



<i>Title:</i> TOS Science Design for Terrestrial Biogeochemistry		<i>Date:</i> 04/06/2022
<i>NEON Doc. #:</i> NEON.DOC.000906	<i>Author:</i> E. Hinckley	<i>Revision:</i> C

TOS SCIENCE DESIGN FOR TERRESTRIAL BIOGEOCHEMISTRY

PREPARED BY	ORGANIZATION	DATE
Eve-Lyn S. Hinckley	FSU	07/07/2014
Samantha R. Weintraub	FSU	04/12/2018

APPROVALS	ORGANIZATION	APPROVAL DATE
Kate Thibault	SCI	04/06/2022

RELEASED BY	ORGANIZATION	RELEASE DATE
Tanisha Waters	CM	04/06/2022

See configuration management system for approval history.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

Change Record

REVISION	DATE	ECO #	DESCRIPTION OF CHANGE
A	10/16/2014	ECO-02284	Initial release
B	04/26/2018	ECO-05314	<ul style="list-style-type: none"> • Updated names of protocols throughout • Section 4: Updated number of gradient sites • Section 5: Minor edits to reflect current sampling and analysis plans • Section 6.1: Clarified purpose, deliverables, and spatial approach for the initial soil characterization effort • Section 6.2: Updated list of analytes to be measured during periodic soil biogeochemical sampling; specified need to use bulk density measurements from initial characterization effort in order to estimate C and N stocks; provided clearer example of ways to link N cycling with microbial measurements; updated plot numbers and temporal sampling approach to reflect current plans • Section 6.3: Updated description of sampling strategy for canopy foliage in forest and grassland plots; revised plot number estimates and temporal sampling strategy for canopy foliage, litter, and roots; removed reference to mini-rhizotrons • Section 6.4: Elaborated on time-sensitive tasks • Tables 4 and 6: Removed archive sample number counts
C	04/06/2022	ECO-06790	<ul style="list-style-type: none"> • Revised logo • Update to reflect change in terminology from relocatable to gradient sites.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

TABLE OF CONTENTS

1 DESCRIPTION..... 1

1.1 Purpose..... 1

1.2 Scope 1

1.3 Acknowledgements..... 1

2 RELATED DOCUMENTS AND ACRONYMS 2

2.1 Applicable Documents 2

2.2 Reference Documents 2

2.3 Acronyms 2

3 INTRODUCTION..... 4

3.1 Overview of the Observatory..... 4

3.2 Components of the Observatory 5

3.3 The Terrestrial Observation System (TOS)..... 5

4 INTRODUCTION TO THE TERRESTRIAL BIOGEOCHEMISTRY SAMPLING DESIGN..... 6

4.1 Background..... 6

4.2 NEON’s Contribution 8

4.3 Purpose and Scope 9

5 SAMPLING FRAMEWORK.....12

5.1 Science Requirements12

5.2 Data Products.....12

5.3 Priorities and Challenges for Terrestrial Biogeochemistry.....12

6 SAMPLING DESIGN FOR TERRESTRIAL BIOGEOCHEMISTRY.....17

6.1 Sampling Design for Terrestrial Biogeochemical Drivers.....17

6.1.1 Sampling Methods19

6.1.2 Spatial Distribution of Sampling20

6.1.3 Temporal Distribution of Sampling20

6.2 Sampling Design for Soil Biogeochemistry.....20

6.2.1 Sampling Methods23

6.2.2 Spatial Distribution of Sampling25

6.2.3 Temporal Distribution of Sampling25

6.3 Sampling Design for Plant Biogeochemistry.....26



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

6.3.1	Sampling Methods	28
6.3.2	Spatial Distribution of Sampling	30
6.3.3	Temporal Distribution of Sampling	31
6.4	Logistics and Adaptability	31
7	REFERENCES	35
APPENDIX A	A STATISTICAL FRAMEWORK FOR EVALUATING NEON SAMPLING DESIGN	41
A.1	Introduction	41
A.2	Methods.....	42
	Response Variable	42
	Historical Data.....	43
A.3	Simulation Design	44
A.4	Data Model.....	44
A.5	Process Model.....	44
A.6	Process Covariance.....	45
A.7	Parameter Models	45
A.8	Analysis of Synthetic Data from the Harvard Forest LTER (NEON Domain 1 Core Site).....	46
A.9	Computational Issues	46
A.10	Results	48
	Trend Parameter	48
	Other Parameters	49
A.11	Discussion.....	49
A.12	References.....	50
A.13	Tables and Figures.....	51
A.14	Code Used for Simulations	62

LIST OF TABLES AND FIGURES

Table 1. Anticipated timescales of change in components (drivers and responses) of biogeochemical cycles.....	13
Table 2. Soil chemical and physical properties measured during the one-time characterization effort at each NEON site.....	19
Table 3. Soil biogeochemical analyses and laboratory methods.	23



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

Table 4. Archived samples from the soil biogeochemistry sampling effort.25

Table 5. Laboratory analyses and methods for plant tissues.29

Table 6. Archived samples from the plant tissue collection efforts.30

Table 7. Summary statistics for peak CO₂ flux estimates obtained from the reduced dataset with four sites and three collars within each site. A single sampling event was used for each year, therefore, there are 12 samples associated with each year.....51

Table 8. Parameter values used to generate synthetic data. Normal distributions are specified with a mean and standard deviation. Two values are specified for the β_{trend} parameter: deciduous = 0.05, and evergreen = 0.10. IGamma is the Inverse Gamma distribution as specified in the MCMCpack library (Martin et al. 2008) in R.51

Table 9. Percentiles for the trend parameter posterior distributions for the 1, 3, and 5 year sampling frequencies of deciduous and evergreen.51

Table 10. Percentiles for the difference in the trend parameter posterior distributions for the 1, 3, and 5 year sampling frequencies. The realizations from the posterior distribution of the difference are also used to provide an estimate of the probability that the difference is greater than zero.52

Figure 1. The seven Grand Challenges defined by the National Research Council (2001). 4

Figure 2. Long-term chemistry records of (A) precipitation at Huntington Forest, NY, and (B) Big Moose Lake, Adirondack State Park, NY (Source: C. Driscoll, in prep; NADP Station NY20, <http://nadp.sws.uiuc.edu/> and Adirondack Lakes Survey Corporation, <http://www.adirondacklakessurvey.org/>). 7

Figure 3. High-level requirements associated with the Terrestrial Biogeochemistry sampling design. Level 1 are overarching design goals, Level 3 are requirements more specific to analyses and sampling approaches. Note that certain Level 3 requirements, such as measurements of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ in algae, plants, and soils, P and cations in litter, and dissolved organic carbon and nitrogen in soils, will not be met at this time. Level 2 requirements are “systems requirements,” which pertain more to infrastructure and sensor-based instrumentation; there are not any level 2 requirements associated with the Terrestrial Biogeochemistry sampling design..... 9

Figure 4. Example of a NEON site (Domain 3, Ordway-Swisher, Florida) with eddy covariance tower location, distributed plots, surface and groundwater sampling locations, and site boundary. Note: tower plots not shown.....11

Figure 5. Steps in the RRQRR statistical framework to determine a stratified random distribution of plots across a NEON site.15

Figure 6. NEON biogeochemistry measurements within a site. Shown are variables that are part of the terrestrial biogeochemistry design (Terrestrial Observation System, TOS), as well as those associated with the tower (TIS), above- and belowground plant biomass (TOS), microbial community (TOS), and ground and surface waters (AOS/AIS). Subsurface hydrological flow paths that are important to constrain in biogeochemical budgets are shown, but will not be measured by NEON.18

Figure 7. Soil (a) net N mineralization and (b) net nitrification in relation to soil C:N (top 10 cm). Figures are from Ollinger et al. (2002).....22



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hincley	Revision: C

Figure 8. Net nitrification in relation to foliar N concentrations for disturbed and undisturbed forest stands. Inset shows the nonlinear response of nitrification. The trends are described by $\log(\text{NO}_3) = 2.75(\text{foliar N}) - 2.96$, ($R^2 = 0.81$, $P < 0.001$) for undisturbed stands; $\log(\text{NO}_3) = 2.94(\text{foliar N}) - 4.28$, ($R^2 = 0.63$, $P < 0.001$) for disturbed stands. Figure is from Ollinger et al. (2002).22

Figure 9. Results from the Domain 1 (Harvard Forest) NEON Prototype in 2012. (a) Percent N and $\delta^{13}\text{C}$ of plant canopy foliar tissues by plant phylogenetic group (i.e., angiosperms, ferns, and gymnosperms) and photosynthetic pathway (i.e., C3 and C4), and (b) Mean (± 1 SE) C:N ratios across the same groups.27

Figure 10. NLCD coverage of Harvard Forest Core Site within Domain 1. Distributed sample locations determined by the RRQRR design are shown for the vegetation types selected for sampling.....53

Figure 11. NLCD coverage of Harvard Forest Core Site within Domain 1. Sample locations are shown for the historical data used to inform the simulation study.54

Figure 12. Boxplots of monthly distributions for CO₂ flux from the control data, collected by PI Davidson using manual soil collars at the Harvard Forest. Metadata for the entire Harvard forest archive are available online (<http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf194>). The grey boxes span the 25th and 75th percentile (i.e., the inter-quartile range, IQR). Medians are depicted with the black horizontal line within each box. Whiskers for each boxplot extend out to the largest value less than the median $\pm 1.5 \times \text{IQR}$. Observations outside the median $\pm 1.5 \times \text{IQR}$ are depicted as open circles.55

Figure 13. Boxplots of annual peak CO₂ flux distributions for the final working dataset from the Harvard Forest. Each year has 12 samples arising from 4 locations with 3 collars each. Note that data are omitted for years 2001 and 2002 due to unequal sample sizes. The data for each year are taken from a single sampling event occurring towards the end of the month of August.56

Figure 14. Boxplots of annual peak CO₂ flux distributions by site within each year for the final working dataset from the Harvard Forest. There is some evidence that inter-annual variability varies among sites. For example, NWN appears to display greater inter-annual variability than NWM.57

Figure 15. Boxplots of annual peak CO₂ flux distributions by site within each year.58

Figure 16. Posterior distributions for the trend parameters at years 3, 18, and 30 for both the deciduous and evergreen vegetation types with annual sampling.58

Figure 17. Posterior distributions for the trend parameters at years 3, 18, and 27 for both the deciduous and evergreen vegetation types with sampling every 3 years.59

Figure 18. Posterior distributions for the trend parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every 5 years.59

Figure 19. Histograms of the posterior distributions of the difference in the trends as a function of sampling frequency and year.....60

Figure 20. Posterior distributions for the intercept parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every year.61

Figure 21. Posterior distributions for the intercept parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every year.....61



<i>Title:</i> TOS Science Design for Terrestrial Biogeochemistry		<i>Date:</i> 04/06/2022
<i>NEON Doc. #:</i> NEON.DOC.000906	<i>Author:</i> E. Hinckley	<i>Revision:</i> C

Figure 22. Posterior distributions for the intercept parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every 5 years.62



1 DESCRIPTION

1.1 Purpose

NEON design documents are required to define the scientific strategy leading to high-level protocols for NEON subsystem components, linking NEON Grand Challenges and science questions to specific measurements. Many NEON *in situ* measurements can be made in specific ways to enable continental-scale science rather than in ways that limit their use to more local or ecosystem-specific questions. NEON strives to make measurements in ways that enable continental-scale science to address the Grand Challenges. Design Documents flow from questions and goals defined in the NEON Science Strategy document, and inform the more detailed procedures described in Level 0 (L0; raw data) protocol and procedure documents, algorithm specifications, and Calibration/Validation (CaVal) and maintenance plans.

1.2 Scope

This document defines the rationale and requirements for Terrestrial Biogeochemistry in the NEON Science Design.

1.3 Acknowledgements

This document was written in collaboration with the Terrestrial Biogeochemistry Technical Working Group members, including: Gordon Bonan (NCAR), Gabriel Bowen (University of Utah), Benjamin Colman (Duke University), Paul Duffy (Neptune Consulting), Christine Goodale (Cornell University), Benjamin Houlton (University of California, Davis), Erika Marín-Spiotta (University of Wisconsin, Madison), Kiona Ogle (Arizona State University), Scott Ollinger (University of New Hampshire), Eldor Paul (Colorado State University), Peter Vitousek (Stanford University), Kathleen Weathers (Cary Institute of Ecosystem Studies), and David Williams (University of Wyoming). Revision B was completed with the assistance of Tanya Chesney, who helped update to a new template.

2 RELATED DOCUMENTS AND ACRONYMS

2.1 Applicable Documents

Applicable documents contain information that is applied in the current document. Examples are higher level requirements documents, standards, rules and regulations.

AD[01]	NEON.DOC.000001	NEON Observatory Design (NOD) Requirements
AD[02]	NEON.DOC.001282	Introduction to the TOS Science Designs
AD[03]	NEON.DOC.000913	TOS Science Design for Spatial Sampling
AD[04]	NEON.DOC.002652	NEON Level 1, Level 2 and Level 3 Data Products Catalog
AD[05]	NEON.DOC.000908	TOS Science Design for Microbial Diversity
AD[06]	NEON.DOC.000912	TOS Science Design for Plant Diversity
AD[07]	NEON.DOC.014048	TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling
AD[08]	NEON.DOC.004130	TOS Standard Operating Procedure: Wetland Soil Sampling
AD[09]	NEON.DOC.000914	TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index
AD[10]	NEON.DOC.001024	TOS Protocol and Procedure: Canopy Foliage Sampling
AD[11]	NEON.DOC.014038	TOS Protocol and Procedure: Plant Belowground Biomass Sampling
AD[12]	NEON.DOC.001710	TOS Protocol and Procedure: Litterfall and Fine Woody Debris
AD[13]	NEON.DOC.000907	TOS Science Design for Plant Phenology
AD[14]	NEON.DOC.000912	TOS Science Design for Plant Diversity

2.2 Reference Documents

Reference documents contain information complementing, explaining, detailing, or otherwise supporting the information included in the current document.

RD [01]	NEON.DOC.000008	NEON Acronym List
RD [02]	NEON.DOC.000243	NEON Glossary of Terms

2.3 Acronyms

Acronym	Definition
C	Carbon
¹² C	Common stable isotope of carbon
¹³ C	Less common stable isotope of carbon
Ca ²⁺	Calcium
CaCl ₂	Calcium chloride
cm	Centimeter
CO ₂	Carbon dioxide
EPA	Environmental Protection Agency
mm	Millimeter
g	Grams
h	Hours



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

K ⁺	Potassium
LTER	Long-term Ecological Research
LTM	Long-term Monitoring
m	Meter
M	Molar
MCMC	Markov Chain Monte Carlo
mg	Milligram
MH	Metropolis-Hastings
ml	Milliliter
mRNA	Messenger Ribonucleic Acid
N	Nitrogen
¹⁵ N	Less common stable isotope of nitrogen
¹⁴ N	Common stable isotope of nitrogen
NADP	National Atmospheric Deposition Program
NCSS	National Cooperative Soil Survey
NRCS	Natural Resources Conservation Service
PO ₄ ³⁻	Phosphate
P	Phosphorus
S	Sulfur
SO ₄ ²⁻	Sulfate
SSURGO	Soil Survey Geographic database
STATSGO	State Soil Geographic database
USDA	United States Department of Agriculture



3 INTRODUCTION

3.1 Overview of the Observatory

The National Ecological Observatory Network (NEON) is a continental-scale ecological observation platform for understanding and forecasting the impacts of climate change, land use change, and invasive species on ecosystem structure and function. NEON was conceived to enable data users, including scientists, planners and policy makers, educators, and the general public, to address the major areas in environmental sciences, known as the Grand Challenges (**Figure 1**). NEON infrastructure and data products are strategically aimed at those aspects of the Grand Challenges for which a coordinated national program of standardized observations and experiments is particularly effective. The open access approach to the Observatory’s data and information products will enable users to explore NEON data in order to map, understand, and predict the effects of humans on the Earth and understand and effectively address critical ecological questions and issues. Detailed information on the NEON design can be found in AD[01], AD[02].

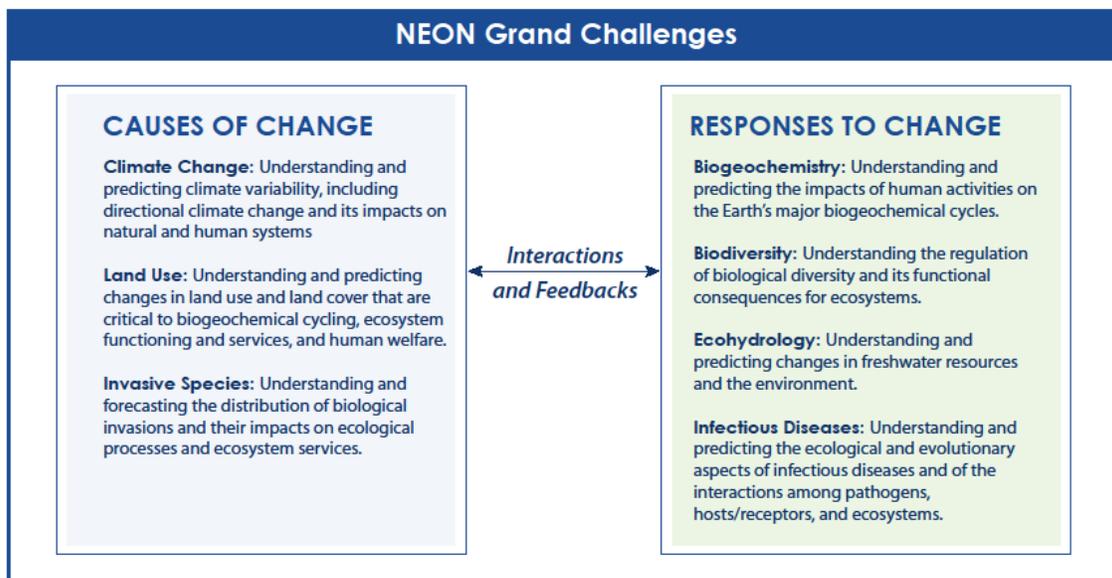


Figure 1. The seven Grand Challenges defined by the National Research Council (2001).



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

3.2 Components of the Observatory

There are five components of the Observatory, the Airborne Observation Platform (AOP), Terrestrial Instrument System (TIS), Terrestrial Observation System (TOS), Aquatic Instrument System (AIS), and Aquatic Observation System (AOS). Collocation of measurements associated with each of these components will allow for linkage and comparison of data products. For example, remote sensing data provided by the Airborne Observation Platform (AOP) will link diversity and productivity data collected on individual plants and stands by the Terrestrial Observation System (TOS) and flux data captured by instruments on the tower (TIS) to that of satellite-based remote sensing. For additional information on these systems, see Keller et al. 2008, Schimel et al. 2011.

3.3 The Terrestrial Observation System (TOS)

The NEON TOS will quantify the impacts of climate change, land use, and biological invasions on terrestrial populations and processes by sampling key groups of organisms (sentinel taxa), infectious disease, soil, and nutrient fluxes across system interfaces (air, land, and water) (AD[01], AD[02]). The sentinel taxa were selected to include organisms with varying life spans and generation times, and wide geographic distributions to allow for standardized comparisons across the continent. Many of the biological measurements will enable inference at regional and continental scales using statistical or process-based modeling approaches. The TOS sampling design captures heterogeneity representative of each site to facilitate this inference when possible. Plot and organism-scale measurements will also be coordinated with the larger-scale airborne measurements, which provide a set of synergistic biological data products at the regional scale. Details of these design elements and algorithms can be found in individual design documents available through the NEON website (www.neonscience.org).

The standardization of protocols across all sites is key to the success of NEON (and its novelty) and must be maintained at all sites through time. Thus, although specific techniques may be required at some sites (e.g., due to different vegetation types), protocols have been developed to ensure data comparability. These details can also be found in individual design documents available through the NEON data portal (<http://data.neonscience.org/documents>).

The TOS Science Designs define the scientific strategies leading to high-level sampling designs for NEON sentinel taxa, terrestrial biogeochemistry, and infectious disease, linking NEON Grand Challenges and science questions to specific measurements (AD[02]). The TOS Spatial Sampling Design document describes the sampling design that collocates observations of the components of the TOS (AD[03]). TOS Science Design documents were developed following input from the scientific community, including discipline-specific Technical Working Groups, and the National Science Foundation (AD[02]). Science Designs will be reviewed periodically to ensure that the data collected by NEON are those best suited to meet the requirements of the observatory (AD[01]), are (to the extent possible) consistent with standards used by the scientific community, and fit within the scope of NEON. Additional information on the development and review process can be found in AD[02].



4 INTRODUCTION TO THE TERRESTRIAL BIOGEOCHEMISTRY SAMPLING DESIGN

4.1 Background

Humans are changing the fundamental chemistry of ecological systems on Earth by altering the global biogeochemical cycles of carbon (C), nitrogen (N), and other elements. These changes are driven by activities that include increasing emissions of carbon dioxide (CO₂) to the atmosphere from fossil fuel combustion, altering the distribution and nature of freshwater resources, changing land cover and land use, and increasing industrial extraction and application of reactive elements (Vitousek et al. 1997). Some of these impacts have been due to technological advances, such as the Haber-Bosch process, which enabled industrial-scale conversion of atmospheric N₂ to nutritionally available ammonia, thereby increasing our ability to provide food to a growing human population. However, such modifications of biogeochemical processes have come with unintended consequences for the biosphere. The human signature on biogeochemical cycles can be seen in nutrient imbalances and ecological impacts at local to global scales: eutrophic surface waters (e.g., Carpenter et al. 1998, Correll 1998, Ryther and Dunstan 1971), declines in health and shifts in the composition of forest species (e.g., Horsley et al. 2002, Shortle et al. 1997), higher incidences of infectious diseases (e.g., MacKenzie and Townsend 2007), and spread of invasive species (e.g., Ashton et al. 2005, Crowl et al. 2008, Lovett et al. 2006, Vitousek and Walker 1989). Examples of these responses can be found in most regions of the world.

Despite having documented the connection between perturbations to biogeochemical cycles and ecosystem effects across the globe, our ability to predict changes in biogeochemical processes over large spatial extents using easily measured ecosystem parameters (e.g., Ollinger et al. 2002) remains limited. Similarly, in only a few areas where intensive, regional studies have occurred, can one determine the degree to which previously impacted ecosystems are recovering in response to policy changes (e.g., Driscoll et al. 2007, 2003, 2001). In addition, there is a growing recognition that in order to understand the responses and feedbacks of ecosystems to global change, researchers must conduct integrated studies of the whole Earth system, including the interactions of climate, hydrology, and biogeochemical cycles (e.g., Gruber and Galloway 2008, Falkowski et al. 2000).

Ability to address these challenges will benefit greatly from spatially extensive, standardized collection of long-term observational data. In several cases, such investments have proven extremely valuable for tackling societally relevant problems. For example, long-term observations of increasing atmospheric CO₂ at Mauna Loa (the Keeling curve) have served both as evidence of human-induced climate change and inspiration for major research efforts. The establishment of experimental research sites such as Hubbard Brook in the 1950s and the Long-term Ecological Research (LTER) network in the 1980s created focal areas for long-term studies in the U.S. Additionally, monitoring networks, such as the National Atmospheric Deposition Program (NADP, <http://nadp.sws.uiuc.edu/>) and the EPA's Long-term Monitoring (LTM) of surface waters (Stoddard et al. 2003), started in response to problems of environmental and ecological importance, including acid rain and mercury deposition.

