The Holocene 19,5 (2009) pp. 757–764

Stomatal-based inference models for reconstruction of atmospheric CO_2 concentration: a method assessment using a calibration and validation approach

W. Finsinger* and F. Wagner-Cremer

(Palaeoecology, Institute of Environmental Biology, University of Utrecht, Laboratory of Palaeobotany and Palynology, Budapestlaan 4, NL-3584 CD Utrecht, The Netherlands)

Received 18 August 2008; revised manuscript accepted 9 February 2009

A HOLOCENE RESEARCH PAPER

Abstract: We investigated changes in atmospheric CO_2 concentration (hereafter [CO_2]) over the period AD 1700–2002 as reconstructed using the inverse relationship between stomatal frequencies (SF) of *Betula nana* leaves from northern Europe and [CO_2]. The predictive ability of SF-inference models was assessed using a method of independent validation that involves two steps: (1) a training set of leaves grown between AD 1843 and 2002 was used to generate inference models; (2) the models were then applied to a fossil SF record of leaves grown between AD 1700 and 2002 that was split into two parts, a validation period (after AD 1850) and a reconstruction period (AD 1700–1850). Although our inference models had uncertainties comparable with other SF-inference models (root mean square error (RMSE) = *c*. 18–19 ppmv), uncertainties arising from the independent validation were larger (RMSE = *c*. 31–34 ppmv). Smoothed SF-inferred [CO_2] values after AD 1850 corresponded better to the industrial [CO_2] increase observed from instrumental records and from high-resolution ice cores, corroborating the accuracy of the reconstruction method in capturing a long-term (decadal- to centennial-scale) signal. This also indicates that in our record higher-frequency signals (eg, [CO_2] maxima around AD 1750) are potentially less reliable. In an attempt to estimate the maximum reconstruction uncertainty (± 67 ppmv), we considered (i) the RMSE of the validation (validation error) and (ii) the maximum difference between reconstructions obtained with different inference models during the validation period (method error). We suggest that reconstruction uncertainties may be reduced by reducing the uncertainty of our inference models, with a subfossil record characterized by lower variability in the SF time series over the validation period, and by smoothing the reconstruction. This study shows that independent validation is an important step to assess the precision and accuracy of quantitative proxy-based reconstructions.

Key words: Stomatal frequency, quantitative reconstructions, calibration, validation, climate change, Sweden.

Introduction

Stomata are small pores located on the surface of plant leaves and they play an important role in the regulation of the uptake of CO_2 for photosynthesis and the loss of water during transpiration. Although the roles of plant-growth regulators and environmental signals in regulating stomatal number are still largely unexplored (Bergmann and Sack, 2007), evidence shows that the concentration of atmospheric CO_2 (hereafter [CO_2]) controls not only stomatal opening (Young *et al.*, 2006) but also stomatal frequencies (SF, eg, Woodward, 1987; Beerling *et al.*, 1998). As shown by recent investigations, SFs are very likely regulated via systemic genetic-signalling pathways (Gray *et al.*, 2000; Lake *et al.*, 2002).

Because SFs are strongly influenced by [CO_2] they have the potential to be used for reconstructing past [CO_2] quantitatively (Roth-Nebelsick, 2005). The cuticle covering the leaf epidermis is made of waxes that are resistant to degradation under anoxic conditions, and thus it mirrors preserved morphological traits of the leaf epidermis inclusive stomata in sedimentary records. Transfer functions have been developed for a range of plant species to reconstruct [CO_2] based on SF (eg, Beerling *et al.*,

*Author for correspondence. Present address: Centre de Bio-Archélogie et d'Ecologie (CBAE), UMR 5059 CNRS – University of Montpellier II, Institut de Botanique, 163 rue Auguste Broussonet, F-34090 Montpellier Cedex, France (e-mail: walter.finsinger@univ-montp2.fr)

10.1177/0959683609105300

1995; McElwain *et al.*, 2002; Rundgren and Björck, 2003; Kouwenberg *et al.*, 2005; Garcia-Amorena *et al.*, 2006). As summarized in a review paper by Wagner *et al.* (2004), Holocene stomatal-based [CO_2] reconstructions are highly comparable in terms of amplitude of changes. In addition, their higher amplitude changes may be reconciled with smaller amplitude changes from ice-core [CO_2] records after smoothing the former time-series (McElwain *et al.*, 2002; Van Hoof *et al.*, 2005). This result supports the idea that in addition to using SF records for exploring greenhouse-gas dynamics over pre-Quaternary times (Jansen *et al.*, 2007), ie, for periods not covered by ice-core [CO_2] records, such records may be also used over more recent timescales.

Even though SFs have been successfully applied as [CO_2] indicators for pre-Quaternary, Lateglacial and Holocene times, very few attempts have been made to compare inferences with ice-core and instrumental records. The most powerful means of validation is to compare the reconstructions, at least for the recent past, against known recorded historical environmental records (Birks, 1998). Validations are increasingly performed for assessing the prediction uncertainty of quantitative palaeoclimate reconstructions based on proxy records (eg, Finsinger *et al.*, 2007; Vermaire and Gregory-Eaves, 2008; Larocque *et al.*, 2009). A first independent methodical assessment of the [CO_2] prediction quality of SF data with herbarium training-sets and historical SF records for three subtropical species showed significant differences in prediction accuracy (Wagner *et al.*, 2005). Although these results were promising, they also showed the importance of validating SF-inferred [CO_2] reconstructions. However, the Wagner *et al.* (2005) study covered a shorter gradient (310–370 ppmv) than the [CO_2] increase during the past ~150 years (290–370 ppmv), ie, the time span used for the calibration of most transfer functions.

Here another validation is attempted using the dwarf birch (*Betula nana*), a common element in the sub-Arctic vegetation whose leaves are occasionally sufficiently well preserved in Quaternary sediments for stomatal counts to be made (eg, Beerling, 1993; Rundgren and Björck, 2003). We assess the predictive ability of inference models using a two-step method of independent validation. (1) Calibration: a training set of modern leaves grown between AD 1843 and 2002 is used to generate inference models that express mathematically the species' response to changes in [CO_2]. (2) Validation: these models are applied to a subfossil-SF record (SF_F) in order to compare inferred with expected [CO_2] values. In addition, the SF_F-inferred [CO_2] are smoothed to different degrees in order to compare their ability in reconstructing [CO_2] at different temporal scales. This independent validation may provide a test to assess the performance of models based on herbarium-training sets to estimate the actual, or measured, values.

In addition, the present study explores [CO_2] changes over the period AD 1700–1850 as inferred from SF of *B. nana*. At present, only two SF-inferred [CO_2] records are available for the period AD 1700–1850. Of these two records, one shows no significant change in reconstructed [CO_2] (Rundgren and Beerling, 1999). By contrast, a more recent study by Kouwenberg *et al.* (2005) inferred a prominent [CO_2] maximum centred at around AD 1750, suggesting that [CO_2] could have served as a forcing factor for climate change also prior to the mid-nineteenth century. The time interval AD 1700–1850 is particularly important because radiative forcing calculations usually take the ice-core [CO_2] value at AD 1750 as the pre-industrial level (Forster *et al.*, 2007) and climate models exploring the influence of natural forcings alone use it as the pre-industrial value of [CO_2] (eg, Tett *et al.*, 2007). According to evidence from several Antarctic ice-core records, [CO_2] did not change by more than 12 ppmv between AD 1000 and 1850 (Siegenthaler *et al.*, 2005; MacFarling Meure *et al.*, 2006).

