

Letter

No need to switch the modified Arrhenius function back to the old form

A comment on Murphy & Stinziano (2020) 'A derivation error that affects carbon balance models exists in the current implementation of the modified Arrhenius function'

The modified Arrhenius function has been used to model a peaked temperature response of the biochemical parameters of the photosynthesis model of Farquhar *et al.* (1980), such as J_{\max} (maximum potential electron transport rate under saturating irradiance) and V_{\max} (maximum carboxylation rate of Rubisco). A form of the modified Arrhenius function for parameter k (where k represents either J_{\max} or V_{\max}) normalised with respect to 25°C is as follows (e.g. Medlyn *et al.*, 2002):

$$k = k_{25} \exp \left[\frac{E}{R} \left(\frac{1}{298} - \frac{1}{T} \right) \right] \frac{1 + \exp \left[\frac{H}{R} \left(\frac{S}{H} - \frac{1}{298} \right) \right]}{1 + \exp \left[\frac{H}{R} \left(\frac{S}{H} - \frac{1}{T} \right) \right]} \quad \text{Eqn 1}$$

where T is temperature (in K), k_{25} is the parameter value at 25°C (i.e. 298 K), E and H are the energy of activation and deactivation (in J mol^{-1} , which indicate the rate of response before and after the peak, respectively), and S is known as an entropy factor, which has the same units as R , the universal gas constant ($8.314 \text{ J K}^{-1} \text{ mol}^{-1}$). There are other forms of the modified Arrhenius equation, but their modelled temperature response is mathematically identical to that of Eqn 1 (see Supporting Information Notes S1). Eqn 1 has also been used for the temperature response of photosynthetic CO_2 -diffusion parameters, mesophyll conductance (Bernacchi *et al.*, 2002) and bundle sheath conductance (Yin *et al.*, 2016). However, recently Murphy & Stinziano (2020) drew our attention to the original modified Arrhenius equation (Johnson *et al.*, 1942), which, if expressed in a form similar to Eqn 1 with the parameter value at the reference temperature of 25°C, is as follows:

$$k = k_{25} \frac{T}{298} \exp \left[\frac{E}{R} \left(\frac{1}{298} - \frac{1}{T} \right) \right] \frac{1 + \exp \left[\frac{H}{R} \left(\frac{S}{H} - \frac{1}{298} \right) \right]}{1 + \exp \left[\frac{H}{R} \left(\frac{S}{H} - \frac{1}{T} \right) \right]} \quad \text{Eqn 2}$$

Compared with Eqn 1, Eqn 2 has an additional term, $T/298$. According to Murphy & Stinziano (2020), this additional term was omitted by Hall (1979) and Farquhar *et al.* (1980), and since then, the form without the term, Eqn 1, has been predominantly used (e.g. Harley *et al.*, 1992; Medlyn *et al.*, 2002). Murphy & Stinziano (2020) compared Eqn 1 and Eqn 2, and found that the

estimated values of k_{25} , H and S were affected minimally, whereas the values of E were impacted substantially by the omission. They then upscaled these impacts using a whole-plant model, and showed that these impacts extended to modelled daily carbon gain. Murphy & Stinziano (2020) believed that the process of simplifying 'generated an error that was propagated for over 40 years', and recommended that the omission error of Eqn 1, which has been used most notably in current large-scale models, should be corrected. Here, I argue that the omission is not an error.

To analyse the differences between the two equations, it is more informative to use generated data because real data can be highly variable. Using a physiologically meaningful set of parameter values, I generated response curves that show a slightly different curvature for the two equations between 5°C and 45°C (Fig. 1a). I fit Eqn 2 to nine data points generated by Eqn 1, and vice versa. In both cases, the fitted E varied slightly, and the fitted H and S hardly changed from the values used to generate the curve (Table 1). As expected, parameter k_{25} , which defines the value at the same reference temperature, did not vary between the two models. Moreover, the curve generated using the fitted parameter values of one equation almost entirely overlapped the curve generated by the other model, and the percentage difference between the two models was negligible compared with the percentage difference between the models when using the same set of parameter values (Fig. 1b). However, the results of Murphy & Stinziano (2020) in their Fig. 2 showed that there were often significant outliers in the plots of estimated parameter values (including k_{25}) of one model vs those of the other model. These outliers presumably resulted from the fact that there were few measurement temperatures (e.g. only five temperatures in their Fig. 1), and estimating four parameters (k_{25} , E , H and S) from actual data at as few as five or six temperatures always carries a risk of overfitting.