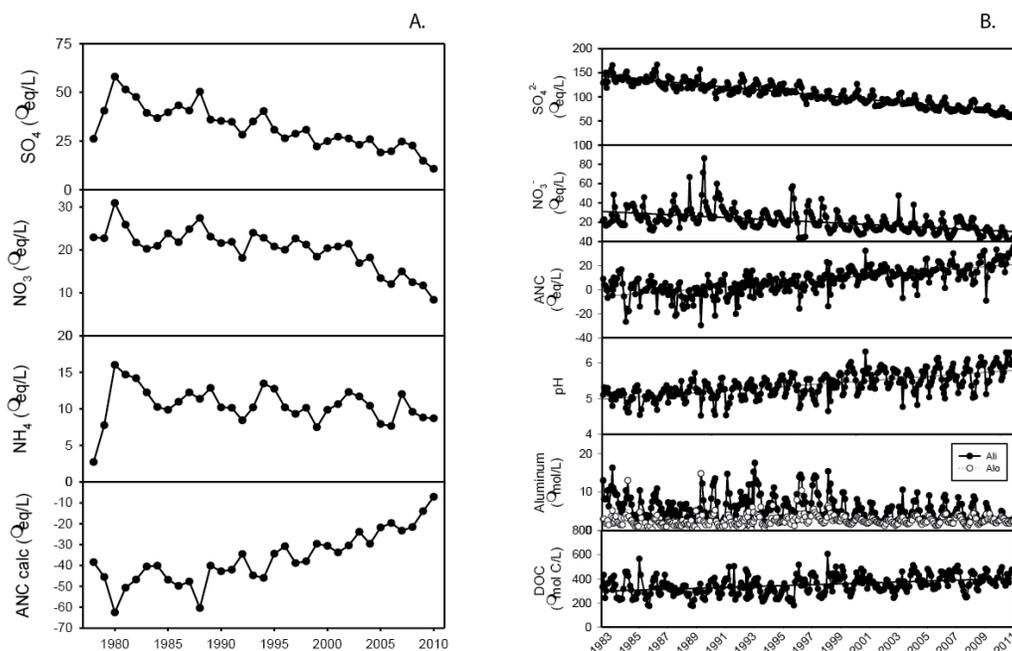


Figure 2. Long-term chemistry records of (A) precipitation at Huntington Forest, NY, and (B) Big Moose Lake, Adirondack State Park, NY (Source: C. Driscoll, in prep; NADP Station NY20, <http://nadp.sws.uiuc.edu/> and Adirondack Lakes Survey Corporation, <http://www.adirondacklakesurvey.org/>).

Decades after the establishment of these study areas and monitoring stations, there are several examples of how collection of long-term biogeochemical data has prompted changes in policy to decrease negative impacts on ecosystems (e.g., Weathers and Lovett 1998, Lovett et al. 2007), and identified surprising ecosystem responses (e.g., Monteith et al. 2007). For example, 30 years of regional precipitation and lake chemistry observations in the Adirondack State Park (New York, USA) established a baseline pattern of the effects of acid rain deposition, followed by decline in the concentrations of nitrate and sulfate in precipitation with passage of the Clean Air Act, and consequent recovery of lakes through the late 1990's and 2000's (e.g., Driscoll et al. 2007, 2003, 2001, unpublished data, Stoddard et al. 1999) (see Figure 2). These and related international monitoring networks detected unexpected increases in stream dissolved organic carbon (DOC) concentrations, a trend attributable to reductions in acid deposition through analysis of corresponding environmental data (Monteith et al. 2007). Other efforts have provided insight into ecosystem nutrient budgets using long-term observations (e.g., Likens et al. 2002) and the effects of nutrient enrichment on ecosystems using long-term experiments at Harvard Forest (e.g., Magill et al. 1997, 2004, Nadelhoffer et al. 2004) and land cover/land use change at Coweeta, Hubbard Brook, and H.J. Andrews research sites, as discussed in Turner et al. (2003).

Coupled to collection of spatially extensive, standardized, long-term observational data must be the assembly of interdisciplinary research teams to drive the advancement of process-based and predictive modeling frameworks (Hinckley et al. 2014). Models provide a means to challenge current knowledge and to guide future field investigations. Investment in such resources will provide a starting point to improve basic understanding of ecosystem biogeochemistry, as well as the ability to predict responses of ecosystems to future changes over large scales.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

4.2 NEON's Contribution

Recent funding of the National Ecological Observatory Network (NEON) provides an opportunity to investigate ecological change at broader spatial and temporal scales than has been achieved before, and to bring new insights beyond those from LTER and other long-term networks where data collection protocols, methods of analysis, and data storage are often diverse and difficult to compare. Previous overview papers (e.g., Field et al. 2006, Keller et al. 2008) outline the focus of the NEON strategy. The goal is to address questions inspired by the 2001 and 2003 National Research Council (NRC) reports that highlighted seven Grand Challenge areas in Earth and environmental sciences (NRC 2001, 2003). The guiding requirements for NEON follow directly from the seven Grand Challenge areas identified by the NRC and the NEON Science Strategy (Schimel et al. 2009), which more specifically summarizes network-wide approaches to sampling across space and time. At the highest level, NEON is designed to improve both understanding of complex ecological systems and the ability of researchers to forecast patterns of ecological change at local, regional, and continental scales. This will be accomplished using standardized and coordinated measurements of ecological taxa and environmental processes at 30-year core ($n = 20$) and 5 to 10-year, gradient locations ($n = 27$) across the U.S. Specific to the terrestrial biogeochemistry component of the observatory, NEON is designed to meet a high-level requirement to "...further our understanding of the Earth's major biogeochemical cycles, [and] evaluate how they are being perturbed by human activities."

In order to construct NEON in an orderly and timely fashion, the project adopted a requirements-based framework to guide design of sampling strategies and infrastructure components. To understand the motivation behind the NEON Terrestrial Biogeochemistry design, it is useful to use the high-level requirements as a starting point. While budgetary constraints prohibit NEON from fulfilling the full suite of Level 3 requirements, the major themes and elements, summarized in **Figure 3**, remain in place. These four main themes are: (1) promote an understanding of biogeochemical stocks and fluxes across air, land, and water systems, (2) use stable isotopes to infer biogeochemical and ecological processes, (3) collocate biogeochemical sampling with other NEON measurement platforms, and (4) measure constituents of interest primarily in plant tissues and soils. It should be noted that fulfilling these requirements relies also on measurement designs developed by the Terrestrial Instrument System (TIS) (i.e., associated with sensor-based tower measurements) and the Aquatic Observation and Instrument Systems (AOS and AIS) (i.e., groundwater and surface water measurements).



Level 1 Science Requirements

NEON's measurement strategy shall include coordinated and co-located measurements of drivers of environmental change, (physical and chemical climate, land use, and biological invaders) and biological responses (matter and energy fluxes, biomass and plant productivity, diversity and genomics of key organismal groups, infectious diseases, and community, phenological and population indicators).

NEON observing strategies shall be designed to support ecological forecasting, including requirements for state and parameter data, and a timely and regular data delivery schedule to support new and ongoing ecological forecast programs.

Level 3 Science Requirements

Algal, plant and litter isotopic ratios ($\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$ and $\delta^2\text{H}$) shall be measured to integrate physiology, nutrient and water sources, and energy balance over the period of growth.

Three sentinel species shall be analyzed for total organic $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$; Water $\delta^{18}\text{O}$, $\delta^2\text{H}$.

Litter shall be analyzed for total organic $\delta^{13}\text{C}$, $\delta^{15}\text{N}$, $\delta^{18}\text{O}$.

Key biogeochemical analyses are required for plant samples, including Total Carbon, Total Nitrogen, and macronutrient concentration (P, K⁺, and Ca²⁺).

Plant biomass shall be analyzed for macronutrient concentration (C, N, K, P, CA, Mg) and lignin.

Litter shall be analyzed for macronutrient concentrations (C, N, P, K⁺, Ca²⁺, Mg²⁺) and lignin.

Foliar tissues shall be analyzed for macronutrient concentrations (C, N, P, K⁺, Ca²⁺, Mg²⁺) and lignin.

Coarse roots shall be analyzed for C and N.

Soils shall be analyzed for total carbon and nitrogen, carbon, nitrogen, and phosphorous fractions, dissolved organic carbon, dissolved organic nitrogen, inorganic nutrient pools, mineral nutrient concentrations, and microbial biomass.

Soil moisture shall be measured at a subset of soil sample locations at least annually.

Soil shall be analyzed for Total Organic C, Total N, Total P, P fractions.

Soils shall be surveyed for standard physical and chemical characteristics, texture, pH, moisture, and temperature.

Soil - Soil isotopic ratios ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) shall be measured to integrate the plant processes over long periods and reflect soil heterotrophic processes.

Soil - Total organic $\delta^{13}\text{C}$, $\delta^{15}\text{N}$; water $\delta^{18}\text{O}$, $\delta^2\text{H}$.

The biogeochemistry measurements shall assess the movement of nutrients through terrestrial and aquatic ecosystems. Measurements shall focus on precipitation and wet deposition, soil, plants, small mammals, surface and groundwater, and aquatic plants and animals. All samples shall be analyzed for chemical and isotopic compositions. Samples shall also be archived to provide a historical record of past environmental and climatic conditions (a reference against which future changes can be evaluated).

The biodiversity and biogeochemistry measurements shall be paired within the FSU in order to connect these fields as much as possible.

Existing analytical resources can provide stable isotope ratios of hydrogen, carbon, nitrogen, oxygen, and sulfur in materials sampled from aquatic, terrestrial, and atmospheric environments. Standardization shall be overseen by the NEON Cal/Val facility.

For both aquatic and terrestrial samples, additional chemical measurements shall be required that shall not be performed at the isotopic facility. These analyses shall be outsourced to certified chemical laboratories with internal QA/QC procedures.

Figure 3. High-level requirements associated with the Terrestrial Biogeochemistry sampling design. Level 1 are overarching design goals, Level 3 are requirements more specific to analyses and sampling approaches. Note that certain Level 3 requirements, such as measurements of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ in algae, plants, and soils, P and cations in litter, and dissolved organic carbon and nitrogen in soils, will not be met at this time. Level 2 requirements are “systems requirements,” which pertain more to infrastructure and sensor-based instrumentation; there are not any level 2 requirements associated with the Terrestrial Biogeochemistry sampling design.

4.3 Purpose and Scope

This document provides an overview of the Terrestrial Biogeochemistry sampling design for implementation at NEON sites. This document describes the motivation and rationale behind the suite of measurements identified to address the Grand Challenge questions and high-level requirements (**Figure 3**), the general field sampling and analytical approaches, and the development of the spatial and temporal sampling design. NEON's broader biogeochemistry measurements include nutrient deposition, C, energy, and water fluxes at tower (TIS) locations; C and nutrient fluxes in surface water sampling locations (AOS/AIS); terrestrial C and nutrient pools and select biogeochemical processes (TOS); and



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

within-site assessments of ecosystem function, including ecosystem productivity (TIS/TOS/AOP), biomass estimates (TOS/AOP), ecosystem (gas and material) exchange (TIS), and the effect of terrestrial systems on productivity and respiration in aquatic systems (TOS/AOS/AIS). Consistent with other sampling components within NEON, the goal of the terrestrial biogeochemistry design is to enable researchers to investigate a number of broad questions using observational and modeling approaches, such as how climate drivers impact ecological stoichiometry at the continental scale, and how regional disturbances (e.g., wildfire and drought) affect coupled C, N, and water cycles. Measurements of biogeochemical stocks and transformations will be made within the tower footprint and in distributed plots across the permitted area of each NEON site; **Figure 4** shows an example of the Domain 3 NEON core site (Ordway-Swisher Biological Station, Florida), including the tower, distributed plots for terrestrial observations, and surface and groundwater sampling locations.

The terrestrial biogeochemistry measurements will be aligned spatially with those in adjacent atmospheric compartments, aquatic systems (certain sites), and other terrestrial ecological measurements. Some measurements will be collocated at the plot-scale (i.e., sub-meter to meter), and others at that of the site (i.e., kilometer to tens of kilometers). Within tower and distributed plots, NEON will conduct sampling of carbon and nutrient concentrations and soil processes, soil microbial communities, and vegetation (structure, biomass, species inventories, and chemistry of tissues). Within sites, these measurements will be collocated with nutrient deposition, meteorological data, C and water fluxes at the tower location, and sample collections of ecological taxa (beetles, mosquitos, birds, and small mammals) and infectious diseases. Nearly half of the NEON terrestrial observation sites are collocated with a NEON aquatic observation site in the same watershed. Collocation of terrestrial and aquatic observations will enable investigators to study C and nutrient linkages in terrestrial and aquatic ecosystems, which is fundamental for understanding ecological connectivity and land-water controls on biogeochemical cycles (Grimm et al. 2003).

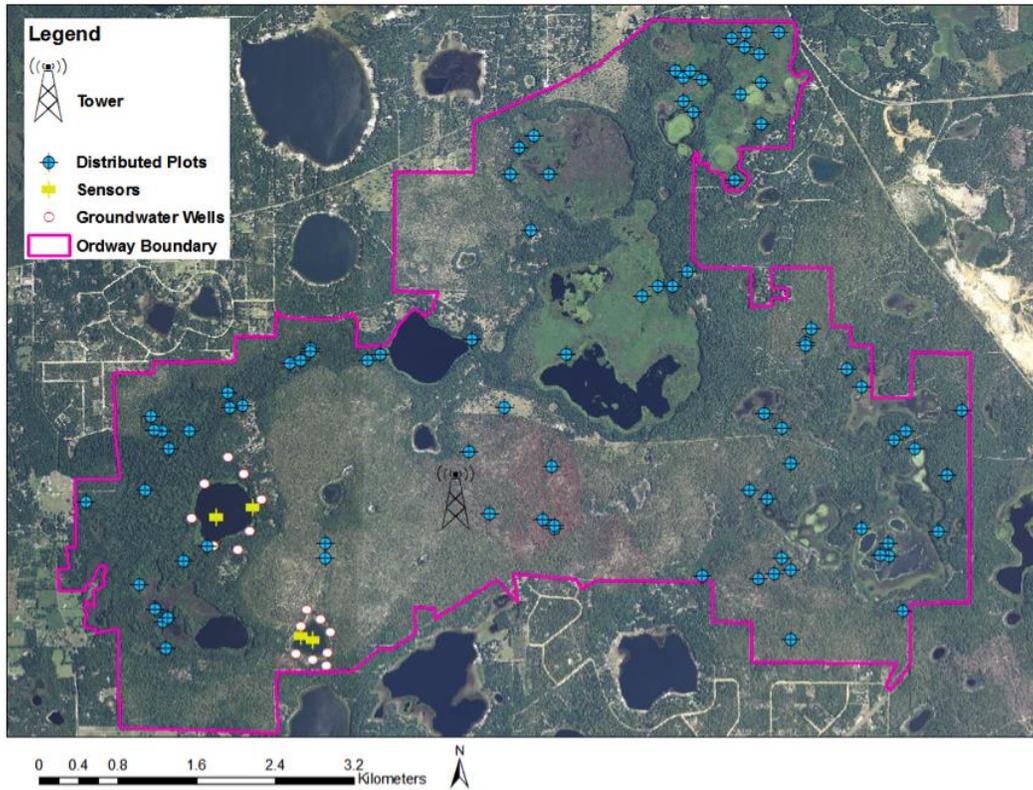


Figure 4. Example of a NEON site (Domain 3, Ordway-Swisher, Florida) with eddy covariance tower location, distributed plots, surface and groundwater sampling locations, and site boundary. Note: tower plots not shown.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

5 SAMPLING FRAMEWORK

The NEON biogeochemistry sampling design focuses on providing data to address regional to continental scale questions regarding long-term changes to ecosystem carbon and nutrient stocks, process rates, major fluxes (measured at tower and aquatic sampling locations), and important feedbacks. There are four general design criteria underlying this program: (1) identification of measurements that will enable investigators to evaluate trends in biogeochemical cycles within a diverse array of ecosystem types distributed across NEON’s spatial purview, and inspire further studies at sub-meter to continental scales; (2) determination of the spatial and temporal sampling design; (3) development of methodologies and QA/QC approaches that are consistent across sites and allow for comparisons with other network observatories; and (4) design of an approach to reevaluate the measurements and sampling strategy over the lifetime of the observatory. Discussion of these criteria is provided in the following sections.

5.1 Science Requirements

This science design is based on Observatory science requirements that reside in NEON’s Dynamic Object-Oriented Requirements System (DOORS). Copies of approved science requirements have been exported from DOORS and are available in NEON’s document repository, or upon request.

5.2 Data Products

Execution of the protocols that stem from this science design procures samples and/or generates raw data satisfying NEON Observatory scientific requirements. These data and samples are used to create NEON data products, and are documented in the NEON Level 1, Level 2 and Level 3 Data Products Catalog (AD[04]).

5.3 Priorities and Challenges for Terrestrial Biogeochemistry

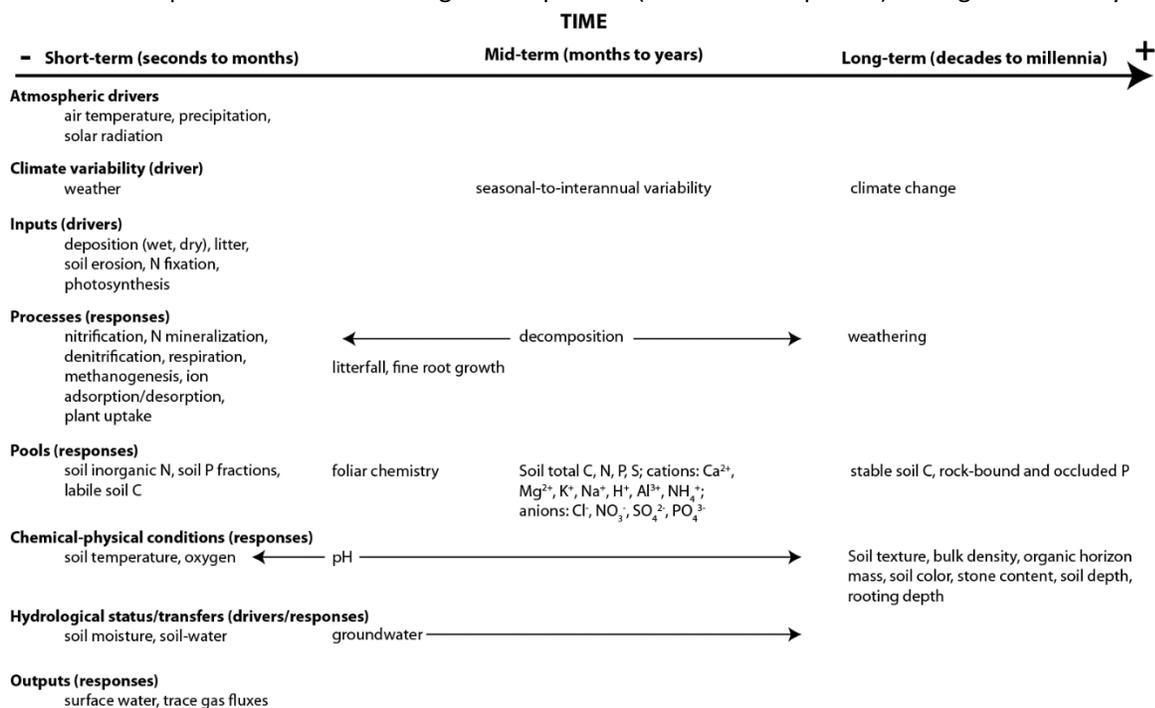
Components of biogeochemical cycles measured by NEON include drivers, such as precipitation, temperature, and soil physical and geochemical characteristics, and responses, such as biogeochemical transformation rates, changes in carbon stocks, and nutrient cycling and uptake by plants and microbes. Drivers, responses, and the feedbacks between them can be viewed through the lens of anticipated timescales of change (see **Table 1**), which inform how they should be sampled over the lifetime of NEON. While most any factor can change abruptly at thresholds, the potential for change is highest at the level of soil biota (microbial communities and larger soil fauna), due to fast turnover times and gene systems that respond to local variations much more so than long-lived vegetation. Soil physical properties (e.g., texture, mineralogy) can be thought of at the other extreme, as they often remain relatively constant for hundreds or thousands of years, in the absence of disturbance. Thus, components of terrestrial biogeochemistry require different sampling strategies, depending upon the scale of measurement and expected spatio-temporal variability; with feedbacks and interactions that produce large-scale patterns that are relatively robust, for example, as the emergence of individual biomes.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

NEON’s terrestrial biogeochemistry sampling strategy could take one of several foci within this framework. Emphasis could be placed entirely on characterizing spatial heterogeneity in C and nutrient stocks across each site, thereby creating a long-term dataset of storage in ecosystem components (soils and plant tissues). This design would rely heavily on collaborators from the larger natural sciences community to conduct short-term process-based studies that would capture the mechanisms underlying these patterns. However, past research has shown that on the order of decades, total ecosystem C and N stocks do not change dramatically in response to chronic ecosystem perturbations like N deposition or climate change (see Magill et al. 1997 compared with Nadelhoffer et al. 2004, Yanai et al. 2003), and heterogeneity in soil physical properties within sites may overwhelm the ability to discern important spatial and temporal patterns. Focusing only on changes in stocks, then, would not address the primary requirements of NEON (**Figure 3**), which include understanding dynamic phenomena within sites. Alternatively, NEON could focus on measuring transformation rates and important fluxes, including soil CO₂ efflux, nitrification, and denitrification, as well as soil water availability and chemistry, which are likely to change in response to climate and other drivers (see Barnett et al. 2005, Cable et al. 2008, Emmett et al. 2004, Hart 2006, Loik et al. 2004). However, focusing exclusively on these measurements would not capture the drivers or feedbacks affecting the results—again, not achieving the stated high-level requirements of NEON—and be extremely challenging from a resource (i.e., labor and financial) perspective.

Table 1. Anticipated timescales of change in components (drivers and responses) of biogeochemical cycles.



Given these considerations, and based on the framework provided in **Figure 3** and **Table 1**, NEON will target key measurements of biogeochemical drivers, responses, and feedbacks at appropriate spatial extents and temporal frequencies. For the terrestrial biogeochemistry sampling design, these key



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

measurements include: a one-time detailed characterization of soil chemical and physical properties, paired with periodic sampling of soil and plant C and nutrient pools, C and N stable isotope ratios, and net N mineralization and net nitrification rates in soils. Additional measurements at the tower and aquatic sampling locations will provide biogeochemical data across systems (air, land, and water). The goal of this approach is to provide a “research backbone” of data resources that will motivate further investigation by the natural sciences community, and provide insight into how ecosystem processes change through time. NEON will also endeavor to illuminate important feedbacks between the drivers and responses of interest. While it would be ideal to sample more of the processes and fluxes known to change over short timescales (i.e., those on the left-most column of **Table 1**) and in response to NEON’s drivers of interest (i.e., climate, land cover/land use, and species composition), the NEON design will only incorporate a limited set of measurements that will grant investigators direct insight into short-term biogeochemical cycling – net N mineralization and net nitrification. Parallel measurements at the tower location, including CO₂ fluxes, will provide further insight into short-term phenomena, and stable isotope ratios will provide information about integrated ecosystem processes over longer timescales. NEON cannot (and should not) measure everything everywhere, and, thus, the sampling strategy for terrestrial biogeochemistry relies on targeting minimal sampling frequencies necessary to capture long-term trends in variables of interest (described below). This approach allows for creating a balance between meeting the scientific requirements of the design and the logistical and financial constraints of conducting field collections and sample processing. The resulting long-term dataset will constitute an important contribution to ecological science, given NEON’s coordinated approach across large spatial and temporal extents.

In general, the intensity of sampling efforts for soil and plant biogeochemistry should be proportional to the anticipated frequencies of change (and variability) for each measurement (**Table 1**). With the exception of soil N transformations, all other measurements within the terrestrial biogeochemistry design are on ecosystem components that are likely to have high spatial variability but relatively low temporal variability. Thus, measurements will be made of plant tissue chemistry (canopy foliage, litter, and roots) and total soil C and N extensively across each site, but only once every five years.

