofuel production, we assert that HLFB return will enable higher rates of residue harvest and increased production of 2G biofuels using crop residues, which may resolve the land use and food vs fuel ambiguity surrounding the sustainability of 2G biofuels.

In terms of future work emerging from this thesis, we are primarily interested in modelling field-scale SOC scenarios in which biologically processed organic matter is continuously added to an agricultural field. Specifically, we are interested in a more conclusive estimation of steady-state SOC in a mature bioenergy cropping scenario. This work is currently in its beginning stages as initiated by Professor Jo Smith at the University of Aberdeen. She endeavors to apply the RothC model of soil carbon dynamics to the isotopic analysis results we discussed in this thesis. Additionally, we are interested in longer and more rigorously controlled soil incubations with even more diverse HLFBs. Specifically, it would be of great interest to test the state-of-the-art HLFBs NREL is producing currently aligned with their most recent process models utilizing alkali-pretreatment. In these future incubations, careful documentation of the various HLFB production processes would be of the utmost importance focusing specifically on the metrics of percent general solubilization, percent carbohydrate

solubilization, general mass yield, and mass yield on a carbon basis. It would also be important to include an appropriate corn stover treatment in each subsequent incubation that would occur for rigorous comparison. Ultimately, in terms of the most important future work still needed on this topic, we are most interested in seeing field-scale experiments observing steady-state SOC where HLFB is returned to agricultural soils that grow the HLFB feedstock over the multi-decade long scale. Though the state-of-the-art of 2G biofuel production may make gathering the amount of HLFB needed to conduct this investigation near impossible, we believe this investigation would be worth the considerable effort and cost. Conclusive results from this sort of field-scale studies of HLFB return are what could turn a current compelling theory into a real agricultural practice.

Appendices

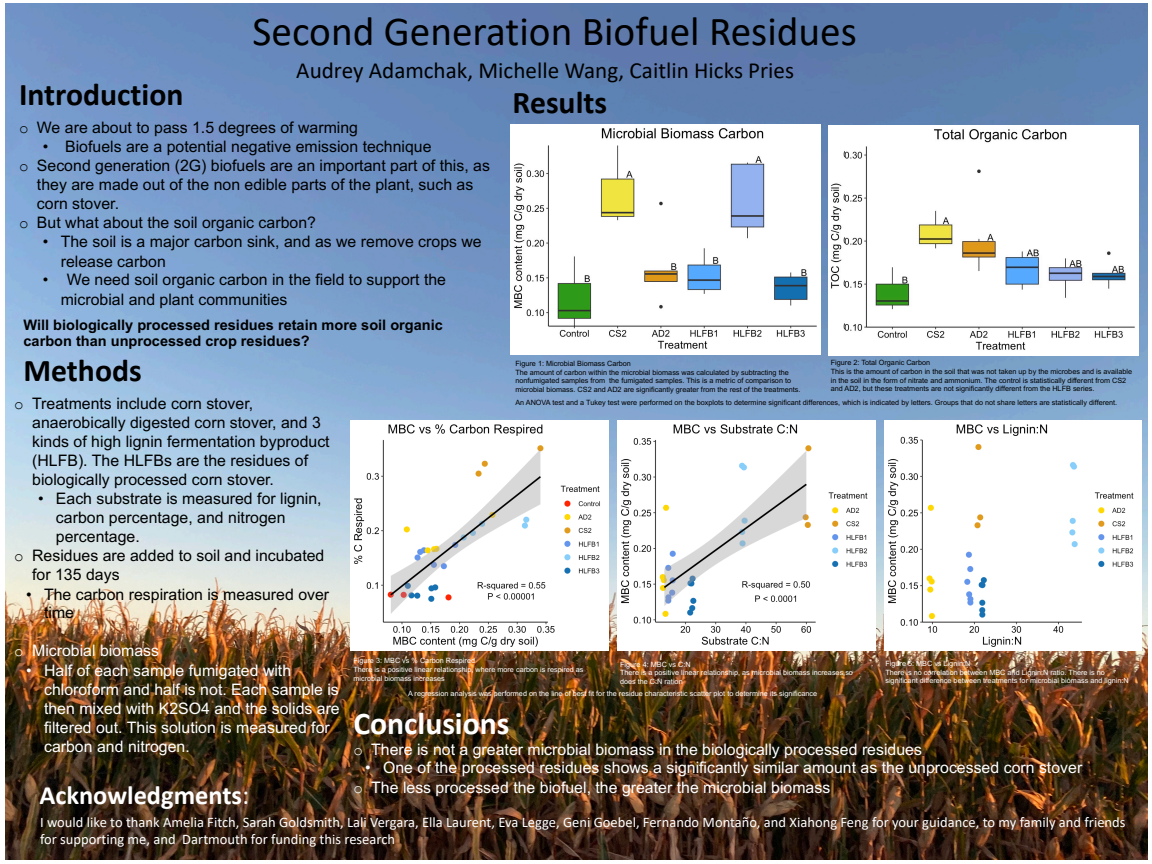
Appendix A – Crop Residue Removal Literature

Literature from 1986-2019 on the topic of residue removal was reviewed. Specifically, literature with enumerated values for crop residue harvest was included.

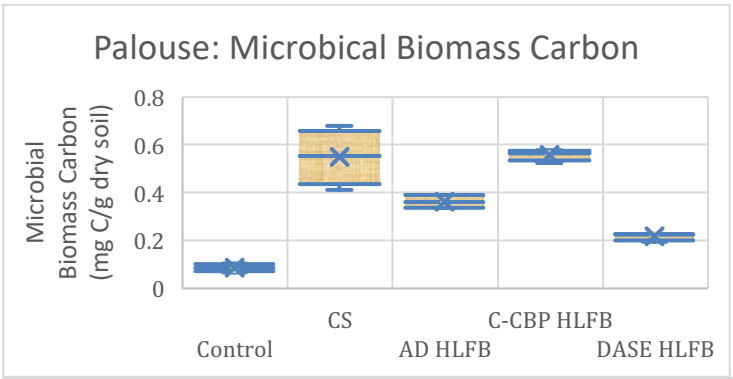
Source	How much crop residue is safe to harvest from a SOC and soil quality perspective?	Soil Context	Complexities
Lindstrom, 1986	.5*USLE, USLE, 2*USLE with the USLE amount being the results of the Universal Soil Loss Equation applied	4 year field study with corn stover on no-till and tilled fields of loam and silty loam in the northwestern Corn Belt	USLE results are site specific. No universal percentage given, however a soil loss tolerance level T is given as 11.2 tons/ha/year.
Blanco-Canqui & Lal, 2009	25% might be available for removal.	4 year field study with corn stover no-till silt loams and clay loam in 3 different fields in Ohio.	Stover removal has the most adverse impacts on sloping and erosion prone soil.
Blum et al., 2010	0<x<50% depending on crop type, soil properties, and climate.	RothC-26.3 model for 40 years, in different soil types across Europe.	For corn, <50% showed increasing SOC stocks but for barley, decreasing for all, but for more root biomass, incr. SOC always.
Karlen et al., 2014	Some level of corn stover harvest may actually be good for productivity.	Meta analysis of 239 site years of field research.	No-till grain yields were significantly lower than with conventional tillage when stover was not removed, but equivalent when it was harvested. Presumably stover harvest helped mitigate many traditional residue management problems such as N immobilization and reduced soil temperatures.
Jin et al., 2015	55% assuming N fertilizer additions. SOC gains limited compared to no-harvest.	12 years, no-till continuous corn system in silt loam in W Corn Belt.	Crop yields and SOC remained equal, but soil stability and erosion protection decrease.
Kenney et al., 2013	<50% assuming 15 cm of stalk left in field on all plots.	3 years, no-till continual corn system in silt loam in Kansas.	>50% incr. Risk of erosion and soil water coupled w/ marginal short term increase grain yield.
Xu et al., 2019	30-40% could minimize adverse impacts of stover removal on SOC.	Meta analysis of 409 global data points.	Stover removal generally reduced SOC stock by 8% in 0-30 cm profile; depth matters and few deep data points.
Gollany et al., 2020	0% only one to incr. SOC, tested 50/100% in till/no-till but all depleted.	10 year field study, then 30 years w/ CQESTR, no-till continual corn in silt loam in W MN.	No till and tillage tested, no till only way for SOC to incr. Work only done 0-30 cm of soil.

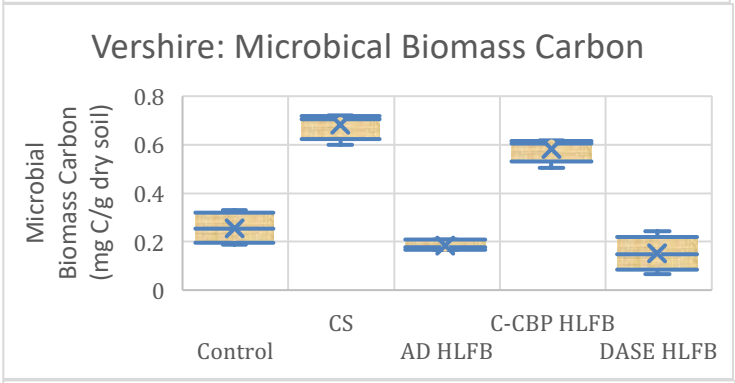
Appendix B – Microbial Biomass Carbon Results

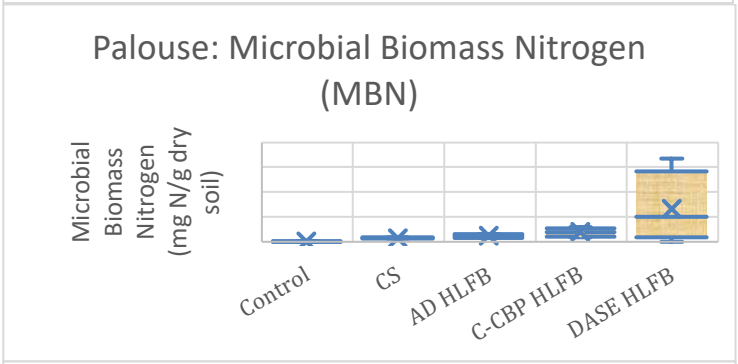
Microbial biomass carbon and nitrogen results from Incubation two. Incubation two poster created by Audrey Adamchak for a senior capstone project in 2023.

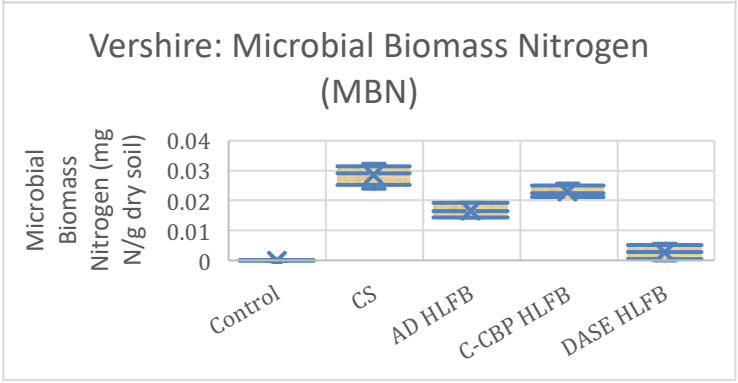


Microbial biomass carbon and nitrogen results from Incubation one.



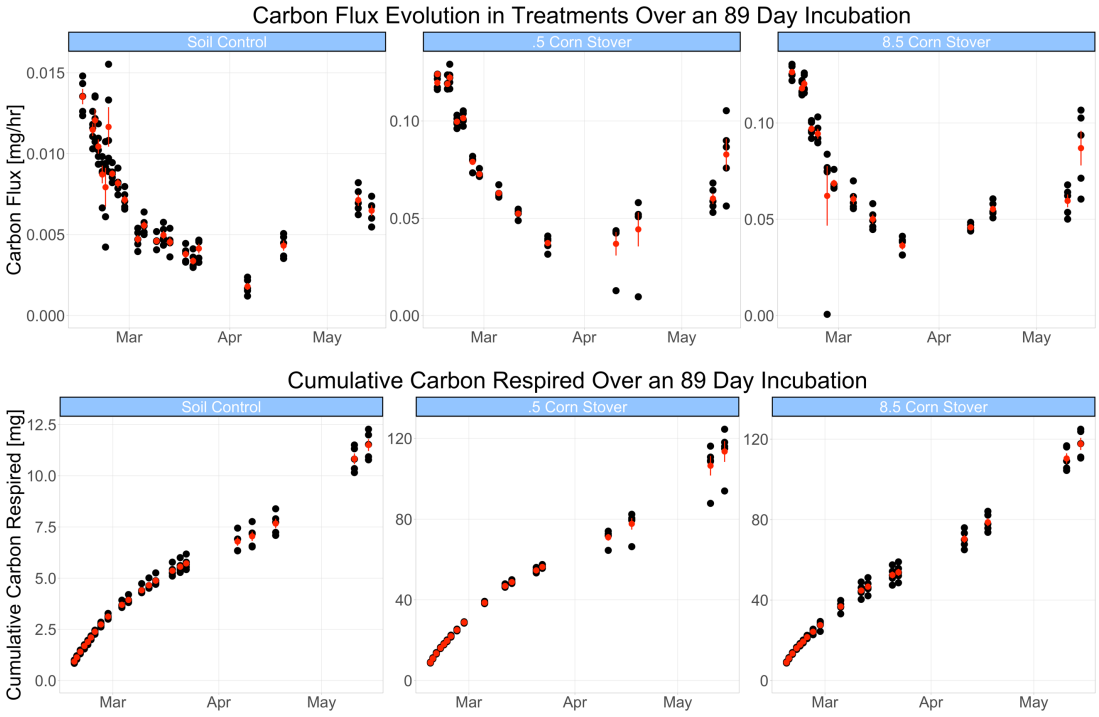






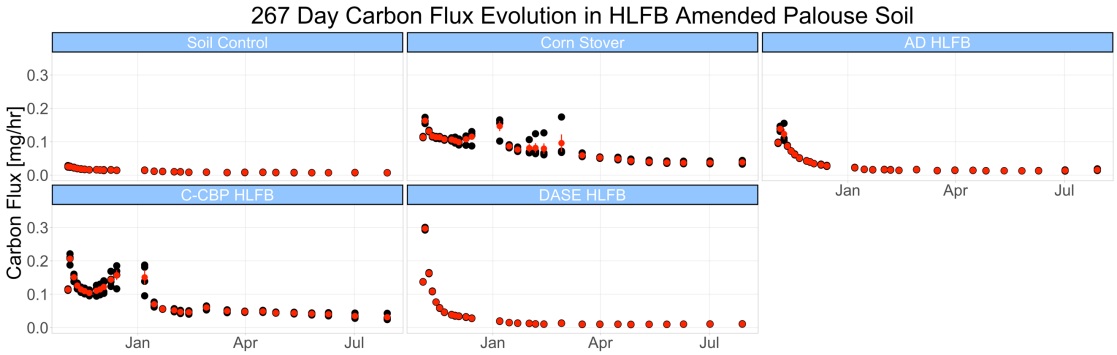
Appendix C – Pre-study Particle Size Incubation Results

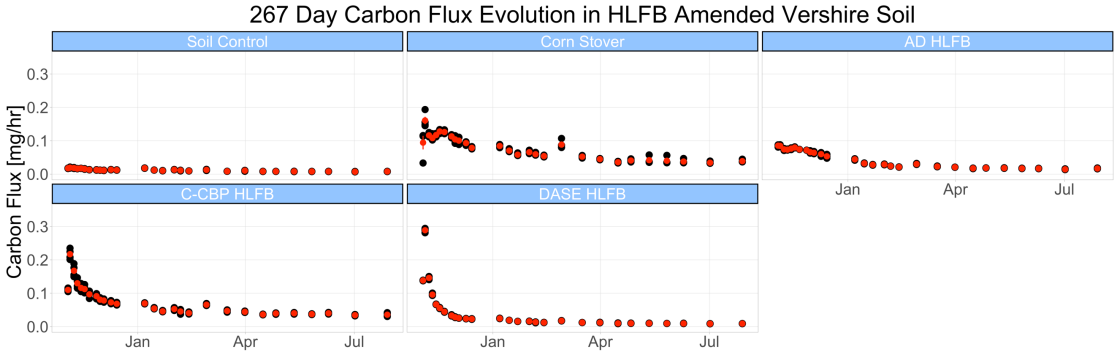
Results from a soil incubation with differently sized corn stover particles (0.5 mm diameter size and 8.5 mm diameter size) is shown. Under similar incubation conditions to this study's Incubations one and two, carbon flux and cumulative amount of carbon respired were statistically insignificant between the different particle size treatments.

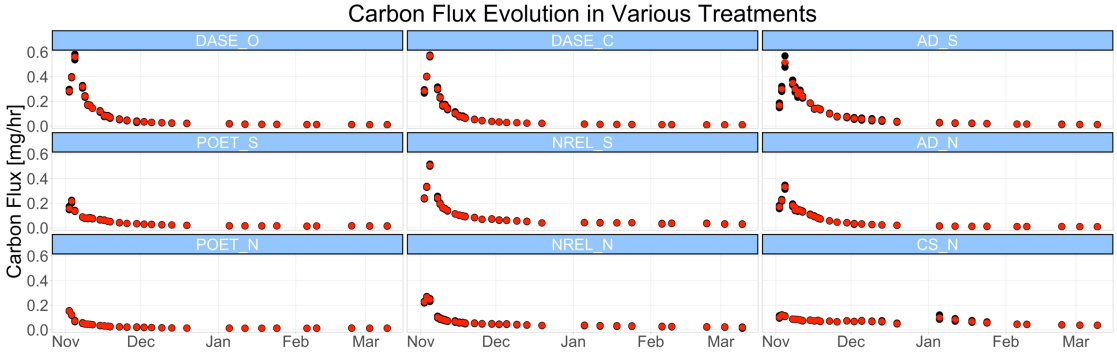


Appendix D – Carbon Flux and Carbon Respired Results

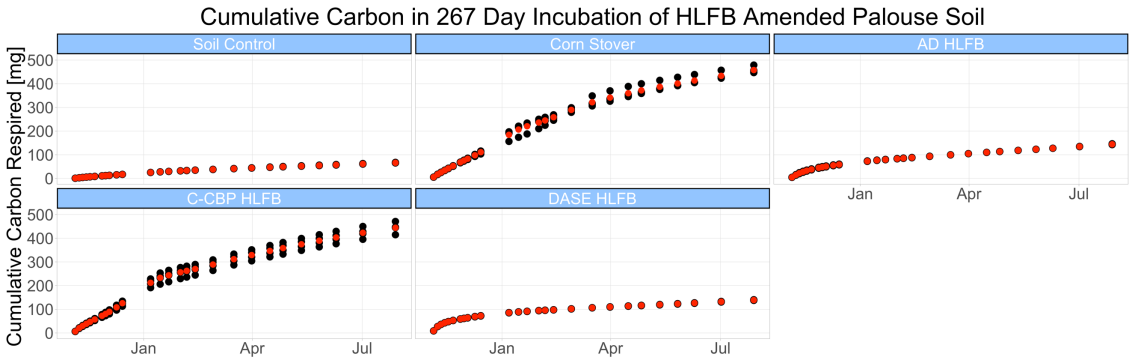
Carbon flux graphs below show exponential decay of C production over time of incubations. 267-day data correspond with Incubation one and 135-day data correspond with Incubation two shown in the final set of graphs on the bottom of the page.

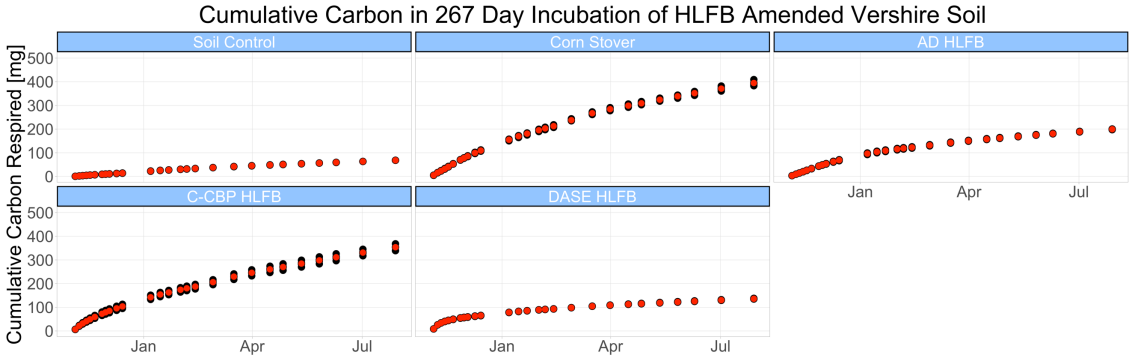


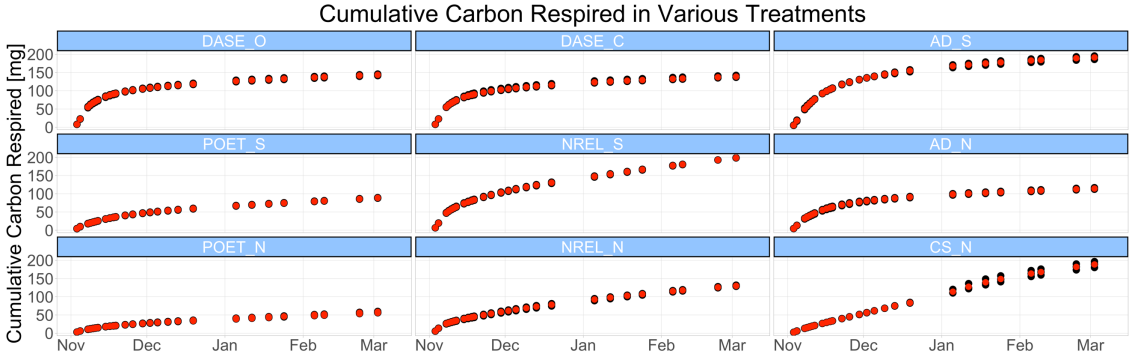




Cumulative carbon respired graphs below show the cumulative production of C released as CO_2 respired during the time of incubations. 267-day data correspond with Incubation one and 135-day data correspond with Incubation two shown in the last set of graphs on the bottom of the page.







Appendix E – Best Fit Parameters for Models

Best fit model parameters for various models from the SoilR Package when applied to our incubation data are shown. Gamma, across these tables, represents the initial partitioning between the two pools of carbon. As an example, for the Palouse Soil control 2 Pool Parallel scenario, gamma indicates that pool 1 contains.205 of the initial total amount of carbon in the system.

