Estimation of past atmospheric carbon dioxide levels using tree-ring cellulose $\delta^{13}C$

Trina Bose$^{1,*}$, Supriyo Chakraborty$^1$, Hemant Borgaonkar$^1$, Saikat Sengupta$^1$ and R. Ramesh$^2$

$^1$Indian Institute of Tropical Meteorology, Pune 411 008, India
$^2$Physical Research Laboratory, Ahmedabad 380 009, India

We study the applicability of the Farquhar model for carbon isotopic discrimination (change in carbon isotopic composition from air CO$_2$ to tree-ring cellulose) in C$_3$ plants to trees growing in the field. Two new carbon isotope datasets from Himalayan conifers with published data from another eight sites across the world show disparate trends in the plot of carbon isotope discrimination versus atmospheric carbon dioxide concentration, in contrast to the model prediction of absence of any trend. This is because the model assumes that the tree adjusts its stomatal conductance for water-use efficiency to maintain a constant ratio of carbon dioxide concentrations inside and outside the leaf and treats the diffusive and biochemical fractionation factors as constants. By introducing a simple linear dependence of these fractionation factors with ambient temperature and humidity, we have enhanced the applicability of the model to naturally growing trees. Further, despite the disparate trends exhibited by the 10 trees, we show using the inverse modelling that it is possible to derive a unique record of past atmospheric CO$_2$ concentrations using tree cellulose $\delta^{13}C$ data. The reconstructions also replicate the summer pCO$_2$ gradient from tropics to mid-latitudes. We also discuss the merits and demerits of the model, and compare the model-derived pCO$_2$ with that of the ice core-based records from Law Dome.

Keywords: Atmospheric carbon dioxide, carbon isotope, cellulose, climate change, tree ring.

The recent progressive increase of the atmospheric carbon dioxide concentration ($c_a$) has earned considerable attention due to its role in global climate change. This has necessitated the identification of sources and sinks of this atmospheric gas. Although a few sources and sinks have been identified from the instrumental data$^4$, such studies are limited due to short-time range of the dataset. The only long-term (1800–1978 CE) annual $c_a$ data that can be validated against instrumental observations are available from ice cores of Law Dome, Antarctica$^5$. The preceeding ice core data with resolution of 2–10 years stop at 1939 (ref. 3). There are some pCO$_2$ data from the Greenland ice cores which reach mid-1900s (ref. 4), but they have very low resolution for comparison in annual scale. The Law Dome ice core data$^5$ is used in this study for comparison because it provides better resolution and maximum overlapping period with instrumental datasets relative to other ice-core records.

In view of the modern pCO$_2$ distribution and fluxes (Rödenbeck et al.$^1$ and references therein), it is apparent that the latitudinal distribution of pCO$_2$ and the difference in various fluxes between the continents cannot be studied using the ice-core data alone. This is because there are no reconstructed CO$_2$ datasets representing the tropics and sub-tropics before 1950s. The paucity of high-resolution data in space and time limits the application of ice-core CO$_2$ data for studying the past carbon fluxes$^1$.

One of the main sinks of atmospheric CO$_2$ are the trees that comprise ~95% of land biomass$^3$. The long-living (up to a few millennia) trees fix carbon in ring cellulose, and are datable to annual to sub-annual resolution$^6$. So stable carbon isotope ratios of cellulose ($\delta^{13}C_c$) may be an independent alternative for the determination of $c_a$. This has been hampered by the dependence of $\delta^{13}C_c$ on multiple environmental and tree-specific parameters$^8$. Carbon isotope records of tree rings have been successfully used for a variety of palaeoclimate and environmental studies$^9$–$^{17}$ including their relationship with pCO$_2$, temperature and humidity$^{18,19}$. However, these studies stop short of reconstructing $c_a$ time series from $\delta^{13}C_c$ of trees growing in natural environments.

Carbon isotope fractionation in trees grown in controlled environments was investigated in detail by Farquhar et al.$^{20}$ in the early 1980s, who formulated a model for calculating the carbon isotope discrimination ($d$) in terms of biophysical parameters of the tree and its surroundings, such as pCO$_2$ and $\delta^{13}C$ of air ($c_a$ and $\delta^{13}C_a$) and intercellular CO$_2$ ($c_i$) in plant materials. The model suggests that $c_i$ is not proportional to $d$ due to the plants adjusting stomatal conductance/density to optimize water-use efficiency so as to maintain a constant ratio of pCO$_2$ in intercellular spaces relative to atmosphere ($c_i/c_a$). Their model, however, does not address possible effects of a sharp rise in atmospheric pCO$_2$ as seen in last few decades, on $d$.

The effect of rising pCO$_2$ was shown experimentally by Evans and Von Caemmerer$^{21}$. They showed that diffusion
of CO₂ from the atmosphere to the confines of the leaf may be considered as that from an infinite reservoir to a finite one. Hence, while c₁ should be linearly proportional to initial values of cₑ, it would saturate at higher values of cₑ when other factors remain constant²¹. Hence, the plot of cᵢ/cₑ versus cₑ should initially remain constant and then fall rapidly with cₑ. By this consideration, the cᵢ/cₑ versus cₑ (or d versus cₑ) plot should show a negative slope at higher values of cₑ.

