



Comparison of Holocene temperature reconstructions based on GISP2 multiple-gas-isotope measurements

Michael Döring^{1,2*} and Markus Christian Leuenberger^{1,2}

¹Climate and Environmental Physics, University of Bern, Switzerland

5 ²Oeschger Centre for Climate Change Research (OCCR), Bern, Switzerland

*Correspondence to: Michael Döring (michael.doering@climate.unibe.ch)

Keywords: temperature reconstruction, ice core, nitrogen isotope, argon isotope, inverse-model, firn-model, accumulation-rate

Abstract: Nitrogen and argon stable-isotope data extracted from ancient air in ice cores provides the possibility to reconstruct Greenland past temperatures when inverting firn-densification and heat-diffusion models (firn-models) to fit the gas-isotope data ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$, $\delta^{15}\text{N}_{\text{excess}}$). This study uses the Döring and Leuenberger (2018) fitting-algorithm coupled on two state of the art firn-models to fit multiple Holocene gas-isotope data measured on the GISP2 ice core. We present for the first time the resulting temperature estimates when fitting $\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$ as single targets with misfits generally in the low permeg level. Whereas the comparison between the reconstructions using $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ shows a high agreement, the use of $\delta^{15}\text{N}_{\text{excess}}$ for reconstructing temperature is problematic, due to higher statistical and systematic data uncertainty influencing especially multi-decadal to multi-centennial signals, and results in an unrealistic temperature estimate that differs significantly from the two other reconstructions. We find evidence for systematic too high $\delta^{40}\text{Ar}$ data in the early- and late-Holocene potentially caused by post coring gas-loss or an insufficient correction of this mechanism. Next, we compare the performance of the Goujon et al. (2003) firn-model and the Schwander et al. (1997) firn-model for Holocene temperature reconstructions. Besides small differences of the reconstructed temperature anomalies – potentially caused by slightly different implementation of firn physics and parameters in the two models – the reconstructed temperature anomalies are highly comparable. We were able to quantify the contribution of the firn-model difference to the uncertainty budget of our reconstruction. Furthermore, the fractions of uncertainty on the reconstructed temperatures, arising from the non-perfect reproducibility of the fitting algorithm and from the remaining final misfits (low permeg level), were quantified. Together with the published measurement uncertainty of the gas-isotope data and the analysis of the impact of accumulation-rate uncertainty on the reconstruction, we were able to calculate the mean uncertainty (2σ) for the nitrogen and the argon based temperature estimates with $2\sigma_T = 0.80 \dots 0.88$ K for $T(\delta^{15}\text{N})$, and $2\sigma_T = 0.87 \dots 1.81$ K for $T(\delta^{40}\text{Ar})$, respectively. Finally, we compare our reconstructed temperatures to two recent reconstructions based on the same gas-isotope data as used here, but following different reconstruction strategies: first the study of Buizert et al. (2018), which uses a combination of $\delta^{18}\text{O}_{\text{ice}}$ -calibration and $\delta^{15}\text{N}$ -fitting, and second the study of Kobashi et al. (2017), where $\delta^{15}\text{N}_{\text{excess}}$ was fitted in order to conduct the temperature reconstruction. We find generally higher agreement between our $T(\delta^{15}\text{N})$ estimate and the Buizert et al. (2018) temperature – in terms of variability and correlation in three investigated periodic-time bands (multi-decadal, multi-centennial and multi-millennial) – as if our $T(\delta^{15}\text{N})$ reconstruction is compared to the Kobashi et al. (2017) temperature. However, all three reconstruction strategies lead to distinct temperature realizations.



1 Introduction

The use of nitrogen ($\delta^{15}\text{N}$) and argon ($\delta^{40}\text{Ar}$) stable-isotope variations in air extracted from ice cores is a relatively new tool for reconstructing past temperature (e.g. Huber et al., 2006b; Kindler et al., 2014; Kobashi et al., 2011; Landais et al., 2006; Orsi et al., 2014; Severinghaus et al., 1998, 2001). This method uses the stability of isotopic compositions of nitrogen and argon in the atmosphere at orbital timescales, as well as the fact that changes are only driven by processes in polar firm (Leuenberger et al., 1999; Mariotti, 1983; Severinghaus et al., 1998) and provides an alternative to the classical calibration of the stable oxygen ($\delta^{18}\text{O}$) and hydrogen isotopes extracted from the ice-core-water samples (Gierz et al., 2017; Johnsen et al., 2001; Steen-Larsen et al., 2011; Stuiver et al., 1995). The isotopic composition of the water samples provides a rather robust proxy for reconstructing paleo-temperatures for times where large temperature variations occur (Gierz et al., 2017). In the Holocene, where temperature variations are comparatively small, changes in seasonal distribution of precipitation as well as of evaporation conditions at the source region may dominate water-isotope-data variations (Huber et al., 2006b; Kindler et al., 2014; Werner et al., 2001). Recent studies (Buizert et al., 2018; Kobashi et al., 2017) used the nitrogen and argon isotopes of the GISP2 ice core (Greenland Ice Sheet Project Two, Meese et al., 1994; Rasmussen et al., 2008; Seierstad et al., 2014) to reconstruct Holocene temperature variations for Greenland summit following different reconstruction strategies. Kobashi et al. (2017) use the second-order parameter $\delta^{15}\text{N}_{\text{excess}}$ ($\equiv \delta^{15}\text{N} - \delta^{40}\text{Ar}/4$) together with the firm-densification and heat-diffusion model from Goujon et al. (2003) to obtain a Holocene temperature estimate. Buizert et al. (2018) reconstructed summit temperatures by calibrating the GISP2 $\delta^{18}\text{O}$ data by forcing the temperature to reproduce the general trend in $\delta^{15}\text{N}$ using a dynamical firm-model. Both methods lead to different temperature estimates. In Buizert et al. (2018) an overall uncertainty of 1.5 K was stated for the reconstructed temperature. Kobashi et al. (2017) estimated the uncertainty of the temperature reconstruction by examining the variance of temperature realizations when shifting the $\delta^{15}\text{N}_{\text{excess}}$ data in the range of analytical uncertainty before using his fitting approach. This approach results in an averaged uncertainty of 1.21 K (1σ). Both approaches have in our view shortcomings, $\delta^{15}\text{N}_{\text{excess}}$ loses information about the Lock-in-depth (LID) and scaling to $\delta^{18}\text{O}_{\text{ice}}$ does not consider side-effects of water isotopes to seasonal distribution of precipitation. Döring and Leuenberger (2018) showed an automated approach enabling fitting gas-isotope data with an outstanding accuracy with mismatches generally below the analytical uncertainty of the isotope measurements. It was shown on synthetic data experiments that in the case of perfectly known accumulation-rate data and neglecting noise, the remaining mismatches would lead to a temperature uncertainty (2σ) below 0.3 K for a single measurement and Holocene-like conditions. This study focuses on the challenges of temperature reconstructions using gas-isotope fitting for real Holocene data. We will discuss different aspects which are in our view integral for the evaluation of the correctness of the reconstructed temperature estimates. First, we will discuss the gas-isotope data in the context of measurement uncertainty and focus on the suitability of the different isotopic quantities for reconstructing robust temperature estimates. Next, we discuss the reproducibility and the contribution of the final misfits to the uncertainty budget using our fitting approach. In addition, we will show the influence of different accumulation-rate estimates on the temperature reconstruction. Finally, we compare the temperature solutions obtained by fitting $\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$ to each other and place them in a context to the estimates of Kobashi et al. (2017) and Buizert et al. (2018). For our reconstruction we used two different firm-models, the models of Schwander et al. (1997) and of Goujon et al. (2003). We will compare our results using both and we will provide an overall uncertainty of our method for the most robust estimate using all available information.



2 Data and method

2.1 Firn-models and inversion algorithm

The observed gas-isotope data mainly relies on firn densification processes combined with gas and heat diffusion (Severinghaus et al., 1998). So the use of firn-densification and heat-diffusion models (from now on referred to as
 5 firn-model) describing the physics of densification and heat and gas transport are necessary for inverting measured gas-isotope data to surface temperatures. In this study we use two classic firn-models. The first model was developed by Schwander et al. (1997) and used for the temperature reconstructions by Huber et al. (2006b) and Kindler et al. (2014). The second one, the model from Goujon et al. (2003) (adapted to the GISP2 site for this study), was used i.e. in the studies by Guillevic et al. (2013), Kobashi et al. (2015), and Kobashi et al. (2017). As
 10 a comparison of temperature reconstructions for Holocene conditions using both models is done for the first time in this study, we will prove the comparability of the solutions gained by using both models.

The conversion of gas-isotope data to surface temperature estimates using a firn-model is an inverse problem. The firn-model acts as a non-linear transfer function, combining temperatures and accumulation-rates with the gas-isotope data. To solve this problem we use the automated fitting algorithm developed by Döring and Leuenberger
 15 (2018). The used algorithm allows the fitting of the gas-isotope data with misfits in the low permeg level, mainly below the analytic measurement uncertainties. For modelling $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$ data, we use, in addition to the details presented in Döring and Leuenberger (2018), the thermal-diffusion constant $\alpha_{T,\text{Ar}}$ and thermal-diffusion sensitivity Ω_{Ar} that have been empirically derived by Grachev and Severinghaus (2003):

$$\alpha_{T,\text{Ar}}(t) = \left(26.08 - \frac{3952}{\bar{T}(t)} \right) \cdot 10^{-3} \quad (1)$$

$$\Omega_{\text{Ar}}(t) = \frac{\alpha_{T,\text{Ar}} \cdot 10^3}{\bar{T}(t)} = \frac{26.08}{\bar{T}(t)} - \frac{3952}{\bar{T}(t)^2} \quad (2)$$

$\bar{T}(t)$ is the mean firn temperature (Leuenberger et al., 1999).

2.2 Timescale, and necessary data

Timescale

For the entire study, the GICC05 chronology is used (Rasmussen et al., 2014; Seierstad et al., 2014). Following Döring and Leuenberger (2018), the temperature input is split into two parts in time. The first part ranges from 10.5 kyr to 35 kyr b2k (“spin-up section”), whereas the second part ranges from 0.02 kyr to 11.5 kyr b2k
 25 (“reconstruction window”) for which we allow the fitting algorithm to change the temperature. The accumulation-rate as well as the surface temperature of the spin-up section remain unchanged during the reconstruction procedure.

Accumulation-rate input data

In addition to temperature, accurate accumulation-rate data are needed to drive the firn-models. As in Döring and Leuenberger (2018), we use as accumulation-rate input the original accumulation-rates (acc) for the GISP2 site,
 30 as reconstructed in Cuffey and Clow (1997), but adapted to the GICC05 chronology. We use for the reconstruction all three given accumulation-rate datasets to analyse the impact of differences between the three accumulation estimates on the reconstructed temperatures. To obtain the accumulation rate estimates three different scenarios of



ice sheet margin retreat were used (50 km, 100 km or 200 km scenario) as boundary conditions for an ice flow model (see Cuffey and Clow (1997) for details; availability of the accumulation rates using the 100 km scenario on GICC05 timescale: //doi.pangaea.de/10.1594/PANGAEA.888997).

Target data and their suitability due to analytical uncertainty

5 The GISP2 gas-isotope data measured by Kobashi et al. (2008b) will be used for the Holocene temperature reconstruction as targets for our fitting algorithm (Döring and Leuenberger, 2018). Figure 1 shows the different gas-isotope targets ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$, $\delta^{15}\text{N}_{\text{excess}}$) set on the GICC05 chronology. In tab. S1 we list measurement uncertainties for the target data given in Kobashi et al. (2008b). As different uncertainties were given in Kobashi et al. (2008b) for different measurement campaigns, we consequently use the minimal and maximal uncertainties

10 given in Kobashi et al. (2008b) for a signal-to-noise analysis of these data for two cases, signals with periodicities $T < 500$ yr (“high frequency”) and $T > 500$ yr (“low frequency”), respectively. Here we focus especially on the faster signal ($T < 500$ yr) in order to analyse the suitability in relation to the analytical uncertainty of the different isotope targets ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$, $\delta^{15}\text{N}_{\text{excess}}$) for reconstructing multi-decadal to multi-centennial temperature variability. We conducted the analysis in the following way: First, we detrended the measured gas-isotope target data by

15 subtracting the respective low-pass-filtered signals using a cut-off-period of 500 yr (fig. S1a-c). In a next step, we identified the local maxima and minima of the high-frequency isotope data (using the Matlab® “findpeaks” algorithm). We defined the high-frequency signals as the difference between successive local maxima and minima (fig. S1d-f) and compared the high-frequency signals to the signal uncertainties calculated from the published measurement uncertainties (Kobashi et al., 2008b). As a signal is defined by at least two points, the signal

20 uncertainties are calculated using Gaussian-error-propagation. We use for our calculations the minimum (red dotted line, fig. S1) and maximum (blue dotted line, fig. S1) uncertainties (tab. S1). The analysis of the high-frequency signals shows that for the minimum uncertainties 78% of the $\delta^{15}\text{N}$ high-frequency signals have amplitudes higher than the uncertainty level (70% for the maximum uncertainty), 74% (or 36%) for $\delta^{40}\text{Ar}$, and only 52% (or 17%) for $\delta^{15}\text{N}_{\text{excess}}$, respectively. Assuming that the “true” uncertainty is in-between the given

25 maximum and minimum uncertainties and keeping in mind that the listed measurement uncertainties are 1σ uncertainty, we argue that only $\delta^{15}\text{N}$ is suitable as a robust reconstruction target in the high-frequency case. The histograms in fig. 1 show a detailed listing of the signal-to-noise ratios (SNRs) for all isotope species and for the minimum and maximum uncertainties, respectively. Here the signals are grouped between integer SNR values. It is clearly visible that for $\delta^{15}\text{N}$ most signals have an SNR between one and two or even higher values for the

30 minimum as well as the maximum uncertainty. In contrast, for $\delta^{40}\text{Ar}$ and especially for $\delta^{15}\text{N}_{\text{excess}}$ the dominant fraction of signals has SNR values lower than the uncertainty values ($\text{SNR} < 1$), which makes it challenging to extract a robust temperature estimate for multi-decadal to multi-centennial signals from these targets.