	PALOUSE Treatments Model Parameters							
	1 Pool	2 Pool Parallel			2 Pool Series			
	k	k1	k2	gamma	k1	k2	alpha21	gamma
Soil Control	0.00060213	0.00385251	3.6207E-09	0.2050224	0.02996006	0.00049075	0.96291063	0.83243589
CS1	0.00186762	0.00626492	9.7936E-10	0.41454358	0.01736454	0.00115017	0.14272221	0.13605892
AD1	0.00098838	0.0096735	1.6263E-08	0.19355018	0.02023506	0.00042587	0.88550346	0.96145974
HLFB1	0.00061527	0.02124474	1.6444E-08	0.09970191	0.10104426	0.00029263	0.8931433	0.46704678

	VERSHIRE Treatments Model Parameters							
	1 Pool	2 Pool Parallel			2 Pool Series			
	k	k1	k2	gamma	k1	k2	alpha21	gamma
Soil Control	0.00026917	0.00385151	1.9114E-10	0.09474933	0.00801508	0.00012956	0.94842326	0.89749085
CS1	0.00113867	0.01384815	0.00057854	0.10019409	0.01598043	0.00063769	0.85398893	0.77818785
AD1	0.00060948	0.00967052	9.4902E-09	0.12416559	0.0411658	0.00037797	0.94357717	0.74899571
HLFB1	0.0004004	0.02123931	1.8425E-08	0.06643432	0.0994277	0.00018848	0.92233811	0.43064804

	1 Pool	2 Pool Parallel			2 Pool Series			
	k	k1	k2	gamma	k1	k2	alpha21	gamma
AD2_S	0.00189853	0.05511563	6.4441E-09	0.15092216	0.09407008	0.00050484	0.29149503	0.15376928
HLFB3_S	0.00075505	0.03121938	9.0043E-09	0.07085624	0.09723211	0.00035762	0.92835663	0.4862196
HLFB2_S	0.00167767	0.03259927	9.5211E-09	0.14763444	0.14805689	0.00088919	0.88689962	0.57334913
AD2_R	0.00152579	0.05737624	1.1037E-08	0.12295175	0.10553739	0.00043866	0.88749601	0.78917944
HLFB3_R	0.00062784	0.02082809	1.0694E-08	0.06901371	0.12593147	0.00038178	0.94835083	0.45548201
HLFB2_R	0.00148818	0.02257428	9.6185E-09	0.1471805	0.14959621	0.00094104	0.91237519	0.53117372
CS2_R	0.00211539	0.00589357	2.9187E-08	0.42621541	2.98690222	0.0019992	0.9205925	0.08886055
SOIL	0.00062089	0.00844859	3.6825E-09	0.1049325	0.04030751	0.0004729	0.95950953	0.47706739
HLFB1_S	0.00138923	0.08706575	1.0785E-08	0.10778758	0.16432275	0.00036241	0.88995359	0.74335243

	INCUBATION 2 R^2 VALUES			
	1P	2PS	2PP	3PP
HLFB1_S	0.72099169	0.99382974	0.9390355	0.99685328
AD2_S	0.78902026	0.99786028	0.97985758	0.97159659
HLFB3_S	0.92355909	0.99763126	0.97311478	0.99721176
HLFB2_S	0.93239192	0.99730001	0.9570505	0.95105699
AD_R	0.80520861	0.99745984	0.97185829	0.96289499
HLFB3_R	0.96648872	0.99607811	0.96974784	0.994541
HLFB2_R	0.96780066	0.99515265	0.97626642	0.97909025
CS2_R	0.9959621	0.9953147	0.9980653	0.99777375
SOIL	0.98402199	0.99639293	0.99950116	0.99904491

Appendix F – Example R Code for One-time Input Modelling

R code used to produce the one-time input modelling results for Incubation two. A similar variation of code was used to produce the modelling results for Incubation one. Further data and complete set of code used in this thesis can be found in the Hicks Pries Lab GitHub.

```
# Incubation 2 Modelling Code by Michelle Wang, edited by Caitlin Hicks Pries for M.S. Thesis
April 2023

library(tidyverse)
library(SoilR)
library(FME)

# Read in data
Cinits <- c(1199.218125, 1198.987771, 1202.890795, 1208.769544, 1280.327308, 869.183259,
871.99067, 907.076619, 880.866037, 540.873971, 535.095807) # these numbers reflect if I
average C per treatment, Information from INC3 -> CombinedIRMS -> Treatment_Calculations
treatment_names <- c('DASE_C', 'DASE_O', 'AD_S', 'POET_S', 'NREL_S', 'AD_N', 'POET_N',
'NREL_N', 'CS_N', 'GWC16', 'GWC20')
inputs_frame = 0

CO2flux_0 <- read.csv("INC2data_mod.csv", header=TRUE)

i=1 # treatment
n=2 # saving number

# DASE O/C combined, so run this code and don't run it in a loop just run the #1 treatment
i = 1 #
Cinits[1] <- (1199.218125+1198.987771)/2 # just averaged DASE together
CO2flux_0 <- read.csv("DASEcomb_INC2data_mod.csv", header=TRUE) # in Excel, I averaged
DASE_C and DASE_O together and then just deleted Num = 2, calling the average Num = 1 so 2
is missing now

#Sample key as follows:
# C/O means closed/open valve
#'DASE_C' = 1 standard dosage
#'DASE_O' = 2 standard dosage
#'AD_S' = 3 standard dosage
#'POET_S' = 4 standard dosage
#'NREL_S' = 5 standard dosage
#'AD_N' = 6 new ie. halved dosage
#'POET_N' = 7 new dosage
#'NREL_N' = 8 new dosage
#'CS_N' = 9 new dosage
#'GWC16' = 10 PALOUSE SOIL CONTROL 1
#'GWC20' = 11 PALOUSE SOIL CONTROL 2

# init saving stuffs
```

```
# AICc
num_treatments = 11
AICc_1p_tot <- numeric(length=num_treatments)
AICc_2ps_tot <- numeric(length=num_treatments)
AICc_2pp_tot <- numeric(length=num_treatments)
AICc_3pp_tot <- numeric(length=num_treatments)
#AICc_3pp_fixed_tot <- numeric(length=num_treatments)

# R
R_1p_tot <-  numeric(length=num_treatments)
R_2ps_tot <-  numeric(length=num_treatments)
R_2pp_tot <-  numeric(length=num_treatments)
R_3pp_tot <-  numeric(length=num_treatments)
#R_3pp_fixed_tot <-  numeric(length=num_treatments)

# parameters
onep_par <- list(length = num_treatments)
twops_par <- list(length = num_treatments)
twopp_par <- list(length = num_treatments)
threepp_par <- list(length = num_treatments)
#threepp_fixed_par <- list(length = num_treatments)

# short term projections w/in incubation, to graph
days=seq(0,135) #Incubation days
short_totalfitCumm <- as.data.frame(matrix(nrow = length(days), ncol = num_treatments*3+1))
short_totalfitCumm[, 1] <- days
colnames(short_totalfitCumm)[1] <- 'days'

# longterm projections
proj_days = seq(1,to= 36500, by = 365/5)
totalfitCumm <- as.data.frame(matrix(nrow = length(proj_days), ncol = num_treatments*3+1))
totalfitCumm[, 1] <- proj_days
colnames(totalfitCumm)[1] <- 'days'

# Inputs every end of year, 99 inputs in dataframe, this only works for inputs w/ time steps of
365/5 days
# inputs_vals <- 1000*c(0.664092011,  # from CombinedIRMS -> Treatment Calculations in
INC2fka3
#                0.664224585,
#                0.668063965,
#                0.674809499,
#                0.746066085,
#                0.333617312,
#                0.337195283,
#                0.372710457,
#                0.345811145,
#                0,
#                0)  # these numbers reflect if I average the residues in each treatment, Information
from INC2 -> IRMS -> "IRMS_summary" -> IRMS_Pre
#
#
# inputs_mainframe <- data.frame(proj_days, matrix(0, length(proj_days),
length(treatment_names)))
# colnames(inputs_mainframe) <- c('days', 'DASE_C', 'DASE_O', 'AD_S', 'POET_S', 'NREL_S',
'AD_N', 'POET_N', 'NREL_N', 'CS_N', 'GWC16', 'GWC20')
# a = 2 # column counter
```

```
# b = 1 # inputs_vals counter
# while (a < 2+length(treatment_names)) {
#   inputs_mainframe[seq(from = 6, to = length(proj_days), by = 5), a] <-  inputs_vals[b]
#   a = a + 1
#   b = b + 1
# }
# write.csv(inputs_mainframe, file = 'inputframe.csv')

while (i < num_treatments+1) {

# begin looping
CO2flux <- CO2flux_0 %>%
  filter(Num == i) %>%    # loop through treatment
  select(time, cummCO2)

plot(x=CO2flux$time, y=CO2flux$cummCO2)

Ctotal= Cinits[i]

# graphing
theme_C <- theme_light() +
  theme(panel.grid.minor = element_blank(),
        #text = element_text(size = 30), #for facetwrapped plots
        strip.background = element_rect(color="black", fill="#93C5FF", size=1.5, linetype="solid"),
        legend.position = "none",
        plot.title = element_text(hjust = 0.5))

# One pool model
eCO2func = function(pars) {
  mod=OnepModel(
    t=days,
    k = pars[1], # GUESSES K1
    C0 = Ctotal,
    In = inputs_frame,
    pass=TRUE
  )
  AccR=getAccumulatedRelease(mod)
  return(data.frame(time=days,cummCO2=rowSums(AccR)))
}

#cost function
eCO2cost=function(pars){
  modelOutput=eCO2func(pars)
  return(modCost(model=modelOutput, obs=CO2flux[,1:2]))
}

inipars=c(k=.0001)  # for Palouse soil control should ~= .0006

# fit model to data
eCO2fit=modFit(f=eCO2cost,p=inipars,method="Nelder-Mead",
               upper=c(Inf),lower=c(0))

onep_par[[i]] <- eCO2fit$par

# rerun model w/ best parameter set for short term
fitmod=OnepModel(t=days, k=eCO2fit$par,
```

```
            In = inputs_frame,
            C0=Ctotal)

fitCumm=getAccumulatedRelease(fitmod)

short_totalfitCumm[, n] <- rowSums(fitCumm)
colnames(short_totalfitCumm)[n] <- '1P'


# plot short-term incubation v. model
fitCumm1 <- rowSums(fitCumm)
fitframe <- data.frame(days, fitCumm1)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe, aes(x = days, y = fitCumm1)) +  # model data
  xlim(0, 135) +
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '1 Pool Model') +
  theme_C
plot1

# save AICc and npars
npars=length(eCO2fit$par)
AIC_1p=(2*npars)-2*log(eCO2fit$ms)
AICc_1p=AIC_1p+(((2*npars^2)+2*npars)/(length(CO2flux[,1])-npars-1))

#pseudo r-squared
fitmod=OnepModel(t=CO2flux$time, k=eCO2fit$par,
            In = inputs_frame,
            C0=Ctotal)

CO2flux$fitCumm1p<-rowSums(getAccumulatedRelease(fitmod))

plot(CO2flux$cummCO2, CO2flux$fitCumm1p)+abline(coef = c(0,1))
test<-summary(lm(cummCO2~fitCumm1p, data=CO2flux))
R_1p<-test$r.squared

# RERUN FOR LONG TERM
fitmod=OnepModel(t=proj_days, k=eCO2fit$par,
            In = inputs_frame,
            C0=Ctotal)

fitCumm=getAccumulatedRelease(fitmod)

totalfitCumm[, n] <- rowSums(fitCumm)
colnames(totalfitCumm)[n] <- '1P'

# LONG TERM
fitCumm2 <- rowSums(fitCumm)
fitframe2 <- data.frame(proj_days, fitCumm2)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe2, aes(x = proj_days, y = fitCumm2)) +  # model data
  xlim(0, 36500) +
```

```
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '1 Pool Model') +
  theme_C
plot1

n <- n + 1

#two pool series
eCO2func=function(pars){
  mod=TwopSeriesModel(
    t=days,
    ks=pars[1:2],
    a21=pars[3]*pars[1],
    C0=Ctotal*c(pars[4],1-pars[4]),
    In=0,
    pass=TRUE
  )
  AccR=getAccumulatedRelease(mod)
  return(data.frame(time=days,cummCO2=rowSums(AccR)))
}

#cost function
eCO2cost=function(pars){
  modelOutput=eCO2func(pars)
  return(modCost(model=modelOutput, obs=CO2flux[,1:2]))
}

inipars=c(k1=0.5,k2=0.05,alpha21=0.5,gamma=0.5)

eCO2fit=modFit(f=eCO2cost,p=inipars,method="Nelder-Mead",
               upper=c(Inf,Inf,1,1),lower=c(0,0,0,0))
options(scipen = 999)
twops_par[[i]] <- eCO2fit$par

#Run the model again with best parameter set
fitmod=TwopSeriesModel(t=days, ks=eCO2fit$par[1:2],
                       a21=eCO2fit$par[3]*eCO2fit$par[1],
                       C0=Ctotal*c(eCO2fit$par[4],1-eCO2fit$par[4]),
                       In=0)
fitCumm=getAccumulatedRelease(fitmod)

short_totalfitCumm[, n] <- rowSums(fitCumm)
colnames(short_totalfitCumm)[n] <- '2PS'

#Plot the results
fitCumm1 <- rowSums(fitCumm)
fitframe <- data.frame(days, fitCumm1)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe, aes(x = days, y = fitCumm1)) +  # model data
  xlim(0, 135) +
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '2 Pool Series Model') +
  theme_C
plot1
```

```
npars=length(eCO2fit$par)
AIC_2ps=(2*npars)-2*log(eCO2fit$ms)
AICc_2ps=AIC_2ps+(((2*npars^2)+2*npars)/(length(CO2flux[,1])-npars-1))

#pseudo r-squared
fitmod=TwopSeriesModel(t=CO2flux$time, ks=eCO2fit$par[1:2],
                       a21=eCO2fit$par[3]*eCO2fit$par[1],
                       C0=Ctotal*c(eCO2fit$par[4],1-eCO2fit$par[4]),
                       In=0)

CO2flux$fitCumm2ps=rowSums(getAccumulatedRelease(fitmod))

plot(CO2flux$cummCO2, CO2flux$fitCumm2ps)+abline(coef = c(0,1))
test<-summary(lm(cummCO2~fitCumm2ps, data=CO2flux))
R_2ps<-test$r.squared

# RERUN FOR LONG TERM
fitmod=TwopSeriesModel(t=proj_days, ks=eCO2fit$par[1:2],
                       a21=eCO2fit$par[3]*eCO2fit$par[1],
                       C0=Ctotal*c(eCO2fit$par[4],1-eCO2fit$par[4]),
                       In=0)
fitCumm=getAccumulatedRelease(fitmod)

totalfitCumm[, n] <- rowSums(fitCumm)
colnames(totalfitCumm)[n] <- '2PS'

# LONG TERM
fitCumm2 <- rowSums(fitCumm)
fitframe2 <- data.frame(proj_days, fitCumm2)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe2, aes(x = proj_days, y = fitCumm2)) +  # model data
  xlim(0, 36500) +
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '2 Pool Series Model') +
  theme_C
plot1

n <- n + 1

#two pool parallel model
eCO2func=function(pars){
  mod=TwopParallelModel(
    t=days,
    ks=pars[1:2],
    gam=pars[3],
    C0=Ctotal*c(pars[3],1-pars[3]),
    In=0,
    pass=TRUE
  )
  AccR=getAccumulatedRelease(mod)
  return(data.frame(time=days,cummCO2=rowSums(AccR)))
}
```

```
eCO2cost=function(pars){
  modelOutput=eCO2func(pars)
  return(modCost(model=modelOutput, obs=CO2flux[,1:2]))
}

inipars=c(k1=0.05,k2=0.000000005,gamma=0.08) #for deeper depths, need different starting
values

eCO2fit=modFit(f=eCO2cost,p=inipars,method="Nelder-Mead",
            upper=c(Inf,Inf,1),lower=c(0,0,0))

twopp_par[[i]] <- eCO2fit$par

#Run the model again with best parameter set
fitmod=TwopParallelModel(t=days, ks=eCO2fit$par[1:2],
                         gam=eCO2fit$par[3],
                         C0=Ctotal*c(eCO2fit$par[3],1-eCO2fit$par[3]),
                         In=0)
fitCumm=getAccumulatedRelease(fitmod)

short_totalfitCumm[, n] <- rowSums(fitCumm)
colnames(short_totalfitCumm)[n] <- '2PP'

#Plot the results
fitCumm1 <- rowSums(fitCumm)
fitframe <- data.frame(days, fitCumm1)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe, aes(x = days, y = fitCumm1)) +  # model data
  xlim(0, 135) +
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '2 Pool Parallel Model') +
  theme_C
plot1

npars=length(eCO2fit$par)
AIC_2pp=(2*npars)-2*log(eCO2fit$ms)
AICc_2pp=AIC_2pp+(((2*npars^2)+2*npars)/(length(CO2flux[,1])-npars-1))

#pseudo r-squared
fitmod=TwopParallelModel(t=CO2flux$time, ks=eCO2fit$par[1:2],
                         gam=eCO2fit$par[3],
                         C0=Ctotal*c(eCO2fit$par[3],1-eCO2fit$par[3]),
                         In=0)

CO2flux$fitCumm2pp=rowSums(getAccumulatedRelease(fitmod))

plot(CO2flux$cummCO2, CO2flux$fitCumm2pp)+abline(coef = c(0,1))
test<-summary(lm(cummCO2~fitCumm2pp, data=CO2flux))
R_2pp<-test$r.squared

# LONG TERM: RERUN MODEL TO PREDICT LONG TERM
fitmod=TwopParallelModel(t=proj_days, ks=eCO2fit$par[1:2],
                         gam=eCO2fit$par[3],
                         C0=Ctotal*c(eCO2fit$par[3],1-eCO2fit$par[3]),
```

```
                    In=0)
fitCumm=getAccumulatedRelease(fitmod)

totalfitCumm[, n] <- rowSums(fitCumm)
colnames(totalfitCumm)[n] <- '2PP'

# LONG TERM: PLOT
fitCumm2 <- rowSums(fitCumm)
fitframe2 <- data.frame(proj_days, fitCumm2)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe2, aes(x = proj_days, y = fitCumm2)) +  # model data
  xlim(0, 36500) +
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '2 Pool Parallel Model') +
  theme_C
plot1

n <- n + 1

#three pool parallel
eCO2func=function(pars){
  mod=ThreepParallelModel(
    t=days,
    ks=pars[1:3],
    gam1=pars[4],
    gam2=pars[5],
    C0=Ctotal*c(pars[4],pars[5],1-pars[4]-pars[5]),
    In=0,
    pass=TRUE
  )
  AccR=getAccumulatedRelease(mod)
  return(data.frame(time=days,cummCO2=rowSums(AccR)))
}

eCO2cost=function(pars){
  modelOutput=eCO2func(pars)
  return(modCost(model=modelOutput, obs=CO2flux[,1:2]))
}

inipars=c(k1=0.005,k2=0.00005,k3=0.000000005,gam1=0.01, gam2=0.1) #for deeper depths,
need different starting values

eCO2fit=modFit(f=eCO2cost,p=inipars,method="Nelder-Mead",
            upper=c(Inf,Inf,Inf,1,1),lower=c(0,0,0,0,0))

threepp_par[[i]] <- eCO2fit$par

#Run the model again with best parameter set
fitmod=ThreepParallelModel(t=days, ks=eCO2fit$par[1:3],
                        gam1=eCO2fit$par[4],
                        gam2=eCO2fit$par[5],
                        C0=Ctotal*c(eCO2fit$par[4],eCO2fit$par[5],1-eCO2fit$par[4]-eCO2fit$par[5]),
                        In=0)
fitCumm=getAccumulatedRelease(fitmod)
```

```
short_totalfitCumm[, n] <- rowSums(fitCumm)
colnames(short_totalfitCumm)[n] <- '3PP'

#Plot the results
plot(CO2flux[,1:2],type="p",xlab="Days",
    ylab="Cummulative respiration (mg C g-1 soil)")
lines(rowSums(fitCumm))

fitCumm1 <- rowSums(fitCumm)
fitframe <- data.frame(days, fitCumm1)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe, aes(x = days, y = fitCumm1)) +  # model data
  xlim(0, 135) +
  #ylim(0, 40) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '3 Pool Model') +
  theme_C
plot1

npars=length(eCO2fit$par)
AIC_3pp=(2*npars)-2*log(eCO2fit$ms)
AICc_3pp=AIC_3pp+(((2*npars^2)+2*npars)/(length(CO2flux[,1])-npars-1))

#pseudo r-squared
fitmod=ThreepParallelModel(t=CO2flux$time, ks=eCO2fit$par[1:3],
                    gam1=eCO2fit$par[4],
                    gam2=eCO2fit$par[5],
                    C0=Ctotal*c(eCO2fit$par[4],eCO2fit$par[5],1-eCO2fit$par[4]-eCO2fit$par[5]),
                    In=0)

CO2flux$fitCumm3pp=rowSums(getAccumulatedRelease(fitmod))

plot(CO2flux$cummCO2, CO2flux$fitCumm3pp)+abline(coef = c(0,1))
test<-summary(lm(cummCO2~fitCumm3pp, data=CO2flux))
R_3pp<-test$r.squared

# LONG TERM: Run the model again with best parameter set
fitmod=ThreepParallelModel(t=proj_days, ks=eCO2fit$par[1:3],
                    gam1=eCO2fit$par[4],
                    gam2=eCO2fit$par[5],
                    C0=Ctotal*c(eCO2fit$par[4],eCO2fit$par[5],1-eCO2fit$par[4]-eCO2fit$par[5]),
                    In=0)
fitCumm=getAccumulatedRelease(fitmod)

totalfitCumm[, n] <- rowSums(fitCumm)
colnames(totalfitCumm)[n] <- '3PP'

# LONG TERM: PLOT
fitCumm2 <- rowSums(fitCumm)
fitframe2 <- data.frame(proj_days, fitCumm2)

plot1 <- ggplot() +
  geom_point(data = CO2flux, aes(x = time, y = cummCO2), shape = 1) + # INC data
  geom_line(data = fitframe2, aes(x = proj_days, y = fitCumm2)) +  # model data
```

```
  xlim(0, 36500) +
  #ylim(0, 100) +
  labs(x = 'Time [days]', y = 'Cumulative CO2 Released [mg]', title = '3 Pool Model') +
  theme_C
plot1

n <- n + 1

# Save outputs INSIDE of loop
AICc_1p_tot[i] <- AICc_1p
AICc_2ps_tot[i] <- AICc_2ps
AICc_2pp_tot[i] <- AICc_2pp
AICc_3pp_tot[i] <- AICc_3pp
#AICc_3pp_fixed_tot[i] <- AICc_3pp_fixed

R_1p_tot[i] <- R_1p
R_2ps_tot[i] <- R_2ps
R_2pp_tot[i] <- R_2pp
R_3pp_tot[i] <- R_3pp
#R_3pp_fixed_tot[i] <- R_3pp_fixed

i = i+1
print(i)
print(n)

}


# Save outputs OUTSIDE of loop
AICc_tot <- data.frame(abs(AICc_1p_tot), abs(AICc_2ps_tot), abs(AICc_2pp_tot),
abs(AICc_3pp_tot))
rownames(AICc_tot) <- treatment_names
colnames(AICc_tot) <- c('1P', '2PS', '2PP', '3PP')

R_tot <- data.frame(abs(R_1p_tot), abs(R_2ps_tot), abs(R_2pp_tot), abs(R_3pp_tot))
rownames(R_tot) <- treatment_names
colnames(R_tot) <- c('1P', '2PS', '2PP', '3PP')

write.csv(AICc_tot, file = 'DASEavg_INC2_365by5_AICc_tot.csv')
write.csv(R_tot, file = 'DASEavg_INC2_365by5_R_tot.csv')

# Export Parameters
write.csv(onep_par, file = 'DASEavg_365by5_onep_par.csv')
write.csv(twops_par, file = 'DASEavg_365by5_twops_par.csv')
write.csv(twopp_par, file = 'DASEavg_365by5_twopp_par.csv')
write.csv(threepp_par, file = 'DASEavg_365by5_threepp_par.csv')

# Export the cummCO2
write.csv(totalfitCumm, file = 'DASEavg_365by5_INC2_multmodels_projectedcummCO2.csv')
write.csv(short_totalfitCumm, file =
'DASEavg_short_365by5_INC2_multmodels_projectedcummCO2.csv')
```

Appendix G – Example Excel Calculations for Annual-Input Modelling

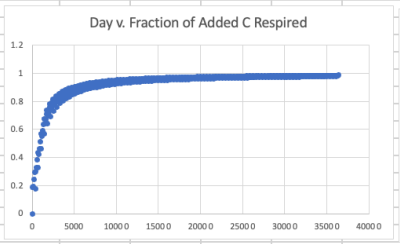
Example of calculations used to produce the annual-input modelling results for Incubation two. Similar calculations were used to produce the modelling results for Incubation one. These calculations account for the respective conversion yields of the various residue inputs by adjusting the one-time input results in "SYST" by either 35% for HLFBs and 50% for ADs and by adjusting the input, "INPUT," by the proper yield as well. The actual data and calculations used in this thesis for annual-input modelling can be found in the Hicks Pries Lab GitHub.

