

The impact of forest management on the temperature sensitivity of SOC decomposition in european forests

Lorenzo Menichetti^{1,2*}, Second Author¹, Third Author³

¹Natural Resources Institute Finland (LUKE), Latokartanonkaari 9 00790 Helsinki

²Swedish University of Agricultural Sciences (SLU), Inst för Ekologi, Box 7044 750 07

³[Third Institution]

*Corresponding author: Lorenzo.Menichetti@luke.fi

Abstract

Background: Provide context and rationale for your study. Briefly explain the problem or gap in knowledge that your research addresses.

Methods: Summarize your experimental design, key methods, and analytical approaches used in the study.

Results: Present the main findings of your research, including key quantitative results and statistical significance where applicable.

Conclusions: State the main conclusions and their broader implications. Highlight the significance of your findings and potential applications.

Keywords: keyword1, keyword2, keyword3, keyword4, keyword5

1. Introduction

1.1. Background and Rationale

1.2. Objectives and Hypotheses

Clearly state your research objectives and hypotheses. For example:

- **Primary objective:** To investigate the relationship between X and Y
- **Secondary objectives:** To evaluate Z and assess W
- **Hypothesis:** We hypothesize that X will significantly affect Y under conditions Z

2. Materials and Methods

2.1. Holisoils project introduction

HoliSoils - Holistic management practices, modelling and monitoring for European forest soils, is an EU funded project (Horizon 2020, Grant Agreements number 101000289), completed in October 2026, with the main aim to develop a harmonised soil monitoring framework.

More specifically, the project aimed at developing a consistent knowledge framework of soil properties, processes, biodiversity and soil microbiota activity, in connection with soil-based ecosystem services (wood production, reduction of greenhouse gas emissions, water supply, soil nutrient retention, avoidance of land degradation), particularly in connection with different management approaches. Modeling was another of the focuses, specifically improving current models and harmonise them into a monitoring framework for estimating ecosystem fluxes. Directly connected with the modeling activity, the project developed a set of standardised sampling and monitoring protocols shared among all actors for greenhouse gas (GHG) reporting. These were used across multiple sampling campaigns across three consecutive years in different locations over Europe, and results have been

collected in a comprehensive database of GHG measurements together with additional edaphic and ecological variables.

2.1.1 Actors and nations involved

The project involved a consortium of 20 project partners, 18 from EU and partners from South America (Uruguay) and Asia (Japan):

- Natural Resources Institute Finland (Luonnonvarakeskus), Luke, Finland
- Institute of Microbiology of the Czech Academy of Sciences (Mikrobiologický ústav AV ČR, v. v. i.), IMIC, Czech Republic
- French National Centre for Scientific Research (Centre National de la Recherche Scientifique Research institute), CNRS, France
- Johann Heinrich von Thünen Institute (Johann Heinrich von Thünen-Institut Research institute), TI, Germany
- Basque Centre for Climate Change (Asociacion BC3 Basque Centre for Climate Change Klima Aldaketa Ikergai), BC3, Spain
- Vrije University Amsterdam (Stichting VU University), VU, Netherlands
- European Forest Institute, EFI, Finland
- Wageningen Research Foundation (Stichting Wageningen), WR, Netherlands
- International Soil Reference & Information Centre (Stichting International Soil Reference & Information Centre), ISRIC, Netherlands
- Stockholm University (Stockholms Universitet), SU, Sweden
- Transylvania University of Brașov (Universitatea Transilvania din Brașov), UTBV, Romania
- University of Barcelona (Universitat de Barcelona), UB, Spain

- University of Aberdeen (The University Court of The University of Aberdeen), UNI-ABDN, UK
- Vytautas Magnus University (Vytauto Didziojo universitetas), VMU, Lithuania
- Aix-Marseille University (Aix-Marseille Université), AMU, France
- Technical University of Munich (Technische Universität München), TUM, Germany
- Technical University in Zvolen (Technická univerzita vo Zvolene), TUZVO, Slovakia
- Forest Science & Technology Centre of Catalonia (Centre de Ciència i Tecnologia Forestal de Catalunya), CTFC, Spain
- National Institute for Agricultural Research (Instituto Nacional de Investigación Agropecuaria), INIA, Uruguay
- Forestry & Forest Products Research Institute, FFPRI, Japan

