Reply to reviewer #1:

We thank reviewer #1 for the report. Hereafter we address the questions and comments mentioned by the reviewer. Reviewer comments are given in italic letters whereas our replies are given in normal letters. Changes in the revised manuscript are given in bold letters.


We added the reference and changed the sentence to:

“A relatively new method for ice-core-based temperature reconstructions uses the thermal fractionation of stable isotopes of air compounds (nitrogen and argon) within a firn layer of an ice sheet (Huber et al., 2006; Kindler et al., 2014; Kobashi et al., 2011; Orsi et al., 2014; Severinghaus et al., 1998, 2001).”

On page 4, line 34: the statement is made that, except for the spin-up, the method is completely independent of water isotopes. If this is really the case, why is the spin-up needed at all? It seems to me that your statement would be more accurate if you said something like "the method is mostly independent of water isotopes." It would also be useful to add a graph in the Supplement showing the resulting temperature histories when a water isotope spin-up was, and was not, used. That would give the reader a feeling for how independent the method really is.

The spin-up is needed to bring the firn state to a well-defined starting point for the optimization. This is necessary as it takes possible memory effects (influence of earlier conditions) of firn states into account as mentioned in the paper at page 10 (“Surface-temperature spin-up”). Without an adequate spin-up it would require a significant extension of allowed amplitudes and time window of adjustment which would lead to an immense increase in computation time. The effect of a “wrong” spin-up can last for several millennia for large differences from the real case and is dependent on the amplitudes of temperature changes in this section. As the synthetic temperature and δ¹⁵N target-scenarios are created to mimic measured Holocene data, a spin-up that includes the large temperature excursions leading into the Holocene and creating partly the falling slope of the isotope data in the early Holocene, is needed. It is not important from which proxy the temperature spin-up is calculated, only the “correctness” of the temperature signal is important. Because of the argumentation above, it is not expedient to try to fit the gas-isotope data without using a spin-up. Also if we would use an unknown spin-up approach, we have to extend the time domain in which we allow the minimization algorithm to work and would need extended target data for this time.

On page 5, line 8: The argument is made that because both ¹⁵N and ⁴⁰Ar are used, there is no possibility to compare one solution against the other. This statement is rather misleading. In fact, the reason that both isotope pairs were used by Orsi et al. was to eliminate the gravitational signal from the data, which substantially reduces the uncertainty in the final derived temperature. When using only one isotope pair, the uncertainty is much larger, due to the need to match the gravitational signal accurately. Unfortunately, the gravitational signal is affected by many non-temperature processes, such as changes in the convective zone thickness, imperfections in the accumulation data accuracy, imperfections in the firn thickness model, and uncertainties in the true lock-in density. These additional uncertainties amount to 2-3 deg C of extra uncertainty in many real-world cases (for example, see Kobashi et al. 2008)

In summary, the use of only one isotope pair is highly inferior to the use of two; and saying that it is an advantage to be able to compare one to the other is like saying that "half of a bicycle is better than a whole bicycle because it is possible to compare the performance of the rider when riding only the front wheel, with the performance of the rider when riding only the back wheel.”

This sentence must therefore be cut prior to publication. Instead, the authors could say something like "The method presented here can be used when no ⁴⁰Ar data are available, which is often the case because ⁴⁰Ar is a more analytically challenging measurement and is not as commonly measured as ¹⁵N."

We changed the corresponding statement as follows based on the reviewer’s suggestion:

Old version reads as follows:

“As both methods rely on δ¹⁵N together with δ⁴⁰Ar, they do not offer the possibility to validate one isotope-based solution against the other. Also these two approaches can only be applied to ice cores where both isotope quantities are measured together with a sufficient precision.”

We changed the corresponding statement as follows based on the reviewer’s suggestion:
“The method presented here can be used when no $\delta^{40}$Ar data are available, which is often the case because $\delta^{40}$Ar is a more analytically challenging measurement and is not as commonly measured as $\delta^{15}$N and further allows a comparison among solutions obtained from any of the available isotope quantities.”

Page 13, line 7: the word "synchronicity" should be replaced with "synchroneity". Synchronicity is a term from Jungian psychology, and I do not think the authors intended this meaning.

We corrected for this and changed the sentence to:

“This is needed to insure synchroneity between the high-frequency temperature variations $\Delta T(t)$ extracted from the mismatch $D_{\delta^{15}N,mc,fin}(t)$ on the ice-age scale and the smooth temperature solution $T_{mc,fin}(t)$.”

Page 29, line 4: The convective zone (CZ) does not include a non-diffusive zone. In fact there is molecular diffusion occurring in the CZ; it is simply overwhelmed by mixing due to convection. I suggest you replace the words "non-diffusive zone" with "well-mixed zone".

We changed the sentence according to the reviewer advice.

“Many studies have shown the existence of a well-mixed zone at the top of the diffusive-firn-column, called convective zone (CZ).”

The authors are thanked for their close attention to my prior comments.

Thanks belong to the reviewer who helped to improve our manuscript.
**Novel automated inversion algorithm for temperature reconstruction using gas isotopes from ice cores**

Michael Döring\(^1,2\)* and Markus Leuenberger\(^1,2\)

\(^1\)Climate and Environmental Physics, University of Bern, Switzerland
\(^2\)Oeschger Centre for Climate Change Research (OCCR), Bern, Switzerland

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

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

**Abstract.** Greenland past temperature history can be reconstructed by forcing the output of a firn-densification and heat-diffusion model to fit multiple gas-isotope data (\(\delta^{15}N\) or \(\delta^{40}Ar\) or \(\delta^{15}N_{\text{excess}}\)) extracted from ancient air in Greenland ice cores using published accumulation-rate (Acc) data-sets. We present here a novel methodology to solve this inverse problem, by designing a fully-automated algorithm. To demonstrate the performance of this novel approach, we begin by intentionally constructing synthetic temperature-histories and associated \(\delta^{15}N\) datasets, mimicking real Holocene data that we use as “true values” (targets) to be compared to the output of the algorithm. This allows us to quantify uncertainties originating from the algorithm itself. The presented approach is completely automated and therefore minimizes the “subjective” impact of manual parameter-tuning, leading to reproducible temperature-estimates. In contrast to many other ice-core-based temperature-reconstruction methods, the presented approach is completely independent from ice-core stable-water-isotopes, providing the opportunity to validate water-isotope-based reconstructions or reconstructions where water isotopes are used together with \(\delta^{15}N\) or \(\delta^{40}Ar\). We solve the inverse problem \(T(\delta^{15}N, \text{Acc})\) by using a combination of a Monte-Carlo-based iterative approach and the analysis of remaining mismatches between modelled and target data, based on cubic-spline-filtering of random numbers and the laboratory-determined temperature-sensitivity for nitrogen isotopes. Additionally, the presented reconstruction approach was tested by fitting measured \(\delta^{40}Ar\) and \(\delta^{15}N_{\text{excess}}\) data, which leads as well to a robust agreement between modelled and measured data. The obtained final mismatches follow a symmetric standard-distribution-function. For the study on synthetic data, 95% of the mismatches compared to the synthetic target-data are in an envelope between 3.0 permeg to 6.3 permeg for \(\delta^{15}N\) and 0.23 K to 0.51 K for temperature (2\(\sigma\), respectively). In addition to Holocene temperature-reconstructions, the fitting approach can also be used for glacial temperature-reconstructions. This is shown by fitting of NGRIP \(\delta^{15}N\) data for two Dansgaard-Oeschger events using the presented approach, leading to results comparable to other studies.
1 Introduction