For the longer-term isotope trends ($T > 500$ yr) it is more challenging to extract a comparable result (fig. S1g-i). Here we divided the analytical uncertainty by a factor of about 5.3 to account for the smoothing (mean data

35 resolution = 17.8yr, $(500 \text{ yr}/17.8 \text{ yr})^{1/2} = 5.3$). The comparison of the minimum or maximum measurement uncertainties (red or blue error bars) indicates that all three gas-isotope quantities are suitable for reconstructing long-term temperature trends. This is particularly correct when only measurement uncertainty is considered, as these uncertainties are in most of the cases lower than the amplitudes of the investigated features. However, $\delta^{15}\text{N}$ is also the most suitable target for reconstructing long-term temperature trends due to its relatively small

40 uncertainty compared to $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$.

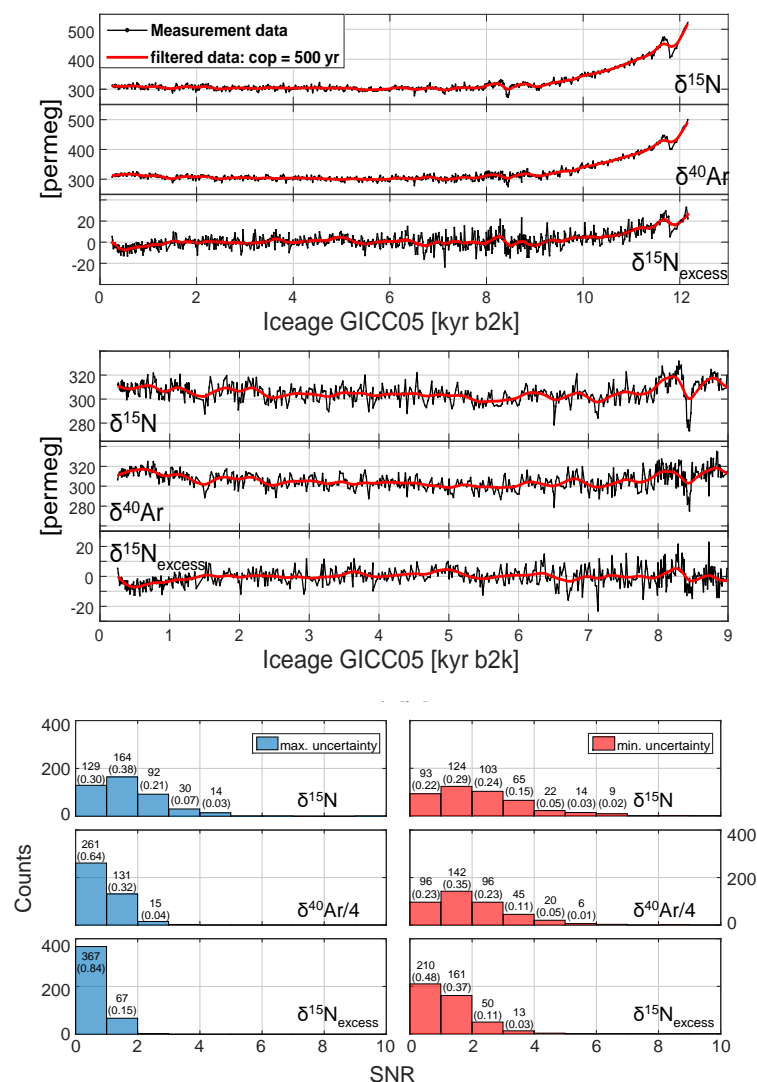


Figure 1: Gas-isotope target data on GICC05 time scale (Kobashi et al., 2008b): Upper three plots: full $\delta^{15}\text{N}$, $\delta^{40}\text{Ar}/4$ and $\delta^{15}\text{N}_{\text{excess}}$ time-series; Middle three plots: zoom in for the recent 9 kyr time window of the same quantities. **5** **Histograms:** Signal-to-noise-ratios (SNR) of the high-frequency signals ($T < 500$ yr) for all reconstruction targets and for the maximum (left hand side) as well as the minimum (right hand side) signal uncertainties (see text). Values in the plots indicate the absolute (relative) number of signals with SNR in between the respective limits.

Problems with $\delta^{15}\text{N}_{\text{excess}}$ due to gas-loss fractionation:

It must be stressed that using $\delta^{15}\text{N}_{\text{excess}}$ only (as single target) for the reconstruction involves the danger of **10** incorporating large drifts in the temperature solution if systematic offsets in $\delta^{15}\text{N}_{\text{excess}}$ are present, since the calculated temperature is solely dependent on the temporal integration of firn-temperature gradients directly calculated from $\delta^{15}\text{N}_{\text{excess}}$. This can alter centennial or even millennial scale temperature variability. It is known that smaller molecules such as Ar suffer from kinetic fractionation due to gas-loss either during bubble close-off or core retrieval and storage which can lead to an enrichment of $\delta^{40}\text{Ar}$ and thus to smaller $\delta^{15}\text{N}_{\text{excess}}$ (Huber et al.,



2006a; Kobashi et al., 2008b, 2010, 2011; Severinghaus et al., 2003; Severinghaus and Battle, 2006). This made a correction of the GISP2 $\delta^{40}\text{Ar}$ data mandatory, which was done based on firm-modelling (Kobashi et al., 2008a) and on $\delta\text{Ar}/\text{N}_2$ (a tracer for potential gas-loss; Kobashi et al., 2010, 2017), leading to smaller $\delta^{40}\text{Ar}$ values compared to the uncorrected $\delta^{40}\text{Ar}$ and thus to higher $\delta^{15}\text{N}_{\text{excess}}$. As we will show in sects. 4.1.2 and 4.3, the executed

5 correction is non-sufficient, especially for the late- and early-Holocene data, which implies that the gas-loss induced fractionation in $\delta^{40}\text{Ar}$ is not constant over the Holocene part of the GISP2 ice core. A better correction approach would be the use of additional isotope quantities ($\delta^{86}\text{Kr}$, $\delta^{136}\text{Xe}$) measured together with $\delta^{40}\text{Ar}$ on the same samples as proposed by Baggenstos (2015). Additionally, the elimination of the gravitational signal using $\delta^{15}\text{N}_{\text{excess}}$ as single target leads to a loss of information (firm column height) and to a less constraint temperature

10 solution with reduced reproducibility and this complicates the firm-model inversion (sect. 3.1.1). The linear dependency (slope) between $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}/4$ (tab. 1) is calculated using geometric-mean-regression (Leng et al., 2007) in order to investigate the contribution of different processes altering the isotope data. Furthermore, the slope is calculated in three periodic-time bands (multi-decadal, multi-centennial, and multi-millennial) and for the raw data. As the variability of the isotope data is created in the firm due to gravitational

15 settling (mass dependent fractionation process) and thermal diffusion (dependent on the firm temperature gradient), the slope between $\delta^{15}\text{N}(\text{y})$ and $\delta^{40}\text{Ar}/4(\text{x})$ should be in the range of one (gravitational settling only) to 1.46 (thermal diffusion only, $\Omega_{\text{N}}/[\Omega_{\text{Ar}}/4]$). Also, both processes affecting $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ are in the same “direction” and in consequence, a generally high correlation is expected. If we compare the results for the different periodic-time bands, it is obvious that the slope calculated for the multi-decadal oscillations cannot be explained neither by

20 gravitational settling nor by thermal diffusion. For these fast oscillations the gravitational background is not expected to change significantly, and the slope should correspond mainly to the thermal diffusion value of 1.46. In contrast, the calculated slope is 0.89 ± 0.05 and the correlation is relatively weak ($r^2 = 0.46$). A slope less than one is pointing to a process which further enriches $\delta^{40}\text{Ar}/4$ compared to $\delta^{15}\text{N}$, which is the case for a potential gas-loss contribution on $\delta^{40}\text{Ar}$ which seems still remaining after the correction of $\delta^{40}\text{Ar}$. The weak correlation shows

25 a decoupling between $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ for multi-decadal variability, which can be partly attributed to analytical uncertainty. For the longer periodicities we find higher correlation ($r^2 > 0.8$) and slopes in the range of expectation (tab. 1). The decrease of the slope from multi-centennial to multi-millennial variability shows that with longer periodic-time the influence of gravitational fractionation due to changes of firm column height gets more and more important compared to the thermal diffusion signal. This is expected as (i) the firm column reacts slowly and (ii)

30 the temperature gradient over the firm column vanishes.



Periodic-time band	r^2	slope
multi-decadal (50-200 yr)	0.46	0.89 ± 0.05
multi-centennial (200 yr-1 kyr)	0.87	1.29 ± 0.04
multi-millennial (1 kyr-4 kyr)	0.87	1.14 ± 0.03
unfiltered data	0.96	1.14 ± 0.02
theoretical values:		
thermal diffusion only		1.46
gravitational settling only		1.00

Table 1: Slopes and correlation coefficients between $\delta^{15}\text{N}(\text{y})$ and $\delta^{40}\text{Ar}/4(\text{x})$ derived using geometric-mean-regression. The theoretical value for thermal diffusion was calculated as ratio of the thermal-diffusion-sensitivities as $\Omega_{\text{N}}/[\Omega_{\text{Ar}}/4]$.

Prior input and model spin-up:

To avoid the influence of possible memory effects (influence of earlier firn-state conditions on later firn-states), a temperature and accumulation-rate spin-up is needed in order to bring the firn-model to a well-defined starting condition. The surface temperature spin-up was obtained by extending the temperature reconstruction for the GISP2 site from Buizert et al. (2014) (interval 10.05 kyr to 20 kyr b2k) to 35 kyr b2k by calibrating the GISP2 $\delta^{18}\text{O}_{\text{ice}}$ data (Grootes et al., 1993; Grootes and Stuiver, 1997; Meese et al., 1994; Steig et al., 1994; Stuiver et al., 1995; data availability: Grootes and Stuiver, 1999), using the slopes and intercepts given in Cuffey and Clow (1997). As prior input for the Holocene section, we simply start with constant temperature using the last value of the spin-up section. The prior temperature input and the spin-up temperature were slightly adjusted in order to match the decreasing flank at the oldest part (9.5 kyr to 12.168 kyr b2k) of the gas-isotope data. The adjustment procedure is described in detail in supplement sect. S2.

3 Results and discussion: gas-isotope fitting

3.1 Gas-isotope fitting results and uncertainty due to methodology

In the following paragraphs we evaluate different factors contributing to the final uncertainty budget of the temperature reconstructions when fitting all different targets ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$). These factors are:

- The reproducibility of the resulting solutions over ten fitting runs (sect. 3.1.1). As we use a Monte-Carlo based method, each fitting run follows a unique pathway. In consequence, the obtained final temperature estimates slightly differ between different fitting runs on the very same target.
- The “goodness” of the fits (sect. 3.1.1). As discussed in Döring and Leuenberger (2018), it is not possible to fit the gas isotope data with an overall misfit of zero. The remaining final mismatches between modelled and measurement data contribute to the uncertainty of the reconstruction.
- The uncertainty in the used accumulation rate data was incorporated in the temperature reconstruction (sect. 3.1.3), as we used all three different accumulation rate data sets for the GISP2 site provided by Cuffey and Clow (1997).



- (iv) Differences in the reconstructed temperatures resulting from the usage of different firn-models (sect. 4.1.1). This is sketched by analysing the differences between the reconstructed temperatures using two different firn-models. It would be beneficial to incorporate more available models to better quantify the uncertainty due to firn-model differences in later studies.
- 5 (v) The fraction of uncertainty in the reconstructed temperatures due to analytical uncertainty in the used isotope data (sect. 4.2).