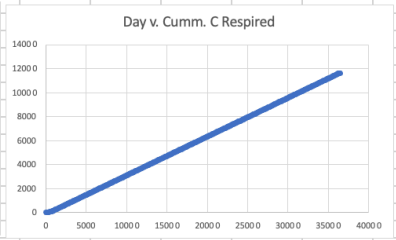
SEGMENT	DAY	INPUT	CUM_INPUT	SOIL	SYST	1	2	3	4	5	6	7	8
1	1	118.018349	118.0183491	0	0	0							
2	74	0	118.0183491	26.1808567	89.3790044	22.1193517							
3	147	0	118.0183491	40.2616577	106.439732	23.162326							
4	220	0	118.0183491	47.8742635	130.140005	28.7930095							
5	293	0	118.0183491	51.9793949	151.282785	34.7561867							
6	366	118.018349	236.0366981	54.1958695	172.355369	41.3558248	0						
7	439	0	236.0366981	55.3919334	192.711962	48.0620101	22.1193517						
8	512	0	236.0366981	56.0375829	212.545888	54.7779067	23.162326						
9	585	0	236.0366981	56.3860824	231.824327	61.4033857	28.7930095						
10	658	0	236.0366981	56.5742379	250.575894	67.9005797	34.7561867						
11	731	118.018349	354.0550472	56.6758861	268.811274	74.2473859	41.3558248	0					
12	804	0	354.0550472	56.7308166	286.545654	80.4351932	48.0620101	22.1193517					
13	877	0	354.0550472	56.7605228	303.792546	86.4612082	54.7779067	23.162326					
14	950	0	354.0550472	56.7766086	320.565418	92.3260834	61.4033857	28.7930095					
15	1023	0	354.0550472	56.7853171	336.877276	98.0321857	67.9005797	34.7561867					
16	1096	118.018349	472.0733962	56.7900873	352.740804	103.582751	74.2473859	41.3558248	0				
17	1169	0	472.0733962	56.7927216	368.168323	108.981461	80.4351932	48.0620101	22.1193517				
18	1242	0	472.0733962	56.7942033	383.171818	114.232165	86.4612082	54.7779067	23.162326				
19	1315	0	472.0733962	56.7950614	397.762961	118.018349	92.3260834	61.4033857	28.7930095				
20	1388	0	472.0733962	56.7955842	411.953065	118.018349	98.0321857	67.9005797	34.7561867				
21	1461	118.018349	590.0917453	56.7959263	425.753153	118.018349	103.582751	74.2473859	41.3558248	0			
22	1534	0	590.0917453	56.7961708	439.173945	118.018349	108.981461	80.4351932	48.0620101	22.1193517			
23	1607	0	590.0917453	56.7963627	452.225868	118.018349	114.232165	86.4612082	54.7779067	23.162326			
24	1680	0	590.0917453	56.7965261	464.919059	118.018349	118.018349	92.3260834	61.4033857	28.7930095			
25	1753	0	590.0917453	56.7966742	477.263378	118.018349	118.018349	98.0321857	67.9005797	34.7561867			
26	1826	118.018349	708.1100943	56.7968141	489.268414	118.018349	118.018349	103.582751	74.2473859	41.3558248	0		
27	1899	0	708.1100943	56.7969495	500.943493	118.018349	118.018349	108.981461	80.4351932	48.0620101	22.1193517		
28	1972	0	708.1100943	56.7970826	512.297682	118.018349	118.018349	114.232165	86.4612082	54.7779067	23.162326		
29	2045	0	708.1100943	56.7972142	523.339803	118.018349	118.018349	118.018349	92.3260834	61.4033857	28.7930095		
30	2118	0	708.1100943	56.7973452	534.078431	118.018349	118.018349	118.018349	98.0321857	67.9005797	34.7561867		
31	2191	118.018349	826.1284434	56.7974757	544.521908	118.018349	118.018349	118.018349	103.582751	74.2473859	41.3558248	0	
32	2264	0	826.1284434	56.7976061	554.678348	118.018349	118.018349	118.018349	108.981461	80.4351932	48.0620101	22.1193517	
33	2337	0	826.1284434	56.7977364	564.555638	118.018349	118.018349	118.018349	114.232165	86.4612082	54.7779067	23.162326	
34	2410	0	826.1284434	56.7978666	574.161451	118.018349	118.018349	118.018349	118.018349	92.3260834	61.4033857	28.7930095	
35	2483	0	826.1284434	56.7979968	583.503249	118.018349	118.018349	118.018349	118.018349	98.0321857	67.9005797	34.7561867	
36	2556	118.018349	944.1467924	56.798127	592.588288	118.018349	118.018349	118.018349	118.018349	103.582751	74.2473859	41.3558248	0
37	2629	0	944.1467924	56.7982571	601.423625	118.018349	118.018349	118.018349	118.018349	108.981461	80.4351932	48.0620101	22.1193517
38	2702	0	944.1467924	56.7983873	610.016124	118.018349	118.018349	118.018349	118.018349	114.232165	86.4612082	54.7779067	23.162326
39	2775	0	944.1467924	56.7985174	618.372458	118.018349	118.018349	118.018349	118.018349	118.018349	92.3260834	61.4033857	28.7930095
40	2848	0	944.1467924	56.7986475	626.499119	118.018349	118.018349	118.018349	118.018349	118.018349	98.0321857	67.9005797	34.7561867
41	2921	118.018349	1062.165141	56.7987777	634.402419	118.018349	118.018349	118.018349	118.018349	118.018349	103.582751	74.2473859	41.3558248
42	2994	0	1062.165141	56.7989078	642.088497	118.018349	118.018349	118.018349	118.018349	118.018349	108.981461	80.4351932	48.0620101
43	3067	0	1062.165141	56.799038	649.563324	118.018349	118.018349	118.018349	118.018349	118.018349	114.232165	86.4612082	54.7779067
44	3140	0	1062.165141	56.7991681	656.832706	118.018349	118.018349	118.018349	118.018349	118.018349	118.018349	92.3260834	61.4033857
45	3213	0	1062.165141	56.7992983	663.902289	118.018349	118.018349	118.018349	118.018349	118.018349	118.018349	98.0321857	67.9005797

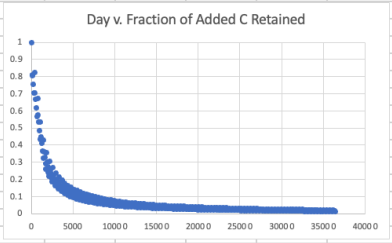
NOTES:
1. First input will be Full Soil + conversion_rate*residue_input
2. Then following inputs are conversion_rate*residue_input
3. Then SYST = conversion_rate*relevant_model_results

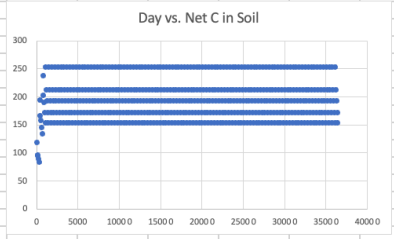
The next image is a horizontal continuation of the same spreadsheet shown above.

100	SUM	FRAC	C_Ret_res	NET_C
	0	0	1	118.018349
	22.1193517	0.18742299	0.81257701	95.8989973
	23.162326	0.19626038	0.80373962	94.8560231
	28.7930095	0.24397062	0.75602938	89.2253396
	34.7561867	0.29449816	0.70550184	83.2621624
	41.3558248	0.1752093	0.8247907	194.680873
	70.1813618	0.29733242	0.70266758	165.855336
	77.9402327	0.33020388	0.66979612	158.096465
	90.1963952	0.38212869	0.61787131	145.840303
	102.656766	0.43491867	0.56508133	133.379932
	115.603211	0.32651197	0.67348803	238.451836
	150.616555	0.42540434	0.57459566	203.438492
	164.401441	0.46433865	0.53566135	189.653606
	182.522479	0.51552006	0.48447994	171.532569
	200.688952	0.5668298	0.4331702	153.366095
	219.185962	0.46430484	0.53569516	252.887435
	259.598016	0.54991028	0.45008972	212.475381
	278.633606	**0.59023366**	0.40976634	193.43979
	300.540828	0.63664004	0.36335996	171.532569
	318.707301	0.67512235	0.32487765	153.366095
	337.204311	0.57144387	0.42855613	252.887435
	377.616365	0.63992823	0.36007177	212.475381
	396.651955	0.67218692	0.32781308	193.43979
	418.559177	0.70931204	0.29068796	171.532569
	436.72565	0.74009788	0.25990212	153.366095
	455.22266	0.64286989	0.35713011	252.887435
	495.634714	0.69994019	0.30005981	212.475381
	514.670304	0.72682244	0.27317756	193.43979
	536.577526	0.75776003	0.24223997	171.532569
	554.743999	0.7834149	0.2165851	153.366095
	573.241009	0.69388848	0.30611152	252.887435
	613.653063	0.74280588	0.25719412	212.475381
	632.688653	0.7658478	0.2341522	193.43979
	654.595875	0.79236574	0.20763426	171.532569
	672.762348	0.81435563	0.18564437	153.366095
	691.259358	0.73215242	0.26784758	252.887435
	731.671412	0.77495514	0.22504486	212.475381
	750.707002	0.79511683	0.20488317	193.43979
	772.614224	0.81832002	0.18167998	171.532569
	790.780697	0.83756118	0.16243882	153.366095
	809.277707	0.76191326	0.23808674	252.887435
	849.689761	0.79996013	0.20003987	212.475381
	868.725351	0.81788162	0.18211838	193.43979
	890.632573	0.83850669	0.16149331	171.532569
	908.799046	0.85560993	0.14439007	153.366095



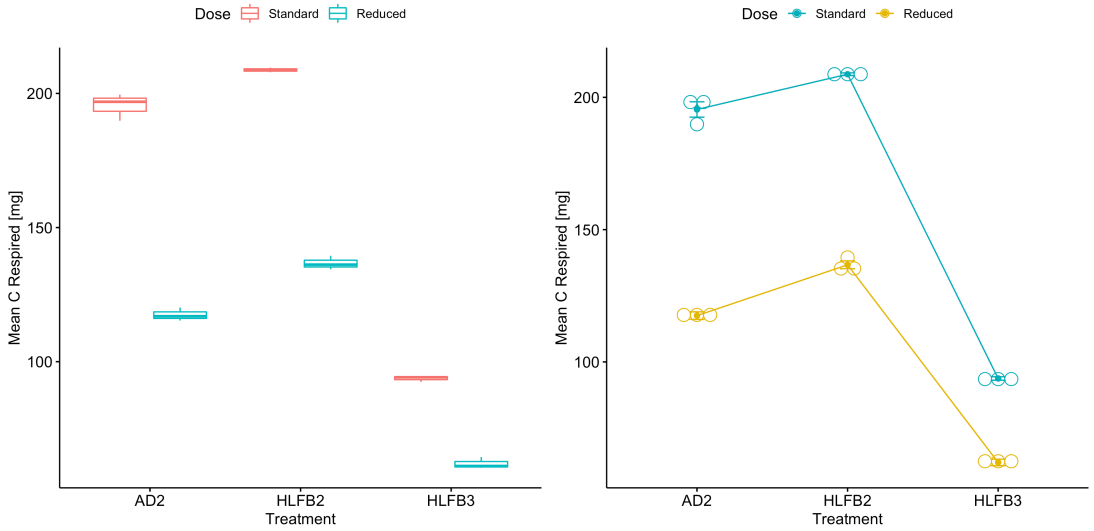






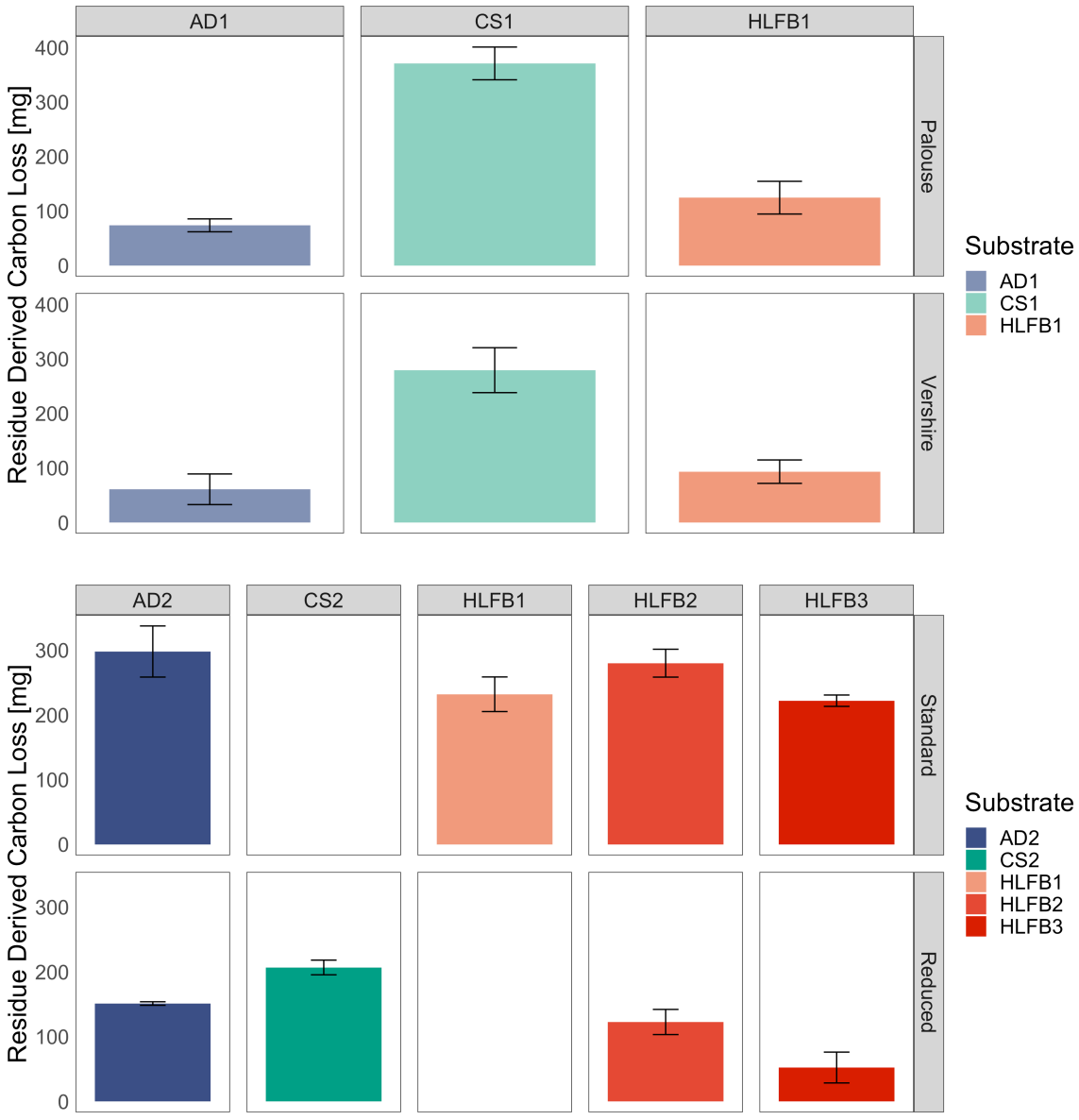
Appendix H – Dosage Incubation Results

Dosage effect is shown on the mean carbon respired for three Incubation two treatments: AD2, HLFB2, and HLFB3. Standard dosage refers to a mass ratio of 1.5 g dry residue to 37.5 g dry soil while the reduced dosage refers to a mass ratio of .75 g dry residue to 37.5 g dry soil.



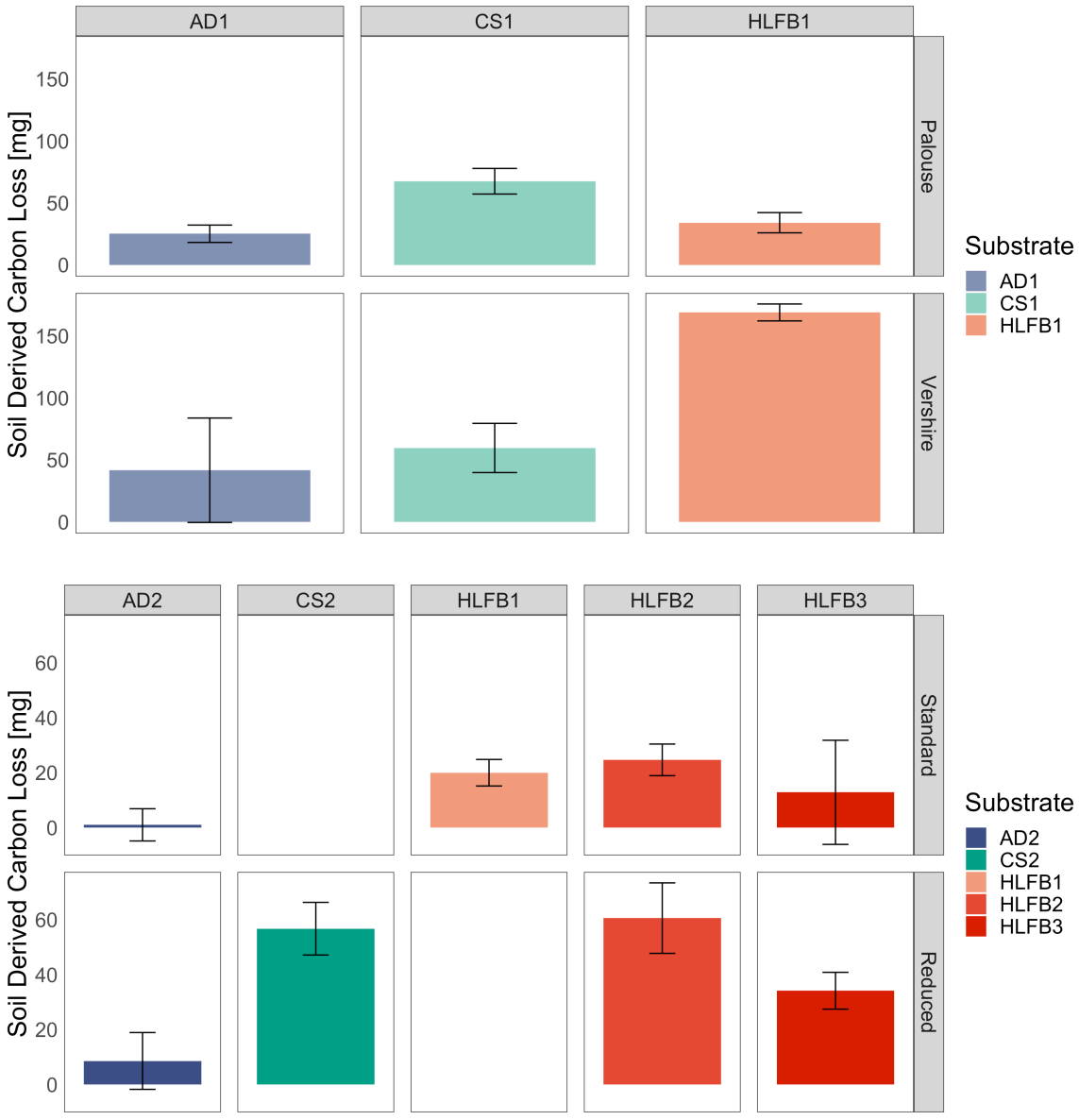
Appendix I – Residue Derived Carbon Losses

Loss in residue derived carbon pre and post incubation timespans is shown here. Top figures represent Incubation one, while bottom graphs represent Incubation two. There is no data for the CS2 * Standard Dosage and DASE1 * Reduced Dosage condition.



Appendix J – Soil Derived Carbon Losses

Soil derived carbon loss pre and post incubation timespans from isotope analysis. Top figures represent Incubation one, while bottom graphs represent Incubation two.



Appendix K – Statistics on Residue Derived Losses

Type III ANOVA test and Tukey comparison on residue derived losses differences for Incubation two separated by Palouse and Vershire soil types and Incubation two separated by standard and reduced dosage.

```
> # RES LOSS, INC 1, SOIL TYPE = PALOUSE
> res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc
 == '1' & soil_type == 'P')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
            Sum Sq Df F value    Pr(>F)
(Intercept)  10867  1   13.96  0.007298 **
treatment   169691  2  109.00 5.312e-06 ***
Residuals     5449  7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = res_stats_data %>% filter(inc == "1"
 & soil_type == "P"))

$treatment
                    diff        lwr       upr     p adj
CS_1P-AD_1P    297.03722  225.87780  368.1966 0.0000139
DASE_1P-AD_1P   50.75816  -20.40126  121.9176 0.1592121
DASE_1P-CS_1P -246.27906 -304.38048 -188.1776 0.0000125
> # RES LOSS, INC 1, SOIL TYPE = VERSHIRE
> res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc
 == '1' & soil_type == 'V')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
            Sum Sq Df F value    Pr(>F)
(Intercept)  11131  1   7.162 0.0367224 *
treatment    95127  2  30.604 0.0007115 ***
Residuals     9325  6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = res_stats_data %>% filter(inc == "1"
 & soil_type == "V"))

$treatment
                     diff        lwr       upr     p adj
CS_1V-AD_1V    218.46148  126.07651 310.84645 0.0008511
DASE_1V-AD_1V   32.32754  -78.09362 142.74870 0.6609537
DASE_1V-CS_1V -186.13394 -290.88865 -81.37923 0.0038190
```

```
> # RES LOSS, INC 2, STANDARD
> res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc
 == '2' & dose == 'S')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
            Sum Sq Df  F value    Pr(>F)
(Intercept) 266441  1 385.6831 4.701e-08 ***
treatment    12170  3   5.8722   0.02026 *
Residuals     5527  8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = res_stats_data %>% filter(inc == "2"
 & dose == "S"))

$treatment
                     diff        lwr        upr     p adj
DASE_AVG-AD_S   -66.041727 -134.76582   2.682366 0.0596310
NREL_S-AD_S     -18.037017  -86.76111  50.687076 0.8341222
POET_S-AD_S     -76.002342 -144.72644  -7.278249 0.0311290
NREL_S-DASE_AVG  48.004710  -20.71938 116.728803 0.1929436
POET_S-DASE_AVG  -9.960615  -78.68471  58.763478 0.9648639
POET_S-NREL_S   -57.965325 -126.68942  10.758768 0.1014367
> # RES LOSS, INC 2, NEW/REDUCED
> res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc
 == '2' & dose == 'N')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
            Sum Sq Df F value    Pr(>F)
(Intercept)  68550  1 247.597 2.658e-07 ***
treatment    37232  3  44.826 2.390e-05 ***
Residuals     2215  8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = res_stats_data %>% filter(inc == "2"
 & dose == "N"))

$treatment
                    diff        lwr        upr     p adj
CS_N-AD_N       55.69003   12.18341   99.19665 0.0145888
NREL_N-AD_N    -28.55127  -72.05789   14.95535 0.2312213
POET_N-AD_N    -98.90187 -142.40849  -55.39525 0.0003916
NREL_N-CS_N    -84.24130 -127.74792  -40.73468 0.0011699
POET_N-CS_N   -154.59190 -198.09852 -111.08528 0.0000150
POET_N-NREL_N  -70.35060 -113.85722  -26.84398 0.0037289
```

Appendix L – Statistics on Soil Derived Differences

Type III ANOVA test and Tukey comparison on soil derived losses differences for Incubation one separated by Palouse and Vershire soil types and Incubation two separated by standard and reduced dosage.

```
> # SOIL LOSS, INC 1, SOIL TYPE = PALOUSE
> res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter
(inc == '1' & soil_type == 'P')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III') # use because of unbalanced design
Anova Table (Type III tests)

Response: diff
            Sum Sq Df F value    Pr(>F)
(Intercept) 1264.7  1  17.399 0.0041830 **
treatment   3291.6  2  22.641 0.0008782 ***
Residuals    508.8  7
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = priming_stats_data %>% filter(inc ==
 "1" & soil_type == "P"))

$treatment
                    diff       lwr       upr     p adj
CS_1P-AD_1P    42.376925  20.63178  64.12207 0.0017627
DASE_1P-AD_1P   8.916821 -12.82832  30.66196 0.4859091
DASE_1P-CS_1P -33.460104 -51.21494 -15.70527 0.0021419
> # SOIL LOSS, INC 1, SOIL TYPE = VERSHIRE
> res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter
(inc == '1' & soil_type == 'V')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
             Sum Sq Df F value   Pr(>F)
(Intercept)  5204.3  1  7.4986 0.033810 *
treatment   21862.9  2 15.7506 0.004096 **
Residuals    4164.2  6
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = priming_stats_data %>% filter(inc ==
 "1" & soil_type == "V"))

$treatment
                   diff       lwr       upr     p adj
CS_1V-AD_1V    17.98285 -43.75376  79.71946 0.6635385
DASE_1V-AD_1V 127.31592  53.52656 201.10529 0.0044273
DASE_1V-CS_1V 109.33308  39.33034 179.33581 0.0072267
```

```
> # SOIL LOSS, INC 2, STANDARD
> res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter
(inc == '2' & dose == 'S')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
            Sum Sq Df F value Pr(>F)
(Intercept)  21.79  1  0.3158 0.5896
treatment   803.75  3  3.8820 0.0555 .
Residuals   552.13  8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = priming_stats_data %>% filter(inc ==
 "2" & dose == "S"))

$treatment
                      diff         lwr      upr     p adj
DASE_AVG-AD_S   17.287565  -4.4343144 39.00944 0.1256071
NREL_S-AD_S     21.974662   0.2527826 43.69654 0.0474473
POET_S-AD_S     13.214072  -8.5078077 34.93595 0.2822660
NREL_S-DASE_AVG  4.687097 -17.0347824 26.40898 0.8977137
POET_S-DASE_AVG -4.073493 -25.7953727 17.64839 0.9290588
POET_S-NREL_S   -8.760590 -30.4824697 12.96129 0.5924416
```

```
> # SOIL LOSS, INC 2, NEW/REDUCED
> res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter
(inc == '2' & dose == 'N')) # test interaction btwn Num and Typ
> Anova(res.aov_priming, type = 'III')
Anova Table (Type III tests)

Response: diff
            Sum Sq Df F value    Pr(>F)
(Intercept)  283.0  1  3.1286 0.1148977
treatment   4962.3  3 18.2839 0.0006121 ***
Residuals    723.7  8
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> TukeyHSD(res.aov_priming)
  Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = diff ~ treatment, data = priming_stats_data %>% filter(inc ==
 "2" & dose == "N"))

$treatment
                     diff         lwr       upr     p adj
CS_N-AD_N       47.01350  22.1439220 71.883081 0.0013713
NREL_N-AD_N     50.85812  25.9885423 75.727701 0.0008107
POET_N-AD_N     24.44041  -0.4291664 49.309992 0.0540461
NREL_N-CS_N      3.84462 -21.0249590 28.714200 0.9579792
POET_N-CS_N    -22.57309 -47.4426677  2.296491 0.0758930
POET_N-NREL_N -26.41771 -51.2872880 -1.548129 0.0378101
```

Appendix M – Soil Characteristics

The material characteristics of the soils used in incubations are shown below. Results are derived from the IRMS, IRGA, and microbial biomass protocols referenced in Appendix B.