Terrestrial biogeochemical measurements will occur in plots located within the tower airshed (i.e., within close proximity to the tower and within the dominant vegetation type) and for some measurements, also in plots distributed across the site in representative vegetation cover classes. The distributed plots (e.g., **Figure 4**) were specified as the foundation of the TOS Science Design for Spatial Sampling (see (AD[03])). The plan for the distributed plot locations was developed using a combination of classical statistical power analysis and spatially balanced sample allocation via the Reverse Randomized Quadrat Recursive Raster approach (RRQRR, Theobald et al. 2007) (see **Figure 5**). In general, a grid of potential, random plot locations is created for each permitted NEON site area, the area is stratified by vegetation community type, and then a stratified random list of plot locations is generated for the site. This approach enables robust site-level estimates of measured parameters, as described in AD[03]. The data collected within the distributed plots are primarily aligned with the design goals associated with stand-alone TOS data products (e.g., all organismal sampling). Sampling within the



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

tower plots supports the design goals associated with the entire TOS platform, but also provides a means to connect sensor-based measurements on and around the tower to the manual observations made across the site.

All plots sampled for terrestrial biogeochemistry will be 40 m x 40 m. They will include a central 20 m x 20 m core reserved for plant biodiversity sampling, which will exclude soil sampling due to its potential for disturbing the area. Thus, soil sampling will occur within the outer ring of this 1600 m² area, and the degree of collocation between plant and soil biogeochemistry measurements is at the scale of the plot.

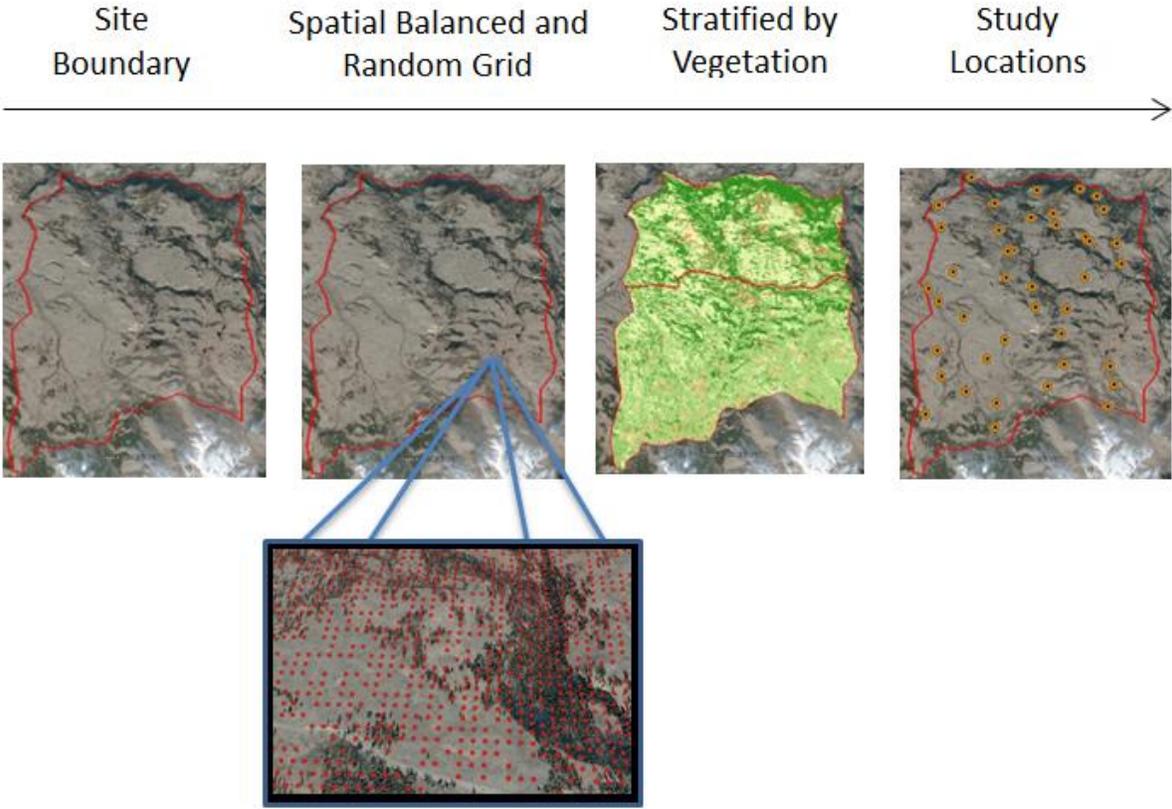


Figure 5. Steps in the RRQRR statistical framework to determine a stratified random distribution of plots across a NEON site.

An important goal for NEON is to provide investigators with sufficient data to distinguish between variability and trends in ecological phenomena across space and time. In order to most effectively address this goal of enabling investigators to resolve complex patterns in space and time, two temporal strategies for sampling will be conducted that incorporate the considerations of temporal variability captured in **Table 1**. These two strategies include 1) an initial soil characterization effort to provide a baseline dataset necessary to inform interpretation of data that fall under the second category, which is 2) regular, periodic sampling during operations of the Observatory (i.e., the lifetime of NEON). Using the framework provided in **Table 1**, the expected timescales of change in drivers and response variables, the measurements described as part of the terrestrial biogeochemistry design fall into these two categories.



<i>Title:</i> TOS Science Design for Terrestrial Biogeochemistry		<i>Date:</i> 04/06/2022
<i>NEON Doc. #:</i> NEON.DOC.000906	<i>Author:</i> E. Hinckley	<i>Revision:</i> C

The following section of this document describes the details of the methodologies, as well as the spatial and temporal sampling strategies for the terrestrial biogeochemistry sampling design.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

6 SAMPLING DESIGN FOR TERRESTRIAL BIOGEOCHEMISTRY

6.1 Sampling Design for Terrestrial Biogeochemical Drivers

NEON's high-priority measurements of controls on biogeochemical cycling include climate variables (e.g., precipitation, air temperature, and solar radiation), precipitation chemistry, soil physical and chemical properties, and soil temperature and moisture. Climate variables and atmospheric deposition will be measured on the tower (part of the TIS) at the majority of sites. In brief, these datasets include: precipitation and dust deposition (amounts and a suite of chemical constituents), stable isotopes of water in precipitation, CO₂ concentration profiles in air and soil (along the length of the tower, and in the top 1m of soil), energy fluxes, continuous soil temperature and moisture, and net ecosystem exchange. In the tower footprint, the TIS also partnered with the Natural Resources Conservation Service (NRCS) to excavate, describe, and sample one soil 'megapit' per site (total maximum depth of 2 m, or bedrock) to provide soil physical and chemical data as well as root depth distributions in the dominant vegetation type.

Measurements that capture a site's *range* of soil physical and chemical characteristics, as well as distributed measurements of soil temperature and moisture, fall under the purview of the Terrestrial Biogeochemistry sampling design. To address the former, prior to full operation of NEON, the NRCS will conduct a one-time, landscape-scale soil characterization effort at each site. Sampling will include the suite of physical and chemical properties summarized in **Table 2** and occur in the majority of plots used to make recurring soil biogeochemical and microbial measurements. These properties create the conditions promoting or suppressing soil microbial activity, affect air space in the soil matrix, influence plant physiological activity and growth, and shape the development of hydrological flow paths, which in turn control redox states and nutrient transport. In addition, they constitute a standard suite of measurements and laboratory analyses made during NRCS soil surveys, and, thus, will be comparable to other soil databases.

Although NRCS and other entities have conducted soil surveys with comparable measurements at the continental scale (e.g., STATSGO and SSURGO), the data are collected at coarser scales than the NEON design (i.e., kilometers versus meters), and the suite of analyses are not consistent across all sample locations. Measurement of soil taxonomic, physical and geochemical characteristics by horizon to at least one meter depth (where bedrock is not less than 1 m from the soil surface) will provide data critical to interpreting shorter-term biogeochemical phenomena (e.g., N transformations), conducting process-based modeling efforts within sites, and broadly characterizing soil environments at the continental scale. At the majority of NEON sites, excavation and description of multiple soil pits by the NRCS will allow for full characterization of the diverse soil forming units present across the landscape (e.g., Vadeboncoeur et al. 2012). However, given permitting restrictions at some NEON sites, soil pits will not be feasible everywhere; in restricted areas, large-diameter soil cores will be collected, described, and analyzed instead. For the distributed soil characterization effort, the number and location of soil pits or cores required to capture the spatial heterogeneity at each site shall be



determined by a team of trained NRCS soil scientists contracted by NEON; a maximum of 34 locations will be sampled per site.

In addition to basic soil characterization data, soil water content and temperature are fundamental controls on biogeochemical processes, and must be measured as part of the baseline data framework. NEON will collect distributed measurements of soil temperature and water content across each site whenever soil cores are collected for biogeochemical or microbial measurements. Previous research by several groups (e.g., Groffman et al. 2012, Savage and Davidson 2001) has documented the importance of soil temperature and water content data, both for the purposes of observation in biogeochemical studies, and for informing models. While continuous data have become increasingly feasible to obtain due to greater affordability of instrumentation, the NEON budget does not include money for purchasing sensors and dataloggers to collect these measurements at multiple locations within each site. Within the tower footprint (i.e., the dominant vegetation type), automated instrumentation to measure soil water and temperature is installed.

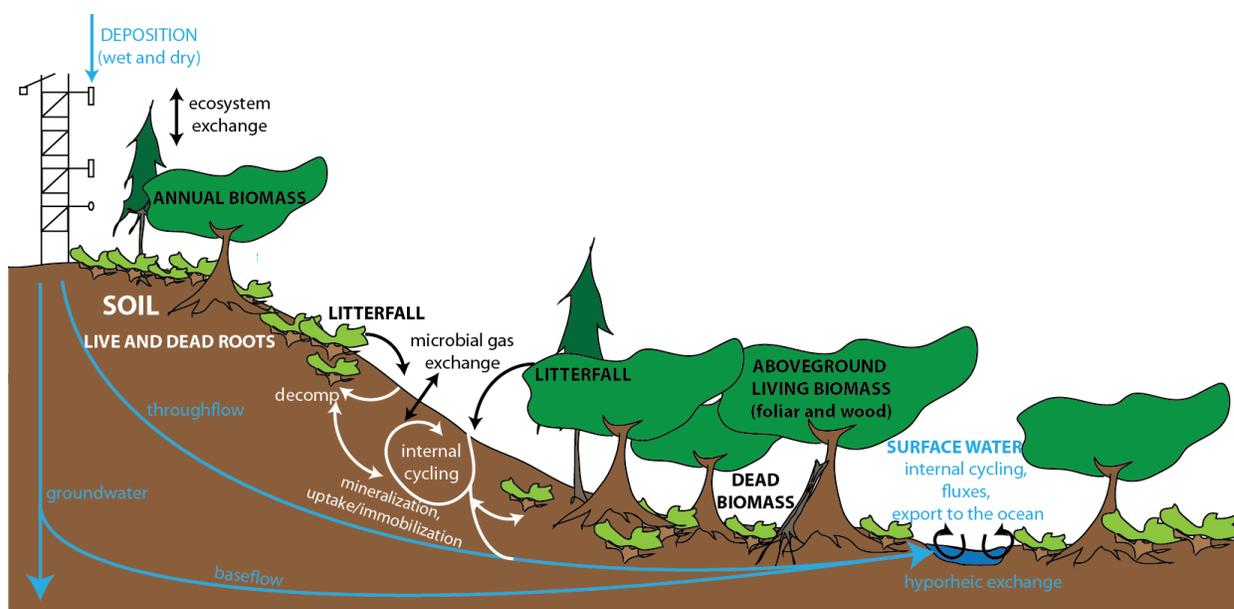


Figure 6. NEON biogeochemistry measurements within a site. Shown are variables that are part of the terrestrial biogeochemistry design (Terrestrial Observation System, TOS), as well as those associated with the tower (TIS), above- and belowground plant biomass (TOS), microbial community (TOS), and ground and surface waters (AOS/AIS). Subsurface hydrological flow paths that are important to constrain in biogeochemical budgets are shown, but will not be measured by NEON.

6.1.1 Sampling Methods

6.1.1.1 Soil Characterization

During the one-time soil characterization effort, field descriptions, soil sampling, and laboratory analyses will be done according to National Cooperative Soil Survey (NCSS) and Natural Resources Conservation Service (NRCS) protocols (see Burt, 2004; Schoeneberger et al., 2012). These approaches represent widely accepted and recognized standards for soil characterization and analysis, and, therefore, are appropriate choices for NEON. **Table 2** shows the soil analyses that will be completed, and the methods that will be used. Contracted soil scientists local to each NEON domain will determine the locations to sample using their knowledge of soils at each site, as well as vegetation community and topographical information. Approximately 10-34 soil pit or coring locations will be chosen, based on the size and spatial heterogeneity of the site (the upper limit of locations is constrained by available budget for this effort). Final reports, including site-level summaries that place the sampled soils in the broader context of soils and geomorphology for the entire NEON site and pedon descriptions containing pit-level observations and field measurements, will be produced from the characterization work. In addition, physical and chemical analyses by soil horizon (listed in **Table 2**) will be made available to NEON data users via the NEON data portal. If resources permit, soil samples by horizon may be archived according to NCSS protocols (air-dried, sieved to $\leq 2\text{mm}$, and stored in a cool, dry location).

Table 2. Soil chemical and physical properties measured during the one-time characterization effort at each NEON site.

Analysis	Method ^a
Particle size distribution	Hydrometer method
Bulk density	By appropriate method; TBD based on sampling restrictions at site
pH	In water and CaCl_2
Total organic C; total N, S	Combustion; elemental analyzer
Inorganic C (CaCO_3) ^b	Acidification with HCl and measurement of evolved CO_2
Phosphorus	By appropriate method; Bray or Mehlich III, based on soil pH
Extractable cations and cation exchange capacity (CEC)	Ammonium acetate extraction
Aluminum ^c	Potassium chloride extraction
Electrical conductivity (EC)	Prediction and saturated paste
Cations and anions ^d	Saturated paste
Fe, Al, Mn, Si, P	Acid oxalate extract
Fe, Al, Mn	Citrate dithionite extract

^a Methods described in brief. All laboratory analyses and QA/QC procedures will follow the Soil Survey Laboratory Methods Manual (Online: http://www.nrcs.usda.gov/wps/portal/nrcs/detail/soils/research/lab/guide/?cid=nrcs142p2_054247)

^b Inorganic C will be measured only for samples containing CaCO_3

^c Extractable Al will be measured only for samples with $\text{pH} < 5.5$

^d Saturated paste extracts will be collected for measurement of cations and anions for samples predicted to have measurable amounts by standard methods.

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

6.1.1.2 Soil Temperature and Water Content

Soil temperature and water content will be measured manually at each site during regular sampling of soils for biogeochemical and microbial measurements (temporal frequencies described below). At each soil sampling location, soil temperature will be measured in the top 10 cm using a surface temperature probe (stainless steel with digital readout). Soil water content will be measured in the laboratory on subsamples of composite soil samples using the widely accepted thermogravimetric method (Topp and Ferré, 2002). Measurements of soil water content will be made on the top 30 cm of soil (see below), separated and analyzed by organic (if present) and mineral horizons. The procedures for field and laboratory components of this sampling are provided in TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling (AD[07]).

6.1.2 Spatial Distribution of Sampling

The group contracted by NEON to conduct the field sampling and laboratory analyses for the one-time soil characterization effort will determine the spatial distribution of sampling locations of soil pits/cores. Based on the spatial heterogeneity (i.e., topography, expected variability of soil types, and vegetation communities) and size of each site, 10-34 NEON base plots will be chosen for descriptions and sampling of soils. These locations will be chosen to represent the range of soil types within each site and are preferentially selected to maximize collocation with NEON plant and soil biogeochemical measurements. The spatial distribution of soil temperature and water content measurements are described below in the section on Soil Biogeochemistry (Section 6.2).

6.1.3 Temporal Distribution of Sampling

During the construction and/or initial operations phase of NEON, the one-time soil characterization survey will occur at each site. The physical and chemical analyses of soils by horizon (listed in Table 2), together with the field descriptions and interpretations, will provide baseline soils data and inform the sampling strategies for regular, periodic measurements of soils by NEON. The temporal distribution of soil temperature and water content measurements match the temporal distribution of NEON periodic soil sampling in general, as anytime soil samples are collected for either microbial or biogeochemical measurements, they are analyzed for temperature and water content. This often equates to three sampling time-points per year, though some sites with short growing seasons only have one. How those time-points are selected is described below in the section on Soil Biogeochemistry (Section 6.2).

6.2 Sampling Design for Soil Biogeochemistry

NEON's high-priority soil biogeochemical measurements include concentrations and stable isotope ratios of total organic C and total N and targeted measurements of N transformation rates. Measurements of the soil microbial community are directly linked (i.e., made on subsamples from the same soil cores) and are described in detail in TOS Science Design for Microbial Diversity (AD[05]). Throughout the lifetime of each observatory site, NEON will measure total organic C and N



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

concentrations, which can be used to estimate total soil C and N stocks when combined with the bulk density data from the soil characterization work described above. A focus on carbon and nitrogen measurements is important as soils are a major reservoir for these elements, C and N concentrations and ratios influence microbial activity and plant growth, and fluctuations in soil C and N dynamics can shed light on ecosystem health and response to global change. Measurement of these chemical constituents in soils is consistent with other components of the NEON biogeochemistry design, including measurements of plant tissues, ecosystem inputs at the tower location, and export in surface waters.

The resulting data will enable investigators to analyze ecosystem stoichiometry, and to infer changes in sources and process rates that may explain patterns in the data. For example, following C:N ratios in soils over time is a useful indicator of microbially-mediated transformations (e.g., Kaye and Hart 1997) such as decomposition (Melillo et al. 1989), and correlate with ecosystem losses (Tietema and Beier 1995, Gundersen et al. 1998, Aitkenhead and McDowell 2000, Lovett et al. 2002). Data on total elemental stocks can also lend insight into identifying where nutrients are stored and released within a landscape and across different ecosystem types within NEON. By combining elemental concentration data with analyses of soil C and N stable isotopes, researchers will be able to address questions related to understanding the mechanisms underlying different nutrient sources (e.g., Phillips and Gregg 2003), soil organic matter turnover rates and decomposition extent (e.g., Marín-Spiotta et al. 2009, Bernoux et al. 1998), and integrated analyses of ecosystem processes (e.g., West et al. 2006, Robinson 2001) and loss pathways (Houlton and Bai 2009, Bai et al., 2012).

The rates of soil biogeochemical processes are expected to change in response to shifts in climate forcing, influencing the amount and forms of C and nutrients that move across systems (Gruber and Galloway 2008). Many previous studies from a wide range of disciplines within Earth and environmental sciences have documented the short-term sensitivity of process rates to changing ecosystem drivers (Barnett et al. 2005, Cable et al. 2008, Emmett et al. 2004, Hart 2006, Loik et al. 2004), as well as long-term controls (e.g., Barford et al. 2000, Richardson et al. 2007). While it would be extremely informative for NEON to quantify a large suite of biogeochemical process rates distributed across each site, including soil CO₂ efflux, nitrification and denitrification, net N mineralization, as well as inorganic nutrient pools, it is neither logistically nor financially feasible in the observatory context. Instead, NEON will provide insight into microbially mediated processing of N through periodic measurement of net N mineralization and net nitrification in soils. Net N mineralization is a measurement of the amount of inorganic N (ammonium and nitrate) released from organic material over time. Net nitrification is the measure of nitrate converted from ammonium over time. In both cases, these rates reflect net accumulation, so do not account for uptake and loss pathways within soils. Nitrogen process rates reflect the lability of substrate in soils, the activity of the microbial community, and the potential for N loss to the atmosphere (via denitrification) or export to aquatic ecosystems (e.g., as nitrate).

There will be many ways that these data can be analyzed by NEON data users, and, at the spatial extent of NEON, they will provide important insight into nutrient cycling in a variety of systems. Previous studies have found that more easily measurable ecosystem parameters, such as foliar N content and soil



C:N ratios, can be used to predict N processing rates (see Ollinger et al. 2002; **Figures 6 and 7**). NEON will be measuring these parameters as well, and will be able to provide data to inform these predictive relationships over a larger suite of ecosystem types and spatial extent than has been done before.

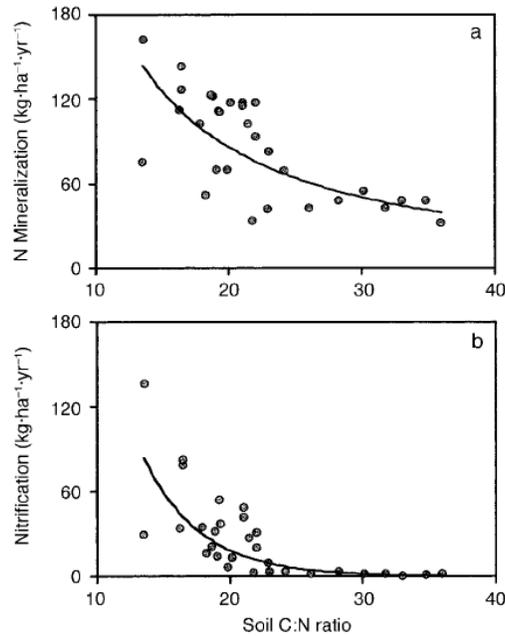


Figure 7. Soil (a) net N mineralization and (b) net nitrification in relation to soil C:N (top 10 cm). Figures are from Ollinger et al. (2002).

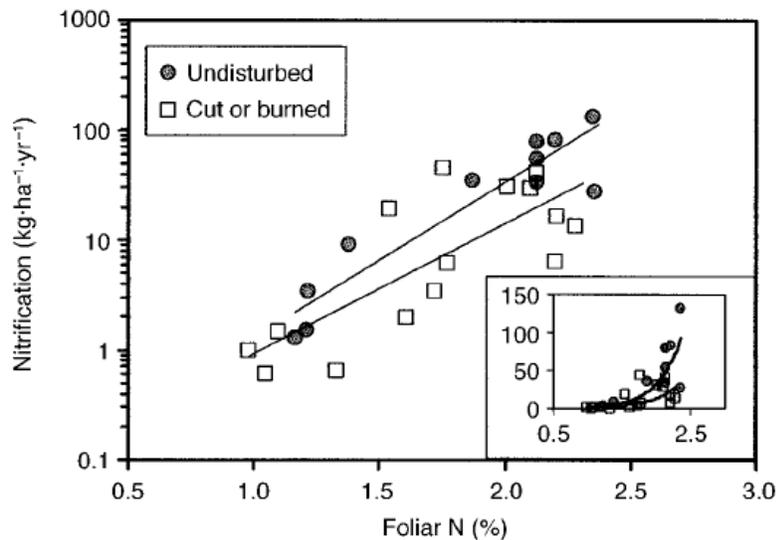


Figure 8. Net nitrification in relation to foliar N concentrations for disturbed and undisturbed forest stands. Inset shows the nonlinear response of nitrification. The trends are described by $\log(\text{NO}_3) = 2.75(\text{foliar N}) - 2.96$, ($R^2 = 0.81$, $P < 0.001$) for undisturbed stands; $\log(\text{NO}_3) = 2.94(\text{foliar N}) - 4.28$, ($R^2 = 0.63$, $P < 0.001$) for disturbed stands. Figure is from Ollinger et al. (2002).



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

Another potential use of these results is to assess relationships between process rates and soil microbial functional genes. Shotgun metagenomic analyses can be used to identify the relative abundance of particular functional genes involved with biogeochemical cycles. Therefore, comparison of direct measurements of biogeochemical transformation rates to functional gene abundance (the latter being part of the microbial ecology component of the NEON design) may help to elucidate links between microbial community functional potential and ecosystem processes. In marine systems, N cycling rates have been associated with levels of functional gene expression (Stewart et al. 2012, Mou et al. 2011, Urich et al. 2008), yet relationships between soil gene expression and process rates are unclear. NEON will preserve samples in a state that will allow others to explore this avenue of research.

