We thank the reviewers for their helpful reviews; the manuscript is stronger because of them. The comments are below, followed (in bold) our responses. All line numbers by us reference the marked-up copy of the revised manuscript.

REVIEWER #1

General comments: Fossil leaf gas-exchange based CO2 models are currently going through the “rigorous testing” phase and as the authors of this paper point out, this mechanistically, rather than empirically calibrated proxy, shows considerable promise. It is therefore of high relevance that studies, such as this one, are presented that provide quantification of potential confounding factors. In this case, the authors test three potential confounding factors (photorespiration, leaf temperature and canopy position) and provide quantifications on how these factors influence final CO2 estimates. They are capable of eliminating two of these factors as insignificantly affecting CO2 estimates (photorespiration and leaf temperature). The third factor, canopy position, is determined to strongly skew CO2 estimates, but the authors point out that it is possible to identify leaves that grew in lower canopy positions, based on leaf micromorphology and an uncharacteristically wide δ13C range. This paper is a relevant contribution towards quantification of the potential error in fossil leaf gas-exchange based CO2 models, and apart from minor suggested amendments, I have no problem with seeing this study being published.

Specific comments: In the materials and methods section, the authors lay out the specific ways that they are testing modern plants for potential bias in reconstructed CO2. In the appendix all the specific plants are listed with their input values and reconstructed CO2. However, from reading the methods section I get the impression that not each plant is being tested for the same potential confounding variable (photorespiration, leaf temperature and canopy position). It would be very helpful if there was a table that outlines specifically which plants were tested for what, or at least that this was made clear in the appendix, because in the main body of text it is hard to follow.

We now include this information in column E of the supplemental table.

In several places in the manuscript, including the abstract, it is mentioned that the random error propagation of the Franks et al. gas exchange model is better than uncertainty estimates of other leading paleo-CO2 proxies. It would be very helpful for the untrained reader to see some proof of this statement in the form of a table that lists 1) the different CO2 proxies, 2) a method of error quantification, 3) the actual amount of uncertainty in those CO2 proxies and 4) the references to the case studies where this was tested. Such a table would lend credibility to the statement that gas-exchange models are quantifiably better than other CO2 proxies.

There are of course two elements of uncertainty: precision (spread of possible solutions) and accuracy (comparison to true answer; can only be quantified for times when CO2 has been measured). The abstract brings up the theme of accuracy (28% mean error rate). In the main text (section 3.1), the mean error rate is compared generally to that in other CO2 proxies by referencing the summary work of Franks et al. (2014).

The error propagation scheme noted by the reviewer is related to precision. We only mention precision in the Introduction by referencing what others have found (Franks et al., 2014). It is not a focal point of the current study.
The reviewer may (also) be referencing the paragraph in the Introduction where we argue that studies using other stomatal-based proxies probably overstate the accuracy and precision of their CO2 estimates (lines 98-106). Our arguments here are conceptual only—there are no data we can summarize in a table, unfortunately. The point we are trying to make is that the reported accuracies and precisions associated with these other methods—when applied to plants living today (not fossils)—are better than what we find with gas-exchange methods. But this is partly because these other methods are based on empirical calibrations with...present-day plants. So excellent accuracies and precisions are not particularly surprising. But when you apply these other methods to fossils that are millions of years old, the present-day empirical calibrations are likely less appropriate.

Final specific comment is on the title itself, for which I would like to suggest that the authors include what specifically is being tested. I.e. “Sensitivity of . . .. CO2 concentration to x, y & z”. There are other variables that the model is sensitive to and I believe the title would be more informative if the specifics were included.

The largest block of data (40 species) is “general” testing, that is, estimating CO2 from field-grown trees without isolating any single confounding factor (summarized in Figure 2). Thus, it would not be fully representative to say that we were only testing the model for the influence of canopy position, temperature, and photorespiration.

Technical corrections: I could not find any spelling or styling errors in the manuscript. The paper is very well constructed and easy to follow.
The authors present a sensitivity analysis of a mechanistic model (Franks model) to predict paleoatmospheric CO2. They explore several specific areas; the effect of gc(op)/gc(max), A0, temperature, photorespiration and leaf canopy position on the accuracy of CO2 estimates produced by the model. In doing so, the paper adds clarity, certainty or recommendations to the model for fossil application, all of which are important additions, especially as this model is being using in a growing number of research projects. Although the paper is an important contribution, it would benefit from clarity or expansion in certain areas:

1) Aims, methods and appendix: The aims and methods section is hard to follow. This may be due to the fact the aims and rationale are mixed in with the methods. It is unclear from the text or appendix data whether all or a subset of the data is being used for each of the analysis performed. A summarised table in the methods section containing the information on the analysis being performed, data source and parameters used or tested would be beneficial (i.e. a summary of the methods in tabular format). Similarly, in the appendix, additional information on the origin of the data, sample number per species, which data points/values are measured vs estimated/assumed and a direct comparison of measured vs model estimated CO2 would greatly improve clarity.

We now present a tabular summary of our study design (new Table 1).

In the Supplemental Table 1, we now give the sample size (column F), the target (i.e., correct) CO2 concentration (column G), and whether the input was measured or inferred (color coding of column headers). And column E gives what part of the study was addressed (general testing, temperature, or canopy position; reviewer #1 also asked for this information). We are not sure what is meant by “additional information on the origin of the data” beyond what is listed in column A and stated in the main-text Methods.

2) Statistical analysis: Accuracy was evaluated by the degree of error rate. These claims can be strengthened by using statistical analysis. How well the model predicts CO2 could be assessed by whether or not the estimates are statistically significant different (or hopefully not) from measured CO2 values.

We have added information about whether individual estimates depart from the target CO2 concentrations (lines 344-346 and 419-421).

3) gc(op)/gc(max) and A0 (section 3.1): This section gives details about when both gc(op)/gc(max) and A0 values are either known or values from Franks et al. 2014 are used, but it would be nice to see these two parameters evaluated separately i.e. how much does gc(op)/gc(max) alone improve estimates and the same for A0. Does one contribute more than the other for improving error rates?

We have added this information (lines 351-352).

Additional comments:  

We have added the qualifier “in many species”.

A Nearest living relative or equivalent approach also get around the issue of extinct taxa.

This is true for the stomatal ratio method, but these CO2 estimates are not meant to be quantitative in the same manner as estimates from the “full calibration” methods or the gas-exchange methods (as noted in the previous paragraph).

Alternative approaches for fossils have been suggested such as estimating fossil A0 using scaling relationships between vein distance and assimilation rate however they are not discussed here (EG Montanez et al., 2016).

We have added a citation to the Montanez paper

Introduction – general comment. Critical published assessments of the Franks model are not cited (eg McElwain et al. 2016) yet they raise issues associated with parametrization of A0 and the insensitivity of CO2 estimates to variation in gamma star values which are both important discussion points in this manuscript in lines 454 -456 and 497-499.

As per a later comment, we have added a citation to McElwain et al. 2016 regarding gamma star on line 466.

Our study does not focus on the parameterization of A0, and so the associated literature does not seem relevant to the Introduction. Our study focuses on temperature, photorespiration, canopy position, as well as a general and broad test of the method.

Paragraph 201-217: A some information is missing here: chamber model/make, duration plants were grown in the chamber, light levels. What were measured vs set chamber conditions for temperature, light and CO2 (i.e. similar to how humidity is reported)

Chamber make/model (lines 212-213) and duration of experiment (line 229) are given. We have added information about light intensity as well as the standard deviations for temperature and CO2 concentrations in lines 213-218.

Lines 232: Stomatal density/stomatal measurements and leaf stable carbon isotopes were performed on the same leaves. Clarify how this was partitioned, e.g. was the leaf divided into 2 or was a whole punch used for carbon isotopes, etc.?

We now clarify our methodology in lines 237-238. We used either a hole punch or razor to remove two adjacent sections of leaf tissue near the leaf centers, avoiding major veins.
As Milligan et al. is in review, I suggest adding more detail here on how δ13Ca of chamber CO2 was calculated. δ13Ca values of supplemented CO2 can be very negative and can vary between cylinders, unless the CO2 gas has a specific δ13Ca. What is the capacity of these cylinder, in L?

This paper is likely to be “in press” soon; we have appended it to the end of this file (after the marked-up copy of our manuscript). In short, a mixing line was established based on direct d13C measurements of lab air, chamber air, and cylinder CO2 (= pure CO2). We were fortunate that the d13C of the cylinder was close to the well-mixed atmosphere (the d13C in most cylinders we have used in other experiments is much more depleted). We used only the single cylinder for the duration of the experiment. The target CO2 concentration (500 ppm) was not much higher than the CO2 concentration inside the lab (~440 ppm), so we did not use much CO2.

Figure 1: Does this need to be on a log scale? 1000 or 2000ppm are not very high values and the log scale visually skews data and error bars. A difference plot between measured and estimates plotted on a non-log scale would improve this figure.

We prefer a log scale because it is easier to differentiate estimates at the low-end of the CO2 scale, and because the uncertainties scale in a logarithmic fashion.

Line 351: Please provide supporting data for this statement in tabular form. What are the error rates of other proxies?

This information was summarized by Franks et al. (2014), so we prefer not to repeat it here.

Line 355: Might be helpful to report standard deviation of CO2 estimates, here and throughout the text.

We now report the range that encompasses two-thirds of all estimates (lines 343-344). (Because the individual estimates are not normally distributed (tail at the high end), reporting a standard deviation can be misleading.)

Line 411 to 413. Reporting of the difference between estimated and measured CO2 here is incomplete. Only means of all species investigated are provided rather than species-based differences or errors. For some species the error is substantial whereas other taxa show very small errors.

As per an earlier comment, we now report the species-level differences on lines 419-421; no individual species-level test was significant (line 408).

Line 454 to 456. This supports the findings of McElwain et al. 2016 Paleo 3 but it is not cited. “This compensation point (Γ* in Eq. (2) is temperature, species and O2 dependent (Ethier and Livingston, 2004) but Franks et al. (2014) account only for the temperature dependency in the new paleo-CO2 proxy model. Allowing Γ* to vary in response to prevailing paleoatmospheric O2 concentration [O2] (Γ* = 1.78 × [O2]), which is known to have varied widely (10% to 30%) through the Phanerozoic (Bergman et al., 2004; Belcher and McElwain, 2008; Berner, 2009), would increase the precision of paleo-CO2 estimates but only fractionally.”

We have added a citation to McElwain et al. (2016 Palaeo3) (line 466).
Lines 500 to 506: A number of papers have suggested methods of estimating A0 to improve the accuracy of CO2 estimates using the Franks model but they are not discussed. This section would provide a good opportunity to discuss the proposed ideas and solutions.

This section deals with living leaves, where A could be measured directly. Measuring A wasn’t part of our study design, unfortunately. In this section we are discussing possible reasons for noise in our mixing-model calculations. With regards to fossils, we are not recommending that our mixing model be used (line 520: “We note that our mixing-model strategy cannot be applied to fossils because…”), so the question of how to constrain A in fossils within the context of the mixing model is moot. Our take-home message for fossil applications is to avoid shade leaves (line 528), and we provide specific measurements that can be made on fossils to make this distinction, including vein density (lines 529-533).