Here, we examine whether the said relationship holds true for naturally growing trees as well. For this, we consider pre-industrial δ¹³Ccₑ to be constant and decreasing continuously since the advent of the industrial revolution²², as a result of fossil-fuel emission. The relationships between d and other meteorological parameters are also studied in this context. Our study shows that Farquhar model does not apply to all cases of natural trees. We further demonstrate that inclusion of variability of meteorological parameters also improves the applicability of the Farquhar model. The modified model is calibrated for the period 1978–2005 CE and tested for congruency with observations during 1958–1977 CE (testing period). Finally, cₑ has been reconstructed for the period 1902–1957 CE, and compared with Law Dome ice core pCO₂ data. This reconstruction fills the pCO₂ data gap in the sub-tropics and tropics from 1958 to 1901. Till now, there were no means to study the summer pCO₂ gradient between the tropics and the mid latitudes, which is reproduced using our reconstruction.

Data choice and testing

To apply the model of Farquhar et al.²⁰ to trees from natural settings, we calculated the slope of d versus cₑ using δ¹³Ccₑ data of such trees. Additionally, we needed to study the effects of meteorological parameters on δ¹³Ccₑ. Such data collected from various sources mentioned below.

Tree ring cellulose δ¹³C

Two of the testing datasets (carbon isotopes) were generated in the present study. These samples were collected from the Western Himalayas; (i) Kanasar, Uttarakhand, India (Cedrus deodara), (ii) Kothi, Himachal Pradesh, India (Picea smithiana) and dated by dendrochronological methods²³,²⁴ before isotopic analysis. The wider rings were separated manually using hammer and chisel, while the thinner rings were separated using a sledge microme- tome (Fritz Hans Schweingruber, Switzerland)²⁵. Cellu- lose extraction was carried out following Leavitt–Danzer method⁷, with minor modifications. The samples were analysed at the Stable Isotope Laboratory, IITM, India using a Delta-V Plus isotope ratio mass spectrometer (Thermo Fisher Scientific) equipped with a Flash EA 1112 Elemental Analyser in continuous flow mode. IAEA primary reference materials [CH₃: δ¹³C = -24.724‰ VPDB and CH₆: δ¹³C = -10.449‰ VPDB] were used for calibration. Standard deviation of the IAEA CH₃ was ± 0.02‰ and that of IAEA-CH₆ was ± 0.08‰. The routine precision based on sample replicates of our measurement was ± 0.1‰. All values are reported against VPDB scale and presented in Figure 1. The black and grey curves show δ¹³C variations of C. deodara and P. smithiana respectively.

Additionally, we have used tree-ring δ¹³Ccₑ data from eight stations across the globe¹²–¹⁷. Sample sets with site locations, code names and lengths of these records are shown in Figure 2 (see also Table 1).

Atmospheric carbon dioxide data

Atmospheric pCO₂ (1957–2008 CE) and δ¹³C of atmospheric CO₂ (1977–2008 CE) data from 11 Scripps Institution of Oceanography (SIO) stations were used in the present study²⁶. For testing the latitudinal gradient we have also used CO₂ data from Cape Ferguson, Australia (http://cdiac.ornl.gov/trends/co2/csiro/csiro-cferg.html). The Australian site was chosen because the SIO sites at similar latitudes were island sites.

The annual average cₑ for all sites shows very low spatial variability (<2 ppm maximum standard deviation, i.e. maximum of spatial standard deviation among all 11 site annual averages and all years). However, the annual min- imum (corresponding to maximum photosynthesis by plants) and maximum (indicating maximum decomposi- tion/plant respiration) values are characterized by higher spatial variability (<4 ppm). The maximum monthly spatial
Table 1. For all sites, chosen values of h for the reconstruction are shown with the corresponding $P_r$ (average absolute difference between observed global annual average $c_a$ and model output $c_{model}$) and $max\sigma$ (maximum reconstruction error between all time steps). The h values of similar species do not differ significantly even with large geographical separation, e.g. the Pinus from India (KOH) and Siberia (MRUS) show similar h values. Samples from nearby sites with different genera, e.g. CASW and VISW show different h values.

<table>
<thead>
<tr>
<th>Site code</th>
<th>Longitude (°)</th>
<th>Latitude (°)</th>
<th>Altitude (m amsl)</th>
<th>Species</th>
<th>Reference</th>
<th>Data range (CE)</th>
<th>$h$ (%/ppm)</th>
<th>$P_r$ (ppm)</th>
<th>$max\sigma$ (ppm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>PMPE</td>
<td>-69</td>
<td>-13</td>
<td>265</td>
<td>Cedrela odorata</td>
<td>16</td>
<td>1820–2004</td>
<td>0.31</td>
<td>1.16</td>
<td>1.16</td>
</tr>
<tr>
<td>BWA</td>
<td>-118</td>
<td>37</td>
<td>3200</td>
<td>Pinus longaeva</td>
<td>17</td>
<td>1085–2005</td>
<td>0.46</td>
<td>0.91</td>
<td>0.99</td>
</tr>
<tr>
<td>MRUS</td>
<td>60</td>
<td>68</td>
<td>40</td>
<td>Picea obovata</td>
<td>12</td>
<td>1905–1996</td>
<td>0.42</td>
<td>0.78</td>
<td>0.56</td>
</tr>
<tr>
<td>LOSW</td>
<td>8</td>
<td>46</td>
<td>2000</td>
<td>Larix decidua</td>
<td>15</td>
<td>1650–2004</td>
<td>0.35</td>
<td>1.18</td>
<td>1.47</td>
</tr>
<tr>
<td>CASW</td>
<td>9</td>
<td>46</td>
<td>900</td>
<td>Quercus petraea</td>
<td>13</td>
<td>1637–2002</td>
<td>0.38</td>
<td>0.76</td>
<td>0.79</td>
</tr>
<tr>
<td>VSW</td>
<td>8</td>
<td>48</td>
<td>1400</td>
<td>Pinus sylvestris</td>
<td>13</td>
<td>1675–2003</td>
<td>0.27</td>
<td>1.29</td>
<td>1.33</td>
</tr>
<tr>
<td>BAGR</td>
<td>75</td>
<td>36</td>
<td>2900</td>
<td>Juniperus excelsa</td>
<td>14</td>
<td>1900–1998</td>
<td>0.24</td>
<td>0.58</td>
<td>0.59</td>
</tr>
<tr>
<td>BOIB</td>
<td>75</td>
<td>37</td>
<td>3900</td>
<td>Juniperus turkestanica</td>
<td>14</td>
<td>1950–1998</td>
<td>0.3</td>
<td>0.71</td>
<td>0.71</td>
</tr>
<tr>
<td>KAUT</td>
<td>78</td>
<td>31</td>
<td>2200</td>
<td>Cedrus deodora</td>
<td>Present work</td>
<td>1925–1993</td>
<td>0.64</td>
<td>1.45</td>
<td>0.60</td>
</tr>
<tr>
<td>KOHP</td>
<td>77</td>
<td>32</td>
<td>2500</td>
<td>Picea smithiana</td>
<td>Present work</td>
<td>1920–1999</td>
<td>0.45</td>
<td>1.16</td>
<td>0.79</td>
</tr>
</tbody>
</table>