Holocene climate variability is of key interest to our society, since it represents a time of moderate natural variations prior to anthropogenic disturbance, often referred to as a baseline for today’s increasing greenhouse effect driven by mankind. Yet, high-resolution studies are still very sparse and therefore limit the investigation of decadal and even centennial climate variations over the course of the Holocene. One of the first studies about changes in the Holocene climate was conducted in the early 1970s by Denton and Karlen (1973). The authors investigated rapid changes in glacier extents around the globe potentially resulting from variations of Holocene climatic conditions. Mayewski et al. (2004) used these data as the base of a multiproxy study identifying rapid climate changes (so called RCCs) globally distributed over the whole Holocene time period. Although not all proxy data are showing an equal behaviour in timing and extent during the quasi-periodic RCC patterns, the authors found evidence for a highly variable Holocene climate controlled by multiple mechanisms, which significantly affects ecosystems (Pál et al., 2016; Beaulieu et al., 2017; Crausbay et al., 2017). Although not all proxy data are showing an equal behaviour in timing and extent during the quasi-periodic RCC patterns, the authors found evidence for a highly variable Holocene climate controlled by multiple mechanisms, which significantly affects ecosystems (Beaulieu et al., 2017; Crausbay et al., 2017; Pál et al., 2016) and human societies (Holmgren et al., 2016; Lespez, L. et al., 2016). Precise high-resolution temperature-estimates can contribute significantly to the understanding of these mechanisms. Ice-core proxy-data offer multiple paths for reconstructing past climate and temperature variability. The studies of Cuffey et al. (1995), Cuffey and Clow (1997) and Dahl-Jensen et al. (1998) demonstrate the usefulness of inverting the measured borehole-temperature profile for surface-temperature-history estimates for the investigated drilling site using a coupled heat- and ice-flow model. Because of smoothing effects due to heat-diffusion within an ice sheet, this method is unable to resolve fast temperature oscillations and leads to a rapid reduction of the time resolution towards the past. Another approach to reconstruct past temperature is based on the calibration of water-stable-isotopes of oxygen and hydrogen ($\delta^{18}$O$_{\text{ice}}$, $\delta$D$_{\text{ice}}$) from ice-core water-samples assuming a constant (and mostly linear) relationship between temperature and isotopic composition due to fractionation effects during ocean evaporation, cloud formation and snow and ice precipitation (Stuiver et al., 1995; Johnsen et al., 2001; Stuiver et al., 1995). This method provides a rather robust tool for reconstructing past temperature for times where large temperature excursions occur when an adequate relationship is used (Dansgaard-Oeschger events, Glacial-Interglacial transitions (Dansgaard et al., 1982; Johnsen et al., 1992)). Also, in the Holocene where Greenland temperature variations are comparatively small, seasonal changes of precipitation as well as of evaporation conditions at the source region may contribute to water-isotope-data variations (Werner et al., 2001; Huber et al., 2006; Kindler et al., 2014). A relatively new method for ice-core-based temperature reconstructions uses the thermal fractionation of stable isotopes of air compounds (nitrogen and argon) within a firm layer of an ice sheet (Severinghaus et al., 1998; Severinghaus et al., 2001; Huber et al., 2006; Kobashi et al., 2011; Kindler et al., 2014). (Huber et al., 2006; Kindler et al., 2014; Kobashi et al., 2011; Orsi et al., 2014; Severinghaus et
The measured nitrogen- and argon-isotope records of air enclosed in bubbles in an ice core can be used as a paleothermometer due to (i) the stability of isotopic compositions of nitrogen and argon in the atmosphere at orbital timescales and (ii) the fact that changes are only driven by firn processes (Mariotti, 1983; Severinghaus et al., 1998; Leuenberger et al., 1999). To robustly reconstruct the surface temperature for a given drilling site, the use of firn models describing gas- and heat-diffusion throughout the ice sheet is necessary to decompose the gravitational from the thermal-diffusion influence on the isotope signals.

This work addresses two issues relevant for temperature reconstructions based on nitrogen and argon isotopes. First, we introduce a novel, entirely automated approach for inverting gas-isotope data to surface-temperature estimates. For that, we force the output of a firn-densification and heat-diffusion model to fit gas-isotope data. This methodology can be used for many different optimization tasks not restricted to ice-core data. As we will show, the approach works besides $\delta^{15}$N for all relevant gas-isotope quantities ($\delta^{15}$N, $\delta^{40}$Ar, $\delta^{15}$N$_{\text{excess}}$) and for Holocene and glacial data as well. Furthermore, the possibility of fitting all relevant gas-isotope quantities, individually or combined, makes it possible for the first time to validate the temperature solution gained from one single isotope species by comparison to the solution calculated from other isotope quantities. This approach is a completely new method which enables the automated fitting of gas-isotope data without any manual tuning of parameters, minimizing any potential “subjective” impacts on temperature estimates as well as working hours. Also, except for the model spin-up, the presented temperature-reconstruction approach is completely independent from water stable-isotopes ($\delta^{18}$O$_{\text{ice}}$, $\delta^{2}D_{\text{ice}}$), which provides the opportunity to validate water-isotope-based reconstructions (e.g. Masson-Delmotte, 2005) or reconstructions where water isotopes are used together with $\delta^{15}$N or $\delta^{40}$Ar (e.g. Landais et al., 2004; Huber et al., 2006; Capron et al., 2010). Capron et al., 2010; Huber et al., 2006; Landais et al., 2004). To our knowledge, there are only two other reconstruction methods independent from water stable-isotopes that have been applied to Holocene gas-isotope data, without a priori assumption on the shape of a temperature change. The studies from Kobashi et al. (2008a, 2017) use the second order parameter $\delta^{15}$N$_{\text{excess}}$ to calculate firn-temperature gradients, which are later temporally integrated from past to future over the time-series of interest using the firn-densification and heat-diffusion model from Goujon et al. (2003). Additionally Orsi et al. (2014) use a linearized firn-model approach together with $\delta^{15}$N and $\delta^{40}$Ar data to extract surface-temperature histories. As both methods rely on $\delta^{15}$N together with $\delta^{40}$Ar, they do not offer the possibility to validate one isotope-based solution against the other. Also these two approaches can only be applied to ice cores where both isotope quantities are measured together with a sufficient precision. The method presented here can be used when no $\delta^{40}$Ar data are available, which is often the case because $\delta^{40}$Ar is a more analytically challenging measurement and is not as commonly measured as $\delta^{15}$N and further allows a comparison among solutions obtained from any of the available isotope quantities.

Second, we investigate the accuracy of our novel fitting approach by examining the method on different synthetic nitrogen-isotope and temperature scenarios. The aim of this work is to study the uncertainties emerging from the algorithm itself. Furthermore the focal question in this study is: what is the minimal mismatch in $\delta^{15}$N for Holocene-like data we can reach and what is the implication for the final temperature mismatches. Studying and moreover answering these questions makes it
mandatory to create well defined $\delta^{15}$N targets and related temperature histories. It is impossible to answer these questions without using synthetic data in a methodology study. The aim is to evaluate the accuracy and associated uncertainty of the inverse method itself to then later apply this method to real $\delta^{15}$N, $\delta^{40}$Ar or $\delta^{15}$N$_{excess}$ datasets, for which of course the original driving temperature histories are unknown.
2. Methods and data

2.1 Reconstruction approach

The problem that we deal with is an inverse problem, since the effect, observed as $\delta^{15}$N variations, is dependent on its drivers, i.e. temperature and accumulation-rate changes. Hence, the temperature that we would like to reconstruct depends on $\delta^{15}$N and accumulation-rate changes. To solve this inverse problem, the firn-densification and heat-diffusion model (from now on referred to as firn model), which is a non-linear transfer function of temperature and accumulation rate to firn states and relates to $\delta^{15}$N values, is run iteratively to match the modelled and measured $\delta^{15}$N values (or other gas species). The automated procedure is significantly more efficient and less time-consuming than a manual approach. The Holocene temperature-reconstruction is implemented by the following four steps (Fig. 01):

Step 1: A prior temperature input (first guess) is constructed, which serves as the starting point for the optimization.

Step 2: A long-term solution which passes through the $\delta^{15}$N data (here synthetic target data) is generated following a Monte-Carlo approach. It is assumed that the smooth solution contains all long-term temperature trends (centuries to millennial) as well as firn-column-height changes (temperature and accumulation-rate dependent) that drive the gravitational background signal in $\delta^{15}$N.

Step 3: The long-term temperature solution is complemented by superimposing short-term information directly extracted from the $\delta^{15}$N data (here synthetic target data). This step adds short-term temperature changes (decadal) in the same time resolution as the data.

Step 4: The gained temperature solution is further corrected using information extracted from the mismatch between the synthetic target and modelled $\delta^{15}$N time-series.

The functionality of the presented inversion algorithm is schematically displayed in Fig. 01. It guides the reader through chapters and documents which variables, listed in Table 01, are in use. In the following a detailed description of each step is given.

**Step 1: prior input**

The starting point of the optimization procedure is the first-guess. To construct the first-guess temperature-input $T_{g0}(t)$, a constant temperature of -29.6 °C is used for the complete Holocene section, which corresponds to the last value of the temperature spin-up (Fig. 02b).
Step 2: Monte-Carlo-type input-generator - Generating long-term solutions

During the second step of the optimization, the prior temperature-input $T_{g,0}(t)$ from step 1 is iteratively (j) changed following a Monte-Carlo approach. The basic idea of the Monte-Carlo approach is to generate smooth temperature-inputs $T_{mc,j}(t)$ by superimpose low-pass-filtered values $\tilde{P}_j$ of uniformly-distributed random values $\tilde{P}_{r,j}$ on the prior input $T_{mc,j-1}$. Then, the new input is fed to the firn model and the mismatch $D_{\delta^{15}N_{mc,j}}$ (with $X = \delta^{15}N_{mc,j}$) between the modelled $\delta^{15}N_{mc,j}$ (here $X_{mod}$), calculated from the model output, and the synthetic $\delta^{15}N_{syn}$ (here $X_{target}$) is computed for every time step (i) of the target data $\delta^{15}N_{syn}$ according to:

$$D_X = \frac{1}{n} \sum_{i=1}^{n} |X_{target,i} - X_{mod,i}|$$

(Note: If not otherwise stated, all mismatches in this study labelled with “D” are calculated similar to eq. (1))

$D_{\delta^{15}N_{mc}}$ serves as the criterion which is minimised during the optimization in step 2. If the mismatch $D_{\delta^{15}N_{mc,j}}$ decreases compared to the prior input ($T_{mc,j-1}$, $D_{\delta^{15}N_{mc,j-1}}$), the new input is saved and used as new guess ($T_{g,j} = T_{mc,j}$). This procedure is repeated until convergence is achieved leading to the final long-term temperature $T_{mc,fin}(t)$. Table 02 lists the number of improvements and iterations performed for the different synthetic datasets.