Section 3.4: Have any values for δ13Ca been measured or are all calculated for this section? Is there any data set (from the literature or otherwise) this could be compared to? i.e. a dataset where known δ13Ca is compared to itself when calculated as per the manuscript? This would strengthen this section. If δ13Ca has only been calculated/inferred for this section without a comparison to measured δ13Ca I think claims on the effect of δ13Ca (or low canopy plants) on the model should be softened.

We made no direct measurements of understory d13Ca (multiple measurements over a growing season, and at different daytime hours, would be needed to calculate a representative mean value). As the reviewer correctly notes, we instead are assuming a well-behaved two end-member mixing model. We have added a note of caution related to this on lines 502-505.

Appendix: The authors used both known and general values for gc(op)/gc(max) and A0 to evaluate error rates but no measured values of either gc(op)/gc(max) or A0 are given in the appendix or text.

The Appendix summarizes all new data presented in the study (with the key graphics being Figures 2, 5, and 7). For these data, we *only* used “default” values of gop/gmax and Ao; that is, we did not measure these inputs on our leaves. As noted in the Introduction, this was a purposeful strategy because we wanted to test the CO2 model in a manner that would be similar to how most (but not all) folks will be applying the model to fossils. A “worst-case” test, if you will.

In the Introduction, we do summarize some of the already-published data (Figure 1). For these estimates, either gop/gmax or A0 were measured, and in most cases both were measured (lines 142-145). These data are not in the Appendix because they are already published and are not central to our study.

As the reviewer noted, we did additionally “degrade” these estimates by re-doing them assuming default values for gop/gmax and A0. We did this so we could compare them more directly to our estimates (lines 349-351).
Sensitivity of a leaf gas-exchange model for estimating paleoatmospheric CO₂ concentration

Dana L. Royer¹, Kylen M. Moynihan¹, Melissa L. McKee¹, Liliana Londoño², and Peter J. Franks³

¹Department of Earth and Environmental Sciences, Wesleyan University, Middletown, Connecticut, USA
²Smithsonian Tropical Research Institute, Balboa, Ancón, Republic of Panamá
³Faculty of Agriculture and Environment, University of Sydney, Sydney, New South Wales, Australia

Correspondence: Dana L. Royer (droyer@wesleyan.edu)

Abstract. Leaf gas-exchange models show considerable promise as paleo-CO₂ proxies. They are largely mechanistic in nature, provide well-constrained estimates even when CO₂ is high, and can be applied to most subaerial, stomata-bearing leaves from C₃ taxa, regardless of age or taxonomy. Here we place additional observational and theoretical constraints on one of these models, the “Franks” model. In order to gauge the model’s general accuracy in a way that is appropriate for fossil studies, we estimated CO₂ from 40 species of extant angiosperms, conifers, and ferns based only on measurements that can be made directly from fossils (leaf δ¹³C and stomatal density and size) and a limited sample size (1-3 leaves per species). The mean error rate is 28%, which is similar to or better than the accuracy of other leading paleo-CO₂ proxies. We find that leaf temperature and photorespiration do not strongly affect estimated CO₂, although more work is warranted on the possible influence of O₂ concentration on photorespiration. Leaves from the lowermost 1-2 m of closed-canopy forests should not be used because the local air δ¹³C value is lower than the global well-mixed value. Such leaves are not common in the fossil record, but can be identified by morphological and isotopic means.

1 Introduction

Leaves on terrestrial plants are well poised to record information about the concentration of atmospheric CO₂. They are in direct contact with the atmosphere and have large surface-area-to-volume ratios, so the leaf internal CO₂ concentration is tightly coupled to atmospheric CO₂ concentration. Also, leaves are specifically built for the purpose of fixing atmospheric carbon into structural tissue, and face constant selection pressure to optimize their carbon uptake relative to water loss. As a result, many components of the leaf system are sensitive to atmospheric CO₂, and these components feedback on one another to reach a new equilibrium when atmospheric CO₂ changes. In terms of carbon assimilation, Farquhar and Sharkey (1982) modeled this system in its simplest form as:

\[ A_n = \frac{A_n}{g_{c(tot)}} \times (c_a - c_i), \]  

where \( A_n \) is the leaf CO₂ assimilation rate (μmol m⁻² s⁻¹), \( g_{c(tot)} \) is the total operational conductance to CO₂ diffusion from the atmosphere to site of photosynthesis (mol m⁻² s⁻¹), \( c_a \) is atmospheric CO₂ concentration (μmol mol⁻¹ or ppm), and \( c_i \) is leaf intercellular CO₂ concentration (μmol mol⁻¹ or ppm) (see also Von Caemmerer, 2000).

Rearranging Eq. (1) for atmospheric CO₂ yields:

\[ c_a = \frac{A_n}{g_{c(tot)} \times (1 - \frac{c_i}{c_a})}. \]
Equation (2) forms the basis of two leaf gas-exchange approaches for estimating paleo-CO2 from fossils (Konrad et al., 2008, 2017; Franks et al., 2014). In the Franks model, conductance is estimated in part from measurements of stomatal size and density, \( c_i/c_a \) from measurements of leaf \( ^{13}C \) along with reconstructions of coeval air \( ^{13}C \) (see also Eq. 9), and \( A_n \) from knowledge of living relatives and its dependency on \( c_o \) (Franks et al., 2014). Following Farquhar et al. (1980), the latter is modeled as (Franks et al., 2014; Kowalczyk et al., 2018):

\[
A_n = A_0 \frac{[\frac{c_i}{c_a} = \Gamma^*][\frac{c_{i_{0}}}{c_{a_{0}}}] + 2\Gamma^*]}{[\frac{c_{i_{0}}}{c_{a_{0}}}][\frac{c_{i_{0}}}{c_{a_{0}}} - \Gamma^*]},
\]

(3)

where \( \Gamma^* \) is the CO2 compensation point in the absence of dark respiration (ppm) and the subscript “0” refers to conditions at a known CO2 concentration (typically present-day). Equations (2) and (3) are then solved iteratively until the solution for \( c_o \) converges.

These gas-exchange approaches grew out of a group of paleo-CO2 proxies based on the CO2 sensitivity of stomatal density \( D \) or the similar metric stomatal index (Woodward, 1987; Royer, 2001). Here, the \( D-c_o \) sensitivity is calibrated in an extant species, allowing paleo-CO2 inference from the same (or very similar) fossil species. These empirical relationships typically follow a power-law function (Wynn, 2003; Franks et al., 2014; Konrad et al., 2017):

\[
c_o = \frac{1}{kD^\alpha},
\]

(4)

where \( k \) and \( \alpha \) are species-specific constants.

The related stomatal ratio proxy is simplified: \( D \) is measured in an extant species \( (D_0, \text{at present-day } c_{a0}) \) and then the ratio of \( D_0 \) to \( D \) in a related fossil species is assumed to be linearly related to the ratio of paleo-\( c_o \) to present-day \( c_{a0} \) (Chaloner and McElwain, 1997; McElwain, 1998):

\[
\frac{c_o}{c_{a0}} = k \frac{D_0}{D}.
\]

(5)

Equation (5) can be rearranged to match Eq. (4) but with \( \alpha \) fixed at 1. Thus, paleo-CO2 estimates using the stomatal ratio proxy are based on a one-point calibration and an assumption that \( \alpha = 1 \); observations do not always support this assumption (e.g., \( \alpha = 0.43 \) for \textit{Ginkgo biloba}; Barclay and Wing, 2016). The scalar \( k \) was originally set at 2 for Paleozoic and Mesozoic reconstructions so that paleo-CO2 estimates during the Carboniferous matched that from long-term carbon cycle models (Chaloner and McElwain, 1997). For younger reconstructions, \( k \) is probably closer to 1 (by definition, \( k = 1 \) for present-day plants). We note that the stomatal ratio proxy was originally conceived as providing qualitative information, only, about paleo-CO2 (McElwain and Chaloner, 1995, 1996; Chaloner and McElwain, 1997; McElwain, 1998) and has not been tested with dated herbaria materials or with CO2 manipulation experiments.

At high CO2, the \( D-c_o \) sensitivity saturates in many species, leading to uncertain paleo-CO2 estimates, often with unbounded upper limits (e.g., Smith et al., 2010; Doria et al., 2011). Stomatal density does not respond to CO2 in all species (Woodward and Kelly, 1995; Royer, 2001), and because \( D-c_o \) relationships can be species-specific (that is, different species in the same genus with different responses; Beerling, 2005; Haworth et al., 2010), only fossil taxa that are still alive today should be used. The gas-exchange proxies partly address these limitations: 1) CO2 estimates remain well-bounded—even at high CO2—and their precision is similar to or better than other leading paleo-CO2 proxies (~+35/-25% at 95% confidence; Franks et al., 2014); 2) the models are mostly mechanistic; that is, they are explicitly
driven by plant physiological principles, not just empirical relationships measured on living plants; 3) because the models retain sensitivity at high CO2 and do not require that a fossil species still be alive today, much of the paleobotanical record is open for CO2 inference, regardless of age or taxonomy; and 4) because the models are based on multiple inputs linked by feedbacks, they can still perform adequately even if one or more of the inputs in a particular taxon is not sensitive to CO2, for example stomatal density (Milligan et al., in review).

We note that the published uncertainties (= precision) associated with the stomatal density proxies are probably too small because they usually only reflect uncertainty in the calibration regression or in the measured values of fossil stomatal density, but not both; when this is done, errors often exceed ±30% at 95% confidence (Beerling et al., 2009). Also, error rates in estimates from extant taxa where CO2 is known (= accuracy) are usually smaller with the stomatal density proxies (e.g., Barclay and Wing, 2016), but this is expected because the same taxa have been calibrated in present-day (or near present-day) conditions. Because the gas-exchange proxies are largely built from physiological principles, they have less “recency” bias; that is, the gas-exchange proxies estimate present-day and paleo-CO2 with similar certainty when the same methods are used to determine the inputs.

2 Study Aims and Methods

Leaf gas-exchange proxies for paleo-CO2 are becoming popular (Konrad et al., 2008, 2017; Grein et al., 2011a, 2011b, 2013; Erdei et al., 2012; Roth-Nebelsick et al., 2012, 2014; Franks et al., 2014; Maxbauer et al., 2014; Montañez et al., 2016; Reichgelt et al., 2016; Tesfamichael et al., 2017; Kowalczyk et al., 2018; Lei et al., 2018; Londoño et al., 2018; Richey et al., 2018; Milligan et al., in review). However, many elements of these models remain understudied. Here we investigate four such elements for the Franks et al. (2014) model: how does the model perform across a large number of phylogenetically diverse taxa; and how is the model affected by temperature, photorespiration, and proximity to the forest floor? We describe next the motivation and details of the study design (see also Table 1 for summary).

Table 1. Summary of data sets.