Figure 2. Location map of the ten tree-ring sites whose $\delta^{13}C$ records have been used in this study. Table 1 provides information on site name, longitude, latitude, altitude, name of the species and duration of $\delta^{13}C$ time series for all sites.

The variability of $c_a$ among the stations is ~4.1 ppm, which possibly corresponds to local differences in peak productivity periods.

Like $c_a$, $\delta^{13}C_a$ also shows very low spatial variability in the annual average (<0.11‰). But the monthly spatial variability maximum for $\delta^{13}C_a$ is 0.23‰. The maximum range in any of the SIO stations is at Point Barrow, Alaska (for all years: average 0.87‰, maximum 0.98‰, minimum 0.74‰) followed by Alert, NWT, Canada (for all years: average 0.79‰, maximum 0.94‰, minimum 0.69‰). The lowest range is seen in American Samoa (for all years: average 0.08‰, maximum 0.16‰, minimum 0.04‰) followed by South Pole (for all years: average 0.79‰, maximum 0.94‰, minimum 0.69‰) and Kermadec Islands (for all years: average 0.79‰, maximum 0.94‰, minimum 0.69‰). This range represents the fact that in the productive period ($c_i$, minimum), trees show a preference towards $^{12}C$, thus leaving enriched in $^{13}C$; this extra $^{13}C$ is released back in air during decomposition making atmospheric CO$_2$ isotopically lighter. Further, Levin et al. showed that this range can vary even among nearby places and depends on altitude or nearby plant concentration.

For input to the model for the time period 1901–2005 CE, a quadratic fit (the linear correlation coefficient, $r = 0.99$) of the instrumental $\delta^{13}C_a$ (global annual average) with time (years) was used to extend the $\delta^{13}C_a$ values back in time with maximum value set at the pre-industrial value of ~6.4% (ref. 22). Global annual average of $\delta^{13}C_a$ was used because of non-availability of $\delta^{13}C_a$ data for most tree-ring sites (nearest available $\delta^{13}C_a$ data are ~500 km from BWA) and no clear pattern of $\delta^{13}C_a$ annual range depending on latitude.

$P_r$ versus $c_a$ slope estimation and analysis

As instrumentally observed $\delta^{13}C_a$ data is only available after 1977 CE, the $d$ and $c_a$ relationship is studied for 1978–2005 CE using the observed values. Figure 3 (details in Table 2) shows that at two locations, $d$ and $c_a$ have no significant correlation ($r = -0.12$ for KOHP to 0.26 for BWA with $P > 0.19$). The remaining eight locations show positive trends ($r = 0.53$ to 0.93; $P < 0.02$). Similar positive slopes were also reported for plants grown under controlled environment. These trends are a serious departure from the model of Farquhar et al. The observed positive trend cannot be explained by the model unless we consider the two fractionation factors $a$ (for diffusion) and $b$ (for carboxylation) to be controlled by meteorological parameters, such as temperature and humidity.

Meteorological data

To test the effect of meteorological parameters on $a$ and $b$, we use the Climate Research Unit (CRU) datasets as: (i) they have the highest resolution; (ii) they are derived mostly from observations that represent the complex site topography better; (iii) they are available at all the site...
locations relatively for longer time duration (1901–
2008 CE) compared to other datasets. Vapour pressure ra-
ther than precipitation is taken as representative of the
moisture here. This is because evapotranspiration from
leaves rather than the amount of soil moisture decides the
 stomatal opening, which in turn decides the rate of
diffusion of CO₂ into the leaf.

