The Stomata Lab

What the past can tell us about our future – using fossil and modern plants to model atmospheric carbon dioxide.

**Objective:** Students will develop a mathematical model of the relationship between atmospheric CO2 and the number of stomata on a leaf (Stomata Index). They will evaluate the model graphically, statistically, and biologically, and then use it to estimate CO2 levels in the distant past.

**Background**

One component in preparing for our future climate is to use the fossil record to study how life on Earth behaved in the past to changing atmospheric carbon dioxide (CO2) levels and the corresponding increase in global temperature. Estimating CO2 levels in the geologic record is challenging but possible. Scientists can use bubbles in ice cores to directly measure gasses in the atmosphere on a timescale of 100s to 100,000s of years ago. To estimate CO2 levels in atmospheres older than 800,000 years ago, however, scientists must use indirect means. Someone can determine that a beaver cut down a tree by observing the shape of the stump left behind. Similarly, scientists can determine ancient atmospheric composition by observing geologic, chemical, and biological proxies.

In this lab, atmospheric CO2 levels 56 million years ago (mya) are investigated. This represents the most recent geologic period to experience a rapid rise in global temperature. Therefore, this geologic period has much to teach about how life on Earth might respond to rapid changes in temperature. Morphology and underlying physiology of plants are used as a proxy to estimate the ancient atmospheric CO2 level.

First, a mathematical model of the relationship between the number of stomata on a leaf and atmospheric CO2 levels today and recent history will be built. Next, experimental data will be incorporated to improve the model. Finally, the improved model will be used to estimate the level of CO2 in the fossil record.

Stomata are a key adaptation in land plants that allow them to avoid desiccation while maintaining gas exchange. CO2 diffuses from the atmosphere through the stomata and into plant tissues. CO2 is used by the plant in the Calvin Cycle (dark reactions) of photosynthesis to build glucose (C6H12O6). However, when stomata are open to let CO2 in, water escapes into the atmosphere at an excessive rate. Think of a plant having to “pay” for carbon with water molecules, and carbon is expensive! When the atmosphere has a high concentration of CO2, the concentration gradient of CO2 between the atmosphere and plant tissues is greater and diffusion rate increases. Therefore, it is easier for plants to obtain CO2 for photosynthesis. Therefore, when atmospheric CO2 concentration is high, fewer stomata are necessary; plants obtain the CO2 they need while minimizing water loss.

To collect data about stomata, Scanning Electron Microscope (SEM) images of the inside surface of the lower epidermis of *Ginkgo biloba* leaves from historical herbarium and modern samples will be used. **Stomatal Index (SI)** of each sample image will be calculated to create a model between the number of stomata on a Ginkgo leaf and the level of CO2 in the atmosphere. That relationship can then be used as a proxy to understand Earth’s atmosphere in the geologic past and what is currently happening today.

The **hypothesis** to be tested is that there a relationship between SI of leaves and CO2 levels in the atmosphere. The **null hypothesis** is there is no relationship – SI is not related to CO2 levels in the atmosphere.

**Stomatal Index (SI)** is the percentage of leaf cells that are stomata in an area of a leaf. It is calculated by dividing the number of stomata in a given area of leaf surface (in this case, 300 µm x 300 µm box) by the total number of epidermal cells in that area.

$$SI=\left(\frac{\# stomata}{\# total cells}\right)\left(100\right)$$

* Given the physiology described above and considering the hypothesis, write a **prediction**.

(If the hypothesis is correct, what will the data look like – will it be a positive or negative relationship?)

Because historical herbarium and modern samples are used, the CO2 concentration of the atmosphere when the leaf sample was collected is known through direct measurements. One way gas concentrations are communicated is by **parts-per-million by volume** (**ppmV**). It may be communicated as just ppm (parts-per-million), but if it refers to a gas mixture then the volume is assumed and just left out for simplicity. This communicates the relative abundance of a gas in a sample. The atmosphere is comprised of several gasses – nitrogen (78%), oxygen (21%), and the last 1% is small amounts of CO2, neon, and hydrogen. Currently, the CO2 concentration is 419 ppmV ([nasa.gov](http://nasa.gov/)).

**Training – How to Count Stomata**

Images of leaf samples are provided for data collection. For each image, the number of stomata included in a 300 µm x 300 µm square will be counted. Then, that information will be used to calculate the **Stomatal Index (SI)**. The total number of epidermal cells in the square is provided.

(Figure 2 from Barclay & Wing, 2016)



An unaltered Scanning Electron Microscope (SEM) image taken of the surface of a *Ginkgo biloba* leaf. The stomata differ in morphology from the other epidermal cells.

On this version of the image, stomata are labeled, the 300 µm x 300 µm box is indicated, and all cells of the epidermis within the box are labeled. Any stomata within the box (even if only partially) are included in the count. This example would have a count of 8 stomata.