<table>
<thead>
<tr>
<th>Factor tested</th>
<th>Number of species</th>
<th>Methods section</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>General testing in a phylogenetically diverse set of species and with a minimal number of leaves measured per species</td>
<td>40</td>
<td>2.1</td>
<td>Leaves come from Panama (published by Londoño et al., 2018), Connecticut, and Puerto Rico</td>
</tr>
<tr>
<td>Temperature</td>
<td>6</td>
<td>2.2</td>
<td>Theoretical calculations and growth chamber experiment</td>
</tr>
<tr>
<td>Photorespiration</td>
<td>NA</td>
<td>2.3</td>
<td>Theoretical calculations</td>
</tr>
<tr>
<td>Canopy position</td>
<td>6</td>
<td>2.4</td>
<td>Leaves come from Panama and Connecticut</td>
</tr>
</tbody>
</table>

2.1 General testing in living plants

Franks et al. (2014) tested the model on four species of field-grown trees (three gymnosperms and one angiosperm) and one conifer grown in chambers at 480 and 1270 ppm CO2. The average error rate (absolute value of estimated CO2 minus measured CO2, divided by measured CO2) was 5%. Follow-up work with three field-grown tree species (Maxbauer et al., 2014; Kowalczyk et al., 2018), CO2
experiments on seven tropical trees species (Londoño et al., 2018), and experiments on two fern and one conifer species (Milligan et al., in review) indicate somewhat higher error rates (Fig. 1). Combined, the average error rate is 20% (median = 13%).

**Figure 1.** Published CO₂ estimates using the Franks model for extant plants where the physiological inputs $A₀$ (assimilation rate at a known CO₂ concentration) and/or $g_{c(op)}/g_{c(max)}$ (ratio of operational to maximum leaf conductance to CO₂) were measured directly. Horizontal lines are the correct CO₂ concentrations. Uncertainties in the estimates correspond to the 16th-84th percentile range. Circles are from Londoño et al. (2018), squares from Milligan et al. (in review), large triangle from Maxbauer et al. (2014), small triangles from Kowalczyk et al. (2018), and diamonds from Franks et al. (2014).

In these studies, two of the key physiological inputs were measured directly with an infrared gas analyzer: the assimilation rate at a known CO₂ concentration ($A₀$) and/or the ratio of operational to maximum stomatal conductance to CO₂ ($g_{c(op)}/g_{c(max)}$, or $ζ$), the latter of which is important for calculating the total leaf conductance ($g_{c(tot)}$). These two inputs cannot be directly measured on fossils; thus, the error rates associated with Figure 1 may not be representative for fossil studies. Franks et al. (2014) argue that within plant functional types growing in their natural environment, mean $A₀$ is fairly conservative, leading to the recommended mean $A₀$ values in Franks et al. (2014) (12 μmol m⁻² s⁻¹ for angiosperms, 10 for conifers, and 6 for ferns and ginkgos). Along similar lines, the mean ratio $g_{c(op)}/g_{c(max)}$ tends to be conserved across plant functional types; Franks et al. (2014) recommend a value of 0.2, which may correspond to the most efficient setpoint for stomata to control conductance (Franks et al., 2012). This conservation of physiological function is one of the underlying principles in the Franks model.

Here we test this assumption by estimating CO₂ from 40 phylogenetically diverse species of field-grown trees. In making these estimates, we use the recommended mean values of $A₀$ and $g_{c(op)}/g_{c(max)}$ from Franks et al. (2014) instead of measuring them directly (see also Montañez et al., 2016 for other ways to infer assimilation rate from fossils). Thus, this dataset should be a more faithful gauge for model accuracy as applied to fossils. Of the 40 species, 21 were previously published in Londoño et
al. (2018), who collected sun-adapted canopy leaves of angiosperms using a crane in Parque Nacional San Lorenzo, Panama. To test the method in temperate forests, we collected leaves from eleven angiosperm and seven conifer species from Dinosaur State Park (Rocky Hill, Connecticut), Wesleyan University (Middletown, Connecticut), and Connecticut College (New London, Connecticut) during the summer of 2015. Here, all trees grew in open, park-like settings; one to three sun leaves were sampled from the lower outside crown of each tree. In January of 2015, we also sampled sun-exposed leaves from the tree fern *Cyathea arborea* in El Yunque National Forest, Puerto Rico (near the Yokahú Tower).

Stomatal size and density were measured either on untreated leaves using epifluorescence microscopy with a 420-490 nm filter, or on cleared leaves (using 50% household bleach or 5% NaOH) using transmitted-light microscopy. For most species, whole-leaf δ¹³C comes from Royer and Hren (2017); the same leaves were measured for δ¹³C and stomatal morphology. The UC Davis Stable Isotope Facility measured some additional leaf samples. Table S1 summarizes for these 40 species all of the inputs needed to run the Franks model, along with the estimated CO₂ concentrations. Uncertainties in the estimates are based on error propagation using Monte Carlo simulations (Franks et al., 2014).

### 2.2 Temperature

The Franks model can be configured for any temperature. Franks et al. (2014) recommend that the photosynthesis parameters $A₀$ and $Γ^*$, and the air physical properties affecting diffusion of CO₂ into the leaf (the ratio of CO₂ diffusivity in air to the molar volume of air, or $d/v$) correspond with the mean daytime growing-season leaf temperature (more precisely, assimilation-weighted leaf temperature). The reasoning behind this is that (i) the assimilation-weighted leaf temperature will correspond with the mean $c/c_o$ derived from fossil leaf δ¹³C; and (ii) both theory (Michaletz et al., 2015, 2016) and observations (Helliker and Richter, 2008; Song et al., 2011) indicate that the control of leaf gas exchange leads to relatively stable assimilation-weighted leaf temperatures (~19-25 °C from temperate to tropical regions) despite large differences in air temperature. This is mostly due to the effects of transpiration on leaf energy balance. Franks et al. (2014) chose a fixed temperature of 25 °C because much of the Mesozoic and Cenozoic correspond to climates warmer than the present-day. When applying the Franks model to known cooler paleoenvironments, improved accuracy may be achieved with leaf-temperature-appropriate values for $A₀$, $Γ^*$, and $d/v$.

Bernacchi et al. (2003) proposed the following temperature sensitivity for $Γ^*$ based on experiments:

$$Γ^* = e^{(19.02 - 3.81 \cdot R/T)}, \quad (6)$$

where $R$ is the molar gas constant (8.31446×10⁻³ kJ K⁻¹ mol⁻¹) and $T$ is leaf temperature (K). Marrero and Mason (1972) describe the sensitivity of water vapor diffusivity to temperature as:

$$d = 1.87 \times 10^{-10} \left(\frac{T^{0.72}}{P}\right), \quad (7)$$

where $P$ is atmospheric pressure, which we fix at 1 atmosphere. Lastly, the temperature sensitivity of the molar volume of air follows ideal gas principles:

$$v = v_{STP} \left(\frac{T}{T_{STP}}\right) \left(\frac{P}{P_{STP}}\right), \quad (8)$$
where $T_{STP}$ is 273.15 K, $P_{STP}$ is 1 atmosphere, and $v_{STP}$ is the air volume at $T_{STP}$ and $P_{STP}$ (0.022414 m$^3$ mol$^{-1}$).

Using Eqs. (6-8), we can describe how, conceptually, the sensitivities of $\Gamma^*$ and $d/v$ to leaf temperature affect estimates of CO$_2$ from the Franks model. We apply these relationships to a suite of 409 fossil and extant leaves from 62 species of angiosperms, gymnosperms, and ferns. These data come from the current study (see Sect. 2.1 and 2.4) and Londoño et al. (2018), Kowalczyk et al. (2018), and Milligan et al. (in review).

To experimentally test more generally how the Franks model is influenced by temperature, we grew six species of plants inside two growth chambers with contrasting temperatures (Conviron E7/2; Winnipeg, Canada). Air temperature was $28 \pm 0.5 \degree C$ (1σ) and $20 \pm 0.3 \degree C$ during the day, and $19 \pm 0.7 \degree C$ and $11 \pm 1.1 \degree C$ during the night. We note that the difference in leaf temperature was probably smaller than that in air temperature during the day (8 °C; see earlier discussion). We held fixed the day length (17 hours with a 30 minute simulated dawn and dusk) and CO$_2$ concentration (500 ± 10 ppm). Light intensity at the heights where we sampled leaves ranged from 100-400 μmol m$^{-2}$ s$^{-1}$. Humidity differed moderately between chambers (76.5 ± 1.8% 1σ and 90.0 ± 3.6%). To minimize any chamber effects, we alternated plants between chambers every two weeks.

Four of the species started as saplings purchased from commercial nurseries: bare-root, one-foot tall saplings of *Acer negundo* and *Carpinus caroliniana*, one-foot tall saplings of *Ostrya virginiana* with a soil ball, and bare-root, four-inch tall saplings of *Ilex opaca*. We grew the other two species from seed: *Betula lenta* from a commercial source, and *Quercus rubra* from a single tree on Wesleyan University’s campus. All seeds were soaked in water for 24 hours and then cold stratified in a refrigerator for 30 and 60 days, respectively.

All seeds and saplings grew in the same potting soil (Promix Bx with Mycorise; Premier Horticulture; Quakertown, Pennsylvania, USA) and fertilizer (Scotts all-purpose flower and vegetable fertilizer; Maryville, Ohio, USA). They were watered to field capacity every other day, and we discarded any excess water passing through the pots. After three months of growth in the chambers, for each species-chamber pair we harvested the three newest fully expanded leaves whose buds developed during the experiment. In most cases, we harvested five plants per species-chamber pair; the one exception was *I. opaca*, where we were limited to three plants in the warm treatment and two in the cool treatment.

We measured stomatal size and density on cleared leaves (using 50% household bleach) with transmitted-light microscopy. Whole-leaf δ$^{13}$C comes from the UC Davis Stable Isotope Facility and the Light Stable Isotope Mass Spec Lab at the University of Florida; the same leaves were measured for δ$^{13}$C and stomatal morphology. We used either a hole punch or razor to remove two adjacent sections of leaf tissue near the leaf centers, avoiding major veins. Because we used the same CO$_2$ gas cylinder as Milligan et al. (in review), we used their two-end-member mixing model to calculate the δ$^{13}$C of the chamber CO$_2$ at 500 ppm (-10.6 ‰). We used the recommended values from Franks et al. (2014) for the physiological inputs $A_0$ and $g_{c(op)}/g_{c(max)}$. Table S1 summarizes all of the inputs from this experiment needed to run the Franks model, along with the estimated CO$_2$ concentrations. The standard errors for the inputs are based on plant means.

To test if leaf δ$^{13}$C and stomatal morphology (stomatal density, stomatal pore length, and single guard cell width) differed between temperature treatments across species, we implemented a mixed model in R (R Core Team, 2016) using the lme4 (Bates et al., 2015) and lmerTest (Kuznetsova et al., 2017) packages, with temperature and species as the two fixed factors. To test if there was a significant difference between CO$_2$ estimates from the two temperature treatments, we ran a Kolmogorov–Smirnov (KS) test in R. For each species, we first estimated CO$_2$ for each plant in the warm and cool treatments based on simulated inputs constrained by their means and variances. In the typical case with five plants per chamber, this produced five CO$_2$ estimates for the warm chamber and the same for the
cool chamber. A KS test was then used to test for a significant temperature effect. We repeated this
procedure 10,000 times, with 10,000 associated KS tests. The fraction of tests with a p-value < 0.05 was
taken as the overall p value. An advantage of this approach is that it incorporates both within- and
across-plant variation.

