Sink or Source? Exploring Forest Carbon Dynamics with NEON Data and AI

NEON Eddy Covariance • Google Colab • Gemini AI

About this module

- **Big question:** Do forests absorb carbon dioxide or release it — and does the answer depend on the time of day?
- **What you will do:** Analyze real data from a mountain forest in Colorado and a New England forest, then use an AI assistant to help you understand the patterns.
- **Time required:** About 3–4 hours total. Your instructor will tell you how this is split across class sessions.
- **What you need:** A Google account and access to the shared Google Colab notebook (link from your instructor). No prior coding experience required — all code is provided.

What you will learn

- How scientists measure carbon exchange between forests and the atmosphere
- How to read and interpret real environmental data — not just textbook examples
- How to use an AI assistant as a tool for data exploration — and when to trust it and when not to
- How to communicate scientific findings for a real audience
- How to estimate the energy cost of your own AI use

PART A: Foundations

Before you start analyzing data or using AI, you need to understand the science and the tools. Part A walks you through both step by step. Take your time here — everything in Parts B and C builds on this foundation.

A1. The Science: How Do Forests Interact with Carbon Dioxide?

You have probably heard that forests help fight climate change by absorbing carbon dioxide (CO_2). But the story is more complicated than that. Forests are constantly both absorbing and releasing CO_2 at the same time — and which one wins depends on the time of day, the season, the weather, and the type of forest.

Two processes are happening simultaneously in every forest:

Photosynthesis — CO_2 IN	Respiration — CO_2 OUT
Plants use sunlight to convert CO_2 and water into sugar and oxygen.	Plants, soil microbes, and decomposers break down organic matter and release CO_2.
Only happens during daylight.	Happens 24 hours a day.
Rate increases with more sunlight.	Rate increases with warmer temperatures.
Net effect: CO_2 removed from atmosphere.	Net effect: CO_2 added to atmosphere.

The net result — the balance of these two processes — is what scientists call the Net Ecosystem Exchange (NEE) or, in the NEON data you will use, the Net Surface-Atmosphere Exchange (NSAE).

The key variable you will work with

- **Name:** data.fluxCo2.nsae.flux
- **What it measures:** The net movement of CO_2 between the whole ecosystem (soil + plants + air) and the atmosphere above it, averaged over 30-minute windows.
- **Units:** micromoles of CO_2 per square meter per second ($\mu mol\ CO_2\ m^{-2}\ s^{-1}$)
- **Negative value:** More CO_2 is going INTO the ecosystem than leaving. The forest is acting as a carbon SINK. → Photosynthesis is winning.
- **Positive value:** More CO_2 is leaving the ecosystem than entering. The forest is acting as a carbon SOURCE. → Respiration is winning.
- **Zero:** The two processes are exactly balanced at that moment.

A1.1 Check Your Understanding

Answer these questions in your own words before moving on. There are no trick questions — just check that the concept makes sense.

1. It is 2 pm on a sunny July afternoon at a forest. Would you expect the CO_2 flux to be negative or positive? Explain your reasoning.

2. It is 2 am. The forest is dark and there is no wind. Would you expect the CO_2 flux to be negative or positive? Explain.

3. A scientist says: “This forest is a carbon sink in summer.” What does that mean in plain language? What would the data look like?

A2. The Sites: Where Does This Data Come From?

The data you will analyze comes from the National Ecological Observatory Network (NEON). NEON operates towers at ecosystems across the United States that continuously measure CO_2 exchange, temperature, humidity, light, and dozens of other variables.

You will work with data from two NEON sites collected in June and July 2018:

	Site	Fast facts
NIWO	**Niwot Ridge, Colorado** *Subalpine forest*	Elevation: ~3,050 m (10,000 ft) in the Rocky Mountains Mostly Engelmann spruce and subalpine fir trees Short growing season — summer is brief and intense

HARV	**Harvard Forest, Massachusetts** *Temperate deciduous forest*	Elevation: ~340 m (1,100 ft) in the New England hills Mostly oak, maple, and birch trees Longer, warmer growing season than NIWO

As you work through the data, keep these differences in mind. Two forests in very different climates and at very different elevations — will they behave the same way?

A2.1 Before You Look at the Data

Make a prediction. Based on what you just read about these two sites and what you know about photosynthesis:

4. Which site do you predict will show stronger carbon uptake (more negative flux values) in summer? Why?

5. What time of day do you predict will show the most negative (highest uptake) flux values? Why?

A3. The Tool: What Is an AI Language Model?

You will use Gemini, an AI assistant made by Google, to help you understand and analyze the data. Before you do, it helps to understand how it works — and why that matters for science.