The modified model

**Modifications**

Isotopic fractionation during carbon fixation is mainly
governed by two processes: (i) gaseous diffusion through
stomata of leaves; (ii) the chemical process of carboxyla-
tion and deposition. Considering these fractionation
pathways, Farquhar et al.\textsuperscript{20} suggested that net \(^{13}\)C change
from air to cellulose \((d)\) can be estimated as

\[
d = \delta^{13}C_a - \delta^{13}C_c = a + (b - a) \frac{c_i}{c_a} - B \frac{c_i - c_a}{c_a}, \tag{1}
\]

where \(a\), \(b\) and \(B\) are fractionation due to diffusion of CO₂
from air to leaves, carboxylation in the leaves forming
sugar, and transport of CO₂ from intercellular spaces of
leaves to the carboxylation respectively. \(c_i\) and \(c_a\) are
partial pressure of CO₂ in the intercellular spaces of leaves
and in the site of carboxylation. It is to be noted that
Farquhar et al.\textsuperscript{20} treated \(b\) and \(B\) as equal and same, but
we have written them as different for process-based
differentiation.

We propose the following modifications to apply the
model of Farquhar et al.\textsuperscript{20} for trees growing in the field.

1. The variables \(c_i\) and \(c_a\) are not measurable in these
cases and need to be estimated by some alternative
means.

   (i) \(c_i\) is about 30% lower than \(c_i\) for many species
when leaves are actively photosynthesizing in high irradiance\textsuperscript{21}. However, this value of 30% will
not be applicable to low irradiance zones. Let \(c_i\) to be about \(x\%\) lower than \(c_i\) for all cases.

   Hence \(c_i \approx (1 - \frac{x}{100}) c_i\) or

   \[
   \frac{c_i - c_a}{c_a} \approx \frac{x}{100} \frac{c_i}{c_a}.
   \]

   So, eq. (1) can be modified as

   \[
d = a + (b - a) \frac{x}{100} \frac{c_i}{c_a}. \tag{2}
\]
Now if we define $b^*$ as the fractionation due to the total chemical process, i.e. $b^* = b - \frac{x}{100}B$ we get

$$d = a + (b^* - a) \frac{c_i}{c_a}, \quad (3)$$

(ii) Since $c_i$ is unknown for natural sites, this term is eliminated by considering $c_i/c_a$ as a simple sigmoidal function of $c_a$, following Evans and Von Caemmerer\cite{EvansVonCaemmerer1978}. This particular function was chosen because it mimics the required variation almost exactly and has a one-to-one correspondence with $c_a$. As the pCO$_2$ concentration in water controls pH values in most natural water bodies through carbonate and bicarbonate ions\cite{Caemmerer1981}, $c_i$ should also be taken as a proxy for the pH of leaf water in the long term (i.e. annual).

$$\frac{c_i}{c_a} = \frac{\gamma}{1 + e^{(a-d-c_i)/100}} = \gamma z, \quad (4)$$

where $sl$ is the level of $c_a$ at which diffusion saturates and $\gamma$ is the constant of proportionality between $c_i/c_a$ and $z$.

2. As discussed above, the reason for some stations showing positive instead of no slope predicted by Farquhar et al.\cite{Farquhar1989}, that $a$ and $b^*$ are not constant but vary with meteorological parameters. Temperature and relative humidity dependence of leaf $\delta^{13}$C was studied in the controlled condition by Edwards et al.\cite{Edwards1998}. Based on this premise, we attempt to modify these factors ($a$ and $b^*$) as:

(i) Diffusion as a physical process should be dependent on temperature (thermal energy provides kinetic energy which translates into velocity) and moisture content of air (which decides the aperture of the stomata). Diffusion is defined as random movement with resultant directional displacement (mean free path), which is governed by the cross-sectional area, thermal energy and concentration gradient. The diffusion coefficient for gases is inversely proportional to square root of the molecular mass and directly proportional to the cube of square root of temperature\cite{Edwards1998}. Further, this movement is integrated over the minimum cross-sectional area in the path. So theoretically, there should be temperature and humidity dependence on isotopic fractionation due to diffusion. Hence, in the relatively small natural temperature range we consider $a$ as a linear function (to a first approximation) of absolute temperature ($T$) and vapour pressure ($e$) with parameter $r_1$ assigned to properties independent of meteorological parameters.

$$a = a_1 T + a_2 e + r_1. \quad (5)$$

(ii) Carboxylation and deposition being biochemical processes are temperature-dependent\cite{Caemmerer1981}. So $b^*$ is considered a linear function (to a first approximation) of $T$. Properties independent of meteorological parameters are represented by $r_2$.

$$b^* = b_1 T + r_2. \quad (6)$$

There are other factors known to affect cellulose formation in trees such as light intensity, metal ion concentration, etc. But, no long-term data of these is available for calibration in this reconstruction. Hence, these effects should be considered to be included in the parameters $r_1$ and $r_2$.

Substituting eqs (4)–(6) into eq. (3) we get

$$d = a_1 T + a_2 e + r_1 + (b_1 - a_1) \gamma Tz - a_2 \gamma ez + (r_2 - r_1) \gamma z, \quad (7)$$

**Estimation of model coefficients and error**

Assuming $Tz = x$, $ez = y$, $(b_1 - a_1) \gamma = Q$, $a_2 \gamma = -R$, $(r_2 - r_1) \gamma = S$ in eq. (7), we get an equation consisting of six variables. This equation is nonlinear, but is linearized by substitution of nonlinear combined variables.

$$d = a_1 T + a_2 e + r_1 + Qx + Ry + Sz. \quad (8)$$

For an accurate estimation of the coefficients of the above equation, we use inverse modelling for the over-determined system in a box of size $n$ (>$6$) for all time series variables, $d$, $T$, $e$ and $z$ (ref. 35).

$$D = MX \Rightarrow M^tD = M^tMX \Rightarrow X = \frac{M^tD}{M^tM}, \quad (9)$$

where $D = [d_i]$ is the column matrix of the $d$ values of a certain tree, $M = [T_i, e_i, 1, x_i, y_i, z_i] = 6 \times n$ matrix given by known variables, where $i = 1$ to $n$ is the number of rows representing years, and $X = [a_1, a_2, r_1, Q, R, S]^t$ is the column matrix of unknown constants to be evaluated.