Material and methods

Training-set leaves

The training set combines 23 *B. nana* and 14 *Betula pubescens* leaf samples, and covers a total of 34 years from the period 1843 to 2007.

B. nana leaves collected in years 1919–1965 originate from specimens stored in the Nationaal Herbarium Nederland (NHN, Utrecht Universiteit branch) and leaves collected in years 1997–2002 originate from mature specimens growing at the Rassejohka garden of the Forest Line Arboretum near the sub-Arctic Research Station Kevo (69°45′N, 27°E, Finland). In addition, *B. nana* leaves picked from a peat monolith that was cut from a mire surface near Kevo and sampled contiguously and at high-resolution (Hicks *et al.*, 2004) were used to fill the gap between 1965 and 1997. The ^{14}C dating of individual *B. nana* leaves and *Sphagnum* leaves, stems and branches from this peat monolith have been wiggle-matched with the atmospheric ^{14}C content across the 1960s bomb peak (Goslar *et al.*, 2005). Based on this, the rate of peat accumulation in the uppermost layers is in excess of 5 mm/yr, so each 2 mm slice is less than 1 year's accumulation. *B. nana* collected from the Kevo peat monolith are dated to years AD 1976 to 1996.

B. pubescens leaves are either herbarium specimens from the National Museum of Denmark (growing between AD 1843 and 1964), or modern leaves collected in northern Fennoscandia during AD 1997 to 2001. All training-set leaves originate from lowland sites (< 200 m a.s.l.) in Fennoscandia between 54° and 69°N. For each year (n = 34), three to ten leaves were analysed.

For the modern training-set herbarium specimens of *B. nana* and *B. pubescens* were combined to test the comparability of the response rate and limit of these two species, which commonly hybridize to the mountain birch *B. pubescens* ssp. *czerepanovii* in northern Scandinavia (Wiegolaski, 2005). The likely presence of leaf fragments of mountain birch in the subfossil assemblages can be counterbalanced by using the training set developed on the parent species of *B. pubescens* ssp. *czerepanovii* (Wagner *et al.*, 2000).

Long stomatal density and stomatal index record

Determination of the age of samples in the Kiruna profile is based on radiocarbon-dated samples consisting of *Sphagnum* stems (Barnekow *et al.*, 2007). Six of the ten samples analysed were deposited immediately prior to, or following, the atomic weapons testing of the 1960s, enabling us to assign calendar ages to the radiocarbon dates with high levels of precision. Based on these control points, an age–depth model was constructed (after Goslar *et al.*, 2005) that was used to subsequently subsample the frozen monolith at an 'annual' resolution, assuming uniform peat accumulation between the control points (Figure 1a). The resulting sample-integration times vary from ~1 year (AD 2003–1955) to 3–7 years (AD 1954–1721) (Figure 1a).

Cuticle analysis

Subfossil leaf fragments of *B. nana* were separated from peat and identified by leaf and cuticle characteristics under a light microscope. For microscopic analysis, all leaf fragments were bleached in a cold 2–4% NaClO solution, washed in water and mounted on slides. Microscopic, computer-aided analysis of epidermal parameters was performed according to standard procedures on a Leica Quantimet 500C/500+ image-analysis system on seven digital images with a field size of 0.035 mm^2 (Wagner *et al.*, 1996). On stomata-bearing alveoles epidermal cell density (ED (n/mm^2)) and stomatal density (SD (n/mm^2)) were analysed and the area-independent stomatal index ($SI(\%) = SD/(SD + ED) \times 100$) was calculated, following Salisbury (1927).

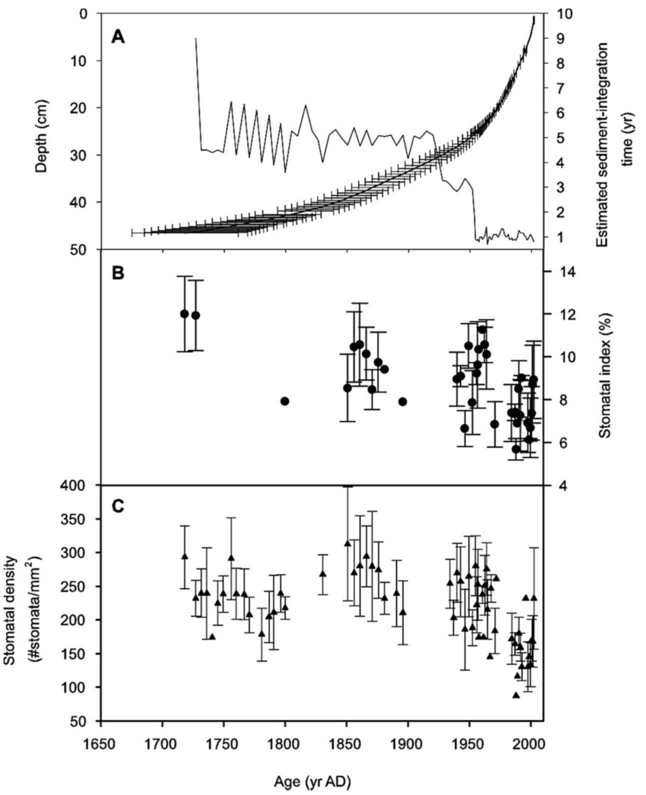

Figure 1 Cuticle parameters of *Betula nana* leaf fragments in the Kiruna profile. (A) Age–depth relationship (whiskers indicate 2σ of calibrated age ranges) and estimated sediment accumulation rate (continuous line), (B) mean stomatal index (SI_F%), (C) mean stomatal density (SD_F)

Atmospheric [CO_2] data

Historical [CO_2] data used for calibration are annual mean values as measured at Mauna Loa (AD 1958–2002) (Keeling and Whorf, 2005) and linearly interpolated [CO_2] measurements from air bubbles trapped in Antarctic ice cores (AD 1850–1957) (MacFarling Meure *et al.*, 2006) that overlap very well with the instrumental Mauna Loa record.

Inference models and statistical analyses

Inference models were developed by formalizing the relation between stomatal frequencies and historical [CO_2] at the year of growth. We used two different methods to develop inference models. In order to account for the non-linear response of stomatal frequencies to changing [CO_2] (eg, Kürschner *et al.*, 1997), SD_T, SI_T and the historical [CO_2] values were log-transformed before fitting a linear response curve through the data sets (models SI_{log} and SD_{log}) (Figure 2A and B). In addition, since the log-transformation and regression tends to underestimate the extremes, thus reducing the amplitude of changes in the reconstruction, we scaled (mean and variance equalization) the SD_T and SI_T timeseries to the same mean and variance of the historical [CO_2] record used for calibration (models SI_{sca} and SD_{sca}) (Figure 2C and D).