The perturbation of the current guess $T_{g,j}$ is conducted in the following way: Let $\tilde{T}_{g,0} = T_{g,0}(t)$ be the vector containing the prior temperature-input. A second vector $\tilde{P}_{r,1}$ with the same number of elements $n_{mc}$ as $\tilde{T}_{g,0}$ is generated containing $n_{mc}$ uniformly-distributed random numbers within the limits of an also randomly (equally-distributed) chosen standard deviation $s$. $s$ is chosen from a range of 0.05-0.50, which means that the maximum allowed perturbation of a single temperature value $T(t_0)$ is in a range of ±5 % to ±50 %. Creating the synthetic frequencies, $\tilde{P}_{r,1}$ is low-pass filtered using cubic-spline-filtering (Enting, 1987) with an equally distributed random cut-off-period (COP) in the range of 500 yr to 2000 yr generating the vector $\tilde{P}_1$. Hereby the low-pass filtering of $\tilde{P}_{r,1}$ reduces the amplitudes of the perturbation of $\tilde{T}_{g,0}$. The new surface temperature input $\tilde{T}_{mc,1}$ is calculated from $\tilde{P}_1$ according to:

$$\tilde{T}_{mc,1} = \tilde{T}_{g,0} \cdot (\tilde{I} + \tilde{P}_1)$$

The superscript “T” stands for transposed and $\tilde{I}$ is the $n$ by 1 matrix of ones.

This approach provides a high potential for parallel computing. In this study, an eight-core computer was used, generating and running eight different inputs of $\tilde{T}_{mc}$ simultaneously, minimizing the time to find an improved solution. For example, during the 706 iterations for scenario S2, about 5600 different inputs were created and tested, leading to 351 improvements (see Table 02). Since it is possible to find more than one improvement per iteration step due to the parallelization on eight CPU’s, the solution giving the minimal misfit $D_{\delta^{15}N_{mc,j}}$ is chosen as new first-guess for the next iteration step. This leads to a decrease of the used improvements for the optimization (e.g. for S2, 172 of the 351 improvements were used). Additionally, a first gas-age scale ($\Delta$age$_{mc,fin}(t)$) is extracted from the model using the last improved conditions, which will then be used in step 3.

(Note: If not otherwise stated, all mismatches in this study labelled with “D” are calculated similar to eq. (1))
Step 3: Adding short-term (high frequency) information

In step 3 the missing short-term temperature history providing a suitable fit between modelled and synthetic $\delta^{15}$N data is directly extracted from the pointwise mismatch $D_{\delta^{15}N,mc,fin}(t)$, between the modelled $\delta^{15}$N$_{mc,fin}(t)$ obtained in step 2 and the synthetic $\delta^{15}$N$_{syn}$ target. Note that for a real reconstruction, this mismatch is calculated using the measured $\delta^{15}$N$_{meas}$ dataset instead of the synthetic one. $D_{\delta^{15}N,mc,fin}(t)$ can be interpreted in first order as the detrended high-frequency signal of the synthetic $\delta^{15}$N$_{syn}$ target. $D_{\delta^{15}N,mc,fin}(t)$ is transferred to the gas-age scale using $\Delta$age$_{mc,fin}(t)$ provided by the firn-model output for the smooth temperature input $T_{mc,fin}(t)$. This is needed to insure synchronicity between the high-frequency temperature variations $\Delta T(t)$ extracted from the mismatch $D_{\delta^{15}N,mc,fin}(t)$ on the ice-age scale and the smooth temperature solution $T_{mc,fin}(t)$. Additionally, the signal is shifted by about 10 yr towards modern values to account for gas diffusion from the surface to the lock-in-depth (Schwander et al., 1993), which is not yet implemented in the firn model. This is necessary for adding the calculated short-term temperature changes $\Delta T(t)$ to the smooth signal $T_{mc,fin}(t)$. The $\Delta T$ values are calculated according to eq. (3):

$$\Delta T_i = \frac{D_{\delta^{15}N,mc,fin,i}}{\Omega_{N_2,i}}$$

(3)

using the thermal-diffusion sensitivity $\Omega_{N_2,i}$ for nitrogen-isotope fractionation from Grachev and Severinghaus (2003):

$$\Omega_{N_2,i} = \frac{8.656 \%}{T_i} - \frac{1232 \%}{T_i^2} \cdot K$$

(4)

$T_i$ is the mean firn temperature in Kelvin which is calculated by the firn model for each time point $i$. To reconstruct the final (high frequency) temperature input $T_{hf}(t)$, the extracted short-term temperature signal $\Delta T(t)$ is simply added to the long-term temperature input $T_{mc,fin}(t)$:

$$T_{hf,i} = T_{mc,fin,i} + \Delta T_i$$

(5)

Step 4: final correction of the surface temperature solution

For a further improvement of the remaining $\delta^{15}$N and resulting surface-temperature misfits ($D_{\delta^{15}N,ad}(t)$, $D_{T,ad}(t)$), it is important to find a correction method that contains information that is also available when using measured data. The benefit of the synthetic data study is that several later unknown quantities can be calculated, and used for improving the reconstruction approach (see Sect. 3 and 4). For instance, it is possible to split the synthetic $\delta^{15}$N$_{syn}$ data in the gravitational and thermo-diffusion parts or to use the temperature misfit, which is unknown in reality. The idea underlying the correction algorithm explained hereafter is that the remaining misfits of $\delta^{15}$N ($D_{\delta^{15}N,ad}(t)$) and temperature ($D_{T,ad}(t)$) are connected to the Monte-Carlo (step 2) and high-frequency part (step 3) of the reconstruction algorithm. In the present inversion framework, it is not possible to find a long-term solution $\delta^{15}$N$_{mc,fin}$ (or $T_{mc,fin}$) which exactly passes through the $\delta^{15}$N$_{syn}$ (or $T_{syn}$) target in the middle of the variance in all parts of the time-series. This leads to a slightly over- or underestimation of $\delta^{15}$N$_{mc,fin}(t)$ and their corresponding temperature values $T_{mc,fin}(t)$. For example, a slightly too low (or too high) smooth temperature estimate $T_{mc,fin}$ leads to a small increase (or decrease) of the firn-column-height, creating a wrong gravitational background signal in
$\delta^{15}\text{N}_{\text{mc,fin}}$ on a later point in time (because the firm column needs some time to react). An additional error in the thermal-diffusion signal is also created due to the high-frequency part of the reconstruction (step 3), because the high-frequency information is directly extracted from the deviation of the synthetic target $\delta^{15}\text{N}_{\text{syn}}(t)$ and the modelled $\delta^{15}\text{N}_{\text{mc,fin}}(t)$ from the final long-term solution $T_{\text{mc,fin}}(t)$ of the Monte-Carlo part. Therefore, this error is transferred into the next step of the reconstruction and partly creates the remaining deviations.

To investigate this problem, the deviations $D_{\delta^{15}\text{N,mc,fin}}(t)$ of the synthetic target data $\delta^{15}\text{N}_{\text{syn}}$ to $\delta^{15}\text{N}_{\text{mc,fin}}$ of the Monte-Carlo part are numerically integrated over a time window of 200 yr (Sect. 4, Supple. S3), and thereafter the window is shifted from past to future in 1 yr steps resulting in a time-series called $\text{IF}(t)$. $\text{IF}(t)$ equals a 200 yr running-mean of $D_{\delta^{15}\text{N,mc,fin}}(t)$. For $t$, the mid position of the window is allocated. The time evolution of $\text{IF}(t)$ is a measure for the deviation of the long-term solution $\delta^{15}\text{N}_{\text{mc,fin}}(t)$ (or $T_{\text{mc,fin}}(t)$) from the perfect middle passage through the target data $\delta^{15}\text{N}_{\text{syn}}(t)$ (or $T_{\text{syn}}(t)$) and for the slightly over- and underestimation of the resulting temperature.

$$\text{IF}(t) = \frac{1}{200} \int_{t-100}^{t+100} (\delta^{15}\text{N}_{\text{syn}}(t) - \delta^{15}\text{N}_{\text{mc,fin}}(t)) \, dt$$

(6)