2.3 Photorespiration

c_i/c_a is estimated in the Franks model following Farquhar et al. (1982):

\[ \Delta_{\text{leaf}} = a + (b - a) \times \frac{c_i}{c_a}, \] (9)

where \( a \) is the carbon isotope fractionation due to diffusion of CO2 in air (4.4‰; Farquhar et al., 1982), \( b \)
is the fractionation associated with RuBP carboxylase (30‰; Roeske and O'Leary, 1984), and \( \Delta_{\text{leaf}} \) is the
net fractionation between air and assimilated carbon ([\( \delta^{13}C_{\text{air}} - \delta^{13}C_{\text{leaf}} \)]/[1+\( \delta^{13}C_{\text{leaf}}/1000 \)]).

Equation (9) can be expanded to include other effects, including photorespiration (Farquhar et al., 1982):

\[ \Delta_{\text{leaf}} = a + (b - a) \times \frac{c_i}{c_a} - f \frac{\Gamma^*}{c_a}, \] (10)

where \( f \) is the carbon isotope fractionation due to photorespiration. Photorespiration occurs when the
enzyme rubisco fixes O2, not CO2 (i.e., RuBP oxygenase). One product of photorespiration is CO2 (Jones, 1992),
whose \( \delta^{13}C \) is lower than the source substrate glycine. If this respired CO2 escapes to the
atmosphere, the \( \delta^{13}C \) of the leaf carbon becomes more positive. Thus, if \( c_i/c_a \) is calculated using Eq. (9),
as is common practice, the calculation may be falsely low, leading to an underprediction of atmospheric
CO2.

Measured values for \( f \) vary from ~9-15‰ (see compilation in Schubert and Jahren, 2018), which
is in line with theoretical predictions (Tcherkez, 2006). At a 400 ppm atmospheric CO2 and \( \Gamma^* \) of 40 ppm,
Eq. (10) implies that ~1‰ of \( \Delta_{\text{leaf}} \) is due to photorespiration, meaning that \( c_i/c_a \) should be ~0.04 higher
relative to Eq. (9). Here, using the suite of fossil and extant leaves described in Sect. 2.2, we explore how
the carbon isotopic fractionation associated with photorespiration affects CO2 estimates with the Franks
model. Because \( c_i/c_a \) is present in both of the fundamental equations (Eqs. 2 and 3), we solve them
iteratively until \( c_i/c_a \) converges.

2.4 Leaves that grow close to the forest floor

The composition of air close to the forest floor can differ considerably from the well-mixed atmosphere.
Of relevance to the Franks model, soil respiration can lead to a locally higher CO2 concentration and
lower \( \delta^{13}C_{\text{air}} \) (Table 24). This effect is strongest at night, when the forest boundary layer is thickest (e.g.,
Munger and Hadley, 2017), but we focus here on daylight hours because that is when most plants take
up CO2. In wet tropical forests, which can have very high soil respiration rates, CO2 during the day near
the forest floor can be elevated by tens-of-ppm, and the \( \delta^{13}C_{\text{air}} \) can be 2-3‰ lower; in temperate forests,
the deviations are smaller (Table 24). Above ~2 m, CO2 concentrations and air \( \delta^{13}C \) during the daytime
largely match the well-mixed atmosphere.
Table 21. Deviations in the δ13C and concentration of CO2 close to a forest floor relative to well-mixed air above the canopy. All measurements were made close to mid-day.

<table>
<thead>
<tr>
<th>Study</th>
<th>δ13Cair relative to well-mixed air (‰)</th>
<th>CO2 relative to well-mixed air (ppm)</th>
<th>Height above forest floor (m)</th>
<th>Forest location</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Tropical forest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Broadmeadow et al. (1992)</td>
<td>-2</td>
<td>+20</td>
<td>0.15-1</td>
<td>Trinidad during dry season</td>
</tr>
<tr>
<td>Buchmann et al. (1997)</td>
<td>-2</td>
<td>+30</td>
<td>0.70-0.75</td>
<td>French Guiana during wet and dry seasons</td>
</tr>
<tr>
<td>Holtum and Winter (2001)</td>
<td>NA</td>
<td>+50</td>
<td>0.10</td>
<td>Panama during wet and dry seasons</td>
</tr>
<tr>
<td>Lloyd et al. (1996)</td>
<td>-3</td>
<td>+70</td>
<td>1</td>
<td>Brazil (Amazon Basin)</td>
</tr>
<tr>
<td>Quay et al. (1989)</td>
<td>-3</td>
<td>+20</td>
<td>2</td>
<td>Brazil (Amazon Basin)</td>
</tr>
<tr>
<td>Sternberg et al. (1989)</td>
<td>-2</td>
<td>+25</td>
<td>1</td>
<td>Panama during wet and dry seasons</td>
</tr>
<tr>
<td><strong>Temperate forest</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Francey et al. (1985)</td>
<td>-1</td>
<td>+20</td>
<td>1</td>
<td>Tasmania</td>
</tr>
<tr>
<td>Munger and Hadley (2017)</td>
<td>NA</td>
<td>+15</td>
<td>1</td>
<td>Massachusetts (Harvard Forest)</td>
</tr>
</tbody>
</table>

As a result, leaves that grow close to the forest floor may cause the Franks model to produce CO2 estimates higher than that of the mixed atmosphere for at least two reasons. First, the concentration of CO2 near the forest floor is elevated; that is, the model may correctly estimate a CO2 concentration that the user is not interested in. Second, because the δ13Cair that a forest-floor plant experiences is lower than the global well-mixed value, if the user chooses the well-mixed value for model input (inferred, for example, from the δ13C of marine carbonate; Tipple et al., 2010), c/ca and thus atmospheric CO2 will be overestimated (see Eq. 2). We sought to test how the Franks model is affected by the forest-floor microenvironment for five tropical angiosperm species and fifteen temperate angiosperm and fern species. The tropical leaves were sampled at ~1-2 m height from Parque Nacional San Lorenzo, Panama. In contrast to the canopy data set from San Lorenzo (Sect. 2.1), these CO2 estimates have not been previously reported. In the summer of 2015, seven fern species were sampled at ~0.5 m height from Connecticut College and Wesleyan University. Also, we used leaf vouchers from Royer et al. (2010), who sampled eight herbaceous angiosperm species at ~0.1-0.2 m height from Reed Gap, Connecticut. For all 20 species, stomatal and carbon isotopic measurements follow the methods described in Sect. 2.1. Table S1 contains all of the inputs needed to run the Franks model, along with the estimated CO2 concentrations.

We also investigated if we could include the forest-floor δ13Cair effect in our estimates of atmospheric CO2. If the only CO2 inputs close to the forest floor are from the soil and well-mixed atmosphere, the system can be modeled as a two-endmember mixing model where δ13Cair has a positive, linear relationship with 1/CO2 (Keeling, 1958). If the CO2 concentration and δ13C of both endmembers are known, the forest-floor microenvironment should fall somewhere on the modelled line. Importantly, the Franks model provides a second constraint on the system. Here, δ13Cair has a negative, nonlinear relationship with 1/CO2 because δ13Cair is positively related to c/ca and CO2. The Franks model thus provides a second calculation for the relationship between δ13Cair and estimated CO2.
concentration. The intersection between the two curves should be the correct $\delta^{13}C_{\text{air}}$ and CO$_2$
concentration for the forest-floor microenvironment.

To estimate the soil CO$_2$ endmember, we measured the $\delta^{13}C$ of soil organic matter collected
from the A horizons of 13 soil sites at San Lorenzo, and of five each at Reed Gap and Connecticut
College. For all soils, we assume a 5000 ppm CO$_2$ concentration for a depth that is below the zone of CO$_2$
diffusion from the atmosphere (~0.3 m; Cerling, 1999; Breecker et al., 2009). The true value for wet
temperate and tropical forest soils may be somewhat less or substantially more than 5000 ppm (Medina
et al., 1986; Cerling, 1999; Hirano et al., 2003; Hashimoto et al., 2004; Sotta et al., 2004). Because the
mixing model uses 1/CO$_2$, a much higher CO$_2$ concentration (e.g., 10000 ppm) has little impact on our
results.

3 Results and Discussion

3.1 General testing in living plants

Estimates of CO$_2$ across the 40 tree species sampled in the field range from 275 to 850 ppm, with a
mean of 478 ppm and median of 472 ppm (Fig. 2); two-thirds of the estimates range between 353 and
585 ppm. In 28% of the tested species, the estimated CO$_2$ concentrations overlap with the target
concentration (400 ppm) at 95% confidence; for these species, the CO$_2$ estimates do not differ
significantly from the target. There are no strong differences across taxonomic orders, nor between
leaves from tropical and temperate forests. The mean error rate across the estimates is 28% (median =
24%), which is higher than estimates that include direct measurements of the physiological inputs $A_0$
and $g_{c(op)}/g_{c(max)}$ (mean = 20%; median = 13%; Fig. 1). Along similar lines, if the estimates presented in Fig.
1 are re-estimated using the values for $A_0$ and $g_{c(op)}/g_{c(max)}$ recommended by Franks et al. (2014), the
mean error rate increases to 37% (median = 33%). If only the default values of $A_0$ are used, the median
test error rate is 27%; and for only default values of $g_{c(op)}/g_{c(max)}$, 22%.

These results indicate that CO$_2$ accuracy is generally improved when $A_0$ and/or $g_{c(op)}/g_{c(max)}$ is
measured. These measurements require expensive gas-exchange equipment and are not always easy or
practical to make. Moreover, $A_0$ and $g_{c(op)}/g_{c(max)}$ cannot be measured on fossils. Some gains in accuracy
are possible by measuring $A_0$ and $g_{c(op)}/g_{c(max)}$ on extant relatives of the fossil species (e.g., the same
genus). Absent of this, our analysis using the recommended mean values of Franks et al. (2014) indicates
an error rate, on average, of approximately 28%. This is comparable to or better than other leading
paleo-CO$_2$ proxies (Franks et al., 2014).

One reliable way to improve accuracy is to estimate CO$_2$ with multiple species because the
falsely-high and falsely-low estimates partially cancel each other out. The grand mean of estimates
presented in Fig. 2 (478 ppm) is 20% from the 400 ppm target, which is less than the 28% mean error
rate of individual estimates.
Figure 2. Estimates of CO₂ based on canopy leaves from 40 tree species. Uncertainties in the estimates correspond to the 16th-84th percentile range. Vertical line is the correct concentration (400 ppm). On the left is an order-level vascular plant phylogeny (APW v.13; Stevens, 2001 onwards).
Dow et al. (2014) observed that $g_{c,\text{op}}/g_{c,\text{max}}$ inversely varies with CO$_2$ in *Arabidopsis thaliana*, but primarily at subambient concentrations (yellow triangles in Fig. 3). At elevated CO$_2$, $g_{c,\text{op}}/g_{c,\text{max}}$ is close to 0.2, which is the value recommended by Franks et al. (2014). Data from eleven species of angiosperms, conifers, and ferns at present-day (or near present-day) and elevated CO$_2$ concentrations support the view of a limited effect at high CO$_2$ (Fig. 3; Franks et al., 2014; Londoño et al., 2018; Milligan et al., in review). More data at subambient CO$_2$ are needed, but for CO$_2$ concentrations similar to or higher than the present-day, we see no strong reason to include a CO$_2$ sensitivity in $g_{c,\text{op}}/g_{c,\text{max}}$. The rather low values for *Cedrus deodara* and many of the tropical angiosperms (<0.1) are likely due to stress imposed by their growth chamber environment; these $g_{c,\text{op}}/g_{c,\text{max}}$ values are probably not representative of field-grown trees, which tend to be closer to 0.2 (Franks et al., 2014).