Let $N_j$ be the number of years a particular station $j$ has data within the calibration period 1978–2005 CE. Hence upon estimation of the coefficients in the sliding box of size $n$, we get $(N_j - n)$ sets of coefficients. These sets of coefficients are then compared for internal variance. To have a significant comparison in this case and the minimum possible error in the coefficient estimation, we ran many trials to decide on $n = 10$ for all the stations. The calibration was thus achieved by sliding this 10-year
window through the period of data availability \((N_r, \text{ which varied from a minimum of 16 to a maximum of 27})\). Matrix \(X\) was estimated using Moore–Penrose pseudo-inverse of matrices\(^{35}\) and the software GNU Octave for \((N_r - n)\) groups of data. All runs were made with annual averages of \(T\) and \(e, d\) was calculated using the annual fitted \(\delta^{13}\)C.

Also from eq. (7), we get
\[
z = \frac{d - (r_1 + a_1T + a_2e)}{[(b_1 - a_1)T - a_2e + (r_2 - r_1)]y}
\]
(10)
\[
= \frac{K}{L} \text{ say;}
\]
then from eq. (4)
\[
c_a = sl - 100 \ast \ln \left(\frac{1}{z} - 1\right).
\]
(11)

Equations (7) and (10) are respectively, the forward and inverse sides of the model. For \((N_r - n)\) sets of matrix \(X\) found from eq. (8), \(z\) was estimated (\(\hat{z}\)) using eq. (10) and the \(X\) giving minimum error in \(z\) (i.e. \(\sum(z - \hat{z})^2\)) was selected for reconstructing \(\hat{z}\) in the total period (1901–2005 CE) using eq. (10). These \(z\)-values are converted to \(c_a\) using eq. (11). Subsequently, the reconstructed \(c_a\) is compared with the observed \(c_a\) values in the testing period (1958–1977 CE).

Using error analysis theorems\(^{36}\) on eq. (10) and considering \(\delta^{13}\)C\(_a\), \(\delta^{13}\)C\(_i\), \(T\) and \(e\) as mutually dependent, we get
\[
\sigma_k^2 = \sigma_{\delta^{13}C_a}^2 + \sigma_{\delta^{13}C_i}^2 + a_1^2\sigma_T^2 + a_2^2\sigma_e^2 + 2a_1\sigma_{\delta^{13}C_a}\sigma_{\delta^{13}C_i} + 2a_2\sigma_{\delta^{13}C_a}\sigma_e + 2a_1\sigma_{\delta^{13}C_i}\sigma_T + 2a_2\sigma_{\delta^{13}C_i}\sigma_e + 2a_1a_2\sigma_T\sigma_e
\]
\[
\sigma_l^2 = \gamma^2[(b_1 - a_1)^2\sigma_T^2 + a_2^2\sigma_e^2 - 2(b_1 - a_1)a_2\sigma_T\sigma_e].
\]
\[
\frac{\sigma_k^2}{z^2} = \frac{\sigma_l^2}{K^2} \times \frac{\sigma_l^2}{L^2} - 2\frac{\sigma_k^2}{KL}
\]
(12)

Also, from eq. (11):
\[
\sigma_{c_a}^2 = \frac{100\gamma^2}{z(1 - z)}.
\]
(13)

where \(\sigma_i^2\) and \(\sigma_{ij}^2\) respectively, represent the variance of parameter \(i\) and the covariance of parameters \(i\) and \(j\).

The reconstruction model error (eq. (13)) is assigned at every time step for all the reconstructed \(c_a\) values. It is to be noted here that the error calculated in eq. (13) is the mathematical error which includes all the variabilities of the input parameters.

In this context it may be mentioned that we have considered globally averaged value of \(\delta^{13}\)C\(_a\) for the following reason. \(\delta^{13}\)C\(_a\) does show significant variability on seasonal timescale, especially in the high latitudes, but the annually averaged value has low variability. Hence we tried to quantify this effect by estimating the error on reconstructed \(c_a\) by taking the maximum variance of \(\delta^{13}\)C\(_a\) as 1%. This did not change the \(c_a\) variance significantly (\(<0.5\ \text{ppm})\).

On the other hand, the instrumental error in observed \(T\) and \(e\) is very small (which we cannot expect prior to 1901). We have tried to generalize the error analysis by taking standard deviation of data as error and not the given error (from CRU). But, the gridded data had low standard deviation or covariance with each other or the isotope data, which led to the \(\pm 1.5\ \text{ppm model error (Table 1). Hence, unless any of the input datasets has a high standard deviation or covariance, we are likely to get similar levels of error.}

**Testing of the modified model**

We have used two parameters to judge the reconstructed \(c_a\): (i) \(P_r\), the average absolute difference between observed global annual average \(c_a\) and model output \(c_a\) during the testing period (1957–1976) and (ii) \(max\sigma\), the maximum reconstruction error among all time steps. The modified model (eqs (7) and (10)) was tested for sensitivity to input \(\delta^{13}\)C\(_a\) data. For example, Figure 4 describes this for four different treatments of \(\delta^{13}\)C\(_a\) for the site KOHP (chosen because it is the longer of the time series

![Figure 4](image-url)