Next, the sample-cross-correlation-function (xcf) (Box et al., 1994) is applied to $\text{IF}(t)$ and the remaining misfits $D_{\delta^{15}\text{N, hf}}(t)$ of $\delta^{15}\text{N}$ after the high-frequency part. The xcf shows two extrema (Fig. 03a), a maximum ($\text{xcf}_{\text{max}}$) and a minimum ($\text{xcf}_{\text{min}}$) at two certain lags ($\text{lag}_{\text{max}}, \delta^{15}\text{N}$ at $\text{xcf}_{\text{max}}, \delta^{15}\text{N}$ and $\text{lag}_{\text{min}}, \delta^{15}\text{N}$ at $\text{xcf}_{\text{min}}, \delta^{15}\text{N}$). Now, the same analysis is conducted for $\text{IF}(t)$ versus the temperature mismatch $D_{\text{T, hf}}(t)$ (Fig. 03b), which shows an equal behaviour (two extrema, $\text{lag}_{\text{max}}, \text{T}$ at $\text{xcf}_{\text{max}}, \text{T}$ and $\text{lag}_{\text{min}}, \text{T}$ at $\text{xcf}_{\text{min}}, \text{T}$). Comparing the two cross correlations show that $\text{lag}_{\text{max}}, \delta^{15}\text{N}$ equals the negative $\text{lag}_{\text{min}}, \text{T}$ and $\text{lag}_{\text{min}}, \delta^{15}\text{N}$ corresponds to the negative $\text{lag}_{\text{max}}, \text{T}$ (Fig. 03d,e). The idea for the correction is that the extrema in the cross-correlation $\text{IF}(t)$ vs. $D_{\delta^{15}\text{N, hf}}(t)$ with the positive lag (positive means here that $D_{\delta^{15}\text{N, hf}}(t)$ has to be shifted to past values relative to $\text{IF}(t)$) creates the misfit of temperature $D_{\text{T, hf}}(t)$ on the negative lag (modern direction) of $\text{IF}(t)$ vs. $D_{\text{T, hf}}(t)$ and vice versa. So $\text{IF}(t)$ yields information about the cause and allows us to correct this effect between the remaining mismatches $D_{\delta^{15}\text{N, hf}}(t)$ and $D_{\text{T, hf}}(t)$ over the whole time-series. The lags are not sharp signals, due to the fact that (i) the cross-correlations are conducted over the whole analysed record, leading to an averaging of this cause and effect relationship as well as that (ii) $\text{IF}(t)$ is a smoothed quantity itself. The correction of the reconstructed temperature after the high-frequency part is conducted in the following way: From the two linear relationships between $\text{IF}(t)$ and $D_{\delta^{15}\text{N, hf}}(t)$ at the two lags ($\text{lag}_{\text{max}}, \delta^{15}\text{N}$ at $\text{xcf}_{\text{max}}, \delta^{15}\text{N}$, $\text{lag}_{\text{min}}, \delta^{15}\text{N}$ at $\text{xcf}_{\text{min}}, \delta^{15}\text{N}$) two sets of $\delta^{15}\text{N}$ correction values ($\Delta\delta^{15}\text{N}_{\text{max}}(t)$ from $\text{xcf}_{\text{max}}, \delta^{15}\text{N}$ and $\Delta\delta^{15}\text{N}_{\text{min}}(t)$ from $\text{xcf}_{\text{min}}, \delta^{15}\text{N}$) are calculated. Then the lags are being inverted (Fig. 03c,e) shifting the two sets of the $\delta^{15}\text{N}$ correction values to the attributed lags of the cross correlation between $\text{IF}(t)$ and $D_{\text{T, hf}}(t)$ (e.g. $\Delta\delta^{15}\text{N}_{\text{min}}(t)$ to lag from $\text{xcf}_{\text{max}}, \text{T}$ from the cross correlation between $\text{IF}(t)$ and $D_{\text{T, hf}}(t)$) therefore changing the time assignments of $\Delta\delta^{15}\text{N}_{\text{min}}(t)$ and $\Delta\delta^{15}\text{N}_{\text{max}}(t)$ to $\Delta\delta^{15}\text{N}_{\text{min}}(t+\text{lag}_{\text{max}}, \text{T})$ and $\Delta\delta^{15}\text{N}_{\text{max}}(t+\text{lag}_{\text{min}}, \text{T})$. Now, the $\Delta\delta^{15}\text{N}_{\text{max}}(t)$ and $\Delta\delta^{15}\text{N}_{\text{min}}(t)$ are component-wise summed up leading to the time-series $\Delta\delta^{15}\text{N}_{\text{cv}}(t)$. From eq. (3) with $\Delta\delta^{15}\text{N}_{\text{cv,i}}$ instead of $D_{\delta^{15}\text{N,mc,fin}}$, the corresponding temperature correction values are calculated and added to the high-frequency temperature solution $T_{\text{hf}}(t)$ giving the corrected temperature $T_{\text{corr}}(t)$. Finally, $T_{\text{corr}}(t)$ is used to run the firm model to calculate the corrected $\delta^{15}\text{N}_{\text{corr}}(t)$ time-series. This cause and effect relationship found in the cross-correlations between $\text{IF}(t)$ and
D_{15N,int}(t), and IF(t) and D_{T,int}(t), is exemplarily shown in Fig. 03 for scenario S1 and was found for all eight synthetic scenarios. The derived correction algorithm leads to a further reduction of the mismatches of about 40 % in $\delta^{15}$N and temperature (see Sect. 3.2).

2.2 Firn densification and heat diffusion model

Surface-temperature reconstruction relies on firn densification combined with gas- and heat-diffusion (Severinghaus et al., 1998). In this study, the firn-densification and heat-diffusion model, developed by Schwander et al. (1997) is used to reconstruct firn parameters for calculating synthetic $\delta^{15}$N values depending on the input time-series. It is a semi-empirical model based on the work of Herron and Langway (1980), Barnola et al. (1991), and implemented using the Crank and Nicholson algorithm (Crank, 1975) and was also used for the temperature reconstructions by Huber et al. (2006) and Kindler et al. (2014). Besides surface-temperature time-series, accurate accumulation-rate data are needed to run the model. The model then calculates the densification and heat-diffusion history of the firn layer and provides parameters for calculating the fractionation of the nitrogen isotopes for each time step, according to the following equations:

$$\delta^{15}\text{N}_{\text{grav}}(z_{\text{LID}}, t) = \left( \frac{\Delta m}{R T(t)} \right) \cdot 1000$$  \hspace{1cm} (7)

$$\delta^{15}\text{N}_{\text{therm}}(t) = \left[ \left( \frac{T_{\text{surf}}(t)}{T_{\text{bottom}}(t)} \right) \cdot g \cdot \rho_{\text{LID}}(t) - 1 \right] \cdot 1000$$ \hspace{1cm} (8)

$$\delta^{15}\text{N}_{\text{mod}}(t) = \delta^{15}\text{N}_{\text{grav}}(t) + \delta^{15}\text{N}_{\text{therm}}(t)$$ \hspace{1cm} (9)

$\delta^{15}\text{N}_{\text{grav}}(t)$ is the component of the isotopic fractionation due to the gravitational settling (Craig et al., 1988; Schwander, 1989) and depends on the lock-in-depth (LID) $z_{\text{LID}}(t)$ and the mean firm temperature $\bar{T}(t)$ (Leuenberger et al., 1999). $g$ is the gravitational acceleration, $\Delta m$ the molar mass-difference between the heavy and light isotopes (equals $10^{-3}$ kg per mol for nitrogen) and R the ideal gas-constant. $z_{\text{LID}}$ is defined as a density threshold $\rho_{\text{LID}}$, which is slightly sensitive to surface temperature, following the formula from Martinierie et al. (1994), with a small offset correction of 14 kg m$^{-3}$ to account for the presence of a non-diffusive zone (Schwander et al., 1997):

$$\rho_{\text{LID}}(kg \cdot m^{-3}) = \frac{1}{\rho_{\text{ice}} - 6.95 \cdot 10^{-7} \cdot \bar{T} - 4.3 \cdot 10^{-5} \cdot \bar{T}^2} - 14$$  \hspace{1cm} (10)

where

$$\rho_{\text{ice}}(kg \cdot m^{-3}) = 916.5 - 0.14438 \cdot \bar{T} - 1.5175 \cdot 10^{-4} \cdot \bar{T}^2$$  \hspace{1cm} (11)

The thermal-fractionation component of the $\delta^{15}$N signal (Severinghaus et al., 1998) is calculated using eq. (8), where $T_{\text{surf}}(t)$ and $T_{\text{bottom}}(t)$ stand for the temperatures at the top and the bottom of the diffusive firn-layer. In contrast to $T_{\text{surf}}(t)$ which is an input parameter for the model, $T_{\text{bottom}}(t)$ is calculated by the model for each time step. The thermal-diffusion constant $\alpha_T$ was measured by Grachev and Severinghaus (2003) for nitrogen (eq. (12)):

$$\alpha_T = \left( 8.656 - \frac{1323}{\bar{T}} \right) \cdot 10^{-3}$$ \hspace{1cm} (12)
The firn model used here behaves purely as a forward model, which means that for the given input time-series the output parameters (here finally $\delta^{15}N_{\text{mod}}(t)$) can be calculated, but it is not easily possible to construct from measured isotope data the related surface-temperature or accumulation-rate histories. The goal of the presented study is an automatization of this inverse-modelling procedure for the reconstruction of the rather small Holocene temperature variations.