6.2.1 Sampling Methods

6.2.1.1 Soil Carbon and Nitrogen Pools

Soil cores for chemical and isotopic analysis will be collected manually in tower and distributed plots at each NEON site. NEON field technicians will be trained in basic soil horizon identification (i.e., organic versus mineral soil horizons, where organic horizons are present) and handling and preliminary processing of soils for biogeochemical analyses. In general, the top 30 cm of the soil profile will be sampled for chemistry, stable isotopes, and N transformations. Where organic horizons are present, they will be separated from mineral soil, and then processed and analyzed independently. Details on the field sampling and laboratory processing of soils are provided in TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling (AD[07]) and TOS Standard Operating Procedure: Wetland Soil Sampling (AD[08]). For analyzing the chemistry of bulk soils, NEON will use analytical methods that are widely accepted in the soil science and ecosystem biogeochemistry communities (i.e., consistent with NRCS and SSSA standards) and are summarized in **Table 3**. All field and laboratory data will be subject to QA/QC procedures detailed in the NEON data product user guides, available on the NEON data portal (<http://data.neonscience.org>). All laboratories contracted with NEON for this work will need to demonstrate established QA/QC protocols and will be periodically audited by NEON Calibration/Validation personnel.

Table 3. Soil biogeochemical analyses and laboratory methods.

Analysis	Method ^a
Total organic C; total N; $\delta^{13}\text{C}$, $\delta^{15}\text{N}$	Combustion, elemental analyzer coupled to isotope ratio mass spectrometer – carbonate removal as required for aridland sites
Ammonium	Potassium chloride extraction, colorimetric analysis
Nitrate	Potassium chloride extraction, colorimetric analysis
pH	In water and CaCl_2

6.2.1.2 Soil Nitrogen Transformations

Soil cores for N transformations will be collected in tower and distributed plots and organic and mineral horizons will be separated (if an organic horizon is present) for analysis. Quantifying net N mineralization



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

and nitrification rates is traditionally accomplished using soil incubations either in the field or laboratory (see Eno 1960, Binkley and Hart 1989). For field incubations, two soil cores are collected, one is transported to the laboratory for immediate extraction and analysis of ammonium and nitrate, while the other is put in a covered core or plastic bag and replaced in the soil borehole. Depending upon the investigator's study question and/or the ecosystem, the incubated core is removed one week to one or more months later and processed in the laboratory using the same process as the initial, paired soil core. To calculate net N mineralization, the final and initial masses of ammonium plus nitrate per unit dry soil (or nitrate only for net nitrification calculations) are then differenced, and a rate of production (usually per day) is reported by dividing the difference by the total incubation period. For laboratory incubations, one soil core is collected in the field and then subsampled for initial extraction of ammonium and nitrate and for incubation. The incubated soil is kept in the laboratory in conditions similar to the field and moisture is maintained according to an initial weight. At the end of the incubation period, a subsample of the core is extracted for ammonium and nitrate, the extractant is analyzed, and calculations are performed as for field incubated soil cores.

There are advantages and disadvantages to both field and lab incubation approaches. Field incubations can be done using a variety of approaches (e.g., close-top core, buried bag, resin). Generally, they maintain field temperature, but the water balance of the core can be influenced by the method. They are thought to provide better quality information than lab incubations (see discussion in Binkley and Hart 1989), but they are more labor intensive because they require an individual to return to the sampling location to retrieve the soil core. In contrast, laboratory incubations are more streamlined in terms of the workflow (i.e., they do not require a return trip to the field site), but it is more difficult to maintain field conditions in the incubated soil core. Some studies have shown strong correlations between lab and field incubated soils for determination of N processing rates (e.g., Ollinger et al. 2002), making the more easily conducted lab incubation viable. However, other studies have found stark differences between the results of incubations under field and lab conditions (e.g., Johnson et al. 1980).

NEON requires standardized approaches in order to achieve the best degree of comparability among datasets from all sites. To this end, NEON will perform field incubations of soils for net N mineralization and net nitrification, not laboratory incubations. In addition, NEON will use standard laboratory processing and analysis of all soils: potassium chloride extraction of ammonium and nitrate at all sites. In the case of the extraction method, there is the potential in some locations for results to be influenced by using one standard approach. However, the alternative of using site-specific protocols initially does not permit data users to determine whether observed differences are due to methodology or real behavior. Once data are available and users can evaluate patterns at the broad spatial scale that NEON represents, they can also conduct follow-up studies to test whether or not methodological effects are present. Constraints at NEON sites may require some site-specific modification of field incubation techniques (e.g., close-top core vs buried bag). The approaches for field sampling and laboratory processing of soils for N transformation rates are described in detail in TOS Protocol and Procedure: Soil Biogeochemical and Microbial Sampling (AD[07]) and TOS Standard Operating Procedure: Wetland Soil Sampling (AD[08]).



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

6.2.1.3 Samples for Archiving

Subsamples of soil collected for analysis of bulk C and N concentrations and stable isotopes as well as N-transformation rates will be air-dried and archived for access by the larger natural sciences community. This will allow researchers to measure other constituents not analyzed by NEON during the course of routine biogeochemical sampling. In addition to these dried archived samples, frozen (i.e., -80°C) soil subsamples from soil microbial collections will be archived, and these samples are splits from the soil cores collected for biogeochemical analyses and/or N-transformation rates when those occur.

Table 4. Archived samples from the soil biogeochemistry sampling effort.

Sample	Processing	Storage conditions	Volume or mass stored per sample
Soil used to measure bulk C and N and N- transformations ^a	Air-dried, sieved to ≤ 2 mm	Ambient, dry cabinet	Remaining soil following all subsampling (up to 250 g)
Soil used to measure bulk C and N and/or N- transformations	Flash-frozen upon field collection (unsieved)	Cryostorage (-80°C or similar)	Several grams fresh mass

^a Subsamples of KCl extracts from N transformation measurements will not be archived.

6.2.2 Spatial Distribution of Sampling

Measurements of soil C and N concentrations, stable isotopes, and soil N transformation rates will be made in four tower plots and six distributed plots at each site. These plot numbers are based on budgetary constraints for field labor and chemical analysis of samples, yet should still provide sufficient information to discern important spatial patterns in phenomena of interest. This design will enable the NEON user community to analyze data in a variety of ways, including comparing variables of interest across vegetation types (e.g., soil C:N in the dominant compared with secondary vegetation communities) and relating concentrations of total C and N to N transformation rates and microbial community composition. Additionally, it will allow users to employ geostatistical approaches to estimate robust site-level means for soil biogeochemical parameters. Within each plot, a center 20 m x 20 m subplot for plant biodiversity sampling will be maintained, and soil sampling will only occur outside of this area within the larger 40 m x 40 m plot. Collection of soil cores will be made at three random locations within plots during each collection event. In the event that the number of soil samples for measuring N transformations is too great for the time budgeted by the Field Operations team, NEON will reduce the number of plots sampled, but preserve the temporal resolution (described below).

6.2.3 Temporal Distribution of Sampling

Soil C and N concentrations and stable isotope ratios are not likely to be highly temporally variable at NEON sites. In other long-term projects, such as the chronic N addition study at Harvard Forest LTER, minimal changes in total N content and distribution were detected in soils after a decade of elevated N deposition (see Magill et al. 1997 and Nadelhoffer et al. 2004). Thus, NEON will sample these



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

constituents once every 5 years in the tower and distributed plots to capture long-term trends at each site. In years when sampling occurs, samples will be collected during each site's period of historic peak greenness, using splits from the same soil cores used for initial N transformation measurements and microbial community analyses.

Greater inter-annual, as well as intra-annual sampling frequencies would be optimal to document the dynamics of soil N transformation rates, which are expected to undergo short-term changes in response to NEON's drivers of interest (see **Table 1**). However, these measurements are also very labor-intensive, and the NEON budget does not accommodate making them every year. To provide useful information on the seasonal changes associated with N transformations, while compensating for the constraints of the project, NEON will sample soils to calculate rates of net N mineralization and net nitrification every 5 years, but within the sampling year, conduct soil core incubations three times. The three sampling periods will be during historic peak greenness at each site, coincident with the soil C and N measurements described above, as well as during seasonal transitions that allow for capturing a range of temporal dynamics. Examples of these transitions include the winter-spring transition in locations where a seasonal snowpack develops (e.g., the Intermountain West and the Northeast) and the dry to wet transition in arid or Mediterranean environments (e.g., the Southwest, California coast). If possible, the initial 3 years of data collected in this manner will be analyzed and reviewed by a group of community experts and/or "NEON power users" in order to optimize the sampling strategy over time.

6.3 Sampling Design for Plant Biogeochemistry

NEON will quantify variation in plant tissue chemistry at plot to site scales using a combination of ground- and airborne-based methods. Ground-based collections will include sampling and analysis of total C and nutrient concentrations (N, P, S, Ca²⁺, Mg²⁺, K⁺, and micronutrients) in sun-lit foliage of dominant and co-dominant canopy species, and C and N of litter and roots. Stable isotopic composition ($\delta^{13}\text{C}$ and $\delta^{15}\text{N}$) of these plant tissues, as well as other chemical characteristics (i.e., chlorophyll in canopy foliage and acid unhydrolyzable residue (chemical extraction of lignin) in canopy foliage and litter) will also be measured (see **Table 5**). These data will provide valuable information about nutrient uptake and storage in terrestrial ecosystems, both above- and belowground. In parallel with plant tissue sampling for chemistry, the data needed to make above- and belowground biomass estimates will be available through the collection of plant species composition and structural data, leaf area index, leaf mass per area, and collection of litterfall and root biomass (see TOS Science Design for Plant Diversity (AD[06]) and TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[09])). These data are necessary for calculating plant nutrient stocks (e.g., Pardo et al. 2005, Yanai et al. 2010) and estimating production rates, which drive biogeochemical models of the C cycle.

In 2012, NEON conducted a prototype effort in Domain 1 (Harvard Forest core site) and sampled canopy foliar tissues across a gradient of plant community structure (i.e., forested, mixed-stature, and disturbed/grassland). This gradient also included differences in plant groups (i.e., phylogenetic and photosynthetic pathways). One objective of the prototype was to determine whether sampling across a gradient of vegetation communities would yield differences in concentrations and stable isotope ratios



of C and N in plant canopy tissues among these groupings. **Figure 9** shows separation of plant groups by $\delta^{13}\text{C}$ values and N content, as well as C:N ratios. Data are separated according to phylogenetic groups, as well as photosynthetic pathways (i.e., the grasses included are C4). These data illustrate some of the patterns across vegetation communities that NEON data users will be able to explore within and among sites.

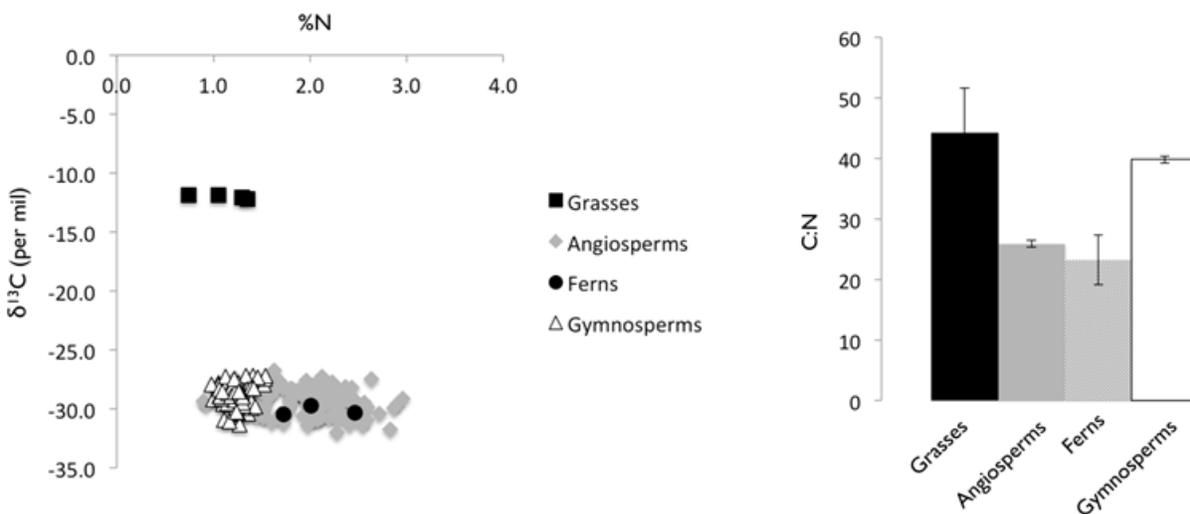


Figure 9. Results from the Domain 1 (Harvard Forest) NEON Prototype in 2012. (a) Percent N and $\delta^{13}\text{C}$ of plant canopy foliar tissues by plant phylogenetic group (i.e., angiosperms, ferns, and gymnosperms) and photosynthetic pathway (i.e., C3 and C4), and (b) Mean (± 1 SE) C:N ratios across the same groups.

Airborne observations of plant canopy chemical and structural characteristics will be collected most years across each NEON site, contingent on availability of funds. The ground-based assessments of plant chemical characteristics will be necessary to interpret the hyperspectral data. Other research groups have successfully integrated airborne remote sensing data with ground-based measurements of canopy nutrient concentrations to describe variation in the composition of some plant nutrients (especially N) and biochemical characteristics (especially pigments) at large spatial scales (Asner and Vitousek 2005, Ollinger et al. 2002, Wang et al. 2010, 2007). Relating the ground-based plant chemistry and structural measurements to the remote sensing data will rely heavily on algorithms developed by the research community. The integration of these data streams will be critical to create large-scale data surfaces for modeling efforts at regional to continental scales. Foliar chemistry data are needed to constrain leaf photosynthetic capacity in ecosystem models. Generally, these data are unavailable for Ameriflux or Fluxnet tower sites. Thus, when model simulations are compared to tower sites, it is not always clear whether or not photosynthetic capacity is represented appropriately; data from NEON sites will help to constrain this and other ecosystem variables.

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

6.3.1 Sampling Methods

6.3.1.1 Plant Tissue Collection

Samples of sun-lit canopy foliar tissues will be collected in both tower and distributed plots at NEON sites during the period of historic peak greenness and coincident with data collection by the NEON remote sensing team. At forested sites, NEON field personnel will obtain samples of dominant and/or co-dominant species using methods permissible at the site (e.g., shotgun, slingshot, line launcher, tree climbers, or pole pruner). At grassland sites, samples (species bulked) will be obtained by clipping all aboveground biomass in a given area. At savanna and shrubland sites/areas, canopy samples will be collected using a combination of these approaches.

In forested and shrubland plant communities, field personnel will identify species and individuals for sampling. In each designated plot, the top three dominant and/or co-dominant canopy species will generally be selected for sampling of leaf mass per area (LMA) and the suite of chemical constituents in **Table 5**. If a plot has low diversity, then less than three unique species may be collected. If a plot has high diversity and no clear dominants (or the dominants have already been sampled with replication), a species that is a member of the site-level over-story but has yet to be sampled may be selected. Priority will be given to individuals tagged for long-term measurements of vegetation structural characteristics, and individuals previously sampled for canopy foliage will be re-sampled in subsequent years. A trained botanist on the NEON domain field staff will identify plant species present and help the canopy foliar sampling team identify target taxa for sampling.

In grasslands, field personnel will use random coordinate lists to locate and sample sun-lit clip harvest strips, generally one per plot (except for large plots, which may have two). Field personnel will cut and collect all sun-lit, standing aboveground biomass in relatively long but narrow strips (0.1 m x 2 m), as this approach enables better sampling of vegetation heterogeneity (Sala and Austin 2000). Wider strips are permitted in sparsely vegetated areas in order to collect enough biomass for analysis. The bulk foliage will be homogenized, then subsampled for leaf mass per area measurements and all laboratory analyses (see **Table 5**). In savanna ecosystems (i.e., grasslands dotted with trees), trees within plots will be sampled as above, and clip harvests of herbaceous biomass will also be performed, ensuring collections are from fully sun-lit areas. Further details on the field sampling of foliar tissues can be found in TOS Protocol and Procedure: Canopy Foliage Sampling (AD[10]).

Within plots located in the tower footprint, NEON will sample root biomass and litterfall to determine components of above- and belowground production. The approaches for sampling these tissues are described in TOS Protocol and Procedure: Plant Belowground Biomass Sampling (AD[11]) and TOS Protocol and Procedure: Litterfall and Fine Woody Debris (AD[12]). Generally, they follow the approaches of the North American Carbon Program (see Hoover 2008), and, thus, data collected over a similarly large spatial scale (i.e., continental) will be comparable across efforts. In the case of litterfall, NEON will obtain samples in traps, material will be sorted to functional group (e.g., needles, leaves), then it will be dried and weighed. During litterfall biogeochemistry bouts, ground, homogenized

subsamples will be analyzed for total C and N concentrations, stable isotopes ratios, and lignin. Root biomass will be sampled from known volumes, then fine roots will be sorted by live and dead into several size classes up to 10 mm diameter. Analyses of C and N concentrations and stable isotopes will be performed on homogenized (dried, ground), live root subsamples.

Table 5. Laboratory analyses and methods for plant tissues.

Plant Tissue	Analysis	Method
Canopy foliage	Total organic C; total N; $\delta^{13}\text{C}$, $\delta^{15}\text{N}$	Combustion, elemental analyzer coupled to isotope ratio mass spectrometer
	P, Ca^{2+} , K^+ , Mg^{2+} , S, micronutrients	Nitric acid and hydrogen peroxide digestion; Inductively Coupled Plasma (ICP)
	Chlorophyll	Spectroscopy
	Lignin	Acid Unhydrolysable Residue (see Ryan et al. 1990 and McClaugherty et al. 1985)
Roots	Total organic C; total N; $\delta^{13}\text{C}$, $\delta^{15}\text{N}$	Combustion, elemental analyzer coupled to isotope ratio mass spectrometer
Litter	Total organic C; total N; $\delta^{13}\text{C}$, $\delta^{15}\text{N}$	Combustion, elemental analyzer coupled to isotope ratio mass spectrometer
	Lignin	Acid Unhydrolysable Residue (see Ryan et al. 1990 and McClaugherty et al. 1985)

The primary decisions involved with choosing methods for sampling plant tissues for biogeochemistry were related to the design rather than to sample acquisition or laboratory techniques. Methods of collection for these types of plant tissues are straightforward (i.e., cut the sample and bag it), and the chosen laboratory techniques are general practice within the community. One decision of note related to the foliar sampling design was whether the sample unit is a georeferenced individual, or the plot. As part of the sampling designs for Phenology and Vegetation Structure (see TOS Science Design for Plant Phenology (AD[13]) and TOS Science Design for Plant Biomass, Productivity, and Leaf Area Index (AD[09])), NEON will tag some individuals within the site for long-term sampling. Providing chemical data at the scale of the individual, with repeat sampling of tagged, canopy-dominant trees and shrubs over time, will be useful in detecting temporal changes in foliar properties while also enabling site-level links with NEON's airborne remote sensing data, given the latter's high spatial resolution (~ 1 m² pixel size). However, this imposes a fixed design from the initial years of the Observatory (i.e., individuals are chosen for sampling in year one) and does not allow for sampling species that may become a dominant part of the community over the course of 30 years. In the case of the latter, incorporating new individuals, if appropriate, will allow for continued evaluation of the most representative species and individuals for canopy foliar chemistry sampling over the lifetime of the Observatory, and will more closely match the airborne data collection if the plant community changes over time.

While the sampling unit is an individual, all trees and shrubs will still be collected from within NEON plots, both for logistical (e.g., permitting) reasons, as well as to facilitate links with long-term vegetation

phenological and structural measurements. Any new or previously sampled individuals will be georeferenced at the stem to within ± 1 m. With this approach, users will be able to combine ground-based and airborne data to analyze foliar trends at the scale of the individual and the site over time.

A second important decision related to plant tissue sampling was whether to sort foliar chemistry clip harvests and litter samples. The decision to not sort foliar chemistry clip harvests at all and to only sort litter to functional groups was largely due to field and laboratory budgetary constraints. It will be impractical and financially unfeasible to sort and chemically analyze foliage samples to the level of species or functional group in diverse grasslands. In addition, the time required to process litter samples for chemical analysis (i.e., homogenizing, grinding, and subsampling) would be too great if required at the level of individual species. NEON’s approach to provide C, nutrient, and stable isotope data bulked across all aboveground biomass for grassland foliage and functional groups for litter will likely provide sufficient detail for most data users, and meets the overall requirements of the NEON Project. Data on species diversity and abundance will be provided by NEON for each site, following the TOS Science Design for Plant Diversity (see (AD[14])), so if separation by species for chemical analyses of plant tissues is desired, the community will have the baseline information to support proposing further research.

Other design-related decisions are for aspects of the spatial distribution and temporal frequency of plant tissue sampling, which are described below.

6.3.1.2 Samples for Archiving

NEON will archive remaining plant material (oven-dried and ground) following subsampling for analysis of C, nutrients, and stable isotopes. In addition, whole leaf specimens from initial plant diversity assessments and stem-mapping efforts will be dried and archived in envelopes.

Table 6. Archived samples from the plant tissue collection efforts.

Sample	Processing	Storage conditions	Volume or mass stored per sample
Canopy foliar tissues	Oven-dried (65°C, 48 hr), ground	Ambient, dry cabinet	20 ml vial
Litterfall	Oven-dried (65°C, 48 hr), ground	Ambient, dry cabinet	20 ml vial
Roots	Oven-dried (65°C, 48 hr), ground	Ambient, dry cabinet	20 ml vial

6.3.2 Spatial Distribution of Sampling

The spatial design for plant tissue collections differs based on the tissue type. Sun-lit canopy foliage samples will be collected from four tower plots and approximately 10-16 stratified, random distributed plots per site, depending on the predominant vegetation cover (as in AD[03]). As described above, in

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

forests and shrublands, individuals will be chosen for sampling that are representative of the canopy community within each plot. In grasslands, one or two random locations for bulk vegetation clip harvest strips will be identified, and in savannahs, a combination of the two methods will be employed.

Collections of litter and root tissues for biomass and chemical analysis will be within the tower plots only. The reasons for this decision are that (1) these measurements will enable estimation of plant belowground biomass and aboveground productivity in the same landsurface area as NEON Tower eddy covariance data, enhancing the ability to intensively study the carbon cycle at each site, and (2) the time needed for field technicians to collect and process these tissues for weights and chemical analyses requires that they be limited in their spatial extent to the dominant vegetation community only (i.e. the one captured in tower plots). Thus, the spatial sampling design for biogeochemical analysis will be to collect root cores from one or two randomly located clip strips, and detrital material in one or two litter baskets, in all 20-30 tower plots per site.

For all tissue types, sampled plot numbers may change based on statistical analyses of early data, if it appears that more or fewer plots are needed to make robust site-level estimates of vegetation biomass or biogeochemical parameters.

6.3.3 Temporal Distribution of Sampling

Every 5 years, NEON will measure the chemistry of plant tissues (foliage, litter, and roots) at a given site. This equates to roughly 20% of sites sampled per year across the Observatory. In the early phases of NEON, the TOS and AOP remote-sensing teams may work together to optimize the design and determine which sites those are, based on community or technical working group feedback. While the chemistry of foliar tissues, in particular, may change annually in response to climate drivers (e.g., rainfall, air temperature), disease, and other pressures, it is not logistically or financially feasible for NEON to sample at a finer temporal resolution. However, it is likely that models linking foliar traits to spectral properties can be developed by the community, allowing for extension of the temporal and spatial coverage of foliar measurements since the NEON AOP collects data for each field site during most years. During plant biogeochemistry sampling years, foliage samples will be collected once during the period of historic peak greenness and in conjunction with AOP overflights, while root samples will be collected once during the period of historic peak biomass at each site. Litterfall traps will be deployed continuously and collected periodically (depending on the vegetation present) at each site in all years. However, subsamples for chemical analysis of litterfall will be prepared once every 5 years, either from a peak senescence collection bout in deciduous sites or a fall bout in coniferous sites. To increase the utility of inter-annual biogeochemical measurements, all of the plant tissue (and soil) sampling will occur at the same sites during the same years.