2.1.2 The greenhouse gas management effort and how it connects to other parts of the project

2.2. Sites and study design

2.2.1 Sites and treatments

Trenching In most sites all treatments have both control and trenched subtreatments. The aim of the trenching is to remove all autotrophic respiration from the total soil respiration. Threnching areas were established (on an area of approximately 2 square meters) by digging trenches at a depth of approximately 30-40 cm and then placing root exclusion fabric around the trenched area. The fabric had around 50 μm mesh size, which prevents roots but allows mychorizae to penetrate the area, as well as water.

2.2.2 Sampling

Soil CO₂ from control and trenched locations was measured approximately every second week with closed chambers and gas concentration analyzers (see 2.2.2), basing the flux estimate on the linear regression of the concentration increase over time. Measurement points for the soil chambers have been established with metal or plastic collars, or in some cases (organic soils) by installing the chamber deeper into the soil, to ensure that there was not air flow between the chamber and the atmosphere.

Devices

Covariates Soil temperature and in some cases moisture have been recorded in the upper layer (5-10 cm) with various commercially available sensors.

2.3. The database

2.3.1 Organization of the published dataset

The database is published as a static resource on Zenodo (ADD LINK HERE!!!!!!).

The dataset is structured in folders, each corresponding to one site. Inside each folder, there are subfolders corresponding to the measurement series. Each measurement series is a sampling campaign, during which each instrument was started and run for several samplings of multiple plots. Inside each series subfolder there are Apache feather files corresponding to each single measurement. Apache Feather is an open columnar binary file format designed for fast data interchange between different programming languages, particularly Python and R. It provides efficient serialization and deserialization of dataframes with minimal overhead, which makes it ideal for storing a large number of relatively small files.

Each feather file, so each measurement, is associated to a line in the metadata file in each site folder. This contains also the respiration fluxes estimated automatically. The

estimates were in some cases trimmed manually by the uploading operators, who selected the window in the time series on which to perform the regression to calculate the fluxes. In other cases, when the operator did not select a manual trimming window, the trimming was done automatically. The metadata files are then collected together in one single general file in the root folder of the dataset (file "All_sites.csv"), where the analyses in this study were performed. The dataset is distributed with both calculated flux values and raw measurements in case the user of the database would want to develop own flux estimation approaches.

2.3.2 Automatic trimming logic

Jani to write this.

2.3.3 Dataset fields and units

A table here, to write together with Jani

2.3.4 Descriptive analysis of the data

Recoding of the treatments

2.4. Modeling SOC decomposition dependency on climate

2.5. The model

Soil respiration was modeled as a function of temperature, moisture, and seasonal effects:

$$R = A \times \xi_{\text{temp}} \times \xi_{\text{moisture}} \times \xi_{\text{seasonal}} \times \lambda \quad (1)$$

where A represents baseline respiration rates (plot-specific), and the environmental multipliers are defined as follows:

Temperature sensitivity following Lloyd and Taylor kinetics with temperature sen-

sitivity (E_0):

$$\xi_{\text{temp}} = \exp\left(-\frac{E_0}{(T + 273.15) - T_0}\right) \quad (2)$$

where T is temperature ($^{\circ}\text{C}$) and $T_0 = 227.13$ K is a reference temperature constant.

Moisture response using a quadratic relationship:

$$\xi_{\text{moisture}} = aM - bM^2 \quad (3)$$

where M is soil moisture content, and a and b are treatment-specific parameters controlling the shape of the moisture response curve.

Seasonal variation through a cosine function:

$$\xi_{\text{seasonal}} = \text{amplitude} \times \cos\left(\frac{2\pi}{365} \times \text{day_year} + \frac{2\pi}{365} \times (\text{peak_day} - 1) - \frac{\pi}{2}\right) \quad (4)$$

where `day_year` is the day of the year (1-365), `peak_day` defines the timing of maximum seasonal effect, and `amplitude` controls the magnitude of seasonal variation.