2.3 Measurement, input data and time scale

Time scale

For the entire study the GICC05 chronology is used (Rasmussen et al., 2014; Seierstad et al., 2014). During the whole reconstruction procedure the two input time-series (surface temperature and accumulation rate) are split into two parts. The first part ranges from 20 yr to 10520 yr b2k (called “Holocene section”) and the second one from 10520 yr to 35000 yr b2k (“spin-up section”). The entire accumulation-rate input, as well as the spin-up section of the surface-temperature input remains unchanged during the reconstruction procedure.

Accumulation-rate data

Besides surface temperatures, accumulation-rate data are needed to drive the firn model. In this study we use the original accumulation rates, reconstructed in Cuffey and Clow (1997) produced using an ice-flow model adapted to the GISP2 location, but adapted to the GICC05 chronology (Rasmussen et al., 2008; Seierstad et al., 2014). A detailed description of the adaption procedure can be found in supplement S1. The raw accumulation-rate data for the main part of the spin-up section (12000 yr to 35000 yr b2k) are linearly interpolated to a 20 yr grid and low-pass filtered with a 200 yr cut-off-period (COP) using cubic-spline-filtering (Enting, 1987). For the Holocene section (20-10520 yr b2k) and the transition part between Holocene and spin-up section (10520 yr to 12000 yr b2k) the raw accumulation-rate data are linearly interpolated to a 1 yr grid to obtain equidistant integer point-to-point distances which are necessary for the reconstruction, and to preserve as much information as possible for this time period (Fig. 02a). Except for these technical adjustments, the accumulation-rate input remains unmodified, assuming high reliability of these data during the Holocene. The accumulation data were reconstructed using annual-layer-counting, and a thinning model which should lead to maximum relative uncertainty of 10 % for the first 1500 m of the 3000 m ice core (Cuffey and Clow, 1997). From the three accumulation-rate scenarios reconstructed in Cuffey and Clow (1997) and adapted here to the GICC05 chronology, the intermediate one is chosen (red curves in Fig. S01). Since the differences between the scenarios are not important for the evaluation of the reconstruction approach, they are not taken into account for this study.

Additionally, two sensitivity experiments were conducted (see supplement S2) in order to investigated (i) the influence of low-pass-filtering of the high-resolution accumulation rates on the model outputs and (ii) the possible contribution of the accumulation-rate variability on the $\delta^{15}N$ data during the Holocene. The first experiment shows that filtering the accumulation-rates with cut-off-periods in the range of 20 yr to 500 yr has nearly no influence on the modelled $\delta^{15}N$ or lock-in-depth as long as the major trends are being conserved. The second experiment leads to the finding that the accumulation-
rate variability explains about 12 % to 30 % of δ¹⁵N variability. 30 % corresponds to the 8.2 kyr event and 12 % for the mean of the whole Holocene period including the 8.2 kyr event. Hence the influence of accumulation changes, excluding the extreme 8.2 kyr event, is generally below 10 % during most parts of the Holocene.

δ¹⁸Oice data

Oxygen-isotope data from the GISP2 ice-core-water samples measured at the University of Washington’s Quaternary Isotope Laboratory are used to construct the surface-temperature input of the model spin-up (12 yr to 35 kyr b2k, Grootes et al., 1993; Meese et al., 1994; Steig et al., 1994; Stuiver et al., 1995; Grootes and Stuiver, 1997; 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).

The raw δ¹⁸Oice data are filtered and interpolated in the same way as the accumulation-rate data for the spin-up part.

Surface-temperature spin-up

The surface-temperature history of the spin-up section (Fig. 02a) is obtained by calibrating the filtered and interpolated δ¹⁸Oice data (eq. (13)) using the values for the temperature sensitivity α¹⁸O and offset β found by Kindler et al. (2014) for the NGRIP ice core assuming a linear relationship of δ¹⁸Oice with temperature.

\[ T_{\text{spin}}(t) = \frac{1}{a_{18O}(t)} \cdot \left[ \delta_{18O}^{\text{ice}}(t) + 35.2 \, \text{‰} \right] - 31.4 {^\circ}C + \beta(t) \] (13)

The values 35.2 ‰ and -31.4 °C are modern-time parameters for the GISP2 site (Schwander et al., 1997; Grootes and Stuiver, 1997; Grootes and Stuiver, 1997). The spin-up is needed to bring the firm model to a well-defined starting condition that takes possible memory effects (influence of earlier conditions) of firm states into account.

Generating synthetic target data

In order to develop and evaluate the presented algorithm, eight temperature scenarios were constructed and used to model synthetic δ¹⁵N data, which serve later as targets for the reconstruction. From these eight synthetic surface-temperature and related δ¹⁵N scenarios (S1-S5 and H1-H3), three data sets (later called Holocene like scenarios H1-H3) were constructed in such a way that the resulting δ¹⁵N time-series are very close to the δ¹⁵N values measured by Kobashi et al. (2008b) in terms of variability (amplitudes) and frequency (data resolution) of the GISP2 nitrogen-isotope data (Fig. 04, Fig. 05).

The synthetic surface-temperature scenarios S1-S5 are created by generating a long-term temperature time-series (T_{syn,smooth}) analogous to the Monte-Carlo part of the reconstruction procedure for only one iteration step (see Sect. 2.1). The values for the cut-off-period used for the filtering of the random values, and the s values (standard deviation of the random values, see Sect. 2.1) for the first five scenarios can be found in Table 03. The long-term temperatures (Fig. 04I) are calculated on a 20 yr grid, which is nearly similar to the time resolution of the GISP2 δ¹⁵N measurement values of about 17 yr (Kobashi et al., 2008b). For the Holocene-like scenarios, the smooth temperature time-series were generated from the temperature reconstruction for the GISP2 δ¹⁵N data (not shown here). The final Holocene surface-temperature solution was filtered with a 100 yr cut-off to obtain the long-term temperature scenario.
Following this, high frequency information is added to the long-term temperature histories. A set of normally-distributed random numbers with a zero mean and a standard deviation (1σ) of 1 K for scenarios S1-S5 and 0.3 K for Holocene-like scenarios H1-H3 is generated on the same 20 yr grid and added up to the long-term temperature time-series. Finally, the resulting synthetic target-temperature-scenarios (Fig. 04II, Fig. 05I) are linearly interpolated to a 1 yr grid.

These synthetic temperatures are combined with the spin-up temperature and are used together with the accumulation-rate input to feed the firn model. From the model output the synthetic δ¹⁵N targets are calculated according to section 2.1. The firn-model output provides ice-age as well as gas-age information. The final synthetic δ¹⁵N target time-series (δ¹⁵Nₚₛₚ) are set intentionally on the ice-age scale to mirror measured data, because no prior information is available for the gas-ice-age difference (ΔAge) for ice-core data.

3. Results

3.1 Monte Carlo type input generator

Figure 06 shows the evolution of the misfit Dδ¹⁵Nₘₚ,ᵢ,j between the synthetic target data (δ¹⁵Nₚₛₚ) versus the modelled output δ¹⁵Nᵢ,ᵢ,j of the Monte-Carlo part (step 2) as a function of the applied iterations (j) for all synthetic scenarios. One can easily see that all scenarios show a steep decline of the mismatch during the first 50 to 200 iterations followed by a rather moderate decrease, which finally leads to a constant value. During the Monte-Carlo part, it was possible to reduce the misfit Dδ¹⁵Nₘₚ compared to the first-guess solution Dδ¹⁵Nₙₑ,₀ by about 15 % to 75 % depending on the scenario and the mismatch of the first-guess solution (see Table 02). This leads to a reduction of the temperature mismatches DTₘₚ compared to the first-guess temperature DTₙₑ,₀ mismatch of about 51 % to 87 %.