To test the 'overfitting' hypothesis, I used the Microsoft EXCEL random-number generator to introduce a set of errors of up to $\pm 20\%$ to the nine Eqn 1-generated data points (Fig. 1a), and then fit Eqn 1 and Eqn 2 either to all nine data points or to only five points (i.e. every other point of the nine). As expected, parameter estimates using nine points differed appreciably not only from those in Table 1 but also from those using five points, and again the obtained parameter estimates differed slightly only for E between the two equations (Table S1). I then further increased the random error magnitude (up to $\pm 30\%$), narrowed the temperature range to only six temperatures (10, 20, 25, 30, 35 and 40°C) to more closely mimic the common measurement-temperature setup, and generated sets of temperature response curves. Of *c.* 100 sets where all four parameters were successfully estimated, predictions using the fitted parameter values of one equation again agreed well with those of the other

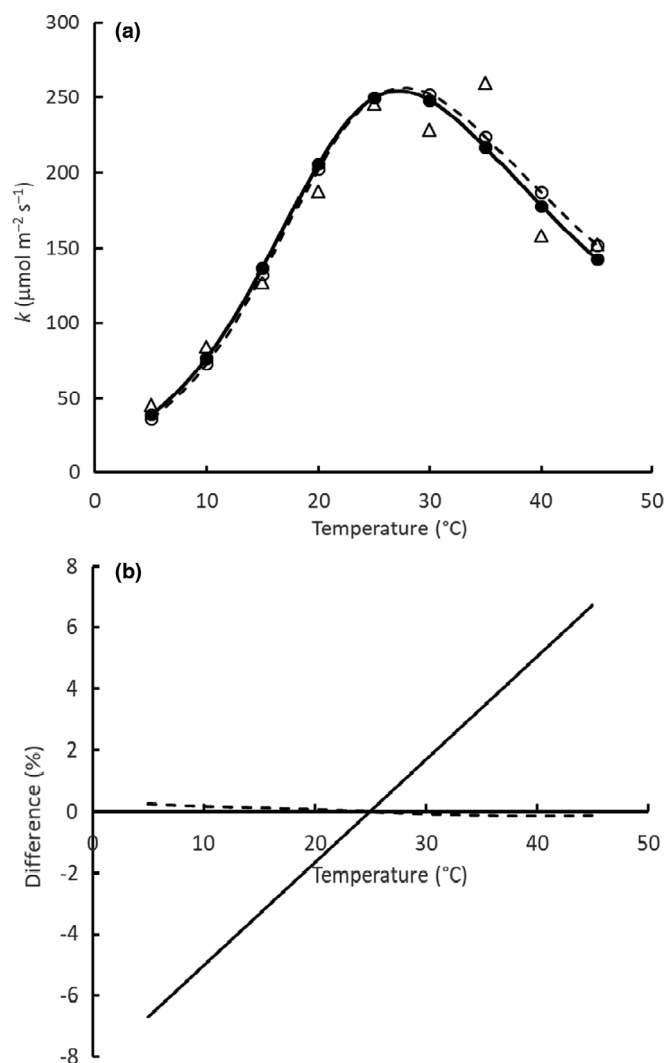


Fig. 1 (a) Temperature response curves generated by Eqn 1 (filled circles and the solid curve) or Eqn 2 (open circles and the dashed curve), using $k_{25} = 250 \mu\text{mol m}^{-2} \text{s}^{-1}$, $E = 95\,000 \text{ J mol}^{-1}$, $H = 136\,000 \text{ J mol}^{-1}$ and $S = 460 \text{ J K}^{-1} \text{ mol}^{-1}$; and Eqn 1-generated data points after introducing a set of random errors of up to $\pm 20\%$ (open triangles); (b) differences (%) between Eqns 1 and 2 when using the same set of parameter values (solid line) or between the two equations when using their own (fitted) parameter values (dashed line). Difference (%) was calculated as $((k_{\text{Eqn}2} - k_{\text{Eqn}1}) / k_{\text{Eqn}1}) \times 100$, where $k_{\text{Eqn}1}$ and $k_{\text{Eqn}2}$ are values of k calculated using Eqn 1 and Eqn 2, respectively. E , activation energy; H , deactivation energy; k_{25} , parameter k (either J_{max} (maximum potential electron transport rate under saturating irradiance) or V_{cmax} (maximum carboxylation rate of Rubisco)) at 25°C ; $k_{\text{Eqn}1}$, parameter k for Eqn 1; $k_{\text{Eqn}2}$, parameter k for Eqn 2; S , entropy factor.