Soil	Source	Method of Preparation	%C	%N	C:N Ratio	13C	Microbial Biomass Carbon (mg C/g dry soil)	Microbial Biomass Nitrogen (mg N/g dry soil)
Palouse Soil	Pullman, WA (Armen)	Soil collected in Pullman, WA	1.39-1.43	.10-0.11	14-13	-26.2	0.086	9.21E-05
Vershire Soil	Vershire, VT (HP Lab)	Soil collected in Vershire, VT	3.30	0.27	12	-25.0	0.256	0

Appendix N – Johnson et al. 2007 Comparison

The actual datasheet and calculations spreadsheet used for this comparison can be found in the Hicks Pries Lab GitHub. An example of the calculation is shown here.

source	inc	num	dose	mean_C_resp_cum	name	alt_name	soil	days	CS/HLFB
WANG	1	1	S	38.1305038	Soil Control	PALOUSE	P	114	/
WANG	1	2	S	289.5531372	CS1	CS1	P	114	1
WANG	1	3	S	93.48874147	AD1	AD1	P	114	3.09719794
WANG	1	5	S	101.7259511	DASE1	HLFB1	P	114	2.84640384
WANG	1	1	S	37.3911763	Soil Control	VERSHIRE	V	114	/
WANG	1	2	S	237.8359599	CS1	CS1	V	114	1
WANG	1	3	S	131.7281239	AD1	AD1	V	114	1.80550632
WANG	1	5	S	98.3180798	DASE1	HLFB1	V	114	2.41904602
WANG	2	1	S	141.7990982	DASE_C	HLFB1	P	112	/
WANG	2	2	S	137.2059965	DASE_O	HLFB1	P	112	/
WANG	2	3	S	189.3087905	AD_S	AD2	P	112	/
WANG	2	4	S	85.85232936	POET_S	HLFB3	P	112	/
WANG	2	5	S	192.7263193	NREL_S	HLFB2	P	112	/
WANG	2	6	R	112.5645784	AD_N	AD2	P	112	1.61651332
WANG	2	7	R	55.35522364	POET_N	HLFB3	P	112	3.28717198
WANG	2	8	R	125.8958677	NREL_N	HLFB2	P	112	1.44533846
WANG	2	9	R	181.9621402	CS_N	CS2	P	112	1
JOHNSON	na	na	1	1.51	HLFB	HLFB	SVEA	112	1.54304636
JOHNSON	na	na	1	2.33	CS	CS	SVEA	112	1
JOHNSON	na	na	1	1.11	HLFB	HLFB	LANGHEI	112	1.43243243
JOHNSON	na	na	1	1.59	CS	CS	LANGHEI	112	1

From Table 3 from Wang Thesis

				Carbon and Nitrogen[1]			Lignin[2]		Structural Sugars[2]					Solubilization[3]
Incubation	Substrate	Substrate	Feedstock	%C	%N	C:N	%Lignin	Lignin: N	%Glucan	%Xylan	%Galactan	%Arabinan	Average Total % Sugar	General Solubilization
1	CS1	Corn Stover	CS1	44 (.37)	0.49 (.03)	89	18 (.05)	37	39 (.02)	26 (.41)	1.5 (.24)	3.5 (.10)	71 (.57)	0%
	AD1	Anaerobic Digestate	CS1	36 (.35)	1.6 (.02)	22	40 (.25)	24	13 (.21)	8 (.08)	1.3 (.01)	2.2 (.03)	24 (.33)	-
	HLFB1	HLFB	CS1	44 (1.1)	3 (03)	15	56 (.33)	19	4.8 (.03)	0.5 (.03)	0 (.00)	1.8 (.04)	7.1 (.01)	74%
2	CS2	Corn Stover	CS2	46 (.14)	0.77 (.04)	60	16 (.08)	21	39 (.16)	28 (.24)	1.8 (.29)	3.9 (.01)	73 (.39)	0%
	AD2	Anaerobic Digestate	CS2	44 (.43)	3.4 (.03)	13	32 (.04)	9.6	19 (.08)	12 (.14)	1.4 (.09)	2.1 (.00)	35 (.13)	65%
	HLFB2*	HLFB	CS2	50 (.04)	1.3 (.02)	39	56	43	26	3.1	0.00	0.00	29	84%
	HLFB3	HLFB	CS**	45 (.03)	2 (.01)	22	45 (.48)	22	23 (.02)	3.9 (.01)	0 (.00)	0.54 (.00)	27 (.03)	52%

From Johnson et al. 2007

that could inhibit plant growth or soil processes is unknown. A byproduct from a similar process, starch fermentation, is dry-distiller's grain, which frequently is used as cattle feed. Our null hypothesis was that HLFB would not have a direct impact on

Table 1. Initial characteristics of two soils (Svea and Langhei) used in the soil incubation study before adding amendment.

Parameter	Svea	Langhei
Total C, g kg^{-1}	27.4	31.5
Inorganic C, g kg^{-1}	7.0	22.7
Organic C, g kg^{-1}	20.4	8.9
Humic acid, g kg^{-1}	17.1	1.72
Total N, g kg^{-1}	1.8	0.8
NH_4^+–N, mg kg^{-1}	5.5	7.2
NO_3^-–N, mg kg^{-1}	12.4	3.7
pH in water	7.9	8.0
pH in $CaCl_2$	7.3	7.4

Table 2. Characteristics of corn stover (CS) and high-lignin fermentation byproduct (HLFB).

Parameter	CS†	HLFB‡
C, g kg^{-1}	470	590
N, g kg^{-1}	7.0	20
Lignin, g kg^{-1}	190	590
Cellulose, g kg^{-1}	360	110
Hemicellulose, g kg^{-1}	230	50
C/N	67	30
Lignin/N	270	30

† Average corn stover values reported by U.S. Department of Energy Biomass Program (2002).

‡ Composition analysis provided by Dan Schell at NREL, Golden CO; this analysis reported 12.4% protein concentration. We estimated N percentage by assuming protein is 16% N.

Appendix O – Excerpt from Lynd et al. 2023 Manuscript

The following text is excerpted from a soon to be published manuscript entitled Liquid Biofuels from Crop Residues with Return of High-Lignin Fermentation Byproduct to the Soil[81].

Analysis and Assessment. We compare alternative strategies for managing a given quantity of above-ground crop residues via two management strategies:

No Harvest (NH), in which above-ground crop residues are left in the field;

Harvest, Process, and Return (HPR), in which above-ground crop residues are harvested, processed biologically, and solid byproduct (digestate or HLFB) produced at fractional carbon yield Y^c is returned to the field.

We assume an unchanging yearly schedule of organic matter input over a sufficient time for SOC to arrive at steady-state. The ratio of steady-state SOC for the NH and HPR strategies for management of above-ground crop residues, R_{AG}, is equal to Y^c multiplied by ε, the relative efficiency of steady-state SOC formation from soil-applied organic matter for the NH and HPR strategies. That is,

$$R_{AG} = \left(\frac{SOC_{ss,HPR}}{SOC_{ss,NH}}\right)_{AG} = \frac{Field-applied\ C,\ HPR}{Field-applied\ C,\ NH} \cdot \frac{\left(\frac{SOC_{ss}}{Field-applied\ C}\right)_{HPR}}{\left(\frac{SOC_{ss}}{Field-applied\ C}\right)_{NH}} = Y^c \varepsilon \qquad [1]$$

It follows that ε must be = 1/ Y^c for steady-state $SOC_{ss,HPR}$ to be equal to $SOC_{ss,NH}$, that is in order for R_{AG} to = 1. For example, if half the mass of agricultural residue C remains after

digestion, $Y^c = 0.5$ and □must = 2 for $R_{AG}$to = 1. If □> 1/ Y^c, then R_{AG}> 1; if □< 1/ Y^c, then $R_{AG} < 1$.

As presented above, literature reports involving manure, crop residues, animal feed components and mixtures of these indicate that long-term SOC levels are similar for field-applied digestates produced by anaerobic digestion and for crop residues left in the field. That is, $R_{AG} \approx 1$. For the Thomsen et al. study, for which $Y^c = 0.2$, $R_{AG} = 1$ implies that □ = 5. For the Smith et al. study, for which Y^c is between 0.2 and 0.31 (average 0.255) and the average steady-state value of R_{AG} is 1.23, the implied value of □is 4 to 6.2 (average 4.83). For the Béguin-Tanneau study, $Y_D = 0.36$, R_{AG} is > 1 over the timeframe evaluated, and the implied value of □is > 2.8.

Analysis of steady-state SOC levels for the HPR and NH strategies can be expanded to consider the contribution of below-ground biomass and a variable fraction of above-ground biomass harvested. For the illustrative case of above-ground and below-ground crop biomass contributing equally to SOC,

$$R_T = \left(\frac{SOC_{ss,HPR}}{SOC_{ss,NH}}\right)_T = 1 - 0.5f + 0.5f\,Y^c\,\varepsilon = 1 + 0.5f(Y^c\varepsilon - 1)$$

[2]

The 1- $0.5f$ term in Equation [2] represents the steady-state SOC that would remain if a fraction of above-ground biomass equal to *f* were harvested without any organic matter returned, normalized to the NH scenario. For f = 1 all above-ground crop residue is removed, and $1 - 0.5f = 0.5$ representing the below-ground contribution to steady-state

SOC. The $0.5Y^{c}\varepsilon$ term represents the steady-state SOC formed as a result of returning digestate or HLFB to the field.

Anticipating the SOC impact of returning HLFB to the soil is limited at present to inference based on results from anaerobic digestion and (incomplete) understanding of organic matter transformation in soils. Factors contributing to this include that soil application of HLFB from liquid cellulosic biofuel production has received vastly less study than soil application of digestates, and that processes for liquid biofuel production are still under development. As developed above, compared to unprocessed crop residues anaerobic digestate has a lower fraction of carbohydrate, higher fractions of lignin and microbial biomass, and substantially greater potential to form long-term SOC per mass applied to the field – that is, □□is substantially greater than 1□□Compared to anaerobic digestate processing the same feedstock, HLFB is expected with a high degree of confidence to have a yet lower fraction of carbohydrate, higher fractions of lignin, and may well have higher fractions of microbial biomass although this is less certain. Based on these characteristics, it is reasonable to hypothesize that the value of □ for HLFB is likely to be greater than that of anaerobic digestate from the same feedstock. Testing this hypothesis is of great interest but requires currently unavailable data from soil incubations and ultimately field studies.

Figure 2 presents R_T as a function of f based on Equation [2]. The $R_T = 1$ line applies to any combination of ε and Y such that $\varepsilon = 1/Y$ as repeatedly observed for anaerobic digestion and consistent with the general hypothesis of Thomsen et al. (2013). The $\varepsilon = 4.8$, $Y = 0.26$ line corresponds to results of Smith et al. (2014). The dashed lines are for $Y =$

0.35, typical of HLFB production accompanying liquid cellulosic biofuels, and a range of speculative values for ε from 2 to 4. For liquid cellulosic biofuel production with $Y = 0.35$, the break-even value of ε is 2.86 with $\square > 2.86$ resulting in $R_T > 1$, that is higher steady-state SOC for HPR than for NR, and $\square < 2.86$ resulting in $R_T < 1$. In general, the sensitivity of SOC to crop residue removal is substantially less with digestate or HLFB return than without removal.

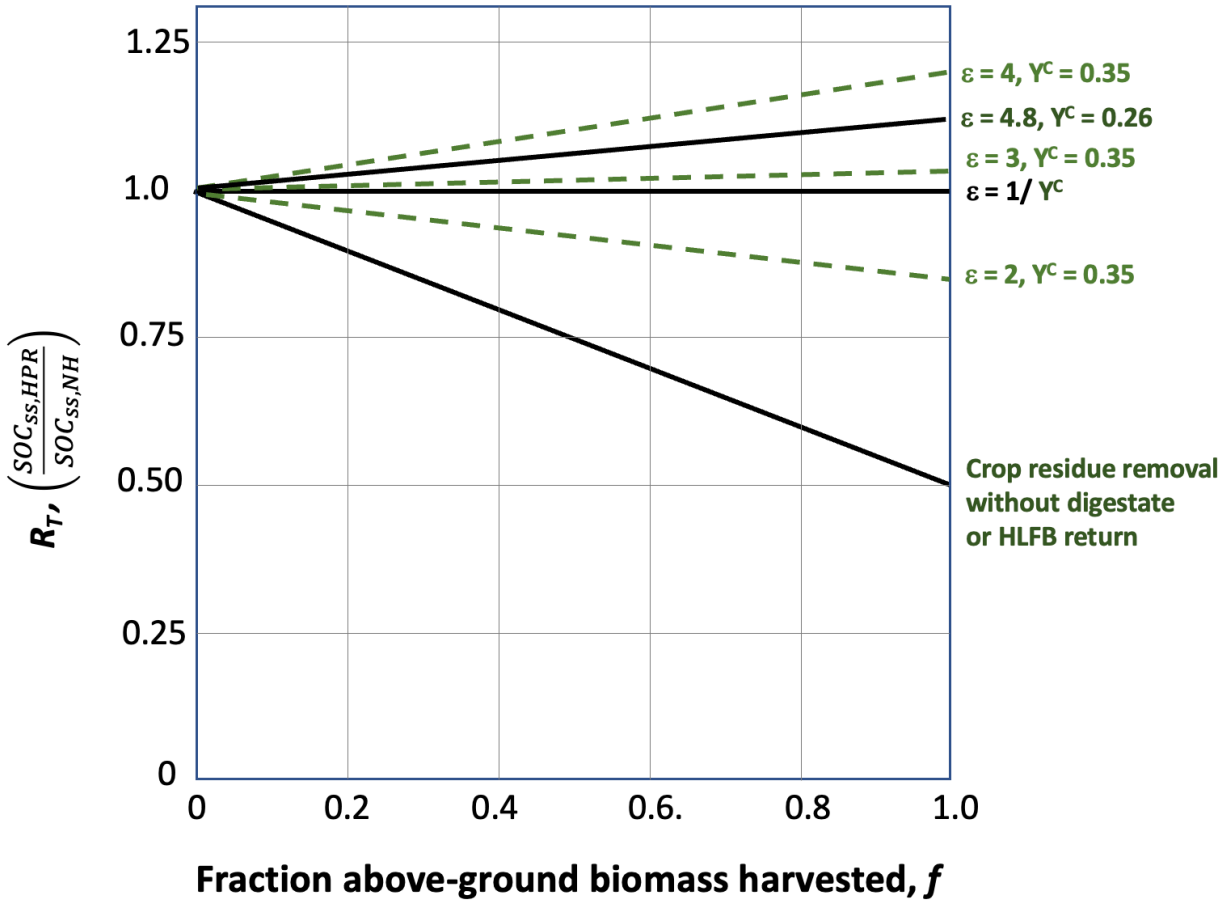


Figure 2. Steady-state SOC levels with and without HLFB return as a function of the fraction of above-ground biomass harvested. Results are calculated using Equation [2].

R_T is the steady-state SOC with harvest, processing, and return (HPR):Steady-state SOC with no harvest (NH).

ε is the relative efficiency of steady-state SOC formation from soil-applied organic matter for the HPR and NH strategies. Y is the carbon yield of solid processing byproduct (digestate or HLFB). See text for added details.

As reviewed above, a substantial literature indicates that SOC can be maintained at constant levels when about half of above-ground corn stover is harvested with no return of HLFB, although in some cases this assumes changes in management practices. The analysis and assumptions embodied in Equation [2] do not negate this possibility. Both empirical (Xu et al., 2019) and modeling (Nguyen et al., 2022) studies document net accrual of SOC for continuous corn or corn-soybean, implying that some fraction of stover could be removed without decreasing SOC at many sites. Corn stover removal with no HLFB or digestate return is, however, expected to have lower steady-state SOC levels than both NH and HPR management.

Appendix P – Example IRMS Data and Calculations

Examples of IRMS analyses data and calculations for Incubation two shown below. The full datasheets and calculations used for this analysis can be found in the Hicks Pries Lab GitHub.

			Mix						
Key	Number	Treatment	%C	%N	stdev%C	stdev%N	N15	stdevC13	stdevN15
DASE_C	1	1A	2.299017653	0.207974776	0.03306348	0.0024819	5.35924124	0.00503665	0.00299761
DASE_C	1	1B	2.399980583	0.215053697	0.03437756	0.00126285	5.33639512	0.08557694	0.25180558
DASE_C	1	1C	2.15783673	0.203778878	0.0043375	0.00025347	5.24181376	0.03539658	0.03675151
DASE_O	2	2A	2.500261473	0.220902922	0.03544522	9.1592E-05	5.48299718	0.01130332	0.07294365
DASE_O	2	2B	2.647312493	0.222440317	0.02536731	0.00135052	5.37444538	0.08602592	0.24769264
DASE_O	2	2C	2.687255744	0.227882862	0.00841711	0.00120674	5.47249094	0.09373921	0.04814539
AD_S	3	3A	2.445805643	0.239714466	0.03579969	0.00302063	3.19084914	0.26166146	0.01749832
AD_S	3	3B	2.365718482	0.229885494	0.0428971	0.00540454	3.3209888	0.55768246	0.07876182
AD_S	3	3C	2.229352688	0.225425275	0.02941685	0.00133646	3.42189325	0.00181215	0.14105478
POET_S	4	4A	2.480917528	0.180973209	0.00685963	0.00048597	4.42006246	0.01826644	0.18409468
POET_S	4	4B	2.583931584	0.18271052	0.135595	0.00688397	4.22899645	0.29017798	0.0199523
POET_S	4	4C	2.442980054	0.177698918	0.05807173	0.00269153	4.43598979	0.07209287	0.09431804
NREL_S	5	5A	2.59799279	0.156196641	0.20714199	0.00701729	4.21935048	0.23006135	0.05391934
NREL_S	5	5B	2.550848287	0.152618979	0.03743044	0.00195226	4.47697595	0.0176881	0.04021947
NREL_S	5	5C	2.45317829	0.151608632	0.00656843	0.00029128	4.50286618	0.06898625	0.00849328
AD_N	6	6A	1.897653716	0.17892705	0.01595081	0.00179746	4.01352119	0.04431561	0.11182448
AD_N	6	6B	1.850997779	0.178036215	0.04811876	0.00157987	4.09005367	0.07283933	0.00883778
AD_N	6	6C	1.849172756	0.173803626	0.03434775	0.0018006	4.1618988	0.2024891	0.1660944
POET_N	7	7A	2.100750266	0.152271496	0.00835555	0.00027198	4.57027875	0.00826872	0.17325221
POET_N	7	7B	1.969516726	0.147011741	0.08918416	0.0021774	4.77871346	0.08331447	0.02593649
POET_N	7	7C	2.125346834	0.152759344	0.00492148	0.00065407	4.33711626	0.02922571	0.12217104
NREL_N	8	8A	1.91209131	0.132206649	0.03574491	0.00044011	5.17205501	0.00045852	0.11032096
NREL_N	8	8B	1.881749203	0.128934732	0.03578339	0.00420388	5.26381244	0.04687628	0.04579674
NREL_N	8	8C	1.943412031	0.131234096	0.08724349	0.00157535	5.26024901	0.00568063	0.13133909
CS_O	9	9A	1.658766737	0.124881799	0.0359243	0.00176815	5.69353683	0.05390635	0.22473042
CS_O	9	9B	1.678399963	0.124141127	0.02183518	0.0017875	5.90858835	0.09910607	0.0750497
CS_O	9	9C	1.583063637	0.123104015	0.02279974	0.00117381	5.66697947	0.03457441	0.02961759
ONESIX	S1	S1A	1.290940657	0.115907689	0.01377814	0.00010681	6.3591591	0.01731189	0.43302555
ONESIX	S1	S1B	1.273391382	0.115548511	0.04285978	0.00154044	5.83968009	0.00335648	0.16186382
ONESIX	S1	S1C	1.277005712	0.114489426	0.00848641	0.0001491	5.02545003	0.01494555	1.47864123
TWENTY	S2	S2A	1.327542475	0.11618582	0.00730606	0.00013595	6.23717176	0.03750419	0.00792923
TWENTY	S2	S2B	1.289364363	0.114794541	0.00622441	0.00031852	6.0246184	0.07763742	0.26415415
TWENTY	S2	S2C	1.322481547	0.114504501	0.00447995	0.00021978	5.93317704	0.0693433	0.04057145

Preincubation									
C13soil_preinc	C13res_preinc	g Initial dry soil	g Initial dry residue	%C in residue	%C in soil	g InitialC_res	g InitialC_soil	total g Initial C	fr
-26.200707	-15.128041	37.6980203	1.5016	44.1912923	1.42599605	0.66357645	0.53757228	1.20114873	0.55245152
-26.200707	-15.128041	37.4232944	1.5039	44.1912923	1.42599605	0.66459285	0.5336547	1.19824755	0.55463735
-26.200707	-15.128041	37.4581233	1.5028	44.1912923	1.42599605	0.66410674	0.53415136	1.1982581	0.55422679
-26.200707	-15.128041	37.4465668	1.5003	44.1912923	1.42599605	0.66300196	0.53398656	1.19698852	0.55389166
-26.200707	-15.128041	37.5265059	1.5078	44.1912923	1.42599605	0.66631631	0.53512649	1.2014428	0.55459678
-26.200707	-15.128041	37.5300127	1.5011	44.1912923	1.42599605	0.66335549	0.5351765	1.19853199	0.55347333
-26.200707	-11.012205	37.5056245	1.5035	44.4289934	1.42599605	0.66798992	0.53482873	1.20281864	0.55535381
-26.200707	-11.012205	37.5503362	1.5023	44.4289934	1.42599605	0.66745677	0.53546631	1.20292308	0.55486238
-26.200707	-11.012205	37.4605143	1.5052	44.4289934	1.42599605	0.66874521	0.53418546	1.20293066	0.55592997
-26.200707	-13.236923	37.431105	1.5031	44.8855593	1.42599605	0.67467484	0.53376608	1.20844092	0.55830188
-26.200707	-13.236923	37.3965152	1.5028	44.8855593	1.42599605	0.67454019	0.53327283	1.20781302	0.55848064
-26.200707	-13.236923	37.5065012	1.5043	44.8855593	1.42599605	0.67521347	0.53484123	1.2100547	0.55800244
-26.200707	-13.808751	37.4923146	1.5044	49.6285562	1.42599605	0.746612	0.53463893	1.28125093	0.58272114
-26.200707	-13.808751	37.4044055	1.5039	49.6285562	1.42599605	0.74636386	0.53338535	1.2797492	0.58321103
-26.200707	-13.808751	37.5007628	1.5016	49.6285562	1.42599605	0.7452224	0.5347594	1.2799818	0.58221328
-26.200707	-11.012205	37.5245134	0.751	44.4289934	1.42599605	0.33366174	0.53509808	0.86875982	0.38406673
-26.200707	-11.012205	37.5710582	0.7508	44.4289934	1.42599605	0.33357288	0.53576181	0.86933469	0.38371054
-26.200707	-11.012205	37.5763981	0.7509	44.4289934	1.42599605	0.33361731	0.53583795	0.86945527	0.38370843
-26.200707	-13.236923	37.527542	0.751	44.8855593	1.42599605	0.33709055	0.53514127	0.87223182	0.38646899
-26.200707	-13.236923	37.5089719	0.7522	44.8855593	1.42599605	0.33762918	0.53487646	0.87250564	0.38696504
-26.200707	-13.236923	37.473346	0.7505	44.8855593	1.42599605	0.33686612	0.53436843	0.87123456	0.38665377
-26.200707	-13.808751	37.5524084	0.751	49.6285562	1.42599605	0.37271046	0.53549586	0.90820632	0.41038082
-26.200707	-13.808751	37.4582827	0.7514	49.6285562	1.42599605	0.37290897	0.53415363	0.9070626	0.41111713
-26.200707	-13.808751	37.4088687	0.7506	49.6285562	1.42599605	0.37251194	0.53344899	0.90596093	0.41117881
-26.200707	-12.045457	37.5018786	0.7501	46.0447133	1.42599605	0.34538139	0.53477531	0.8801567	0.39240898
-26.200707	-12.045457	37.5065809	0.7514	46.0447133	1.42599605	0.34597998	0.53484236	0.88082234	0.39279201
-26.200707	-12.045457	37.5559949	0.7516	46.0447133	1.42599605	0.34607207	0.535547	0.88161907	0.39254149
-26.200707	NA	37.931376	0	0	1.42599605	0	0.54089992	0.54089992	0
-26.200707	NA	37.877616	0	0	1.42599605	0	0.54013331	0.54013331	0
-26.200707	NA	37.979676	0	0	1.42599605	0	0.54158868	0.54158868	0
-26.200707	NA	37.5374248	0	0	1.42599605	0	0.5352822	0.5352822	0
-26.200707	NA	37.5301721	0	0	1.42599605	0	0.53517877	0.53517877	0
-26.200707	NA	37.5054651	0	0	1.42599605	0	0.53482645	0.53482645	0

Postincubation											
C13mix	fr	fs	g_drymix	g C from residue	g C from soil	total g C	g C from residue lost	g C from soil lost	% C from residue lost	% C from soil lost	corrected %C from soil lost
-21.53552	0.42132461	0.57867539	38.9195556	0.37698755	0.5177799	0.89476745	0.28658889	0.01979238	43.19%	3.68%	3.68%
-21.271381	0.4451797	0.5548203	38.6568529	0.41301857	0.5147384	0.92775696	0.25157428	0.0189163	37.85%	3.54%	3.54%
-22.105171	0.3698781	0.6301219	38.6940747	0.30883155	0.5261234	0.83495496	0.35527519	0.00802796	53.50%	1.50%	1.50%
-20.938325	0.47525883	0.52474117	38.5562215	0.45815253	0.50585382	0.96400635	0.20484943	0.02813274	30.90%	5.27%	5.27%
-20.650513	0.50125188	0.49874812	38.6637561	0.51305658	0.51049386	1.02355044	0.15325972	0.02463263	23.00%	4.60%	4.60%
-20.620244	0.50398549	0.49601451	38.6206859	0.52305459	0.51478201	1.0378366	0.1403009	0.02039449	21.15%	3.81%	3.81%
-19.589817	0.43525621	0.56474379	38.3575895	0.40833652	0.52981557	0.93815209	0.2596534	0.00501316	38.87%	0.94%	0.94%
-20.005191	0.4079083	0.5920917	38.5864066	0.37235736	0.54048839	0.91284575	0.29509941	-0.0050221	44.21%	-0.94%	0.00%
-20.386097	0.38282977	0.61717023	38.601415	0.32944863	0.53111306	0.86056168	0.33929658	0.0030724	50.74%	0.58%	0.58%
-20.159907	0.46597509	0.53402491	38.8359607	0.44896148	0.51452668	0.96348816	0.22571336	0.0192394	33.46%	3.60%	3.60%
-20.232221	0.46039697	0.53960303	38.905204	0.46282964	0.54245421	1.00528385	0.21171054	-0.0091814	31.39%	-1.72%	0.00%
-20.125285	0.4686458	0.5313542	39.0075994	0.44659502	0.50635286	0.95294787	0.22861845	0.02848837	33.86%	5.33%	5.33%
-20.191392	0.48493679	0.51506321	38.5433769	0.48559347	0.51576069	1.00135415	0.26101853	0.01887824	34.96%	3.53%	3.53%
-20.250483	0.48016825	0.51983175	38.3471981	0.46969043	0.50848842	0.97817885	0.27667343	0.02489693	37.07%	4.67%	4.67%
-20.407216	0.46752035	0.53247965	38.6234074	0.44297602	0.50452502	0.94750104	0.30224638	0.03023438	40.56%	5.65%	5.65%
-22.363827	0.25261741	0.74738259	37.9626636	0.18198556	0.53841434	0.7203999	0.15167618	-0.0033163	45.46%	-0.62%	0.00%
-22.199719	0.26342219	0.73657781	38.0187185	0.18537695	0.51834869	0.70372564	0.14819594	0.01741312	44.43%	3.25%	3.25%
-22.31786	0.25564385	0.74435615	38.0772415	0.18000241	0.52411157	0.70411398	0.1536149	0.01172639	46.05%	2.19%	2.19%
-21.438464	0.3673498	0.6326502	38.0407223	0.29356413	0.50557645	0.79914057	0.04352642	0.02956482	12.91%	5.52%	5.52%
-21.746994	0.3435504	0.6564496	38.1338119	0.25802415	0.49302766	0.7510518	0.07960503	0.0418488	23.58%	7.82%	7.82%
-21.326999	0.37594801	0.62405199	37.9485106	0.30321615	0.50332132	0.80653747	0.03364997	0.03104712	9.99%	5.81%	5.81%
-22.104009	0.3305934	0.6694066	38.0000248	0.24020757	0.4863876	0.72659517	0.13250289	0.04910826	35.55%	9.17%	9.17%
-22.064127	0.33381172	0.66618828	37.8240715	0.23759188	0.47416229	0.71175416	0.13531709	0.05999135	36.29%	11.23%	11.23%
-21.595992	0.37158909	0.62841091	37.734326	0.2724987	0.46083473	0.73333343	0.10001324	0.07261426	26.85%	13.61%	13.61%
-23.053829	0.22231178	0.77768822	37.6926144	0.13899656	0.48623599	0.62523255	0.20638484	0.04853932	59.76%	9.08%	9.08%
-22.831277	0.23803398	0.76196602	37.6148892	0.15027758	0.4810507	0.63132829	0.19570239	0.05379166	56.56%	10.06%	10.06%
-23.166539	0.2143493	0.7856507	37.6043092	0.12760217	0.46769797	0.59530015	0.21846989	0.06784903	63.13%	12.67%	12.67%
-25.59506	0	1	37.1523959	0	0.47961538	0.47961538	0	0.06128454	0	11.33%	11.33%
-25.714771	0	1	37.2471475	0	0.47430197	0.47430197	0	0.06583134	0	12.19%	12.19%
-25.69073	0	1	37.2112319	0	0.47518956	0.47518956	0	0.06639912	0	12.26%	12.26%
-25.695118	0	1	37.1650263	0	0.49338151	0.49338151	0	0.04190069	0	7.83%	7.83%
-25.646527	0	1	37.1762705	0	0.47933758	0.47933758	0	0.05584119	0	10.43%	10.43%
-25.594024	0	1	37.1476094	0	0.49127028	0.49127028	0	0.04355617	0	8.14%	8.14%

Appendix Q – Example IRGA Data and R Code

Example of IRGA measurements shown below. The complete datasheets and code used for all our IRGA data can be found in the Hicks Pries Lab GitHub.