What Gemini is

- Gemini is a large language model (LLM). It was trained on a huge amount of text from the internet, books, and scientific papers.
- When you type a question, it predicts what a helpful, knowledgeable response would look like — based on patterns in that training data.
- It does not have eyes or a web browser during your conversation. It cannot see your data unless you paste information into the prompt.

- It can be very helpful for explaining concepts, writing code, and interpreting patterns — but it can also sound confident when it is wrong.

What Gemini is NOT

- It is not looking at your NEON data unless you tell it the numbers.
- It is not a database or a search engine — it is generating text, not retrieving facts.
- It does not know about your specific notebook, your specific plots, or your specific site unless you tell it.
- It is not always right. Scientists (including you) need to check its answers.

This is why the most important skill in this module is not knowing how to use AI — it is knowing how to CHECK what AI tells you. You will practice this throughout.

A3.1 The Grounding Principle

The single most important thing to understand about prompting an AI for data analysis is this:

If you give Gemini no context about your data,
it will give you a textbook answer, not a data answer.
Grounding = giving Gemini your actual data context in the prompt.

The notebook is already set up to do this for you automatically — a function called build_neon_context() packages up the site name, variable descriptions, sample data rows, and statistics and passes them to Gemini with every question. You will see how this works in A4.

A4. Getting Started in the Colab Notebook

Open the shared Colab notebook. Each step below corresponds to a section in the notebook. Run each code cell by clicking the play button (▶) on the left side. A green checkmark means it finished successfully.

Step-by-step: running the notebook

- **Step 1 — Install packages:** Run the first cell (pip install neonutilities). This installs the tools needed to download NEON data. It takes about 30 seconds.

- **Step 2 — Import libraries:** Run the imports cell. This loads Python tools for data analysis and plotting.
- **Step 3 — Add your API key:** Go to the key icon (🔑) in the left sidebar of Colab. Add a secret named GOOGLE_API_KEY and paste in your Gemini API key. Your instructor will provide this or guide you through getting one.
- **Step 4 — Download NEON data:** Run the nu.zips_by_product cell. This downloads eddy covariance data for Niwot Ridge (NIWO) and Harvard Forest (HARV) for June–July 2018. This can take 1–2 minutes.
- **Step 5 — Stack and load the data:** Run the nu.stack_eddy cell. This organizes the downloaded files into a format Python can work with.
- **Step 6 — Set up Gemini:** Run the Gemini setup cell. This connects to the API and builds the context block from your data.

A4.1 Explore the Dataset

After running the setup cells, look at the output of flux['NIWO'].head and flux['variables'] in the notebook. Then answer these questions:

6. Look at the timeBgn column. What time interval is between measurements? (Hint: look at the first few rows.)

7. Find data.fluxCo2.nsae.flux in the variables table. Write down the full description of what this variable measures.

8. Run the .describe() cell for NIWO. What is the mean CO_2 flux value? Is it positive or negative? What does that tell you about this forest in summer?

9. The maximum value is around 310 μmol CO_2 m^{-2} s^{-1}. Does that seem reasonable for a forest? Compare it to the 75th percentile value. What does this suggest about that data point?

A4.2 Your First Plots

The notebook includes two pre-built plots. Run the cells to generate them, then describe what you see.

Plot 1: CO_2 flux over time (the full June–July time series)

Plot 2: CO_2 flux over two days (July 7–9, 2018, zoomed in)

Describe the pattern within each day. When is flux most negative? When is it positive? Does this match your prediction from A2.1?

Plot 3: CO_2 flux vs. PAR (light level). PAR stands for Photosynthetically Active Radiation — it measures how much light is reaching the treetops.

Describe the shape of the scatter plot. As light increases, what happens to CO_2 flux?

A4.3 First AI Interaction: Variable Explanation

Now you will use Gemini for the first time. The notebook has a function called ask_about_variable(). Use it to ask about data.fluxCo2.nsae.flux.

Prompt Scaffold A — Copy this structure into your prompt log

Using the NEON data loaded above, I want to understand the variable `data.fluxCo2.nsae.flux`. Specifically, I am confused about why some values are negative and some are positive, and what that means for what is happening at Niwot Ridge in summer 2018. Please explain it using the actual data statistics provided.

Read Gemini's full response. Then complete the verification checklist below before moving on.