6.4 Logistics and Adaptability

One of the strengths of NEON is that a standard suite of measurements with the same methodological approaches will be implemented at all sites. Therefore, a prioritization scheme does not limit the



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

number of NEON sites at which samples will be collected, but identifies the measurements and priority analyses that should be conducted at all sites. The design targets minimum within-site sampling frequencies and key temporal periods of measurement necessary to provide a meaningful dataset. Sampling timing and frequency will be site-specific for some measurements, in order to measure responses to particular seasons or in response to growing season constraints. Sampling number will also be site-specific to capture spatial heterogeneity. Once this sampling design has been reviewed and implemented at NEON sites, initial data may indicate that there are pressing site- or region-specific measurements that should be added to address an important ecological issue or question, or other measurements should be removed; such modifications to the design approach—including changes in the priority of data types, sampling timing, sampling frequency, or spatial representation—may be incorporated once the initial design has been put in place across the observatory sites and upon review of the initial data streams by NEON personnel and community experts.

Generally, for the terrestrial biogeochemistry design, budgetary constraints drive limitations in the suite of analyses listed in Tables 3 and 5, the minimum inter- and intra-annual sampling frequencies for measurements of soil N transformations, in particular, and the ability to incorporate measurements of biogeochemical cycling into the design, more generally. Over the lifetime of the Observatory, it may be necessary to reallocate resources among the focal measurement areas within the terrestrial biogeochemistry design (or across NEON measurement platforms more broadly) to optimize the value and usability of the available data. In particular, it is likely that more frequent measurements of short-term biogeochemical transformations and fluxes (e.g., soil water and trace gases), which are not currently part of the design, may be necessary. These measurements are highly variable in space and time, as well as sensitive to small changes in environmental drivers. They are also a critical part of the science design, with respect to meeting the high-level NEON science requirements listed in Figure 3; the design described herein allows for the measurement of two key N cycling processes, in order to get some insight into short-term behavior at each site. In terms of detecting change over decadal time scales, and providing useful information to understand how ecosystems respond to global change drivers, these data will likely be a very valuable part of the terrestrial biogeochemistry design. As discussed above, the one-time baseline soil characterization effort and the regular, inter-annual measurements of soil and plant tissue C, nutrients, and stable isotopes are important for providing a baseline dataset, and resources must be allocated to these efforts, at minimum. They will add value to the interpretation of several other NEON datasets, such as soil microbial community analysis, chemistry of surface waters, and hyperspectral data from the airborne observations, as well as many projects by NEON collaborators.

The first assessment of design feasibility will come from the NEON domain field personnel in charge of collecting and processing samples for analysis by contracted laboratory facilities. Most sampling efforts for the terrestrial biogeochemistry design will require several days to a week of dedicated field time (i.e., soil and plant tissue sampling) for two technicians to complete 1 site, depending upon the complexity of each site. In addition to field time, domain personnel will also conduct sample processing (e.g., soil sieving, drying, subsampling) prior to sending samples to contracted laboratories. The most time-



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

sensitive post-field processing will be for soil N transformation rates. This procedure requires that technicians sieve soil cores, weigh subsamples, extract subsamples in 2M potassium chloride solution, and then filter the extractant for shipment to a contracted laboratory within 24 hours of sample collection. Scanning for leaf mass per area and measurement of soil moisture are also time-sensitive tasks that must occur within several days of sample collection. Careful balancing of field time (i.e., number of soil cores or foliar samples collected per day) with the required post-field processing will be necessary to complete these tasks successfully. NEON's ability to make these measurements will require iteration between field personnel and staff scientists to determine an approach that is time-efficient and does not compromise the data.

Aside from logistical review of the design by NEON field crews, one of the tools that can be used to assess the initial sampling design is a statistical framework developed to aid in optimizing the sampling strategy across the observatory. The spatial and temporal components that define where and when soils and plant tissues will be sampled have implications for how the data can be analyzed on their own and in relationship to other datasets. However, it is useful to assess the adequacy of the sampling plans using both a classical statistical power analysis and analyses of simulated data within a Bayesian hierarchical data assimilation framework. The Bayesian hierarchical approach allows for the specification of both deterministic trend components and covariance structures with both spatial and temporal terms. In this approach, data are simulated in accordance with deterministic components and spatial-temporal covariance structures of interest. Statistical data assimilation is used to analyze the simulated data and estimate posterior distributions for parameters of interest (e.g., temporal trends). In reality, the number of samples and frequency of collections are driven by an understanding of baseline data needs (i.e., characterization), published data on inter-annual variability in C and nutrient stocks (i.e., periodic sampling of soils and plant tissues), published data on intra-annual variability in soil biogeochemical processes (i.e., soil N transformations), and budgetary constraints of the project. However, this statistical framework will provide a useful means to check the quality of information and to help justify an update of the design, if resources permit. An example of a use case for the Bayesian statistical framework is given in Appendix 8.1.

As NEON operations begin, not only internal review by NEON science staff, but also regular iteration with the larger natural sciences community will be critical, in order to verify that the design and scientific priorities are aligned. In particular, the terrestrial biogeochemistry component of the observatory will benefit from regular workshops to evaluate data streams and discuss the degree to which NEON is providing useful resources to the rest of the field, including observational and modeling communities. As outlined in this document, biogeochemical transformations are inherently subject to short-term changes and are very heterogeneous spatially and temporally; in most instances, they also require labor-intensive, manual measurements. Therefore, flexibility in the NEON design strategy will be critical to assure that personnel efforts and financial resources are not wasted or misdirected. Periodic evaluation of data streams, discussion of technological and methodological advances, and general iteration between NEON staff scientists and the ecosystem biogeochemistry community should occur



neon
Operated by Battelle

<i>Title:</i> TOS Science Design for Terrestrial Biogeochemistry		<i>Date:</i> 04/06/2022
<i>NEON Doc. #:</i> NEON.DOC.000906	<i>Author:</i> E. Hinckley	<i>Revision:</i> C

every one to three years during the lifetime of the observatory. This approach will help to create the transformative data use and community research experience that is central to the NEON mission.



7 REFERENCES

- Ashton, I.W., L.A. Hyatt, K.M. Howe, J. Gurevitch, and M.T. Lerdau. 2005. Invasive species accelerate decomposition and litter nitrogen loss in a mixed deciduous forest. *Ecological Applications* 15: 1263-1272.
- Asner, G.P. and P.M. Vitousek. 2005. Remote analysis of biological invasion and biogeochemical change. *PNAS* doi: 10.1073/pnas.0500823102.
- Bai, E., B.Z. Houlton, and Y.P. Wang. 2012. Isotopic identification of nitrogen hotspots across natural terrestrial ecosystems. *Biogeosciences* 9: 3287-3304, doi: 10.5194/bg-9-3287-2012.
- Barford, C.C., S.C. Wofsy, and M.L. Goulden. 2001. Factors controlling long- and short-term sequestration of atmospheric CO₂ in a mid-latitude forest. *Science* 294: 1688–1691.
- Barnett, T.P., J.C. Adam, D.P. Lettenmaier. 2005. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* 438: 303-309.
- Bernoux, M., C.C. Cerri, C. Neill, and J.F.L. de Moraes. 1998. The use of stable carbon isotopes for estimating soil organic matter turnover rates. *Geoderma* 82(1-3): 43-58.
- Binkley, D. and S.C. Hart. 1989. The Components of Nitrogen Availability Assessments in Forest Soils. Pages 57-112 In: B.A. Stewart (Ed.) *Advances in Soil Science*, Vol. 10. Springer. 200 pp.
- Cable, J.M. K. Ogle, D.G. Williams, J.F. Weltzin, and T.E. Huxman. 2008. Soil texture drives responses of soil respiration to precipitation pulses in the Sonoran Desert: Implications for climate change. *Ecosystems*. 11(6): 961-979.
- Carpenter, S.R., N.F. Caraco, D.L. Correll, R.W. Howarth, A.N. Sharpley, and V.H. Smith. 1998. Nonpoint pollution of surface waters with phosphorus and nitrogen. *Ecological Applications* 8: 559-568.
- Correll, D.L. 1998. The role of phosphorus in the eutrophication of receiving waters: A review. *Journal of Environmental Quality* 27(2): 261-266.
- Crowl, T.A., T.O. Crist, R.R. Parmenter, G. Belovsky, and A.E. Lugo. 2008. The spread of invasive species and infectious disease as drivers of ecosystem change. *Frontiers in Ecology and Environment* 6: 238-246.
- Driscoll, C.T., K.M. Driscoll, K.M. Roy, J. Dukett. 2007. Changes in the chemistry of lakes in the Adirondack region of New York following declines in acidic deposition. *Applied Geochemistry* 22: 1181-1188.
- Driscoll, C.T., K.M. Driscoll, K.M. Roy, M.J. Mitchell. 2003. Chemical response of lakes in the Adirondack Region of New York to declines in acidic deposition. *Environmental Science and Technology* 37: 2036-2042.
- Driscoll, C.T., G. Lawrence, A. Bulger, T. Butler, C. Cronan, C. Eagar, K. Lambert, G. Likens, J. Stoddard, and K. Weathers. 2001. Acidic deposition in the northeastern US: Sources, inputs, ecosystem effects, and management strategies. *BioScience* 51:180-198.



- Emmett, B.A., C. Beier, M. Estiarte, A. Tietema, H.L. Kristensen, D. Williams, J. Peñuelas, I. Schmidt, and A. Sowerby. 2004. The response of soil processes to climate change: Results from manipulation studies of shrublands across an environmental gradient. *Ecosystems* 7(6): 625-637.
- Eno, C.F. 1960. Nitrate production in the field by incubating the soil in polyethylene bags. *Soil Science Society of America Journal* 24(4): 277-279.
- Falkowski, P., R.J. Scholes, E. Boyle, J. Canadell, D. Canfield, J. Elser, N. Gruber, K. Hibbard, P. Högberg, S. Linder, F.T. Mackenzie, B. Moore III, T. Pedersen, Y. Rosenthal, S. Seitzinger, V. Smetacek, W. Steffen. 2000. The global carbon cycle: A test of our knowledge of Earth as a system. *Science* 291-296.
- Field, C.B., R. DeFries, D. Foster, et al. 2006. Integrated science and education plan for the National Ecological Observatory Network. www.neoninc.org/documents/ISEP_2006Oct23.pdf. Viewed 21 March 2012.
- Grimm, N, S. Gergel, W. McDowell, E. Boyer, C. Dent, P. Groffman, S. Hart, J. Harvey, C. Johnston, E. Mayorga, M. McClain, and G. Pinay. 2003. Merging aquatic and terrestrial perspectives of nutrient biogeochemistry. *Oecologia* 137: 485-501.
- Groffman, P.M., L.E. Rustad, P.H. Templer, J.L. Campbell, L.M. Christenson, N.K. Lany, A.M. Socci, M.A. Vadeboncoeur, P.G. Schaberg, G.F. Wilson, C.T. Driscoll, T.J. Fahey, M.C. Fisk, C.L. Goodale, M.B. Green, S.P. Hamburg, C.E. Johnson, M.J. Mitchell, J.L. Morse, L.H. Pardo, and N.L. Rodenhouse. 2012. Long-term integrated studies show complex and surprising effects on climate change in the Northern Hardwood Forest. *BioScience* 62(12), doi: 10.1525/bio.2012.62.12.7
- Gruber, N. and J.N. Galloway. 2008. An Earth-system perspective of the global nitrogen cycle. *Nature* 451: 293-296.
- Gundersen, P., B.A. Emmet, O.J. Kjonaas, C.J. Koopmans, and A. Tietema. 1998. Impact of nitrogen deposition on nitrogen cycling in forests: A synthesis of NITREX data. *Forest Ecology and Management* 101: 37-56.
- Hart, S.C. 2006. Potential impacts of climate change on nitrogen transformations and greenhouse gas fluxes in forests: A soil transfer study. *Global Change Biology* 12(6): 1032-1046.
- Hinckley, E.S., W.R. Wieder, N. Fierer, and E.A. Paul. 2014. Digging deeper into the world beneath our feet: Bridging across scales in the age of global change. *Eos Transactions* 95(11): 96-97.
- Hoover, C.M. (Ed.) 2008. *Field Measurements for Forest Carbon Monitoring*. New York, NY: Springer. 240 pp.
- Horsley, S.B., R.P. Long, S.W. Bailey, R.A. Hallett, and P.M. Wargo. 2002. Health of eastern North American sugar maple forests and factors affecting decline. *Northern Journal of Applied Forestry* 11:34-44.
- Houlton, B.Z. and E. Bai. 2009. Imprint of denitrifying bacteria on the global terrestrial biosphere. *PNAS* 106(51): 21713-21716.



- Johnson, D.W., N.T. Edwards, and D.E. Todd. 1980. Nitrogen mineralization, immobilization, and nitrification following urea fertilization of a forest soil under field and laboratory conditions. *Soil Science Society of America Journal* 44(3): 610-616.
- Kaye, J.P. and S.C. Hart. 1997. Competition for nitrogen between plants and soil microorganisms. *TREE* 12(4): 139-143.
- Keller, M., D.S. Schimel, W.W. Hargrove, and F.M. Hoffman. 2008. A continental strategy for the National Ecological Observatory Network. *Frontiers in Ecology and Environment* 6(5): 282-285.
- Likens, G.E., C.T. Driscoll, D.C. Buso, M.J. Mitchell, G.M. Lovett, S.W. Bailey, T.G. Siccama, W.A. Reiners, and C. Alewell. The biogeochemistry of sulfur at Hubbard Brook. *Biogeochemistry* 60(3): 235-316.
- Loik, M.E., D.D. Breshears, W.K. Lauenroth, and J. Belnap. 2004. A multi-scale perspective of water pulses in dryland ecosystems: climatology and ecohydrology of the western U.S.A. *Oecologia* 141(2): 269-281.
- Lovett, G.M., D.A. Burns, C.T. Driscoll, J.C. Jenkins, M.J. Mitchell, L. Rustad, J.B. Shanley, G.E. Likens, and R. Haeuber. 2007. Who needs environmental monitoring? *Frontiers in Ecology and Environment* 5(5): 253-260.
- Lovett, G.M., K.C. Weathers, and M.A. Arthur. 2002. Control of nitrogen loss from forested watersheds by soil carbon:nitrogen ratio and tree species composition. *Ecosystems* 5: 712-718.
- Lovett, G.M., C.D. Canham, M.A. Arthur, K.C. Weathers, and R.D. Fitzhugh. 2006. Forest ecosystem responses to exotic pests and pathogens in eastern North America. *BioScience* 56(5): 395-405.
- MacKenzie, V.J. and A.R. Townsend. 2007. Parasitic and infectious disease responses to changing global nutrient cycles. *EcoHealth* 4(4): 384-396.
- Magill, A.H., J.D. Aber, W.S. Currie, K.J. Nadelhoffer, M.E. Martin, W.H. McDowell, J.M. Melillo, and P. Steudler. 2004. Ecosystem response to 15 years of chronic nitrogen additions at the Harvard Forest LTER, Massachusetts, USA. *Forest Ecology and Management* 196(1): 7-28.
- Magill, A.H., J.D. Aber, J.J. Hendricks, R.D. Bowden, J.M. Melillo, and P.A. Steudler. 1997. Biogeochemical response of forest ecosystems to simulated chronic nitrogen deposition. *Ecological Applications* 7: 402-415, doi: 10.1890/10-51-0761(1997)007[0402:BROFET]2.0.CO;2.
- Marín-Spiotta, E., W.L. Silver, C.W. Swanston, and R. Ostertag. 2009. Soil organic matter dynamics during 80 years of reforestation of tropical pastures. *Global Change Biology* 15(6): 1584-1597.
- McClougherty, C.A., J. Pastor, J.D. Aber, and J.M. Melillo. 1985. Forest litter decomposition in relation to soil nitrogen dynamics and litter quality. *Ecology* 66(1): 266-275.
- Melillo, J.M., J.D. Aber, A.E. Linkins, A. Ricca, B. Fry, and K.J. Nadelhoffer. 1989. Carbon and nitrogen dynamics along the decay continuum: Plant litter to soil organic matter. *Plant and Soil* 115: 189-198.
- Monteith, D.T., J.L. Stoddard, C.D. Evans, H.A. deWit, M. Forsius, T. Høgåsen, A. Wilander, B.L. Skjelkvåle, D.S. Jeffries, J. Vuorenmaa, B. Keller, J. Kopáček, J. Vesely. 2007. Dissolved organic carbon trends



resulting from changes in atmospheric deposition chemistry. *Nature* 450: 537-540, doi: 10.1038/nature06316.

Mou, X., M. Vila-Costa, S. Sun, W. Zhao, S. Sharma & M.A. Moran. 2011. Metatranscriptomic signature of exogenous polyamine utilization by coastal bacterioplankton. *Environmental Microbiology Reports* 3:798-806.

Nadelhoffer, K.J., B.P. Colman, W.S. Currie, A.H. Magill, and J.D. Aber. 2004. Decadal-scale fates of ¹⁵N tracers added to oak and pine stands under ambient and elevated N inputs at the Harvard Forest (USA). *Forest Ecology and Management* 196: 89-107.

National Research Council. 2001. *Grand Challenges in Environmental Sciences*. 107 pp. The National Academies Press, Washington, D.C.

National Research Council. 2003. *NEON - Addressing the nation's environmental challenges*. The National Academies Press, Washington, D.C.

Ollinger, S.V., M.L. Smith, M.E. Martin, R.A. Hallett, C.L. Goodale, and J.D. Aber. 2002. Regional variation in foliar chemistry and N cycling among forests of diverse history and composition. *Ecology* 83: 339-355.

Pardo, L.H., M. Robin-Abbott, N. Duarte, and E.K. Miller. 2005. *Tree chemistry database (version 1.0)*. Gen. Tech. Rep. NE-324. Newtown Square PA: U.S. Department of Agriculture, Forest Service, Northeastern Research Station. 45 p.

Phillips, D.L. and J.W. Gregg. 2003. Source partitioning using stable isotopes: coping with too many sources. *Oecologia* 136(2): 261-269.

Richardson, A.D., D.Y. Hollinger, J.D. Aber, S.V. Ollinger, and B.H. Braswell. 2007. Environmental variation is directly responsible for short-but not long-term variation in forest-atmosphere carbon exchange. *Global Change Biology* 13:788-803.

Robinson, D. 2001. $\delta^{15}\text{N}$ as an integrator of the nitrogen cycle. *TREE* 16(3): 153-162.

Ryan, M.G., J.M. Melillo, and A. Ricca. 1990. A comparison of methods for determining proximate carbon fractions in litter. *Canadian Journal of Forest Research* 20: 166-171.

Ryther, J.H. and W.M. Dunstan. 1971. Nitrogen, phosphorus, and eutrophication in the coastal marine environment. *Science* 171(3975): 1008-1013.

Sala, O.E. and A.T. Austin. 2000. Methods of estimating aboveground net primary productivity, In: O.E. Sala, R.B. Jackson, H.A. Mooney and R.W. Howarth (eds.) *Methods in Ecosystem Science*, p 31-43, Springer-Verlag.

Savage, K.E. and E.A. Davidson. 2001. Interannual variation of soil respiration in two New England forests. *Global Biogeochemical Cycles* 15:337-350.

Schimel, D., W. Hargrove, F. Hoffman, and J. MacMahon. 2007. NEON: a hierarchically designed national ecological network. *Frontiers in Ecology and the Environment* 5:59-59.



- Schimel, D.S., M. Keller, and P. Duffy. 2009. NEON Science Strategy. The National Ecological Observatory Network, Boulder, CO. 50 pp.
- Shortle, W.C., K.T. Smith, R. Minocha, G.B. Lawrence, and M.B. David. 1997. Acidic deposition, cation mobilization, and biochemical indicators of stress in healthy red spruce. *Journal of Environmental Quality* 26: 871-876.
- Stewart, F.J., O. Ulloa, and E.F. DeLong. 2012. Microbial metatranscriptomics in a permanent marine oxygen minimum zone. *Environmental Microbiology* 14: 23-40.
- Stoddard, J.L., D.S. Jeffries, A. Lükewille, T.A. Clair, P.J. Dillon, C.T. Driscoll, M. Forsius, M. Johannessen, J.S. Kahl, J.H. Kellogg, A. Kemp, J. Mannio, D.T. Monteith, P.S. Murdoch, S. Patrick, A. Rebsdorf, B.L. Skjelkvale, M.P. Stainton, T. Traaen, H. VanDam, K.E. Webster, J. Wieting, and A. Wilander. 1999. Regional trends in aquatic recovery from acidification in North America and Europe. *Nature* 401: 575-578.
- Stoddard, J.L., J.S. Kahl, F.A. Deviney, D.R. DeWalle, C.T. Driscoll, A.T. Herlihy, J.H. Kellogg, P.S. Murdoch, J.R. Webb, and K.E. Webster. 2003. Response of surface water chemistry to the Clean Air Act Amendments of 1990. EPA/620/R-03/001. Research Triangle Park, NC.
- Theobald, D.M., D.L. Stevens, D. White, N.S. Urquhart, A.R. Olsen, and J.B. Norman. 2007. Using GIS to generate spatially balanced random survey designs for natural resource applications. *Environmental Management* 40:134-146.
- Tietema, A. and C. Beier. 1995. A correlative evaluation of nitrogen cycling in the forest ecosystems of the EC projects NITREX and EXMAN. *Forest Ecology and Management* 71: 143-152.
- Topp, C. and P.A. Ferré. 2002. The Soil Solution Phase. Pages 417-1071. In J.H. Dane and G.C. Topp (Eds). *Methods of Soil Analysis, Part 4—Physical Methods*. SSSA Book Series: 5. Soil Science Society of America. 1692 pp.
- Turner, M.G., S.L. Collins, A.L. Lugo, J.J. Magnuson, T.S. Rupp, and F.J. Swanson. 2003. Disturbance dynamics and ecological response: The contribution of long-term ecological research. *BioScience* 53(1): 46-56, doi: 10.1641/0006-3568(2003)053[0046:DDAERT]2.0.CO;2.
- Urich, T., A. Lanzen, J. Qi, D.H. Huson, C. Schleper, and S.C. Schuster. 2008. Simultaneous assessment of soil microbial community structure and function through analysis of the metatranscriptome. *PLoS ONE* 3:e2527
- Vadeboncoeur, M.A., S.P. Hamburg, J.D. Blum, M.J. Pennino, R.D. Yanai, and C.E. Johnson. 2012. The quantitative soil pit method for measuring belowground carbon and nitrogen stocks. *Soil Sci. Soc. Am. J.* 76(6): 2241-2255.
- Vitousek, P.M., H.A. Mooney, J. Lubchenco, and J.M. Melillo. 1997. Human domination of Earth's ecosystems. *Science* 277(5325): 494-499.
- Vitousek, P.M. and L.R. Walker. 1989. Biological invasion by *Myrica faya* in Hawai'i: Plant demography, nitrogen fixation, ecosystem effects. *Ecological Monographs* 59: 247-265.