The performance of the models is evaluated based on (i) their root-mean-squared error (RMSE), (ii) the RMSE of the validation and (iii) the Reduction of Error statistic (RE). The latter was used to estimate the strength of the linear relationship between validation and observation. The RE statistic, which ranges from $-\infty$ to +1, compares the mean-square error (MSE) of the reconstruction to the MSE of a reconstruction that is constant in time with a value equivalent to the sample mean for the calibration data. If the

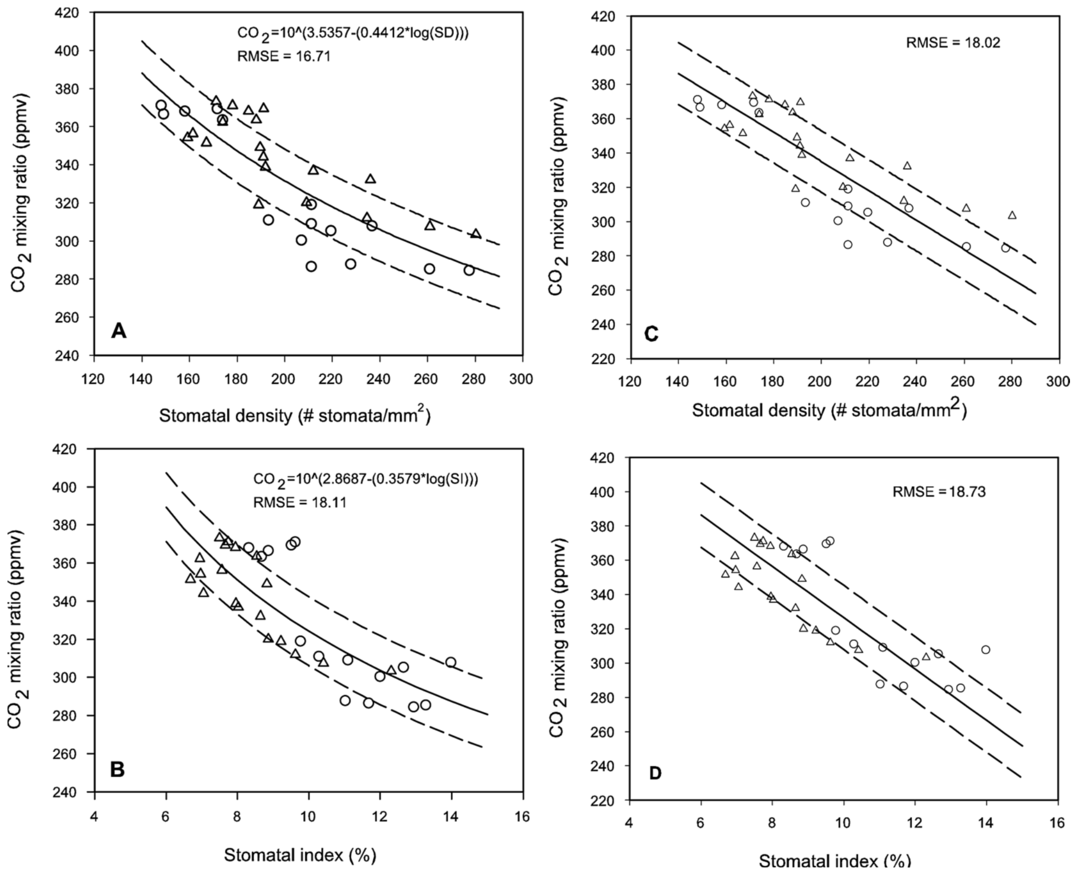

Figure 2 (A) and (B) Continuous line: [CO_2] inference models based on linear regression of log-transformed stomatal density (SD_{log}) (A) and stomatal index (SI_{log}) data (B) for *Betula pubescens/nana*. (C) and (D): Continuous line: [CO_2] inference models based on scaling of stomatal density (SD_{sca}) (C) and stomatal index (SI_{sca}) (D) data for *Betula pubescens/nana. Betula pubescens* (open circles) and *Betula nana* leaves (open triangles). Dashed lines indicate ± 1 RMSE

Table 1 Performances of stomatal-based [CO_2] inference models for the calibration period (AD 1843–2002) and for the validation period (AD 1851–2002)

Variable	Calibration (1843–2002)			Validation data set (1851–2000) Kiruna record		
	Method	r^2	RMSE	RMSE (RE)	RMSE (RE) 3-point smoothed	RMSE (RE) 5-point smoothed
SD	log/log	0.68	16.71	33.57 (–0.27)	20.57 (0.52)	15.77 (0.72)
	scaling	–	18.02	33.84 (–0.30)	18.87 (0.60)	13.46 (0.80)
SI	log/log	0.64	18.11	31.01 (–0.09)	26.91 (–0.18)	25.82 (0.25)
	scaling	–	18.73	32.60 (–0.20)	28.97 (–0.02)	28.15 (0.05)

reconstruction has any predictive value, one would expect it to do better than just the sample average over the calibration period; that is, one would expect RE to be greater than zero (see North *et al.*, 2006).

Results

Modern training set

The mean SD and SI of the training set (SD_T and SI_T) over the [CO_2] range from 284 ppmv to 373 ppmv changed by almost 50% from 280 n/mm² (14 SI%) to 148 n/mm² (6.7 SI%). These changes largely exceed the mean standard deviation of sampled leaves from individual years, reflecting (i) the plastic response of the two species studied to changes in [CO_2] and (ii) the potential of the training set to serve for reconstructions of past [CO_2]. The smallest RMSE was found for the SD_{log} model (16.71 ppmv CO_2) while other models (SI_{log}, SI_{sca} and SD_{sca}) are characterized by RMSEs in the range of 18–19 ppmv CO_2 (Table 1).

Subfossil record

Over the period AD 1850–2003, subfossil *B. nana* leaves show a successive long-term reduction in fossil-stomatal frequencies (SD_F and SI_F) (Figure 2b and c). Mean SD_F (SI_F) in the validation period (AD 1850–2002) range from 312 n/mm² (12.0%) to 86 n/mm² (5.6%). Before AD 1850, few data points of SI_F are available because of the poor preservation of subfossil cuticles and only SD_F were measured. SD_F values before AD 1850 range from 174 n/mm² to 293 n/mm².