☐	Did Gemini correctly state the units as μmol CO_2 m^{-2} s^{-1}?
☐	Did it correctly say that NEGATIVE = uptake (not release)?
☐	Did it reference the actual mean or min/max values from your .describe() output?
☐	Did it say something specific about Niwot Ridge, subalpine conditions, or elevation — not just generic forest facts?
✓	**If ANY of the above are wrong or missing: revise your prompt to give more context, then re-run.**

Part A Deliverables — check before you move on

- Written answers to all questions in A1.1, A2.1, A4.1, and A4.2 (in this workbook or in your notebook markdown cells).
- Your prompt log entry for the A4.3 interaction: the prompt you used and a 2–3 sentence summary of Gemini’s response.
- Verification checklist A4.3 completed and signed off.

PART B: Inquiry and Analysis

Now you take the lead. You will choose a research question, analyze the data, and use Gemini as a tool to help you — but your reasoning comes first.

B1. Choose Your Research Question

Pick one of the questions below, or write your own with instructor approval. Write it here:

Research Question Options

- Is Niwot Ridge a net carbon sink or source in summer 2018? Does the answer change depending on the time of day?
- How does the amount of light (PAR) affect CO_2 uptake at NIWO? Does the relationship look like what you would predict from photosynthesis?
- Are there any days in the dataset where the ecosystem behaves strangely? Can you identify what might cause that?
- Does Niwot Ridge absorb more or less carbon than Harvard Forest in the same time period? What might explain the difference?
- [Your own question — must be answerable with data already loaded; check with instructor first]

My research question:

B2. Your Pre-AI Observation (Required Before Using Gemini)

Before you use Gemini for any analysis in Part B, you must complete this step. Make a plot relevant to your research question, look at it carefully, and write down what YOU observe. This protects you from just copying Gemini's interpretation without thinking.

10. Make a plot related to your question (the notebook has tools to help). Describe the plot here: what is on each axis, and what range of values do you see?

11. What pattern do you see? Describe it in plain language.

12. Based on your ecology background, what do you think is driving that pattern?

13. What are you most uncertain about? What do you NOT know how to interpret?

B3. AI-Assisted Analysis

Now use Gemini to deepen your analysis. The notebook gives you four tools:

Tool	When to use it
ask_about_variable()	You encounter a column name you do not understand
explain_plot()	After making a visualization — to get an ecological interpretation
ask_for_plot()	You want Gemini to write code for a new visualization
suggest_extensions()	You want ideas for where to take the analysis next

Prompt Scaffold B — Analysis Request

My research question is: [your question here] I have already observed that: [your pre-AI observation from B2] Using the NEON data loaded above, please [specific request — e.g., explain this pattern / write code to make a plot of X vs Y / suggest what is driving this]. Flag any data quality issues I should check before trusting the result.

B3.1 When Gemini Writes Code: Read Before You Run

If you use ask_for_plot() and Gemini generates Python code, do NOT click run immediately. First read through it and check:

☐	Does it use flux['NIWO'], fxpr, or other variable names already defined in the notebook?
☐	Does it include axis labels and a title that make scientific sense?
☐	Does it handle missing data (look for dropna() or similar)?
☐	Is there any line you do not understand? If so, ask Gemini to explain that one line.
☐	After running: does the plot look plausible? Compare it to the .describe() statistics you already know.

B3.2 Peer Discussion Checkpoint

Share your best plot with a partner. Discuss these questions together and record your answers:

14. Does Gemini's ecological explanation match what you observed before using AI?

15. Is there anything in Gemini's response that does not seem right to you? What would you push back on?

16. What would you need to look up or verify to be more confident in the interpretation?

B4. Extend Your Analysis

Use suggest_extensions() to get three ideas for next analyses, then choose one to implement.

The extension I chose and why:

What the extended analysis showed:

Part B Deliverables

- Completed B2 pre-AI observation (all four questions answered before any B3 AI use).
- Prompt log with at least 3 Gemini interactions from Part B (prompt + summary of response + what you did with it).
- At least two plots with captions written in your own words (not pasted from Gemini).
- Written answer to your research question in 3–5 sentences, based on the data.

PART C: Synthesis and Communication

In Part C you bring your findings together and share them with a real audience. You also reflect honestly on how AI helped you and where your own judgment was essential.

C1. Choose Your Communication Format

Select the format that your instructor has assigned, or choose one with their approval. Write your chosen format and audience here:

Format	Who you are writing for	Core requirement
Scientific poster	Other students / conference	One finding per panel, data evidence, your interpretation
One-page brief	A forest manager or park ranger	Translate findings into plain-language implications
Infographic	General public	Explain carbon cycling visually, no jargon
Annotated notebook	Future students	Clean code + your explanations + honest AI reflection
3-minute audio explainer	Podcast listeners	Accurate science, story-driven, cite your data

C2. Required Elements for Every Format

Whatever format you choose, your product must include all four of these:

17. **Environmental claim:** Your central finding stated clearly: Is the ecosystem a sink or source? What drives the pattern you found?
18. **Data evidence:** At least one data visualization with a caption you wrote yourself.
19. **AI transparency:** A brief explanation of how you used Gemini, what it helped with, and at least one place where you had to check or correct it.
20. **Uncertainty:** One honest limitation: what your analysis cannot tell you, or a question the data cannot answer.