Trenching effect distinguishing trenched from non-trenched plots:

$$\lambda = \begin{cases} 1 & \text{if trenched} \\ \text{linear_mult} & \text{if not trenched} \end{cases} \quad (5)$$

where `linear_mult` represents the multiplicative effect of root presence in non-trenched plots.

2.6. The Bayesian framework

We employed a two-stage Bayesian inference approach to ensure robust parameter estimation and convergence in this complex hierarchical model. The procedure consisted of an initial variational approximation phase followed by full MCMC sampling.

2.6.1 Variational inference for MAP estimation

Prior to MCMC sampling, we obtained maximum a posteriori (MAP) parameter estimates using stochastic variational inference (SVI) implemented in NumPyro. This pre-estimation step serves multiple purposes: it provides good starting values for the MCMC chains, reduces burn-in time, and helps identify potential convergence issues before computationally expensive full sampling. The SVI approach uses an AutoNormal guide, which assumes a mean-field variational approximation where each parameter is independently normally distributed. This approximation, while simpler than the true posterior, provides a computationally efficient way to locate the posterior mode. The optimization was performed using the Adam optimizer with a learning rate of 0.01, running for 3,000 iterations to ensure convergence of the variational parameters.

The variational objective minimizes the evidence lower bound (ELBO):

$$\text{ELBO} = \mathbb{E}_{q_\phi(\theta)}[\log p(y|\theta) + \log p(\theta) - \log q_\phi(\theta)] \quad (6)$$

where $q_\phi(\theta)$ is the variational approximation with parameters ϕ , $p(y|\theta)$ is the likelihood, and $p(\theta)$ represents the prior distributions.

2.6.2 MCMC initialization and sampling

Using the MAP estimates from SVI, we initialized 24 parallel MCMC chains by adding small random perturbations (noise scale = 0.05) around the MAP values. This initialization strategy, known as "jittering," ensures that chains start from different but reasonable locations in parameter space, facilitating better exploration of the posterior distribution while avoiding poor local optima that can occur with random initialization. The MCMC sampling employed the No-U-Turn Sampler (NUTS) algorithm [2]. NUTS is particularly well-suited for high-dimensional problems like our hierarchical model, as it can efficiently navigate complex posterior geometries. We configured NUTS with a target acceptance probability of 0.9. Each chain was run for 20,000 iterations total: 5,000 warmup sam-

ples, followed by 15,000 retained samples for posterior inference. The warmup phase allows NUTS to learn the local geometry of the posterior distribution, which is crucial for efficient sampling in subsequent iterations. The parallel chain approach enables both computational efficiency (especially in our case, run on a highly parallelized cluster, see Section 2.8) and robust convergence diagnostics through between-chain comparisons.

2.6.3 Prior distributions

Prior distributions were informed by ecological knowledge and previous studies:

$$E_0 \sim \mathcal{N}(398.5, 20^2) \quad (\text{activation energies typical for soil processes}) \quad (7)$$

$$A \sim \mathcal{N}(400, 50^2) \quad (\text{baseline respiration rates}) \quad (8)$$

$$a \sim \mathcal{N}(3.11, 1^2) \quad (\text{moisture response parameters}) \quad (9)$$

$$b \sim \mathcal{N}(2.42, 1^2) \quad (\text{based on soil-water relationships}) \quad (10)$$

$$\text{amplitude} \sim \mathcal{N}(0, 1^2) \quad (\text{seasonal variation magnitude}) \quad (11)$$

$$\text{peak_day} \sim \mathcal{N}(196, 10^2) \quad (\text{mid-growing season timing}) \quad (12)$$

$$\text{linear_mult} \sim \mathcal{N}(1, 0.5^2) \quad (\text{trenching effect multiplier}) \quad (13)$$

$$\sigma \sim \text{Exp}(1) \quad (\text{observation error}) \quad (14)$$

2.6.4 Likelihood

The likelihood function assumes that observed respiration measurements follow a normal distribution around the model predictions:

$$R_{\text{obs},i} \sim \mathcal{N}(R_{\text{pred},i}, \sigma^2) \quad (15)$$

where $R_{\text{obs},i}$ is the i -th observed respiration measurement, $R_{\text{pred},i}$ is the corresponding model prediction, and σ represents the standard deviation of observation errors. This formulation accounts for measurement uncertainty and natural variability not captured by the environmental predictors, with the assumption that residuals are normally distributed

and homoscedastic across the range of predicted values.