equation between 1°C and 45°C (Fig. 2). There was essentially no difference in estimated k_{25} , H or S between the two equations, while values of E estimated by Eqn 2 were 2–3% lower than those estimated by Eqn 1 (Fig. S1). The slightly lower E of Eqn 2 is expected, because its extra term $T/298$ makes the curve rise slightly faster with increasing temperature (Fig. 1) and the estimated E , as the parameter defining the rate of increase of the curve, has to be lower compared to the estimated E of Eqn 1 (Tables 1, S1; Fig. S1) to achieve their best fit to the same data.

Table 1 Parameter values (standard error values in parentheses) of Eqn 2 estimated from data generated by Eqn 1, and those of Eqn 1 estimated from data generated by Eqn 2.

| | Eqn 2 fit to Eqn 1-generated data ^a | Eqn 1 fit to Eqn 2-generated data ^b |
|---|--|--|
| k_{25} ($\mu\text{mol m}^{-2} \text{s}^{-1}$) | 250.0 (0.02) | 250.0 (0.02) |
| E (J mol^{-1}) | 92561.7 (28.1) | 97443.2 (33.2) |
| H (J mol^{-1}) | 136144.0 (17.8) | 135869.0 (21.4) |
| S ($\text{J K}^{-1} \text{ mol}^{-1}$) | 460.5 (0.07) | 459.5 (0.08) |
| R^2 | 1.0 | 1.0 |

Parameter values used for both Eqn 1 and Eqn 2 to generate the data points were: $k_{25} = 250 \mu\text{mol m}^{-2} \text{s}^{-1}$; $E = 95\,000 \text{ J mol}^{-1}$; $H = 136\,000 \text{ J mol}^{-1}$; and $S = 460 \text{ J K}^{-1} \text{ mol}^{-1}$. E , activation energy; H , deactivation energy; k_{25} , parameter k (either J_{max} (maximum potential electron transport rate under saturating irradiance) or V_{cmax} (maximum carboxylation rate of Rubisco)) at 25°C ; S , entropy factor.

^aNine data points generated by Eqn 1 are shown as closed circles in Fig. 1.

^bNine data points generated by Eqn 2 are shown as open circles in Fig. 1.

The opposite and much more significant difference Murphy & Stinziano (2020) showed in their fig. 2(a,b) – that is, that E of Eqn 1 (their M2002 model) was *c.* 83–85% of E of Eqn 2 (their J1942 model) – is hard to explain. But this and the aforementioned outliers in their Fig. 2 were presumably associated with high variabilities in the temperature response of their estimated V_{cmax} and J_{max} . In my generated data with up to $\pm 30\%$ random errors, there were often cases – *c.* 30% of the total – where H could not be estimated and other parameter estimates went up to extreme values (e.g. $E > 2\,000\,000 \text{ J mol}^{-1}$, $S > 8000 \text{ J K}^{-1} \text{ mol}^{-1}$, in a manner analogous to some of the extremely high estimates in Fig. 2 of Murphy & Stinziano, 2020). It was in these cases (where data were highly variable) that my estimated E or S values from one equation deviated significantly from those of the other equation. Only when one parameter (H or S) was prefixed could the remaining parameters be estimated to have normal values, with which the predicted curves of the two equations coincided again. This agrees with a common strategy in the literature (Harley *et al.*, 1992; Medlyn *et al.*, 2002; Cai *et al.*, 2020) of avoiding an overfitting by prefixing one or two parameters. Uncertainties in the temperature response of V_{cmax} and J_{max} in the study of Murphy & Stinziano (2020) may not only stem from the measured data from multiple sources but may also occur at the fitting step when estimating the two parameters. The strong impact of the missing term in Eqn 1 on the modelled carbon productivity described in their study could result from the carry-over effect of an overfitting occurrence, rather than the real difference between the equations. When data points are few and highly variable, one can never judge which model is more favourable over the other. In my opinion, for such a case, the parsimony principle applies – the simpler model, Eqn 1, should be chosen.