Figure 07 provides the comparison between the first-guess (g,₀; step 1) and Monte-Carlo (mc, fin; step 2) solution versus the synthetic target data (syn) for the modelled δ¹⁵N (a-c) and surface-temperature values (d-f) for scenario S5. Subplots (a) and (d) show the time-series of the synthetic target (black dotted line), the first-guess solution (blue line) and the Monte-Carlo solution (red line) for δ¹⁵N and temperature. In subplots (b) and (e), the distribution of the pointwise mismatch Dᵢ of the first-guess (blue) and the Monte-Carlo solution (red) versus the synthetic target data for δ¹⁵N (Dδ¹⁵Nₚₛₚ) and temperature (DT) can be found. Subplots (c) and (f) contain the time-series for Dδ¹⁵N,ᵢ,j and DT,ᵢ,j. The Dδ¹⁵N,ᵢ,j,₀(ᵢ) data (red) are used to calculate the high-frequency signal, that is superimposed to the long-term temperature solution Tₘₚ,ᵢ,j according to eq. (3) and eq. (5) (see Sect. 2.1, step 3). From Fig. 07 it can be concluded that the Monte-Carlo part of the reconstruction algorithm (step 2) leads to two major improvements of the first-guess solution. First, it is obvious that the Monte-Carlo approach corrects the offsets of the first-guess input (g,₀), which shifts the midpoint of the distributions of Dδ¹⁵N,ᵢ,j and DT,ᵢ,j to zero (see blue against red in Fig. 07b,e). The second improvement is that the distributions become more symmetric and the misfit is overall reduced (the distributions become narrower) compared to the first-guess, due to the middle passage through the δ¹⁵Nₚₛₚ targets. These improvements can be observed for all eight synthetic scenarios, showing the robustness of the Monte-Carlo part (see Table 02, Fig. 07).
3.2 High frequency step and final correction

Figure 08 provides the comparison between the Monte-Carlo (mc, fin; step 2), the high-frequency (hf; step 3) and the correction (corr; step 4) parts of the reconstruction procedure for the scenarios S5. Additional data for all other scenarios can be found in Table 04. The upper four plots (a-d) illustrate each reconstruction step and their effect on the modelled $\delta^{15}$N; the bottom four plots (e-h) show the corresponding results on the temperature. Plots (a) and (e) contain the time-series of the synthetic $\delta^{15}$N$_{syn}$ or $T_{syn}$ target (syn; black dotted line), the high-frequency solution (hf; blue line), and the final solution after the correction part (corr; red line). For visibility reasons, subplots (b) and (f) display a zoom-in for a randomly chosen time-window of about 500 yr for the same quantities, which shows the excellent agreement in timing and amplitudes of the modelled $\delta^{15}$N and temperature compared to the synthetic target data. Histograms (c) and (g) and subplots (d) and (h) show the distribution and the time-series of the pointwise mismatches ($D_{\delta^{15}N,i}$ for $\delta^{15}$N; $D_{T,i}$ for temperature) between the modelled and the synthetic target data in $\delta^{15}$N and temperature for each reconstruction step.

Compared to the Monte-Carlo solution, the high-frequency part leads to a large refinement of the reconstructions. For the mean $\delta^{15}$N misfits $D_{\delta^{15}N}$, the improvement between the Monte-Carlo and the high-frequency parts is in the range of 64 % to 76 % (see Table 04). This leads to a reduction of the temperature mismatches $D_{T}$ of 43 % to 67 %. The standard deviations (1$\sigma$) of the pointwise mismatches (Fig. 08c,d,g,h) in $\delta^{15}$N and temperature after the high-frequency parts are in the range of about 2.7 permeg to 5.4 permeg (one permeg equals $10^{-6}$) for $\delta^{15}$N and 0.22 K to 0.40 K for the reconstructed temperatures depending on the scenario, which is clearly visible in the decreasing width of the histograms (subplots (c) and (g) of Fig. 08, blue against grey).

The mismatches after the correction part of the reconstruction approach show clearly a further decrease of the misfits. This means that the width of the distributions of the pointwise mismatches $D_{\delta^{15}N,i}$ as well as $D_{T,i}$ is further reduced, and the distributions become more symmetric (long tales disappear; see histogram (c) and (g); red against blue of Fig. 08). The time series of the mismatches (subplots (d) and (h) of Fig.08) clearly illustrate that the correction approach mainly tackles the extreme deviations (sharp reduction of extreme values occurrence in the red distribution compared to the blue distribution) leading to a further improvement of about 40 % in $\delta^{15}$N and temperature. Finally, the 95 % quantiles ($2\sigma_{\delta^{15}N,corr,95}$, $2\sigma_{T,corr,95}$) of the remaining pointwise mismatches of $\delta^{15}$N and temperature ($D_{\delta^{15}N,i}$ or $D_{T,i}$) were calculated for the final solutions for all scenarios and are used as an estimate for the 2$\sigma$ uncertainty of the reconstruction algorithm (see Fig. 08c,g and Table 04). The final uncertainties (2$\sigma$) are in the order of 3.0 permeg to 6.3 permeg for $\delta^{15}$N and 0.23 K to 0.51 K for the surface temperature misfits. It is noteworthy that the measurement uncertainties (per point) of state of the art $\delta^{15}$N measurements are in the same order of magnitude, i.e. 3 permeg to 5 permeg (Kobashi et al., 2008b), highlighting the effectiveness of the presented fitting approach. Table 05 contains the final mismatches (2$\sigma$) in $\Delta$age between the synthetic target and the final modelled data after the correction step for all scenarios and shows that with a known accumulation rate and assumed perfect firm physics, it is possible to fit the $\Delta$age history in the Holocene with mean uncertainties better than 2 yr. In other words, the uncertainty in $\Delta$age reconstruction due to the inversion algorithm alone is in the order of 2 yr.
4. Discussion

4.1 Monte Carlo type input generator

Figure 09 shows the distribution of the cut-off-periods (COP) (I) and s values (II) used to create the improvements (Sect. 2.1, step 2) for all scenarios. The cut-off-periods are more or less evenly distributed, which shows that nearly the whole of the allowed frequency range (500 yr to 2000 yr) was used to create the improvements during the iterations. In contrast, the distributions of the s values show clearly that mostly small s values are used to create the improvements, which implies that iterations with small perturbations more likely lead to an improvement than larger ones.

Figure 06 reveals a weak point of the Monte-Carlo part, namely the absence of a suitable termination criterion for the optimization. The implementation until now is conducted such that the maximum number of iterations is given by the user or the iterations are terminated after a certain time (e.g. 15 h). Figure 06 shows that for nearly all scenarios it would be possible to stop the optimization after about 400 iterations, due to rather small additional improvements later on. This would decrease the time needed for the Monte-Carlo part to about 10 h (a single iteration needs about 90 s). Since the goal of the Monte-Carlo part is to find a temperature realisation that leads to an optimal middle passage through the $\delta^{15}$N target data, it would be possible to use the mean difference between the $\delta^{15}$N target and spline-filtered $\delta^{15}$N data using a certain cut-off-period as a termination criterion. This issue is under investigation at the moment. Another possibility to decrease the time needed for the Monte-Carlo part could be an increase in the numbers of CPUs used for the parallelization of the model runs. For this study an eight-core parallelization was used. A further increase in numbers of workers would improve the speed of the optimization.

4.2 High frequency step and final correction

To investigate the timing and contributions of the remaining mismatches in $\delta^{15}$N and temperature for scenario S1 after the high-frequency (step 3) and correction part (step 4), different cross-correlation experiments were conducted (see supplement S3). The experiments lead to equal results. The major fraction of the final mismatches of $\delta^{15}$N emerges from mismatches in the thermal-diffusion component $D_{\delta^{15}N_{\text{therm}}}$. Also a cancelation effect between the gravitational component $D_{\delta^{15}N_{\text{grav}}}$ and $D_{\delta^{15}N_{\text{therm}}}$ of the total mismatch in $\delta^{15}$N became obviously, affecting the calculation of $\text{lag}_{\text{max},\delta^{15}N}$ and $\text{lag}_{\text{min},\delta^{15}N}$ and most likely leading to a fundamental residual uncertainty in the low-permeg level for the corrected $\delta^{15}$N data. The same analyses were conducted for all synthetic scenarios, leading to similar results.

Additionally, the influence of the window length, used for the calculation of IF(t), on the correction was analysed, showing that for all investigated window lengths the correction reduces the mismatches of $\delta^{15}$N and temperature, whatever correction mode was used (calculated with $\text{xcf}_{\text{max}}$, $\text{xcf}_{\text{min}}$ or both quantities). Moreover, the correction is most efficient for window lengths in the range of 100 yr to 300 yr with an optimum at 200 yr for all cases.
4.3 Key points to be considered for the application to real data

Benefits of the novel gas isotope fitting approach

In addition to the fitting of $\delta^{15}$N data, the algorithm is able to fit $\delta^{40}$Ar and $\delta^{15}$N$_{\text{excess}}$ data as well using the same basic concepts (Fig. 10). Here the $\delta^{40}$Ar and $\delta^{15}$N$_{\text{excess}}$ data from Kobashi et al. (2008b) were used as the fitting targets. We reach final mismatches (2σ) of 4.0 permeg for $\delta^{40}$Ar/4 and 3.7 permeg for $\delta^{15}$N$_{\text{excess}}$, which are for both quantities below the analytical measurement uncertainty of 4.0 permeg to 9.0 permeg for $\delta^{40}$Ar/4 and 5.0 permeg to 9.8 permeg for $\delta^{15}$N$_{\text{excess}}$ measured data (Kobashi et al., 2008b).

The automated inversion of different gas-isotope quantities ($\delta^{15}$N, $\delta^{40}$Ar, $\delta^{15}$N$_{\text{excess}}$) provides a unique opportunity to study the differences in the gained solutions using different targets and to improve our knowledge about the uncertainties of gas-isotope-based temperature reconstructions using a single firn model. Next, the presented algorithm is not dependent on the firn model, which leads to the implication that the algorithm can be coupled to different firn models describing firn physics in different ways. Furthermore, an automated reconstruction algorithm avoiding manual manipulation and leading to reproducible solutions makes it possible for the first time, to study and learn from the differences between solutions matching different targets. Finally, differences obtained by applying different firn physics (densification equations, convective zone, etc.) but the very same inversion algorithm may help to assess firn model shortcomings, resulting in more robust uncertainty estimates than it was ever possible before.