*Figure 4.* Time series of atmospheric \(c_a\), reconstructed from *P. smithiana* from Kothi, India (KOHP – grey lines) using our model, for four cases of pre-processing of \(\delta^{13}\)C: (i) Raw data; (ii) Smoothed by 30 years spline; (iii) Corrected using a fitted correction factor \((c_f = \delta^{13}\)C\(_a + hc)\) with \(d = -6.4\text{‰} \delta^{13}\)C\(_a\); (iv) A combination of (ii) and (iii). The instrumentally observed \(c_a\) is shown by black line.
generated in this work). For comparison, we have also included in this figure the global annual average of observed \( c_s \) (2008–1957 CE). The same plot for the site LOSW is given in the Supplementary Information (Figure S1; see online). The following cases were tested:

(i) No treatment done on \( \delta^{13}C_s \); the reconstructed \( c_s \) data do not show a good fit for the calibration period. The gaps in the reconstruction are due to the 5 ppm model error cap.

(ii) The data have been corrected for variability in tree growth pattern (e.g. juvenile/ canopy effects) with a spline function at various levels. Again, such input did not yield a better match with the observations.

(iii) Atmospheric pCO\(_2\) rise due to anthropogenic activities affects plant \( \delta^{13}C_s \) in two ways: (a) burning of fossil fuels depleted in \(^{13}\)C leads to plants receiving pCO\(_2\) with lowered \( \delta^{13}C_s \) values; (b) The increase in \( c_s \) causes further discrimination against the heavier isotope (may be due to the change in pH of leaf water). Feng and Epstein\(^{37}\) suggested a correction factor \( c_i = \delta^{13}C_s + h c_s \) to be applied to the \( \delta^{13}C_s \) to compensate for these effects, where, \( h (\% / \text{ppm}) \) is the decrease in \( \delta^{13}C_s \) due to (anthropogenic) increase in \( c_s \). This correction factor \( c_i \) is calculated for the calibration years (with annual average \( c_s \) and \( \delta^{13}C_s \) assuming \( h \) values (varied) and a quadratic fit was made as a function of years to extend it to 1901. This fitted value was used to correct \( \delta^{13}C_s \) values for the years 1901–2005 CE (ref. 14). As \( \delta^{13}C_s \) is already used in the correction, pre-industrial value of \( \delta^{13}C_s = -6.4\% \) is used for this case for calculating \( d \). The \( h \) value was varied from 0 to 1.4 with increment 0.001 to compensate for high \( P_t \) from earlier levels of \( h \) used (<0.05)\(^{14,37\%}. \) \( h \) value used in the final reconstruction was determined by minimizing \( P_t \). It is emphasized here that use of the observed \( c_s \) was limited to the calibration period alone and for the reconstruction and testing periods we have not used any observed \( c_s \) values.

(iv) \( \delta^{13}C_s \) was spline corrected to 30 years to remove the effect of tree growth and then subjected to case (iii) correction. This yielded a smoother \( c_s \), which in turn resulted in the least value of \( P_t \).

From each site \( \delta^{13}C_s \) data we reconstructed \( c_s \) and these data (Figure 5) with associated uncertainties are given in http://doi.pangaea.de/10.1594/PANGAEA.800072. The reconstructed data show that despite disparate trends observed in the original tree \( \delta^{13}C_s \) datasets, our method way of accounting for the temperature and vapour pressure effects at each site results in mutually concordant \( c_s \) values.

**Uncertainty estimation**

Further, sensitivity analysis was done by varying \( h \) value from 0.001 to 1.4 in increment of 0.001 and studying the resultant pattern of \( P_t \) and \( mxr\sigma \). For all stations, the \( h \) versus \( P_t \) plots (Figure 6) show that \( P_t \) decreases exponentially with increasing \( h \), but characterized with some intermittent variations. The \( h = \text{mxr}\sigma \) plot, in contrast is more irregular and \( \text{mxr}\sigma \) approaches zero for high \( h \) but ends with a few spikes. Since we cannot expect error values to vary widely from near zero to very high values due to small perturbations in \( h \), such values of \( h \) have been avoided. Instead, the value of \( h \) that yields both \( P_t \) and \( \text{mxr}\sigma \) lower than 1.5 ppm and close to each other is selected for reconstruction (Table 1). This makes \( P_t \) and \( \text{mxr}\sigma \) equivalent in magnitude.

Figure 7 shows the sensitivity on reconstructed pCO\(_2\) for different values of \( h \) for the site KOHP. Lower values (0.2 and 0.25) of \( h \) (light grey) show very high variability in the testing period (1958–1977). This leads to high values of \( P_t \), as seen in the Figure 6 for the same site. If we decrease \( h \) further, then \( P_t \) will increase exponentially (Figure 6). But the change in variability is minimal in the range \( h = 0.3 \) to 0.65 (dark grey) and an optimum value is obtained at \( h = 0.45 \) (marked with filled dots). This shows that in this range the model is not sensitive to the variation in \( h \) and we can use any of these values without losing much accuracy of reconstruction.

We varied the saturation levels (sl) between 200 and 1000 ppm and calibrated for the minimum error in \( \Delta \left( \sqrt{\sum(z-\bar{z})^2} \right) \). Among all samples, only PMPE gave resulting saturation level of 306 ppm. The rest showed a inverse relationship between sl and the error. This might be due to: (i) the effect of temperature rise on conductance in the leaf\(^{38\%}; \) (ii) the sites showing inverse relationship are all conifers and the saturation behaviour in conifers is different than what is assumed here; this can be confirmed by inputing \( \delta^{13}C \) data from more non-conifers in the future.