3.1.1 Reproducibility and final misfits:

In order to analyse the reproducibility of the gas-isotope fitting algorithm coupled to the Schwander et al. (1997) firn-model, ten fitting runs were conducted for each gas-isotope species ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$). From the ten solutions for each isotope a mean-solution was calculated as the average of the ten temperature solutions. This mean-solution was processed by the firn-model leading to the final modelled isotope solution. The gained data were analysed for the reproducibility between the ten runs (top six plots of fig. 2) to estimate the possible spread in absolute temperature and isotope solutions for each isotope target. In addition, we compare the final (mean) solution with the best-fit solution (out of ten) regarding the remaining mismatches (fig. 3). Table S2 contains additional information for the reproducibility (“rep”) as well as the goodness of the fits (“fit”). It is clearly visible that the spread of the isotope solutions between the ten runs is in the low permeg level for all targets with mean spreads over the whole reconstruction time-window (0.255 kyr to 12.168 kyr iceage b2k) of 1.68 ± 1.14 permeg for $\delta^{15}\text{N}$, 2.58 ± 1.77 permeg for $\delta^{40}\text{Ar}/4$, and 1.28 ± 2.14 permeg for $\delta^{15}\text{N}_{\text{excess}}$. The low variances between the modelled isotope solutions show the robust performance of our gas-isotope fitting algorithm. Similarly, the spread between the gained temperatures was analysed (bottom 6 plots of fig. 2). Using $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ as single targets lead to temperature solutions varying in a narrow band of 0.17 ± 0.12 K for $T(\delta^{15}\text{N})$ and 0.26 ± 0.18 K for $T(\delta^{40}\text{Ar})$. In contrast to the excellent reproducibility of $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$, the fitting of $\delta^{15}\text{N}_{\text{excess}}$ generates a wide spread of possible temperature realizations which contrasts with the robustness of the modelled isotope solutions. Whereas the variance between the isotope solutions is comparable for all three species, the spread of $T(\delta^{15}\text{N}_{\text{excess}})$ is about 10 times (2.04 ± 1.90 K) higher as for $T(\delta^{15}\text{N})$ or $T(\delta^{40}\text{Ar})$. This is due to the removal of all information about the gravitational component in the isotope signals when calculating $\delta^{15}\text{N}_{\text{excess}}$, and therefore the information about the height of the firn column is lost, which leads to a significantly extended space of firn-states or absolute temperatures on which the fitting algorithm can work to yield very similar modelled isotope solutions. Besides the spread in absolute temperature, the relative temperature variations (deviation from the trend) gained by fitting $\delta^{15}\text{N}_{\text{excess}}$ are very much comparable to each other. However, the different densification backgrounds lead to differences in the time evolution of Δage of about 20 yr to 30 yr, and therefore in asynchronous temperature estimates (fig. S3). Due to this, fitting of $\delta^{15}\text{N}_{\text{excess}}$ as single target makes it challenging to determine the right timing of temperature changes on a multi-decadal scale. Interestingly, fitting of $\delta^{15}\text{N}_{\text{excess}}$ with the Goujon et al. (2003) model leads to a smaller spread of possible temperature estimates compared to the Schwander et al. (1997) model. The reproducibility between 10 runs (fig. S4) is 2.7 times better when the Goujon model is used. The reason for that difference was not found yet. The implementation of the geothermal heat flux in the Goujon model provides a negative constant fraction to ΔT_{firn} and may lead to this stabilization effect. Figure S3 shows the modelled Δage and LID data for all fitting targets ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$, $\delta^{15}\text{N}_{\text{excess}}$) and all 10 runs. Whereas the variance between the modelled Δage and LID is very small for $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$, fitting of $\delta^{15}\text{N}_{\text{excess}}$ creates a variety of LID and Δage states. While running the gas-isotope fitting algorithm several times on the same target, we notice a boundary



effect for the last 500 yr to 1 kyr to today. Here different temperature solutions emerge, while the rest of the time-series is highly reproducible. To overcome this boundary effect, a stabilization of the solutions for the last 500 yr to today was necessary as described in supplement sect. S4. To solve this issue, additional information is needed, which was added by using the measured borehole temperature profile for the GISP2 site as an additional constraint.

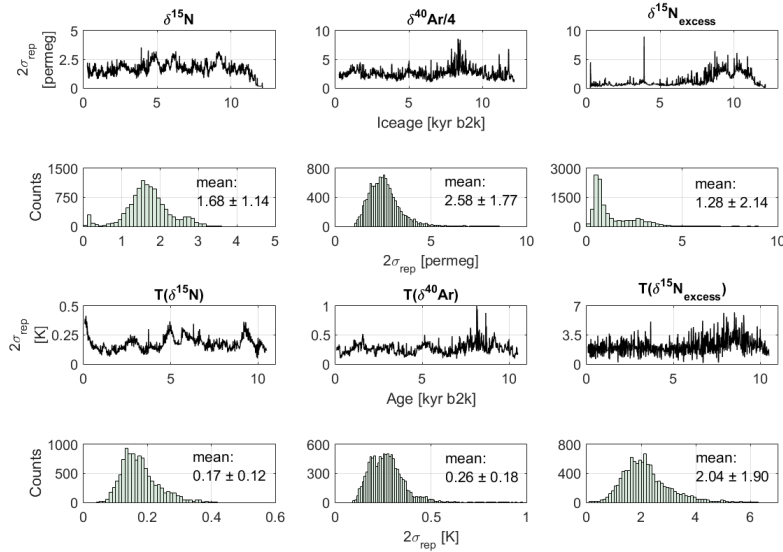


Figure 2: Reproducibility between 10 runs for each target (first row: $\delta^{15}\text{N}$ and $T(\delta^{15}\text{N})$ using the fitting algorithm coupled to the firn-model from Schwander et al. (1997); second row: $\delta^{40}\text{Ar}$ and $T(\delta^{40}\text{Ar})$; third row: $\delta^{15}\text{N}_{\text{excess}}$ and $T(\delta^{15}\text{N}_{\text{excess}})$). Top 6 plots: reproducibility for the modelled isotopes per yr. Bottom 6 plots: reproducibility for the reconstructed temperatures per yr. Values are mean \pm 2 σ . See also table S2.

The evaluation of the mismatches between the measured and modelled time-series gives a constrain on the uncertainty budget of the final temperature. It is obvious that for all targets the mismatches between the measured and modelled isotope data are at least comparable or below the analytic uncertainty (which is 1 σ) of the measurement data (tab. S1). Using the average temperatures, we reach final mismatches (2 σ) of 3.65 permeg for $\delta^{15}\text{N}$, 2.79 permeg for $\delta^{40}\text{Ar}/4$, and 5.43 permeg for $\delta^{15}\text{N}_{\text{excess}}$ (fig. 3). Interestingly, averaging the $\delta^{15}\text{N}$ - and $\delta^{40}\text{Ar}$ temperature solutions leads to a further decrease of the mismatch compared to the best fit out of the ten runs (4.5 % for $\delta^{15}\text{N}$, 18.9 % for $\delta^{40}\text{Ar}$). It seems that the averaging of the ten temperature solutions corrects some of the remaining (potentially randomly distributed) mismatches. Obviously, a larger number of runs (> 10) would slightly improve the mean solution. As discussed above, the averaging of the temperature solutions gained from $\delta^{15}\text{N}_{\text{excess}}$ -fitting is problematic, due to the wider spread between the temperature solutions and thereof worse constrained Δage . Consequently, the averaging leads to an increase of the mismatches compared to the single fits. The averaging of the single $T(\delta^{15}\text{N}_{\text{excess}})$ solutions leads to more than a doubling (factor 2.55) of the mismatches compared to the best-fit solution.

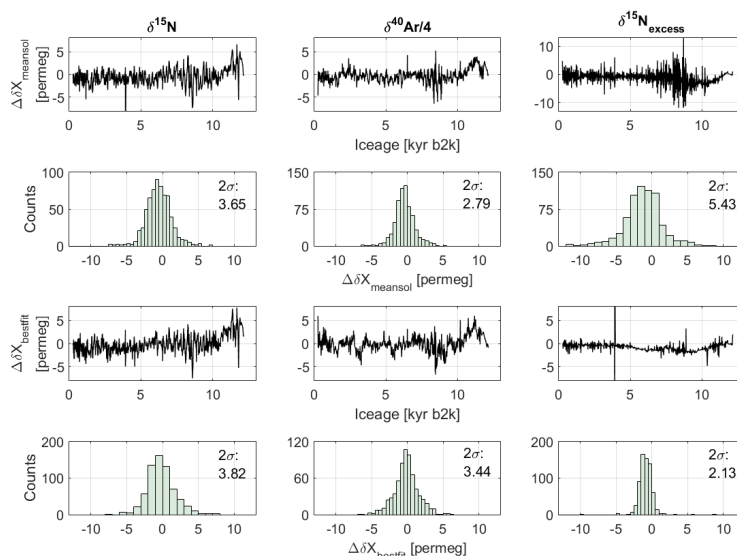


Figure 3: Top six plots: Mismatches between the modelled and measured isotopes of the final (mean) solution. The numbers in the histograms are the 2σ values of all pointwise mismatches. Bottom 6 plots: Mismatches between the modelled and measured isotopes of the best fit out of ten runs. The numbers in the histograms are the 2σ values of all pointwise mismatches. Interesting is the decrease of the mismatches of the mean solution compared to the best fit for $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$. $\delta^{15}\text{N}_{\text{excess}}$ shows a reverse behaviour due to the wider spread of “possible” temperature solutions (figure 2 and table S2).

3.1.2 Comparison of misfits among the reconstructions

The histograms on the right-hand-side of fig. 4 show the mean mismatches for all gas-isotope quantities when fitting a single isotope target using the Schwander et al. (1997) (black bars) or the Goujon et al. (2003) (blue bars) firn-model. E.g. precise fitting of $\delta^{15}\text{N}$ (top plot, where 96% of the mismatches (2σ) are smaller than 3.7 permeg using Schwander et al. (1997) firn-model) leads to insufficient fits for the other gas-isotope targets with mismatches (2σ values) of 11.3 permeg for $\delta^{40}\text{Ar}/4$ and 11.1 permeg for $\delta^{15}\text{N}_{\text{excess}}$, keeping in mind that the data uncertainty (1σ) of the measured gas isotope data is 3.0 to 4.0 permeg for $\delta^{15}\text{N}$, 4.0 to 9.0 permeg for $\delta^{40}\text{Ar}/4$ and 5.0 to 9.8 permeg for $\delta^{15}\text{N}_{\text{excess}}$. In other words, precise fitting of $\delta^{15}\text{N}$ does not automatically lead to accurate fits for $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$. The same is true for all other single fits and when using both firn-models. It is not possible to find a temperature estimate leading to modelled isotope regimes that provide a sufficient agreement for all isotopic targets together. This finding is pointing to the issues that the GISP2 gas-isotope data suffers from fractionations which are not captured by the used firn-models (e.g. gas-loss).

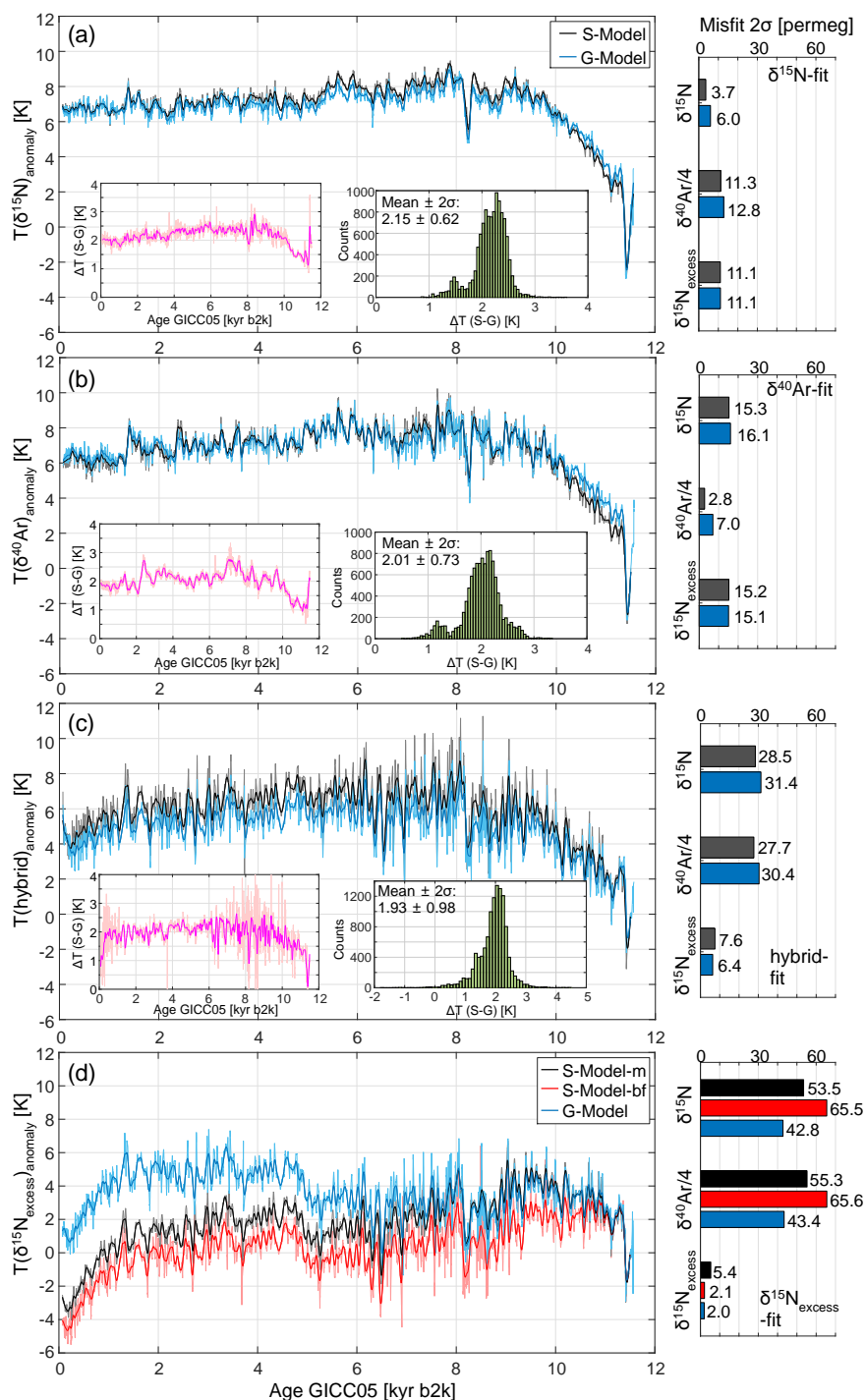


Figure 4: Temperature solutions for all targets with differences between Schwander and Goujon modelling (histogram in the plots), and mean misfits for all species (histogram right hand side).