2.6.5 Sampling stratification

The model sampling relied on stratification to ensure proper inference across the huge variety of treatments. Our sampling design incorporated three levels of stratification:

Treatment level: The model accounts for treatment-specific parameters (E_a , a , b , amplitude, peak_day) that capture systematic differences in respiration responses between experimental treatments. This allows for treatment-specific temperature sensitivities, moisture responses, and seasonal patterns, based on the assumption that treatments will influence decomposition kinetics.

Plot level: Baseline respiration rates (A) were modeled as plot-specific parameters to account for spatial heterogeneity in soil properties, microbial communities, and other plot-level factors not explicitly measured. This approach recognizes that plots within the same treatment may have different inherent respiration capacities due to local environmental conditions. The most important plot-level source of variation we expect is substrate amount, meaning C stocks, and possibly local substrate quality.

Observation level: Individual respiration measurements were treated as independent observations conditional on the plot and treatment effects, with measurement error captured through the likelihood function. This level accounts for temporal variation within plots and measurement uncertainty.

Treatment effects are estimated by pooling information across all plots within each treatment, while plot-level heterogeneity is explicitly modeled rather than ignored. This stratification ensures that statistical inference properly reflects the experimental design and sampling structure of the study.

2.7. Missing Values Imputation Approach

Due to sampling issues and diversity of the sampling approaches between sites, our dataset contained missing values for soil temperature (18.5% of observations) and particumarily

soil moisture (41.8% of observations). Rather than using traditional approaches like mean imputation or listwise deletion, which can introduce bias or reduce statistical power, or dropping completely the moisture function discarding some available information, we implemented a hierarchical Bayesian missing data model that accounts for site-specific patterns while properly propagating uncertainty. Due to computational constraints, missing values were modeled using site-level parameters, assuming that temperature and moisture are relatively constant within a same site: soil moisture followed $M_{\text{mean},s} \sim \text{Beta}(2, 2)$ scaled to [0.1, 0.9] and temperature followed $T_{\text{mean},s} \sim \text{Normal}(15, 10)$, where s indexes monitoring sites. Missing observations were then imputed from their corresponding site-specific distributions within the main Bayesian model, ensuring uncertainty propagation through all parameter estimates. This approach reduced the missing data parameter space from potentially thousands of individual parameters to 4 parameters per site ($N = Z$ sites), resulting in substantial computational efficiency gains while maintaining biological realism and proper uncertainty quantification. A model imputing missing values independently for each value would of course be more accurate in theory, but would imply an unfeasible number of Bayesian parameters in the model (many thousands) with consequent prohibitive memory and power requirements.

2.8. Implementation and hardware

All analyses were conducted in Python using JAX for numerical computing and NumPyro for Bayesian inference [3]. Computations were performed on the Puhti supercomputer cluster operated by CSC - IT Center for Science, Finland. The model was run on CPU nodes with 25 parallel processing cores allocated per job to ensure computational efficiency and optimal resource utilization within the high-performance computing environment.

Convergence was assessed through visual inspection of trace plots, effective sample size diagnostics, Gelman-Rubin convergence diagnostic (\hat{R}) [1, 4], and absence of divergent transitions. CPU utilization and performance were monitored throughout the sampling process to ensure efficient use of computational resources. The parallel chain implementa-

tion allowed for robust convergence diagnostics by comparing results across independent chains.