Sample	Date	Time	Date Time	C_ppm	Flush	Notes	Day
CO2 FREE	11/1/22	7:10 PM	11/1/22 19:10	60	0		0
2008	11/1/22	7:10 PM	11/1/22 19:10	2008	0		
2%	11/1/22	7:10 PM	11/1/22 19:10	16078	0		
1A	11/1/22	7:10 PM	11/1/22 19:10	756	0		
1B	11/1/22	7:10 PM	11/1/22 19:10	759	0		
1C	11/1/22	7:10 PM	11/1/22 19:10	726	0		
2A	11/1/22	7:10 PM	11/1/22 19:10	712	0		
2B	11/1/22	7:10 PM	11/1/22 19:10	733	0		
2C	11/1/22	7:10 PM	11/1/22 19:10	694	0		
3A	11/1/22	7:10 PM	11/1/22 19:10	720	0		
3B	11/1/22	7:10 PM	11/1/22 19:10	645	0		
3C	11/1/22	7:10 PM	11/1/22 19:10	690	0		
4A	11/1/22	7:10 PM	11/1/22 19:10	682	0		
4B	11/1/22	7:10 PM	11/1/22 19:10	698	0		
4C	11/1/22	7:10 PM	11/1/22 19:10	667	0		
5A	11/1/22	7:10 PM	11/1/22 19:10	698	0		
5B	11/1/22	7:10 PM	11/1/22 19:10	697	0		
5C	11/1/22	7:10 PM	11/1/22 19:10	706	0		
6A	11/1/22	7:10 PM	11/1/22 19:10	633	0		
6B	11/1/22	7:10 PM	11/1/22 19:10	652	0		
6C	11/1/22	7:10 PM	11/1/22 19:10	651	0		
7A	11/1/22	7:10 PM	11/1/22 19:10	663	0		
7B	11/1/22	7:10 PM	11/1/22 19:10	979	0		
7C	11/1/22	7:10 PM	11/1/22 19:10	1027	0		
8A	11/1/22	7:10 PM	11/1/22 19:10	1086	0		
8B	11/1/22	7:10 PM	11/1/22 19:10	1042	0		
8C	11/1/22	7:10 PM	11/1/22 19:10	1014	0		
9A	11/1/22	7:10 PM	11/1/22 19:10	829	0		
9B	11/1/22	7:10 PM	11/1/22 19:10	775	0		
9C	11/1/22	7:10 PM	11/1/22 19:10	762	0		
CO2 FREE	11/1/22	7:10 PM	11/1/22 19:10	29	0		
2008	11/1/22	7:10 PM	11/1/22 19:10	1881	0		
2%	11/1/22	7:10 PM	11/1/22 19:10	15991	0		

Example of R code used to process the IRGA data for Incubation two.

```
## SOIL INCUBATION CALCULATIONS, JAN 2023
## Author: Michelle S. Wang, michelle.s.wang.th@dartmouth.edu

# Load packages + functions
library(tidyverse)
# library(SoilR)
library(FME)
library(ggpubr)

# Read in data
data <- read.csv("IRGA_Measurements.csv", stringsAsFactors = FALSE, header = TRUE) # scan
in document formatted like example
last_day <- max(data$Day, na.rm = TRUE) # [days] final day of measurement for this datasheet

#############################################################################
# GENERAL CALCULATIONS
#########################################################
# Constants
R <- 82.05746   # [mL*atm/(K*mol)]

# Room Parameters
Pr <- .98 # [atm]
Tr <- 22 + 273  # [K]

# Jar/Soil Parameters
Vjar_P <- 473.176 - 46  # [mL] pint jar - filled sample cup, from Google Sheet 'Incubation
Initializations <- Bulk Density'

# n [mol] air inside jar
n_P <- (Pr*Vjar_P) / (R*Tr)     # [mol] Palouse

# Moles/Mass of C inside jar
molmass_C <- 12.011*10^3  # [mg/mol] molar mass of C
data_C <- data %>%
  mutate(moles_C_P = C_ppm*n_P/(10^6)) %>%  # [mol] moles of C in air in jar
  mutate(mass_C_P = moles_C_P*molmass_C)    # [mg] mg of C in air in jar

# Removed air inside syringe
Vrem <- 30      # [mL] CO2 rich air removed from jar
Trem <- 25 + 273  # [K] temp of air removed since in incubator
nrem <- (Pr*Vrem) / (R*Trem)  # [mol] moles of air removed from jar

# CLEAN DATA ############################################################
# Flux
num_labs <- c('DASE_C', 'DASE_O', 'AD_S', 'POET_S', 'NREL_S', 'AD_N', "POET_N",
'NREL_N', 'CS_N', 'GWC16', 'GWC20')
names(num_labs) <- c('1', '2', '3', '4','5', '6', '7', '8', '9', 'S1', 'S2')

data_all <- data_C %>%
  filter(!Sample %in% c('CO2 FREE', '2008', '2%')) %>% # filters out controls
  separate(Sample, c("Num", "Lett"), sep=cumsum(c(1,1)), remove = FALSE) %>%  # separates
out Palouse/Vershire soil and treatments
```

```
  mutate(Date.Time = as.POSIXct(Date.Time, format = '%m/%d/%y %H:%M')) %>% # converts
Date.Time from characters to date-time format
  mutate(Date = as.Date(Date)) %>%
  group_by(Sample, Flush) %>%  # group by flush
  arrange(Date.Time) %>%  # arrange in ascending order
  mutate(time_diff = as.numeric(Date.Time - lag(Date.Time, default = first(Date.Time)), units =
'hours')) %>% # [hours] find time difference in flush groups
  mutate(mass_diff = as.numeric(mass_C_P - lag(mass_C_P, default = first(mass_C_P)))) %>%
# [mg] find mass_C difference in flush groups for Palouse and Vershire
  mutate(rem_moles_C = C_ppm*nrem/(10^6)) %>%  # [mol] moles of C in removed air
  mutate(rem_mass_C = rem_moles_C*molmass_C)  %>% # [mg] mg of C in removed air
  mutate(adj_mass_diff = as.numeric(ifelse(time_diff != '0', as.numeric(mass_diff +
lag(rem_mass_C, default = first(rem_mass_C))), '0'))) %>% # [mg] find adjusted mass by
including removal mass_C difference in flush groups
  filter(!time_diff == '0') %>% # delete used values
  mutate(flux = adj_mass_diff/time_diff) #%>% # this depends on prev. line being right
  #filter_if(~is.numeric(.), all_vars(!is.infinite(.))) # keeps the "last" day of a flux measurement ie.
gets rid of the "first" day of each session, that's what we graph


# Respired
data_resp <- data_all %>%
  group_by(Sample) %>%
  ungroup(Flush) %>%  # ungroup Flush but keep groups by Sample
  #select(Flush, Sample, Date.Time, flux) %>%  # clean it up
  arrange(Date.Time) %>% # rearrange in ascending order
  mutate(time_hours = (Date.Time - lag(Date.Time, k = 1))) %>%     # time difference btwn flux
measurements in days
  #mutate(time_hours = (Date.Time - lag(Date.Time, k = 1))*24) %>% # time difference btwn flux
measurements in hours
  mutate(C_resp = .5*(time_hours)*(flux+lag(flux))) %>% # [mg] trapezoidal area calculation to get
C respired
  drop_na(C_resp) %>% # drops rows w/ NAs which arise from the first trapezoid area
measurement
  mutate(C_resp_cum = cumsum(as.numeric(C_resp))) %>% # [mg] cumulatively add together
trapezoids
  #mutate(invC_resp_cum = total - C_resp_cum)this doesn't work, but it could be used to
generate the inv figure lee thinks abt
  mutate(time = as.numeric(Date.Time - first(Date.Time), units = 'days')) # calculate time
difference from first in group in [days]

#write.csv(data_resp, file="respdata1.csv", row.names = FALSE)

stats_resp <- data_resp %>% # output averages plotted in RESP graphs
  group_by(Sample, Num) %>%
  summarise(max_C_resp_cum = max(C_resp_cum)) %>%
  #group_by(Num) %>%   # comment this in/out if you want it broken up to replicates or not
  summarise(mean_C_resp_cum = mean(max_C_resp_cum)) #%>% # [mg]
  #mutate(Name = num_labs)    # comment this in/out if you want it broken up to replicates or not

stats_resp2 <- data_resp %>% # output averages plotted in RESP graphs
  group_by(Sample, Num) %>%
  summarise(max_C_resp_cum = max(C_resp_cum)) %>%
  group_by(Num) %>%   # comment this in/out if you want it broken up to replicates or not
  summarise(mean_C_resp_cum = mean(max_C_resp_cum), stdev = sd(max_C_resp_cum)) #
[mg C]
```

```
  #mutate(Num = num_labs)    # comment this in/out if you want it broken up to replicates or not

print(paste("The incubation period currently spans", last_day, "days!"))
#write.csv(stats_resp2, file = 'INC2summary_cumCresp.csv', row.names = FALSE) # CHECK
THAT THIS IS CORRECT NAME

# Check for ~112 days to compare with Johnson
stats_resp112 <- data_resp %>% # output averages plotted in RESP graphs
  filter(time < 113) %>%
  group_by(Sample, Num) %>%
  summarise(max_C_resp_cum = max(C_resp_cum)) %>%
  group_by(Num) %>%   # comment this in/out if you want it broken up to replicates or not
  summarise(mean_C_resp_cum = mean(max_C_resp_cum)) #%>% # [mg]
#mutate(Name = num_labs)    # comment this in/out if you want it broken up to replicates or not

write.csv(stats_resp112, file = 'INC2summary112_cumCresp.csv', row.names = FALSE) #
CHECK THAT THIS IS CORRECT NAME

#########################################################################
# STATISTICS ############################################################

# 2 WAY ANOVA for INC1, test if treatment and soil type have an effect on mean C resp/fraction
of C retained by soil/fraction of C retained by residue by end of incubation

# recode Num to factors
thirteenC_data <- read.csv("INC2_2wayanova_13C.csv", stringsAsFactors = FALSE, header =
TRUE) # scan in document formatted like example
soil_C_data <- read.csv('summary_SOILcumCresp.csv', stringsAsFactors = FALSE, header =
TRUE)
stats_resp <- rbind(stats_resp, soil_C_data)

twowayanova_data <- merge(stats_resp, thirteenC_data, by = 'Sample') # if this excludes the soil
data, check to make sure the 'Sample' column for both sheets is labelled correctly

twowayanova_data$Num <- factor(twowayanova_data$Num,
                               levels = names(num_labs),
                               labels = num_labs)

## Open vs. Closed, paired t-test
oc_data <- twowayanova_data %>%
  filter(Num == 'DASE_O' | Num == 'DASE_C')

t.test(mean_C_resp_cum ~ Valve, data = oc_data, paired = TRUE)
t.test(fr ~ Valve, data = oc_data, paired = TRUE)
# We see that for both 13C and inc data, p>.05 or that there is an insignificant difference between
Open v. Closed.

## Dosage, 2-way ANOVA
dose_data <- twowayanova_data %>%
  filter(Sub == 'AD' | Sub == 'NREL' | Sub == 'POET') %>% # balanced design since same
number of observations per treatment
  mutate(Sub = factor(Sub, levels = c('AD', 'NREL', 'POET'), labels = c("AD2","HLFB2",
"HLFB3"))) %>%
  mutate(Dose = factor(Dose, levels = c('S', 'N'), labels = c('Standard', 'Reduced')))

dose_summary <- dose_data %>%
```

```
  group_by(Num, Sub, Dose) %>%
  summarize(act_mean_C_resp_cum = mean(mean_C_resp_cum))

dose_plot1 <- ggboxplot(dose_data, x = 'Sub', y = 'mean_C_resp_cum', color = 'Dose', # boxplot
shows various treatments and how they compare to each other + soil type
         xlab = 'Treatment',
         ylab = 'Mean C Respired [mg]')
dose_plot1
ggsave("dose_plot1.png", plot = dose_plot1, width = 15, height = 15, units = "cm")


dose_plot2 <- ggline(dose_data, x = "Sub", y = "mean_C_resp_cum", color = "Dose",
       add = c("mean_se", "dotplot"),
       palette = c("#00AFBB", "#E7B800"),
       xlab = 'Treatment',
       ylab = 'Mean C Respired [mg]')
dose_plot2
ggsave("dose_plot2.png", plot = dose_plot2, width = 15, height = 15, units = "cm")


ggline(dose_data, x = "Sub", y = "fr", color = "Dose",
       add = c("mean_se", "dotplot"),
       palette = c("#00AFBB", "#E7B800"))

res.aov_dose <- aov(mean_C_resp_cum ~ Dose * Sub, data = dose_data) # test interaction btwn
Num and Typ
summary(res.aov_dose)

res.aov_dose2 <- aov(fr ~ Dose * Sub, data = dose_data) # test interaction btwn Num and Typ
summary(res.aov_dose2)

# Tukey-Kramer maybe
TukeyHSD(res.aov_dose)  # unclear if this is taking into account unbalanced design, I think it's
using Tukey Kramer

## INC1 V INC2, t-test
INC2_DASE_C_data <- twowayanova_data %>%
  #filter(Sub == 'DASE' & Valve == 'C') %>%
  filter(Sub == 'DASE' & Valve == 'C') %>%
  select(Num, mean_C_resp_cum, fr) %>%
  mutate(Inc = '2')

INC1_anovadata <- read.csv("INC1_twowayanova_data.csv", stringsAsFactors = FALSE, header
= TRUE) # scan in document formatted like example
INC1_DASE_data <- INC1_anovadata %>%
  filter(Typ == 'P' & Num == 'DASE HLFB') %>%
  select(Num, mean_C_resp_cum, fr) %>%
  mutate(Inc = '1')

DASE_data <- rbind(INC2_DASE_C_data, INC1_DASE_data)

t.test(mean_C_resp_cum ~ Inc, data = DASE_data, paired = FALSE)
t.test(fr ~ Inc, data = DASE_data, paired = FALSE)

# Check SOIL controls
INC2_soil_data <- twowayanova_data %>%
```

```
  filter(Sub == 'SOIL') %>%
  select(Num, mean_C_resp_cum) %>%
  mutate(Inc = '2')

INC1_soil_data <- INC1_anovadata %>%
  filter(Typ == 'P' & Num == 'Soil Control') %>%
  select(Num, mean_C_resp_cum) %>%
  mutate(Inc = '1')

SOIL_data <- rbind(INC1_soil_data, INC2_soil_data)

t.test(mean_C_resp_cum ~ Inc, data = SOIL_data, paired = FALSE)

################################################################################
# PLOTTING ####################################################

# Theme and Labels
theme_C <- theme_light() +
  theme(panel.grid.minor = element_blank(),
  text = element_text(size = 30), #for facetwrapped plots
  strip.background = element_rect(color="black", fill="#93C5FF", size=1.5, linetype="solid"),
  legend.position = "none",
  plot.title = element_text(hjust = 0.5)
  )

# CHANGE THESE DATES FOR YOUR GRAPHING PLEASURE!
end_date = '2023-03-09 14:00' # !!!! CHANGE THIS TO EXTEND GRAPH !!!
  #'2022-12-13 9:15' <- this is for 42 days
# start_date = #"2022-11-07 7:30" # <- this is for ignoring the initial 6 day bump
start_date = "2022-11-01 19:10" # ACTUAL FIRST MEASUREMENT
lims <- as.POSIXct(strptime(c(start_date, end_date), format = "%Y-%m-%d %H:%M"))

# FLUX: Mean and SE Each
p1P<- ggplot(data_all, aes(x=Date.Time, y=flux)) +
  geom_point(aes(size = .8)) +
  scale_x_datetime(limits = lims) +
  stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
  facet_wrap(~Num, labeller = labeller(Num = num_labs)) +
  #facet_wrap(~Num, scales = 'free', labeller = labeller(Num = num_labs)) + # free scale bc 1 is
so small
  theme_C +
  #scale_y_continuous(limits=c(0,.35)) +  # sets all plots start at 0 go to .3
  labs(x = '', y = 'Carbon Flux [mg/hr]', title = 'Carbon Flux Evolution in Various Treatments')
p1P

#ggsave("flux_mean&se.png", plot = p1P, width = 60, height = 20, units = "cm")  # change this
accordingly

# RESPIRED: Mean and SE Each
p2P <- ggplot(data_resp, aes(x=Date.Time, y=C_resp_cum)) +
  geom_point(aes(size = .8)) +
  scale_x_datetime(limits = lims) +
  stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
  facet_wrap(~Num, labeller = labeller(Num = num_labs)) + # NON FREE SCALE
  ##facet_wrap(~Num, scales = 'free', labeller = labeller(Num = num_labs)) + # free scale bc 1 is
so small
```

```
  ##geom_vline(xintercept = as.POSIXct(as.Date(c('2021-03-22', '2021-04-22'))), linetype =
'dashed', color = 'blue', size = 2) +  # when water was added, comment this out for no lines
  theme_C +
  #scale_y_continuous(limits=c(0,500)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon Respired in
Various Treatments')
p2P

#ggsave("resp_mean&se.png", plot = p2P, width = 60, height = 20, units = "cm")


# GRAPHS OF RESPIRED ONLY OF THE NEW RATIOS
edit_data_resp <- data_resp %>%
  filter(Num == '6' | Num == '7'| Num == '8'| Num == '9')

edit_p2P <- ggplot(edit_data_resp, aes(x=Date.Time, y=C_resp_cum)) +
  geom_point(aes(size = .8)) +
  scale_x_datetime(limits = lims) +
  stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
  facet_wrap(~Num, labeller = labeller(Num = num_labs)) +
  theme_C +
  labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon Respired in 50%
Residue Dosage Treatments')
edit_p2P
ggsave("newratio_resp_mean&se.png", plot = edit_p2P, width = 60, height = 20, units = "cm")

# GRAPHS OF RESPIRED ONLY OF OLD RATIOS
edit_data_resp <- data_resp %>%
  filter(Num == '3' | Num == '4'| Num == '5')

edit_p2P <- ggplot(edit_data_resp, aes(x=Date.Time, y=C_resp_cum)) +
  geom_point(aes(size = .8)) +
  scale_x_datetime(limits = lims) +
  stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
  facet_wrap(~Num, labeller = labeller(Num = num_labs)) +
  theme_C +
  labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon Respired in
Normal Residue Dosage Treatments')
edit_p2P
ggsave("oldratio_resp_mean&se.png", plot = edit_p2P, width = 60, height = 20, units = "cm")

# GRAPHS OF DASE O/C
OCdata_resp <- data_resp %>%
  filter(Num == '1' | Num == '2')

OC_p <- ggplot(OCdata_resp, aes(x=Date.Time, y=C_resp_cum)) +
  geom_point(aes(size = .8)) +
  scale_x_datetime(limits = lims) +
  stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
  facet_wrap(~Num, labeller = labeller(Num = num_labs)) +
  theme_C +
  labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon Respired in DASE
O/C Treatments')
OC_p
ggsave("DASEOC_resp_mean&se.png", plot = OC_p, width = 60, height = 20, units = "cm")
```

```
# OC_p2 <- ggplot(OCdata_resp, aes(x=Date.Time, y=C_resp_cum)) +
#   geom_point(aes(size = .8, col = Num)) +
#   scale_x_datetime(limits = lims) +
#   stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
#   #facet_wrap(~Num, labeller = labeller(Num = num_labs)) +
#   theme_C +
#   labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon Respired in
DASE O/C Treatments') +
#   legend
# OC_p2
# ggsave("sameDASEOC_resp_mean&se.png", plot = OC_p2, width = 60, height = 20, units =
"cm")

#lumped figure w/ geom smooth of C respired
theme_lump <- theme_light() +
  theme(panel.grid.minor = element_blank(),
        text = element_text(size = 30), #for facetwrapped plots
        strip.background = element_rect(color="black", fill="#93C5FF", size=1.5, linetype="solid"),
        legend.position = "bottom",
        plot.title = element_text(hjust = 0.5)
  )

data_resp_old <- data_resp %>%
  filter(Num == '2' |Num == '3' | Num == '4'| Num == '5')

lumped1 <- ggplot(data_resp_old, aes(x=Date.Time, y=C_resp_cum)) +
  geom_smooth(aes(color = Num), se = TRUE) +
  scale_x_datetime(limits = lims) +
  theme_lump +
  scale_color_manual("Treatments", labels = c("DASE1", "AD2", "DASE2", "DASE3"), values =
c("2", "3", "4", "5")) +
  scale_y_continuous(limits=c(0,250)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon Respired in
INCUBATION 2 of HLFB Amended Palouse Soil')
lumped1

ggsave("Plumped_scale_mean&se.png", plot = lumped_P, width = 60, height = 20, units = "cm")

#C retained throughout Incubation 2
datainitC <- read.csv("justinitC.csv", stringsAsFactors = FALSE, header = TRUE) # scan in
document formatted like example
data3 <- left_join(data_resp, datainitC, by = 'Sample')
data3 <- data3 %>%
  mutate(invC_resp_cum = init_C*1000 - C_resp_cum)  %>%
  filter(Num == '2' |Num == '3' | Num == '4'| Num == '5') %>%
  mutate(invC_resp_cum_ADJ = case_when(Num == '3' ~ .5*invC_resp_cum,
                                       Num == '2' | Num == '4' ~ .35*invC_resp_cum,
                                       Num == '5' ~ .35*invC_resp_cum)) %>%
  mutate(ID = case_when(Num == '2' ~ 'DASE1_2',
                        Num == '3' ~ 'AD2',
                        Num == '4' ~ 'DASE2',
                        Num == '5' ~ 'DASE3'))

lumped_2 <- ggplot(data3, aes(x=Date.Time, y=invC_resp_cum)) +
  geom_smooth(aes(color = Num), se = TRUE) +
  scale_x_datetime(limits = lims) +
```

```
  theme_lump +
  scale_color_manual("Treatments", labels = c("DASE1", "AD2", "DASE2", "DASE3"), values =
c("2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,250)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = '', y = 'Cumulative Carbon Retained [mg]', title = 'Carbon Retained in Treatments
Throughout INCUBATION 2 of HLFB Amended Palouse Soil')
lumped_2

ggsave("Cret4res.png", plot = lumped_2, width = 60, height = 20, units = "cm")


lumped_3 <- ggplot(data3, aes(x=Date.Time, y=invC_resp_cum_ADJ)) +
  geom_smooth(aes(color = Num), se = TRUE) +
  scale_x_datetime(limits = lims) +
  theme_lump +
  scale_color_manual("Treatments", labels = c("DASE1", "AD2", "DASE2", "DASE3"), values =
c("2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,250)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = '', y = 'Cumulative Carbon Retained [mg]', title = 'Carbon Retained in Treatments
Throughout INCUBATION 2 of HLFB Amended Palouse Soil')
lumped_3

ggsave("Cret4resADJ.png", plot = lumped_3, width = 60, height = 20, units = "cm")

# combine
inc2data <- read.csv("INC2_invC_resp.csv", stringsAsFactors = FALSE, header = TRUE) # scan
in document formatted like example

retdata1 <- data3 %>%
  select(Sample, ID, time, invC_resp_cum, invC_resp_cum_ADJ)

retdata2 <- inc2data %>%
  filter(time < 135) %>%
  select(Sample, ID, time, invC_resp_cum, invC_resp_cum_ADJ)

retdata_comb <- rbind(data_frame(retdata1), data_frame(retdata2))

lumped_4 <- ggplot(retdata_comb, aes(x=time, y=invC_resp_cum)) +
  geom_smooth(aes(color = ID), se = TRUE) +
  #scale_x_datetime(limits = lims) +
  theme_lump +
  #scale_color_manual("Treatments", labels = c("DASE1", "AD2", "DASE2", "DASE3"), values =
c("2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,250)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = 'Time [days]', y = 'Cumulative Carbon Retained [mg]', title = 'Carbon Retained in 135
Day Incubations of HLFB Amended Palouse Soil')
lumped_4

ggsave("totalCret4res.png", plot = lumped_4, width = 60, height = 20, units = "cm")

lumped_5 <- ggplot(retdata_comb, aes(x=time, y=invC_resp_cum_ADJ)) +
  geom_smooth(aes(color = ID), se = TRUE) +
  #scale_x_datetime(limits = lims) +
  theme_lump +
  #scale_color_manual("Treatments", labels = c("DASE1", "AD2", "DASE2", "DASE3"), values =
c("2", "3", "4", "5")) +
```

```
  #scale_y_continuous(limits=c(0,250)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = 'Time [days]', y = 'Cumulative Carbon Retained [mg]', title = 'Carbon Retained in 135
Day Incubations of HLFB Amended Palouse Soil')
lumped_5

ggsave("ADJtotalCret4res.png", plot = lumped_5, width = 60, height = 20, units = "cm")

#Show Initial C
pV_init <- ggplot(data_resp, aes(x=Date.Time, y=C_resp_cum)) +
  geom_smooth(aes(size = .8)) +
  scale_x_datetime(limits = lims) +
  #stat_summary(fun.data = "mean_se", colour = "red", size = .8) +
  facet_wrap(~Num, labeller = labeller(Num = num_labs)) + # NON FREE SCALE
  #geom_vline(xintercept = as.POSIXct(as.Date(c('2021-03-22', '2021-04-22'))), linetype =
'dashed', color = 'blue', size = 2) +
  #facet_wrap(~Num, scales = 'free', labeller = labeller(Num = num_labs)) + # free scale bc 1 is
so small
  ##geom_vline(xintercept = as.POSIXct(as.Date(c('2021-03-22', '2021-04-22'))), linetype =
'dashed', color = 'blue', size = 2) +  # when water was added, comment this out for no lines
  theme_C +
  scale_y_continuous(limits=c(0,NA)) +  # sets all plots start at 0 go to unique maxes for each
  labs(x = '', y = 'Cumulative Carbon Respired [mg]', title = 'Cumulative Carbon in 267 Day
Incubation of HLFB Amended Vershire Soil')
pV_init

# Calculate C retained as percentage of residue C and total treatment C
datainitC <- read.csv("justinitC.csv", stringsAsFactors = FALSE, header = TRUE) # scan in
document formatted like example
data3 <- left_join(data_resp, datainitC, by = 'Sample')
data3 <- merge(data3, soil_resp, by = c('Typ', 'Flush'))  # matches soil resp. to each
measurement at a time point
data3 <- data3 %>%
  mutate(init_totC = init_C*1000) %>%         # [mg C] initial C (soil+res) in each treatment on
average
  mutate(init_resC = init_resC*1000) %>%      # [mg C] initial C (res) in each treatment on average
  mutate(Ctot_ret = 100*(init_totC - C_resp_cum)/init_totC) %>% # [% total C] C retained from
total treatment
  mutate(Cres_ret = 100*(init_resC-(C_resp_cum-mean_soil_cum))/init_resC) %>% # [% residue
C] C retained from residue in each treatment
  group_by(Num, Typ, Date.Time) %>%
  summarize(meanCtot_ret = mean(Ctot_ret), meanCres_ret = mean(Cres_ret)) %>% # [% total
C] average of prev. calculations per treatment on specific days
  ungroup() %>% # necessary to add row after
  add_row(Typ = 'P', Num = c('1', '2', '3', '4', '5'), Date.Time = as.POSIXct('2021-11-01 15:00:00'),
meanCres_ret = 100, meanCtot_ret = 100) %>% # add initial anchor point of 100% for all
treatments (when incubation began)
  add_row(Typ = 'V', Num = c('1', '2', '3', '4', '5'), Date.Time = as.POSIXct('2021-11-01 15:00:00'),
meanCres_ret = 100, meanCtot_ret = 100)

data3P <- data3 %>%
  filter(Typ == 'P') %>%
  mutate(Ctot_Label = round(ifelse(Date.Time == max(Date.Time), meanCtot_ret, NA), 0)) %>%
# add labels to last point of each line
  mutate(Cres_Label = round(ifelse(Date.Time == max(Date.Time), meanCres_ret, NA), 0))

data3V <- data3 %>%
```