**Figure 3.** Literature compilation of the sensitivity of $g_{c,\text{op}}/g_{c,\text{max}}$ (ratio of operational to maximum leaf conductance to CO$_2$) to atmospheric CO$_2$ concentration.

### 3.2 Temperature

The temperature sensitivities of the ratio of diffusivity of CO$_2$ in air to the molar volume of air ($d/v$) and the CO$_2$ compensation point in the absence of dark respiration ($\Gamma^*$) have little effect on estimated CO$_2$ in the Franks model (Fig. 4). Given that assimilation-weighted leaf temperature only varies about 7 °C across plants today, the differences shown in Fig. 4—which are based on leaf temperatures of 25 °C and 15 °C—are likely a maximum effect. As such, we consider the use of a fixed leaf temperature (e.g., 25 °C) in the model to be a defensible simplification.
Figure 4. Estimates of CO₂ at leaf temperatures of 25 °C and 15 °C. Each symbol is an extant or fossil leaf. The difference in estimated CO₂ for any leaf is due to the theoretical effects of temperature on gas diffusion (\(d/v\)) and the CO₂ compensation point in the absence of dark respiration (\(Γ^*\)) (Eqs. 6-8).

Other inputs in the model may respond to temperature, though. In our growth chamber experiments where daytime air temperatures were 28 °C and 20 °C, the effect on estimated CO₂ was mixed (Fig. 5). In five out of six species, estimated CO₂ was higher in the warm treatment, but for all species these differences were not statistically significant (\(P > 0.05\) based on a KS test; see Methods). Incorporating the temperature sensitivities in \(d/v\) and \(Γ^*\) had little effect ("adj" estimates in Fig. 5), as expected from Fig. 4.

None of the measured inputs—stomatal density, stomatal pore length, single guard cell width, and leaf δ¹³C—were significantly affected by temperature across all species (\(P > 0.05\) for each of the four inputs based on a mixed model; see Methods). These small differences probably cannot account for the differences in estimated CO₂ between temperatures. It is more likely that some of the inputs that we did not directly measure, such as assimilation rate (\(A_0\)), the \(g_{c(op)}/g_{c(max)}\) ratio, or mesophyll conductance (\(g_m\)), differ from the true mean value. In the cases for the five species where estimated CO₂ is higher in the warm treatment, our mean \(A_0\) for the warm plants must be falsely high, or \(g_{c(op)}/g_{c(max)}\) or \(g_m\) falsely low.

In summary, we see no strong reason to expand the parameterization of temperature in the model, though more growth-chamber experiments may be warranted. We note that in three out of the six species from the warm treatment, the estimated CO₂ cannot be distinguished from the 500 ppm target at 95% confidence; for the cool treatment, this is true for four of the species. Also, the across-species means of estimated CO₂ for the warm and cool treatments are reasonably close to the 500 ppm target (590 and 502 ppm, respectively) and overall have a mean error rate of 25%. This level of uncertainty is similar to our field estimates where we did not measure \(A_0\) or \(g_{c(op)}/g_{c(max)}\) (28%; see Fig. 2).
This too provides support for our recommendation that it is not critical to include a broader treatment of temperature in the model.

**Figure 5.** Estimates of CO₂ for plants grown inside growth chambers at daytime air temperatures of 28 °C and 20 °C. Also shown are estimates after taking into account the temperature sensitivity of gas diffusion (d/v) and the CO₂ compensation point in the absence of dark respiration (Γ*) (“adj”; see also Fig. 4). Dashed line is the correct CO₂ concentration (500 ppm). Uncertainties in the estimates correspond to the 16th–84th percentile range.

### 3.3 Photorespiration

The theoretical effects of photorespiration do not strongly impact estimates of CO₂ in the Franks model. The average effect for our 409 extant and fossil leaves is to increase estimated CO₂ by 2.2% plus 38 ppm (Fig. 6). At 1000 ppm, for example, estimates would increase by 60 ppm. This calculation assumes a photorespiration fractionation (f) of 9.1‰, which is the value estimated for *Arabidopsis thaliana* (Schubert and Jahren, 2018). If a fractionation towards the upper bound of published estimates is used instead (15‰), estimated CO₂ increases on average by 3.8% plus 61 ppm. Across this range in f, the associated uncertainty in estimated CO₂ is well within the method’s overall precision (~+35/-25% at 95% confidence; Franks et al., 2014). As such, CO₂ estimates made without these photorespiration effects (i.e. using Eq. 9 instead of Eq. 10) should be reliable, although some improvement is possible using Eq. 10 in cases where f is accurately known.
Figure 6. Estimates of CO₂ with and without a photorespiration effect ($f = 9.1\%$; see Eq. 10). Each symbol is an extant or fossil leaf. Dashed line is $y=x$.

We note that both $f$ and $\Gamma^*$ are also affected by atmospheric O₂ concentration. Because O₂ is directly responsible for photorespiration, $f$ should scale with O₂ (or, more precisely, the O₂:CO₂ molar ratio). Unfortunately, this effect is poorly constrained (Beerling et al., 2002; Berner et al., 2003; Porter et al., 2017). In contrast, the theoretical effect of O₂ on $\Gamma^*$ is known: it is linear with an approximate slope of 2 (Farquhar et al., 1982; see their Eq. B13). During the Phanerozoic, O₂ likely ranged from 10-30%, with lows during the early Paleozoic and early Triassic, and highs during the Carboniferous to early Permian and Cretaceous (Berner, 2009; Glasspool and Scott, 2010; Arvidson et al., 2013; Mills et al., 2016; Lenton et al., 2018). Assuming a present-day $\Gamma^*$ of 40 ppm (at 21% O₂), $\Gamma^*$ would be 60 ppm at 30% O₂ and 20 ppm at 10% O₂. Running the Franks model on our library of 409 extant and fossil leaves, we find little effect on estimated CO₂: estimates are 7.4% higher on average at 30% O₂ than at 10% O₂ (see also McElwain et al., 2016).

3.4 Leaves that grow close to the forest floor

CO₂ estimates for tropical understory leaves from five species at San Lorenzo, Panama, are very high, ranging from 1903 to 18863 ppm (species mean = 6837 ppm). For two of the species Londoño et al. (2018) also analyzed canopy leaves from trees nearby, and these within-species comparisons highlight the vast discrepancy (Ocotea sp.: 541 vs. 5737 ppm; Tontelea sp.: 622 vs. 18863 ppm). The primary difference in the inputs between the canopy and understory leaves is the $\delta^{13}C_{leaf}$: Londoño et al. (2018) report a species-mean $\delta^{13}C_{leaf}$ of -30.0% for the 21 canopy species versus -35.6% for the five understory
species. This difference leads to very different mean estimates of $c_i/c_a$: 0.69 for canopy leaves versus a highly unrealistic (Diefendorf et al., 2010) 0.93 for understory leaves.

It is likely that the high CO$_2$ estimates from understory leaves are mostly driven by falsely high $\delta^{13}C_{air}$ inputs. Following the mixing model strategy outlined in Sect. 2.4 (and based on a soil organic matter $\delta^{13}C$ of -28.2‰ measured at San Lorenzo), we calculate a species-mean $\delta^{13}C_{air}$ of -16.7‰ (mean of intersection points in Fig. 7). When this $\delta^{13}C_{air}$ is used to estimate CO$_2$ with the Franks model (instead of -8.5‰), the species mean drops to 699 ppm. This is somewhat higher than the species mean from canopy leaves in the same forest (563 ppm; red triangles in Fig. 2; Londoño et al., 2018).

Understory leaves from Connecticut temperate forests show similar but less dramatic patterns, which we attribute to a more open canopy with stronger atmospheric mixing. CO$_2$ estimates for the 15 species range from 447 to 1567 ppm (mean = 794 ppm). Our intersection method identifies a mean

**Figure 7.** Sensitivity of estimated CO$_2$ in the Franks model to the $\delta^{13}C$ of atmospheric CO$_2$. Estimates come from leaves of five angiosperm species that grew close to the forest floor in Parque Nacional San Lorenzo, Panama. The step in $\delta^{13}C_{air}$ between estimates is 0.5‰. The dashed line is a two-endmember mixing model for CO$_2$ between the soil and well-mixed atmosphere. The intersection between the mixing model and the Franks model should correspond to the CO$_2$ concentration and $\delta^{13}C_{air}$ of the forest-floor microenvironment.
δ¹³C_air of -11.2‰ for the Wesleyan and Connecticut College campuses (based on a soil δ¹³C of -27.6‰ measured at Connecticut College) and -10.3‰ for Reed Gap (soil δ¹³C = -26.4‰). Using these adjusted δ¹³C_air, the species mean of estimated CO₂ drops to 566 ppm, which is somewhat higher than the species mean from canopy leaves in the same areas (449 ppm; blue circles in Fig. 2).

We acknowledge that this analysis is too simple. First, we did not measure the understory δ¹³C_air (this would require repeated measurements during different daytime hours over a growing season to calculate a time-integrated value); instead, we assumed a simple two end-member mixing model between the soil and well-mixed atmosphere. Second, other factors probably contribute to the differences in estimated CO₂ between canopy and understory leaves. Our predicted δ¹³C_air values are too low (~8‰ and 2‰ lower than the well-mixed atmosphere for the tropical and temperate forests) and our estimated CO₂ too high (~100 ppm higher than that from canopy leaves). In the lowermost 1-2 meters of the canopy, previous work suggests up to a -3‰ and +70 ppm deviation in tropical forests and -1‰ / +20 ppm in temperate forests (Table 1). One input that could help to resolve this discrepancy is the assimilation rate (A₀). We assumed a fixed A₀ of 12 µmol m⁻² s⁻¹ for all leaves, regardless of canopy position. Shade leaves often have lower assimilation rates than sun leaves (Givnish, 1988). Substituting lower A₀ values for understory leaves would lower estimated CO₂ roughly in proportion (Eqs. 2-3). Using lower A₀ values for shade leaves in the model is appropriate, but determining the best value is difficult. Typical A₀ values for leaves growing at the top of the canopy in full sun are far more consistent because photosynthesis in these leaves is usually at its maximum capacity (saturated at full sunlight) for the prevailing atmospheric CO₂ concentration. Because the degree of shadiness near the forest floor is highly variable, photosynthesis (A₀) in these leaves will be acclimated to some fraction of the full-sun maximum in a sun exposed leaf, but careful thought must go into determining what this fraction is.