Model outputs provided posterior distributions for all parameters, enabling quantification of uncertainty in treatment effects and environmental sensitivities. The high-performance computing infrastructure enabled extensive sampling (15,000 samples \times 24 chains) necessary for reliable posterior inference in this complex hierarchical model.

3. Results

3.1. Descriptive analysis

3.1.1 Soil respiration and temperature

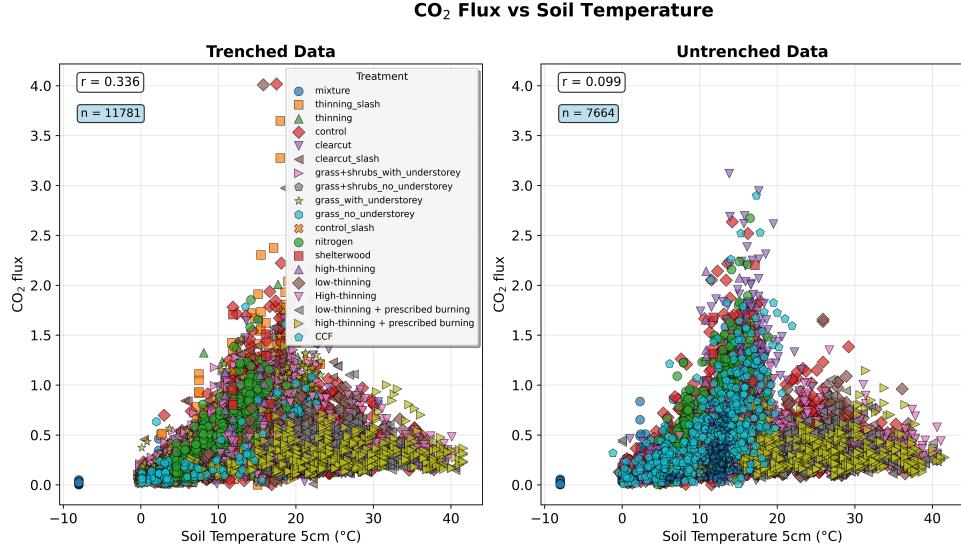


Figure 1: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean. * $p < 0.05$, ** $p < 0.01$.

3.1.2 Soil respiration and moisture

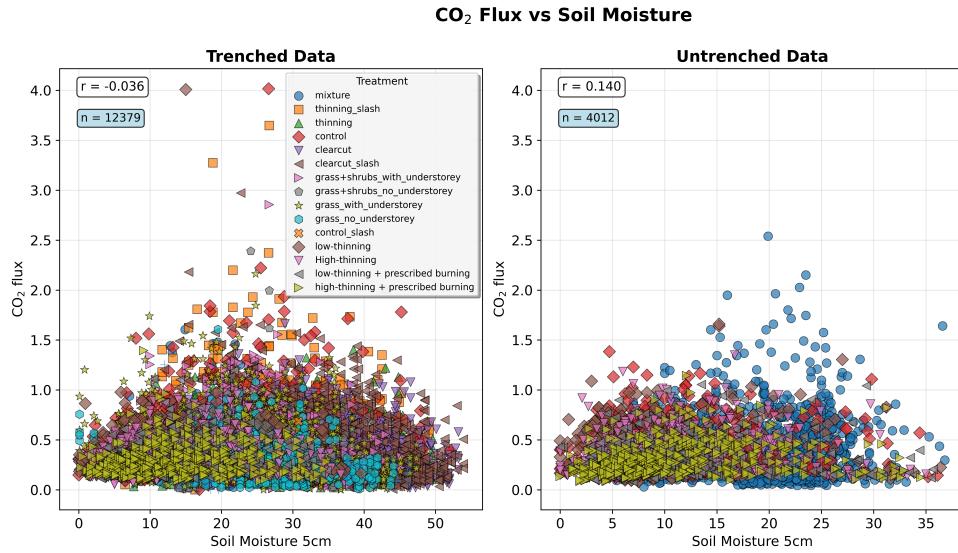


Figure 2: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean. * $p < 0.05$, ** $p < 0.01$.