- Wang, L., G.S. Okin, and S.A. Macko. 2010. Remote sensing of nitrogen and carbon isotope compositions in terrestrial ecosystems, In: J.B. West, G.J. Bowen, T.E. Dawson and K.P. Tu (eds.) *Isoscapes: Understanding Movement, Pattern, and Process on Earth Through Isotope Mapping*, p. 51-70, Springer, doi:10.1007/978-90-481-3354-3_3.
- Wang, L., G.S. Okin, J. Wang, H. Epstein, and S.A. Macko. 2007. Predicting leaf and canopy ¹⁵N compositions from reflectance spectra. *Geophysical Research Letters* 34, L02401, doi:10.1029/2006GL02850
- Weathers, K.C. and G.M. Lovett. 1998. Acid deposition research and ecosystem ecology: Synergistic successes, In: M.L. Pace and P.M. Groffman (eds.). *Successes, Limitations and Frontiers in Ecosystem Science*, p. 195-219.
- West, J.B., G.J. Bowen, T.E. Cerling, and J.R. Ehleringer. 2006. Stable isotopes as one of nature's ecological recorders. *TREE* 21(7): 408-414.
- Yanai, R.D., J.J. Battles, A.D. Richardson, E.B. Rastetter, D.M. Wood, and C. Blodgett. 2010. Estimating uncertainty in ecosystem budget calculations. *Ecosystems* 13(2): 239-248.
- Yanai, R.D., S.V. Stehman, M.A. Arthur, C.E. Prescott, A.J. Friedland, T.G. Siccama, and D. Binkley. 2003. Detecting change in forest floor carbon. *Soil Sci. Soc. Am. J.* 67:1583-1593.

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

APPENDIX A A STATISTICAL FRAMEWORK FOR EVALUATING NEON SAMPLING DESIGN

A.1 Introduction

The purpose of this appendix is to provide details regarding analyses performed in support of the specification of the terrestrial biogeochemistry sample design. NEON's high-level requirements provide guidance regarding the overarching science goals; however, the determination of measurements deemed relevant in the context of the design specification requires additional consideration. Flexibility within the design is required in order to accommodate unique aspects of ecological responses measured by each component of NEON. Computational analyses performed in support of the development of design constraints are intended to assess the sufficiency of the datasets that will be collected to accomplish the goals identified for specific components of NEON. The framework of computational analyses must also maintain the generality necessary to ensure relevance across numerous data uses considered.

At the most general level, the minimally sufficient sampling effort corresponds to the sample design specified by the plant biodiversity analysis to determine locations of distributed plots at NEON sites. These minimum sample sizes are characterized using a frequentist approach to a power analysis. This power analysis is based on the test of a difference between the slopes characterizing linear change through time at two locations. In this setting, repeat measurements through time are taken on the same sampling units within each group. With a minimum sample size of approximately 10 (depending on the specification of tolerable error levels, etc.), distributed sample locations were spatially located using the Reverse Randomized Quadrat Recursive Raster (RRQRR) method. Given this initial allocation, modifications to the terrestrial biogeochemistry sampling design can be made in accordance with the high-level requirements through the application of this simulation tool, which uses a Bayesian hierarchical modeling (BHM) framework for the analysis of simulated data. This analysis framework can also be applied to data products generated by multiple components of NEON.

The BHM framework provides the capability to depict complexity with respect to ecosystem processes and interactions (Cressie et al., 2009). This also presents a challenge in that the degree of generality, and hence relevance across a range of ecological responses, is inversely proportional to the level of complexity for a given model structure. Two general paths for the depiction of complex space-time dynamics using BHM are: 1) place a significant degree of effort into the characterization of these dynamics into the deterministic component of the process model, and 2) keep the deterministic component of the process model relatively simple and account for the space-time dynamics using more complex covariance functions. Wikle and Hooten (2010) provide a nice discussion of some issues related to the development of a modeling approach in this context.

The latter approach, sometimes called the second-moment approach (due to its relative focus on covariance structures), is the one selected as a basis for the quantification of sample design adequacy in the computational analyses presented here. This approach was chosen for two reasons. First, the minimal assumptions associated with the process model component allow for a greater degree of

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

generality; this feature is essential, due to the large number of response variables that will ultimately be considered. The intent is that this will allow for broader application of these analysis tools to different responses. More specific process models (i.e., the first approach) require that more complex models be built that describe specific dynamics of each individual response of interest. This also increases the degree of subjectivity associated with the model building and unnecessarily obfuscates the intent of the work presented here. To be clear, the development of complex process models is of fundamental importance when studying a specific system in detail; however, this approach does not align with the intent of this design characterization work as well as the second-moment approach. The second reason to implement the development of more complex space-time covariance functions that evolve from the second approach is that NEON will generate data products with a high degree of spatial and temporal coverage. The BHM framework allows for the iterative optimization of the data collection (for both NEON sampling and PI driven projects) through the application of statistical data assimilation to update the parameter distribution models. Together, these two objectives drive the selection and application of the latter approach.

Historical data play an important role in the work presented here. In general, subsequent application of these tools should be performed in concert with analyses of existing NEON data, both to inform the generation of synthetic data and provide context for analyses of data products. In this work, historical data are used to inform parameter values that specify various statistical properties of the simulated data. Thus, the simulated data provide a reasonably realistic depiction of the types of data products that will be generated by NEON. The data simulated in this work are analyzed in the BHM framework to characterize the impact of various design configurations on posterior distributions of parameters of interest. This approach allows for the quantitative assessment of various measures of sample adequacy under scenarios of interest, as determined by NEON's high-level science requirements.

A.2 Methods

Response Variable

The response variable of interest for this simulation study is peak annual CO₂ flux. The question of interest for this response is: Can a difference in the rate of change in peak annual CO₂ flux through time be detected between vegetation types? The proposed sampling that is of interest for characterization consists of a single annual sampling event where data are collected at each of the 40 locations specified by the RRQRR based sample locations. Data from the Harvard forest LTER (i.e., NEON Domain 1 core site) are used in this work (**Figure 10**) where the two largest vegetation types are deciduous and evergreen forest classes, which have 12 and 13 RRQRR sample locations, respectively. The simulation study focuses on quantifying the impact of varying sample frequencies (one, three and five years) of soil CO₂ fluxes on subsequent analyses performed on synthetic data generated from the RRQRR sample locations within each of these two vegetation classes.

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

Historical Data

In order to simulate data that are as realistic as possible, historical soil CO₂ flux data collected from the Harvard Forest site were assembled (**Figure 11**). Metadata are available online (<http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf194>). Exploratory analyses of these historical data were performed and a subset of the data was assembled in order to provide the highest degree of consistency with the types of data that will be collected by NEON.

The first round of data reduction dealt with the exclusion of treatment plots; only data from control (i.e., not experimental treatment) plots were used. In order to minimize impacts that may be a consequence of chosen sample methods not articulated in the metadata, the selected dataset only includes measurements made by a single investigator using consistent methodology. The data associated with PI Davidson provide a ten-year record with good representation for the summer months when peak flux occurs, and uses manual collars to obtain measurements at the locations in **Figure 11**. Multiple measurements were taken within each month during the ten-year sampling period. **Figure 12** shows the monthly distribution of CO₂ flux measurements for this dataset.

The historical data from Harvard Forest show that the largest monthly flux distributions are present in the months of July and August. The rest of this analysis is based on data available for the month of August. Within each month, multiple samples were taken across space and through time. With respect to time, samples were collected at approximately weekly intervals. Hence, using only the samples from the last sampling event for which data were collected in August further reduces the dataset by removing intra-monthly variability. Across space, samples were collected from multiple collars at multiple locations; this analysis excludes those with anomalously large or small values. The final historical data set had approximately 12 samples for each year, creating a subset of data with the corresponding variability that would be expected from deciduous and evergreen forest classes with 12 and 13 samples, respectively. Boxplots of the peak CO₂ flux distribution by year from the final working dataset of historical data are presented in **Figure 13**. **Figure 14** shows boxplots for these same data by sampling location within each year, providing a sense of the spatial variability among sites. **Figure 14** also depicts the relative differences in inter-annual variability among sampling locations through time.

Figure 15 shows the variability among collars within sampling locations. The collars within a sampling location do not have distinct spatial locations. This variability, thus, cannot be resolved with spatial covariance modeling; in this sense, the between collar variability within each site is analogous to the nugget in a variogram model. Exploratory analysis of the spatial correlation of the full historical dataset suggests that there is, at best, weak spatial correlation at the distances that will be encountered in the design locations. It is also worth noting that these exploratory spatial analyses are somewhat tenuous due to the relatively small sample sizes. Ultimately, the potential impact of a reduction in effective sample size as a function of the relatively shorter distances between the collars does not seem to be a critical issue. If the spatial correlation of these data were stronger at the scale of the collars, then this discrepancy in distances would need to be accommodated. Summary statistics for the final subset of

historical data are presented in **Table 7**. These data are used to specify values for parameters in the statistical model, which generates synthetic data for the simulation study.

A.3 Simulation Design

The primary focus of the work presented here is an assessment of the impact of sampling at one, three, and five year intervals. Hence, multiple realizations of annual peak flux were generated and analyzed at the one, three, and five year sampling frequency. The number of realizations was determined through iterative assessment of the stability of the results, in conjunction with the computational burden. For the annual sampling results, there were ten realizations, for the three year there were 100 realizations, and for the five year sampling there were 300 realizations. Each realization has 5000 samples from the posterior distributions generated via Markov Chain Monte Carlo (MCMC) methods.

Each realization in the simulation study is a time series of multivariate observations. The multivariate aspect of the data comes from the spatially explicit observations. Synthetic data are generated using a separate process model and a data (measurement) model. The process model characterizes the underlying behavior of the system, incorporating spatial and temporal correlation as well as structural changes in the mean, due to other known or measured quantities. The measurement model then characterizes the variation and bias (if any) that relates the latent underlying process to the actual measurements. Separating the process and data model provides several advantages including better characterization of uncertainty associated with parameter estimates (Calder et al. 2003). Parameters are specified as fixed values for the generation of synthetic data. However, distributional models for parameters are specified as part of the BHM used to analyze the synthetic data (**Table 8**). Each of the model components is discussed in detail below.

A.4 Data Model

For this application, the measurement model is assumed to provide an unbiased measurement of the unobservable peak CO₂ flux with a fixed but unknown measurement variance, σ_m^2 . Bias parameters can be readily accommodated in the data model if it is desirable to assess whether there is bias in the measurements (i.e., include the bias parameter and see if the data suggest that it is effectively no different from zero). Assuming measurement error to be normally distributed conditional on the unobserved latent process, Y , the data model can be written as:

$$\mathbf{Z} | \mathbf{Y}, \sigma_m^2 \sim MVN(\boldsymbol{\mu}(\mathbf{t}), \sigma_m^2 \mathbf{I}) \quad [1]$$

Where, \mathbf{Z} is the vector of observations made on the underlying latent process Y , and $\boldsymbol{\mu}(\mathbf{t})$ is the mean of the latent process, and \mathbf{t} is the vector of times associated with the measurements, \mathbf{I} is the identity matrix, and σ_m is a scalar that represents the overall magnitude of the measurement error in the same units as \mathbf{Z} .

A.5 Process Model



The process model for the simulation of peak CO₂ flux through time (i.e., $t = 1, 2, \dots, 30$) is specified as $\mu(t)$, conditional on the parameters and is defined as,

$$\mu(\mathbf{t}) | \sigma_p^2, SR, TR \sim MVN(0, \mathbf{C}_p) \quad [2]$$

where \mathbf{C}_p is the space-time covariance matrix defined below, σ_p^2 is the variance of the process, SR is the spatial range that specifies the decay rate of correlation for spatial distance (in meters) between measurement (d_{ij}) for a given point in time, and TR is the time range that specifies the decay rate of correlation for times (in years) between measurements ($|t_i - t_j|$) at a given location. A linear deterministic functional form was used to characterize the mean of the process model for the simulation of peak CO₂ flux through time (i.e. $t = 1, 2, \dots, 30$).

$$\mu(\mathbf{t}) = \beta_0 + \beta_1 t \quad [3]$$

The mean component eqn [3] could also include covariate information, such as precipitation, but there is little difference in formulation if an additive linear increase in signal over time is the change of interest. Better process models using covariate information would, however, lead to a reduction in the magnitude of the variance associated with the process model. This deterministic component of the process model is considered to change uniformly across space for the work presented here. In practice, data from NEON will likely suggest a spatially varying mean for the process model. In this simulation study, the specification of a spatially varying mean surface would only serve to obfuscate the results.

A.6 Process Covariance

While the deterministic component of the process model changes uniformly as a function of space, the simulation design allows for the specification of a stationary, but separable spatio-temporal process covariance structure. In this sense, ‘separable’ implies that the covariance can be expressed as a product of two terms, each corresponding exclusively to a spatial and temporal component. This approach assumes that there will be spatial and temporal covariance in the response that is unaccounted for by the deterministic terms of the process model. The covariance matrix corresponding to this structure is shown in Eq [4].

$$C_{ij} = \sigma_p^2 \cdot \frac{\exp\left\{-\left(\frac{d_{ij}}{SR}\right) / \left(\frac{|t_i - t_j|}{TR} + 1\right)\right\}}{\left(\frac{|t_i - t_j|}{TR} + 1\right)} \quad [4]$$

It follows that the process equation for the underlying true ecological process of interest can be represented as,

$$Y_t = (\mu)\mathbf{t} + \mathbf{C}_p \quad [5]$$

A.7 Parameter Models

There are parameters associated with both the data and process models. For the sake of clarity parameters are grouped accordingly, such that $\theta_t^Z = \{\sigma_m\}$, $\theta_t^Y = \{\sigma_p, b_f, SR, TR\}$, $\theta_t = \{\theta_t^Z, \theta_t^Y\}$. For the generation of synthetic data, scalar values must be specified for the parameters. In order to analyze the synthetic data in the BHM framework, prior distributions are specified for each of the parameters. Prior distributions are assigned for each of the elements of θ_t (see **Table 8**). The parameters are assumed to be independent such that the joint distribution of the parameters is the product of the individual marginal distributions of each of the parameters in eqn [6].

$$P[\theta_t] = \prod_{i=1}^n P[\theta_i] \quad [6]$$

Where n is the total number of parameters from both the data and process models.

For each of the sampling frequencies considered, distributions for parameters are updated by assimilating the current year's observations. This process begins at year 3 for all of the sampling frequencies considered; prior distributions for each of these parameters need to be specified at year 2. The year 2 priors for θ_t are informed based on values that are reasonably inferred from exploratory data analysis of the first two years data (**Table 8**).

A.8 Analysis of Synthetic Data from the Harvard Forest LTER (NEON Domain 1 Core Site)

The simulation framework is designed to provide a suite of tools that can be used to address the detection and quantification of trends in climatically driven ecological responses by bracketing the plausible ranges for flexible components of the design. In this application, each simulated times series contains realizations that correspond to annual sampling at distributed locations (**Figure 19**). At each time step of interest, simulated data are analyzed using a MCMC approach to update the prior distributions and estimate the posterior distributions of parameters. In this application, the distribution of the difference of the trend parameters for the different vegetation types is of primary interest. The intent is to mimic the analysis of data collected sequentially through time with the goal being to determine which functional form for the process model best characterizes the linkage between the ecological response of interest and the climate driver.

A.9 Computational Issues

Given the models for the data, process, and parameters, the focus of this exercise is on the posterior distribution of the process and parameters conditional on the observed data, which can be generally expressed as

$$p(Y, \theta_t | Z) \propto p(Z | Y, \theta_t) p(Y | \theta_t) p(\theta_t) \quad [7]$$

where θ_t represents the collection of all parameters of interest at time t. This application evaluates the posterior distribution of the process and the parameters by conditioning on a functional form (eqn[2]) for the deterministic component of the process. The multivariate normal distribution form in eqn[1] and



eqn[2] allows some simplification of the expression for the posterior distribution of the process and the parameters when each process is considered separately. The posterior of interest is expressed as

$$p(\theta_t|Z, Y) \propto p(Z|Y, \theta_t)p(\theta_t) \quad [8]$$

Samples from the posterior distributions for parameters in θ_t (eqn [6]) are obtained using a MCMC approach with a Metropolis-Hastings (MH) sampling algorithm from the MCMCpack library in R (Martin et al. 2008). The first update occurs at year three; therefore, priors for the parameter distributions are assigned at year two. Results are presented for years three and 18 for all three sampling frequencies. The final year of analysis for the sampling frequencies of one, three, and five years was 30, 30 and 28, respectively.

If samples from the posterior distribution for the process, conditional on the data and parameters, are desired, then they can be constructed post-hoc, by conditioning on the MCMC posterior samples for θ_t . The multivariate normal distributions for the process and measurement lead to a closed form for the process conditional on the parameters and data:

$$p(Y|\theta_t, Z) \sim MVN\left(V^* \left[V^{-1} \mu(t|\theta_t) + \frac{1}{\sigma^2} IZ\right], V^*\right) \quad [9]$$

where

$$V^* = \left[V^{-1} + \frac{1}{\sigma^2} I\right]^{-1} \quad [10]$$

The posterior predictive distribution for the process at a future time point, $t + i$, can be constructed in similar fashion by extending the mean $\mu(t|\theta_t)$ and covariance matrix V for the joint distribution through time $t + i$, and augmenting the I matrix with zeroes for times yet to be measured.

Due to the nature of the covariance structure of the posterior distributions and the potential high-dimensionality of the parameters (when including the process values Y), it is not a simple matter to summarize the posterior distributions succinctly. Marginal distributions might be fit to some parameters, but a flexible multivariate model for capturing the potentially non-linear relationships between the parameters is generally not available. Thus, when new data become available, there is typically not a simple method of updating the posterior with the new data conditional on the previous results.

Consequently, for the analyses presented here, when synthetic data are assimilated at subsequent time steps, the model is expanded to accommodate the new data, and the MCMC is re-run from scratch (i.e., updating the year two priors). The dimensionality of the model can lead to slow computation as the data size increases. However, since the change in the parameters is likely to be relatively small as the sample size increases, this allows for previous MCMC runs to be potentially be used to construct starting values and Metropolis proposal distributions for increased efficiency of subsequent runs.

A.10 Results

Trend Parameter

The study question motivating this exercise is, Can a difference in the rate of change in peak annual CO₂ flux through time be detected between vegetation types? The relative ability to detect this difference was assessed for sampling frequencies of one, three, and five years. For the simulation and analysis approach presented above, there were noticeable differences among the analyses of the different sampling frequencies with respect to the ability to detect a difference (in the rate of change in peak annual CO₂ flux between vegetation types). Specifically, the certainty associated with the detection of the relative differences increased with sample frequency. These differences are represented in the posterior distributions of the trend parameters for the deciduous and evergreen sampling locations (**Figures 16-18** and **Table 9**).

The measures of central tendency of the posterior distributions for the trend parameters are close to the specified values (**Table 8**) by year 18 (**Table 9**) across all sampling frequencies. There is little change in the mean of the posterior distributions between year 18 and the last year of updating (i.e., year 30 for the one and three year sampling frequencies and year 28 for the five year sampling frequency).

Between the updates at years three and 18, there is roughly an order of magnitude decrease in the standard deviation of the posterior distributions of the trend parameters for all sampling frequencies. The decrease is larger for the annual sampling frequency than it is for the three and five year sampling frequency scenarios. The standard deviation at year 30 is roughly 50% greater for the three year sampling frequency than that of the annual sampling. Although not an exact comparison of years, the standard deviation of the five year sampling frequency at year 28 is roughly 100% greater than that of the annual sampling frequency at year 30.

The posterior distribution of the difference between trend parameters for the deciduous and evergreen vegetation classes shows similar results (**Figure 19** and **Table 10**), leading to a probability statement. Specifically, the probability that the difference in the trends is less than zero can be used as a measure of the ability of specific sampling scenario to detect a trend (**Table 10**). The year 18 results show a decrease in the probability that the difference is less than zero with increased sampling frequency. The probability corresponding to the three year sampling scenario is roughly four times greater than that of the annual sampling. Similarly, the probability corresponding to the five year sampling interval is over six times greater. By year 30, with the annual sampling frequency, there were no MCMC samples where the difference in trend corresponded to a value less than or equal to zero. For the three year sampling scenario, the probability of the trend being less than or equal to zero was approximately 8%. Interestingly, for the five year sampling frequency, the probability at year 28 is only 2%.

Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

Other Parameters

For the annual sampling, the posterior distribution of the intercept parameter displays reduced variance with increasing time (**Figure 11**). For the three and five year sampling intervals, there is a negligible difference in the variance of the posterior distributions between year 18 and the final year of updating (**Figures 12 and 13**). The error parameters in the annual sampling case show slight shifts toward the values used to specify the simulated data; however, for the three and five year sampling cases, there is little noticeable movement away from the prior distributions.

A.11 Discussion

By simulating data based on characteristics gleaned from analyses of regional historical data in conjunction with the implementing spatial locations for the Harvard Forest site, a flexible framework for assessing various aspects of the sufficiency of the design is presented and tested. Specifically, these results demonstrate the impact of several sampling frequency scenarios on the ability to detect a difference in trend magnitudes of the peak CO₂ flux between vegetation types. Differences between the sampling frequency scenarios are quantified and provide a means to assess the sufficiency of various sampling frequency scenarios. The difference in slopes considered here reflects the uncertainty associated with differences in the rates of change in the peak flux as a function of the different vegetation types through time. Subsequent analyses can easily be performed to assess the sensitivity of the design to varying trend magnitudes.

The parameters associated with the deterministic component of the process model (i.e., intercept and slope) are well informed (as measured by a reduction in the variance of the posterior distributions) by the annual sampling scenario. Specifically, the trend and intercept parameters for the annual sampling case demonstrate a consistent reduction in the variance of the posterior distributions with increased data collection through time. In contrast, the posterior distributions for the three and five years sample frequency scenarios demonstrate little change in the variance associated with the posterior distributions between year 18 and the final year of sampling. This result suggests there is negligible increase in the information content associated with these parameters that corresponds to the data collected between year 18 and final year of sampling.

The design does not appear to provide much useful information with respect to the spatial and temporal covariance structures considered here. In this context, it is worth noting that there was little evidence in the historical data to suggest anything beyond a weak temporal and spatial correlation structure should be imposed. It is possible that there was too much signal (deterministic) to noise (error) in these data to pick up the weak spatial and temporal covariance structure imposed. In addition, this situation makes the specification of the priors more influential on the final analyses since the data do not provide much information to make the posteriors meaningfully different from the priors. This finding may indicate that these terms should be eliminated from the model. In practice, for the sake of model building, that is a reasonable approach to take. For the purpose of this design characterization work, these analyses



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

present a useful baseline for comparison to subsequent scenario of design implementation to be considered.

There are several additional issues that should be considered when interpreting the results of this analysis. The first is that as NEON data are collected through time, there will be advances in the understanding of the system that correspond to more sophisticated process models. For example, there is variability associated with the date of occurrence of the peak flux. As models that account for additional complexity such as this are developed, less of the observed variability in the response will be accounted for by the covariance terms in both the process and data models. Ideally, the increased sophistication of the process models developed and implemented through time result in the non-independent component of the process covariance to be negligible. This is unlikely, so the utility of the more complex space-time covariance structure considered in this framework should persist.

Second, the consistency of the response between vegetation types for a given year (i.e. 'species uncertainty') is an interesting and important factor that may be worth considering in subsequent analyses. Essentially, this factor would address the question, For the climate observed in a given year, do we expect to see similar responses in peak annual CO₂ flux between the vegetation types of interest? For example, in a hot and dry year, would one observe similar responses in peak flux for the different vegetation types of interest? In general, the capability to have the response be either similar or different between vegetation types will likely be useful for at least some situations. If the two vegetation types of interest are both primarily deciduous (e.g. 'deciduous open canopy' and 'deciduous closed canopy') then one might expect the responses to be quite similar. If however, the vegetation types under consideration are physiologically different, (e.g. 'conifer closed canopy' and 'deciduous closed canopy') then it may be of interest to have the responses be somewhat different within a given year. Functionally, for the case where there is between-vegetation type variability, separate time series realizations are generated for each vegetation type. This means there is a different sample taken from the process covariance for each vegetation type. If between-vegetation type variability is not desired in the model, then the realization from the process covariance is selected treating samples from both vegetation types as realizations from the same process. Then the data from each vegetation type are analyzed separately.