More importantly, it is not only the simplicity but also the mathematical elegance that makes Eqn 1 more attractive. Both Eqn 1 and Eqn 2 are formulated atop the simpler, non-peaked Arrhenius equation (see Eqn S1.1 in Notes S1) in order to describe

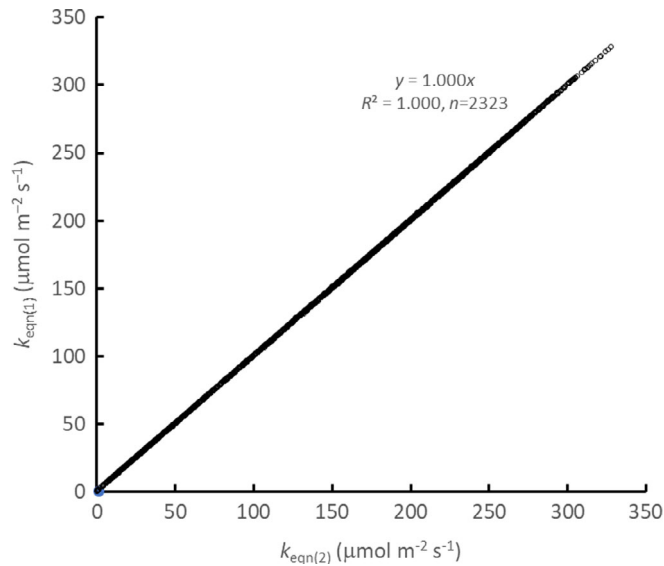


Fig. 2 Comparison of values of k calculated using Eqn 1, $k_{\text{Eqn 1}}$, with those calculated using Eqn 2, $k_{\text{Eqn 2}}$, every 2°C from 1°C to 45°C (so, at 23 temperatures), using their own parameter values estimated from fitting to *c.* 100 sets of curves generated with up to $\pm 30\%$ random errors for the responses to six temperatures (10, 20, 25, 30, 35 and 40°C). k (either J_{max} (maximum potential electron transport rate under saturating irradiance) or V_{cmax} (maximum carboxylation rate of Rubisco)); k_{Eqn1} , parameter k for Eqn 1; k_{Eqn2} , parameter k for Eqn 2.

response curves having an optimum temperature, T_{opt} . In the case of Eqn 1, T_{opt} (in K) can be solved analytically, as given by Hall (1979) and Farquhar *et al.* (1980):

$$T_{\text{opt}} = \frac{H}{S + R \log_e \left(\frac{H}{E} - 1 \right)} \quad \text{Eqn 3}$$

Once E , H and S are estimated by fitting Eqn 1 to data, T_{opt} can be calculated immediately from Eqn 3. Because T_{opt} often shifts among species or in plants grown under different environmental conditions (Medlyn *et al.*, 2002; Cai *et al.*, 2020), it is always useful to have an analytical solution for the characterisation of temperature optima. Eqn 3 also suggests that H must be $> E$, otherwise T_{opt} cannot be solved (the logarithm of zero or a negative number does not exist). This may have an implication in setting initial values of these two parameters in the fitting procedure. If the estimated H is $< E$ (see e.g. Boyd *et al.*, 2015), one should re-examine the data and reconsider whether to choose nonpeaked or peaked Arrhenius equations.

In the case of Eqn 2, the formula for T_{opt} can be expressed as follows (Notes S2):

$$\frac{S}{H} - \frac{1}{T_{\text{opt}}} = \frac{R}{H} \log_e \frac{E + RT_{\text{opt}}}{H - E - RT_{\text{opt}}} \quad \text{Eqn 4}$$

As T_{opt} appears in both logarithmic and other terms, it is impossible to solve for T_{opt} analytically (note that the RT_{opt} terms within the logarithmic function in Eqn 4 stem from the additional term in Eqn 2, see Notes S2; and Eqn 4 without the RT_{opt} terms would become Eqn 3). T_{opt} can only be calculated numerically

from Eqn 4 after E , H and S are estimated, which is technically challenging for general plant scientists to implement.