3.1.3 Moisture and temperature interactions

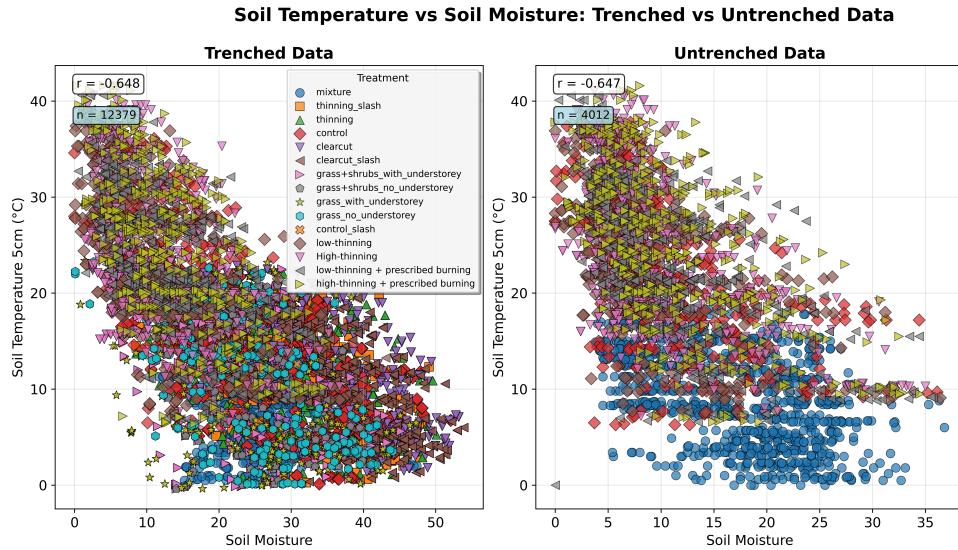


Figure 3: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean. * $p < 0.05$, ** $p < 0.01$.

3.1.4 Autotrophic and heterotrophic fluxes

3.1.5 Moisture and temperature interactions

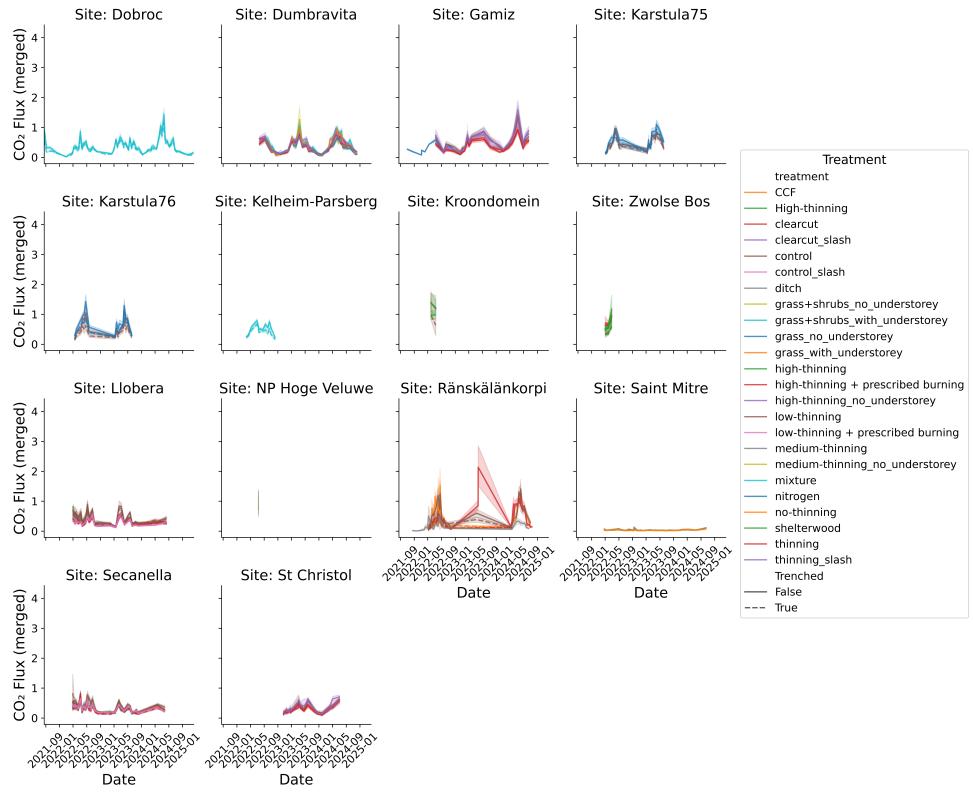


Figure 4: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean. * $p < 0.05$, ** $p < 0.01$.