A.12 References

- Andrew D. Martin, Kevin M. Quinn, Jong Hee Park (2011). MCMCpack: Markov Chain Monte Carlo in R. *Journal of Statistical Software*. 42(9): 1-21. URL <http://www.jstatsoft.org/v42/i09/>.
- Calder, C., M. Lavine, P. Muller, and J. S. Clark. 2003. Incorporating multiple sources of stochasticity into dynamic population models. *Ecology* 84:1395-1402.
- Cressie, N., C. A. Calder, J. S. Clark, J. M. Ver Hoef, and C. K. Wikle, 2009. Accounting for uncertainty in ecological analysis: the strengths and limitations of hierarchical statistical modeling, *Ecological Applications* 19, 553-570.



Title: TOS Science Design for Terrestrial Biogeochemistry		Date: 04/06/2022
NEON Doc. #: NEON.DOC.000906	Author: E. Hinckley	Revision: C

Wikle, C.K. and M.B. Hooten, 2010: A general science-based framework for spatio-temporal dynamical models. Invited discussion paper for Test. 19, 417-451

A.13 Tables and Figures

Table 7. Summary statistics for peak CO2 flux estimates obtained from the reduced dataset with four sites and three collars within each site. A single sampling event was used for each year, therefore, there are 12 samples associated with each year.

Year	Min	1st Qu.	Median	Mean	3rd Qu.	Max	SD
1997	2.53	3.65	4.38	4.52	4.90	6.89	1.35
1998	4.49	5.95	6.51	7.44	7.87	15.52	2.92
1999	2.39	2.91	3.34	3.68	3.89	6.53	1.15
2000	2.74	4.55	5.54	5.61	6.16	8.85	1.82
2003	2.89	3.95	4.25	4.89	5.26	10.28	2.03
all years	2.39	3.72	4.66	5.23	6.23	15.52	2.28

Table 8. Parameter values used to generate synthetic data. Normal distributions are specified with a mean and standard deviation. Two values are specified for the β_{trend} parameter: deciduous = 0.05, and evergreen = 0.10. IGamma is the Inverse Gamma distribution as specified in the MCMCpack library (Martin et al. 2008) in R.

	β_0	β_{trend}	σ_m^2	σ_p^2	SR	TR
Value used in Simulation	0.47	Deciduous = 0.05, Evergreen=0.10	0.47	0.5	500	0.05
Prior Distribution	Normal (5.3, 1)	Normal (0, 0.5)	Uniform (0.5, 5.0)	IGamma (5, 3)	IGamma (5, 1000)	IGamma (5, 2)

Table 9. Percentiles for the trend parameter posterior distributions for the 1, 3, and 5 year sampling frequencies of deciduous and evergreen.

Frequency	Vegetation	Year	Min.	1st	Median	Mean	3rd	Max.	SD
1yr	Deciduous	3	-1.000	-0.240	-0.046	-0.040	0.158	0.989	0.287
	Evergreen	3	-1.093	-0.255	-0.067	-0.069	0.119	0.791	0.276
	Deciduous	18	0.021	0.080	0.094	0.094	0.108	0.165	0.021
	Evergreen	18	-0.026	0.032	0.045	0.045	0.058	0.111	0.019
	Deciduous	30	0.067	0.093	0.099	0.099	0.106	0.136	0.010
	Evergreen	30	0.014	0.042	0.048	0.048	0.055	0.086	0.010
3yr	Deciduous	3	-1.164	-0.258	-0.065	-0.062	0.133	1.094	0.290
	Evergreen	3	-1.101	-0.294	-0.103	-0.100	0.093	1.152	0.286
	Deciduous	18	-0.035	0.074	0.096	0.096	0.118	0.227	0.033
	Evergreen	18	-0.093	0.024	0.043	0.042	0.062	0.185	0.029
	Deciduous	30	0.038	0.088	0.099	0.099	0.109	0.164	0.033
	Evergreen	30	-0.005	0.038	0.048	0.048	0.057	0.107	0.029



5yr	Deciduous	3	-1.181	-0.264	-0.068	-0.069	0.126	1.229	0.290
	Evergreen	3	-1.338	-0.287	-0.093	-0.092	0.101	1.100	0.288
	Deciduous	18	-0.068	0.073	0.097	0.097	0.121	0.248	0.036
	Evergreen	18	-0.103	0.022	0.045	0.045	0.068	0.189	0.034
	Deciduous	28	0.021	0.087	0.099	0.100	0.113	0.181	0.019
	Evergreen	28	-0.037	0.035	0.048	0.047	0.060	0.128	0.019

Table 10. Percentiles for the difference in the trend parameter posterior distributions for the 1, 3, and 5 year sampling frequencies. The realizations from the posterior distribution of the difference are also used to provide an estimate of the probability that the difference is greater than zero.

Frequency	Year	Min.	1st	Median	Mean	3rd	Max.	SD	Prob (diff<0)
1yr	3	-1.007	-0.157	0.026	0.030	0.209	1.091	0.278	0.467
	18	-0.026	0.035	0.050	0.049	0.064	0.126	0.022	0.020
	30	0.009	0.043	0.050	0.050	0.057	0.083	0.011	0.000
3yr	3	-1.155	-0.146	0.037	0.038	0.222	1.305	0.280	0.445
	18	-0.099	0.028	0.054	0.054	0.079	0.209	0.038	0.080
	30	-0.035	0.037	0.051	0.051	0.064	0.138	0.020	0.006
5yr	3	-1.407	-0.169	0.024	0.023	0.216	1.375	0.296	0.466
	18	-0.191	0.022	0.052	0.052	0.083	0.222	0.046	0.125
	28	-0.080	0.036	0.053	0.052	0.069	0.154	0.025	0.021

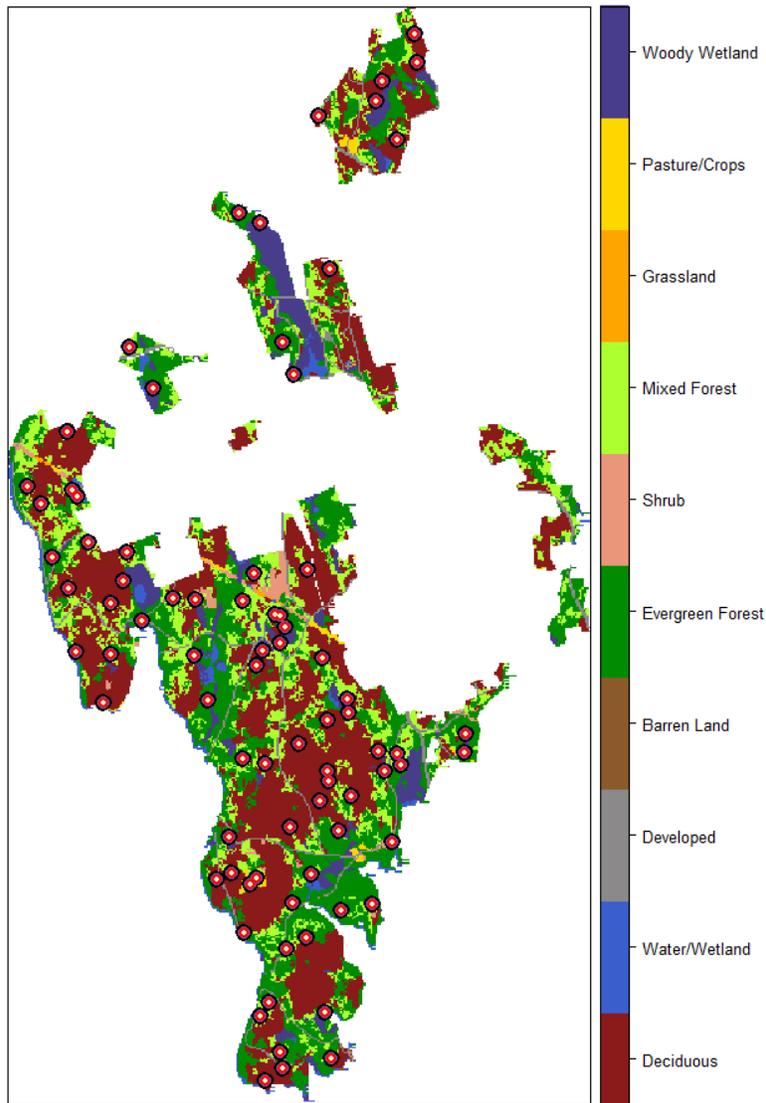


Figure 10. NLCD coverage of Harvard Forest Core Site within Domain 1. Distributed sample locations determined by the RRQRR design are shown for the vegetation types selected for sampling.

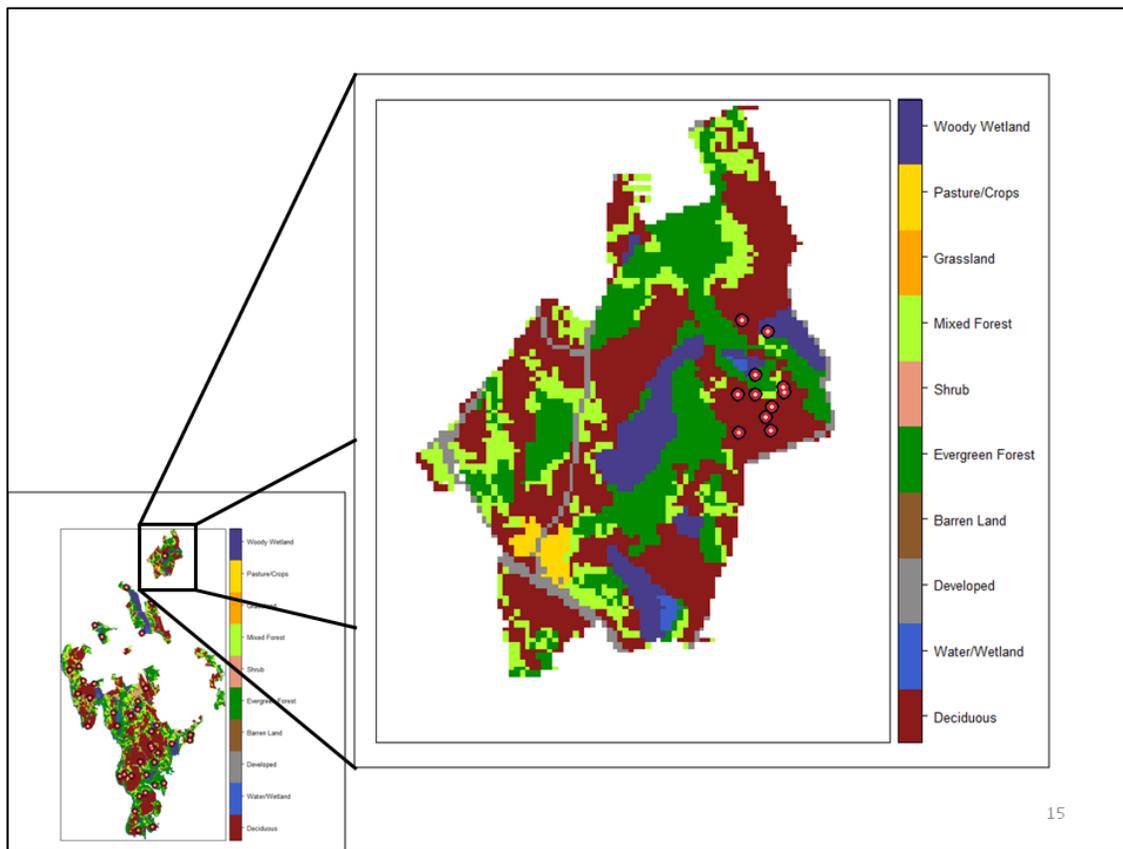


Figure 11. NLCD coverage of Harvard Forest Core Site within Domain 1. Sample locations are shown for the historical data used to inform the simulation study.

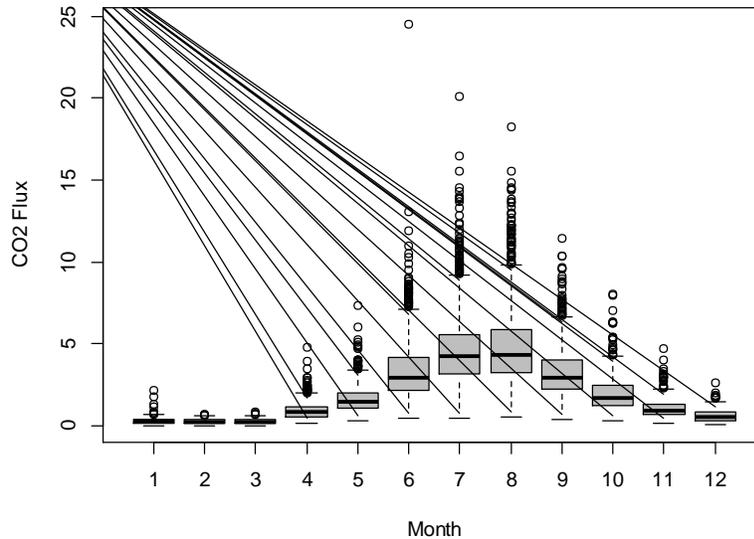


Figure 12. Boxplots of monthly distributions for CO₂ flux from the control data, collected by PI Davidson using manual soil collars at the Harvard Forest. Metadata for the entire Harvard forest archive are available online (<http://harvardforest.fas.harvard.edu:8080/exist/xquery/data.xq?id=hf194>). The grey boxes span the 25th and 75th percentile (i.e., the inter-quartile range, IQR). Medians are depicted with the black horizontal line within each box. Whiskers for each boxplot extend out to the largest value less than the median $\pm 1.5 \times \text{IQR}$. Observations outside the median $\pm 1.5 \times \text{IQR}$ are depicted as open circles.

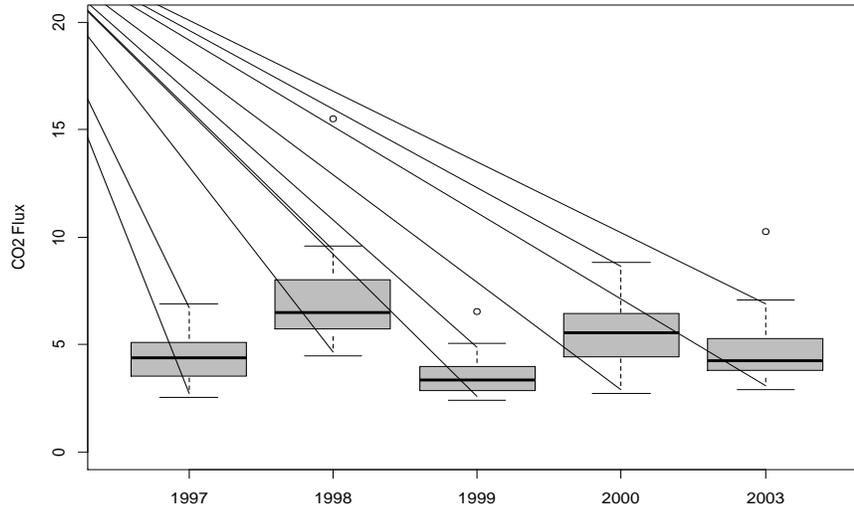


Figure 13. Boxplots of annual peak CO₂ flux distributions for the final working dataset from the Harvard Forest. Each year has 12 samples arising from 4 locations with 3 collars each. Note that data are omitted for years 2001 and 2002 due to unequal sample sizes. The data for each year are taken from a single sampling event occurring towards the end of the month of August.

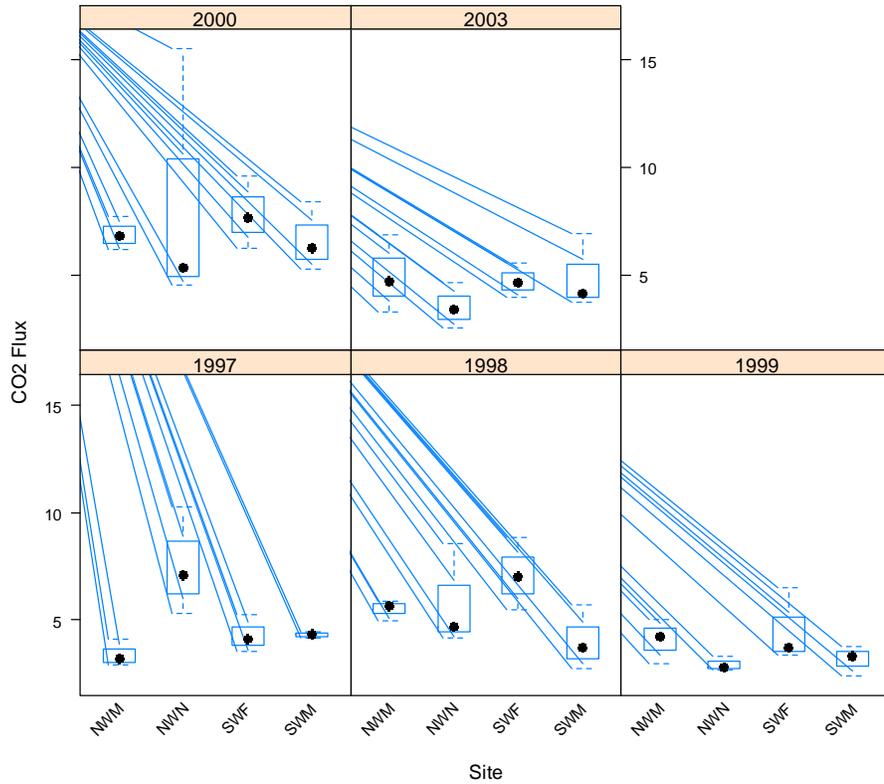


Figure 14. Boxplots of annual peak CO₂ flux distributions by site within each year for the final working dataset from the Harvard Forest. There is some evidence that inter-annual variability varies among sites. For example, NWN appears to display greater inter-annual variability than NWM.

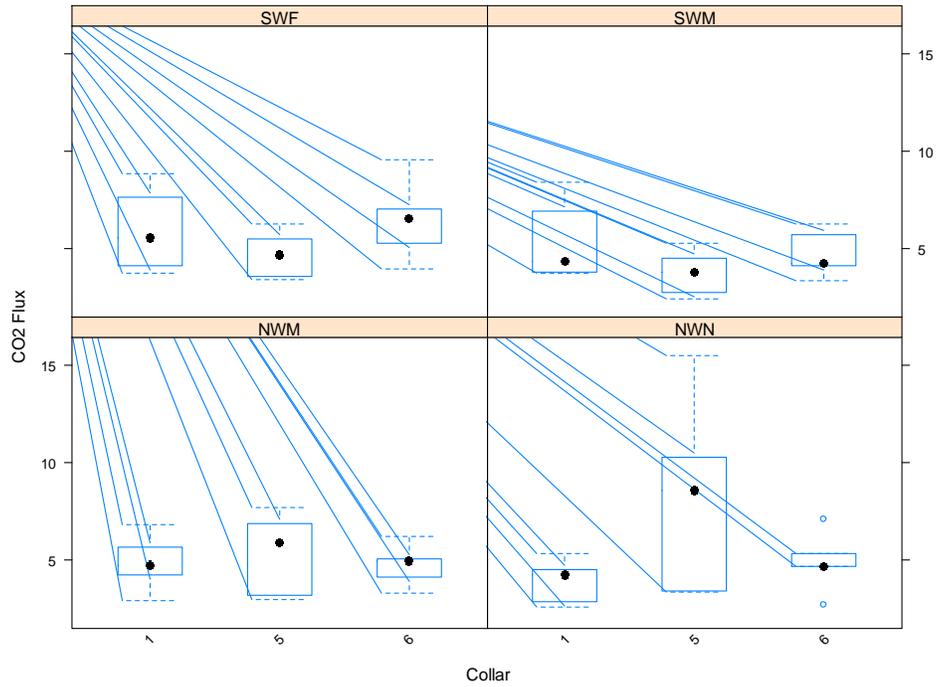


Figure 15. Boxplots of annual peak CO2 flux distributions by site within each year.

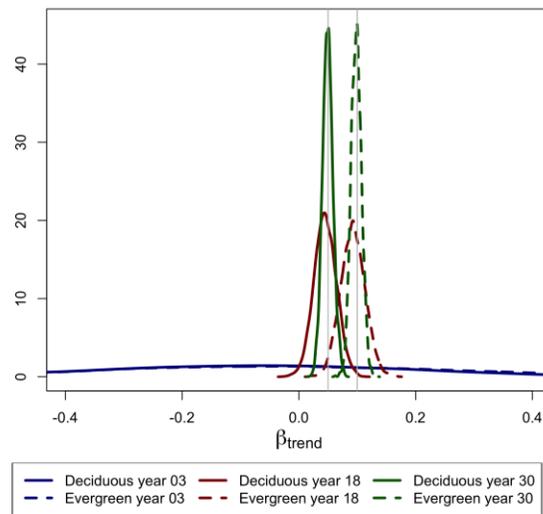


Figure 16. Posterior distributions for the trend parameters at years 3, 18, and 30 for both the deciduous and evergreen vegetation types with annual sampling.

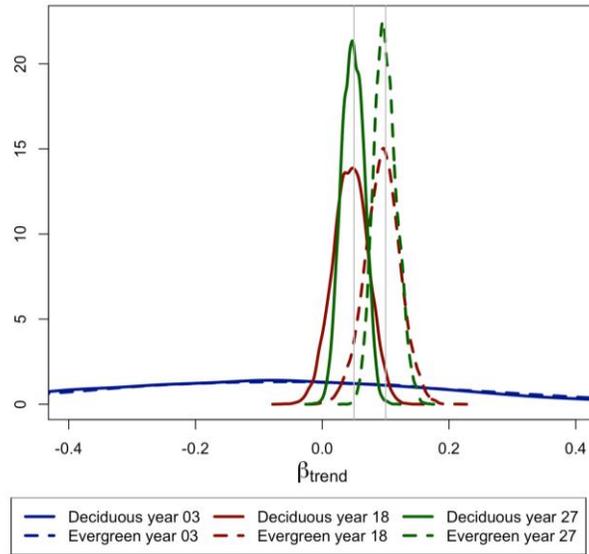


Figure 17. Posterior distributions for the trend parameters at years 3, 18, and 27 for both the deciduous and evergreen vegetation types with sampling every 3 years.

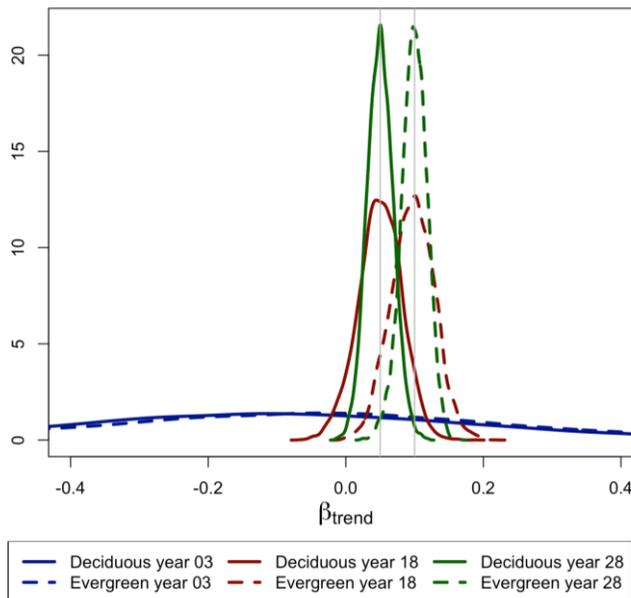


Figure 18. Posterior distributions for the trend parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every 5 years.