To avoid this numerical approach, one can combine Eqn 2 and Eqn 4 in a fitting procedure so as to estimate T_{opt} instead of S . The better option is to derive a single equation that uses T_{opt} as a parameter, thereby bypassing the parameter S , which is not easily interpreted (Medlyn *et al.*, 2002). This equation for the case of Eqn 1 is available in the literature:

$$k = k_{\text{opt}} \frac{H \cdot \exp \left[\frac{E}{R} \left(\frac{1}{T_{\text{opt}}} - \frac{1}{T} \right) \right]}{H - E \left\{ 1 - \exp \left[\frac{H}{R} \left(\frac{1}{T_{\text{opt}}} - \frac{1}{T} \right) \right] \right\}} \quad \text{Eqn 5}$$

where k_{opt} is the value of k at T_{opt} . Eqn 5 is Eqn 2 from Harley & Baldocchi (1995; note an error in their equation, whereby the parameter 'deactivation energy' within the exponential part of the denominator in Eqn 5 was typeset as the 'activation energy') and Eqn 18 in Medlyn *et al.* (2002).

Here I derive the Eqn 2-based function with T_{opt} as its parameter (Notes S3):

$$k = k_{\text{opt}} \frac{T}{T_{\text{opt}}} \frac{H \cdot \exp \left[\frac{E}{R} \left(\frac{1}{T_{\text{opt}}} - \frac{1}{T} \right) \right]}{H - (E + RT_{\text{opt}}) \left\{ 1 - \exp \left[\frac{H}{R} \left(\frac{1}{T_{\text{opt}}} - \frac{1}{T} \right) \right] \right\}} \quad \text{Eqn 6}$$

The value of T_{opt} can then be fitted together with k_{opt} , E and H . When fitting to the same data, the E and H estimated using Eqn 5 and Eqn 6 were the same as the estimates produced using Eqn 1 and Eqn 2, respectively (Table S1). Again there was little difference in the estimated T_{opt} , k_{opt} and H between Eqn 5 and Eqn 6 (Table S1), or in predicted temperature response curves between the two models when using their respective parameter values.

To avoid error propagation, I would also like to comment on Murphy & Stinziano's use of the equation of Farquhar *et al.* (1980) for the electron transport-limited rate of CO_2 -assimilation A_j :

$$A_j = \frac{J(C - \Gamma_*)}{4(C + 2\Gamma_*)} - R_d \quad \text{Eqn 7}$$

where C is the CO_2 partial pressure at the carboxylating sites, Γ_* is the CO_2 compensation point in the absence of day respiration (R_d) and J is the potential electron transport rate. The stoichiometric coefficient 4 in Eqn 7 refers to four electrons required per CO_2 fixed. However, in the equivalent equation in Table 3 in Murphy & Stinziano (2020), the 4 was missing. As usual, Murphy & Stinziano (2020) expressed J as a function of incident irradiance (Q_{in}):

$$\theta J^2 - (0.5\alpha\phi Q_{\text{in}} + J_{\text{max}})J + 0.5\alpha\phi Q_{\text{in}}J_{\text{max}} = 0 \quad \text{Eqn 8}$$


where α is absorptance, ϕ is the maximum quantum efficiency of Photosystem II electron transport under limiting light and θ is the curvature factor. Their value of ϕ was $0.08 \text{ mol mol}^{-1}$ (see their Table 2). But biologically sensible values of ϕ should be *c.* 0.8 mol mol^{-1} (e.g. Baker, 2008). Hopefully these errors were

only typos (but note that the Corrigendum published later in November 2020 indicated that the errors actually occurred in their modelling, and the impacts of the difference between Eqn 1 and Eqn 2 after correction became smaller than their initially reported values).

In summary, despite the importance of respecting the classics, there is no need to switch the currently used Eqn 1 back to the older form, Eqn 2. The extra term, $T/298$, in Eqn 2 has little influence on the function, which, as Hall (1979) stated, was one of the main reasons why he intentionally omitted the term. Here I have proposed a more important reason – that is, Eqn 1 is mathematically elegant in that the optimum temperature T_{opt} can be calculated analytically from the estimated E , H and S (see Eqn 3), whereas this is not possible with Eqn 2. To minimise this weakness of the old equation, I provide a form, Eqn 6, with which one can fit T_{opt} . The predominantly used form, Eqn 1 or Eqn 5, and the old form, Eqn 2 or Eqn 6, should give nearly identical temperature response curves when using their respective parameter values if the parameters are successfully fitted to data (the SAS scripts I used for the parameterization of the four equations are provided in Notes S4). The significant difference shown initially by Murphy & Stinziano (2020) when using Eqn 1 and Eqn 2 to predict whole-plant carbon gains must result from the carry-over artefact of overfitting the equations to few yet highly variable data points (plus their mis-handling of the Farquhar *et al.* (1980) model).