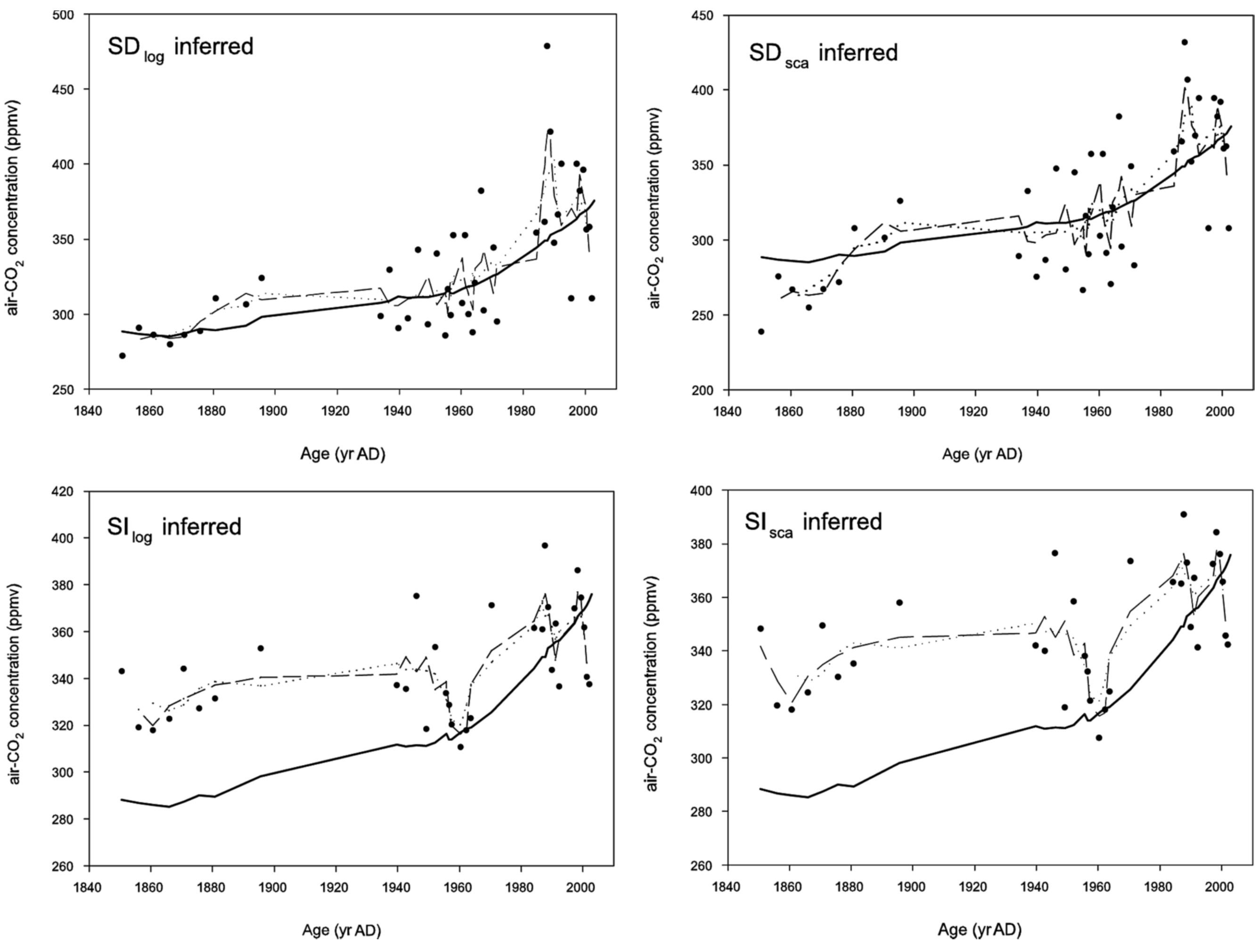

Figure 3 Comparison between historical record of [CO_2] changes (Mauna Loa record and Law Dome record: continuous lines) and SF-inferred [CO_2] reconstructions obtained with different inference models. In order to highlight decadal-scale changes in SF-inferred [CO_2] changes, the record has been smoothed with a 3-point and with a 5-point moving average (dashed lines and dotted lines, respectively)

Validation of inference models

The performance of inference models was investigated, applying them to the SD_F and SI_F records. The SF_F-inferred reconstructions reflect the observed long-term trend of [CO_2] over the period AD 1850–2002 as recorded by high-resolution ice cores and by the instrumental record (Figure 3). However, the low reconstruction skill is indicated by higher RMSEs than in the calibration set and by negative RE values (Table 1). In an attempt to investigate the performance of the inference models in capturing longer-term trends of [CO_2] changes, *n*-point running means ($n = 3$ and $n = 5$) were calculated and predictions compared with the unsmoothed observed values. By doing this, RMSEs of smoothed reconstructions obtained with the SD_{log} and SD_{sca} models substantially decreased (Table 1). Notably, RMSEs of SD models were similar or even lower than those of inference models and REs considerably improved. In terms of RMSEs, reconstructions obtained with SI_{log} and SI_{sca} models did not improve much, while REs of smoothed SI_F-inferred [CO_2] reconstructions remained low. Despite the improvements obtained by smoothing, SD_F-inferred [CO_2] are overestimated in parts of the records (eg, around AD 1990) or underestimated (eg, AD 1850–1880) (Figure 3).

Reconstruction of past [CO_2]

Because of the lack of SI_F values before AD 1850 and the better reconstruction skill of SD_{log} and SD_{sca} models, an attempt was made to reconstruct [CO_2] for the period AD 1850–1700 using the SD_F record. SD_F values in the reconstruction period range from 174 n/mm² to 293 n/mm², ie, within the range of SD_T values used for the development of inference models (280 n/mm² to 148 n/mm²). Thus, reconstructions are not based on extrapolation of the models into the unknown, which may increase the reconstruction uncertainty (North *et al.*, 2006).

In an attempt to estimate reconstruction uncertainties, two types of error were identified: uncertainty arising from (i) the different calibration methods that can be chosen for modelling the relationship between SF and historical [CO_2] (method error), and (ii) the unexplained variance in the reconstructions derived from validation against historical [CO_2] (validation error). The range between the maximum and minimum values for the reconstructions during the validation period is utilized to represent the method error. The validation error is estimated by means of the RMSE. These two error terms were quantified separately and combined (Figure 4) to allow for an estimation of the overall error of the historical [CO_2]. The combined uncertainty is represented by the square root of the summed and squared individual errors, following Esper *et al.* (2007). For the unsmoothed validation data set, the combined uncertainty is about ± 68 ppmv and decreases with increasing smoothing. For example, with a three-point smoothed data set the combined uncertainty reduces by *c.* 50% and equals *c.* ± 36 ppmv.

Reconstructed [CO_2] is characterized by high-amplitude variability (average amplitude is 70–100 ppmv). Lowest values are inferred with the SD_{sca} model that does not overestimate low [CO_2] as the SD_{log} does (Figures 3 and 4). Prominent minima in [CO_2] of about 260–270 ppmv occur around AD 1830, AD 1750 and AD 1720. In between these minima, [CO_2] maxima of up to 340–350 ppmv around AD 1780 and AD 1740 are reconstructed (Figure 4).