3.1.3 Influence of different accumulation-rate estimates

To investigate the contribution of the uncertainty in the used accumulation-rate data, we use all three available accumulation-rate estimates for the GISP2 site (sect. 2.2) to reconstruct temperature on the base of $\delta^{15}\text{N}$ -fitting. It is visible that the deviation between the three different scenarios can be up to more than $\pm 10\%$ in the early-Holocene and decreases over time (fig. 4b). The deviations between the three scenarios have a minor influence on the modelled Δage during the Holocene (fig. 5c,d). Starting with a maximum difference of about 30 yr in the early-Holocene, the Δage difference decreases until 5 kyr b2k following the decrease in the difference between the accumulation-rate data. From the mid-Holocene until today the modelled Δage difference becomes less than 5 yr. The same is true for the spread in LID (fig. 5e). Here also the effect of the adjustment is visible, forcing the same firn-state for all accumulation-rate scenarios at the beginning of the reconstruction window (supplement sect. S2). Whereas the reconstructed temperatures when using the 50 km and 100 km accumulation-rate scenarios lead mainly to the same temperature trend, the reconstruction using the 200 km scenario shows a slightly larger cooling over the Holocene starting from about 7.5 kyr b2k (fig. 5f). This is exactly the point in time where the decrease in accumulation-rate for the 200 km scenario compared to the averaged scenario starts to accelerate (fig. 5b). As we do not see an equal (but opposite) behaviour for the 50 km scenario, we may have found a non-linear response between temperature and accumulation-rate change. From about 6 kyr to today, this faster cooling starting at 7.5 kyr levels out and leads to a constant offset of about 0.3 K between the 200 km solution and the two others, which is nearly 15% of the whole cooling trend of about 2 K referred to the warmest part of the reconstructed temperatures at around 7.8 kyr b2k. Besides this, the shapes and amplitudes of the faster signals of the long-term fractions ($T > 500$ yr) are highly comparable. The relatively large deviation of Δage between the scenarios in the early-Holocene leads to a slightly asynchronous behaviour of the short-term temperature variations ($T < 500$ yr, fig. 5g,h). But the shapes and amplitudes of the signals are independent from the deviations between the accumulation-rate scenarios. The decreasing deviation between the accumulation-rate scenarios and thereof between the modelled Δage over the Holocene leaves no differences between the short-term fractions of the reconstructed temperatures in the late-Holocene section (fig. 5h). To sum up, the deviation between the three different accumulation-rate scenarios do not have a major impact on the reconstructed temperature anomalies. The differences between the accumulation-rates lead to slightly different modelled Δage in the early-Holocene and to a 0.3 K larger cooling for the higher accumulation-rate scenario compared to the two other ones.

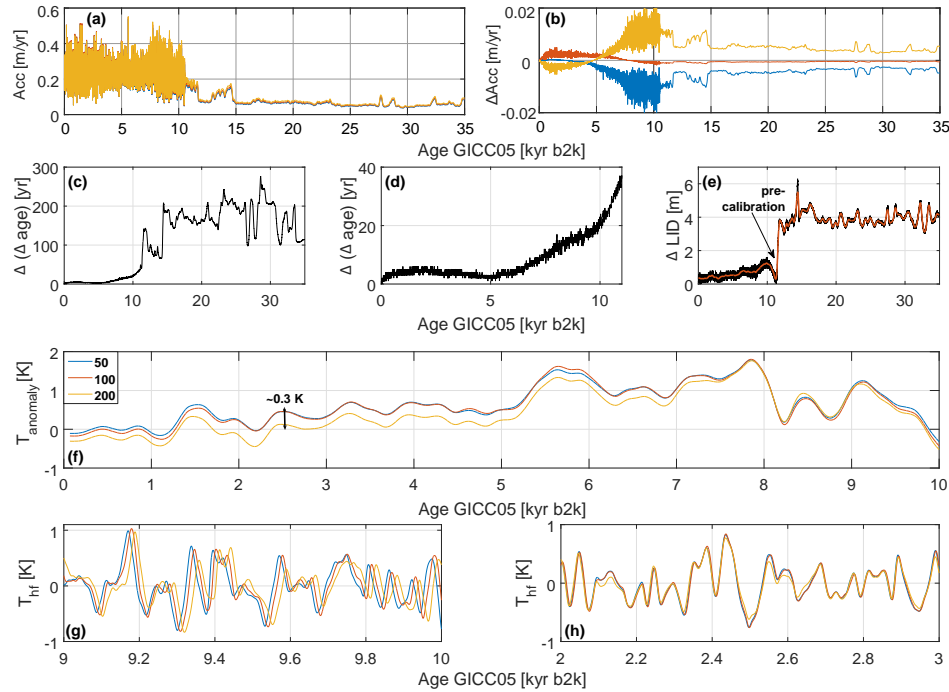


Figure 5: Influence of different accumulation-rate scenarios ((a): blue: 50 km, red: 100 km, yellow: 200 km) on the reconstruction. B: Deviation of each single scenario to the average of all three scenarios. (c): Maximum spread in modelled Δ age using the three different accumulation-rate scenarios for the whole input (Holocene and spin-up). (d): Zoom-in for (a) for the Holocene part. (e): Maximum spread in the modelled Lock-In-Depth (LID), the pre-calibration leads to a convergence of the LID and therefore to the same gravitational background for the isotope signals. (f): Long-term temperature trend T_{anomaly} (low-pass: > 500 yr) of the reconstructed temperatures modelled using the three accumulation-rate scenarios. (g): Short-term temperature signals T_{hf} (high-pass: < 500yr) showing asynchrony in the early-Holocene, as a result of the spread in Δ age. (h): Short-term temperature signals T_{hf} (high-pass: < 500yr) showing synchrony from the mid- to late-Holocene, as a result of the decreasing spread in Δ age and the decreasing difference between the accumulation-rates.

4 Results and discussion: temperature reconstruction

In this section we show and discuss the temperature estimates emerging when fitting the different gas-isotope targets independently ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$, $\delta^{15}\text{N}_{\text{excess}}$) using the Schwander et al. (1997) and Goujon et al. (2003) firn-model (sects. 4.1. and 4.2.). All temperature estimates are shown as anomalies (relative to 11.3 kyr b2k). The modelled firn parameters (Δ age, LID) for each target are shown in fig. S3. Additionally, a so-called hybrid solution is created using $\delta^{15}\text{N}$ together with $\delta^{15}\text{N}_{\text{excess}}$, as follows: The hybrid solution is created from the mean temperature solution of $\delta^{15}\text{N}$ -fitting, low-pass filtered with a cut-off-period of 500 yr giving a long-term temperature trend. This long-term temperature trend is superimposed by adding high-frequency information calculated from $\delta^{15}\text{N}_{\text{excess}}$. The high-frequency temperatures are calculated by translating the mismatches of modelled – using the long-term temperature trend (from $\delta^{15}\text{N}$) – and measured $\delta^{15}\text{N}_{\text{excess}}$ data ($\Delta\delta^{15}\text{N}_{\text{excess,hf}}$) into temperature by using the temperature sensitivities Ω_{N} of $\delta^{15}\text{N}$ and Ω_{Ar} of $\delta^{40}\text{Ar}$ as follows:

$$\Delta T_{\text{hybrid,hf}}(t) = \frac{\Delta\delta^{15}\text{N}_{\text{excess,hf}}(t)}{(\Omega_{\text{N}}(t) - \Omega_{\text{Ar}}(t)/4)} \quad (3)$$

The hybrid solution is used to imitate the temperature reconstruction by Kobashi et al. (2017) but with a different strategy. Kobashi et al. (2017) fitted the temperature gradient over the diffusive firn-column ΔT_{firn} (eq. 4)



calculated from $\delta^{15}\text{N}_{\text{excess}}$. Due to the high relative uncertainty of $\delta^{15}\text{N}_{\text{excess}}$ together with the yearly calculation of surface temperature from the modelled bottom temperature values using the ΔT_{firn} -integration-method (Kobashi et al., 2008a, 2010), strong drifts in the reconstructed temperature can occur (sect. 4.1.2). To overcome this issue, Kobashi et al. (2017) forces the $\delta^{15}\text{N}_{\text{excess}}$ temperature to also fit the general trend of $\delta^{15}\text{N}$ by allowing constant shifts in ΔT_{firn} in nine certain time windows of 1500 yr length. This correction reduces the goodness of the $\delta^{15}\text{N}_{\text{excess}}$ -fit in some parts of the time-series and adds additional uncertainty to the reconstructed temperature. Additionally, the correction can change millennial scale trends, because changing the mean ΔT_{firn} in a 1500°yr window directly changes the temperature trend in this section. Also, allowing sharp shifts between the windows can create short term temperature signals (jumps with 50-200 yr durations), which could be misinterpreted as real temperature changes. As our method provides the possibility to know the long-term $T(\delta^{15}\text{N})$ trend, it is interesting to compare the hybrid temperature solution to the Kobashi et al. solution, and to investigate the differences which should mainly occur due to the “window correction” method.

4.1 Temperature solution comparison

4.1.1 Model comparison

The comparison of the fitting results using the Schwander et al. (1997) or the Goujon et al. (2003) firn-model (fig. 4a-d) on an individual fitting target reveals that except for $T(\delta^{15}\text{N}_{\text{excess}})$, the solutions gained by using the two models are highly comparable. They show high correlation ($r > 0.9$) in all considered periodic-time bands (supplement section S6, tabs. S3-S6). The absolute temperatures show an offset of about 2 K between the two models. This offset can be explained by the implemented convective zone in the Goujon et al. (2003) model. As the convective zone lowers the height of the diffusive firn column, a colder temperature (compared to the Schwander et al. (1997) model) is needed, decelerating the densification and leading to the LID needed to fit the gravitational fraction of the gas-isotope data. The temperature differences between the estimates using both models are not fully constant. Besides the offset we find mean differences (2σ) of 0.62 K for $T(\delta^{15}\text{N})$ and 0.73 K for $T(\delta^{40}\text{Ar})$, which are partly driven by remaining mismatches of the isotope fits. In the early-Holocene (9.5 kyr to 11.5 kyr) the $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ reconstructions using the Goujon et al. (2003) firn-model show a faster warming, leading to slightly higher temperature anomalies compared to the Schwander et al. (1997) model estimates. The rest of the time-series shows the reverse behaviour. Here the Schwander et al. (1997) model estimates are slightly warmer than the Goujon et al. (2003) model estimates. The fact that the trend in the time-series of the differences between the Schwander et al. (1997) and Goujon et al. (2003) model temperature solution follows the trend in the reconstructed temperature anomalies is pointing to a temperature sensitive fraction of these differences. An explanation for this could be that the densification itself is slightly temperature depended. The 2 K colder absolute temperature in the Goujon et al. (2003) model inputs leads to a different densification pathway compared to the Schwander et al. (1997) model runs. Nevertheless, both models provide highly comparable temperature estimates if temperature anomalies are considered.