3.1.6 Modeling analysis

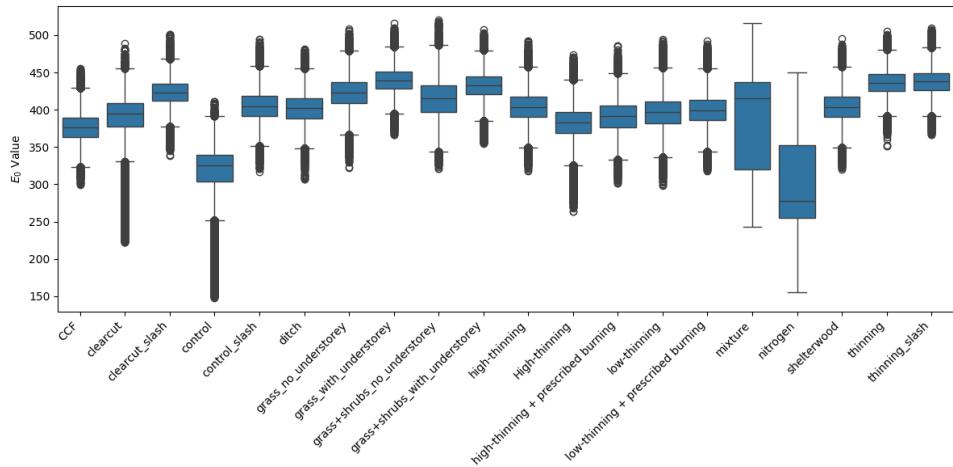


Figure 5: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean.
 $*p < 0.05$, $**p < 0.01$.

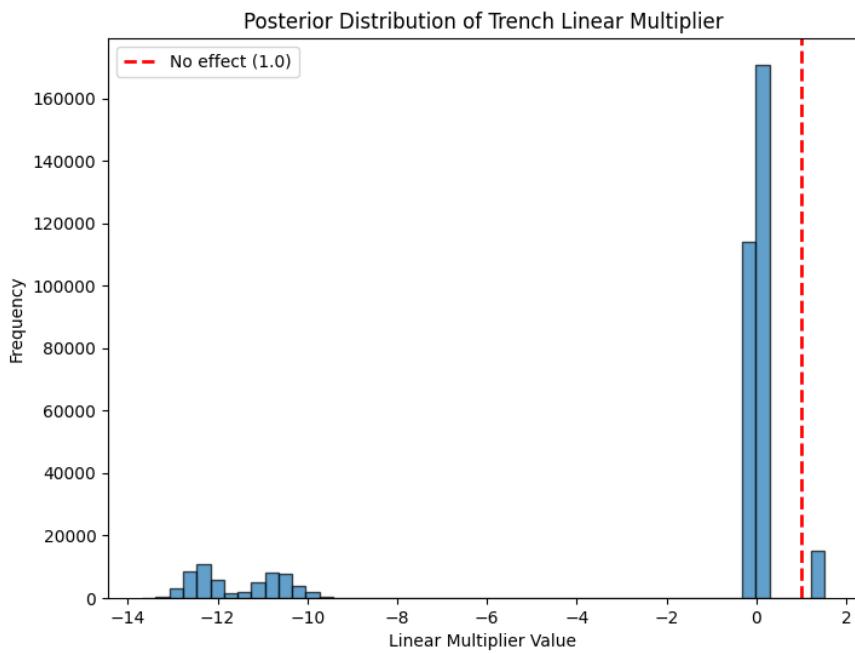


Figure 6: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean.
 $*p < 0.05$, $**p < 0.01$.