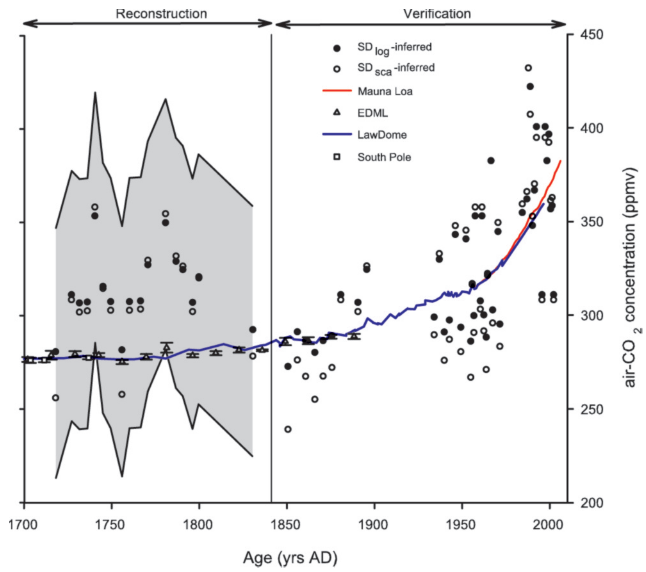

Figure 4 Verification (right) and reconstruction (left) of SD_F-based [CO_2] with *Betula nana* leaf fragments from the Kiruna profile (northern Sweden) and high-resolution ice-core derived [CO_2] for the period AD 1700–1996. Open triangles, Dronning Maud Land (EDML, Siegenthaler *et al.*, 2005); blue line, Law Dome (MacFarling Meure *et al.*, 2006); red line, instrumental record of atmospheric [CO_2] for the period AD 1958–2007, Mauna Loa Observatory (Keeling and Whorf, 2005); grey band, combined uncertainty for the SD_F-based reconstruction (method error + validation error)

Discussion

[CO_2] controls both stomatal opening (Young *et al.*, 2006) and stomatal numbers (SD and SI) in a wide variety of plants (Woodward, 1987; Beerling *et al.*, 1998). However, SD also changes with light intensity (Salisbury, 1927; Tichà, 1982; Poole and Kürschner, 1999) and water stress (Tichà, 1982) through lateral cell expansion of the epidermal cells surrounding the stomata. Thus, when the effects of epidermal size are taken into account in the area-independent SI the effects of light intensity and water stress may largely disappear (Salisbury, 1927; Poole and Kürschner, 1999). Hence, SI is generally preferred to SD for studying the plants' response to changing [CO_2] (eg, Royer, 2001). In our records, however, changes in SD were similar to changes in SI over the calibration period (training-set samples: $r^2 = 0.51$; subfossil record: $r^2 = 0.55$), indicating that the effect of [CO_2] was on stomatal development. We may, in addition, exclude the influence of light conditions for *B. nana* because this shade-intolerant plant grows in open tundra heath environments (De Groot *et al.*, 1997) and self-shading is very limited (Niinemets *et al.*, 2002). In addition, water availability is unlikely to be an issue because the plants analysed in this study were growing on a mire. In fact, a testate-amoebae record from the Kiruna peat profile indicates that from AD 1700 to 1900 the bog was mostly characterized by wet conditions (K. Schoning, personal communication, 2007).

Validation of stomatal-based [CO_2] reconstructions: are high-amplitude signals reliable?

To produce absolute quantitative measures of environmental changes it is necessary to place considerable reliance on the accuracy and precision of reconstructions (Briffa and Osborn, 1999). It is notable that the inference-model uncertainties (RMSE) are fully comparable with models based on similar numerical techniques from other plant species (16.4–13.9 ppmv; Wagner *et al.*, 2005). They are, however, larger than those obtained with training sets of *Ilex cassine* (8.9 ppmv) (Wagner *et al.*, 2005) and with combined training sets of *Quercus robur/petraea* and *Betula pendula/pubescens* (10.2 and 9.6 ppmv, respectively) (Wagner *et al.*, 2004; Van Hoof *et al.*, 2006), while they are lower than the RMSE of a training set for *Tsuga heterophylla* (42.8 ppmv) (Wagner *et al.*, 2004).

The evaluation and validation of palaeoenvironmental reconstructions is an important step to be taken prior to proxy-based reconstructions of a variable of interest because all (palaeoenvironmental) reconstruction procedures will produce a result in the form of computer output (Birks, 1998). Here we evaluate SD_F-inferred [CO_2] reconstructions with historical [CO_2] records used to establish the inference models. Ideally, SD_F-inferred [CO_2] values should match with these historical [CO_2] values. We observe that increasing SD_F-inferred [CO_2] after AD 1850 (Figure 3) corresponds well to the industrial [CO_2] increase seen in instrumental and ice-core records (Keeling and Whorf, 2005; MacFarling Meure *et al.*, 2006). This corroborates the accuracy of the reconstruction in broadly capturing the long-term trend over the past *c.* 150 years. In addition, SF_F-inferred [CO_2] obtained with the SF_{sca} models are not biased by over- or underestimation of reconstructed [CO_2] values as is common with SF_{log} models (eg, Wagner *et al.*, 2005). On the other hand, lower RMSEs and higher REs of smoothed SF_F-inferred [CO_2] time series in the validation (Table 1) indicate that the apparent high-amplitude signals in the unsmoothed time series are less reliable. The higher variability of the unsmoothed SF_F-inferred [CO_2] time series may have three potential causes: first, the inherent variability in the phenotypic plasticity of the plants may contribute a stochastic error in the reconstruction (eg, RMSE of inference models). Second, the [CO_2] record used to develop the inference models (Law Dome record) is in itself partly a smoothed record (smoothing resulting from enclosure of air in the ice is about 10 years at DE08, Law Dome, Trudinger *et al.*, 2003; MacFarling Meure *et al.*, 2006). However, SD_F-inferred [CO_2] would suggest a higher variability of [CO_2] changes than the annual instrumental record of Mauna Loa (Keeling and Whorf, 2005). This clearly indicates that the higher amplitude of our SD_F-inferred [CO_2] record in comparison with ice-core records can not be related to the smoothing effect of processes involved in air trapping but also to uncertainties in the SF_F-based reconstructions, as was previously suggested (eg, McElwain *et al.*, 2002; Van Hoof *et al.*, 2005). Third, the variability of the SD_F time-series is higher than that of the SD_T series. We may hypothesize that the variability in the training set is lower because within-leaf variation of SF was minimized by means of a well-defined sampling strategy (Poole *et al.*, 1996). Subfossil leaves from the Kiruna profile were rarely complete and fragments analysed could not always be assigned to the central leaf area. This potentially influences the estimate of the combined reconstruction uncertainty, which appears to be *c.* three times as large as the RMSE of the inference models.

SD-inferred [CO_2] changes AD 1700–1850

Knowledge of past [CO_2] variations is of utmost importance for our understanding of global biogeochemical cycles and for climate modelling. The most direct way of investigating past variations of [CO_2] before AD 1958, when continuous [CO_2] measurements started (Keeling and Whorf, 2005), is the analysis of air extracted from suitable ice cores. Detailed [CO_2] records from Antarctica (eg, Dronning Maud (DML, Siegenthaler *et al.*, 2005) and Law Dome (MacFarling Meure *et al.*, 2006)) indicate that between AD 1000 and 1850 variations of up to 12 ppmv occurred. Therefore it is generally assumed that climate changes prior to *c.* AD 1850 (eg, the 'Little Ice Age', AD 1300–1860) were mainly caused by changes in solar radiation and stratospheric-reaching aerosols from volcanic eruptions (eg, Crowley, 2000; Denman *et al.*, 2007).