Each image sheet will also provide the CO2 levels for the year the sample was collected. The calculated SI and the CO2 data will be entered into an Excel template file.

**Herbarium and Modern Data**

1. Locate the *Herbarium* tab of the Excel template.
2. Collect data from the herbarium and modern leaf image sample sheets (10 images labeled 1A – 1J). Start with the first two images (1A and 1B), and check with the instructor before continuing.
3. Enter the SI and CO2 levels recorded on the data sheets.
4. The Excel template is programed to create a scatterplot with a regression line and will provide a goodness of fit statistic called the **coefficient of determination** (*R*2).

Interpretation – Thinking about Mathematical Models

* Write the equation for the regression line found on the graph:

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Notice that the equation is in the *y* = *mx* + *b* format. (Remember *m* is the slope, and *b* is the *y*-intercept).

The **hypothesis** is that there is a biological relationship between the level of atmospheric CO2 and the SI value of leaves. The regression line represents that relationship, and *R*2 is a goodness of fit statistic which tests that hypothesis.

The Excel template provides the value of *R*2. It indicates how well the line fits the data. *R*2 ranges from 0 to 1.

$R^{2}=\frac{variance explained by the model}{total variance} $,

which can also be written as:

$$R^{2}=\frac{SS\_{total}}{SS\_{error}+SS\_{total}} ,$$

(***SS*** is an abbreviation for ***sum of squares***, which is a measure of variance or deviation from the mean or expected value.)

If *R*2 = 1, the model perfectly fits the data and explains 100% of the variance. Therefore, the higher the value for R2, the greater the fit. If the *R*2 is high and the *p*-value is significant (less than 0.05), then the relationship between the two variables (*x* and *y*) is explained by the hypothesis. \*see notes Remember, *x* is the independent variable (CO2 levels) and *y* is the dependent variable (SI).

* What is the value of *R*2?

*R*2 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* What is the *p*-value (provided by instructor)? \*

*p-*value = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

The equation for the regression line is a mathematical model that expresses the relationship between SI and atmospheric CO2 levels.

* Based on *R*2 and the *p*-value, do you **reject** or **fail to reject** the null hypothesis *(circle one)*

reject null hypothesis fail to reject null hypothesis

* Based on *R*2 and the *p*-value, do you **support** or **not support** the alternative hypothesis? *(circle one)*

support alternative hypothesis do not support alternative hypothesis

* Ponder the line, and describe what it represents in biological terms. In other words, what is the relationship between SI and atmospheric CO2 levels?
* Follow the line down to the *x*-axis. At what CO2 level does the line intercept the x-axis?
	+ What does this mean biologically?

Throughout the ancient past, CO2 levels have naturally fluctuated. During the Cambrian Period, 500 mya, most of life was contained in the oceans. Early in the process of life moving onto land, CO2 levels in the atmosphere is estimated to have been as high as 3,000 ppmV! As of 2022, CO2 levels is 419 ppmV, but pre-industrial levels (before 1850) were 280 ppmV.

* All models have limitations. Based on this context and the biological interpretation of the *x*-intercept, is this linear model based on historical (100+ years ago) and modern data a good model to use in order to estimate CO2 levels in the geologic past or for our own future? Explain your reasoning demonstrating your understanding of the model and the biology.

**Experimental Data**

Scientists at the Smithsonian’s Environmental Research Center (SERC) in Edgewater, Maryland set up an experiment to better understand how big changes in Earth’s atmosphere and climate in the geologic past affect plant physiology, which predict future conditions. The experiment investigates how modern Ginkgo tree growth is affected by increasing levels of CO2. Growth chambers were built with CO2 levels at 400 ppmV (modern ambient conditions) and at 600 and 800 ppmV (<https://fossilatmospheres.wixsite.com/fossilatmospheres>). Ginkgo trees are used because they are considered a “living fossil. “Only one species is living today, but the genus evolved during the Triassic Period, approximately 250 mya, before dinosaurs. Morphologically, the ancient Ginkgo leaves and modern Ginkgo leaves appear to be almost identical.

Exploring how modern Gingko SI changes with CO2 levels can establish a better understanding of the atmosphere in the geologic past and allow for better preparation for the current climate trajectory.

Experimental data will now be included in the model to determine if that changes the relationship between atmospheric CO2 and the SI.

1. Locate the *Experimental* tab in the Excel template. The data from the *Herbarium* tab(1A –1J) are already present.
2. Collect data from the experimental leaf image sample sheets (8 images labeled 2A – 2H).
3. Enter the SI and the CO2 levels recorded on the data sheets.
4. The Excel template is programed to create a scatterplot with a regression line and will provide a value for *R*2.

Interpretation – Thinking about Mathematical Models

The regression line changed; it is now a curve. A curve is necessary to fit the new data.

* What is the value of *R*2?