4.1.2 Comparison of $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ reconstruction

The comparison between the $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ based temperatures (fig. 4a,b and fig. S6) shows that the general trends between both reconstructions are very similar. Also, the shapes of many of the shorter-term temperature features are in a good agreement, but $T(\delta^{40}\text{Ar})$ points to higher amplitudes of these anomalies. We find high correlations (supplement sect. S6, tabs. S3-S6) for the low pass filtered data (cut-off: 50 yr, $r = 0.96$) and the multi-



millennial band ($r = 0.94$), which was expected due to the high agreement in the long-term trends between both temperatures. The multi-centennial band shows a lower but still high correlation ($r = 0.87$). In the multi-decadal band, the correlation between $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ is weak ($r = 0.69$, lag = -4 yr or $r = 0.67$ for lag = 0 yr), and equals the correlation between the measured isotopes ($\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$) in the same band with $r = 0.68$. This result is not surprising because: (i) The high-frequency fraction of the reconstructed temperatures is directly calculated from the high-frequency fraction of the respective isotope targets (Döring and Leuenberger, 2018) and (ii) the accumulation-rate input has only a minor effect on the reconstructed temperatures in multi-decadal band. The mean offset between $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ over the whole time-series is 0.28 K and the standard deviation (2σ) of the differences is 1.00 K for the maximum resolution case (mean resolution of the isotope data is 17.8 yr). In the early to mid (6.4 kyr-11.5 kyr) and late-Holocene (0.07 kyr-1.3 kyr b2k), the $\delta^{15}\text{N}$ reconstruction leads to a slightly higher absolute temperature compared to $T(\delta^{40}\text{Ar})$ with mean-offsets of about 0.46 K for the early- to mid-Holocene and 0.61 K for the late-Holocene, whereas the rest of the temperature time-series is showing a similar trend (1.3 kyr-6.4 kyr b2k, mean offset: 0.06 K). The lower temperatures in $T(\delta^{40}\text{Ar})$ in the early- to mid- and late-Holocene can be explained by an enrichment in $\delta^{40}\text{Ar}$ due to gas-loss still remaining after the applied corrections. The LID estimates of both reconstructions are in good agreement (fig. S6b). The differences between them vary in a narrow band of -2 m to 1 m driven by the temperature differences between the reconstructions. The comparison of modelled firn temperature gradients ΔT_{firn} (fig. S6c) of the $\delta^{15}\text{N}$ - and $\delta^{40}\text{Ar}$ -fits shows a high agreement in the general trends and in the shapes of the shorter-term features. The differences between them are less than ± 1 K in most of the case. Additionally, $\Delta T_{\text{firn, meas}}$ calculated directly from the measured isotope data (dotted line, meas) together with its maximum ($1\sigma_{\text{max}}$) and minimum ($1\sigma_{\text{min}}$) uncertainty, is compared to the modelled estimates. $\Delta T_{\text{firn, meas}}$ was calculated analogous to Kobashi et al. (2010, 2011, 2017) from $\delta^{15}\text{N}_{\text{excess}}$ according to:

$$\Delta T_{\text{firn, meas}} = \frac{\delta^{15}\text{N}_{\text{excess}}}{(\Omega_{\text{N}} - \Omega_{\text{Ar}}/4)} = \frac{\delta^{15}\text{N} - \delta^{40}\text{Ar}/4}{(\Omega_{\text{N}} - \Omega_{\text{Ar}}/4)} \quad (4)$$

The uncertainties ($1\sigma_{\text{max}}$, $1\sigma_{\text{min}}$) of $\Delta T_{\text{firn, meas}}$ were calculated using Gaussian-error-propagation on eq. 4 together with the uncertainties of $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ as stated in tab. S1. If we compare the modelled ΔT_{firn} to $\Delta T_{\text{firn, meas}}$ we find a good agreement of the general trend in the late- to mid-Holocene (1.3 kyr-6.4 kyr b2k), which is exactly the part in time where also the trends in $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ are showing the smallest offset. In the early- to mid-Holocene (6.4 kyr-11.5 kyr), ΔT_{firn} modelled from $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ significantly exceeds $\Delta T_{\text{firn, meas}}$, which can be a sign for systematic too high $\delta^{40}\text{Ar}$ in this section, reducing $\delta^{15}\text{N}_{\text{excess}}$ and $\Delta T_{\text{firn, meas}}$. The same is true for the late-Holocene (0.07 kyr-1.3 kyr b2k). Comparing the amplitudes of the faster signals of the measured and modelled ΔT_{firn} shows that the modelled signals underestimates the amplitudes of $\Delta T_{\text{firn, meas}}$, which leads to the assumption that $\Delta T_{\text{firn, meas}}$ and therefore $\delta^{15}\text{N}_{\text{excess}}$ is potentially more driven by noise in the isotope data than temperature.

4.1.2 Comparison of $\delta^{15}\text{N}$ and $\delta^{15}\text{N}_{\text{excess}}$ reconstruction

Precise fitting of $\delta^{15}\text{N}_{\text{excess}}$ and therefore ΔT_{firn} (fig. 4d and fig. S7) results in a distinct temperature regime compared to $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$. In the early-Holocene (9 kyr-11.5 kyr) the fitting of $\delta^{15}\text{N}_{\text{excess}}$ leads to a flat temperature with nearly no trend. This is not only in disagreement with $T(\delta^{15}\text{N})$ or $T(\delta^{40}\text{Ar})$ but also with the reconstructions from Kobashi et al. (2017) and Buizert et al. (2018) for the GISP2 site (fig. 6). The flat temperatures in this section when precisely fitting $\delta^{15}\text{N}_{\text{excess}}$ is the result of too low $\delta^{15}\text{N}_{\text{excess}}$ and therefore ΔT_{firn} , which is driven by too high $\delta^{40}\text{Ar}$ in that section. In the late-Holocene, $T(\delta^{15}\text{N}_{\text{excess}})$ shows a large cooling of about -3.6 K/kyr (0.16 kyr-1.25 kyr), which is highly unrealistic, pointing also to too low $\delta^{15}\text{N}_{\text{excess}}$ and ΔT_{firn} . Also the



$\delta^{15}\text{N}_{\text{excess}}$ -fit of Kobashi et al. (2017) shows large disagreement between their modelled $\delta^{15}\text{N}_{\text{excess}}$ and the measured ones during this sections (see supplementary information fig. S3 in Kobashi et al. (2017)), which means that the quality of the $\delta^{15}\text{N}_{\text{excess}}$ -fit has to be reduced significantly to extract a potentially meaningful temperature estimate. As we have used the corrected $\delta^{40}\text{Ar}$ to calculate $\delta^{15}\text{N}_{\text{excess}}$ and ΔT_{fir} , we have to argue that the presently available correction (Kobashi et al., 2015b, 2017) is non-sufficient, especially for the late- and early-Holocene data. This result is somehow surprising as it was argued in Buizert et al. (2018) that the influence of possible gas-loss on $\delta^{40}\text{Ar}$ is most severe within the bubble-clathrate transition zone (about 800 m to 1500 m depth of the GISP2 core, equals 3.8-9.3 kyr BP ice age, see supplement p. X-2 in Buizert et al. (2018)). In the following we compare two of the major mid-Holocene cooling trends in $T(\delta^{15}\text{N}_{\text{excess}})$ with the simultaneous trends in $T(\delta^{15}\text{N})$. In the time range 2 kyr to 4.8 kyr b2k, where the trends in $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ are highly comparable, $T(\delta^{15}\text{N}_{\text{excess}})$ shows a cooling rate of -0.52 K/kyr. The cooling trend of $T(\delta^{15}\text{N})$ in the same time range is -0.18 K/kyr and therefore about 3 times smaller. In the second time range from 6.3 kyr to 8.1 kyr, where $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ show a significant offset, the cooling in $T(\delta^{15}\text{N}_{\text{excess}})$ exceeds the cooling in $T(\delta^{15}\text{N})$ by a factor of about 2.5 (-1.58 K/kyr for $T(\delta^{15}\text{N}_{\text{excess}})$, -0.62 K/kyr for $T(\delta^{15}\text{N})$). Also, these results could be explained by too low $\delta^{15}\text{N}_{\text{excess}}$ and ΔT_{fir} . Next, we compare the correlations of $T(\delta^{15}\text{N}_{\text{excess}})$ with $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ in the same periodic-time bands as it was done in the previous section (supplement sect. S6, tabs. S3-S6). In all periodic-time bands the correlation of $T(\delta^{15}\text{N}_{\text{excess}})$ with $T(\delta^{15}\text{N})$ as well as $T(\delta^{15}\text{N}_{\text{excess}})$ with $T(\delta^{40}\text{Ar})$ is weak. The highest correlation between $T(\delta^{15}\text{N}_{\text{excess}})$ and $T(\delta^{15}\text{N})$ was found for the multi-millennial time-band (tab. S4) with $r = 0.61$ for the best fit $T(\delta^{15}\text{N}_{\text{excess}})$ solution and $r = 0.68$ for the averaged $T(\delta^{15}\text{N}_{\text{excess}})$ solution. The correlation with $T(\delta^{40}\text{Ar})$ is even weaker ($r = 0.48$, best fit; $r = 0.54$, mean solution). In the multi-decadal periodic-time band (tab. S6) we find a weak negative correlation between $T(\delta^{15}\text{N}_{\text{excess}})$ and $T(\delta^{40}\text{Ar})$ with $r = -0.41$ and $r = -0.36$ for the best fit and the mean $T(\delta^{15}\text{N}_{\text{excess}})$ solution, respectively. The result for the multi-decadal time band can be explained, since the multi-decadal oscillations in $T(\delta^{15}\text{N}_{\text{excess}})$ are mainly driven by $\delta^{40}\text{Ar}$ with less influence from $\delta^{15}\text{N}$ due to higher variability of $\delta^{40}\text{Ar}/4$ compared to $\delta^{15}\text{N}$.

4.2 Final uncertainty of reconstructed temperature

Using all information from the previous sections we can calculate an overall limit of the mean uncertainty for reconstructed temperatures using the following equation:

$$(2\sigma_T)^2 = \left(\frac{2\sigma_{\text{miss}}}{\Omega_X}\right)^2 + (2\sigma_{\text{rep}})^2 + \left(\frac{2\sigma_{\text{meas}}}{\Omega_X}\right)^2 + (2\sigma_{\text{model}})^2 \quad (5)$$

σ_T is the uncertainty of the reconstructed temperature, σ_{miss} is the remaining mean mismatch after fitting the isotope values (sect. 3.1.1), σ_{rep} is the fraction of uncertainty due to the reproducibility of the fitting method (sect. 3.1.1), σ_{meas} is the analytical uncertainty of the measured data (tab. S1) and σ_{model} is the standard deviation of the differences of the temperature anomalies between the used models (sect. 4.1.1). Ω_X is the thermal diffusion sensitivity of the respective isotope species. For $T(\delta^{15}\text{N})$ the calculated uncertainty is $2\sigma_T = 0.80 \dots 0.88$ K, the range is due to the minimal or maximal analytical uncertainty of the measured data. The uncertainty of the $\delta^{40}\text{Ar}$ reconstruction is $2\sigma_T = 0.87 \dots 1.81$ K. This final uncertainty is attributed to each single temperature point in time (the mean data resolution was 17.8 yr), so a smoothing or running mean calculation will decrease the uncertainty due to the averaging over a certain amount of points. For example, a smoothing with 100 yr cut-off will reduce the uncertainty with a factor of $1/n^{1/2} = 1/(100 \text{ yr}/17.8 \text{ yr})^{1/2} = 0.42$ as it equals an averaging over 5.6 points. This is important to keep in mind when discussing filtered versions of these reconstructed temperatures. For $T(\delta^{15}\text{N}_{\text{excess}})$ we are not able to provide a final uncertainty, as we do not understand the reason for the different behaviour when



fitting the data with the two models (stable Goujon- vs. unstable Schwander-solutions). However, as fitting of $\delta^{15}\text{N}_{\text{excess}}$ leads to highly distinct temperature estimates compared to $T(\delta^{15}\text{N})$ or $T(\delta^{40}\text{Ar})$ and also when compared to other reconstructions (comparison of fig. 4d with fig. 6a), we do not recommend to use $T(\delta^{15}\text{N}_{\text{excess}})$ for any climatic interpretation yet. As $\delta^{15}\text{N}$ is easier to measure due to the higher abundance of nitrogen in air and less susceptible to gas-loss induced fractionation (Huber et al., 2006a), we argue that $T(\delta^{15}\text{N})$ provides the most robust temperature estimate compared to $T(\delta^{40}\text{Ar})$ and especially $T(\delta^{15}\text{N}_{\text{excess}})$.