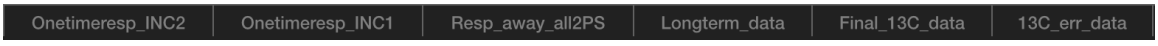
```
  filter(Typ == 'V') %>%
  mutate(Ctot_Label = round(ifelse(Date.Time == max(Date.Time), meanCtot_ret, NA), 0)) %>%
# add labels to last point of each line
  mutate(Cres_Label = round(ifelse(Date.Time == max(Date.Time), meanCres_ret, NA), 0))

# Graph C retained graphs
# C retained of only residue graphs
Cret_P <- ggplot(data3P, aes(x=Date.Time, y=meanCres_ret)) +
  geom_line(aes(color = Num), size = .5) +
  geom_point(size = .25, color = 'black') +
  scale_x_datetime(date_breaks = '1 month', labels = date_format("%b")) +
  ylim(35, 100) +
  theme_lump +
  scale_color_manual("Treatments", labels = c("Soil Control", "CS", "AD", "C-CBP", "DASE"),
values = c("1", "2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,500)) +  # sets all plots start at 0 go to 500
  labs(x = '', y = 'Carbon Retained in Residue \n [% of Initial Residue C]', title = 'Palouse Soil
Incubations') +   # Carbon Retained in Residue in \n 267 Day Incubation of Biofuel Residues in
Palouse Soil
  geom_label_repel(aes(label = Cres_Label), min.segment.length = 0, size = 2, force = 2.1,
direction = 'y', hjust = 'left', label.padding = unit(0.1, "lines"), na.rm = TRUE)  # labels last point
with final percentage of each line
Cret_P

# C retained of total treatment graphs
Cret_P2 <- ggplot(data3P, aes(x=Date.Time, y=meanCtot_ret)) +
  geom_line(aes(color = Num), size = .5) +
  geom_point(size = .25, color = 'black') +
  scale_x_datetime(date_breaks = '1 month', labels = date_format("%b")) +
  ylim(60, 100) +
  #scale_x_datetime(limits = lims) +
  theme_lump +
  scale_color_manual("Treatments", labels = c("Soil Control", "CS", "AD", "C-CBP", "DASE"),
values = c("1", "2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,500)) +  # sets all plots start at 0 go to 500
  labs(x = '', y = 'Carbon Retained in Treatment \n [% of Initial Treatment C]', title = 'Palouse Soil
Incubations') + # 'Carbon Retained in Treatment in \n 267 Day Incubation of Biofuel Residues in
Palouse Soil'
  geom_label_repel(aes(label = Ctot_Label), min.segment.length = 0 , size = 2, force = .6,
direction = 'y', hjust = 'left', label.padding = unit(0.1, "lines"), na.rm = TRUE)  # labels last point
with final percentage of each line
Cret_P2

# C retained of only residue graphs
Cret_V <- ggplot(data3V, aes(x=Date.Time, y=meanCres_ret)) +
  geom_line(aes(color = Num), size = .5) +
  geom_point(size = .25, color = 'black') +
  scale_x_datetime(date_breaks = '1 month', labels = date_format("%b")) +
  ylim(35, 100) +
  #scale_x_datetime(limits = lims) +
  theme_lump +
  #geom_label_repel(aes(), nudge_x = 1, na.rm = TRUE) +  # CHANGE THIS SO THE LABEL
WORKS, PAGE IS SAVED IN GOOGLE
  scale_color_manual("Treatments", labels = c("Soil Control", "CS", "AD", "C-CBP", "DASE"),
values = c("1", "2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,500)) +  # sets all plots start at 0 go to 500
```

```
  labs(x = '', y = 'Carbon Retained in Residue \n [% of Initial Residue C]', title = 'Vershire Soil
Incubations') +   # Carbon Retained in Residue in \n 267 Day Incubation of Biofuel Residues in
Vershire Soil
  geom_label_repel(aes(label = Cres_Label), min.segment.length = 0, size = 2, force = .5,
direction = 'y', hjust = 'left', label.padding = unit(0.1, "lines"), na.rm = TRUE)  # labels last point
with final percentage of each line
Cret_V

# C retained of total treatment graphs
Cret_V2 <- ggplot(data3V, aes(x=Date.Time, y=meanCtot_ret)) +
  geom_line(aes(color = Num), size = .5) +
  geom_point(size = .25, color = 'black') +
  scale_x_datetime(date_breaks = '1 month', labels = date_format("%b")) +
  ylim(60, 100) +
  #scale_x_datetime(limits = lims) +
  theme_lump +
  scale_color_manual("Treatments", labels = c("Soil Control", "CS", "AD", "C-CBP", "DASE"),
values = c("1", "2", "3", "4", "5")) +
  #scale_y_continuous(limits=c(0,500)) +  # sets all plots start at 0 go to 500
  labs(x = '', y = 'Carbon Retained in Treatment \n [% of Initial Treatment C]', title = 'Vershire Soil
Incubations') + # 'Carbon Retained in Treatment in \n 267 Day Incubation of Biofuel Residues in
Vershire Soil'
  geom_label_repel(aes(label = Ctot_Label), min.segment.length = 0, size = 2, force = .6,
direction = 'y', hjust = 'left', label.padding = unit(0.1, "lines"), na.rm = TRUE)  # labels last point
with final percentage of each line
Cret_V2

ggsave("CretP_scale_mean&se.png", plot = Cret_P, width = 60, height = 20, units = "cm")
ggsave("CretV_scale_mean&se.png", plot = Cret_V, width = 60, height = 20, units = "cm")

#############################################################################
# SOIL MODELLING ############################################################
# Based off of https://www.bgc-jena.mpg.de/TEE/optimization/2015/12/09/Fractions-Incubations/
# Context from https://escholarship.org/uc/item/9h72f7hk

# Clean data for modelling
data_mod <- data_resp %>%
  ungroup(Sample) %>%   # now, not grouped as anything
  select(c('time','Num', 'C_resp_cum'))  %>% # select these columns for ease
  group_by(Num, time) %>%  #
  summarize(cummCO2 = mean(C_resp_cum)) # sd gives an error for some reason: Stderr =
sd(C_resp_cum))   # [mg] amount of carbon respired cumulatively, not in terms of mg C/g soil
  #summarize(cummCO2 = mean(C_resp_cum)/50, Stderr = sd(C_resp_cum/50)) %>%  # /50 so
it's in [g C/g soil] since we start w/ ~50g soil, summarizing by all incubations def. loses precision
since it's not a rate, it's an absolute amount?, but also it's based off of rate anyways
write.csv(data_mod, file = 'INC2data_mod.csv')
```

Appendix R – Example Overall Graphing Data and R Code

Key explaining the sheets in the overall datasheet used to generate the graphs referenced in this thesis is shown below. The complete datasheet and R code for the graphs presented in this thesis can be found in the Hicks Pries Lab GitHub.

The first two sheets “Onetimeresp_INC2”, “Resp_away_all2PS,” and “Onetimeresp_INC1” were used to generate the one-time input graphs shown in Figure 3. The “Longterm_data” sheet was used to generate the annual-input modelling graphs shown in Figure 4. The “Final_13C_data” and “13C_err_data” sheet was used to generate the 13C partitioning graphs in Figure 1 and priming / partitioned graphs shown in Figure 2 and Appendices I and J .