It is well known that ice-core [CO_2] records are naturally affected by smoothing because of enclosure of air in the ice that removes high frequency variations from the record, so the true atmospheric variation may have been slightly larger than represented in ice core–air records (Trudinger *et al.*, 2003; MacFarling Meure *et al.*, 2006). In fact, air in a sample of ice does not correspond to a single time in the past, but is a mix of air parcels that left the atmosphere over a range of times (Trudinger *et al.*, 2003). The main processes involved in the air trapping are (1) diffusion through the firn layer and (2) the gradual trapping of air into bubbles in ice in the lock-in zone (Trudinger *et al.*, 2003; Siegenthaler *et al.*, 2005). The age spread from bubble close off is largely determined by the snow-accumulation rate, with higher accumulation-rate sites having narrower age distributions than low accumulation-rate sites. For three high-resolution ice-core records spanning the time interval AD 1700–present, the accumulation rate is highest at DE08 (Law Dome record AD 1844–1973, Etheridge *et al.*, 1996; MacFarling Meure *et al.*, 2006). At DE08 the smoothing corresponds broadly to a 10-year running mean (Trudinger *et al.*, 2003) and at the DSS site (Law Dome record: AD 1–1884, MacFarling Meure *et al.*, 2006) the age distribution for [CO_2] in ice is almost certainly wider than at DE08 (Trudinger *et al.*, 2003). In comparison with the DSS site, the accumulation rate at EDML is assumed very low (Siegenthaler *et al.*, 2005). Hence, if the amplitude and the rapidity of SD_F-inferred [CO_2] changes between AD 1750 and 1850 were real, they should better match with [CO_2] changes recorded at Law Dome (DSS site).

It is notable that between AD 1700 and 1850, mean SD_F-inferred [CO_2] are generally higher than mean [CO_2] values estimated from ice-core records, which range between 276 and 288 ppmv (Figure 4). Consistently higher SF-inferred [CO_2] than ice-core [CO_2] records have been observed for the Holocene by McElwain *et al.* (2002), who related the discrepancy to (i) problems with the transfer functions used to predict [CO_2] from past SF changes or to (ii) the influence of environmental variables on SD. We can exclude that high [CO_2] in the reconstruction are due to the transfer function alone because lowest values were correctly predicted with the use of the SD_{log} model in the validation (possibly underestimated with the SD_{sca} model). The influence of other environmental factors (eg, changing light intensity and water stress) are minimized (see discussion above).

In addition to generally higher values in comparison with ice-core [CO_2] records, our SD_F-inferred reconstruction also shows two maxima of larger magnitude (*c.* 60–70 ppmv) and of short duration (less than 100 years) centred between AD 1700 and 1800 (Figure 4). A prominent maximum (within uncertainty bands) similar in timing (considering age uncertainties, Figure 1) and in magnitude centred at around AD 1750 was recorded in an SD-inferred [CO_2] reconstruction based on *T. heterophylla* needles from northeastern America by Kouwenberg *et al.* (2005) (Figure 5). Because the [CO_2] variation also shows similarities with terrestrial air temperature trends, Kouwenberg *et al.* (2005) suggested that [CO_2] could have served as a forcing factor for climate change also prior to the mid-nineteenth century. By contrast, a *Salix herbacea* SI-inferred [CO_2] record from Fennoscandia (not shown) indicated no significant change in reconstructed values between AD 1700 and 1850 (Rundgren and Beerling, 1999). In the latter record, reconstructed values were within the 95% confidence limits of the reconstruction (*c.* ± 20 ppmv). In our new SD_F-inferred [CO_2] record, reconstruction uncertainties are larger than that (combined reconstruction uncertainty is *c.* ± 67 ppmv), and can therefore not confirm the conclusion of significantly higher [CO_2] values during the period AD 1700–1850 with any level of confidence (Figure 4). The reconstruction uncertainty is almost as large as the [CO_2] changes that occurred during glacial terminations as revealed by analysis of CO_2 trapped in Antarctic ice cores (Lüthi *et al.*, 2008). It is essential to reduce the reconstruction uncertainty. This may be achieved in at least two ways: (i) the reduction of the uncertainty of our inference models by adding more samples used in the calibration of the proxy (eg, Lucy *et al.*, 2008), and (ii) with a subfossil record characterized by lower variability in the SF time series over the validation period. The latter may be achieved if a record with higher abundance of well-preserved cuticles and complete leaves can be found. In addition, a higher datapoint density in the reconstruction period would enable the smoothing of the record thereby reducing the uncertainties of the high-frequency record.

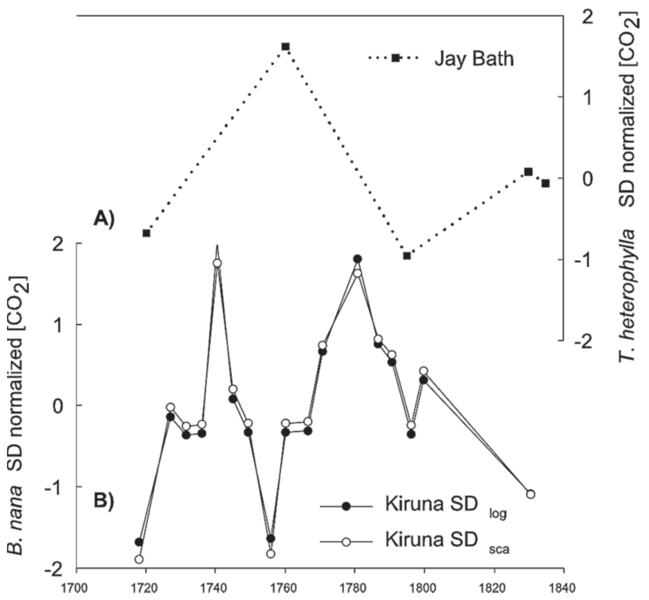

Figure 5 Normalized SF-inferred [CO_2] records (zero mean and unit standard deviation) from Jay Bath (Kouwenberg *et al*., 2005) and from the Kiruna record (see Figure 4)

Conclusions

With ever-increasing interest in past climate changes, it is important to assess the confidence in the quality of proxy-based reconstructions of environmental variables. Here we attempt to investigate the performance of inference models for the prediction of [CO_2] changes based on subfossil stomata of *Betula nana* leaves. We generated a training set with leaves collected from herbariums and from a short peat section and historical [CO_2] records from instrumental and ice-core records. Subsequently, inference models were applied to an independent subfossil SF record. The latter was obtained from a sediment core deposited over the time span from AD 1700 to 2003 that was accurately sampled allowing for a near-annual resolution between AD 1950 and present and a *c*. 5 year resolution for older samples. This well-dated record enabled the independent validation of SF-based [CO_2] reconstructions by comparison between observed versus predicted [CO_2] values showing that (i) scaling-based inference models perform better than regression-based models, (ii) the reconstruction uncertainty is larger (by a factor of two) than the uncertainty arising from the training set, and (iii) that the long-term (decadal- to centennial-scale) changes are better retained than the higher frequency changes. Hence, although our SD_F-inferred [CO_2] reconstruction resembles in its long-term temporal pattern another SF-inferred [CO_2] record from northeastern America (Kouwenberg *et al.*, 2005), it is clearly essential to reduce the reconstruction uncertainty to enhance the accuracy and precision of stomatal-based [CO_2] reconstructions.

Acknowledgements

We are grateful to Timme Donders, Hiro Yamazaki and Danny McCarroll for discussions and to two anonymous reviewers for constructive comments. Atmospheric [CO_2] from Law Dome and South Pole ice cores were downloaded from the Pangaea database (http://dx.doi.org/10.1594/PANGAEA.472477 and http://dx.doi.org/10.1594/PANGAEA.472477). WF was financially supported by the European Union 017008 (GOCE) 'MILLENNIUM' project. This is Netherlands Research School of Sedimentary Geology (NSG) publication no. 20090403.