4.3 Comparison of $T(\delta^{15}\text{N})$ with the reconstructions of Kobashi et al. 2017 and Buizert et al. 2018

Figure 6 shows the comparison of $T(\delta^{15}\text{N})$ (blue lines) with the reconstructions of Buizert et al. (2018) (red lines) and Kobashi et al. (2017) (black lines) and the comparison between the hybrid-temperature (green lines, fig. 6b) with the Kobashi et al. (2017) solution. As stated before, the temperature reconstructions of Kobashi et al. (2017) were conducted using $\delta^{15}\text{N}_{\text{excess}}$ to obtain temperature, which is in our opinion problematic due to the high relative uncertainty of $\delta^{15}\text{N}_{\text{excess}}$ and the systematic offsets to too high $\delta^{40}\text{Ar}$. Buizert et al. (2018) use the calibration of water-stable-isotope $\delta^{18}\text{O}_{\text{ice}}$ to fit the long-term trend of $\delta^{15}\text{N}$ for the early- to mid-Holocene (until 4 kyr b2k). From 4 kyr to today they use the temperature from Kobashi et al. (2017) superimposed with a larger cooling trend (fig. S8). First, we compare the variance of the temperatures (tab. 2) for two time-windows (0.5–4.0 kyr b2k and 4.0–11.2 kyr b2k) and two periodic-time bands (bands: 100 yr to 500 yr and 500 yr to 4 kyr). As the temporal resolution of the Buizert et al. (2018) estimate is 20 yr, we resampled our data and the Kobashi et al. (2017) data to the same grid before band-pass filtering. Also, we cut out the 8.2k-event, as it dominates the variance of the temperature data. In the early- to mid-Holocene (4.0–11.2 kyr b2k, w/o 8.2k-event), the standard deviation (2σ) of $T(\delta^{15}\text{N})$ in the 100–500 yr periodic-time band is 0.47 K which is nearly equal to the Buizert et al. (2018) reconstruction with a value of 0.49 K. The Kobashi et al. (2017) reconstruction has a more than double as large variance in that section with $2\sigma = 1.17$ K. For the mid- to late-Holocene (0.5–4.0 kyr b2k), the standard deviation of $T(\delta^{15}\text{N})$ is about 20% smaller ($2\sigma = 0.37$ K) than for the early- to mid-Holocene (4.0–11.2 kyr). The variance of the Kobashi et al. (2017) reconstruction is slightly smaller (17%, $2\sigma = 0.97$ K) during this time. The Buizert et al. (2018) reconstruction nearly equals the Kobashi et al. (2017) estimate here, with $2\sigma = 1.00$ K. This is not unexpected as Buizert et al. (2018) use the Kobashi et al. (2017) data with slight modifications. But it is not quite reasonable that a doubling of the variance between the two parts of the Holocene is realistic. For longer periodicities (band: 0.5–4.0 kyr) we also see a reduction of the variance for the late-Holocene compared to the early-Holocene in all three reconstructions. During the early- to mid-Holocene the variance of $T(\delta^{15}\text{N})$ and the Buizert et al. (2018) reconstruction are in good agreement ($2\sigma_{T(\delta^{15}\text{N})} = 0.58$ K, $2\sigma_{\text{Buizert}} = 0.65$ K), whereas the Kobashi et al. (2017) estimate is pointing to higher variability ($2\sigma_{\text{Kobashi}} = 1.26$ K), an equal behaviour as it was found for the faster periodicities. For the mid- to late-Holocene, $T(\delta^{15}\text{N})$ shows the smallest variability ($2\sigma = 0.35$ K) compared to the Buizert et al. (2018) ($2\sigma = 0.58$ K) and Kobashi et al. (2017) ($2\sigma = 0.81$ K) reconstructions. We conducted the same analysis for the accumulation-rate data and for the gas-isotope data (tab. 3). Especially for the 0.5–4.0 kyr periodic-time band, we find an equal reduction of the data variance between the early- to mid- and the mid- to late-Holocene. For $\delta^{15}\text{N}$ the standard deviation (2σ) in the early- to mid-Holocene is about 52% higher as in the mid- to late-Holocene, 34% for $\delta^{40}\text{Ar}$ and 41% for $\delta^{15}\text{N}_{\text{excess}}$, respectively. The accumulation-rate data show this deviation of the variance between the two time-sections in this periodic-time band with a reduction of 46%. For the faster periodicities (band-pass 100–500 yr), the gas-isotope data also shows that behaviour, but with less disagreement between the two time-sections. In contrast, the accumulation-rate data



show a slightly higher variance in the mid- to late-Holocene if compared to the early- to mid-Holocene. Based on those findings we conclude that the difference in the variability of our temperature estimates between the early- to mid- and the mid- to late-Holocene is a direct result of that behaviour of the gas-isotope target data and the accumulation-rate input. Interestingly, the difference between the Buizert et al. (2018) and Kobashi et al. (2017) estimates (fig. S8) for the time section 0.5-4.0 kyr b2k, points to the necessity of modifications on the Kobashi et al. (2017) estimate when used in the study of Buizert et al. (2018). A reason for that could be the use of a different firn-model or possible memory effects occurring due to the differences in the temperature estimates between the Buizert et al. (2018) and Kobashi et al. (2017) estimates in the early- to mid-Holocene. Figure 6b shows the comparison between the Kobashi et al. (2017) estimate and our hybrid temperature. It is obvious that both estimates agree well, especially for the faster features. In four time-sections we find larger offsets. These sections are: 9-10.5 kyr, shortly after the 8.2k-event (6.6-8.1 kyr), 5.3-6.1 kyr and 0.07-1.8 kyr. All sections start with a fast warming or cooling trend in the Kobashi et al. (2017) estimate with duration of about 100-200 yr. It is highly probable that these shifts are introduced by the “window correction method” used by Kobashi et al. (2017). In our view, this correction method is highly critical as the choice of the window-length, the window positions and the found offsets are arbitrary, but on the other side crucial for the reconstruction. Additionally, we correlated the Buizert et al. (2018) and Kobashi et al. (2017) temperature estimates with each other and with our data after low-pass filtering (cop = 50 yr) and in all three investigated periodic-time bands (multi-decadal, multi-centennial, multi-millennial, tab. S7). For the Buizert et al. (2018) reconstruction the correlations were calculated only for the early- to mid-Holocene values (4.0-11.5 kyr b2k), whereas the correlations between our and the Kobashi et al. (2017) estimates were calculated for 0.5-11.5 kyr b2k. In the low-pass filtered case (general trend) the Buizert et al. (2018) reconstruction shows the highest correlations with $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ with $r > 0.9$ and a correlation of $r = 0.82$ with the Kobashi et al. (2017) estimate, due to the high agreement in the general trend between the three studies. The comparison between the Kobashi et al. (2017) temperature and our estimates has the highest correlation with the hybrid temperatures ($r = 0.87$, $r = 0.83$), and $T(\delta^{15}\text{N})$ ($r = 0.81$) in that case. Besides the general trend, the correlations become reduced for faster oscillations. In all periodic-time bands the Buizert et al. (2018) reconstruction has the highest correlation with $T(\delta^{15}\text{N})$ with $r_m = 0.67$, $r_c = 0.61$ and $r_d = 0.30$ for multi-millennial, multi-centennial and multi-decadal signals, respectively. The correlation coefficients between the Buizert et al. (2018) and the Kobashi et al. (2017) reconstruction are $r_m = 0.52$ ($p = 0.03$), $r_c = 0.48$ and $r_d = 0.05$ ($p = 0.25$). That finding implies that the agreement between the Buizert et al. (2018) $\delta^{18}\text{O}_{\text{ice}}$ -based and our $\delta^{15}\text{N}$ -based reconstruction is generally higher as for the Buizert et al. (2018) reconstruction compared to the $\delta^{15}\text{N}_{\text{excess}}$ -based one of Kobashi et al. (2017). On the other hand, the correlations are relatively weak, especially for the multi-decadal band.



time section [kyr b2k]	Band-pass: 100 yr-500 yr			Band-pass: 0.5 kyr-4.0 kyr		
	2σ [K] T($\delta^{15}\text{N}$)	2σ [K] Buizert 2018	2σ [K] Kobashi 2017	2σ [K] T($\delta^{15}\text{N}$)	2σ [K] Buizert 2018	2σ [K] Kobashi 2017
0.5-11.2 w/o 8.2k-event	0.44	0.72	1.10	0.51	0.63	1.12
0.5-4.0	0.37	1.00	0.97	0.35	0.58	0.81
4.0-11.2 w/o 8.2k-event	0.47	0.49	1.17	0.58	0.65	1.26

Table 2: Standard deviations (2σ) of temperature estimates of T($\delta^{15}\text{N}$) (this study), Buizert et al. (2018) and Kobashi et al. (2017) when band-pass filtered for two periodic-time bands and calculated without the 8.2k-event.

time section [kyr b2k]	Band-pass: 100 yr-500 yr				Band-pass: 0.5 kyr-4.0 kyr			
	2σ [permeg] $\delta^{15}\text{N}$	2σ [permeg] $\delta^{40}\text{Ar}$	2σ [permeg] $\delta^{15}\text{N}_{\text{excess}}$	2σ [mm/yr] acc	2σ [permeg] $\delta^{15}\text{N}$	2σ [permeg] $\delta^{40}\text{Ar}$	2σ [permeg] $\delta^{15}\text{N}_{\text{excess}}$	2σ [mm/yr] acc
0.5-11.2 w/o 8.2k- event	7.2	6.9	5.2	11.7	6.6	5.5	3.0	8.2
0.5-4.0	6.6	5.9	3.9	13.0	3.7	4.0	2.0	5.1
4.0-11.2 w/o 8.2k- event	7.5	7.3	5.7	11.0	7.7	6.1	3.4	9.4

Table 3: Standard deviations (2σ) of gas-isotope measured data and accumulation-rates when band-pass filtered for two periodic-time bands and calculated without the 8.2k-event.

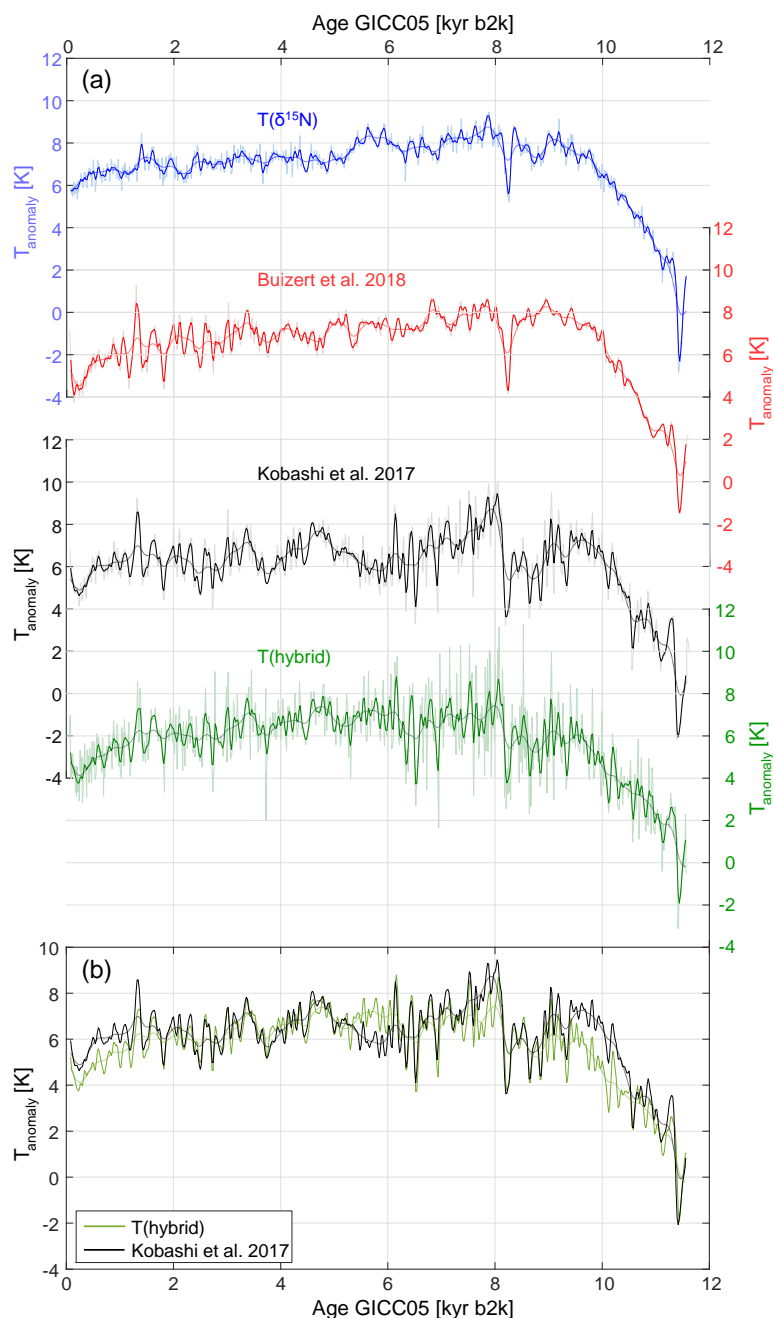


Figure 6: (a) Comparison of $T(\delta^{15}N)$ (blue lines) and $T(\text{hybrid})$ (green lines) modelled using the Schwander et al. model with the temperature reconstructions (anomalies rel. to 11.500 kyr b2k) for the GISP2 site from Buizert et al. 2018 (red lines) and Kobashi et al. 2017 (black lines). Thin lines show unfiltered data, thick lines show low-pass filtered data using cop = 100 yr, dotted lines show low-pass filtered data using cop = 500 yr. (b) Comparison between the reconstruction from Kobashi et al. 2017 (black lines) and $T(\text{hybrid})$ (lines) in the same plot.