Example of the R code used to generate the graphs in this thesis is shown below.

```
## GRAPHS FOR WANG THESIS 2023
## Author: Michelle S. Wang, michelle.s.wang.th@dartmouth.edu

# Load packages + functions
library(tidyverse)
library(ggsci)
library(ggrepel)
library(scales)

library(FME)
library(ggpubr)
library(car)

# Nature color palette: https://nanx.me/ggsci/reference/pal_npg.html;
show_col(pal_npg("nrc")(10))

# Theme
theme_C <- theme_light() +
  theme(panel.grid.minor = element_blank(),
        text = element_text(size = 20), #for facetwrapped plots
        strip.background = element_rect(color="black", fill="#93C5FF", size=1.5, linetype="solid"),
        #legend.position = "none",
```

```
      plot.title = element_text(hjust = 0.5),
  )

# Theme
theme_bar <-  theme_bw() +
  theme(
    plot.title = element_text(hjust = 0.5),  # center title
    panel.background = element_blank(),
    panel.grid.major = element_blank(),
    panel.grid.minor = element_blank(),
    axis.ticks = element_blank(),
    axis.text.x = element_blank(),
    axis.title.x = element_blank(),
    panel.spacing = unit(.9, 'lines'),
    text = element_text(size = 20)
  )

INC1_colors <- c('SOIL' = '#7E6148FF', 'CS1' = '#91D1C2FF', 'AD1' = '#8491B4FF', 'HLFB1' =
'#F39B7FFF')
INC2_colors <- c('SOIL' = '#7E6148FF', 'CS2' = '#00A087FF', 'AD2' = '#3C5488FF', 'HLFB1' =
'#F39B7FFF', 'HLFB2' = '#E64B35FF', 'HLFB3' = '#DC0000FF')
INCtot_colors <- c('SOIL' = '#7E6148FF', 'CS1' = '#91D1C2FF', 'AD1' = '#8491B4FF', 'CS2' =
'#00A087FF', 'AD2' = '#3C5488FF', 'HLFB1' = '#F39B7FFF', 'HLFB2' = '#E64B35FF', 'HLFB3' =
'#DC0000FF')
################################################################################
####################
# ONE TIME INPUT GRAPHS

# Initials
ADconv = .5
DASEconv = .35

# Read in data
onetime_data0 <- read.csv("onetime_data.csv", stringsAsFactors = FALSE, header = TRUE) #
scan in document formatted like example

onetime_data0 <- onetime_data0 %>%
  #select(-'X', -'X.1') %>% # get rid of weird extra column
  select(-'GWC20', -'CCBP_P', -'CCBP_V', -'DASE_C', -'DASE_O')

# total treatments: c('DASE_C',  'DASE_O',        'DASE_AVG',    'AD_S', 'POET_S',
            'NREL_S',         'AD_N', 'POET_N',        'NREL_N',         'CS_N', 'GWC16',
            'GWC20',          'PALOUSE',      'CS_1P',          'AD_1P',          'CCBP_P',
            'DASE_1P',        'VERSHIRE',     'CS_1V',          'AD_1V',          'CCBP_V',
            'DASE_1V')

CinitsINC2 <- c((1199.218125+1198.987771)/2, 1202.890795, 1208.769544, 1280.327308,
869.183259, 871.99067, 907.076619, 880.866037, 540.873971) # these numbers reflect if I
average C per treatment, Information from INC3 -> CombinedIRMS -> Treatment_Calculations
# CinitsINC2 key = 'DASE_AVG',          'AD_S', 'POET_S',        'NREL_S',         'AD_N',
            'POET_N',         'NREL_N',         'CS_N', 'GWC16'

CinitsP <- c(514.4336596, 1145.656643, 1059.347782, 1188.126723) # these numbers reflect if I
average C per treatment, Information from INC2 -> IRMS -> "IRMS_summary" -> IRMS_Pre
CinitsV <- c(1113.366093, 1752.370126, 1651.682688, 1783.200554)
CinitsINC1 <- c(CinitsP, CinitsV)
```

```
# CinitsINC1 key =  'PALOUSE', 'CS_1P',          'AD_1P',          'DASE_1P',       'VERSHIRE',
          'CS_1V',          'AD_1V',          'DASE_1V'

onetime_data <- onetime_data0 %>%
  pivot_longer(
    cols = c('DASE_AVG',          'AD_S', 'POET_S',        'NREL_S',        'AD_N', 'POET_N',
          'NREL_N',        'CS_N', 'GWC16',          'PALOUSE',      'CS_1P',          'AD_1P',
          'DASE_1P',       'VERSHIRE',     'CS_1V',          'AD_1V',          'DASE_1V'),
    names_to = 'treatment',
    values_to = 'cummCO2resp'
  )

# ONE TIME INPUT INC1
onetime_data_INC1 <- onetime_data %>%
  filter(treatment == 'PALOUSE' | treatment ==     'CS_1P' | treatment == 'AD_1P' | treatment ==
          'DASE_1P' | treatment ==          'VERSHIRE' | treatment ==        'CS_1V' | treatment ==
          'AD_1V' | treatment ==   'DASE_1V') %>%
  mutate(soil = ifelse(treatment == c('PALOUSE', 'CS_1P',          'AD_1P',          'DASE_1P'), 'P',
'V')) %>%
  mutate(type = case_when(treatment == 'PALOUSE' | treatment == 'VERSHIRE' ~ 'SOIL',
                  treatment == 'CS_1P' | treatment == 'CS_1V' ~ 'CS1',
                  treatment == 'AD_1P' | treatment == 'AD_1V' ~ 'AD1',
                  treatment == 'DASE_1P' | treatment == 'DASE_1V' ~ 'HLFB1')) %>%
  mutate(cummCO2respCONV = case_when(type == 'SOIL' ~ cummCO2resp,
                               type == 'CS1' ~ cummCO2resp,
                               type == 'AD1' ~ ADconv*cummCO2resp,
                               type == 'HLFB1' ~ DASEconv*cummCO2resp)) %>%
  mutate(soil = factor(soil, labels = c('Palouse', 'Vershire'))) %>%
  mutate(init_totalC = case_when(treatment == 'PALOUSE' ~ CinitsINC1[1],
                           treatment == 'CS_1P' ~ CinitsINC1[2] ,
                           treatment == 'AD_1P' ~ CinitsINC1[3],
                           treatment == 'DASE_1P' ~ CinitsINC1[4],
                           treatment == 'VERSHIRE' ~ CinitsINC1[5],
                           treatment == 'CS_1V' ~ CinitsINC1[6],
                           treatment == 'AD_1V' ~ CinitsINC1[7],
                           treatment == 'DASE_1V' ~ CinitsINC1[8])) %>%
  mutate(Cret = init_totalC - cummCO2resp) %>%
  mutate(init_totalCCONV = case_when(type == 'SOIL' ~ init_totalC,
                               type == 'CS1' ~ init_totalC,
                               type == 'AD1' ~ ADconv*init_totalC,
                               type == 'HLFB1' ~ DASEconv*init_totalC)) %>%
  mutate(CretCONV = init_totalCCONV - cummCO2respCONV)

data_ends <- onetime_data_INC1 %>%
  group_by(treatment) %>%
  top_n(1, YEAR)

#to isolate each soil type use this code in ggplot data =
#onetime_data_INC1 %>%
 # filter(soil == 'Palouse')

onetime_INC1_plot <- ggplot(data = onetime_data_INC1, aes(x = YEAR, y = cummCO2resp,
group = treatment)) +
  geom_line(aes(col = type, linetype = factor(soil)), size = 2) +
  scale_linetype_manual(values = c('Palouse' = 'solid', 'Vershire' = 'dotdash')) +
```

```
  scale_color_manual(values = INC1_colors) + #c('SOIL' = '#7E6148FF', 'CS2' = '#91D1C2FF',
'AD2' = '#8491B4FF', 'DASE1' = '#F39B7FFF', 'DASE2' = '#E64B35FF', 'DASE3' = '#DC0000FF'))
+
  theme_C +
  #xlim(0, 25) +
  labs(x = 'Years', y = 'Total C Respired [mg]',
       #title = '100 Year Projections of C Respired in Incubation 1 Treatments',
       col = 'Treatment', linetype = 'Soil Type')
onetime_INC1_plot
ggsave("onetimeINC1_plot.png", plot = onetime_INC1_plot, width = 30, height = 20, units = "cm")
# change this accordingly

# residue conversion yield accounted for
onetimeCONV_INC1_plot <- ggplot(data = onetime_data_INC1 , aes(x = YEAR, y =
cummCO2respCONV, group = treatment)) +
  geom_line(aes(col = type, linetype = soil), size = 2) +
  scale_color_manual(values = INC1_colors) +
  theme_C +
 # xlim(0, 25) +
  labs(x = 'Years', y = 'Total C Respired [mg]',
       # title = '100 Year Projections of C Respired \n in Incubation 1 Treatments with Conversion
Rates',
       col = 'Treatment', linetype = 'Soil Type')
onetimeCONV_INC1_plot
ggsave("VERonetimeCONV_INC1_plot.png", plot = onetimeCONV_INC1_plot, width = 30, height
= 20, units = "cm")  # change this accordingly

# INC1 Retained version of above plots
onetimeRET_INC1_plot <- ggplot(data = onetime_data_INC1, aes(x = YEAR, y = Cret, group =
treatment)) +
  geom_line(aes(col = type, linetype = factor(soil)), size = 2) +
  scale_color_manual(values = INC1_colors) +
  theme_C +
  xlim(0, 50) +
  labs(x = 'Years', y = 'Total C Retained [mg]', title = '100 Year Projections of C Retained in
Incubation 1 Treatments', col = 'Treatment', linetype = 'Soil Type')
onetimeRET_INC1_plot
ggsave("onetimeRET50_INC1_plot.png", plot = onetimeRET_INC1_plot, width = 30, height = 20,
units = "cm")  # change this accordingly

# residue conversion yield accounted for
onetimeRET_CONV_INC1_plot <- ggplot(data = onetime_data_INC1, aes(x = YEAR, y =
CretCONV, group = treatment)) +
  geom_line(aes(col = type, linetype = soil), size = 2) +
  scale_color_manual(values = INC1_colors) +
  theme_C +
  xlim(0, 50) +
  labs(x = 'Years', y = 'Total C Retained [mg]', title = '100 Year Projections of C Retained \n in
Incubation 1 Treatments with Conversion Rates', col = 'Treatment', linetype = 'Soil Type')
onetimeRET_CONV_INC1_plot
ggsave("onetimeRET50_CONV_INC1_plot.png", plot = onetimeRET_CONV_INC1_plot, width =
30, height = 20, units = "cm")  # change this accordingly

# ONE TIME INPUT INC2
# Nature color palette
#E64B35FF # orange red
```

```
#F39B7FFF # salmon red
#DC0000FF # deep red

onetime_data_INC2 <- onetime_data %>%
  filter(treatment == 'DASE_AVG' | treatment ==  'AD_S' | treatment == 'POET_S' | treatment ==
           'NREL_S' | treatment =='AD_N' | treatment ==     'POET_N' | treatment ==
           'NREL_N' | treatment ==            'CS_N' | treatment ==     'GWC16') %>%
  mutate(dose = case_when(treatment == 'GWC16' ~ 'SOIL',
                          treatment == 'CS_N' | treatment == 'AD_N'
                          | treatment == 'POET_N' | treatment == 'NREL_N' ~ 'N',
                          treatment == 'DASE_AVG' | treatment == 'AD_S'
                          | treatment == 'POET_S' | treatment == 'NREL_S' ~ 'S')) %>%
  mutate(type = case_when(treatment == 'GWC16' ~ 'SOIL',
                          treatment == 'CS_N' ~ 'CS',
                          treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD',
                          treatment == 'DASE_AVG' | treatment == 'POET_N' | treatment == 'NREL_N' |
treatment == 'POET_S' | treatment == 'NREL_S' ~ 'DASE')) %>%
  mutate(type2 = case_when(treatment == 'GWC16' ~ 'SOIL',
                          treatment == 'CS_N' ~ 'CS2',
                          treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD2',
                          treatment == 'DASE_AVG' ~ 'HLFB1',
                          treatment == 'POET_N' | treatment == 'POET_S' ~ 'HLFB3',
                          treatment == 'NREL_N' | treatment == 'NREL_S' ~ 'HLFB2')) %>%
  mutate(cummCO2respCONV = case_when(type == 'SOIL' ~ cummCO2resp,
                                type == 'CS' ~ cummCO2resp,
                                type == 'AD' ~ ADconv*cummCO2resp,
                                type == 'DASE' ~ DASEconv*cummCO2resp)) %>%
  mutate(dose = factor(dose, labels = c('Reduced', 'Standard', 'Soil'))) %>%
  mutate(init_totalC = case_when(treatment == 'DASE_AVG' ~ CinitsINC2[1],
                            treatment == 'AD_S' ~ CinitsINC2[2] ,
                            treatment == 'POET_S' ~ CinitsINC2[3],
                            treatment == 'NREL_S' ~ CinitsINC2[4],
                            treatment == 'AD_N' ~ CinitsINC2[5],
                            treatment == 'POET_N' ~ CinitsINC2[6],
                            treatment == 'NREL_N' ~ CinitsINC2[7],
                            treatment == 'CS_N' ~ CinitsINC2[8],
                            treatment == 'GWC16' ~ CinitsINC2[9])) %>%
  mutate(Cret = init_totalC - cummCO2resp) %>%
  mutate(init_totalCCONV = case_when(type == 'SOIL' ~ init_totalC,
                                type == 'CS' ~ init_totalC,
                                type == 'AD' ~ ADconv*init_totalC,
                                type == 'DASE' ~ DASEconv*init_totalC)) %>%
  mutate(CretCONV = init_totalCCONV - cummCO2respCONV)

# code so only dosage group
# onetime_data_INC2 %>%
#   filter(soil == 'Vershire')

onetime_INC2_plot <- ggplot(data = onetime_data_INC2, aes(x = YEAR, y = cummCO2resp,
group = treatment)) +
  geom_line(aes(linetype = factor(dose), colour = type2), size = 2) +
  scale_linetype_manual(values = c('Standard' = 'solid', 'Reduced' = 'twodash', 'Soil' = 'dotted')) +
  scale_color_manual(values = INC2_colors) +
  theme_C +
  xlim(0, 25) +
  labs(x = 'Years', y = 'Total C Respired [mg]',
```

```
      #title = '25 Year Projections of C Respired in Incubation 2 Treatments',
      linetype = 'Dosage', color = 'Treatment')
onetime_INC2_plot
ggsave("onetimeINC2_plot.png", plot = onetime_INC2_plot, width = 30, height = 20, units = "cm")
# change this accordingly

# residue conversion yield accounted for
onetimeCONV_INC2_plot <- ggplot(data = onetime_data_INC2 , aes(x = YEAR, y =
cummCO2respCONV, group = treatment)) +
  geom_line(aes(linetype = dose, colour = type2), size = 2) +
  scale_linetype_manual(values = c('Standard' = 'solid', 'Reduced' = 'twodash', 'Soil' = 'dotted')) +
  scale_color_manual(values = INC2_colors) +
  theme_C +
  xlim(0, 25.2) +
  labs(x = 'Years', y = 'Total C Respired [mg]',
      # title = '25 Year Projections of C Respired \n in Incubation 2 Treatments with Conversion
Rates',
      linetype = 'Dosage', col = 'Treatment')
onetimeCONV_INC2_plot
ggsave("onetimeCONV_INC2_plot.png", plot = onetimeCONV_INC2_plot, width = 30, height =
20, units = "cm")  # change this accordingly

# INC2 RETAINED VERSION OF ABOVE PLOTS
onetimeRET_INC2_plot <- ggplot(data = onetime_data_INC2, aes(x = YEAR, y = Cret, group =
treatment)) +
  geom_line(aes(linetype = factor(dose), colour = type2), size = 2) +
  scale_color_manual(values = INC2_colors) +
  theme_C +
  xlim(0, 25) +
  labs(x = 'Years', y = 'Total C Retained [mg]', title = '25 Year Projections of C Retained in
Incubation 2 Treatments', linetype = 'Dosage', color = 'Treatment')
onetimeRET_INC2_plot
ggsave("onetimeRET_INC2_plot.png", plot = onetimeRET_INC2_plot, width = 30, height = 20,
units = "cm")  # change this accordingly

# residue conversion yield accounted for
onetimeRET_CONV_INC2_plot <- ggplot(data = onetime_data_INC2, aes(x = YEAR, y =
CretCONV, group = treatment)) +
  geom_line(aes(linetype = dose, colour = type2), size = 2) +
  scale_color_manual(values = INC2_colors) +
  theme_C +
  xlim(0, 25.2) +
  labs(x = 'Years', y = 'Total C Respired [mg]', title = '25 Year Projections of C Respired \n in
Incubation 2 Treatments with Conversion Rates', linetype = 'Dosage', col = 'Treatment')
onetimeRET_CONV_INC2_plot
ggsave("onetimeRET_CONV_INC2_plot.png", plot = onetimeRET_CONV_INC2_plot, width = 30,
height = 20, units = "cm")  # change this accordingly


##########################################################################################
#######################
# LONGTERM MODELLING GRAPHS
longterm_data0 <- read.csv("Longterm_data.csv", stringsAsFactors = FALSE, header = TRUE) #
scan in document formatted like example

longterm_data0 <- longterm_data0 %>%
```

```
  select(-'CCBP_P', -'CCBP_V', -'GWC16', -'PALOUSE', -'VERSHIRE')

# c('DASE_C',   'DASE_O',        'DASE_AVG',    'AD_S', 'POET_S',        'NREL_S',        'AD_N',
           'POET_N',        'NREL_N',        'CS_N', 'GWC16',         'GWC20',         'PALOUSE',
           'CS_1P',         'AD_1P',         'CCBP_P',        'DASE_1P',      'VERSHIRE',
           'CS_1V',         'AD_1V',         'CCBP_V',        'DASE_1V')

longterm_data <- longterm_data0 %>%
  pivot_longer(
    cols = c('DASE_AVG',          'AD_S', 'POET_S',        'NREL_S',        'AD_N', 'POET_N',
           'NREL_N',        'CS_N', 'CS_1P',         'AD_1P',         'DASE_1P',      'CS_1V',
           'AD_1V',         'DASE_1V'),
    names_to = 'treatment',
    values_to = 'cummCO2resp'
  )

# longterm INC1
longterm_data_INC1 <- longterm_data %>%
  filter(treatment ==        'CS_1P' | treatment == 'AD_1P' | treatment ==    'DASE_1P' | treatment
==        'CS_1V' | treatment ==   'AD_1V' | treatment ==   'DASE_1V') %>%
  mutate(soil = ifelse(treatment == c('CS_1P',      'AD_1P',          'DASE_1P'), 'P', 'V')) %>%
  mutate(type = case_when(treatment == 'CS_1P' | treatment == 'CS_1V' ~ 'CS1',
                          treatment == 'AD_1P' | treatment == 'AD_1V' ~ 'AD1',
                          treatment == 'DASE_1P' | treatment == 'DASE_1V' ~ 'HLFB1')) %>%
  mutate(cummCO2respCONV = case_when(type == 'CS1' ~ cummCO2resp,
                                     type == 'AD1' ~ ADconv*cummCO2resp,
                                     type == 'HLFB1' ~ DASEconv*cummCO2resp)) %>%
  mutate(soil = factor(soil, labels = c('Palouse', 'Vershire')))

data_ends <- onetime_data_INC1 %>%
  group_by(treatment) %>%
  top_n(1, YEAR)

longterm_INC1_plot <- ggplot(data = longterm_data_INC1,
                             # %>% filter(soil == 'Vershire')   ,
                              aes(x = YEAR, y = cummCO2resp, group = treatment)) +
  geom_smooth(aes(col = type, linetype = soil), size = 2, alpha = 0) +
  scale_linetype_manual(values = c('Palouse' = 'solid', 'Vershire' = 'dotdash')) +
  scale_color_manual(values = INC1_colors) +
  theme_C +
  coord_trans( y="log2") +  # otherwise DASE1 overwhelms plot
  #xlim(0, 25) +
  labs(x = 'Years', y = 'Total C Retained in Treatments \n log2([mg C])',
       #title = '100 Year Steady State Projections \n of C Retained in Incubation 1 Treatments',
       linetype = 'Soil Type', color = 'Treatment')
longterm_INC1_plot
ggsave("VERlongtermINC1_plot.png", plot = longterm_INC1_plot, width = 30, height = 20, units =
"cm")  # change this accordingly

# check INC1 data
ggp <- ggplot(longterm_data_INC1, aes(x = YEAR, y = cummCO2resp, group = treatment)) +
  stat_smooth(aes(col = treatment))
ggp
ggp_data <- ggplot_build(ggp)
head(ggp_data$data[[1]])
```

```
write.csv(ggp_data$data[[1]], 'inc1ggp_data.csv') # spits out data in color order, so graph it to see
what number corresponds w/ what
# palouse, standard -> 1: AD1, 3: CS1, 5: HLFB1
# vershire, standard -> 2: AD1, 4: CS1, 6: HLFB1

# longterm INC2
longterm_data_INC2 <- longterm_data %>%
  filter(treatment == 'DASE_AVG' | treatment ==  'AD_S' | treatment == 'POET_S' | treatment ==
           'NREL_S' | treatment =='AD_N' | treatment ==     'POET_N' | treatment ==
           'NREL_N' | treatment ==             'CS_N' | treatment ==     'GWC16') %>%
  mutate(dose = case_when(treatment == 'CS_N' | treatment == 'AD_N'
                      | treatment == 'POET_N' | treatment == 'NREL_N' ~ 'N',
                      treatment == 'DASE_AVG' | treatment == 'AD_S'
                      | treatment == 'POET_S' | treatment == 'NREL_S' ~ 'S')) %>%
  mutate(type = case_when(treatment == 'CS_N' ~ 'CS',
                      treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD',
                      treatment == 'DASE_AVG' | treatment == 'POET_N' | treatment == 'NREL_N' |
treatment == 'POET_S' | treatment == 'NREL_S' ~ 'DASE')) %>%
  mutate(type2 = case_when(treatment == 'CS_N' ~ 'CS2',
                       treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD2',
                       treatment == 'DASE_AVG' ~ 'HLFB1',
                       treatment == 'POET_N' | treatment == 'POET_S' ~ 'HLFB3',
                       treatment == 'NREL_N' | treatment == 'NREL_S' ~ 'HLFB2')) %>%
  mutate(cummCO2respCONV = case_when(type == 'CS' ~ cummCO2resp,
                                type == 'AD' ~ ADconv*cummCO2resp,
                                type == 'DASE' ~ DASEconv*cummCO2resp)) %>%
  mutate(dose = factor(dose, labels = c('Reduced', 'Standard')))

longterm_INC2_plot <- ggplot(data = longterm_data_INC2
                         %>% filter(dose == 'Reduced'),
                         aes(x = YEAR, y = cummCO2resp, group = treatment)) +
  scale_linetype_manual(values = c('Standard' = 'solid', 'Reduced' = 'twodash')) +
  geom_smooth(aes(linetype = dose, colour = type2), size = 2, alpha = 0) +
  scale_color_manual(values = INC2_colors) +
  theme_C +
  #coord_trans( y="log2") +  # otherwise DASE3 overwhelms plot
  labs(x = 'Years', y = 'Total C Retained in Treatments [mg C]',
      #title = '100 Year Steady State Projections \n of C Retained in Incubation 2 Treatments',
      linetype = 'Dosage', color = 'Treatment')
longterm_INC2_plot
ggsave("REDlongtermINC2_plot.png", plot = longterm_INC2_plot, width = 30, height = 20, units =
"cm")  # change this accordingly

# output geom_smooth data to check SS values
ggp <- ggplot(longterm_data_INC2, aes(x = YEAR, y = cummCO2resp, group = treatment)) +
  stat_smooth(aes(col = treatment))
ggp
ggp_data <- ggplot_build(ggp)
head(ggp_data$data[[1]])

write.csv(ggp_data$data[[1]], 'inc2ggp_data.csv') # spits out data in color order, so graph it to see
what number corresponds w/ what
# reduced -> groups 1: AD2, 3: CS2, 5: HLFB2, 7: HLFB3

##############################################################################
#######################
```

```
# 13C PARTITIONING

# Read in data
thirteenC_data0 <- read.csv("thirteenC_data.csv", stringsAsFactors = FALSE, header = TRUE) #
scan in document formatted like example

thirteenC_data <- thirteenC_data0 %>%
  filter(treatment != 'CCBP_1P' & treatment != 'CCBP_1V' & treatment != 'TWENTY' ) %>% #&
treatment != 'ONESIX' & treatment != 'PALOUSE' & treatment != 'VERSHIRE'
  #select(-'X', -'X.1', -'X.2', -'X.3', -'X.4') %>%
  filter(sample != 'V5C' & sample != 'V5A' & sample != 'P3B' & sample != 'V1A') %>%
  group_by(treatment, time, source) %>%
  summarize_all(list(mean, sd)) %>%
  select(-'num_fn1', -'sample_fn1', -'rep_fn1', -'num_fn2', -'sample_fn2', -'rep_fn2', -'inc_fn2')
#%>%
  #mutate(source = factor(source, levels = c('soil', 'res')))

thirteenC_err_data0 <- read.csv("thirteenC_err_data.csv", stringsAsFactors = FALSE, header =
TRUE) # scan in document formatted like example

thirteenC_err_data <- thirteenC_err_data0 %>%
  filter(treatment != 'CCBP_1P', treatment != 'CCBP_1V', treatment != 'ONESIX', treatment !=
'TWENTY', treatment != 'PALOUSE', treatment != 'VERSHIRE') %>%
  filter(sample != 'V5C' & sample != 'V5A' & sample != 'P3B' & sample != 'V1A') %>%
  group_by(treatment, time) %>%
  summarize_all(list(mean, sd)) %>%
  select(-'num_fn1', -'sample_fn1', -'rep_fn1', -'num_fn2', -'sample_fn2', -'rep_fn2', -'inc_fn2')

# priming data
priming_data <- thirteenC_data %>%
  mutate(amt_fn1 = 1000*amt_fn1) %>%
  mutate(amt_fn2 = 1000*amt_fn2) %>%
  group_by(treatment, source) %>%
  arrange(time) %>%
  mutate(diff = amt_fn1 - lag(amt_fn1, default = first(amt_fn1))) %>%
  mutate(diff = -1*diff) %>%
  mutate(stdev = sqrt(amt_fn2^2 + (lag(amt_fn2, default = first(amt_fn2)))^2)) %>% # standard
deviation
  filter(diff != 0) %>%
  mutate(diff = ifelse(diff < 0, 0, diff))  # if priming says somehow soil gained carbon from
incubation, correct to 0



#write.csv(priming_data, file="priming_data.csv", row.names = FALSE)

# PRIMING STATISTICS
loss_data <- thirteenC_data0 %>%
  filter(treatment != 'CCBP_1P' & treatment != 'CCBP_1V' & treatment != 'ONESIX' & treatment !=
'TWENTY' & treatment != 'PALOUSE' & treatment != 'VERSHIRE') %>%
  #select(-'X', -'X.1', -'X.2', -'X.3', -'X.4') %>%
  filter(sample != 'V5C' & sample != 'V5A' & sample != 'P3B' & sample != 'V1A') %>%
  mutate(amt = 1000*amt) %>%
  group_by(sample, source) %>%
  arrange(time) %>%
  mutate(diff = amt - lag(amt, default = first(amt))) %>%
```

```
  mutate(diff = -1*diff) %>%
  filter(diff != 0) %>%
  mutate(diff = ifelse(diff < 0, 0, diff)) #if priming says somehow soil gained carbon, correct to 0


priming_stats_data <- loss_data %>%
  filter(source == 'soil')

res_stats_data <- loss_data %>%
  filter(source == 'res')

# SOIL LOSS, INC 1, SOIL TYPE = PALOUSE
res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter(inc == '1' &
soil_type == 'P')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III') # use because of unbalanced design
TukeyHSD(res.aov_priming)

# SOIL LOSS, INC 1, SOIL TYPE = VERSHIRE
res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter(inc == '1' &
soil_type == 'V')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)

# SOIL LOSS, INC 2, NEW/REDUCED
res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter(inc == '2' & dose
== 'N')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)

# SOIL LOSS, INC 2, STANDARD
res.aov_priming <- aov(diff ~ treatment, data = priming_stats_data %>% filter(inc == '2' & dose
== 'S')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)

##############################################################################
####################################

# RES LOSS, INC 1, SOIL TYPE = PALOUSE
res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc == '1' & soil_type ==
'P')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)

# RES LOSS, INC 1, SOIL TYPE = VERSHIRE
res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc == '1' & soil_type ==
'V')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)

# RES LOSS, INC 2, NEW/REDUCED
res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc == '2' & dose ==
'N')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)
```

```
# RES LOSS, INC 2, STANDARD
res.aov_priming <- aov(diff ~ treatment, data = res_stats_data %>% filter(inc == '2' & dose ==
'S')) # test interaction btwn Num and Typ
Anova(res.aov_priming, type = 'III')
TukeyHSD(res.aov_priming)


# 13C INC1
thirteenC_data_INC1 <- thirteenC_data %>%
  filter(inc_fn1 == '1') %>%
  #filter(treatment ==       'CS_1P' | treatment == 'AD_1P' | treatment ==     'DASE_1P' | treatment
==         'CS_1V' | treatment ==   'AD_1V' | treatment ==   'DASE_1V') %>%
  mutate(soil = ifelse(treatment == 'CS_1P' | treatment ==             'AD_1P' | treatment ==
           'DASE_1P', 'P', 'V')) %>%
  mutate(type = case_when(treatment == 'CS_1P' | treatment == 'CS_1V' ~ 'CS1',
                          treatment == 'AD_1P' | treatment == 'AD_1V' ~ 'AD1',
                          treatment == 'DASE_1P' | treatment == 'DASE_1V' ~ 'HLFB1')) %>%
  mutate(amt_fn1 = 1000*amt_fn1) %>%
  mutate(amt_fn2 = 1000*amt_fn2)  %>%
  mutate(err_max = amt_fn1+amt_fn2) %>%
  mutate(err_min = amt_fn1-amt_fn2) %>%
  mutate(soil = factor(soil, levels = c('P', 'V'), labels = c('Palouse', 'Vershire')))

thirteenC_err_data_INC1 <- thirteenC_err_data %>%
  filter(inc_fn1 == '1') %>%
  #filter(treatment ==       'CS_1P' | treatment == 'AD_1P' | treatment ==     'DASE_1P' | treatment
==         'CS_1V' | treatment ==   'AD_1V' | treatment ==   'DASE_1V') %>%
  mutate(soil = ifelse(treatment == 'CS_1P' | treatment ==             'AD_1P' | treatment ==
           'DASE_1P', 'P', 'V')) %>%
  mutate(type = case_when(treatment == 'CS_1P' | treatment == 'CS_1V' ~ 'CS1',
                          treatment == 'AD_1P' | treatment == 'AD_1V' ~ 'AD1',
                          treatment == 'DASE_1P' | treatment == 'DASE_1V' ~ 'HLFB1')) %>%
  mutate(resC_fn1 = 1000*resC_fn1) %>%
  mutate(soilC_fn1 = 1000*soilC_fn1) %>%
  mutate(resC_fn2 = 1000*resC_fn2) %>%
  mutate(soilC_fn2 = 1000*soilC_fn2)

thirteenC_data_INC1 <- merge(x = thirteenC_data_INC1, y = thirteenC_err_data_INC1, by =
c('treatment', 'time'))
thirteenC_data_INC1 <- thirteenC_data_INC1 %>%
  mutate(v_adj = ifelse(source == 'soil', 0, soilC_fn1))  #this needs to be amt_fn1 of the other
source, 0))

thirteenC_INC1_plot <- ggplot(thirteenC_data_INC1, aes(fill=factor(source, levels = c('soil', 'res')),
y=amt_fn1, x=time)) +
  geom_bar(position = position_stack(reverse = TRUE), stat="identity") +
  scale_y_continuous(expand = c(0,0),
                     limits = c(0,2000)) +
  geom_errorbar(aes(ymin=err_min+v_adj, ymax=err_max+v_adj), col = 'black', width=.2, position
= 'identity') + # in aes(col = factor(source, levels = c('soil', 'res'))) to check if err bars are on right
bars
              #position=position_dodge(.9))  +
  facet_grid(soil.x ~ type.x) +
  #scale_fill_discrete(limits = c("res", "source"), labels = c("Residue", "Soil")) +
  scale_fill_npg( labels = c("Soil", "Residue")) +
  labs(y = 'C in Treatments [mg C]',
```

```
    # title = 'C Partitioning of Initial and Remaining C in Treatments Pre and Post Incubation 1',
     fill = 'Source') +
  theme_bar
  #geom_text(aes(label = amt_fn2))
thirteenC_INC1_plot
ggsave("thirteenC_INC1_plot.png", plot = thirteenC_INC1_plot, width = 30, height = 15, units =
"cm")  # change this accordingly

# priming INC1
priming_INC1 <- priming_data %>%
  filter(inc_fn1 == '1') %>%
  mutate(soil = ifelse(treatment == 'CS_1P' | treatment ==            'AD_1P' | treatment ==
           'DASE_1P' | treatment == 'PALOUSE', 'P', 'V')) %>%
  mutate(type = case_when(treatment == 'CS_1P' | treatment == 'CS_1V' ~ 'CS1',
                          treatment == 'AD_1P' | treatment == 'AD_1V' ~ 'AD1',
                          treatment == 'DASE_1P' | treatment == 'DASE_1V' ~ 'HLFB1')) %>%
  filter(source == 'soil') %>%
  mutate(soil = factor(soil, levels = c('P', 'V'), labels = c('Palouse', 'Vershire'))) %>%
  rename(old_diff = diff) %>%
  mutate(control = ifelse(soil == 'Palouse',            21, 0)) %>% # values here from diff column of
priming data of PALOUSE and VERSHIRE
  mutate(diff = old_diff-control) %>%
  filter(treatment != 'PALOUSE', treatment != 'VERSHIRE')

priming_INC1_plot <- ggplot(priming_INC1, aes(x = type, y = diff, fill = type)) + # change "diff" to
"old_diff" if you want to just see soil derived losses
  geom_bar(stat = 'identity') +
  geom_errorbar(aes(ymin=diff-stdev, ymax=diff+stdev), col = 'black', width=.2, position =
'identity') +
  theme_bar +
  labs(y = 'Soil Priming [mg]', fill = 'Substrate') +
  #labs(y = 'Soil Derived Carbon Loss [mg]', fill = 'Substrate') +
  #theme(axis.text.x = element_text()) +
  facet_grid(soil ~ type, scales = "free_x") +
  scale_fill_manual(values = INC1_colors)
priming_INC1_plot
ggsave("adj_priming_INC1_plot.png", plot = priming_INC1_plot , width = 30, height = 15, units =
"cm")  # change this accordingly

# RESIDUE LOSS
priming_INC1 <- priming_data %>%
  filter(inc_fn1 == '1') %>%
  mutate(soil = ifelse(treatment == 'CS_1P' | treatment ==            'AD_1P' | treatment ==
           'DASE_1P', 'P', 'V')) %>%
  mutate(type = case_when(treatment == 'CS_1P' | treatment == 'CS_1V' ~ 'CS1',
                          treatment == 'AD_1P' | treatment == 'AD_1V' ~ 'AD1',
                          treatment == 'DASE_1P' | treatment == 'DASE_1V' ~ 'HLFB1')) %>%
  filter(source == 'res') %>%
  mutate(soil = factor(soil, levels = c('P', 'V'), labels = c('Palouse', 'Vershire')))

priming_INC1_plot <- ggplot(priming_INC1, aes(x = type, y = diff, fill = type)) +
  geom_bar(stat = 'identity') +
  geom_errorbar(aes(ymin=diff-stdev, ymax=diff+stdev), col = 'black', width=.2, position =
'identity') +
  theme_bar +
  labs(y = 'Residue Derived Carbon Loss [mg]', fill = 'Substrate') +
```

```
  #theme(axis.text.x = element_text()) +
  facet_grid(soil ~ type, scales = "free_x") +
  scale_fill_manual(values = INC1_colors)
priming_INC1_plot
ggsave("resloss_INC1_plot.png", plot = priming_INC1_plot , width = 30, height = 15, units = "cm")
# change this accordingly


# 13C INC2
thirteenC_data_INC2 <- thirteenC_data %>%
  filter(inc_fn1 == '2') %>%
  mutate(dose = case_when(treatment == 'CS_N' | treatment == 'AD_N'
                          | treatment == 'POET_N' | treatment == 'NREL_N' ~ 'N',
                          treatment == 'DASE_AVG' | treatment == 'AD_S'
                          | treatment == 'POET_S' | treatment == 'NREL_S' ~ 'S')) %>%
  mutate(type = case_when(treatment == 'CS_N' ~ 'CS',
                          treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD',
                          treatment == 'DASE_AVG' | treatment == 'POET_N' | treatment == 'NREL_N' |
treatment == 'POET_S' | treatment == 'NREL_S' ~ 'DASE')) %>%
  mutate(type2 = case_when(treatment == 'CS_N' ~ 'CS2',
                           treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD2',
                           treatment == 'DASE_AVG' ~ 'HLFB1',
                           treatment == 'POET_N' | treatment == 'POET_S' ~ 'HLFB3',
                           treatment == 'NREL_N' | treatment == 'NREL_S' ~ 'HLFB2')) %>%
  mutate(dose = factor(dose, levels = c('S', 'N'), labels = c('Standard', 'Reduced'))) %>%
  mutate(amt_fn1 = 1000*amt_fn1) %>%
  mutate(amt_fn2 = 1000*amt_fn2)  %>%
  mutate(err_max = amt_fn1+amt_fn2) %>%
  mutate(err_min = amt_fn1-amt_fn2)

thirteenC_err_data_INC2 <- thirteenC_err_data %>%
  filter(inc_fn1 == '2') %>%
  mutate(dose = case_when(treatment == 'CS_N' | treatment == 'AD_N'
                          | treatment == 'POET_N' | treatment == 'NREL_N' ~ 'N',
                          treatment == 'DASE_AVG' | treatment == 'AD_S'
                          | treatment == 'POET_S' | treatment == 'NREL_S' ~ 'S')) %>%
  mutate(type = case_when(treatment == 'CS_N' ~ 'CS',
                          treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD',
                          treatment == 'DASE_AVG' | treatment == 'POET_N' | treatment == 'NREL_N' |
treatment == 'POET_S' | treatment == 'NREL_S' ~ 'DASE')) %>%
  mutate(type2 = case_when(treatment == 'CS_N' ~ 'CS2',
                           treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD2',
                           treatment == 'DASE_AVG' ~ 'HLFB1',
                           treatment == 'POET_N' | treatment == 'POET_S' ~ 'HLFB3',
                           treatment == 'NREL_N' | treatment == 'NREL_S' ~ 'HLFB2')) %>%
  mutate(dose = factor(dose, levels = c('S', 'N'))) %>%
  mutate(resC_fn1 = 1000*resC_fn1) %>%
  mutate(soilC_fn1 = 1000*soilC_fn1) %>%
  mutate(resC_fn2 = 1000*resC_fn2) %>%
  mutate(soilC_fn2 = 1000*soilC_fn2)

thirteenC_data_INC2 <- merge(x = thirteenC_data_INC2, y = thirteenC_err_data_INC2, by =
c('treatment', 'time'))
thirteenC_data_INC2 <- thirteenC_data_INC2 %>%
  mutate(v_adj = ifelse(source == 'soil', 0, soilC_fn1))  #this needs to be amt_fn1 of the other
source, 0))
```

```
thirteenC_INC2_plot <- ggplot(thirteenC_data_INC2, aes(fill=factor(source, levels = c('soil', 'res')),
y=amt_fn1, x=time)) +
  geom_bar(position = position_stack(reverse = TRUE), stat="identity") +
  scale_y_continuous(expand = c(0,0),
                     limits = c(0,1500)) +
  facet_grid(dose.x ~ type2.x) +
  geom_errorbar(aes(ymin=err_min+v_adj, ymax=err_max+v_adj),col = 'black', width=.2, position
= 'identity') + # in aes(col = factor(source, levels = c('soil', 'res'))) to check if err bars are on right
bars
  #scale_fill_discrete(limits = c("res", "source"), labels = c("Residue", "Soil")) +
  scale_fill_npg(labels = c("Soil", "Residue")) +
  labs(y = 'C in Treatments [mg C]',
       #title = 'C Partitioning of Initial and Remaining C in Treatments Pre and Post Incubation 2',
       fill = 'Source') +
  theme_bar
thirteenC_INC2_plot
ggsave("thirteenC_INC2_plot.png", plot = thirteenC_INC2_plot, width = 30, height = 15, units =
"cm")  # change this accordingly

# priming INC2
priming_INC2 <- priming_data %>%
  filter(inc_fn1 == '2') %>%
  mutate(dose = case_when(treatment == 'CS_N' | treatment == 'AD_N'
                          | treatment == 'POET_N' | treatment == 'NREL_N' ~ 'N',
                          treatment == 'DASE_AVG' | treatment == 'AD_S'
                          | treatment == 'POET_S' | treatment == 'NREL_S' ~ 'S')) %>%
  mutate(type = case_when(treatment == 'CS_N' ~ 'CS',
                          treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD',
                          treatment == 'DASE_AVG' | treatment == 'POET_N' | treatment == 'NREL_N' |
treatment == 'POET_S' | treatment == 'NREL_S' ~ 'DASE')) %>%
  mutate(type2 = case_when(treatment == 'CS_N' ~ 'CS2',
                           treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD2',
                           treatment == 'DASE_AVG' ~ 'HLFB1',
                           treatment == 'POET_N' | treatment == 'POET_S' ~ 'HLFB3',
                           treatment == 'NREL_N' | treatment == 'NREL_S' ~ 'HLFB2')) %>%
  mutate(dose = factor(dose, levels = c('S', 'N'), labels = c('Standard', 'Reduced'))) %>%
  filter(source == 'soil') %>%
  rename(old_diff = diff) %>%
  mutate(control = 65) %>%
  mutate(diff = old_diff-control) %>%
  filter(treatment != 'ONESIX')

priming_INC2_plot <- ggplot(priming_INC2, aes(x = type2, y = diff, fill = type2)) + # change "diff"
to "old_diff" if you are interested in just soil derived losses
  geom_bar(stat = 'identity') +
  geom_errorbar(aes(ymin=diff-stdev, ymax=diff+stdev), col = 'black', width=.2, position =
'identity') +
  theme_bar +
  labs(y = 'Soil Priming [mg]', fill = 'Substrate') +
  #labs(y = 'Soil Derived Carbon Loss [mg]', fill = 'Substrate') +
  #theme(axis.text.x = element_text()) +
  facet_grid(dose ~ type2, scales = "free_x") +
  scale_fill_manual(values = INC2_colors)
priming_INC2_plot
```

```
ggsave("adj_priming_INC2_plot.png", plot = priming_INC2_plot , width = 30, height = 15, units =
"cm")  # change this accordingly

# old priming plot w/ new color, no control correction
priming_INC2_plot <- ggplot(priming_INC2, aes(x = type2, y = old_diff, fill = type2)) +
  geom_bar(stat = 'identity') +
  geom_errorbar(aes(ymin=old_diff-stdev, ymax=old_diff+stdev), col = 'black', width=.2, position =
'identity') +
  theme_bar +
  labs(y = 'Soil Derived Carbon Loss [mg]', fill = 'Substrate') +
  #theme(axis.text.x = element_text()) +
  facet_grid(dose ~ type2, scales = "free_x") +
  scale_fill_manual(values = INC2_colors)
priming_INC2_plot
ggsave("priming_INC2_plot.png", plot = priming_INC2_plot , width = 30, height = 15, units =
"cm")  # change this accordingly


# res loss
priming_INC2 <- priming_data %>%
  filter(inc_fn1 == '2') %>%
  mutate(dose = case_when(treatment == 'CS_N' | treatment == 'AD_N'
                          | treatment == 'POET_N' | treatment == 'NREL_N' ~ 'N',
                          treatment == 'DASE_AVG' | treatment == 'AD_S'
                          | treatment == 'POET_S' | treatment == 'NREL_S' ~ 'S')) %>%
  mutate(type = case_when(treatment == 'CS_N' ~ 'CS',
                          treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD',
                          treatment == 'DASE_AVG' | treatment == 'POET_N' | treatment == 'NREL_N' |
treatment == 'POET_S' | treatment == 'NREL_S' ~ 'DASE')) %>%
  mutate(type2 = case_when(treatment == 'CS_N' ~ 'CS2',
                           treatment == 'AD_N' | treatment == 'AD_S' ~ 'AD2',
                           treatment == 'DASE_AVG' ~ 'HLFB1',
                           treatment == 'POET_N' | treatment == 'POET_S' ~ 'HLFB3',
                           treatment == 'NREL_N' | treatment == 'NREL_S' ~ 'HLFB2')) %>%
  mutate(dose = factor(dose, levels = c('S', 'N'), labels = c('Standard', 'Reduced'))) %>%
  filter(source == 'res')

priming_INC2_plot <- ggplot(priming_INC2, aes(x = type2, y = diff, fill = type2)) +
  geom_bar(stat = 'identity') +
  geom_errorbar(aes(ymin=diff-stdev, ymax=diff+stdev), col = 'black', width=.2, position =
'identity') +
  theme_bar +
  labs(y = 'Residue Derived Carbon Loss [mg]', fill = 'Substrate') +
  #theme(axis.text.x = element_text()) +
  facet_grid(dose ~ type2, scales = "free_x") +
  scale_fill_manual(values = INC2_colors)
priming_INC2_plot
ggsave("ressloss_INC2_plot.png", plot = priming_INC2_plot , width = 30, height = 15, units =
"cm")  # change this accordingly

################################################################################
######################
# CORRELATIONS

correlations_data0 <- read.csv("correlations_data.csv", stringsAsFactors = FALSE, header =
TRUE) # scan in document formatted like example
```

```
correlations_data <- correlations_data0 %>%
  mutate(perSol = 100*perSol) %>%
  mutate(Cret_tot = 100*Cret_tot)

theme_point <- theme_bw() + theme(
  text = element_text(size = 20)
)

# C:N plot / analysis
C2N_plot <- ggplot(correlations_data, aes(x = C2N, y = Cret_tot)) +
  geom_point(aes(col = id), size = 2) +
  stat_smooth(method = "lm",
              formula = y ~ x,
              geom = "smooth", alpha = .25) +
  theme_point +
  labs(x = 'C:N of Substrate',
       y = '% of Carbon Retained in \n Treatment Containing Substrate [%]',
       col = 'Substrate') +
  ylim(65, 100) +
  scale_color_manual(values = INCtot_colors)
C2N_plot
ggsave("C2N_plot.png", plot = C2N_plot , width = 20, height = 15, units = "cm")  # change this
accordingly

mod_C2N <- lm(Cret_tot ~ C2N, data = correlations_data)
anova(mod_C2N)
summary(mod_C2N)

# %lignin plot / analysis
lig_plot <- ggplot(correlations_data, aes(x = perLig, y = Cret_tot)) +
  geom_point(aes(col = id), size = 2) +
  stat_smooth(method = "lm",
              formula = y ~ x,
              geom = "smooth", alpha = .25) +
  theme_point +
  labs(x = '% Lignin of Substrate',
       y = '% of Carbon Retained in \n Treatment Containing Substrate [%]',
       col = 'Substrate') +
  ylim(65, 100) +
  scale_color_manual(values = INCtot_colors)
lig_plot
ggsave("lig_plot.png", plot = lig_plot , width = 20, height = 15, units = "cm")  # change this
accordingly

mod_lig <- lm(Cret_tot ~ perLig, data = correlations_data)
anova(mod_lig)
summary(mod_lig)

# %solubilization / analysis
sol_plot <- ggplot(correlations_data, aes(x = perSol, y = Cret_tot)) +
  geom_point(aes(col = id), size = 2) +
  stat_smooth(method = "lm",
              formula = y ~ x,
              geom = "smooth", alpha = .25) +
  theme_point +
```

```
  labs(x = '% Solubilization of Substrate',
       y = '% of Carbon Retained in \n Treatment Containing Substrate [%]',
       col = 'Substrate') +
  ylim(65, 100) +
  scale_color_manual(values = INCtot_colors)
sol_plot
ggsave("sol_plot.png", plot = sol_plot , width = 20, height = 15, units = "cm")  # change this
accordingly

mod_sol <- lm(Cret_tot ~ perSol, data = correlations_data)
anova(mod_sol)
summary(mod_sol)
```