References

Barnekow, L., Loader, N.J., Hicks, S., Froyd, C.A. and **Goslar, T.** 2007: Strong correlation between summer temperature and pollen accumulation rates for *Pinus sylvestris*, *Picea abies* and *Betula* spp. in a high-resolution record from northern Sweden. *Journal of Quaternary Science* 22, 653–58.

Beerling, D.J. 1993: Changes in the stomatal density of *Betula nana* leaves in response to increases in atmospheric carbon dioxide concentrations since the late-glacial. *Special Papers in Palaeontology* 49, 181–87.

Beerling, D.J., Birks, H.H. and **Woodward, F.I.** 1995: Rapid late-glacial atmospheric CO_2 changes reconstructed from the stomatal density record of fossil leaves. *Journal of Quaternary Science* 10, 379–84.

Beerling, D.J., McElwain, J.C. and **Osborne, C.P.** 1998: Stomatal responses of the 'living fossil' *Ginkgo biloba* L. to changes in atmospheric CO_2 concentrations. *Journal of Experimental Botany* 49, 1603–607.

Bergmann, D.C. and **Sack, F.D.** 2007: Stomatal development. *Annual Review of Plant Biology* 58, 163–81. doi:10.1146/annurev.arplant.58.032806.104023.

Birks, H.J.B. 1998: Numerical tools in palaeolimnology – progress, potentialities, and problems. *Journal of Paleolimnology* 20, 307–32.

Briffa, K.R. and **Osborn, T.J.** 1999: Seeing the wood from the trees. *Science* 284, 926–27. doi: 10.1126/science.284.5416.926.

Crowley, T.J. 2000: Causes of climate change over the past 1000 years. *Science* 289, 270–77.

De Groot, W.J., Thomas, P.A. and **Wein, R.W.** 1997: *Betula nana* L. and *Betula glandulosa* Michx. *Journal of Ecology* 85, 241–64.

Denman, K.L., Brasseur, G., Chidthaisong, A., Ciais, P., Cox, P.M., Dickinson, R.E., Hauglustaine, D., Heinze, C., Holland, E., Jacob, D., Lohmann, U., Ramachandran, S., da Silva Dias, P.L., Wofsy, S.C. and **Zhang, X.** 2007: Couplings between changes in the climate system and biogeochemistry. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L., editors, *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 499–587.

Esper, J., Frank, D., Büntgen, U., Verstege, A., Luterbacher, J. and **Xoplaki, E.** 2007: Long-term drought severity variations in Morocco. *Geophysical Research Letters* 34, L17702. doi:10.1029/2007GL030844.

Etheridge, D.M., Steele, L.P., Langenfelds, R.L., Francey, R.J., Barnola, J.-M. and **Morgan, V.I.** 1996: Natural and anthropogenic changes in atmospheric CO_2 over the last 1000 years from air in Antarctic ice and firn. *Journal of Geophysical Research D: Atmospheres* 101, 4115–28.

Finsinger, W., Heiri, O., Valsecchi, V., Tinner, W. and **Lotter, A.F.** 2007: Modern pollen assemblages as climate indicators in southern Europe. *Global Ecology and Biogeography* 16, 567–82. doi:10.1111/j.1466–8238.2007.00313.x.

Forster, P., Ramaswamy, V., Artaxo, P., Berntsen, T., Betts, R., Fahey, D.W., Haywood, J., Lean, J., Lowe, D.C., Myhre, G., Nganga, J., Prinn, R., Raga, G., Schulz, M. and **Van Dorland, R.** 2007: Changes in atmospheric constituents and in radiative forcing. In. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L., editors, *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*. Cambridge University Press, 129–234.

Garcia-Amorena, I., Wagner, F., van Hoof, T.B. and **Gomez-Manzaneque, F.** 2006: Stomatal responses in deciduous oaks from southern Europe to the anthropogenic atmospheric CO_2 increase; refining the stomatal-based CO_2 proxy. *Review of Palaeobotany and Palynology* 141, 303–12.

Goslar, T., Van Der Knaap, W. O., Hicks, S., Andric, M., Czernik, J., Goslar, E., Räsänen, S. and **Hyötylä, H.** 2005: Radiocarbon dating of modern peat profiles: pre- and post-bomb ^{14}C variations in the construction of age-depth models. *Radiocarbon* 47, 115–34.

Gray, J.E., Holroyd, G.H., Van der Lee, F.M., Bahrami, A.R., Sijmons, P.C., Woodward, I., Schuch, W. and **Hetherington, A.M.** 2000: The HIC signalling pathway links CO_2 perception to stomatal development. *Nature* 408, 713–16.

Hicks, S., Goslar, T. and **van der Borg, K.** 2004: A near annual record of recent tree line dynamics from Northern Finland. *Acta Palaeobotanica* 223–224, 5–11.

Jansen, E., Overpeck, J., Briffa, K.R., Duplessy, J.-C., Joos, F., Masson-Delmotte, V., Olago, D., Otto-Bliesner, B., Peltier, W.R., Rahmstorf, S., Ramesh, R., Raynaud, D., Rind, D., Solomina, O., Villalba, R. and **Zhang, D.** 2007: Palaeoclimate. In Solomon, S., Qin, D., Manning, M., Chen, Z., Marquis, M., Averyt, K.B., Tignor, M. and Miller, H.L., editors, *Climate change 2007: the physical science basis. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change*, Cambridge University Press, 433–97.

Keeling, C.D. and **Whorf, T.P.** 2005: Atmospheric CO_2 records from sites in the SIO air sampling network. In *Trends: a compendium of data on global change*. Carbon Dioxide Information Analysis Center, Oak Ridge National Laboratory, U.S. Department of Energy, Oak Ridge, TN, USA, Available: http://cdiac.ornl.gov/trends/co2/sio-mlo.html

Kouwenberg, L., Wagner, R., Kürschner, W. and **Visscher, H.** 2005: Atmospheric CO_2 fluctuations during the last millennium reconstructed by stomatal frequency analysis of *Tsuga heterophylla* needles. *Geology* 33, 33–36.

Kürschner, W.M., Wagner, F., Visscher, E.H. and **Visscher, H.** 1997: Predicting the response of leaf stomatal frequency to a future CO_2-enriched atmosphere: constraints from historical observations. *Geologische Rundschau* 86, 512–17.

Lake, J.A., Woodward, F.I. and **Quick, W.P.** 2002: Long-distance CO_2 signalling in plants. *Journal of Experimental Botany* 53, 183–93.

Larocque, I., Grosjean, M., Heiri, O., Bigler, C. and **Blass, A.** 2009: Comparison between chironomid-inferred July temperatures and meteorological data AD 1850–2001 from varved Lake Silvaplana, Switzerland. *Journal of Paleolimnology* 41, 329–42. doi:10.1007/s10933-008-9228-0.