We note that our mixing-model strategy cannot be applied to fossils because the global atmospheric CO₂ concentration is needed (one endpoint for dashed line in Fig. 7). Instead, our motivation for the analysis is to demonstrate that: 1) leaves growing in the lowermost 2 m of the canopy should be considered with caution in the context of the Franks model; and 2) the failure of the model is due to faulty inputs (mostly δ¹³C_air), not the model itself.

In most fossil leaf deposits, shade morphotypes are comparatively rare (e.g., Kürschner, 1997; Wang et al., 2018) because—relative to sun leaves—they are not as tough, do not travel as far by wind, and are produced at a slower rate (Dilcher, 1973; Roth and Dilcher, 1978; Spicer, 1980; Ferguson, 1985; Burnham et al., 1992). Our recommendation is to exclude such leaves. There are several ways to differentiate sun vs. shade morphotypes: overall shape (Talbert and Holch, 1957; Givnish, 1978; Kürschner, 1997; Sack et al., 2006), shape of epidermal cells (larger and with a more undulated outline in shade leaves; Kürschner, 1997; Dunn et al., 2015), vein density (lower in shade leaves; Uhl and Mosbruger, 1999; Sack and Scoffoni, 2013; Crifo et al., 2014; Londoño et al., 2018), and range in δ¹³C_leaf (high when both sun and shade leaves are present, for example in our study; Graham et al., 2014). Not all shade leaves grow within 2 m of the forest floor, but excluding all such leaves would eliminate the forest-floor bias.

4 Conclusions

The Franks model is reasonably accurate (~28% error rate) even when the physiological inputs A₀ (assimilation rate at a known CO₂ concentration) and g_c(op)/g_c(max) (ratio of operational to maximum leaf conductance to CO₂) are inferred, not measured. Accuracy does improve when these inputs are measured (~19% error rate), but such measurements are not possible with fossils and may not always be feasible with nearest living relatives. A 28% error rate is broadly in line with (or better than) other leading paleo-CO₂ proxies.
Most of the possible confounding factors that we investigated appear minor. The temperature sensitivities of $d/v$ (related to gas diffusion) and $F^*$ (CO$_2$ compensation point in the absence of dark respiration) have a negligible impact on estimated CO$_2$. Our temperature experiments in growth chambers point to larger differences in some species, which must be related to incorrect values for inputs that were not directly measured, such as $A_0$, $g_{c(op)}/g_{c(max)}$, and $g_m$ (mesophyll conductance). Overall, though, we find that the differences in estimated CO$_2$ imparted by temperature are generally smaller than the overall 28% error rate.

Incorporating the covariance between CO$_2$ concentration and photorespiration leads to only small changes in estimated CO$_2$. O$_2$ concentration affects photorespiration and thus may confound CO$_2$ estimates from the Franks model, but presently the effect is poorly quantified. The effect of O$_2$ on $F^*$ is better known, and imparts only small changes in estimated CO$_2$ across a feasible range in Phanerozoic O$_2$ of 10-30%.

Leaves from the lowermost 1-2 m of the canopy experience slightly elevated CO$_2$ concentrations and lower air $\delta^{13}$C during the daytime relative to the well-mixed atmosphere. We find that if we use the well-mixed air $\delta^{13}$C to estimate CO$_2$ from leaves that grew near the forest floor, estimates are too high, especially in dense tropical canopies. When we use a two-endmember mixing model to calculate the correct local air $\delta^{13}$C, the falsely-high CO$_2$ estimates largely disappear. For fossil applications, shade leaves from the bottom of the canopy should be avoided. Shade leaves are typically rare in the fossil record (relative to sun leaves), and can be identified by their overall shape, the shape of their epidermal cells, their low leaf $\delta^{13}$C, and their low vein density.

Conceptually, the Franks model holds considerable promise for quantifying paleo-CO$_2$: it is mechanistically grounded and can be applied to most fossil leaves. Our tests of the model’s accuracy and sensitivity to temperature and photorespiration largely uphold this promise.

Author contribution. DR, KM, MM, and LL designed and conducted the experiments; all authors interpreted the data; DR prepared the manuscript with contributions from all co-authors.

Competing interests. The authors declare that they have no conflict of interest.

Acknowledgements. We thank G. Dreyer and P. Siver for logistical support at Connecticut College, S. Wang for lab assistance, and C. Crifò and A. Baresh for collecting the tropical samples. Support for LL was provided by the Smithsonian Tropical Research Institute; the Mark Tupper Fellowship; National Science Foundation grants EAR 0824299 and OISE, EAR, DRL 0966884; the Anders Foundation; and the Gregory D. and Jennifer Walston Johnson and 1923 Fund.

References


No evidence for a large atmospheric CO$_2$ spike across the Cretaceous-Paleogene boundary

Joseph N. Milligan$^{1,2}$, Dana L. Royer$^1$, Peter J. Franks$^3$, Garland R. Upchurch$^4$, Melissa L. McKee$^1$

$^1$Department of Earth and Environmental Sciences, Wesleyan University, Middletown, Connecticut, USA.

$^2$Department of Geology, Baylor University, Waco, TX, USA.

$^3$School of Life and Environmental Sciences, University of Sydney, Sydney, New South Wales, Australia.

$^4$Department of Biology, Texas State University, San Marcos, TX, USA.

Corresponding author: Joseph Milligan (Joseph_Milligan@baylor.edu)
Key Points:

- Understanding atmospheric CO$_2$ across the Cretaceous-Paleogene boundary has been limited due to deficiencies in existing records
- Our study highlights the utility of a proxy based on leaf gas-exchange principles
- We record a small transient rise in atmospheric CO$_2$ that is more in line with modeled estimates of both Deccan volcanism and a bolide impact
Abstract

Currently there is only one paleo-CO$_2$ record from plant macrofossils that has sufficient stratigraphic resolution to potentially capture a transient spike related to rapid carbon release at the Cretaceous-Paleogene (K-Pg) boundary. Unfortunately, the associated measurements of stomatal index are off-calibration, leading to a qualitative interpretation of $>2300$ ppm CO$_2$. Here we re-evaluate this record with a paleo-CO$_2$ proxy based on leaf gas-exchange principles. We also test the proxy with three living species grown at 500 and 1000 ppm CO$_2$, including the nearest living relative of the K-Pg fern, and find a mean error rate of $\sim 22\%$, which is comparable to other leading paleo-CO$_2$ proxies. Our fossils record a $\sim 250$ ppm increase in CO$_2$ across the K-Pg boundary from $\sim 625$ to $\sim 875$ ppm. A small CO$_2$ spike associated with the end-Cretaceous mass extinction is consistent with many temperature records and with carbon cycle modeling of Deccan volcanism and the meteorite impact.

Plain Language Summary

Currently there is only one paleo-CO$_2$ record close enough to the Cretaceous-Paleogene (K-Pg) boundary to record a rapid release in atmospheric CO$_2$, a greenhouse gas. This record is based on the stomatal frequencies of fern fossils at the K-Pg boundary and Ginkgo fossils before and after the boundary. Unfortunately, due to deficiencies with the method, the CO$_2$ inferences are only qualitative. Here we look at the same fossils with a proxy based on leaf gas-exchange principles (i.e. photosynthesis). We first test the proxy with three living species grown at 500 and 1000 ppm CO$_2$, including the nearest living relative of the K-Pg fern, and find a comparable accuracy to other quantitative paleo-CO$_2$ proxies. The fossils record a modest $\sim 250$ ppm increase in CO$_2$ across the K-Pg boundary. These estimates are consistent with most temperature records and with carbon cycle modeling of Deccan volcanism and the meteorite impact.
1 Introduction

The Cretaceous–Paleogene (K-Pg) boundary ~66 Ma marks one of the largest mass extinctions in Earth’s history (Alroy, 2008; Brusatte et al., 2015; McElwain and Punyasena, 2007; Raup and Sepkoski, 1982). The concentration of atmospheric CO$_2$ may have risen abruptly at this time, contributing to the biological upheaval (Beerling et al., 2002). Removal of an instantaneous release of CO$_2$ to the atmosphere typically requires up to 100-200 kyrs, following exponential decay due to silicate weathering (Archer, 2005; Colbourn et al., 2015; Schaller et al., 2011; Zeebe and Zachos, 2013). Adequate constraints on atmospheric CO$_2$ from proxy records during this critical period have been missing, mostly because of a lack in sufficient stratigraphic resolution to definitively identify individual records occurring <100 kyrs after the extinction event. This is because either the stratigraphic section is too coarse to resolve 100 kyrs of time (Steinthorsdottir et al., 2016) or because definitive markers of the boundary (e.g., iridium spike, presence of microspherules) are missing (Huang et al., 2013; Nordt et al., 2003; Zhang et al., 2018).

One exception is the study of Beerling et al. (2002), who used stomatal indices (SI, stomatal density normalized by the number of epidermal cells) to estimate CO$_2$ from fern macrofossils (aff. Stenochlaena) that occur 5-25 cm above the K-Pg boundary in the Raton Basin, New Mexico. In this stratigraphic section, the K-Pg boundary is identified by an iridium spike and shocked quartz, and the fossils come from sediments that contain, and lie directly above, the fern spore spike. This fern spike is present across the globe (Vajda et al., 2001) and likely occurred within $10^3$ yrs after the K-Pg boundary (Clyde et al., 2016). Thus, the aff. Stenochlaena fossils should record any transient rise in atmospheric CO$_2$ associated with the Chicxulub impact and K-Pg boundary. Indeed, the fossils likely capture close to the peak in CO$_2$
change because after an instantaneous release, CO$_2$ will remain significantly elevated for hundreds of years (Solomon et al., 2009; Zeebe, 2013). Unfortunately, the measured stomatal indices fall well below the present-day calibration of *S. palustris*, leading Beerling et al. (2002) to interpret a CO$_2$ concentration that exceeded the calibrated space (>2300 ppm), considerably higher than latest Cretaceous and earliest Paleocene CO$_2$ values of ~350-550 ppm inferred from *Ginkgo* fossils (Beerling et al., 2002; 2009). The Beerling et al. (2002) study thus suggests a very large, but poorly constrained, CO$_2$ pulse.

Leaf gas-exchange models are an alternative to stomatal density (SD) and SI proxies for estimating paleo-CO$_2$ concentration (Franks et al., 2014; Konrad et al., 2008, 2017). The model developed by Franks et al. (2014) depends on the well-established relationship between the rate of CO$_2$ assimilation of plants (A), leaf conductance to CO$_2$ ($g_{ctot}$), and the difference between atmospheric ($c_a$) and leaf intercellular CO$_2$ ($c_i$) (Farquhar and Sharkey, 1982):

$$A = g_{ctot} (c_a - c_i)$$  \hspace{1cm} (1)

Equation 1 can be rearranged to solve for atmospheric CO$_2$ (Equation 2). The three input variables needed are the average assimilation rate (determined from a nearest living relative), average total leaf conductance (determined largely from SD and stomatal size measured on the fossil), and average $c_i/c_a$ (determined from the fossil leaf and air carbon isotopic composition combined with knowledge of the fractionation process) (Franks et al., 2014):

$$c_a = \frac{A}{g_{ctot} \left(1 - c_i/c_a \right)}$$  \hspace{1cm} (2)

The model has been used to reconstruct CO$_2$ during the Phanerozoic (Franks et al., 2014), including the late Paleozoic (Montañez et al., 2016), middle Cretaceous (Richey et al., 2018), Late Cretaceous (Martínez et al., 2018), early Paleocene (Kowalczyk et al., 2018), middle
Eocene (Maxbauer et al., 2014; Wolfe et al., 2017), Oligocene-Miocene boundary (Reichgelt et al., 2016; Tesfamichael et al., 2017) and early Miocene (Londoño et al., 2018).