![Figure 5. Reconstructed pCO2 data for all stations compared with the instrumentally observed data (solid black) and ice-core data from Law Dome (dotted black).](Image 50x120 to 290x295)
In all analyses, the maximum allowed estimation error was 5 ppm and the breaks seen in Figure 4(i) and (ii) are due to higher analytical errors at those points. It is to be noted that as the maximum error in reconstruction, i.e. \( P_r \) or \( mxr\sigma \) is \(~1.5\) ppm, the reconstructions are to be considered accurate within this uncertainty (Table 1).

**Comparison with other datasets**

To test the reliability of the reconstructions, \( c_a \) data were compared with those from instrumental and ice-core datasets. The maximum spatial standard deviation of reconstructed \( c_a \) averaged globally at any time step yields...
~5 ppm, which is comparable to ~4 ppm values for the observed data. The difference between global annual average observed \(c_a\) and reconstructed \(c_r\) at any site has a mean ~0.31 and standard deviation ~1.37 ppm with skewness ~0.21, showing that reconstruction errors are primarily random.

Assumption of a smooth global curve for \(\delta^{13}C_a\) may be the cause of the wiggles in the reconstructed \(c_r\), as variations in the observed \(\delta^{13}C_a\) are not as smooth as those of observed \(c_a\). This is the reason \(P_t\) cannot be decreased below a threshold. Many of the wiggles are coherent, which indicates that the value of \(\delta^{13}C_a\) assumed in that period may not be strictly valid. This can possibly provide corrections for the \(\delta^{13}C_a\) curve by backtracking, as this information is unavailable in high-resolution ice-core data.

The applicability of our model may be verified by comparing the observed pCO\(_2\) gradient (between the tropics and the mid latitudes) with that of the model-derived data. As only one of the sites (PMPE) is in the tropics, we depict this gradient by plotting its values with respect to the reconstructed value for the site BWCA (Figure 8). For comparison with observations, we selected two datasets in comparable latitudes: (a) Cape Fergusson, Australia (CSIRO) and (b) La Jolla Pier, California, USA (SIO). The two grey points represent the mean value for the months of June during 1992–2006 for these sites. The mean (1902–1957) of the reconstructed data is given in black colour. This plot compares that the latitudinal gradient in observed (slope: ~0.146 ppm/°; intercept: 362.977 ppm) and reconstructed data (slope: ~0.150 ppm/°; intercept: 307.830 ppm). A good agreement between the observed and reconstructed slope gives credence to our model.

Comparison with the ice-core data from Law Dome (Figure 9) shows that the difference (\(D\)) between pCO\(_2\) derived from ice core and all tree-ring sites remains mostly 0 ± 4 ppm till 1960s; prior to this it rises until 1931–1940 CE with different amplitudes for different sites. This amplitude of \(D\) correlates well with latitude (\(r = 0.55, P < 0.10\), implying its dependence on the distance from Antarctica. Also, the adjacent sites show similar amplitude values. The maximum \(D\) observed was ~12.5 ppm at BWCA. Before 1931, it decreases gradually to reach a near-zero level around 1910. In some sites it goes below ~5 ppm at ~1908 (e.g. PMPE site). We rule out the possibility of any model bias being responsible for the \(D\) values because: (a) the difference is not seen throughout the reconstruction period; (b) the timing of maximum difference does not coincide among sites, and (c) the magnitudes of the differences are not identical. Possible reasons for this discrepancy are discussed later.

Discussion and conclusion

Testing of various processing schemes of \(\delta^{13}C_a\) data reveals that the model is sensitive to the processing of input

**Figure 7.** Sensitivity of the model is tested against \(h\) (‰/ppm) for the \(\delta^{13}C_a\) from Kothi, India (KOHPI). a. The whole period (1920–2000). b. The reconstruction period, including testing period (1920–1978). The lightest shades are used for the lowest values of \(h\) (0.2 and 0.25) as they cause the most deviation. The rest show small, random spreads around the best estimation of \(c_r\) for \(h = 0.45\) (shadow circled).

**Figure 8.** Comparison of model-derived pCO\(_2\) against observed June pCO\(_2\) values showing latitudinal gradient between tropics and mid-latitudes. Two sites each have been chosen from the southern tropics and northern mid-latitudes (as these have the highest gradient) based on availability. The observed data (mean of 1992–2006 June months) are taken from CSIRO and SIO pCO\(_2\) databases respectively, and are shown in grey. The model-derived data (mean of 1902–1957) are shown in black.
$\delta^{13}C_v$. This implies that the value of $d$ in natural samples not only depends on the ratio $c_i/c_a$, but also on the absolute values of $c_a$. Hence, $hc_a$ should be considered an inherent part of the model and not a mere correction applied to remove anthropogenic effect on $c_v$. So, the final form of $d$ turns out to be

$$d + 6.4 - hc_a = a + (b - a) \frac{c_i}{c_a}$$

$$= a_T + a_x + r_1 + (h_i - a_1)yz - a_2yz + (r_2 - r_1)z.$$  \hspace{1cm} (14)

Notably, Feng and Epstein\cite{15} proposed the correction for $c_v$ as the correction for the higher $c_a$ values in the latter part of the 20th century. Instead, our study shows the factor ($hc_a$) in eq. (14) needs to be applied for both higher ($\sim 380$ ppm) and lower ($\sim 280$ ppm) ranges of $c_a$ spanning the entire last century (Figure 4). This means that the modified value is important in terms of tree chemistry and is observed only in the presence of a significant and persistent temporal trend in $c_v$. Although this high value of $h$ appears to be a limitation of the model, it is justified by considering the fractionation factors $(a, b)$ as not constant but temperature- and humidity-dependent; the last few decades experienced unprecedented rate in rise in both temperature and $c_a$, which must have affected the tree chemistry and as a result the carbon isotope fractionation. The earlier estimates for this constant involved trials to increase the linear correlation with the meteorological parameters\cite{14,37} and nonlinear variabilities were not considered. Also, this correction does not have a linear effect on reconstruction as that is derived primarily from a nonlinear (exponential) function.