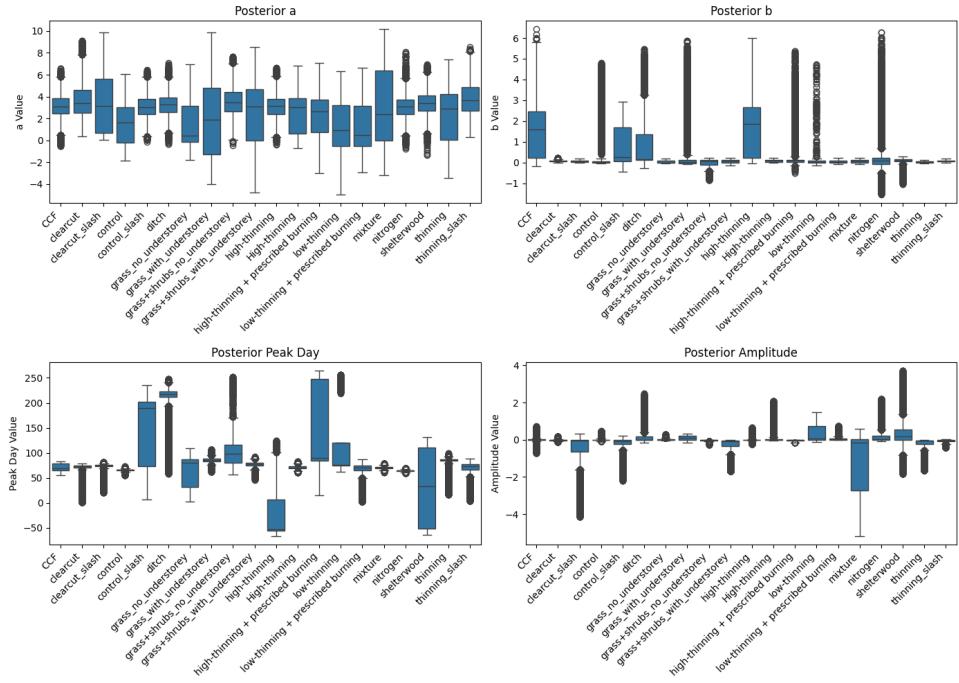


Figure 7: Example figure caption. Describe what the figure shows, including experimental conditions, sample sizes, and statistical tests. Error bars represent standard error of the mean. * $p < 0.05$, ** $p < 0.01$.

4. Discussion

4.1. Interpretation of Results

4.2. Limitations

4.3. Implications and Future Directions

5. Conclusions

Summarize the key findings and their significance. Avoid simply repeating the abstract; instead, provide a synthesis that emphasizes the contribution of your work to the field.

Acknowledgments

Acknowledge funding sources, institutional support, and individuals who contributed to the work but are not listed as authors.

Funding

This work was supported by [Grant Agency] under Grant [Number]. [Author Name] was supported by [Fellowship/Scholarship].

Conflicts of Interest

The authors declare no conflicts of interest.

Data and Code Availability

The data that support the findings of this study are available as Open Data . . .

The code used for this analysis is available (for repeatability purposes, but the code is specific to the cluster we run it onto and needs some adaptation to run on other machines) at https://github.com/ilmenichetti/Holisoils_fitting

References

- [1] Andrew Gelman and Donald B Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7(4):457–472, 1992.
- [2] Matthew D Hoffman and Andrew Gelman. The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1):1593–1623, 2014.
- [3] Du Phan, Neeraj Pradhan, and Martin Jankowiak. Composable effects for flexible and accelerated probabilistic programming in numpyro. *arXiv preprint arXiv:1912.11554*, 2019.
- [4] Aki Vehtari, Andrew Gelman, Daniel Simpson, Bob Carpenter, and Paul-Christian Bürkner. Rank-normalization, folding, and localization: An improved \hat{R} for assessing convergence of mcmc. *Bayesian Analysis*, 16(2):667–718, 2021.