Leaf gas-exchange models provide at least five crucial advantages over other stomatal approaches: (1) they are based mechanistically on physiological principles, not empirical, species-specific calibrations; (2) measurements of SD, a component of $g_{clos}$, are typically more reliable and easier to make than SI because epidermal cells can be difficult to count (Barclay and Wing, 2016); (3) they are less sensitive to the saturating effect that can limit other stomatal methods to <500-1000 ppm CO$_2$ (e.g. Doria et al., 2011); (4) they can be applied to most subaerial leaves from C$_3$ species, regardless of age or taxonomy; and (5) they are not restricted to species whose SD or SI is sensitive to CO$_2$ because the models have multiple physiological inputs with well-understood sensitivities to CO$_2$. Importantly, these gas-exchange methods open up much of the paleobotanical record for quantitative CO$_2$ inference, not just to fossil taxa that are still living today. While the Franks et al. (2014) model shows promise, more extensive testing will improve confidence in the CO$_2$ estimates. Specifically, model validation with extant species has been limited to mostly angiosperms and a few gymnosperms, neglecting major clades such as ferns and lycophytes. Additionally, the model has been tested at elevated CO$_2$ on only a few species (Franks et al., 2014; Londoño et al., 2018).

Here we test the model using growth-chamber experiments at elevated CO$_2$ (500 and 1000 ppm) for two ferns (*Osmundastrum cinnamomeum* (L.) C. Presl and a close living relative to the K-Pg fern, *Stenochlaena palustris* (Burm.f.) Bedd.), and one conifer (*Cedrus deodara* (Roxb.) Loud. We then use the same fossils of aff. *Stenochlaena* and *Ginkgo* from Beerling et al. (2002) to re-evaluate atmospheric CO$_2$ across the K-Pg boundary using the gas-exchange model of Franks et al. (2014).
2 Materials and methods

For detailed methods and all data, see the supporting information.

2.1 Growth chamber experiments

All plants were potted with Premier Horticulture "Pro-mix" Bx with Mycorise and grown in two Conviron E7/2 growth chambers. Plants were watered to field capacity daily and given Scotts all-purpose flower and vegetable fertilizer (10-10-10) every two months. The chamber conditions were set to a 17-hour photoperiod with a 30-minute simulated dawn and dusk. Temperature was $25 \pm 0.2^\circ$C (1σ) during the day and $20 \pm 1^\circ$C (1σ) at night. The relative humidity was $84 \pm 5\%$ (1σ) and the CO$_2$ concentration was either $500 \pm 25$ (1σ) or $1000 \pm 15$ (1σ) ppm. Growth light levels (photosynthetically active radiation, or PAR) varied between 100-500 µmol m$^{-2}$ s$^{-1}$ depending on plant height. Plants were rotated between the two chambers every two weeks to negate any chamber effects (e.g., Porter et al., 2015).

2.2 Fossil leaves

The fossils come from Beerling et al. (2002). The aff. *Stenochlaena* fossils were collected at the Clear Creek South locality in the Raton Basin, New Mexico (Wolfe and Upchurch, 1987). The fossils represent an extinct (and currently unnamed) genus related to *Stenochlaena* (Wolfe and Upchurch, 1987), with identification based on venation, tooth and frond architecture, and stomatal anatomy, especially maceration-resistant cutin lamellae on the guard cells (Beerling et al., 2002; Wolfe and Upchurch, 1987). The stratigraphic interval containing the fern fossils includes the top of the fern spore spike and the overlying level where dicot pollen returns to dominance.

The latest Cretaceous and earliest Paleocene *Ginkgo adiantoides* fossils were obtained by loan from the Denver Museum of Nature and Science (DMNH) and the Yale Peabody Museum.
(YPM), respectively. The Cretaceous fossils come from the Hell Creek Formation in the Williston Basin of North Dakota (DMNH site 566), 33.5 m below the K-Pg boundary (Johnson, 2002). Based on constraints from geochronology, magnetostratigraphy, and sedimentation rates, Hicks et al. (2002) consider the locality 0.5 Myrs older than the K-Pg boundary. The early Paleocene fossils come from the Fort Union Formation in the Bighorn Basin of Wyoming (YPM locality 7659), 4 m above the K-Pg boundary; based on sedimentation rates, Wing et al. (1995) interpret the site to post-date the K-Pg boundary by 0.5 Myrs. We assume a K-Pg boundary age of 66 Ma (Gradstein et al., 2012; Renne et al., 2013).

2.3 Leaf gas-exchange model

The Franks et al. (2014) leaf gas-exchange model has 16 inputs that are used to calculate the average assimilation rate, total leaf conductance, and $c_i/c_a$ (Equation 2). When possible, we measured the inputs directly, including SD, stomatal pore length, single guard cell width, and leaf $\delta^{13}$C (Table S1). For living plants, the assimilation rate, $A$, and operational stomatal conductance to CO$_2$, $g_{c(op)}$, were also measured with a LI-COR 6400 portable photosynthesis system. These measurements were made under environmental conditions identical (or nearly identical) to the growth chamber environment. Leaves first equilibrated inside the leaf chamber for 10 to 30 minutes. All reported results are means of the most stable individual measurements (typically <5% variance across measurements).

For the fossil leaves, nearest living relatives were used to assign taxon-specific values of $A_0$ (assimilation rate at a known CO$_2$ concentration) and $g_{c(op)}/g_{c(max)}$ (ratio of operational to maximum stomatal conductance to CO$_2$). For aff. Stenochlaena, values come from S. palustris reported here; for G. adiantoides, values come from field-grown G. biloba at ~400 ppm CO$_2$ (Kowalczyk et al., 2018). For other inputs not directly measured, we used the recommended
values from Franks et al. (2014) or appropriate values from the literature (see Dataset S1). To
solve for atmospheric CO$_2$, we use the Kowalczyk et al. (2018) code written in R (v.3.4.4; R core
team, 2018).

As with the Beerling et al. (2002) study, our atmospheric CO$_2$ reconstruction comes from
two different species at three different localities. Because this potentially introduces species and
environmental effects, we performed a sensitivity analysis by estimating CO$_2$ after sequentially
varying each input parameter across a range typical for C$_3$ plants. Consistent with previous work
(Kowalczyk et al., 2018; Maxbauer et al., 2014; McElwain et al., 2016) we find that among the
inputs that cannot be measured directly on fossils, changes in $A_0$ and $g_{c(op)}/g_{c(max)}$ have the biggest
impact on estimated CO$_2$ (Figure S16). As such, we explored how different value choices for
these inputs may affect our CO$_2$ estimates. For example, because a one-step change in CO$_2$ may
not induce the same physiological response as a slow-and-steady CO$_2$ increase over geological
time, we evaluated the model both with the measured physiological inputs (discussed earlier) and
generic values recommended by Franks et al. (2014) (Table S2).

We note that the Franks et al. (2014) leaf gas-exchange model is based on leaf
temperature, not air temperature. Both theory (Michaletz et al., 2015; Michaletz et al., 2016) and
observations (Helliker and Richter, 2008; Song et al., 2011) indicate that the control of leaf gas
exchange leads to relatively stable assimilation-weighted leaf temperatures (~19-25 °C from
temperate to tropical regions; i.e., thermoregulation). Thus, despite significant changes (e.g.,
several degrees) in global mean air temperature, as often observed across the K-Pg boundary,
daytime leaf temperature during the growing season should stay relatively constant. If instead
leaf temperature did vary substantially, it could have mixed effects on many model inputs ($A$,
$g_{c(op)}/g_{c(max)}$, SD, stomatal size, $c_i/c_a$); for example, an increase in $A$ with no changes to other
inputs will cause an equally proportional increase in estimated CO₂ (Figure S16B). While assimilation rates can increase with leaf temperature within seconds to hours (e.g. Berry and Björkman, 1980); C₃ plants generally exhibit stable assimilation rates when acclimated to a range of growth temperatures (i.e., temperature homeostasis of photosynthesis, Yamori et al., 2014). With regards to the Franks et al. (2014) model, tests on six species grown at 20 and 28 °C air temperature show only a mild effect on the ability of the model to estimate CO₂ (Royer et al., 2018). For these reasons, we argue that changes in mean global temperature probably have little impact on the reliability of our CO₂ reconstructions.

2.4 Statistics

A one-sample Kolmogorov-Smirnov test identified that most of our inputs did not have normal distributions (Dataset S1). Thus, for our experiments, we used a two-sample Kolmogorov-Smirnov test to test for differences between CO₂ treatments in the inputs. All analyses were done within R and performed at the plant level.

3 Results and Discussion

3.1 Growth chamber experiments

The median CO₂ estimates for the three living species in the 500 ppm CO₂ treatment range from 584-686 ppm, and in the 1000 ppm treatment from 1016-1442 ppm (Figure 1; Table S2). Across all species, the 500 and 1000 ppm CO₂ treatments have a mean error rate \[ \frac{\text{|estimated CO₂-observed CO₂|}}{\text{observed CO₂}} \] of ~25% and ~19%, respectively. This is higher than elevated CO₂ experiments of Wollemia nobilis at 480 and 1270 ppm (7%; Franks et al., 2014), but is comparable to other paleo-CO₂ proxies at present day CO₂ such as alkenones (12.4%; Pagani, 2002), boron isotopes (8.2%; Henehan et al., 2013; Hönisch and Hemming, 2005), and pedogenic carbonates (67%; Ekart et al., 1999). Additionally, the precision of estimates within this study are comparable or better than other paleo-CO₂ proxies, especially at
elevated CO\textsubscript{2} (Beerling et al., 2009; Montañez et al., 2011; Royer, 2014). Using the generic values recommended by Franks et al. (2014) for A\textsubscript{0} and g\textsubscript{c(op)}/g\textsubscript{c(max)}, median CO\textsubscript{2} estimates increase for \textit{S. palustris} and \textit{O. cinnamomeum} while decreasing for \textit{C. deodara}, with a mean error rate of 44\% and 21\% for the 500 and 1000 ppm CO\textsubscript{2} treatments (Table S2). Note, however, that the generic values recommended by Franks et al. (2014) were obtained for field conditions which may differ slightly from growth chambers. Plants in growth chambers typically experience lower light and higher humidity, which affect A\textsubscript{0} and g\textsubscript{c(op)}/g\textsubscript{c(max)} via g\textsubscript{c(op)}.