C3. Reflection Questions

Answer each of these in a few sentences. This is submitted alongside your communication product.

21. What was the most useful thing Gemini did in your analysis? What would have taken you longer without it?

22. Describe a moment when Gemini's output was wrong or needed fixing. How did you catch it?

23. Look at your prompt log from Part A to Part B. How did your prompts change? What did you learn about effective prompting?

24. Where in this module did your own scientific reasoning matter most — where could Gemini NOT replace your thinking?

25. If a friend said "I can just use AI to analyze environmental data without knowing the science," what would you tell them?

Part C Deliverables

- Your communication product (meeting all four required elements in C2).
- Your written reflections for all five C3 questions.
- Your complete prompt log from all three parts.
- Your completed Energy and Carbon Tracker from the Extension section.

Extension: How Much Energy Did Your AI Use?

Every time you send a prompt to Gemini, it uses energy. Every time you run a code cell in Colab, your computer and Google's servers use energy. This section asks you to estimate — not precisely measure — the energy cost of your work in this module.

Why estimate rather than measure? Because energy use in cloud computing is genuinely hard to measure exactly. The honest scientific answer is: we use what data we have, state our assumptions clearly, and acknowledge the uncertainty. That is exactly what you have been doing with the NEON flux data.

E1. Start and End Times

Record the times you spent actively running the notebook in Colab. You can find elapsed time in Colab by looking at how long each cell took, or by checking your session start time in the Runtime menu.

Session	Approx. runtime (minutes)	CPU or GPU?
Part A		*CPU / GPU (circle one)*
Part B		*CPU / GPU (circle one)*
Part C		*CPU / GPU (circle one)*
TOTAL		

If you are not sure whether you used CPU or GPU: in Google Colab, go to Runtime → Change runtime type. The default is CPU. GPU is only enabled if you specifically chose it.

E2. Prompt Log

Record every Gemini prompt you sent during the module in the table below. Mark each as Simple (short question, text only) or Complex (long prompt, multi-part, or pasting in large data).

Prompt #	Simple or complex?	What you asked about
1		
2		
3		
4		
5		

6		
7		
8		

Total number of prompts: ______ Simple: ______ Complex: ______

E3. Calculate Your Energy Estimate

Use the formulas below. These are teaching estimates based on publicly available data from Google and computer science research — they are not exact measurements. You will reflect on that uncertainty in E4.

Where these numbers come from
0.24 Wh per Gemini text prompt: Google reported this as the median energy use for a Gemini Apps text prompt. Simple prompts use less; longer multimodal prompts can use more. **0.003 Wh per complex prompt surcharge:** A rough additional estimate for prompts that are significantly longer or more computationally demanding. **Colab CPU ≈ 25 W:** A classroom approximation for a typical cloud CPU session. Real power draw varies by utilization. **Colab GPU ≈ 100 W:** A classroom approximation for a GPU session. Real GPU sessions vary widely (60–150+ W).

Step 1: Gemini energy

Formula	Your calculation
`Gemini Wh = (total prompts × 0.24 Wh) + (complex prompts × 0.003 Wh)`	Gemini Wh = ______

Step 2: Colab energy

Formula	Your calculation
`Colab Wh = (total runtime hours) × (power in watts)` *Use 25 W for CPU | 100 W for GPU* *Reminder: convert minutes to hours by ÷ 60*	Colab Wh = ______

Step 3: Total and conversion

Formula	Your result

`Total Wh = Gemini Wh + Colab Wh`
`Total kWh = Total Wh ÷ 1000`
CO_2 `(grams) ≈ Total kWh × 400`
(using ~400 g CO_2 per kWh, a US average estimate)
	Total Wh = ______
	Total kWh = ______
	CO_2 ≈ ______ grams

E4. Put Your Numbers in Context

These comparisons use rough but commonly cited figures. Your instructor may provide more precise values.

Activity	Approximate energy
LED light bulb on for 1 hour	10 Wh
Laptop running for 1 hour	30–60 Wh
Charging a smartphone fully	10–15 Wh
Streaming video for 1 hour	36 Wh
Driving a car 1 mile	~500 Wh (gasoline equivalent)
Boiling a kettle once	~100 Wh

How does your total energy use compare to these everyday activities?