5 Conclusion

In this study we applied the Döring and Leuenberger (2018) gas-isotope fitting-algorithm to Holocene $\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$ data measured on the GISP2 ice core (Kobashi et al., 2008b) using two state of the art firn-densification and heat-diffusion models (Goujon et al., 2003; Schwander et al., 1997). The results of this study are summarized as follows:

Signal-to-noise study:

As starting point, a signal-to-noise (SNR) study was conducted, investigating the suitability of the three gas-isotope tracers ($\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$) for temperature reconstructions in context of the in Kobashi et al. (2008b) stated measurement uncertainty. It was shown that $\delta^{15}\text{N}$ is most favoured to reconstruct Holocene temperature due to its higher SNR compared to $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$, especially for multi-decadal to multi-centennial signals.

Gas-isotope fitting results and uncertainties:

To evaluate the performance of the Döring and Leuenberger (2018) gas-isotope fitting-algorithm on measured Holocene data and to constrain the uncertainty of the reconstructed temperatures, the reproducibility of the fitting algorithm was tested using the Schwander et al. (1997) firn-model. The results are showing excellent performance for $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ -fits. The reproducibility between ten runs of the fitting algorithm was quantified to be (mean \pm 2σ) 1.68 ± 1.14 permeg for $\delta^{15}\text{N}$, 2.58 ± 1.77 permeg for $\delta^{40}\text{Ar}/4$, and 1.28 ± 2.14 permeg for $\delta^{15}\text{N}_{\text{excess}}$ for each single point in time. The translation in temperature also shows good reproducibility with (0.17 ± 0.12 K) for $T(\delta^{15}\text{N})$ and (0.26 ± 0.18 K) for $T(\delta^{40}\text{Ar})$. In contrast, the spread for $\delta^{15}\text{N}_{\text{excess}}$ temperature was about ten times higher (2.04 ± 1.90 K) due to the loss of information about the firn-column-height when calculating $\delta^{15}\text{N}_{\text{excess}}$, which leads to a variety of different possible temperature solutions fitting $\delta^{15}\text{N}_{\text{excess}}$ with the same precision. In addition, the contribution of remaining misfits to the uncertainty budget of the reconstructed temperatures was quantified. We reach final misfits (2σ) of 3.7 permeg for $\delta^{15}\text{N}$, 2.8 permeg for $\delta^{40}\text{Ar}/4$, and 2.1 permeg for $\delta^{15}\text{N}_{\text{excess}}$ using the Schwander et al. (1997) model, and 6.0 permeg, 7.0 permeg and 2.0 permeg for $\delta^{15}\text{N}$, $\delta^{40}\text{Ar}/4$ and $\delta^{15}\text{N}_{\text{excess}}$ respectively using the Goujon et al. (2003) model. Additionally, the influence of three different accumulation-rate estimates on the temperature reconstructions was investigated. The accumulation-rate uncertainty leads to asynchronous multi-decadal temperature signals due to deviations in the modelled Δage regimes, with maximum differences of 30 yr in the early-Holocene but keeping the amplitudes of the signals unchanged. From the mid-Holocene until today, the modelled Δage difference becomes less than 5 yr. For longer term temperature trends, the higher accumulation-rate scenario leads to a 0.3 K larger cooling compared to the lower and intermediate accumulation-rate scenarios.

Comparison of model results:

Next, we compared the temperature estimates calculated by fitting $\delta^{15}\text{N}$, $\delta^{40}\text{Ar}$ and $\delta^{15}\text{N}_{\text{excess}}$ with each other. We found significant and high correlation between $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ for multi-centennial and multi-millennial signals, but $T(\delta^{40}\text{Ar})$ points to higher amplitudes for some of these temperature anomalies. For multi-decadal signals the correlation between $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ is weak but still significant and equals the correlation between the isotope data ($\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$). The comparison of the temperature gradient over the diffusive firn column ΔT_{firn} modelled from $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ with the $\Delta T_{\text{firn, meas}}$ – calculated from the measured $\delta^{15}\text{N}_{\text{excess}}$ – documents a good agreement of the general trend in the late- to mid-Holocene (1.3 kyr-6.4 kyr b2k). However, in the early to mid-Holocene (6.4 kyr-11.5 kyr) and the late-Holocene (0.07 kyr-1.3 kyr b2k), ΔT_{firn} modelled from $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ significantly exceeds $\Delta T_{\text{firn, meas}}$, which might be an indication for systematic too high $\delta^{40}\text{Ar}$ in this section, potentially caused by gas-loss induced fractionation of $\delta^{40}\text{Ar}$. The temperature calculated by fitting $\delta^{15}\text{N}_{\text{excess}}$ differs



significantly from the coherent $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$, especially for the early- and late-Holocene. The analysis of these differences also suggests that the $\delta^{40}\text{Ar}$ values are too high in these sections. The correlations between $T(\delta^{15}\text{N}_{\text{excess}})$ with $T(\delta^{15}\text{N})$ or $T(\delta^{40}\text{Ar})$ are weak for all three analysed periodic-time bands (multi-decadal, multi-centennial, multi-millennial). For multi-decadal signals we find a weak negative correlation between $T(\delta^{15}\text{N}_{\text{excess}})$ and $T(\delta^{40}\text{Ar})$ and no correlation between $T(\delta^{15}\text{N}_{\text{excess}})$ and $T(\delta^{15}\text{N})$, which implies that the multi-decadal oscillations in $T(\delta^{15}\text{N}_{\text{excess}})$ are mainly driven by $\delta^{40}\text{Ar}$ with less influence from $\delta^{15}\text{N}$ due to a higher noise contribution on $\delta^{40}\text{Ar}/4$ compared to $\delta^{15}\text{N}$. In addition, we calculated the slope between $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$ using geometric-mean-regression for all three periodic-time bands. Especially for multi-decadal signals the slope significantly underestimates the theoretical value (calculated using the empirical derived thermal sensitivities of $\delta^{15}\text{N}$ and $\delta^{40}\text{Ar}$) by 53%. These results are pointing to too high $\delta^{40}\text{Ar}$ that may be influenced by noise or fractionation which is not captured by the firn-models (e.g. gas loss). For the multi-centennial and the multi-millennial bands, the slope equals the theoretical expectation.

Comparison of model results of two different firn-models:

Next, we compared reconstructed temperatures obtained by using the Schwander et al. (1997) firn-model with the solutions of the Goujon et al. (2003) model. For $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ the temperature estimates show high correlation ($r > 0.9$) in all considered periodic-time bands (multi-decadal, multi-centennial, multi-millennial) and in amplitudes and shapes of many of the shorter-term features. In the early-Holocene (9.5 kyr to 11.5 kyr) the $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$ reconstructions using the Goujon et al. (2003) firn-model show a faster warming, leading to slightly warmer temperature anomalies compared to the Schwander et al. (1997) model estimates. The rest of the time-series shows the reverse behaviour. Here the Schwander et al. (1997) model estimates are slightly warmer than the Goujon et al. (2003) model estimates. The result that the difference in the temperature anomalies between both models follows the general trend of the temperature anomalies themselves, is pointing to a temperature dependence which is attributed to the temperature dependence of the densification due to the difference in absolute temperatures of about 2 K. The variance of the differences (2σ) between the temperature anomalies obtained by using both models were quantified to be 0.62 K for $T(\delta^{15}\text{N})$ and 0.73 K for $T(\delta^{40}\text{Ar})$.

Uncertainty estimation:

Using all results presented in this study we estimated the mean uncertainty for $T(\delta^{15}\text{N})$ and $T(\delta^{40}\text{Ar})$. The final uncertainty budget of the reconstructed temperature anomalies is dependent on four terms: (i) the mismatch between the measured and modelled isotope data; (ii) the reproducibility of the isotope fits; (iii) the measurement uncertainty of the isotope data; and (iv) the difference between the temperature estimates using different firn-models (here between two firn-models). Adding up these terms leads to a final uncertainty of $2\sigma_T = 0.80 \dots 0.88$ K for $T(\delta^{15}\text{N})$, and $2\sigma_T = 0.87 \dots 1.81$ K for $T(\delta^{40}\text{Ar})$.

Comparison to other published temperature reconstructions:

Finally, we compared our temperature estimates to temperature reconstructions for the GISP2 site of Buizert et al. (2018) and Kobashi et al. (2017). First, the variance of the temperature anomalies was analysed in two periodic-time bands (100-500 yr and 0.5-4 kyr). We found a high agreement of the variance between $T(\delta^{15}\text{N})$ and the Buizert et al. (2018) estimate in both considered bands for the time 4.0-11.5 kyr b2k. In contrast, the variance of the Kobashi et al. (2017) temperature anomalies is nearly twice as high as our $T(\delta^{15}\text{N})$ or the Buizert et al. (2018) estimate. Interestingly, all three reconstructions are pointing to a decrease of the variance in the mid- to late-Holocene (0.5-4.0 kyr b2k) compared to the early- to mid-Holocene (4.0-11.5 kyr b2k) for multi-centennial to multi-millennial signals. For our reconstruction this result is attributed to an equal behaviour of the gas-isotope



and the accumulation-rate data. Finally, we compared the correlations between the three reconstructions for three periodic-time bands (multi-decadal, multi-centennial, multi-millennial). We find generally higher agreement in all bands between our $\delta^{15}\text{N}$ -based reconstruction and the $\delta^{18}\text{O}$ - and $\delta^{15}\text{N}$ -based reconstruction of Buizert et al. (2018) as between our $T(\delta^{15}\text{N})$ estimate and the Kobashi et al. (2017) reconstruction.

5 Competing interests

The authors declare that they have no competing financial interests.

Acknowledgement

We would like to thank Gunnar Jansen for his IT support. This study was financed by the Swiss National Science Foundation through the icoCEP project (172550) as well as the project Climate and Environmental Physics (159563).

References

- Baggenstos, D.: Taylor Glacier as an archive of ancient ice for large-volume samples: Chronology, gases, dust, and climate, UNIVERSITY OF CALIFORNIA, SAN DIEGO. [online] Available from: <https://escholarship.org/uc/item/9wn8789k> (Accessed 5 August 2020), 2015.
- 15 Buizert, C., Gkinis, V., Severinghaus, J. P., He, F., Lecavalier, B. S., Kindler, P., Leuenberger, M., Carlson, A. E., Vinther, B., Masson-Delmotte, V., White, J. W. C., Liu, Z., Otto-Bliesner, B. and Brook, E. J.: Greenland temperature response to climate forcing during the last deglaciation., *Science*, 345(6201), 1177–80, doi:10.1126/science.1254961, 2014.
- Buizert, C., Keisling, B. A., Box, J. E., He, F., Carlson, A. E., Sinclair, G. and DeConto, R. M.: Greenland-Wide Seasonal Temperatures During the Last Deglaciation, *Geophys. Res. Lett.*, 45(4), 1905–1914, doi:10.1002/2017GL075601, 2018.
- 20 Cuffey, K. M. and Clow, G. D.: Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition, *J. Geophys. Res. Ocean.*, 102(C12), 26383–26396, doi:10.1029/96JC03981, 1997.
- Döring, M. and Leuenberger, M. C.: Novel automated inversion algorithm for temperature reconstruction using gas isotopes from ice cores, *Clim. Past*, 14(6), 763–788, doi:10.5194/cp-14-763-2018, 2018.
- 25 Gierz, P., Werner, M. and Lohmann, G.: Simulating climate and stable water isotopes during the Last Interglacial using a coupled climate-isotope model, *J. Adv. Model. Earth Syst.*, 9(5), 2027–2045, doi:10.1002/2017MS001056, 2017.
- Goujon, C., Barnola, J.-M. and Ritz, C.: Modeling the densification of polar firn including heat diffusion: Application to close-off characteristics and gas isotopic fractionation for Antarctica and Greenland sites, *J. Geophys. Res. Atmos.*, 108(D24), n/a–n/a, doi:10.1029/2002JD003319, 2003.
- 30 Grachev, A. M. and Severinghaus, J. P.: Determining the thermal diffusion factor for $^{40}\text{Ar}/^{36}\text{Ar}$ in air to aid paleoreconstruction of abrupt climate change, *J. Phys. Chem. A*, 107(23), 4636–4642, doi:10.1021/jp027817u, 2003.
- 35 Grootes, P. M. and Stuiver, M.: Oxygen $^{18}/^{16}$ variability in Greenland snow and ice with 10-3- to 105-year time resolution, *J. Geophys. Res. Ocean.*, 102(C12), 26455–26470, doi:10.1029/97JC00880, 1997.