*R*2 = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

* What is the *p*-value (provided by instructor)? \*

*p-*value = \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Evaluating the New Model

* Based on *R*2 and the *p*-value, do you **reject** or **fail to reject** the null hypothesis *(circle one)*

reject null hypothesis fail to reject null hypothesis

* Based on *R*2 and the *p*-value, do you **support** or **not support** the alternative hypothesis? *(circle one)*

support alternative hypothesis do not support alternative hypothesis

* Look at the old *x*-intercept for the Herbarium data. What happens in the new model at that level of CO2?
* Is this a more accurate model? Explain and incorporate the biology.

**Using the Model – Fossil Data**

Paleontologists observed a dramatic change in communities of organisms (plants and animals) in the fossil record as the Paleocene ended and the Eocene began, around 56 mya. Scientists wanted to understand what was the cause of the pattern they were seeing in the fossils. This was a transformative period in the Earth’s climate and is now identified as the Paleocene-Eocene Thermal Maximum (PETM). It is the most recent time the planet had a quickly warming climate. During this geological period, the climate warmed more rapidly than at any other time in Earth’s history (except now). The temperature increased by about 5 to 8°C (9 to 14°F) over only a few thousand years (10,000 years maximum).

The PETM was a volatile time for life on Earth. The fossil record suggests that the ranges of plants and animals shifted toward the poles. While some groups declined or went extinct (especially in the oceans), other groups survived and even experienced origination (new species evolving). The first primates evolved during this time, and by the end of the PETM, all modern orders of mammals had evolved.

These characteristics of the PETM has drawn understandable attention from scientists as an example of how our planet has responded to rapid warming in its past, helping to prepare for the future. One of the big questions is: **what was the atmospheric CO2 level during the PETM**? Fossil leaves may hold the answer.

Now, data will be collected from images of fossil Ginkgo leaves that are 56 million years old. The model will then be used to estimate the CO2 levels during the PETM.

1. Locate the *Fossil* tab of the Excel template. The graph and regression from the *Experimental* tab have been copied here.
2. Collect data from the fossil leaf image sample sheets (4 images labeled 3A – 3D).
3. Enter SI you calculated from 3A – 3D.
4. To estimate the CO2 levels in the earth’s atmosphere 56 mya, drag the orange arrows provided to indicate where on the curve CO2 levels should be based on the SI. Then, look at the *x*-axis and estimate what the atmospheric CO2 levels might have been 56 mya.
* What range of CO2 levels did your fossil data indicate?

**Explore More**

* Fossil atmospheres

<https://www.si.edu/fossil-atmospheres/about>

<https://fossilatmospheres.wixsite.com/fossilatmospheres>

* Articles on the PETM
	+ Smithsonian Magazine

<https://www.smithsonianmag.com/science-nature/ancient-earth-warmed-dramatically-after-one-two-carbon-punch-180953610/#:~:text=Many%20researchers%20think%20the%20Paleocene,led%20to%20spiraling%20gloal%20temperatures>

* + Washington Post

<https://www.washingtonpost.com/news/speaking-of-science/wp/2018/03/27/this-ancient-climate-catastrophe-is-our-best-clue-about-earths-future/>

* Barclay, R. S., & Wing, S. L. (2016). Improving the Ginkgo CO2 barometer: Implications for the early Cenozoic atmosphere. *Earth & Planetary Science Letters*, 439, 158–171.

**Notes**

For future statistical knowledge: R2 is the coefficient of determination for a regression. It tells you how well your regression explains the variation in your data. If it explained 100% of the variation all your data points would be directly on the line of regression.

$$R^{2}=\frac{signal}{noise}=\frac{variance explained by the model}{total variance}= \frac{SS\_{total}}{SS\_{error}+SS\_{total}}$$

This can also be expressed as the following (and will give the same answer):

$$R^{2}=1-\frac{SS\_{error}}{SS\_{total}}$$

$SS$ = Sum of Squares

$SS\_{total}$ = Signal, difference between fitted points on the regression line and the mean of the data

$$\sum\_{}^{}\left(y\_{i} - \overbar{y}\right)^{2}$$

$SS\_{error}$ = Noise, difference between each data point and the regression line.

$$\sum\_{}^{}\left(y\_{i} - f\_{i}\right)^{2}$$

$f\_{i}$ = the $y$ value of the expected data points, if the data were to perfectly fit a straight line. This is often represented by $\hat{y}$. (We use $f$ because there are too many *y*’s!)

However, often in regression analyses you will see the statistic *r* which is the correlation coefficient. The correlation coefficient (*r*) is simply the square root of the coefficient of determination (*R*2).

$$r= √R^{2}$$

Generally, if you have a *r* value of 0.7 or greater (r ≥ 0.7) your results are significant (your *p*-value will be less than 0.05), and the relationship between your two variables (*x* and *x* and *y*) is real (and explained by your hypothesis… hopefully…).

ANOVA - an ANOVA is an analysis of variance when there are multiple independent factors.