If the Niwot Ridge forest absorbs carbon at its average rate (~–1.55 μmol CO_2 m^{-2} s^{-1}), approximately how many square meters of forest for one hour would offset the CO_2 from your Gemini prompts? (This is a rough estimation exercise — show your work.)

E5. Reflection on Uncertainty and Workflow

Honest scientific practice means being transparent about the limits of your estimates.

26. List two assumptions built into your energy estimate. Which one introduces the most uncertainty, and why?

27. Compare your prompt count with a partner's. Who used more prompts? What does that tell you about different working styles?

28. If you did this module again, what would you do differently to use AI more efficiently? Would fewer, better-designed prompts be more energy-efficient AND more scientifically useful?

29. The Niwot Ridge forest absorbed CO_2 all through June and July 2018. Your AI use released a small amount of CO_2. Does this comparison change how you think about AI tools in environmental science work? Why or why not?

AI Use Guide

This module encourages you to use Gemini as a learning tool — not as a shortcut. Here is what that means in practice.

Encouraged	Not permitted
Asking Gemini to explain what a variable name means	Asking Gemini to write your reflection or answer a conceptual question for you
Using Gemini to generate plot code, then reading it carefully before running	Submitting Gemini's interpretation word-for-word as your own caption or explanation
Asking Gemini to interpret a plot's ecological meaning, then checking if you agree	Running AI-generated code without reading it first
Using Gemini to suggest next analysis steps	Using Gemini to decide your research question or draw your scientific conclusions
Asking Gemini to explain a Python error message	Ignoring data quality flags because Gemini didn't mention them

Your prompt log is a required part of your submission. It should record every Gemini interaction: (1) your full prompt, (2) a brief summary of the response, and (3) what you did with or changed about the output.

Assessment Rubric

Your work is evaluated across four equal dimensions. Your prompt log is required evidence for the AI Literacy dimension.

Dimension	Excellent (4)	Proficient (3)	Developing (2)	Beginning (1)
Environmental Understanding	Accurately explains carbon flux, ecosystem metabolism, and Niwot Ridge ecology using correct units.	Mostly accurate with minor gaps.	Basic understanding, some errors.	Significant gaps.
Data Literacy	Correctly interprets patterns, applies quality flags, draws justified conclusions.	Mostly correct with minor gaps.	Some correct interpretation; limited quality flag use.	Conclusions not justified by data.
AI Literacy	Evaluates all AI outputs critically; complete prompt log; clearly explains AI vs. own reasoning.	Evaluates most outputs; log mostly complete.	Some evaluation; limited documentation.	Accepts AI uncritically; no log.
Communication	Clear, accurate, well-supported for the target audience.	Mostly clear with minor gaps.	Partially clear; some unsupported claims.	Unclear or unsupported.

Instructor Guide (remove from student version)

Anticipated difficulties at the 100 level

Common issues

- **Students do not know what a DataFrame is:** The notebook has a comment explaining this. If students are confused by flux['NIWO'].head, tell them to think of it as a spreadsheet where each row is a 30-minute window.
- **Gemini gives generic answers:** Almost always caused by forgetting to pass NEON_CONTEXT to the function. Check that the function call includes context=NEON_CONTEXT.
- **Students cannot find qfFinl column:** Run flux['NIWO'].filter(like='qf').columns to show all quality flag columns. The final quality flag may be named differently across NEON product versions.
- **Energy tracker numbers seem very small:** This is the correct result — individual AI prompts use small amounts of energy. The teaching point is that scale matters: millions of prompts per day is a different story.
- **Harvard Forest PAR data:** prtop_harv uses verticalPosition=='040'. If no data returns, run par30_harv['verticalPosition'].unique() to find the correct position label for HARV.

Discussion questions for class debrief

30. You had to tell Gemini the site name, the units, and the statistics before it gave a good answer. What does that tell you about what AI actually ‘knows’?
31. Gemini said the forest was a carbon sink. You checked with the data. Were the numbers what you expected? What would change your conclusion?
32. Your energy estimate had a lot of assumptions in it. How is that different from the NEON flux measurements? Are the NEON measurements assumption-free?
33. NIWO and HARV are in very different climates. Did the data support your prediction from A2.1? What surprised you?

Implementation options

Format	Recommendation
Single 3-hr lab	Part A as pre-lab reading; run setup together; B and C in class
Two 75-min classes	Day 1: A1–A4; Day 2: B + C + energy tracker
Homework over a week	Part A day 1–2, Part B days 3–5, Part C + extension days 6–7
No-code version	Pre-run all cells; students analyze printed output and use explain_plot() only
Shorter version (60 min)	A1–A4 + one B3 analysis + energy tracker only