- Grootes, P. M. and Stuiver, M.: GISP2 Oxygen Isotope Data, [online] Available from: <https://doi.org/10.1594/PANGAEA.56094>, 1999.
- Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. and Jouzel, J.: Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores, *Nature*, 366(6455), 552–554, doi:10.1038/366552a0, 1993.
- 5 Guillevic, M., Bazin, L., Landais, A., Kindler, P., Orsi, A., Masson-Delmotte, V., Blunier, T., Buchardt, S. L., Capron, E., Leuenberger, M., Martinerie, P., Prié, F. and Vinther, B. M.: Spatial gradients of temperature, accumulation and $\delta^{18}\text{O}$ - ice in Greenland over a series of Dansgaard-Oeschger events, *Clim. Past*, 9(3), 1029–1051, doi:10.5194/cp-9-1029-2013, 2013.
- Huber, C., Beyerle, U., Leuenberger, M., Schwander, J., Kipfer, R., Spahni, R., Severinghaus, J. P. and Weiler, K.: Evidence for molecular size dependent gas fractionation in firn air derived from noble gases, oxygen, and nitrogen measurements, *Earth Planet. Sci. Lett.*, 243(1–2), 61–73, doi:10.1016/j.epsl.2005.12.036, 2006a.
- 10 Huber, C., Leuenberger, M., Spahni, R., Flückiger, J., Schwander, J., Stocker, T. F., Johnsen, S., Landais, A. and Jouzel, J.: Isotope calibrated Greenland temperature record over Marine Isotope Stage 3 and its relation to CH_4 , *Earth Planet. Sci. Lett.*, 243(3–4), 504–519, doi:10.1016/j.epsl.2006.01.002, 2006b.
- 15 Johnsen, S. J., Dahl-Jensen, D., Gundestrup, N., Steffensen, J. P., Clausen, H. B., Miller, H., Masson-Delmotte, V., Sveinbjörnsdottir, A. E. and White, J.: Oxygen isotope and palaeotemperature records from six Greenland ice-core stations: Camp Century, Dye-3, GRIP, GISP2, Renland and NorthGRIP, *J. Quat. Sci.*, 16(4), 299–307, doi:10.1002/jqs.622, 2001.
- Kindler, P., Guillevic, M., Baumgartner, M., Schwander, J., Landais, A. and Leuenberger, M.: Temperature reconstruction from 10 to 120 kyr b2k from the NGRIP ice core, *Clim. Past*, 10(2), 887–902, doi:10.5194/cp-10-887-2014, 2014.
- 20 Kobashi, T., Severinghaus, J. P. and Barnola, J. M.: $4 \pm 1.5^\circ\text{C}$ abrupt warming 11,270 yr ago identified from trapped air in Greenland ice, *Earth Planet. Sci. Lett.*, 268(3–4), 397–407, doi:10.1016/j.epsl.2008.01.032, 2008a.
- Kobashi, T., Severinghaus, J. P. and Kawamura, K.: Argon and nitrogen isotopes of trapped air in the GISP2 ice core during the Holocene epoch (0–11,500 B.P.): Methodology and implications for gas loss processes, *Geochim. Cosmochim. Acta*, 72(19), 4675–4686, doi:10.1016/j.gca.2008.07.006, 2008b.
- 25 Kobashi, T., Severinghaus, J. P., Barnola, J. M., Kawamura, K., Carter, T. and Nakaegawa, T.: Persistent multi-decadal Greenland temperature fluctuation through the last millennium, *Clim. Change*, 100(3), 733–756, doi:10.1007/s10584-009-9689-9, 2010.
- 30 Kobashi, T., Kawamura, K., Severinghaus, J. P., Barnola, J. M., Nakaegawa, T., Vinther, B. M., Johnsen, S. J. and Box, J. E.: High variability of Greenland surface temperature over the past 4000 years estimated from trapped air in an ice core, *Geophys. Res. Lett.*, 38(21), n/a–n/a, doi:10.1029/2011GL049444, 2011.
- Kobashi, T., Box, J. E., Vinther, B. M., Goto-Azuma, K., Blunier, T., White, J. W. C., Nakaegawa, T. and Andresen, C. S.: Modern solar maximum forced late twentieth century Greenland cooling, *Geophys. Res. Lett.*, 35(14), 5992–5999, doi:10.1002/2015GL064764, 2015a.
- 35 Kobashi, T., Ikeda-Fukazawa, T., Suwa, M., Schwander, J., Kameda, T., Lundin, J., Hori, A., Motoyama, H., Döring, M. and Leuenberger, M.: Post-bubble close-off fractionation of gases in polar firn and ice cores: Effects of accumulation rate on permeation through overloading pressure, *Atmos. Chem. Phys.*, 15(24), 13895–13914, doi:10.5194/acp-15-13895-2015, 2015b.



- Kobashi, T., Menviel, L., Jeltsch-Thömmes, A., Vinther, B. M., Box, J. E., Muscheler, R., Nakaegawa, T., Pfister, P. L., Döring, M., Leuenberger, M., Wanner, H. and Ohmura, A.: Volcanic influence on centennial to millennial Holocene Greenland temperature change, *Sci. Rep.*, 7(1), 1441, doi:10.1038/s41598-017-01451-7, 2017.
- Landais, A., Barnola, J. M., Kawamura, K., Caillon, N., Delmotte, M., Van Ommen, T., Dreyfus, G., Jouzel, J.,
 5 Masson-Delmotte, V., Minster, B., Freitag, J., Leuenberger, M., Schwander, J., Huber, C., Etheridge, D. and Morgan, V.: Firn-air $\delta^{15}\text{N}$ in modern polar sites and glacial-interglacial ice: A model-data mismatch during glacial periods in Antarctica?, *Quat. Sci. Rev.*, 25(1–2), 49–62, doi:10.1016/j.quascirev.2005.06.007, 2006.
- Leng, L., Zhang, T., Kleinman, L. and Zhu, W.: Ordinary least square regression, orthogonal regression, geometric mean regression and their applications in aerosol science, *J. Phys. Conf. Ser.*, 78(1), 012084, doi:10.1088/1742-
 10 6596/78/1/012084, 2007.
- Leuenberger, M. C., Lang, C. and Schwander, J.: Delta ^{15}N measurements as a calibration tool for the paleothermometer and gas-ice age differences: A case study for the 8200 B.P. event on GRIP ice, *J. Geophys. Res.*, 104(D18), 22163–22170, doi:10.1029/1999jd900436, 1999.
- Mariotti, A.: Atmospheric nitrogen is a reliable standard for natural ^{15}N abundance measurements, *Nature*,
 15 303(5919), 685–687, doi:10.1038/303685a0, 1983.
- Meese, D. A., Alley, R. B., Flacco, R. J., Germani, M. S., Gow, A. J., Grootes, P. M., Illing, M., Mayewski, P. A., Morrison, M. C., Ram, M., Taylor, K., Yang, Q. and Zielinski, G. A.: Preliminary depth-age scale of the GISP2 ice core. [online] Available from:
 https://scholar.google.ch/scholar?q=Meese%2C+D.A.%2C+R.B.+Alley%2C+R.J.+Fiacco%2C+M.S.+German
 20 i%2C+A.J.+Gow%2C+P.M.+Grootes%2C+M.+Illing%2C+P.A.+Mayewski%2C+M.C.+Morrison%2C+M.+R
 am%2C+K.C.+Taylor%2C+Q.+Yang%2C+and+G.A.+Zielinski.+1994.+Preliminary+depth-ag (Accessed 21 March 2017), 1994.
- Orsi, A. J., Cornuelle, B. D. and Severinghaus, J. P.: Magnitude and temporal evolution of Dansgaard-Oeschger event 8 abrupt temperature change inferred from nitrogen and argon isotopes in GISP2 ice using a new least-squares inversion, *Earth Planet. Sci. Lett.*, 395, 81–90, doi:10.1016/j.epsl.2014.03.030, 2014.
 25
- Rasmussen, S. O., Seierstad, I. K., Andersen, K. K., Bigler, M., Dahl-Jensen, D. and Johnsen, S. J.: Synchronization of the NGRIP, GRIP, and GISP2 ice cores across MIS 2 and palaeoclimatic implications, *Quat. Sci. Rev.*, 27(1–2), 18–28, doi:10.1016/j.quascirev.2007.01.016, 2008.
- Rasmussen, S. O., Bigler, M., Blockley, S. P., Blunier, T., Buchardt, S. L., Clausen, H. B., Cvijanovic, I., Dahl-Jensen, D., Johnsen, S. J., Fischer, H., Gkinis, V., Guillevic, M., Hoek, W. Z., Lowe, J. J., Pedro, J. B., Popp, T., Seierstad, I. K., Steffensen, J. P., Svensson, A. M., Vallelonga, P., Vinther, B. M., Walker, M. J. C., Wheatley, J. J. and Winstrup, M.: A stratigraphic framework for abrupt climatic changes during the Last Glacial period based on three synchronized Greenland ice-core records: Refining and extending the INTIMATE event stratigraphy, *Quat. Sci. Rev.*, 106, 14–28, doi:10.1016/j.quascirev.2014.09.007, 2014.
 30
- Schwander, J., Sowers, T., Barnola, J. M., Blunier, T., Fuchs, A. and Malaizé, B.: Age scale of the air in the summit ice: Implication for glacial-interglacial temperature change, *J. Geophys. Res.*, 102(D16), 19483–19493, doi:10.1029/97JD01309, 1997.
 35



- Seierstad, I. K., Abbott, P. M., Bigler, M., Blunier, T., Bourne, A. J., Brook, E., Buchardt, S. L., Buizert, C., Clausen, H. B., Cook, E., Dahl-Jensen, D., Davies, S. M., Guillevic, M., Johnsen, S. J., Pedersen, D. S., Popp, T. J., Rasmussen, S. O., Severinghaus, J. P., Svensson, A. and Vinther, B. M.: Consistently dated records from the Greenland GRIP, GISP2 and NGRIP ice cores for the past 104ka reveal regional millennial-scale $\delta^{18}\text{O}$ gradients with possible Heinrich event imprint, *Quat. Sci. Rev.*, 106, 29–46, doi:10.1016/j.quascirev.2014.10.032, 2014.
- Severinghaus, J. P. and Battle, M. O.: Fractionation of gases in polar ice during bubble close-off: New constraints from firn air Ne, Kr and Xe observations, *Earth Planet. Sci. Lett.*, 244(1–2), 474–500, doi:10.1016/j.epsl.2006.01.032, 2006.
- Severinghaus, J. P., Sowers, T. and Alley, R. B.: Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice, *Nature*, 39(January), 141–146, doi:10.1002/jqs.622, 1998.
- Severinghaus, J. P., Grachev, A. and Battle, M.: Thermal fractionation of air in polar firn by seasonal temperature gradients, *Geochemistry, Geophys. Geosystems*, 2(7), n/a–n/a, doi:10.1029/2000GC000146, 2001.
- Severinghaus, J. P., Grachev, A., Luz, B. and Caillon, N.: A method for precise measurement of argon $40/36$ and krypton/argon ratios in trapped air in polar ice with applications to past firn thickness and abrupt climate change in Greenland and at Siple Dome, Antarctica, *Geochim. Cosmochim. Acta*, 67(3), 325–343, doi:https://doi.org/10.1016/S0016-7037(02)00965-1, 2003.
- Steen-Larsen, H. C., Masson-Delmotte, V., Sjolte, J., Johnsen, S. J., Vinther, B. M., Bréon, F. M., Clausen, H. B., Dahl-Jensen, D., Falourd, S., Fettweis, X., Gallée, H., Jouzel, J., Kageyama, M., Lerche, H., Minster, B., Picard, G., Punge, H. J., Risi, C., Salas, D., Schwander, J., Steffen, K., Sveinbjörnsdóttir, A. E., Svensson, A. and White, J.: Understanding the climatic signal in the water stable isotope records from the NEEM shallow firn/ice cores in northwest Greenland, *J. Geophys. Res. Atmos.*, 116(6), D06108, doi:10.1029/2010JD014311, 2011.
- Steig, E. J., Grootes, P. M. and Stuiver, M.: Seasonal precipitation timing and ice core records, *Science* (80-.), 266(5192), 1885–1886, doi:10.1126/science.266.5192.1885, 1994.
- Stuiver, M., Grootes, P. M. and Braziunas, T. F.: The GISP2 $[\delta^{18}\text{O}]$ Climate Record of the Past 16,500 Years and the Role of the Sun, Ocean, and Volcanoes, *Quat. Res.*, 44(3), 341–354, doi:10.1006/qres.1995.1079, 1995.
- Werner, M., Heimann, M. and Hoffmann, G.: Isotopic composition and origin of polar precipitation in present and glacial climate simulations, *Tellus Ser. B Chem. Phys. Meteorol. B*, 53(1), 53–71, doi:10.1034/j.1600-0889.2001.01154.x, 2001.

30