In this publication we show the functionality and the basic concepts of the automated inversion algorithm using well known synthetic $\delta^{15}$N fitting targets. In this “perfect world scenario” the forward problem, converting surface temperature to $\delta^{15}$N, as well as the inverse problem, converting $\delta^{15}$N to surface temperature, is completely described by the used firn model. Consequently all sources of signal noise are ignored. For the later use of the algorithm on $\delta^{15}$N, $\delta^{40}$Ar or $\delta^{15}$N$_{\text{excess}}$ measured data this will not be the case anymore due to different sources of signal noise in the used measured data. As a result, differences between temperature solutions obtained from individual targets ($\delta^{15}$N, $\delta^{40}$Ar, $\delta^{15}$N$_{\text{excess}}$) will become obvious. These differences will allow to quantify the uncertainties associated with different unconstrained processes. Next, we will list and discuss potential sources of uncertainties and try to provide suggestions for their handling and quantification in our approach.

Measurement uncertainty and firn heterogeneity (cm-scale variability):

Many studies have investigated the influence of firn heterogeneity (or density fluctuations) on measurements of air compounds and quantities (e.g. $\delta^{15}$N, $\delta^{40}$Ar, CH$_4$, CO$_2$, O$_2$/N$_2$ ratio, air content) extracted from ice cores resulting in cm-scale variability and leading to additional noise on the measured data (e.g., Etheridge et al., 1992; Huber and Leuenberger, 2004; Fujita et al., 2009; Capron et al., 2010; Hörhold et al., 2011; Rhodes et al., 2013, 2016; Fourteau et al., 2017; Capron et al., 2010; Etheridge et al., 1992; Fourteau et al., 2017; Fujita et al., 2009; Hörhold et al., 2011; Huber and Leuenberger, 2004; Rhodes et al., 2013, 2016). Using discrete measurement technique instead of continuous sampling methods makes it difficult to quantify these effects. However, during discrete analyses of ice-core air-data it is common to measure replicates for given
depths, from which the measurement uncertainties of the gas-isotope data are calculated using pooled-standard-deviation (Hedges L. V., 1985). Often it is not possible to take real replicates (same depth) and instead the replicates are taken from nearby depths. Hence, any potential cm-scale variability is to some degree already included in the measurement uncertainty, because each measurement point represents the average over a few centimetres of ice. This is especially the case for low-accumulation sites or glacial ice samples for which the vertical length of a sample (e.g., 10-25 cm long for the glacial part of the NGRIP ice core, Kindler et al., 2014) covers the equivalent of 20 yr to 50 yr of ice at approximately 35 kyr b2k. Increasing the depth resolution of the samples would increase our knowledge of cm-scale variability, for e.g. identifying anomalous entrapped gas-layers that could have been rapidly isolated from the surface due to an overlying high-density layer (e.g., Rosen et al., 2014). As this variability is likely due to heterogeneity in the density profile, modelling such heterogeneities (if possible at all) may not help to better reconstruct a meaningful temperature history, but rather to reproduce the source of noise. This means that the potential cm-scale variability, in many cases, is already incorporated in the analytical noise obtained from gas-isotope measurements, due to analytical techniques themselves. Assuming the measurement uncertainty as Gaussian distributed, it is easy to incorporate this source of uncertainty in the inverse-modelling approach presented here. This will increase the uncertainty of the temperature according to eq. (3). The same equation can also be used for the calculation of the uncertainty in temperature related to measurement uncertainty in general.

To answer the pertinent question of how to better extract a meaningful temperature history from a noisy ice-core record, an excellent – but costly – solution is of course to use multiple ice cores. For example, a δ¹⁵N-based temperature reconstruction from the combination of data from the GISP2 ice core with the “sister ice core” GRIP drilled 30 kilometres apart is likely one of the best ways to overcome potential cm-scale variability. A comparison of ice cores that were drilled even closer might be even more advantageous.

**Smoothing effects due to gas diffusion and trapping:**

It is known that gas-diffusion and trapping processes in the firm can smooth out fast signals and result in a damping of the amplitudes of gas-isotope signals (e.g. Spahni et al., 2003; Grachev and Severinghaus, 2005). The duration of gas diffusion from the top of the diffusive column to the bottom where the air is closed off in bubbles is for Holocene conditions in Greenland approximately in the order of 10 yr (Schwander et al. 1997), whereas the data resolution of the synthetic targets was set to 20 yr to mimic the measurement data from Kobashi et al. (2008b) with a mean data resolution of about 17 yr (see Sect. 2.3: “Generating synthetic target data”). In the study of Kindler et al. (2014) it was shown that a glacial Greenland lock-in-depth leads to a damping of the δ¹⁵N signal of about 30 % for a 10 K temperature rise in 20 yr. We further assume that the smoothing according to the lock-in process is negligible for Greenland Holocene conditions according to the much smaller amplitude signals and shallower LID. Yet, for glacial conditions it requires attention.

**Accumulation rate uncertainties:**
For the synthetic data study presented in this paper it is assumed that the used accumulation-rate data are well known with zero uncertainty. This simplification is used to show the functionality and basic concepts of the presented fitting algorithm in every detail on well-known δ¹⁵N and temperature targets and to focus on the final uncertainties originating from the presented fitting algorithm itself. For the later reconstruction using measured gas-isotope data together with the published accumulation-rate scenarios shown in supplement S1 this will not be the case anymore. Uncertainties in layer-counting and corrections for ice thinning lead to a fundamental uncertainty. Especially in the early Holocene, this can easily exceed 10%. As the accumulation-rate data are used to run the firm model, all potential accumulation uncertainties are in part incorporated into the temperature reconstruction. On the other hand, as we discussed in supplement S2, the accumulation rate variability has a minor impact compared to the input temperature on the variability of δ¹⁵N data in the Holocene. The influence of these quantities, accumulation rate or temperature, on the temperature reconstruction is not equal; during the Holocene, accumulation-rate variability explains about 12% to 30% of δ¹⁵N variability. 30% corresponds to the 8.2 kyr event and 12% for the mean of the whole Holocene period including the 8.2 kyr event. Hence the influence of accumulation changes, excluding the extreme 8.2 kyr event, is generally below 10% during the Holocene. If the accumulation is assumed to be completely correct then the missing part will be assigned to temperature variations. Nevertheless for the fitting of the Holocene measurement-data we will use all three accumulation-rate scenarios as shown in S1. The difference in the reconstructed temperatures arising from the differences of these three scenarios will be used for the uncertainty calculation as well and is most likely higher than the uncertainty arising from uncertainties due to the process of producing the accumulation-rate data and from the conversion of the accumulation-rate data to the GICC05 timescale.

**Convective zone variability:**

Many studies have shown the existence of a non-diffusive well-mixed zone at the top of the diffusive-firm-column, called convective zone (CZ). The CZ is formed by strong katabatic winds and pressure gradients between the surface and the firm (e.g., Kawamura et al., 2006, 2013; Severinghaus et al., 2010). The existence of a CZ changes the gravitational background signal in δ¹⁵N and δ⁴₀Ar as it reduces the diffusive-column-height. The presented fitting algorithm was used together with the two most frequently used firm models for temperature reconstructions based on stable isotopes of air, the Schwander et al. (1997) model which has no CZ build in (or better a constant CZ of 0 m) and the Goujon firm model (Goujon et al., 2003) (which assumes a constant convective zone over time, that can easily be set in the code). (Goujon et al., 2003) (which assumes a constant convective zone over time, that can easily be set in the code). This difference between the two firm models only changes significantly the absolute temperature rather than the temperature anomalies as it was shown by other studies (e.g., Guillevic et al., 2013, Fig. 3). In the presented work, we show the results using the model from Schwander et al. (1997), because the differences between the obtained solutions using the two models are negligible besides a constant temperature offset. Also, noteworthy is that there is no firm model at the moment which uses a dynamically changing CZ. Indeed, this should be investigated but requires additional intense work. Additionally, the knowledge of the time-evolution of CZ changes for time periods of millennia to several hundreds of
millennia (in frequency and magnitude) is too poor to estimate the influence of this quantity on the reconstruction. In
principle it is possible to cancel-out the influence of a potentially changing CZ by using $\delta^{15}N_{\text{excess}}$ data for temperature
reconstruction, as due to the subtraction of $\delta^{40}\text{Ar}/4$ from $\delta^{15}N$ the gravitational term of the signals is eliminated. From that
point of view it will be interesting to compare temperature solutions gained from $\delta^{15}N_{\text{excess}}$ fitting with the solutions based on
$\delta^{15}N$ or $\delta^{40}\text{Ar}$ alone. This can offer a useful tool for quantifying the magnitude and frequency of CZ changes in the time
interval of interest.