References

1. Trivedi, P., Singh, B. P. & Singh, B. K. Chapter 1 - Soil Carbon: Introduction, Importance, Status, Threat, and Mitigation. in *Soil Carbon Storage* (ed. Singh, B. K.) 1–28 (Academic Press, 2018). doi:10.1016/B978-0-12-812766-7.00001-9.
2. Pribyl, D. W. A critical review of the conventional SOC to SOM conversion factor. *Geoderma* **156**, 75–83 (2010).
3. Schmidt, M. W. I. *et al.* Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).
4. Karlen, D. L. *et al.* Soil Quality: A Concept, Definition, and Framework for Evaluation (A Guest Editorial). *Soil Science Society of America Journal* **61**, 4–10 (1997).
5. Cruse, R. M., Cruse, M. J. & Reicosky, D. C. Soil quality impacts of residue removal for biofuel feedstock. in *Soil Quality and Biofuel Production* 45–62 (CRC Press, 2009).
6. Lal, R. Soil health and carbon management. *Food and Energy Security* **5**, 212–222 (2016).
7. Lal, R. Soil quality impacts of residue removal for bioethanol production. *Soil and Tillage Research* **102**, 233–241 (2009).
8. *Dealing With Corn Stalks #706 (Air Date 10/19/11)*. (2011).
9. Ertl, D. *Sustainable Corn Stover Harvest: A publication of the Iowa Corn Promotion Board*. (2013).
10. Pennington, D. Corn stover: What is its worth? *MSU Extension* https://www.canr.msu.edu/news/corn_stover_what_is_its_worth (2013).

11. To Harvest Stover or Not: Is it Worth it? • farmdoc daily. *farmdoc daily* https://farmdocdaily.illinois.edu/2016/02/to-harvest-stover-or-not-is-it-worth-it.html (2016).
12. Humbird, D. *et al. Process Design and Economics for Biochemical Conversion of Lignocellulosic Biomass to Ethanol: Dilute-Acid Pretreatment and Enzymatic Hydrolysis of Corn Stover*. NREL/TP-5100-47764, 1013269 http://www.osti.gov/servlets/purl/1013269/ (2011) doi:10.2172/1013269.
13. Riahi, K. *et al.* Mitigation pathways compatible with long-term goals. in (Cambridge University Press (CUP), 2022). doi:10.1017/9781009157926.005.
14. Field, J. L. *et al.* Robust paths to net greenhouse gas mitigation and negative emissions via advanced biofuels. *PNAS* **117**, 21968–21977 (2020).
15. Jaspers, B. C., Kuo, P.-C., Amladi, A., van Neerbos, W. & Aravind, P. V. Negative CO2 Emissions for Transportation. *Frontiers in Energy Research* **9**, 151 (2021).
16. Smith, P. *et al.* Biophysical and economic limits to negative CO2 emissions. *Nature Clim Change* **6**, 42–50 (2016).
17. Eisentraut, A. Sustainable Production of Second-Generation Biofuels. *International Energy Agency* 221 (2010).
18. Kramer, D. Whatever happened to cellulosic ethanol? *Physics Today* **75**, 22–24 (2022).
19. Blanco-Canqui, H. & Lal, R. Corn Stover Removal for Expanded Uses Reduces Soil Fertility and Structural Stability. *Soil Science Society of America Journal* **73**, 418–426 (2009).

20. Blanco-Canqui, H. Crop Residue Removal for Bioenergy Reduces Soil Carbon Pools: How Can We Offset Carbon Losses? *Bioenerg. Res.* **6**, 358–371 (2013).
21. Liska, A. J. *et al.* Biofuels from crop residue can reduce soil carbon and increase CO2 emissions. *Nature Clim Change* **4**, 398–401 (2014).
22. Amundson, R. *et al.* Soil science. Soil and human security in the 21st century. *Science* **348**, 1261071 (2015).
23. Koch, A. *et al.* Soil Security: Solving the Global Soil Crisis. *Global Policy* **4**, 434–441 (2013).
24. Lehmann, J. & Kleber, M. The contentious nature of soil organic matter. *Nature* **528**, 60–68 (2015).
25. Sanderman, J., Hengl, T. & Fiske, G. J. Soil carbon debt of 12,000 years of human land use. *Proceedings of the National Academy of Sciences* **114**, 9575–9580 (2017).
26. Borrelli, P. *et al.* An assessment of the global impact of 21st century land use change on soil erosion. *Nat Commun* **8**, 2013 (2017).
27. Rumpel, C. *et al.* The 4p1000 initiative: Opportunities, limitations and challenges for implementing soil organic carbon sequestration as a sustainable development strategy. *Ambio* **49**, 350–360 (2020).
28. Minasny, B. *et al.* Soil Carbon 4 per mille. *Goderma* **292**, 59–86 (2017).
29. Sykes, A. J. *et al.* Characterising the biophysical, economic and social impacts of soil carbon sequestration as a greenhouse gas removal technology. *Global Change Biology* **26**, 1085–1108 (2020).

30. Beauregard, M. S., Hamel, C., Atul-Nayyar & St-Arnaud, M. Long-Term Phosphorus Fertilization Impacts Soil Fungal and Bacterial Diversity but not AM Fungal Community in Alfalfa. *Microb Ecol* **59**, 379–389 (2010).
31. Kim, N., Zabaloy, M. C., Guan, K. & Villamil, M. B. Do cover crops benefit soil microbiome? A meta-analysis of current research. *Soil Biology and Biochemistry* **142**, 107701 (2020).
32. Ogle, S. M. *et al.* Climate and Soil Characteristics Determine Where No-Till Management Can Store Carbon in Soils and Mitigate Greenhouse Gas Emissions. *Sci Rep* **9**, 11665 (2019).
33. Follett, R. Soil management concepts and carbon sequestration in cropland soils. *Soil and Tillage Research* **61**, 77–92 (2001).
34. Huggins, D. R. *et al.* Soil organic C in the tallgrass prairie-derived region of the corn belt: effects of long-term crop management. *Soil and Tillage Research* **47**, 219–234 (1998).
35. Lu, X. A meta-analysis of the effects of crop residue return on crop yields and water use efficiency. *PLOS ONE* **15**, e0231740 (2020).
36. Paustian, K. *et al.* Agricultural soils as a sink to mitigate CO2 emissions. *Soil Use and Management* **13**, 230–244 (1997).
37. Kenney, I. *et al.* Soil and crop response to stover removal from rainfed and irrigated corn. *GCB Bioenergy* **7**, (2013).
38. Jin, V. *et al.* Twelve Years of Stover Removal Increases Soil Erosion Potential without Impacting Yield. *Publications from USDA-ARS / UNL Faculty* (2015).

39. Xu, H. *et al.* A global meta-analysis of soil organic carbon response to corn stover removal. *GCB Bioenergy* **11**, 1215–1233 (2019).

40. Karlen, D. L. *et al.* Multilocation Corn Stover Harvest Effects on Crop Yields and Nutrient Removal. *Bioenerg. Res.* **7**, 528–539 (2014).

41. Blum, W. *et al.* Ecological consequences of biofuels. in *Soil Quality and Biofuel Production* 63–92 (2010).

42. Wilhelm, W. W., Johnson, J. M. F., Karlen, D. L. & Lightle, D. T. Corn Stover to Sustain Soil Organic Carbon Further Constrains Biomass Supply. *Agron.j.* **99**, 1665–1667 (2007).

43. Adler, P. R. *et al.* Integrating biorefinery and farm biogeochemical cycles offsets fossil energy and mitigates soil carbon losses. *Ecological Applications* **25**, 1142–1156 (2015).

44. Pourhashem, G., Adler, P. R., McAloon, A. J. & Spatari, S. Cost and greenhouse gas emission tradeoffs of alternative uses of lignin for second generation ethanol. *Environ. Res. Lett.* **8**, 025021 (2013).

45. Reijnders, L. Ethanol production from crop residues and soil organic carbon. *Resources, Conservation and Recycling* **52**, 653–658 (2008).

46. Kumar, D. & Murthy, G. S. Impact of pretreatment and downstream processing technologies on economics and energy in cellulosic ethanol production. *Biotechnology for Biofuels* **4**, 27 (2011).

47. Kazi, F. K. *et al. Techno-Economic Analysis of Biochemical Scenarios for Production of Cellulosic Ethanol*. NREL/TP-6A2-46588, 982937 http://www.osti.gov/servlets/purl/982937-qA6THq/ (2010) doi:10.2172/982937.

48. Liu, X.-J. A. *et al.* The soil priming effect: Consistent across ecosystems, elusive mechanisms. *Soil Biology and Biochemistry* **140**, 107617 (2020).
49. Kuzyakov, Y., Friedel, J. K. & Stahr, K. Review of mechanisms and quantification of priming effects. *Soil Biology and Biochemistry* **32**, 1485–1498 (2000).
50. Stewart, C. E., Moturi, P., Follett, R. F. & Halvorson, A. D. Lignin biochemistry and soil N determine crop residue decomposition and soil priming. *Biogeochemistry* **124**, 335–351 (2015).
51. Guenet, B., Leloup, J., Raynaud, X., Bardoux, G. & Abbadie, L. Negative priming effect on mineralization in a soil free of vegetation for 80 years. *European Journal of Soil Science* **61**, 384–391 (2010).
52. Gayathri, R. *et al.* A review on biological carbon sequestration: A sustainable solution for a cleaner air environment, less pollution and lower health risks. *Journal of King Saud University - Science* **33**, 101282 (2021).
53. Bhattacharyya, S. S., Ros, G. H., Furtak, K., Iqbal, H. M. N. & Parra-Saldívar, R. Soil carbon sequestration – An interplay between soil microbial community and soil organic matter dynamics. *Science of The Total Environment* **815**, 152928 (2022).
54. Hall, S. J., Huang, W., Timokhin, V. I. & Hammel, K. E. Lignin lags, leads, or limits the decomposition of litter and soil organic carbon. *Ecology* **101**, e03113 (2020).
55. Taylor, B. R., Parkinson, D. & Parsons, W. F. J. Nitrogen and Lignin Content as Predictors of Litter Decay Rates: A Microcosm Test. *Ecology* **70**, 97–104 (1989).
56. Talbot, J. M., Yelle, D. J., Nowick, J. & Treseder, K. K. Litter decay rates are determined by lignin chemistry. *Biogeochemistry* **108**, 279–295 (2012).

57. Austin, A. T. & Ballaré, C. L. Dual role of lignin in plant litter decomposition in terrestrial ecosystems. *Proceedings of the National Academy of Sciences* **107**, 4618–4622 (2010).

58. Wu, J. *et al.* Reduced Lignin Decomposition and Enhanced Soil Organic Carbon Stability by Acid Rain: Evidence from 13C Isotope and 13C NMR Analyses. *Forests* **11**, 1191 (2020).

59. Thevenot, M., Dignac, M.-F. & Rumpel, C. Fate of lignins in soils: A review. *Soil Biology and Biochemistry* **42**, 1200–1211 (2010).

60. Johnson, J., Sharratt, B., Reicosky, D. C. & Lindstrom, M. Impact of High-Lignin Fermentation Byproduct on Soils with Contrasting Organic Carbon Content. *Soil Science Society of America Journal* **71**, (2007).

61. Beghin, R., Guérin, F., Guiresse, M., Kleiber, D. & Scheiner, J. Carbon sequestration in soil amended with anaerobic digested matter. *Soil and Tillage Research* **192**, 87 (2019).

62. Chen, R. *et al.* Decomposition of biogas residues in soil and their effects on microbial growth kinetics and enzyme activities. *Biomass and Bioenergy* **45**, 221–229 (2012).

63. Smith, J. *et al.* What is the potential for biogas digesters to improve soil carbon sequestration in Sub-Saharan Africa? Comparison with other uses of organic residues. *Biomass and Bioenergy* **70**, 73–86 (2014).

64. Iocoli, G. A. *et al.* Towards Sustainable Dairy Production in Argentina: Evaluating Nutrient and CO2 Release from Raw and Processed Farm Waste. *Agronomy* **11**, 2595 (2021).

65. Egene, C. E. *et al.* Solid fraction of separated digestate as soil improver: implications for soil fertility and carbon sequestration. *J Soils Sediments* **21**, 678–688 (2021).
66. Sluiter, A. D. *et al. Determination of Structural Carbohydrates and Lignin in Biomass*. https://www.nrel.gov/docs/gen/fy13/42618.pdf (2008).
67. Wang, M. A Foray into Laboratory Scale Soil Incubations with Corn Stover and High Lignin Fermentation Byproduct. *ENGS 88 Honors Thesis (AB Students)* (2021).
68. R: A language and environment for statistical computing. R Foundation for Statistical Computing. (2020).
69. 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).
70. Langholtz, M. H., Stokes, B. J. & Eaton, L. M. *2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy*. DOE/EE-1440, ORNL/TM-2016/160, 1271651 http://www.osti.gov/servlets/purl/1271651/ (2016) doi:10.2172/1271651.
71. Martin, J. *Project Liberty: Launch of an Integrated Bio-Refinery with Eco-Sustainable and Renewable Technologies. Conversion of Corn Stover Biomass to Bio-Ethanol, Final Report*. https://www.osti.gov/biblio/1866610 (2021) doi:10.2172/1866610.
72. Davis, R. E. *et al. Process Design and Economics for the Conversion of Lignocellulosic Biomass to Hydrocarbon Fuels and Coproducts: 2018 Biochemical Design Case Update; Biochemical Deconstruction and Conversion of Biomass to*

Fuels and Products via Integrated Biorefinery Pathways. NREL/TP--5100-71949, 1483234 http://www.osti.gov/servlets/purl/1483234/ (2018) doi:10.2172/1483234.

73. Kuhn, E. M., Chen, X. & Tucker, M. P. Deacetylation and Mechanical Refining (DMR) and Deacetylation and Dilute Acid (DDA) Pretreatment of Corn Stover, Switchgrass, and a 50:50 Corn Stover/Switchgrass Blend. *ACS Sustainable Chem. Eng.* **8**, 6734–6743 (2020).
74. Bernal, M. P., Sánchez-Monedero, M. A., Paredes, C. & Roig, A. Carbon mineralization from organic wastes at different composting stages during their incubation with soil. *Agriculture, Ecosystems & Environment* **69**, 175–189 (1998).
75. Schievano, A., D'Imporzano, G., Salati, S. & Adani, F. On-field study of anaerobic digestion full-scale plants (part I): an on-field methodology to determine mass, carbon and nutrients balance. *Bioresour Technol* **102**, 7737–7744 (2011).
76. Templeton, D. W., Sluiter, A. D., Hayward, T. K., Hames, B. R. & Thomas, S. R. Assessing corn stover composition and sources of variability via NIRS. *Cellulose* **16**, 621–639 (2009).
77. Yang, B. & Wyman, C. E. Pretreatment: the key to unlocking low-cost cellulosic ethanol. *Biofuels, Bioproducts and Biorefining* **2**, 26–40 (2008).
78. Steinbach, D. *et al.* Steam Explosion Conditions Highly Influence the Biogas Yield of Rice Straw. *Molecules* **24**, 3492 (2019).
79. Wickham, H. ggplot2: Elegant Graphics for Data Analysis. (2016).
80. Manlay, R. J., Feller, C. & Swift, M. J. Historical evolution of soil organic matter concepts and their relationships with the fertility and sustainability of cropping systems. *Agriculture, Ecosystems & Environment* **119**, 217–233 (2007).

81. Lynd, L. *et al.* Liquid Biofuels from Crop Residues with Return of High-Lignin Fermentation Byproduct to the Soil. (2023).