ORCID

Xinyou Yin  <https://orcid.org/0000-0001-8273-8022>

Xinyou Yin 

Centre for Crop Systems Analysis, Wageningen University & Research, PO Box 430, Wageningen, 6700 AK the Netherlands (email: Xinyou.yin@wur.nl)

References

- Baker NR. 2008. Chlorophyll fluorescence: a probe of photosynthesis *in vivo*. *Annual Review of Plant Biology* **59**: 89–113.
- Bernacchi CJ, Portis AR, Nakano H, von Caemmerer S, Long SP. 2002. Temperature response of mesophyll conductance. Implication for the determination of Rubisco enzyme kinetics and for limitations to photosynthesis *in vivo*. *Plant Physiology* **130**: 1992–1998.
- Boyd RA, Gandin A, Cousins AB. 2015. Temperature response of C_4 photosynthesis: biochemical analysis of Rubisco, phosphoenolpyruvate carboxylase and carbonic anhydrase in *Setaria viridis*. *Plant Physiology* **169**: 1850–1861.
- Cai C, Li G, Di L, Ding Y, Fu L, Guo X, Struik PC, Pan G, Chen W, Luo W *et al.* 2020. The acclimation of leaf photosynthesis of wheat and rice to seasonal temperature changes in T-FACE environments. *Global Change Biology* **26**: 539–556.
- Farquhar GD, von Caemmerer S, Berry JA. 1980. A biochemical model of photosynthetic CO_2 assimilation in leaves of C_3 species. *Planta* **149**: 78–90.
- Hall AE. 1979. A model of leaf photosynthesis and respiration for predicting carbon dioxide assimilation in different environments. *Oecologia* **143**: 299–316.
- Harley PC, Baldocchi DD. 1995. Scaling carbon dioxide and water vapour exchange from leaf to canopy in a deciduous forest. I. Leaf model parameterization. *Plant, Cell & Environment* **18**: 1146–1156.
- Harley PC, Thomas RB, Reynolds JF, Strain BR. 1992. Modelling photosynthesis of cotton grown in elevated CO_2 . *Plant, Cell & Environment* **15**: 271–282.
- Johnson FH, Eyring H, Williams RW. 1942. The nature of enzyme inhibitions in bacterial luminescence: sulfanilamide, urethane, temperature and pressure. *Journal of Cellular and Comparative Physiology* **20**: 247–268.
- Medlyn BE, Dreyer E, Ellsworth D, Forstreuter M, Harley PC, Kirschbaum MUF, Le Roux X, Montpied P, Strassmeyer J, Walcroft A *et al.* 2002. Temperature response of parameters of a biochemically based model of photosynthesis. II. A review of experimental data. *Plant, Cell & Environment* **25**: 1167–1179.
- Murphy BK, Stinziano JR. 2020. A derivation error that affects carbon balance models exists in the current implementation of the modified Arrhenius function. *New Phytologist*. doi: 10.1111/nph.16883
- Yin X, van der Putten PEL, Driever SM, Struik PC. 2016. Temperature response of bundle-sheath conductance in maize leaves. *Journal of Experimental Botany* **67**: 2699–2714.

Supporting Information

Additional Supporting Information may be found online in the Supporting Information section at the end of the article.

Fig. S1 Comparison of parameter values estimated by Eqns 1 and 2.

Notes S1 Various forms of the (peaked) Arrhenius equation.

Notes S2 Derivation of Eqn 4.

Notes S3 Derivation of Eqn 6.

Notes S4 SAS scripts for parameterising Eqn 1, Eqn 2, Eqn 5 and Eqn 6.

Table S1 Parameter estimates from generated data into which random errors of up to $\pm 20\%$ were introduced.

Please note: Wiley Blackwell are not responsible for the content or functionality of any Supporting Information supplied by the authors. Any queries (other than missing material) should be directed to the *New Phytologist* Central Office.

Key words: modified Arrhenius function, optimum temperature, parameter estimation, photosynthetic parameters, temperature response.

Received, 22 September 2020; accepted, 26 February 2021.