Table 1 shows that the $h$ values of samples from the same genus are similar even when growing in contrasting climates, e.g. KOHP in India ($P. smithiana$) and MRUS in Siberia ($Picea obovata$). On the other hand, samples from different genera from nearby sites show quite different $h$ values, e.g. CASW ($Quercus petraea$) and VISW ($Pinus sylvestris$). This implies that $h$ value is primarily species-dependent.

Figure 9 shows that $D$ is high during 1960–1910, with maximum in 1940–1931 at all the sites. Possible causes for this anomaly include:

(i) Decrease in industrial production in the northern hemisphere during this period, as the times of maxima coincide with the Great Depression and the beginning of the Second World War.

(ii) Increase in plant productivity due to the temperature maximum in the northern hemisphere\cite{39}.

(iii) Possible increased CO$_2$ supply from deep ocean near Law Dome area (as known in present times from Takahashi et al.\cite{40}). But due to lack of contemporary data we cannot confirm this hypothesis.
Our model for reconstructing $c_\text{r}$ from tree-ring cellulose has a few advantages over the $c_\text{r}$ data conventionally retrieved from ice cores. First, ice sheets accumulate only in the polar regions during the present. Therefore, $c_\text{r}$ values reconstructed from ice cores are mostly confined to polar regions. Second, ice-core records have limited overlap with the period of instrumental data (1958 and later), which limits their scope for validation. Third, their resolution decreases with time as snow compacts to fern and further to ice. Moreover, trapping of CO$_2$ in ice layers depends on snowfall rate that may have seasonality, causing bias in flux estimation. Lastly, ice-core dating requires substantial logistic support and manpower. The tree-ring-based reconstruction of $c_\text{r}$, on the other hand, overcomes many of these problems. Dating of tree rings is a straightforward process and can be achieved with minimal resources. Trees can be dated with annual resolution as far back as 11,000 years in time. They extend geographically from the tropics to high latitudes. Our study presents a reconstructed dataset which depicts the observed summer latitudinal gradient emphasizing the compatibility of the reconstruction with the observations. This study, therefore, expands the scope of carbon dynamics study spatially as well as temporally from the 1950s to about the early Holocene.

The present work is limited by the availability of temperature and vapour pressure data, i.e. to only till the beginning of the 20th century for most places. For earlier times, temperature reconstructions are available from various sources, including tree-ring-based ones from similar sites with annual resolutions. Also available are the drought estimates from tree-ring sites with implied moisture levels. Using these datasets in the presented model, it is possible to reconstruct $c_\text{r}$ for the entire duration of available tree-ring chronologies, albeit with higher error levels (as the error in the present reconstruction includes the errors in temperature and vapour pressure inputs). However, as the error in our model are taken as the variance of the input data, this increase might not be very high as most temperature reconstructions do not have high standard deviation.

The proposed modifications to the carbon isotope model for tree rings add complications to the original model by inclusion of recently available data and information. Further parameterization, including various tree (e.g. metal ion concentration), soil (e.g. chemical contamination) or physical parameters (e.g. light intensity) can be added in future when long-term data of these become available for calibration in the annual to multi-decadal scale for use in the palaeoclimatic studies. Even with limited amount of added complexities, the presented model reconstructs the $c_\text{r}$ data for all latitudes, from tropical (PMPE) to boreal (MRUS) with reasonable coherence.

Most importantly, this study has the potential to reconstruct the seasonal pCO$_2$ variation across the latitudes facilitating carbon cycle study in high temporal resolution. Hence, this model can further be tuned from annual to seasonal resolution on the basis of tree-ring samples separated into early and late woods in the growing season. In addition to estimation of past $c_\text{r}$ in the tropics and subtropics, our model separates meteorologically dependent effects in carbon isotopic fractionation from the independent ones, i.e. effects of species, location, etc. Removal of meteorologically independent coefficients may facilitate studies of future tree–atmosphere interactions vis-à-vis climate.


ACKNOWLEDGEMENTS. We thank M. Saurer (PSI, Switzerland) for providing data and all those who have contributed their data to the NOAA palaeoeclimate database. We are grateful to V. K. Gaur (CM-MACs, Bangalore), for active support in formulating and coding of the model; Y. K. Tiwari for providing valuable input to atmospheric pCO₂ data interpretation; S. Mahapatra, S. Saha, C. Gnanaseelan, S. Singh, N. R. Deshpande, M. V. S. Ramarao and S. Haldar for help in various stages of analysis, and Prof. B. N. Goswami (Director, Indian Institute of Tropical Meteorology, Pune) for encouragement. IITM is fully funded by the Ministry of Earth Science, New Delhi. Tree-ring samples from Western Himalayas were collected under ISRO–GBP–Dendro project.

Received 25 April 2014; revised accepted 27 June 2014