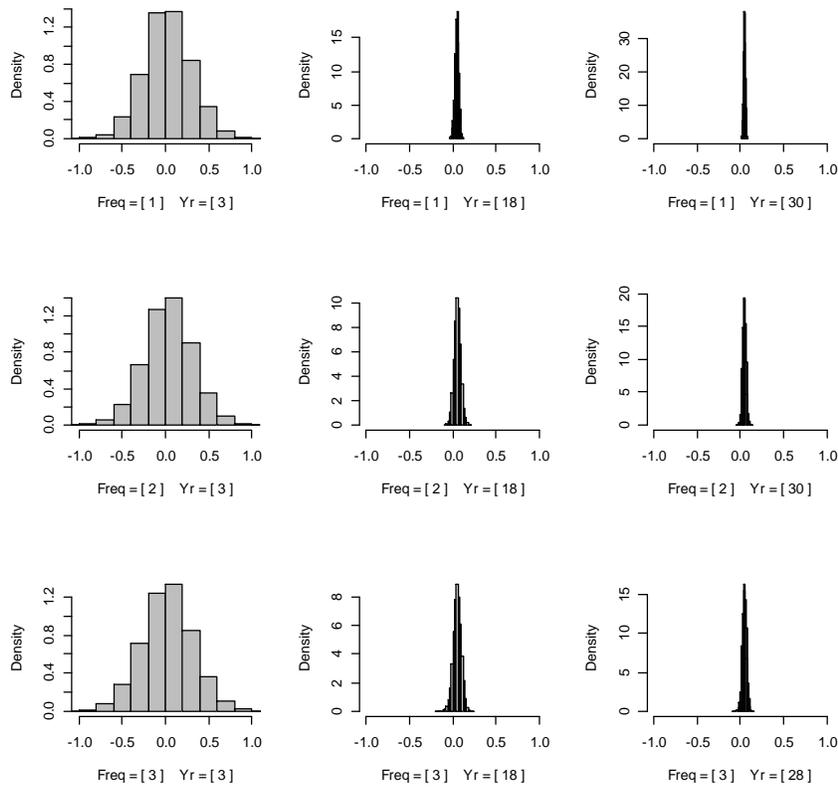


Figure 19. Histograms of the posterior distributions of the difference in the trends as a function of sampling frequency and year.

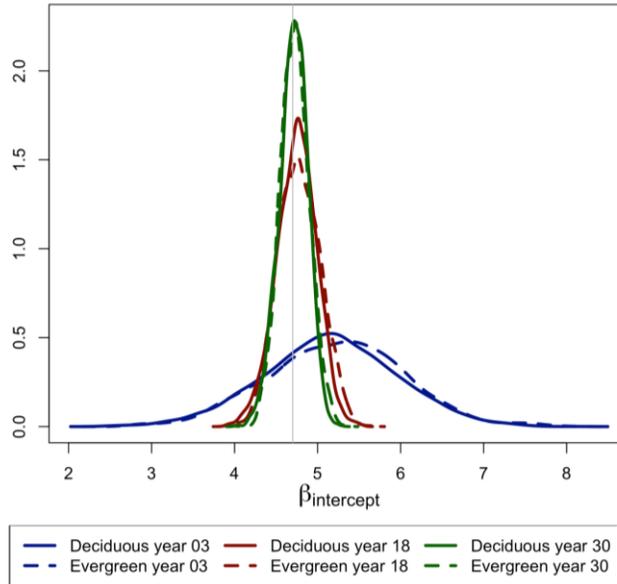


Figure 20. Posterior distributions for the intercept parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every year.

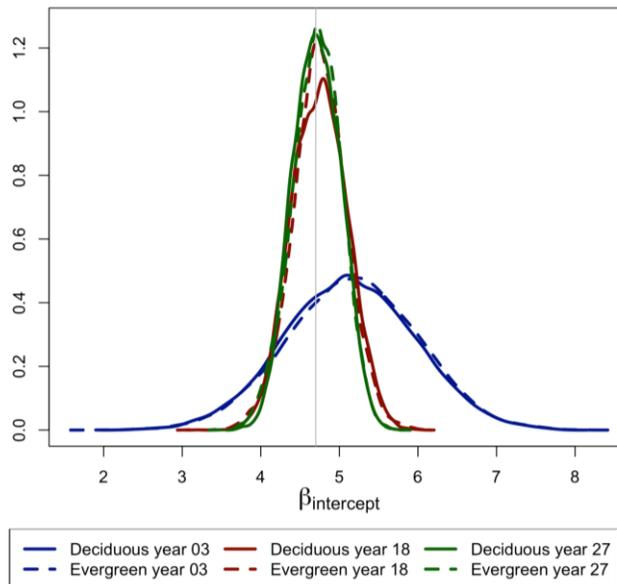


Figure 21. Posterior distributions for the intercept parameters at years 3, 18, and 27 for both the deciduous and evergreen vegetation types with sampling every 3 years.

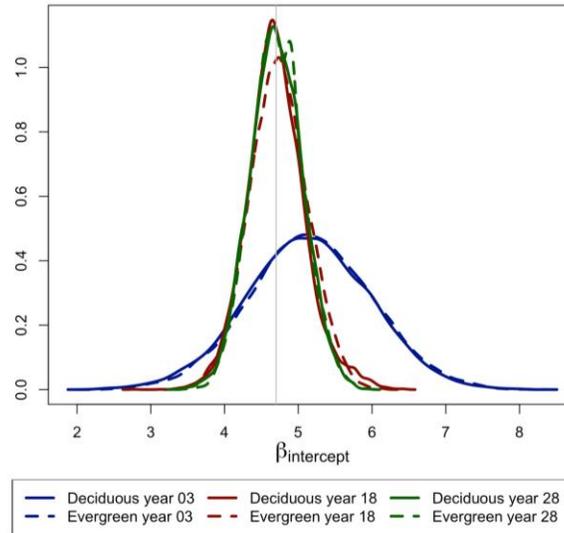


Figure 22. Posterior distributions for the intercept parameters at years 3, 18, and 28 for both the deciduous and evergreen vegetation types with sampling every 5 years.

A.14 Code Used for Simulations

#collection of functions used for simulating and analyzing data

```
spatio_temporal.post = function(theta.params,form,data.df,priors,post=TRUE,separable=NULL){
  n.obs = nrow(data.df)
  sites.sp = data.df[!duplicated(data.df[,c("x","y")]),]
  sites.n = nrow(sites.sp)
  sites.t = unique(data.df$t)
  t.n = length(sites.t)
  n.beta = length(greexpr("beta",form)[[1]])
  beta = as.numeric(theta.params[1:n.beta])
  sigma.meas = ifelse(is.na(as.numeric(theta.params[n.beta+1])),1,as.numeric(theta.params[n.beta+1]))
  sigma.process = ifelse(is.na(as.numeric(theta.params[n.beta+2])),1,as.numeric(theta.params[n.beta+2]))
  h.range = as.numeric(theta.params[n.beta+3]) # spatial covariance range
  u.rho = as.numeric(theta.params[n.beta+4]) # temporal AR1 coefficient
  if(separable==F){sp.time.int = as.numeric(theta.params[n.beta+5])}# space-time interaction coefficient

  priors = sum(unlist(lapply(1:length(theta.params),
    function(i,theta){
      x=theta[i]; dens=(eval(parse(text=priors[[i]])); ifelse(is.na(dens),log(0.01),dens);
    },
    theta=theta.params)))
  if(priors==-Inf|priors=="NaN"){priors = -2e7}
  if( sigma.meas > 0 & sigma.process > 0 & sigma.process < 3 & h.range > 0 & h.range < 15000 & u.rho > 0 &
    u.rho < 1 ){
    spdmat = as.matrix( dist( data.df[,c("x","y")], diag=TRUE, upper=TRUE ) )
    tdistmat = as.matrix( dist( data.df[, "t"], diag=TRUE, upper=TRUE ) )
    nspdist = spdmat / h.range
    ntndist = tdistmat / u.rho + 1
    if(separable==T){sp.time.int<-0}
    V.process = exp( -nspdist / ntndist^(sp.time.int/2) ) / ntndist
  }
  V.meas<- diag(nrow(data.df))
  V.all<- sigma.process*V.process + sigma.meas*V.meas
  y = data.df[, "response_1"]
}
```



```
yhat = with(data.df,eval(parse(text=form)))
loglik= dmvnorm(yhat,y,V.all,log=T)
if(post){
  logpost = loglik + priors
}else{
  logpost = loglik
}
} else {
  logpost = -1e5
}
if(logpost==-Inf|logpost=="NaN"){logpost = -3e5}
logpost
}
```

#####

```
model_choice.mcmc = function(mcmc.samples,separable=NULL){
  n.samples = nrow(mcmc.samples$params)
  w=rep(1/n.samples,n.samples)
  theta.mean = crossprod(w,mcmc.samples$params)
  df = length(theta.mean)
  theta = attr(mcmc.samples$data,"theta")
  theta.median = apply(mcmc.samples$params,2,median)

  loglik.med = spatio_temporal.post(theta.params=theta.median,form=mcmc.samples$form,separable=separable,
    data.df=mcmc.samples$data,priors=mcmc.samples$posteriors,post=T)

  loglik.mu = spatio_temporal.post(theta.params=theta.mean,form=mcmc.samples$form,separable=separable,
    data.df=mcmc.samples$data,priors=mcmc.samples$posteriors,post=T)
  loglik.sum = 0
  lik.marg = 0
  post = numeric(n.samples)
  post = apply(mcmc.samples$params,1,spatio_temporal.post,mcmc.samples$form,separable=separable,
    data.df=mcmc.samples$data,priors=mcmc.samples$posteriors,post=T)
  center.post = mean(post)
  alpha = var(post)
  lmax = center.post + alpha
  lik.marg = sum(1/post)^1
  pD = -2*(center.post - loglik.mu)
  DIC = pD - (2*center.post)
  lik.marg = (lik.marg/n.samples)^(-1)
  AIC = -2*mcmc.samples$loglik + 2 * df
  BIC = -2*mcmc.samples$loglik + df * log(n.samples)
  list(AIC=list(AIC=AIC,df=df,loglik=mcmc.samples$loglik),
    DIC=list(DIC=DIC,pD.med=-2*(median(post) - loglik.med),dev.med=-2*loglik.med,med.dev=-2*median(post),DIC.med = pD - 2*median(post),
      mean.dev=-2*center.post,dev.mean=-2*loglik.mu,pD=pD),
    BIC=list(BIC=BIC,pD=pD,loglik=mcmc.samples$loglik,n=n.samples),
    lik_marg=list(lik.marg=lik.marg,alpha=alpha,lmax=lmax))
}
```

#####

```
mcmc.samples.plot = function(obj,yr=1,separable=F){
  #if(yr==5){obj$priors[]<-sub("log","",obj$priors[])}
  obj$priors[]<-sub("log","",obj$priors[])
  params.df = stack(as.data.frame(do.call("cbind",lapply(obj$priors,
    function(distr,n){
      distr=gsub("d","r",distr)
      distr=gsub(",log=TRUE","",distr)
      distr=gsub("x","n",distr)
      distr=gsub("renp","rexp",distr)
    },n=20000))))
  eval(parse(text=distr))
},n=20000)))
```



```

params.df$type="prior"
tmp=stack(as.data.frame(obj$params))
tmp$type="posterior"
params.df = rbind(params.df,tmp);rm(tmp)
params.df$test = rep(NA,nrow(params.df))
params.df$test[params.df$ind=="beta0"] = c("Beta 0")
params.df$test[params.df$ind=="beta1"] = c("Beta 1")
params.df$test[params.df$ind=="beta2"] = c("Beta 2")
params.df$test[params.df$ind=="sigma.meas"] = c("Measurement Sigma")
params.df$test[params.df$ind=="sigma.process"] = c("Process Sigma")
params.df$test[params.df$ind=="spat.range"] = c("Spatial Range")
params.df$test[params.df$ind=="time.range"] = c("Temporal Range")
params.df$test[params.df$ind=="sp.time.int"] = c("Space Time Int")

print(densityplot(~values | test,
  group=type,adjust=2.0,
  data=params.df,lwd=2,
  panel = function(...) {
    panel.fill(col = "#E1EBE3FF")
    panel.densityplot(...)},
  scales=list(x=list(relation="free"),y=list(relation="free")),
  plot.points=F,main=paste("Year",yr),
  auto.key=T,xlab="Parameter Values")
)
}

#####

spattemp.mcmc = function(yr.begin=3,yr.end=3,data,force.samp=T,facr=1e10,separable=NULL,update.interval=5,
  formula=NULL,name="exp",mcmc=1000,burnin=1000,thin=1,tune=NULL,plot.t=F){

  library(ramps);library(lattice);library(sp);
  library(MCMCpack);library(mvtnorm);library(SuppDists)
  base_name = paste(name,"_post_yr",sep="")

  saved.samples = ls(pat=base_name,envir=.GlobalEnv)
  remove.samples = saved.samples[saved.samples %in% paste(base_name,formatC(yr.begin:10,digits=1,flag=0),"mcmc",sep="")]
  theta = attr(data,"theta")
  sigma.meas = theta[["sigma.meas"]]
  sigma.process = theta[["sigma.process"]]
  if(is.null(formula)) {
    formula = attr(data,"formula")
    beta = theta[["beta"]]
    n.betas = length(beta)
  } else {
    n.betas = length(gregexpr("beta",formula)[[1]])
    beta = c(rnorm(1,mean(data[data$t<5,"response_1"]),sqrt(var(data[data$t<5,"response_1"])/3),runif(n.betas-1,-1,1.5))
  }
  param.names = c(paste("beta",0:(n.betas-1),sep=""),"sigma.meas","sigma.process","spat.range","time.range","sp.time.int")

  if(separable==T){
    param.names<-param.names[-length(param.names)]
  }

  n.params = length(param.names)
  if(is.null(tune)){
    tune = rep(1,n.params)
  } else if(length(tune)<n.params) {
    tune = rep(tune,n.params)
  }
  beta=c(4.7,1.)
  post_yr2.mcmc = list(
    theta.init = c(
      rnorm(1,beta[1],0.25), # beta0

```



```
rep(0,length(beta)-1), # beta1
0.5, # sigma.meas
0.47, # sigma.process
500, # spt.range
0.05, # time.range
0.01 # sp.time.int
),
posterior = as.list(
c(paste("log(dnorm(x,5.3,1))",sep=""), # beta0
rep("log(dnorm(x,0,.5))",n.betas-1), # beta1
"log(dunif(x,0.01, 2.5))", # sigma.meas
"log(dinvgamma(x,shape = 5, scale = 3))", # sigma.process
"log(dinvgamma(x,shape = 3, scale = 1000))", # spat.range
"log(dinvgamma(x,shape = 5, scale = 0.2))", # time.range
"log(dinvgamma(x,shape = 10, scale = 2))" # sp.time.int
)
)

if(separable==T){
post_yr2.mcmc[["posteriors"]] = post_yr2.mcmc[["posteriors"]][-n.params]
post_yr2.mcmc[["theta.init"]] = post_yr2.mcmc[["theta.init"]][-n.params]
}

names(post_yr2.mcmc[["posteriors"]]) = param.names
assign(paste(base_name,"02.mcmc",sep=""),post_yr2.mcmc)

year.vec<-seq(yr.begin,yr.end,update.interval)

for(yr in year.vec){
if(yr==3){update.interval=1}
var.name = paste(base_name,formatC(yr,width=2,flag=0),".mcmc",sep="")
var_prev.name = paste(base_name,formatC(yr-update.interval,width=2,flag=0),".mcmc",sep="")
var_prev.name = ifelse(exists(var_prev.name),var_prev.name,
ls(pat=base_name,envir=.GlobalEnv)[length(ls(pat=base_name,envir=.GlobalEnv))]
)

data.update<-data[data$T%in%year.vec,]
data.update<-data.update[data.update$T<=yr,]
priors.update<-post_yr2.mcmc[["posteriors"]]
low.vec<-c(3,rep(-1, n.betas-1), .01, .01, 1, 0.01, 0.01)
up.vec<-c(7,rep(1, n.betas-1), 4, 4, 10000, 1, 1)

if(separable==T){
low.vec<-low.vec[-n.params]
up.vec<-up.vec[-n.params]
}

MLEs = optim(get(var_prev.name)[["theta.init"]],
spatio_temporal.post,
separable=separable,
form=formula,
data.df=data.update,
priors=priors.update,
post=F,hessian=T,
method = "L-BFGS-B",
lower = low.vec,
upper = up.vec,
control=list(fnscale=-1,factr=factr,maxit=20)
)

if(is.null(get(var_prev.name)[["params"]])) {
V.proposal = NULL
} else {
V.proposal = var(get(var_prev.name)[["params"]])
}
```



```

}

params = MCMCmetrop1R(spatio_temporal.post,
separable=separable,
  theta.init=get(var_prev.name)[["theta.init"]],
  form=formula,
  data.df=data.update,
  priors=priors.update,
  mcmc=mcmc, burnin=burnin, thin=thin, force.samp=force.samp,
  tune=tune,
  V = V.proposal,
  verbose=T,
  optim.method = "L-BFGS-B",
  optim.lower = low.vec,
  optim.upper = up.vec
)

dimnames(params) = list(NULL,param.names)
posterior = lapply(param.names,
  function(param,data){
    tmp = try( JohnsonFit(data[,param]) )
    if( inherits( tmp, "try-error" ) ){
      return( paste( "log(dnorm(",mean(data[,param]),",",sd(data[,param]),")")" )
    } else {
      return( paste("dJohnson(x, list(",paste(tmp[-5],collapse=","),
        ",","tmp"),log=TRUE)",sep="" )
    }
  },
  data=params)
names(posteriors) = param.names
mcmc.samples = list(
  form = formula,
  yr=yr,
  data = data.update,
  mles = MLEs$par,
  loglik = MLEs$value,
  theta.init = apply(params,2,median),
  params = params,
  posteriors = posteriors,
  priors = priors.update,
prev.post = get(var_prev.name)[["posteriors"]]
)

mcmc.samples$model_choice = model_choice.mcmc(mcmc.samples,separable=separable)
assign(var.name,mcmc.samples,envir=.GlobalEnv)
if(plotit){
  win.graph()
  mcmc.samples.plot(obj=mcmc.samples,yr,separable=separable)
}
}
}

#####

spattemp_sim.data = function(formula="beta[1] + beta[2] * t",data,
  n.reals=1,separable=F,
  beta=c(100,1.1),sigma.process=0.15,sigma.meas=0.1,
  h.range=0.01,u.rho=0.05,sp.time.int = 0.05){
require(ramps)
# generate the spatial covariance function
spdmat = as.matrix( dist( data[,c("x","y")], diag=TRUE, upper=TRUE ) )
tdistmat = as.matrix( dist( data[, "t"], diag=TRUE, upper=TRUE ) )
nspdist = spdmat / h.range
ntdist = tdistmat / u.rho + 1

```



```

    if(separable==T){ sp.time.int<-0}
    cormat = exp( -nspdist / ntdist^(sp.time.int/2) ) / ntdist
    V = sigma.process * cormat + sigma.meas * diag(nrow(data))
    # one check if the covariance matrix is valid (add other checks based on Gneiting 2002)
    cat(ifelse(min(eigen(V)$values) > -.Machine$double.eps,"Covariance matrix appears valid","Covariance function not positive semi-
    definite"),"\n")
    # generate the response (need the mvtnorm library to generate mvnrm variates)
    responses = with(eval(parse(text=formula)),data=data) + t(mvtnorm::rmvnorm(n.reals,rep(0,nrow(data)),V))
    responses = as.data.frame(responses)
    names(responses) = paste("response_",1:ncol(responses),sep="")
    data = as.data.frame(cbind(responses,data))
list(beta=beta,sigma.process=sigma.process,sigma.meas=sigma.meas,h.range=h.range,u.rho=u.rho,sp.time.int=sp.time.int)
    data$response_1[data$response_1<0]<-0
    return(data)
}

# Code used to generate realizations for this work
mcmcRdata = NULL
for(realization in realizations){
  for(sigProc in sigProcs){ # sigProc = sigProcs[1]
    combined_sites_lin.df = spattemp_sim.data(formula="beta[1]+beta[2]*t",data=rbind(dataDeciduous,dataEvergreen),separable=T,
      beta=c(beta0, beta1[1]), sigma.process=sigProc, sigma.meas=sigmeas,
      h.range=spatrange, u.rho=timerange, sp.time.int = sptimeint)
    decid_sites_lin.df = combined_sites_lin.df[1:nrow(dataDeciduous),]
    assign(paste("decid_sites_",sigProc,"_",realization,".df",sep=""),decid_sites_lin.df)
    if(is_full_stoch){
      green_sites_lin.df = spattemp_sim.data(formula="beta[1]+beta[2]*t",data=dataEvergreen,separable=T,
        beta=c(beta0, beta1[2]), sigma.process=sigProc, sigma.meas=sigmeas,
        h.range=spatrange, u.rho=timerange, sp.time.int = sptimeint)
    }else{
      green_sites_lin.df = decid_sites_lin.df
      green_sites_lin.df$response_1 = green_sites_lin.df$response_1 * 2
    }
    assign(paste("green_sites_",sigProc,"_",realization,".df",sep=""),green_sites_lin.df)
    for(interval in intervals){ # interval = intervals[2]
      for(vegType in vegTypes){
        cat("----- ",sigProc,interval,vegType," -----","\n")
        spattemp.try
      }
      try(spattemp.mcmc(yr.begin=3,yr.end=30,data=eval(parse(text=paste(vegType,"_sites_lin.df",sep=""))),formula="beta[1]+beta[2]*t",
        separable=T,
        name=paste(vegType,"_",interval,"_",sigProc,sep=""),burnin=2000,mcmc=5000,thin=10,update.interval=interval,tune=c(1,1,1,1,1),plotit=F))
      if(inherits(spattemp.try,"try-error")) break
      save.image(file=paste(host,"-",is_full_stoch,"-",interval,"-r",realization,".Rdata",sep=""))
    }
    if(interval==5) {
      yrs = c("03","08","13","18","23","28")
    }else if(interval==3) {
      yrs = c("03","06","09","12","15","18","21","24","27","30")
    }else if(interval==1){
      yrs = c(paste("0",3:9,sep=""),10:30)
    }
    fileRdata = paste(host,"-",is_full_stoch,"-",interval,"-",min(realizations),"-",max(realizations),".Rdata",sep="")
    for(yr in yrs){
      beta2Diffs2 = rbind(beta2Diffs2,c(realization,sigProc,interval,yr,
        eval(parse(text=paste(vegTypes[1],"_",interval,"_",sigProc,"_post_yr",yr,".mcmc$params",sep="")))[,2],
        eval(parse(text=paste(vegTypes[2],"_",interval,"_",sigProc,"_post_yr",yr,".mcmc$params",sep="")))[,2]
      ))
    }
    if(exists(paste(vegType,"_",interval,"_",sigProc,"_post_yr",yr,".mcmc$params",sep=""))){
      mcmcRdata = rbind(mcmcRdata,
        get(paste(vegTypes[1],"_",interval,"_",sigProc,"_post_yr",yr,".mcmc$params",sep="")),
        get(paste(vegTypes[2],"_",interval,"_",sigProc,"_post_yr",yr,".mcmc$params",sep="")))
    }
  }
}# yr

```



<i>Title:</i> TOS Science Design for Terrestrial Biogeochemistry		<i>Date:</i> 04/06/2022
<i>NEON Doc. #:</i> NEON.DOC.000906	<i>Author:</i> E. Hinckley	<i>Revision:</i> C

```
save(mcmcRdata,file=fileRdata)
write.csv(beta2Diffs2,paste("beta2Diffs2-",host,"-",is_full_stoch,"-",interval,"-",min(realizations),"-",max(realizations),".csv",sep=""),row.names=FALSE)
}# interval
}
}
```