\textit{O. cinnamomeum} and \textit{C. deodara} show no significant differences to CO\textsubscript{2} in SD, guard cell length, stomatal pore length, single guard cell width, and g\textsubscript{c(op)}/g\textsubscript{c(max)} (P>0.05), but both have significantly higher A at 1000 ppm CO\textsubscript{2} (P=0.03; P=0.02; Figure 2). SD in \textit{S. palustris} declines significantly by 21\% at high growth CO\textsubscript{2} (P=0.048), but with no significant change in guard cell length, stomatal pore length, or guard cell width (P>0.05). \textit{C. deodara} and \textit{S. palustris} exhibit a significant increase in c\textsubscript{i}/c\textsubscript{a} at elevated CO\textsubscript{2} (P=0.004; P=0.048), while \textit{O. cinnamomeum} does not.

The disparate physiological and morphological responses to CO\textsubscript{2} highlight an advantage of leaf gas-exchange proxies over other stomatal proxies. If SD or SI does not respond to CO\textsubscript{2}, then by definition the SD and SI methods cannot be used (see Reichgelt et al., 2016). For leaf gas-exchange models, this is not necessarily true if other inputs do respond to CO\textsubscript{2}. This is in fact the case with \textit{O. cinnamomeum} and \textit{C. deodara}, which produced reasonable CO\textsubscript{2} estimates for both treatments despite no changes in SD. Part of the issue with the other stomatal proxies is that they depend on a calibrated response, and the timescale associated with these responses (typically months to years) may not be sufficiently long, especially at higher-than-present CO\textsubscript{2} concentrations (Royer, 2001; see multi-year response from Hincke et al., 2016).
3.2 K-Pg boundary CO₂

The leaf gas-exchange estimates of CO₂ from *G. adiantoides* are similar for the Late Cretaceous (66.5 Ma; 624 ppm; 95% confidence interval 454-882 ppm) and early Paleocene (65.5 Ma; 630 ppm; 95% confidence interval 408-1181 ppm) (Figure 3; Table S2). The larger uncertainty with the Paleocene estimate is mostly due to having to model both stomatal pore length and single guard cell width because we were unable to measure them (Table S1). The leaf gas-exchange estimate of CO₂ from *aff.* *Stenochlaena* directly after the K-Pg boundary is 873 ppm (95% confidence interval 550-1414 ppm). By comparison, the estimates from Beerling et al. (2002) (updated by Beerling et al., 2009) are 539 ppm for the Late Cretaceous, >2300 ppm for the fern layer, and 343 ppm for the early Paleocene.

It is possible that all three of our estimates are falsely-high because the model overestimates present-day CO₂ for *G. biloba* (Barclay and Wing, 2016; Kowalczyk et al., 2018; but see Franks et al., 2014) and *S. palustris* at both 500 and 1000 ppm CO₂ (Figure 1). The relative temporal patterns, though, are more likely to be robust. If we use the generic inputs for $A_0$ and $g_{\text{c(op)}}/g_{\text{c(max)}}$ recommended by Franks et al. (2014), all three estimates increase by ~200-500 ppm (Table S2 and Figure 3), but the increase in CO₂ between the Late Cretaceous and fern layer does not change by very much (+250 ppm) and remains fundamentally different from the original interpretation of Beerling et al. (2002) (Figure 3).

A source of uncertainty for the aff. *Stenochlaena* CO₂ estimate is the atmospheric δ¹³C directly at the K-Pg boundary, which affects the calculation of $c_i/c_a$. Measured carbon isotopic excursions at the K-Pg boundary range from 0 to -3‰ (Arens and Jahren, 2000; Beerling et al., 2001; Maruoka et al., 2007; Schimmelmann and DeNiro, 1984; Schulte et al., 2010). Where examined in detail, the excursion in terrestrial sections begins immediately above the K-Pg
boundary clay in the fern spike interval, with the most negative values in the early phase of dicot recovery, and a return to pre-excursion values no higher than 2-3 m up section (reviewed in Upchurch et al., 2007). For our initial modeling we assume -2‰ (Text S2). If we instead assume an excursion of 0‰, comparable to the value at the top of the K-Pg boundary clay, or -3‰, the median CO$_2$ is 1170 and 762 ppm, respectively. Neither of these changed estimates strongly affect our key interpretations.

Our CO$_2$ record implies a transient change of ~+250; if we take the extreme scenario of comparing the lower and upper bounds of the 95% confidence intervals, this change could range from -333 to +1032 ppm. Critically, we provide the first fully-bounded CO$_2$ estimate from the top of the fern spike interval, and thus likely from within the first 10$^3$ years after the bolide impact. Our Ginkgo estimates bracket the event by roughly 500 kyrs, meaning that we do not know the CO$_2$ concentration directly before the bolide impact. This is an important deficiency because global temperatures rose ~300 kyrs before the K-Pg boundary and subsequently fell leading up to the boundary (Barnet et al., 2017; Nordt et al., 2003; Petersen et al., 2016; Wilf et al., 2003; Zhang et al., 2018). Zhang et al. (2018) estimate with the pedogenic carbonate proxy a CO$_2$ concentration of 700 ppm 110 kyrs before the K-Pg boundary (Figure S1), suggesting that Deccan volcanism caused an elevation in CO$_2$ before the boundary (Courtillot et al., 1986; Tobin et al., 2017 and sources cited within) and therefore the CO$_2$ spike we report may not be contributed entirely by the bolide-impact.

The Chicxulub bolide impact would release CO$_2$ almost instantaneously via the vaporization of target carbonate bedrock (Artemieva and Morgan, 2017; O'Keefe and Ahrens, 1989) and wildfires (Durda and Kring, 2004; Wolbach et al., 1990). A recent model for the vaporization of target carbonate bedrock at Chicxulub suggests a modest 54 ppm rise in
atmospheric CO₂ (Artemieva et al., 2017). Global wildfires may have caused CO₂ to increase by 315 ppm (Toon et al., 2016), but the extent of these fires is contentious and may have been far less (Belcher et al., 2003, 2004, 2005, 2009, 2015; Belcher, 2009; Harvey et al., 2008; Morgan et al. 2013).

Establishing a link between Deccan volcanism and CO₂ change at the K-Pg boundary is difficult because: 1) age uncertainties of the lava flows are on the order of 10⁴-10⁵ yrs (e.g., Renne et al., 2015; Schoene et al., 2015; Schoene et al., 2019; Sprain et al., 2019); and 2) constraining the amount and rate of CO₂ release is challenging (Jay and Widdowson, 2008; Self et al., 2006). Deccan volcanism clearly brackets the K-Pg boundary, but whether there was a pulse of activity within 10²-10³ yrs of the boundary is unresolved (Schoene et al., 2019; Sprain et al., 2019). Using existing constraints on the magnitude and pacing of CO₂ release for the Deccan, Tobin et al. (2017) demonstrate that it is possible, in principle, to raise CO₂ concentrations by several hundred ppm. Future work may provide clarity.

Temperature records spanning the first 10²-10³ yrs after the K-Pg boundary are sparse, but most modeling and high resolution marine data are not consistent with a large change in CO₂. After a brief “impact winter” (months to decades; Bardeen et al., 2017; Brugger et al., 2017; Taylor et al., 2018; Vellekoop et al., 2014, 2015, 2016), temperatures increased between ~1-6 °C depending on paleolatitude and geographic location, with the largest increases often at higher paleolatitudes (Macleod et al., 2018; Taylor et al., 2018; Vellekoop et al., 2014; Zhang et al., 2018). Terrestrial temperature trends inferred from leaf fossils are somewhat ambiguous and model dependent (Upchurch et al., 2007). Among marine records and most relevant to our study, Taylor et al. (2018) document a 2.5-4 °C warming during the fern spike interval in the southern mid-latitudes (present-day New Zealand). Together, these reconstructions best fit a scenario with
a modest 1-3 °C rise in global mean surface temperature. If we assume an Earth-system
sensitivity of 3 °C or higher per CO₂ doubling (Royer 2016), these records imply—at most—one
CO₂ doubling. One exception is a ~5 °C warming within ~100,000 yrs after the K-Pg boundary
at the global stratotype El Kef, Tunisia (~20 °N paleolatitude; MacLeod et al., 2018). This
subtropical temperature record appears incompatible with our record, suggesting that either CO₂
directly before the K-Pg boundary was substantially lower (<400 ppm) than what our and most
other reconstructions imply (Zhang et al., 2018; see also Figure S17) or local changes in ocean
chemistry biased the temperature estimates.

In summary, we find no strong evidence for a large pulse of atmospheric CO₂ coincident
with the K-Pg boundary. Our CO₂ record from within or directly above the fern spike is most
consistent with a CO₂ rise of no more than ~500 ppm and more likely ~250 ppm or less. This is
in keeping with the balance of evidence from temperature records and from the carbon cycle
modeling of impact vaporization of target bedrock, widespread wildfire, and Deccan volcanism.

Acknowledgements

We thank S. Sultan for use of her LI-COR gas-exchange analyzer, T. Ku for providing
laboratory space, and D. Penman and one anonymous reviewer for helpful comments. Data used
in this study is available in the Supplementary Information and DatasetSI.
References


Maxbauer, D. P., Royer, D. L., & LePage, B. A. (2014). High Arctic forests during the middle Eocene supported by moderate levels of atmospheric CO2, Geology, 42, 1027-1030. https://doi.org/10.1130/G36014.1


Figure 1. Atmospheric CO\textsubscript{2} estimates and probability density function using the leaf gas-exchange model of Franks et al. (2014) with Cedrus deodara (C.d.), Osmundastrum cinnamomeum (O.c.), and Stenochlaena palustris (S.p.), grown at two CO\textsubscript{2} treatments (500 and 1000 ppm CO\textsubscript{2}). Dotted lines represent the target CO\textsubscript{2} concentrations. Estimates are the median and 95% confidence interval.

Figure 2. Measured inputs for Cedrus deodara (C.d.), Osmundastrum cinnamomeum (O.c.), and Stenochlaena palustris (S.p.) grown at two CO\textsubscript{2} concentrations (500 and 1000 ppm) and fossil Ginkgo adiantoides (G.a.) and aff. Stenochlaena. Abbreviations: K, Cretaceous; Pg, Paleogene. For multiple comparisons different letters indicate significantly different values at the 0.05 level.
* P ≤ 0.05, ** P ≤ 0.01, *** P ≤ 0.001.

Figure 3. Atmospheric CO\textsubscript{2} estimates from the Cretaceous-Paleogene boundary. Estimates from the leaf gas-exchange model (this study) are based on the same fossils whose stomatal index was used to estimate CO\textsubscript{2} by Beerling et al. (2002). The gray squares are based on the recommended values from Franks et al. (2014) for assimilation rate and the ratio of operational to maximum stomatal conductance. Error bars represent the 95% confidence interval.
Figure 1.
Figure 3.
Atmospheric CO₂ (ppm) vs Age (Ma)

- **Fossil model (this study)**
- **Fossil model** (Franks et al. (2014) defaults)
- **Beerling et al. (2002)** (Updated by Beerling et al. (2009))

**Density**

- 0.0050
- 0.0025
- 0.0000

**Atmospheric CO₂ (ppm)**

- 3000
- 2000
- 1000
- 500
- 300

**Age (Ma)**

- 67
- 66
- 65