It is known that for some very low accumulation-rate sites in areas with strong katabatic winds (e.g. “Megadunes”,
Antarctica) extremely deep CZs can occur, which are potentially able to smooth-out even decadal-scale temperature
variations (Severinghaus et al., 2010). For this its deepness would need to be of several dozens of meters, which is highly
unrealistic even for glacial Summit conditions (Guillevic et al., 2013, see discussion in Annex A4, p. 1042) as well as for the
rather stable Holocene period in Greenland for which no low accumulation and strong katabatic-wind situations are to be
expected.

4.4 Proof of concept for glacial data

For glacial conditions the task of reconstructing temperature (with correct frequency and magnitude) without using $\delta^{18}O_{\text{ice}}$
information is much more challenging due to the highly variable gas-age - ice-age differences ($\Delta\text{age}$) between stadial and
interstadial conditions. Here, contrary to the rather stable Holocene period, the $\Delta\text{age}$ can vary by several hundreds of years.
Also the accumulation-rate data are more uncertain than for the Holocene. To prove that the presented fitting algorithm also
works for glacial conditions we inverted the $\delta^{15}N$ data measured for the NGRIP ice core by Kindler et al. (2014) for two
Dansgaard-Oeschger events, namely DO6 and DO7. Since the magnitudes of those events are higher and the signals are
smoother than in the Holocene we only had to use the Monte-Carlo-type input-generator (see Sect. 2.3.2) for changing the
temperature inputs. To compare our results to the $\delta^{18}O_{\text{ice}}$ based and manually calibrated values from Kindler et al. (2014) we
use the ss09sea06bm time scale (NGRIP members: Andersen et al., 2004; Johnsen et al., 2001) as it was done in the Kindler
et al. publication. For the model spin-up we use the accumulation-rate and temperature data from Kindler et al. (2014) for the
time span 36.2 kyr to 60 kyr. The reconstruction window (containing DO6 and DO7) is set to 32 kyr to 36.2 kyr. As the first-
guess (starting point) of the reconstruction we use the accumulation-rate data ($\text{Acc}_{z,0}$) for NGRIP from the ss09sea06bm
time-scale together with a constant temperature of about -49 °C for this time window. As minimization-criterion $D_g$ for the
reconstruction we simply use the sum of the root-mean-squared-errors of the $\delta^{15}N$ and $\Delta\text{age}$ mismatches weighted with their
uncertainties ($\text{wRMSE}$) according to the following equation, instead of the mean $\delta^{15}N$ misfit alone as used for the Holocene
(eq. (1)).

$$D_{g_1} = \text{wRMSE}(\delta^{15}N) + \text{wRMSE}(\Delta\text{age})$$

$$= \frac{1}{N} \sum_{i=1}^{N} \left[ \frac{\delta^{15}N_{\text{meas},i} - \delta^{15}N_{\text{mod},i}}{\epsilon_{\delta^{15}N,i}} \right]^2 + \frac{1}{M} \sum_{k=1}^{M} \left[ \frac{\Delta\text{age}_{\text{meas},k} - \Delta\text{age}_{\text{mod},k}}{\epsilon_{\Delta\text{age},k}} \right]^2$$

$$18$$
Here εδ^{15}N_i and εΔage_k are the uncertainties in δ^{15}N and Δage for the measured values i or k (Δage match points: Guillevic, M. (2013), p.65, Tab. 3.2) and N, M the number of measurement values. We set εδ^{15}N = 20 permeg for all i (Kindler et al., 2014) and εΔage_k = 50 yr for all k. The relative uncertainties in Δage can easily reach up to 50 % and more in the Glacial using the ss09sea06bm time-scale which results in a pre-eminence of the δ^{15}N misfits over the Δage misfits (10 % to 20 % when using GICC05 time-scale; Guillevic, M., 2013, p. 65 Tab. 3.2). Due to this issue we have to set Δage uncertainties to 50 yr to make both terms equally important for the fitting algorithm. In Fig. 11 we show preliminary results. The δ^{15}N and Δage fitting (a, b) and the resulting gained temperature and accumulation-rate solutions (c, d) using the presented algorithm are completely independent from δ^{18}O_ice which provides the opportunity to evaluate the δ^{18}O_ice-based reconstructions. In this study the algorithm was used in three steps. First, starting with the first-guess (constant temperature), the temperature was changed as explained before. The accumulation rate was changed in parallel to the temperature allowing a random offset shift (up and down) together with a stretching or compressing (in y direction) of the accumulation-rate signal over the whole time-window (32 kyr to 36.2 kyr). This first step leads to the “Monte-Carlo Solution 0” (MCS0) which provides a first approximation and is the base for the next step. For the next step, we fixed the accumulation rate and let the algorithm only change the temperature to improve the δ^{15}N-fit (MSC1). Finally, we allow the algorithm to change the temperature together with the accumulation rate using the Monte-Carlo-type input-generator for both quantities. This allows to change the shape of the accumulation-rate data. This final step can be seen as a fine tuning of the gained solutions from the steps before. The obtained mismatches in δ^{15}N and Δage of all steps are at least of the same quality or better than the δ^{18}O_ice-based manual method from Kindler et al. (2014) (see Table 06). The gained temperature solutions show a very good agreement in timing and magnitude compared to the reconstruction of Kindler et al. (2014). Also the accumulation-rate solutions show that the accumulation has to be reduced significantly compared to the ss09sea06bm data to allow a suitable fit of the δ^{15}N and Δage target data, a result highly similar to Guillevic et al. (2013) and Kindler et al. (2014). The mismatches in δ^{15}N and Δage of the final MCS FIN solution show a 15 % smaller misfit of δ^{15}N (2σ) and an about 31 % smaller misfit of Δage (2σ) compared to the Kindler et al. (2014) solution. Keeping in mind that the used approach is completely independent from δ^{18}O_ice strengthens the functionality and quality of the presented gas-isotope fitting approach also for glacial reconstructions.

As this section contains a proof-of-concept of the presented automated gas-isotope fitting algorithm on glacial data, preliminary results and ongoing work were shown here. Furthermore as the presented fitting algorithm was developed and tested in first order for Holocene-like data, it is highly probable that the functionality of the algorithm using glacial data will be further extended and adjusted in future studies.

5. Conclusion

A novel approach is introduced and described for inverting a firn-densification and heat-diffusion model to fit small gas-isotope-data variations as observed throughout the Holocene. From this new fitting method, it is possible to extract the surface-temperature history that drives the firn status which in turn leads to the gas-isotope time-series. The approach is a
combination of a Monte-Carlo-based iterative method and the analysis of remaining mismatches between modelled and target data. The procedure works fully automated and provides a high potential for parallel computing for time consumption optimization. Additional sensitivity experiments have shown that accumulation-rate changes have only a minor influence on short-term variations of δ15N, which themselves are mainly driven by high-frequency temperature variations. To evaluate the performances of the presented approach, eight different synthetic δ15N time-series were created from eight known temperature histories. The fitting approach leads to an excellent agreement in timing and amplitudes between the modelled and synthetic δ15N and temperature data. The obtained final mismatches follow a symmetric standard-distribution-function. 95% of the mismatches compared to the synthetic data are in an envelope between 3.0 permeg to 6.3 permeg for δ15N and 0.23 K to 0.51 K for temperature, depending on the synthetic temperature scenarios. These values can therefore be used as a 2σ estimate for the reconstruction uncertainty arising from the presented fitting algorithm itself. For δ15N the obtained final uncertainties are in the same order of magnitude as state of the art measurement uncertainty. The presented reconstruction approach was also successfully applied to δ40Ar and δ15Nexcess measured data. Moreover, we have shown that the presented fitting approach can also be applied to glacial temperature reconstructions with minor algorithm modifications. Based on the demonstrated flexibility of our inversion methodology, it is reasonable to adapt this approach for reconstructions of other non-linear physical processes.

**Data availability**

The synthetic δ15N and temperature targets, the reconstructed δ15N and temperature data, and the used accumulation rates can be found in the data supplement of this paper. The GISP2 δ18Oice data used in this study for calculating the temperature spin-up can be found in Grootes and Stuiver (1999). The source code for the inversion algorithm and additional auxiliary data are available upon request.

**Competing interests**

The authors declare that they have no competing financial interests.

**Acknowledgements**

This work was supported by the SNF grants (SNF-159563 and SNF-172550). We would like to thank Drs. Takuro Kobashi, Philippe Kindler and Myriam Guillevic for helpful discussions about the ice core data and model peculiarities.
References


Guillèvre, M.: Characterisation of rapid climate changes through isotope analyses of ice and entrapped air in the NEEM ice core, PhD thesis, University of Copenhagen, Université de Versailles Saint Quentin en Yvelines., 2013.
Guilevic, 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
$\delta^{18}O_{\text{ice}}$ in Greenland over a series of Dansgaard-Oeschger events, Clim. Past, 9(3), 1029–1051,


Holmgren, K., Gogou, A., Izdebski, A., Luterbacher, J., Sicre, M. A. and Xoplaki, E.: Mediterranean Holocene climate,


Huber, C. and Leuenberger, M.: Measurements of isotope and elemental ratios of air from polar ice with a new on-line