Lucy, D., Robertson, I., Aykroyd, R.G. and **Pollard, A.M.** 2008: Estimates of uncertainty in the prediction of past climatic variables. *Applied Geochemistry* 23, 2961–65. doi:10.1016/j.apgeochem.2008.04.008.

Lüthi, D., Le Floch, M., Bereiter, B., Blunier, T., Barnola, J.-M., Siegenthaler, U., Raynaud, D., Jouzel, J., Fischer, H., Kawamura, K. and **Stocker, T.F.** 2008: High-resolution carbon dioxide concentration record 650,000–800,000 years before present. *Nature* 453, 379–82.

MacFarling Meure, C., Etheridge, D., Trudinger, C., Steele, P., Langenfelds, R., Van Ommen, T., Smith, A. and **Elkins, J.** 2006: Law Dome CO_2, CH_4 and N_2O ice core records extended to 2000 years BP. *Geophysical Research Letters* 33, L14810. Doi:10.1029/2006GL026152.

McElwain, J.C., Mayle, F.E. and **Beerling, D.J.** 2002: Stomatal evidence for a decline in atmospheric CO_2 concentration during the younger dryas stadial: a comparison with Antarctic ice core records. *Journal of Quaternary Science* 17, 21–29.

Niinemets, Ü., Portsmuth, A. and **Truus, L.** 2002: Leaf structural and photosynthetic characteristics, and biomass allocation to foliage in relation to foliar nitrogen content and tree size in three *Betula* species. *Annals of Botany* 89, 191–204.

North, G.E., Biondi, F., Bloomfield, P., Christy, J.R., Cuffey, K.R., Dickinson, R.E., Druffel, E.R.M., Nychka, D., Otto-Bliesner, B., Roberts, N., Turekian, K.K. and **Wallace, J.M.** 2006: *Surface temperature reconstructions for the last 2,000 years*. National Academies Press, Available: http://www.nap.edu/catalog/11676.html

Poole, I. and **Kürschner, W.M.** 1999: Stomatal density and index: the practice. In Jones, T.P. and Rowe, N.P., editors, *Fossil plants and spores: modern techniques*. Geological Society, 257–60.

Poole, I., Weyers, J.D.B., Lawson, T. and **Raven, J.A.** 1996: Variations in stomatal density and index: implications for palaeoclimatic reconstructions. *Plant, Cell and Environment* 19, 705–12.

Roth-Nebelsick, A. 2005: Reconstructing atmospheric carbon dioxide with stomata: possibilities and limitations of a botanical pCO_2-sensor. *Trees – Structure and Function* 19, 251–65.

Royer, D.L. 2001: Stomatal density and stomatal index as indicators of paleoatmospheric CO_2 concentration. *Review of Palaeobotany and Palynology* 114, 1–28.

Rundgren, M. and **Beerling, D.** 1999: A Holocene CO_2 record from the stomatal index of subfossil *Salix herbacea* L. leaves from northern Sweden. *The Holocene* 9, 509–13.

Rundgren, M. and **Björck, S.** 2003: Late-glacial and early Holocene variations in atmospheric CO_2 concentration indicated by high-resolution stomatal index data. *Earth and Planetary Science Letters* 213, 191–204.

Salisbury, E.J. 1927: On the causes and ecological significance of stomatal frequency, with special reference to the woodland flora. *Philosophical Transactions B* 216, 1–65.

Siegenthaler, U., Monnin, E., Kawamura, K., Spahni, R., Schwander, J., Stauffer, B., Stocker, T.F., Barnola, J.-M. and **Fischer, H.** 2005: Supporting evidence from the EPICA Dronning Maud Land ice core for atmospheric CO_2 changes during the past millennium. *Tellus B* 57, 51–57. doi:10.1111/j.1600–0889.2005.00131.x.

Tett, S.F.B., Betts, R., Crowley, T.J., Gregory, J., Johns, T.C., Jones, A., Osborn, T.J., Öström, E., Roberts, D.L. and **Woodage, M.J.** 2007: The impact of natural and anthropogenic forcings on climate and hydrology since 1550. *Climate Dynamics* 28, 3–34.

Tichà, I. 1982: Photosynthetic characteristics during ontogenesis of leaves: 7. Stomata density and sizes. *Photosynthetica* 16, 375–471.

Trudinger, C.M., Rayner, P.J., Enting, I.G., Heimann, M. and **Scholze, M.** 2003: Implications of ice core smoothing for inferring CO_2 flux variability. *Journal of Geophysical Research D: Atmospheres* 108(D16), 4492.

Van Hoof, T.B., Kaspers, K.A., Wagner, F., Van De Wal, R.S.W., Kürschner, W.M. and **Visscher, H.** 2005: Atmospheric CO_2 during the 13th century AD: reconciliation of data from ice core measurements and stomatal frequency analysis. *Tellus, Series B: Chemical and Physical Meteorology* 57, 351–55.

Van Hoof, T.B., Kürschner, W.M., Wagner, F. and **Visscher, H.** 2006: Stomatal index response of *Quercus robur* and *Quercus petraea* to the anthropogenic atmospheric CO_2 increase. *Plant Ecology* 183, 237–43.

Vermaire, J.C. and **Gregory-Eaves, I.** 2008: Reconstructing changes in macrophyte cover in lakes across the northeastern United States based on sedimentary diatom assemblages. *Journal of Paleolimnology* 39, 477–90.

Wagner, F., Below, R., De Klerk, P., Dilcher, D.L., Joosten, H., Kürschner, W.M. and **Visscher, H.** 1996: A natural experiment on plant acclimation: lifetime stomatal frequency response of an individual tree to annual atmospheric CO_2 increase. *Proceedings of the National Academy of Sciences of the United States of America* 93, 11 705–708.

Wagner, F., Neuvonen, S., Kürschner, W.M. and **Visscher, H.** 2000: The influence of hybridization on epidermal preperties of birch species and the consequences for palaeoclimatic interpretations. *Plant Ecology* 148, 61–69.

Wagner, F., Kouwenberg, L.L.R., Van Hoof, T.B. and **Visscher, H.** 2004: Reproducibility of Holocene atmospheric CO_2 records based on stomatal frequency. *Quaternary Science Reviews* 23, 1947–54.

Wagner, F., Dilcher, D.L. and **Visscher, H.** 2005: Stomatal frequency responses in a hardwood-swamp vegetation from Florida during a 60 years continuous CO_2 increase. *American Journal of Botany* 92, 690–95.

Wiegolaski, F.E. 2005: History and environment of the Nordic mountain birch. In Wiegolaski, F.E., Karlsson, P.S., Neuvonen, S. and Thannheiser, D., editors, *Plant ecology, herbivory, and human impact in Nordic mountain birch forests*. Ecological Studies, 180, Springer Verlag, 3–18.

Woodward, F.I. 1987: Stomatal numbers are sensitive to increases in CO_2 from pre-industrial levels. *Nature* 327, 617–18.

Young, J.J., Mehta, S., Israelsson, M., Godoski, J., Grill, E. and **Schroeder, J.I.** 2006: CO_2 signaling in guard cells: calcium sensitivity response modulation, a Ca^{2+}-independent phase, and CO_2 insensitivity of the gca2 mutant. *Proceedings of the National Academy of Sciences of the United States of America* 103, 7506–11.